

Active Matter

Editors: Giovanni Volpe, Nuno Araújo, Giorgio Volpe, Agnese Callegari

March 4, 2025

Contents

1	Introduction	1
GIOVANNI VOLPE, NUNO ARAÚJO, AGNESE CALLEGARI, GIORGIO VOLPE		
1.1	What does it mean to be far from equilibrium?	2
1.2	The importance of dissipation in active matter	3
1.3	Overview of the Book	4
1.3.1	Part I: Theory	4
1.3.2	Part II: Simulation	6
1.3.3	Part III: Experiments	7
1.3.4	ActiveMatter Game	8
Bibliography		8
I	Theory	9
2	What is Soft Matter?	10
MARGARIDA MARIA TELO DA GAMA, JOSÉ MARIA TAVARES, MYKOLA TASINKEVYCH		
2.1	Condensed matter	11
2.1.1	Phases of matter	11
2.1.2	Thermal fluctuations and criticality	12
2.1.3	Broken symmetry and phase transitions	13
2.2	Soft condensed matter	13
2.2.1	Building blocks and interactions	14
2.2.2	Some examples	15
2.3	Basic theoretical tools	17
2.3.1	Landau theory for the ferromagnetic–paramagnetic transition	17
2.3.2	Landau-de Gennes theory for the nematic-isotropic transition	20
2.4	Further readings	23
Bibliography		23
3	Nonequilibrium Physics of Small Systems	24
FELIX RITORT		
3.1	Nonequilibrium phenomena	25
3.2	Small systems and single-molecule experiments	25
3.3	Nonequilibrium force spectroscopy	27
3.4	Fluctuation theorems	29
3.5	Energy, information and the Maxwell demon	31
3.6	Outlook	34
3.7	Problems	36
Bibliography		37
4	From Passive to Active Brownian Motion	44

DEMIAN LEVIS, NUNO A. M. ARAÚJO, IGNACIO PAGONABARRAGA	
4.1 Brownian motion: degrees of freedom, length and time scales	45
4.2 Langevin approach	47
4.3 Fokker–Planck equation	49
4.4 Active Brownian motion	50
4.4.1 Active Brownian particles	50
4.4.2 Underdamped active Brownian particles	56
4.4.3 Run and tumble motion	57
4.4.4 Active Ornstein–Uhlenbeck particles	57
4.4.5 Active particles with an energy depot	58
4.4.6 Monte Carlo and lattice gas models of self-propelled particles	58
4.5 Derivation of continuum descriptions	59
4.5.1 Linear stability analysis	62
Bibliography	64
5 Stochastic Differential Equations and Active Matter	70
JAN WEHR	
5.1 Stochastic differential equations	71
5.1.1 Itô integral	71
5.1.2 The Itô formula	73
5.1.3 The Ornstein–Uhlenbeck equation	75
5.2 Stratonovich integral	76
5.3 General SDE systems	77
5.4 Small mass limit	79
5.4.1 Noise-induced drift	79
5.4.2 Generalizations	80
5.5 Motion of light-sensitive robots with sensorial delay	80
Bibliography	83
6 Anomalous Diffusion and Ergodicity Breaking	84
CARLO MANZO, GORKA MUÑOZ-GIL	
6.1 Introduction	85
6.2 Anomalous diffusion and ergodicity breaking	87
6.3 Anomalous diffusion models	88
6.3.1 Fractional Brownian motion	89
6.3.2 Continuous time random walk	90
6.4 Estimators	90
6.4.1 Anomalous diffusion exponent	90
6.4.2 Ergodicity breaking	91
6.5 Machine learning	92
6.5.1 Convolutional neural network for trajectory analysis	92
6.5.2 Application of a CNN for trajectory analysis	93
6.6 Problems	94
Bibliography	94
7 Low Reynolds Number Hydrodynamics	97
HOLGER STARK	
7.1 Stokes equations and low Reynolds numbers	98
7.2 Oseen tensor	99
7.3 Friction coefficients	100
7.3.1 Stokes friction	101
7.3.2 Friction of a particle with arbitrary shape	102
7.3.3 Friction of a slender rod	103
7.4 Resistive force theory	104

7.5	Faxén theorem	104
7.6	Hydrodynamic interactions	105
7.6.1	Point particles	105
7.6.2	Rotne-Prager approximation	106
7.6.3	Application to microswimmers	107
7.7	Hydrodynamic singularities (multipoles)	107
7.7.1	Force dipoles, pushers, and pullers	107
7.7.2	Fundamental singular solutions	108
7.8	Lorentz reciprocal theorem	110
7.9	Blake tensor	111
7.10	Time-dependent Stokes equations	111
7.11	Problems	112
	Bibliography	112
8	Phoretic and Self-phoretic Transport in Colloidal Science	114
	IVO BUTTINONI	
8.1	Fluid transport at small length scales	114
8.2	Interfacial transport	116
8.2.1	Transport in electric fields	116
8.2.2	Transport in chemical gradients	119
8.2.3	Transport in temperature gradients	121
8.3	Self-phoretic colloids	122
8.3.1	From phoresis to autophoresis	123
8.3.2	Self-electrophoretic colloids	123
8.3.3	Self-phoretic transport in AC fields	124
8.3.4	Self-diffusiophoretic colloids	125
8.3.5	Self-thermophoretic colloids	127
8.4	Experiments with active Brownian motion	128
8.4.1	Methods	128
8.4.2	MSD analysis	130
8.4.3	Experimental data analysis	131
	Bibliography	133
9	Dynamical Density Functional Theory	137
	HARTMUT LÖWEN	
9.1	Density functional theory in equilibrium	138
9.1.1	Basics	138
9.1.2	Approximations for the density functional	138
9.2	Classical dynamical density functional theory	139
9.2.1	Brownian dynamics and Smoluchowski equation	139
9.2.2	Derivation	140
9.3	Polar particles	141
9.3.1	Density functional theory	141
9.3.2	Dynamic density functional theory	142
9.4	Dynamical density functional theory for dry active matter	142
9.5	Dynamical density functional theory for wet active matter	143
9.5.1	The swimmer model	143
9.5.2	Derivation	145
9.5.3	Applications	146
9.6	Final remarks	147
9.7	Problems	148
	Bibliography	148
10	Motility-Induced Phase Separation	152

JOAKIM STENHAMMAR

10.1	Pattern formation in active particles with spatially varying speeds	153
10.2	MIPS in quorum-sensing active particles	154
10.3	Microscopic particle models for MIPS	155
10.4	MIPS in active Brownian particles	157
10.4.1	Kinetic model of MIPS in ABPs	158
10.4.2	Similarities between MIPS and equilibrium phase coexistence	159
10.5	Mechanical pressure in ABPs	160
10.5.1	Pressure in non-interacting ABPs	160
10.5.2	Pressure of interacting ABPs	161
10.5.3	Pressure and phase coexistence in ABPs	163
10.6	Current research directions	165
	Bibliography	166

11 Liquid Crystals 170

GARETH P. ALEXANDER

11.1	Landau–de Gennes theory of the isotropic–nematic transition	171
11.1.1	Director field	171
11.2	Director gradients and Frank free energy	172
11.2.1	Frank free energy	173
11.2.2	Electric fields and Frederiks transition	174
11.3	Topological defects	174
11.3.1	Defects in two dimensions	174
11.3.2	Disclination lines and defect loops in three dimensions	175
11.4	Hydrodynamics of nematic liquid crystals	176
11.4.1	Active nematics	178
11.5	Layered phases including smectics and cholesterics	178
11.5.1	Undulational instability and active layered phases	180
11.6	Final remarks	180
11.7	Problems	181
	Bibliography	182

12 Active Nematics 186

JULIA M YEOMANS

12.1	Active nematohydrodynamics	187
12.1.1	Equations of motion	187
12.1.2	The active stress	188
12.1.3	Active instabilities	189
12.2	Active turbulence	189
12.2.1	Motile topological defects	190
12.2.2	Activity-driven active turbulence	191
12.2.3	Experimental systems	192
12.3	Confinement	193
12.3.1	Friction	194
12.4	Three dimensions	194
12.5	Final remarks	196
	Bibliography	197

13 Activity in the Cytoskeleton 200

FRANCOIS NEDELEC	
13.1 Introduction	201
13.2 Kinesins and other molecular motors	202
13.2.1 Kinesin movement mechanism	203
13.2.2 Kinesin's forward and backward steps	203
13.3 Bell's law and Kramers' theory of transitions under load	203
13.4 Simulation of stochastic events with varying rates	206
13.5 Final remarks and examples	208
Bibliography	211
14 From Cells to Tissues	213
ROBERTO CERBINO, FABIO GIAVAZZI	
14.1 Epithelial tissues as active networks	214
14.2 Geometrical properties and mechanics of single cells	215
14.2.1 Fluid drop model	215
14.2.2 Effective surface tension	215
14.2.3 Cell-cell adhesion	216
14.2.4 Traction forces and cell motility	218
14.3 Geometrical properties and mechanics of tissues	219
14.3.1 Cell tissues as active networks	219
14.3.2 The energy of a monolayer	220
14.3.3 Ground states of a monolayer	221
14.3.4 The effect of cell division	221
14.3.5 Rigidity transition	222
14.3.6 Adding cell motility	225
14.3.7 Adding cell-cell alignment	225
14.4 Final Remarks	229
14.5 Problems	229
Bibliography	229
II Simulation	232
15 Numerical Models for Active Matter	233
LUKAS HECHT, JOSÉ CARLOS UREÑA MARCOS, BENNO LIEBCHEN	
15.1 The effect of the solvent	234
15.2 Dry active particles	235
15.2.1 Active Brownian particle	235
15.2.2 Run-and-tumble model	240
15.2.3 Active Ornstein-Uhlenbeck model	240
15.2.4 Chiral particle model	241
15.2.5 Models with alignment interactions	241
15.2.6 Advantages and limitations	241
15.3 Dry continuum models	242
15.4 Wet active particles	247
15.5 Hydrodynamic interactions	249
15.5.1 Minimal models	249
15.5.2 Squirmers models	251
15.5.3 Explicit simulation of the solvent	252
15.6 Wet continuum models	253
Bibliography	253
16 Molecular Dynamics	262

AGNESE CALLEGARI, GIORGIO VOLPE, GIOVANNI VOLPE	
16.1 Basic algorithms	263
16.1.1 Euler algorithm	264
16.1.2 Leapfrog algorithm	266
16.1.3 Velocity Verlet algorithm	268
16.2 Interacting particles	269
16.2.1 Hard-core potential	269
16.2.2 Yukawa-like potential	269
16.2.3 Lennard-Jones potential	270
16.2.4 Truncated Lennard-Jones potential	270
16.3 Controlling the simulation	273
16.3.1 Thermostats	274
16.3.2 Barostats	276
16.4 Boundary conditions	278
16.4.1 Reflecting boundary condition	278
16.4.2 Periodic boundary condition	279
16.4.3 Adsorbing boundary condition	280
16.5 Problems	281
Bibliography	283
17 Brownian Dynamics	285
AYKUT ARGUN, AGNESE CALLEGARI, GIORGIO VOLPE, GIOVANNI VOLPE	
17.1 Random walks	286
17.2 White noise	288
17.3 Brownian motion	289
17.4 Active Brownian motion	291
17.4.1 Active Brownian particles	291
17.4.2 Chiral active Brownian particles	292
17.4.3 Run-and-tumble particles	293
17.4.4 Active Ornstein-Uhlenbeck particles	294
17.4.5 Non-spherical active particles	295
17.5 Complex and crowded environments	296
17.5.1 External force fields	296
17.5.2 Obstacles	297
17.5.3 Interacting particles	299
17.6 Multiplicative noise	304
17.6.1 Spurious drift	304
17.6.2 At thermodynamic equilibrium	305
17.6.3 Out of equilibrium	306
Bibliography	310
18 Accelerating Molecular Dynamics	313
JOOST DE GRAAF	
18.1 Relevant scales	314
18.1.1 Other approaches for solving hydrodynamic equations	315
18.2 Neighbor lists	315
18.2.1 Verlet list	317
18.2.2 Cell list	317
18.2.3 Linked list	319
18.3 Event-driven molecular dynamics	320
18.4 Problems	322
Bibliography	322
19 Implementation with LAMMPS	324

CRISTOVÃO S. DIAS

19.1 Basic algorithms	325
19.1.1 Velocity Verlet algorithm	325
19.1.2 Force computation	325
19.1.3 Parallel computing	325
19.1.4 Langevin dynamics	326
19.1.5 Reduced Units	327
19.2 Hands-on practice	328
19.2.1 Compiling LAMMPS	328
19.2.2 Running LAMMPS	328
19.2.3 Brownian motion	330
19.3 Customize LAMMPS	332
19.4 Final Discussion	335
19.5 Problems	335
Bibliography	336
20 Interaction Rules in Models of Active Matter	338
DAVID M. SUSSMAN	
20.1 Agent-based numerical modeling	339
20.2 Won't you be my neighbor?	340
20.3 Metric and topological flocking models	340
20.4 Voronoi neighbor lists	341
20.5 Geometric models of dense tissue	343
20.5.1 Vertex model forces	344
20.5.2 Voronoi model forces	344
20.6 Discussion and outlook	345
20.7 Problems	346
Bibliography	346
21 Mesoscopic Modelling of Epithelial Tissues	349
KEVIN HÖLLRING, ANA-SUNČANA SMITH	
21.1 Modeling tissues at various levels of granularity	349
21.1.1 Cellular Potts models	350
21.1.2 Phase field models	350
21.1.3 Active network/vertex/Voronoi models	350
21.1.4 Particle models	351
21.1.5 Continuum models	351
21.1.6 Bridging scales with adaptive numerical simulations	351
21.2 Vertex model of tissues	351
21.2.1 The physical foundation of the vertex model as an energy functional	352
21.2.2 Tissue boundary conditions and system size	353
21.3 Implementing the vertex model relaxation procedure	353
21.3.1 Data structures for simulation	353
21.3.2 Outline of the relaxation procedure	354
21.3.3 The individual steps of the vertex model relaxation procedure	355
21.3.4 Energy calculation	356
21.3.5 Equilibration and gradient descent step	356
21.3.6 T1-transition step	357
21.3.7 Check for convergence	360
21.3.8 Storage of tissue state	360
21.3.9 Technical and analytical limitations	360
21.3.10 Possible tweaks and optimizations	361
21.4 Final remarks	361
21.5 Problems	362

Bibliography	363
22 Active Fluids	365
LUCA BIANCOFIORE	
22.1 Equations for an active fluid	365
22.1.1 Motion of a Janus particle in a fluid	366
22.1.2 Continuum model for active suspensions	366
22.1.3 Active suspension flow using generalized Navier–Stokes equations	367
22.2 Navier–Stokes equations	367
22.2.1 Non-dimensionalization	367
22.3 Discretization of transport equation via the finite volume method	368
22.3.1 Finite volume method	368
22.4 Discretization of Navier–Stokes	373
22.4.1 Pressure gradient	373
22.4.2 Discretization of the Navier–Stokes momentum equation	374
22.4.3 Discretization of the continuity equation	375
22.4.4 SIMPLE algorithm	375
22.5 Properties of numerical schemes	376
22.5.1 Accuracy	376
22.5.2 Stability	377
22.5.3 Boundedness	377
22.5.4 Conservativeness	377
22.5.5 Transportiveness	378
22.6 Problems	378
Bibliography	380
23 Multiparticle Collision Dynamics	382
MARISOL RIPOLL	
23.1 Multiparticle collision dynamics method	383
23.1.1 MPC algorithm	384
23.1.2 Capillary flow	386
23.2 Solvent properties	387
23.2.1 Viscosity	387
23.2.2 Diffusion coefficient	388
23.2.3 Dimensionless numbers	388
23.2.4 Equation of state	389
23.3 MPC alternative implementations	389
23.3.1 MPC with Andersen thermostat	389
23.3.2 MPC with angular momentum conservation	390
23.3.3 Non-ideal MPC	390
23.4 Coupling with particles and obstacles	390
23.5 Random MPC solvent	392
23.6 Active matter examples	393
23.6.1 Flagellated swimmers	393
23.6.2 Phoretic swimmers	394
23.6.3 Squirmers	394
23.7 Concluding remarks	394
23.8 Problems	395
Bibliography	395
24 Lattice Boltzmann Method	401

RODRIGO COELHO, NUNO ARAÚJO, IGNACIO PAGONABARRAGA	
24.1 Fundamental concepts	403
24.1.1 Distribution function	403
24.1.2 Boltzmann equation	404
24.1.3 Collective dynamics	404
24.1.4 Gauss–Hermite quadrature	406
24.2 Numerical implementation	407
24.2.1 Discretization of the phase space	407
24.2.2 Algorithm and code hints	408
24.2.3 Boundary conditions	409
24.2.4 Applied forces	410
24.3 Moving particles in a fluid	411
24.4 Problems	411
Bibliography	413
25 Lattice Boltzmann for Active Fluids	417
ZIGA KOS AND MIHA RAVNIK	
25.1 Lattice Boltzmann in active liquid crystal continuum models	418
25.1.1 Active nematics	418
25.1.2 Active polar fluids	418
25.1.3 Active (nematic) binary fluids	418
25.2 Lattice Boltzmann approaches	418
25.2.1 Numerical implementation	419
25.2.2 LUDWIG code	420
25.3 Hybrid lattice Boltzmann approaches	420
25.3.1 Lattice Boltzmann time step	420
25.3.2 Finite difference time step	423
25.3.3 Separation of scales	423
25.3.4 Numerical implementation	425
25.3.5 Computational requirements	426
25.3.6 Visualization of fields	426
25.4 Discussion, overview, and challenges	427
25.5 Problems	428
Bibliography	428
26 Phase Field Models of Active Systems	434
ROMAIN MUELLER, AMIN DOOSTMOHAMMADI	
26.1 Phase field description of diffuse interfaces	435
26.1.1 Free-energy functional	435
26.1.2 Hydrodynamics	437
26.2 Biphase active systems	438
26.2.1 Active nematic interfaces	438
26.2.2 Active matter in a viscoelastic medium	440
26.2.3 Self-deforming active surfaces	441
26.3 Cell monolayers	443
26.3.1 Cells as active deformable droplets	443
26.3.2 Minimal model of cell monolayers	444
26.3.3 Numerical implementation	445
26.3.4 Active polar driving	446
26.3.5 Active nematic driving	447
26.4 Summary	450
26.5 Problems	450
Bibliography	451

III Experiments	457
27 Chemistry of Micromotors	458
JULIANE SIMMCHEN	
27.1 Basic physical chemistry of micromotors	459
27.1.1 Single-component phase transitions	459
27.1.2 Binary mixtures	460
27.1.3 Chemical reactions	461
27.2 Chemical reactions driving active motion	462
27.2.1 Catalytic micromotors	462
27.2.2 Redox reactions	466
27.2.3 Polymerization reactions	468
27.3 Energy conversion efficiency	468
27.4 A note on nanomotors and enzymes	470
27.5 Problems	471
Bibliography	471
28 Experiments with Artificial Active Particles	475
CHUN-JEN CHEN, FÉLIX GINOT	
28.1 Janus particles as active particles	476
28.1.1 Fabrication of Janus particles	476
28.1.2 Propulsion mechanisms of Janus particles	477
28.2 Janus particles in critical mixtures	478
28.2.1 Simple active motion	479
28.2.2 Active alignment and phototaxis	480
28.2.3 Individual control	481
28.3 Image processing and active particle detection	482
28.3.1 Position detection	482
28.3.2 Orientation detection	483
28.4 Introduction to the motion analysis of active particles	483
28.4.1 Trajectory and velocity	484
28.4.2 Velocity and orientation autocorrelation	485
28.4.3 Mean square displacement	485
28.4.4 Pair correlation function	486
28.4.5 Polar and rotational order parameters	487
28.5 Problems	487
Bibliography	488
29 Active Particles with Feedback Interactions	491
FRANK CICHOS	
29.1 Why feedback control of active particles	492
29.2 Switchable thermophoretic swimmers	493
29.2.1 Thermo-osmosis vs thermophoresis	493
29.2.2 Self-thermophoretic swimmer	494
29.3 Feedback control	495
29.3.1 Photon nudging	495
29.3.2 Experimental setup	496
29.3.3 Photon-nudging theory	498
29.3.4 Advanced control with symmetric active particles	502
29.4 Applications of feedback control	503
29.4.1 Active particle pressure and activity landscapes	503
29.4.2 Information-controlled self-assembly	506
29.4.3 Machine learning with microswimmers	508
Bibliography	510

30 Active Colloids at Fluid Interfaces	514
LUCIO ISA	
30.1 Why active colloids at fluid interfaces	515
30.2 Particle adsorption at fluid interfaces	515
30.3 Contact angle measurement	520
30.4 Measurement of interfacial tension	522
30.5 Self-propulsion at fluid interfaces	524
30.5.1 Catalytic particles at fluid interfaces	524
30.5.2 Marangoni propulsion	526
30.6 Problems	529
Bibliography	529
31 Optical Tweezers	533
DAVID BRONTE CIRIZA, MARIA G. DONATO, ONOFRIO M. MARAGÒ	
31.1 Theory of optical trapping	534
31.1.1 Geometrical optics	534
31.1.2 Rayleigh approximation	535
31.1.3 Electromagnetic theory	536
31.1.4 Enhanced simulations using deep learning	537
31.2 Experimental setups	538
31.2.1 Microscope	538
31.2.2 Laser	538
31.2.3 Particle tracking	539
31.3 Calibration	540
31.3.1 Power spectrum analysis	540
31.3.2 Autocorrelation function analysis	542
31.3.3 FORMA	542
31.4 Applications	543
31.4.1 Mechanical properties of red blood cells	543
31.4.2 Micromachines	544
31.4.3 Stochastic thermodynamics	544
31.5 Problems	544
Bibliography	545
32 Atomic Force Microscopy	551
SHIVAPRAKASH N. RAMAKRISHNA, LUCIO ISA	
32.1 Surface analysis in active matter systems	552
32.1.1 A brief history of scanning probe microscopy (SPM)	552
32.2 Atomic force microscopy (AFM)	553
32.2.1 Forces between the tip and the sample surface	554
32.2.2 Operational modes of the AFM	556
32.3 Force spectroscopy in AFM	558
32.3.1 Nanoindentation	559
32.3.2 Force-distance curve-based imaging	561
32.4 Surface roughness	561
32.4.1 Measuring roughness	562
32.5 Applications of AFM to active matter	564
32.6 Outlook	566
32.7 Problems	571
Bibliography	572
33 Spectroscopy	577

PIETRO GIUSEPPE GUCCIARDI, ANTONINO FOTI	
33.1 IR and Raman spectroscopy	578
33.1.1 Molecular vibrations	578
33.1.2 Vibrational transitions and spectroscopies	579
33.2 Surface enhanced Raman spectroscopy	582
33.2.1 Electromagnetic enhancement	582
33.2.2 Chemical enhancement	584
33.3 Raman tweezers	584
33.3.1 Applications of Raman tweezers for Raman and SERS	586
33.4 Tip-enhanced Raman spectroscopy	590
33.4.1 Field propagation, diffraction, and spatial resolution	590
33.4.2 Field enhancement and spatial confinement in TERS	592
33.4.3 Experimental configurations for TERS	594
33.4.4 TERS applications	597
Bibliography	599
34 Experimental Active Nematics	605
BERTA MARTÍNEZ-PRAT, JORDI IGNÉS-MULLOL, FRANCESC SAGUÉS	
34.1 Preparation of a 2D active nematic	606
34.1.1 Kinesin expression	606
34.1.2 Microtubules polymerization	606
34.1.3 Active gel preparation	607
34.1.4 Experimental setup for the assembly of the 2D active nematic	610
34.1.5 Glass-slides functionalization	611
34.1.6 Active nematics for particle image velocimetry	611
34.2 Imaging and data acquisition	612
34.2.1 Particle image velocimetry	612
34.2.2 Director determination	614
34.3 Data analysis	615
34.3.1 Filtering of the nematic director field	615
34.3.2 Defect location and classification. Winding number	616
34.3.3 Analysis of patterns in active nematics	617
34.3.4 Spectral analysis	617
34.3.5 Location of vortices	618
34.3.6 Correlation functions	619
34.4 Conclusions	620
34.5 Problems	620
Bibliography	621
35 Planktonic Active Matter	622
ANUPAM SENGUPTA	
35.1 Overview of planktonic active matter	623
35.2 Physical ecology of plankton	624
35.2.1 Morphology and organelles of phytoplankton	626
35.2.2 Encounter rates and kernels	627
35.3 Gravitaxis in planktonic active matter	629
35.3.1 Gravity-sensing mechanisms	630
35.3.2 Biomechanics of gravitaxis	630
35.3.3 Quantifying cell morphology in experiments	635
35.3.4 Influence of intracellular organelles on gravitactic stability	637
35.4 Gravitaxis-related phenomena	638
35.4.1 Gyrotaxis: An interplay of gravitactic and viscous torques	638
35.4.2 Phytoplankton swimming in turbulent eddies	640
35.4.3 Bioconvection	642

35.5 Outlook	644
35.6 Problems	646
Bibliography	647
36 Microfluidics	654
CAROLINE BECK ADIELS, REZA MAHDAVI	
36.1 Advantages and limitations of microfluidics	655
36.2 Fluid Mechanics	655
36.2.1 Navier-Stokes equations and Reynolds number	655
36.2.2 Concentration distribution	656
36.2.3 Simulations using COMSOL Multiphysics	657
36.3 Materials and fabrication methods	657
36.3.1 PDMS fabrication procedure	658
36.3.2 Handling of fluids	660
36.3.3 Detection	661
36.4 Applications	662
36.4.1 Single-cell studies	662
36.4.2 Coordinated cell behavior	664
36.5 Problems	664
Bibliography	666
37 3D Printing	669
STEPHEN EBBENS, PIYUSH KUMAR, DAVID A. GREGORY, PATRICK SMITH, XIUBO ZHAO	
37.1 Introduction to 3D printing	670
37.2 3D inkjet printing	671
37.2.1 Actuation methods	671
37.2.2 Continuous and drop-on-demand inkjet printing	671
37.2.3 Reactive 3D ink jet printing	673
37.2.4 Printability	674
37.2.5 Droplet spacing and coverage	674
37.3 Printing active rockets and stirrers	675
37.3.1 The printing equipment and controllers	676
37.3.2 Computer aided designs of fibroin structures	677
37.3.3 Printing details	677
37.3.4 Device performance	679
37.4 Additional examples of printed active matter	681
37.5 Future directions and conclusions	682
37.6 Problems	682
Bibliography	682
38 Active Mechanical Metamaterials	684
CORENTIN COULAIS	
38.1 Mathematical description of active mechanical metamaterials	685
38.2 Examples of active mechanical metamaterials	686
38.2.1 One-dimensional odd chain	686
38.2.2 One dimensional chain with time-modulated stiffness	687
38.2.3 Metamaterial with time-modulated rest-length that crawls	688
38.2.4 Odd elastic lattice	688
38.3 Properties of active mechanical metamaterials	691
38.3.1 Spectral analysis	691
38.3.2 Continuum limit	695
38.4 Final Remarks	696
38.4.1 Difference between active mechanical metamaterials and active matter	696
38.4.2 Blurring active mechanical metamaterials and active matter	696

Bibliography	700
39 Living Robotics	702
MARIA GUIX, RAFAEL MESTRE	
39.1 Advancements and challenges in bio-hybrid robotic systems	703
39.2 Key design features	703
39.2.1 Design versus actuation mode	706
39.2.2 Motion control	709
39.3 Actuation by electrical pulse stimulation	710
39.4 Force measurement	712
39.5 Unique capabilities and main challenges	714
39.6 Ethical concerns	715
39.7 Conclusions	717
39.8 Problems	718
Bibliography	718
40 Locomotor Transitions in Complex Terrain	724
RATAN OTHAYOTH, QIYUAN FU, SEAN W. GART, QIHAN XUAN, YAQING WANG, CHEN LI	
40.1 The importance of locomotor transitions	725
40.2 Legged locomotion	726
40.2.1 Experimental methods	726
40.2.2 Modelling approaches	728
40.2.3 Insights on legged locomotion	728
40.2.4 Towards multi-pathway locomotor transitions	732
40.3 Limbless locomotion	732
40.3.1 Experimental tools	733
40.3.2 Insights from snake traversing large steps	733
40.4 Common features of legged and limbless locomotion	737
40.5 Outlook	737
40.6 Problems	737
Bibliography	737
41 Vibrational Robots in Complex Environments	742
ALESSANDRO MAGAZZÙ	
41.1 Macroscopic active agents in complex environments	743
41.2 Setting a complex environment for macro agents	743
41.2.1 Building the arena	743
41.2.2 Hexbugs as active agents	744
41.2.3 Setting the passive particles	745
41.2.4 The right boundary frame	746
41.3 Hexbugs moving in the arena	746
41.4 What information can we extract from the Hexbug trajectories?	747
41.4.1 Calculating the mean square displacement	747
41.4.2 Calculating the Hexbug velocity	748
41.5 Collective behaviors of the Hexbugs	749
41.5.1 Anti-cooperative behaviors	749
41.5.2 Cooperative behaviors	750
41.5.3 Formation of temporary channels	750
41.6 Problems	752
Bibliography	752
42 Experiments with Robot Swarms	754

VITO TRIANNI	
42.1 Introduction to robot swarms	755
42.2 Exploration of an unknown environment	755
42.2.1 Random exploration with a motion bias	756
42.2.2 Collective exploration strategies	758
42.3 Self-organized aggregation	759
42.4 Coordinated motion	761
42.5 Conclusions	762
42.6 Problems	763
Bibliography	764
IV Game	768
43 Active Matter Game	769
LAURA NATALI, JESUS DOMINGUEZ, DAVIDE BREONI, SANDRINE HEIJNEN, DAVID BRONTE CIRIZA, CAROLINA VAN BAALEN, ALIREZA KHOSHZABAN, AGNESE CALLEGARI	
43.1 Background	770
43.2 Playing the game	771
43.2.1 Preparation	771
43.2.2 Let the game start	772
43.2.3 Navigate the board	773
43.2.4 Interaction with other swimmers	774
43.2.5 Interaction with obstacles	775
43.2.6 The game end	778
43.3 Variations	778
43.3.1 Multiple players	779
43.3.2 Mystery mode	779
43.4 Additional materials	779
43.4.1 Personalized boards	780
43.5 Problems	781

Chapter 1

Introduction

GIOVANNI VOLPE, NUNO ARAÚJO, AGNESE CALLEGARI, GIORGIO VOLPE

Active matter systems are biological and artificial systems made of individual components that are capable of autonomously utilizing energy in their environment to drive themselves *far from equilibrium*, as shown in Fig. 1.1a. Because of this fundamental property, they can develop interactions with their environment

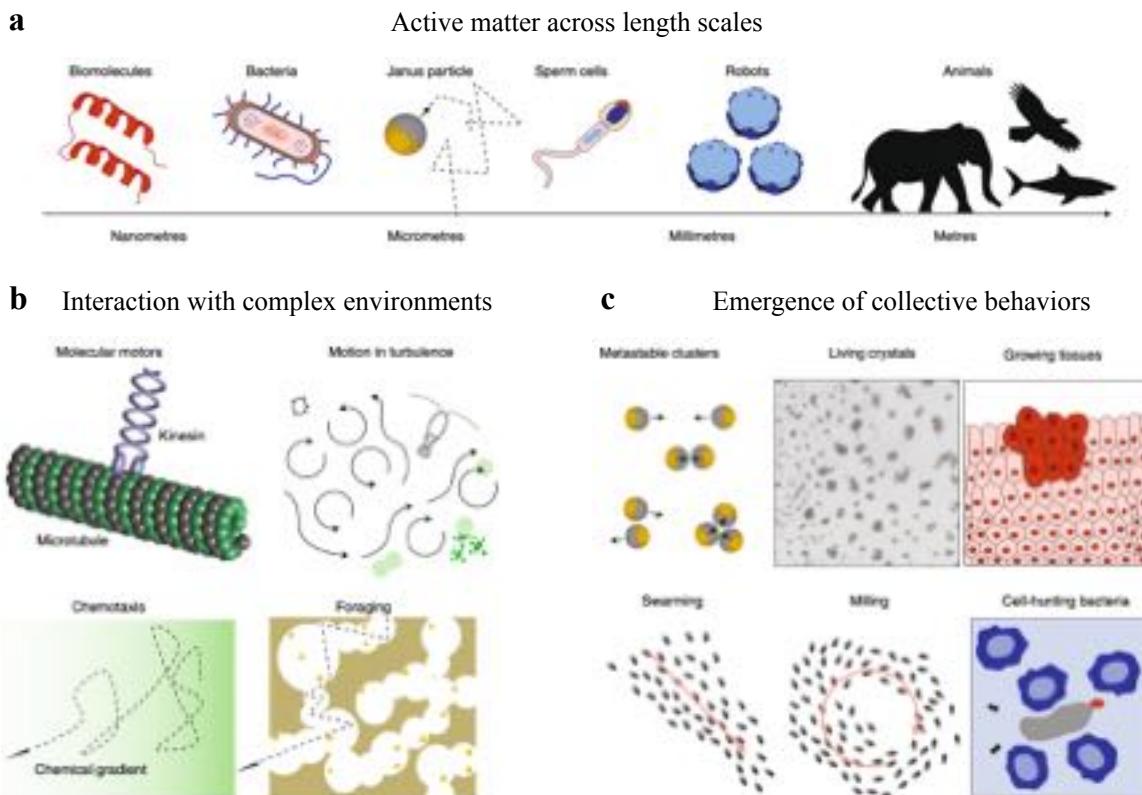


Figure 1.1: **Active matter systems.** (a) Active particles range from micrometers to meters. (b) Active particles react to environmental signals and optimize their behavior to reach certain goals, e.g., biomolecular motors move along microtubules, microorganisms swim in turbulent flows, motile cells respond to chemotactic gradients, and animals look for food (foraging). (c) Interactions between active particles may lead to complex collective behaviors, e.g., the growth of metastable clusters of particles, and to the emergence of collective dynamics such as swarming and milling. Reproduced from Ref. [1].

(Fig. 1.1b) and collective behaviors (Fig. 1.1c) that are fundamentally different from those of matter at thermodynamic equilibrium (their passive counterpart). Examples in the natural world include the beautiful spontaneous self-organized patterns shown by groups of animals, such as schools of fish and flocks of birds.

During the last decade, the field of active matter has developed enormously because it offers tantalizing options to perform tasks not easily achievable with other available techniques on the micro- and nanoscale. The study of active matter therefore provides a forefront route to designing smart devices and materials, e.g., for environmental remediation, for sensing, for drug delivery and gene therapy. In addition, manmade active matter exhibit striking similarities with the dynamical properties of living systems, providing an ideal model system to understand phenomena such as emergence and self-organization in living matter.

1.1 What does it mean to be far from equilibrium?

This is the proverbial one-million dollar question. Simply put, a system is far from equilibrium when it is not at equilibrium (not even near to it). The reality is that answering this question is non trivial at all as there is no unified framework or theory for non-equilibrium physics (yet!). Luckily, the task of identifying non-equilibrium systems is simplified by the fact that classical statistical physics dealing with equilibrium systems is a very well defined subject.

The whole concept of “being far from equilibrium” is better understood considering a system of particles subject to thermal fluctuations (e.g., molecules or colloids). Classical statistical physics tells us that, for a group of interacting particles that has reached thermal equilibrium, the energy $U(k)$ of a given microscopic configuration k and the probability $p(k)$ of observing the system in that configuration are related by the *Boltzmann distribution*, i.e., a simple exponential relationship:

$$p(k) \propto e^{-\frac{U(k)}{k_B T}}, \quad (1.1)$$

where T is the absolute temperature (in Kelvin) and $k_B = 1.38 \times 10^{-23} \text{ J K}^{-1}$ is the Boltzmann constant. If the group of particles is now driven away from equilibrium, this simple relationship does no longer hold and it becomes increasingly more challenging to express the likelihood of a given configuration in familiar thermodynamic quantities (e.g., internal energy, enthalpy entropy) and variables (e.g., pressure, volume, and temperature).

For pedagogical purposes, this breaking of the Boltzmann distribution can be visualized considering *active Brownian particles* (i.e., active colloids subject to Brownian fluctuations) confined within a finite two-dimensional space such as a pore, where *non-Boltzmann particle distributions* emerge as a signature of the differences between passive matter and active matter, as illustrated in Fig. 1.2.

Active Brownian particles move under the drive of a non-equilibrium force that makes them propel in a directional manner with a velocity v . These particles are also subject to Brownian fluctuations, which tend to randomize their orientation on a characteristic time scale $\tau_r = D_r^{-1}$, which is inversely proportional to the particle rotational diffusion coefficient D_r . The motion of these particles is therefore characterized by a finite *persistence length*, $\ell = v\tau_r$: on average, active Brownian particles will move along the direction of their initial orientation for a distance ℓ before their direction of motion is randomized. If $v = 0$, the motion of the particles is purely Brownian and we are looking at the equilibrium case (passive Brownian particles).

When active Brownian particles are free to move within a circular pore (Fig. 1.2), they explore over time the confined space available to them. Importantly, they do so in the absence of any external potential $U(x,y)$. Passive Brownian particles ($v = 0$) uniformly explore the space available to them to move in. Active particles instead ($v > 0$) spend more time near the pore walls. In fact, when an active particle encounters a boundary, it keeps moving along it until the rotational diffusion orients the particle away from it. The chance that an active particle encounters the pore boundary in one of its straight runs increases as its velocity v (and thus its persistence length ℓ) increases. These observations can be made more quantitative by looking at the particle probability distribution. The histograms on the bottom of Figure 1.2 show the probability distribution of finding the particle along a diameter of the pore: they clearly demonstrate that the likelihood of finding the particles near the pore boundaries increases with the particles’ velocity.

This fact (i.e., that active particles tend to accumulate at the boundaries) is a consequence of their non-equilibrium nature. As we mentioned above, for a passive Brownian particle at equilibrium with its environment, the probability distribution $p(x,y)$ is related to an external potential $U(x,y)$ by the Boltzmann relation.

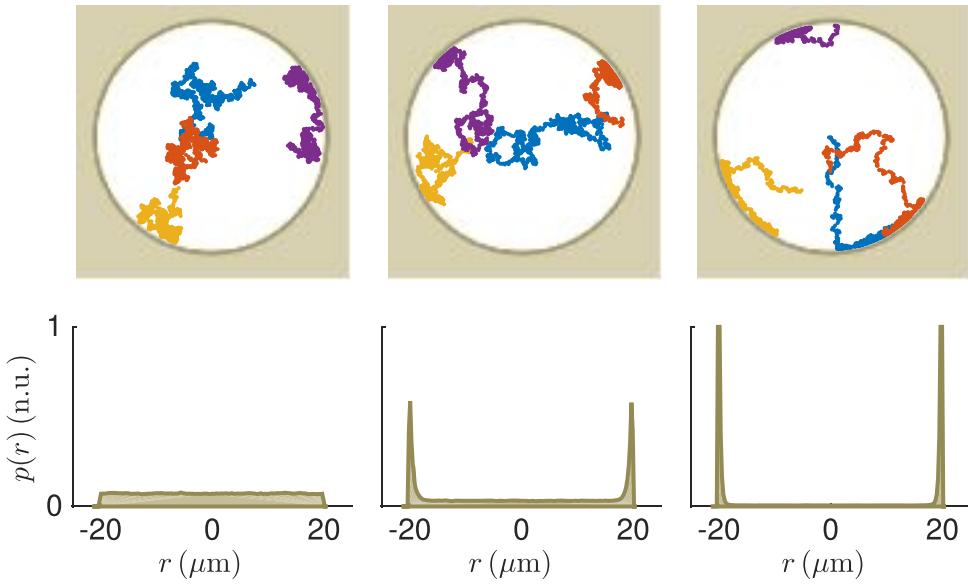


Figure 1.2: Non-Boltzmann statistics of active matter. (Top row) Trajectories of active Brownian particles moving in a circular pore with increasing level of activity from left to right. The case on the very left is equivalent to passive Brownian motion (no activity). (Bottom row) Corresponding probability distributions of finding the particles along one pore diameter. The higher the particle activity, the more this probability is peaked near the pore boundaries. Reproduced from Ref. [2].

In the cases considered in Figure 1.2, as there are no external potentials, the corresponding Boltzmann distributions is uniform (as for $v = 0$). However, the fact that the distributions for $v > 0$ are not uniform, despite the constant potential, is a clear deviation from Boltzmann statistics, i.e., from the behavior of passive matter at thermodynamic equilibrium.

1.2 The importance of dissipation in active matter

Now that we have established that active matter belongs to the (not fully understood) category of non-equilibrium systems, we can see how its characteristic emergent phenomena and collective behaviors are a direct consequence of this observation.

For consistency of formalism, we will continue considering the case of microscopic systems in the presence of thermal fluctuations. Nonetheless, the final conclusions of this section are entirely general and can be extended to active matter systems at all scales. In particular, we can consider what happens to our (microscopic) active matter system (i.e., subject to a non-equilibrium driving force), in contact with a thermal reservoir at temperature T , when it changes from a configuration k to a configuration j and back. For this, we will consider the path $\gamma \equiv [\mathbf{x}(t), \mathbf{v}(t)]$ that takes the system from k to j over a finite time τ and the exact time-reversed path $\gamma^* \equiv [\mathbf{x}(\tau-t), \mathbf{v}(\tau-t)]$ from j to k .

In the presence of a non-equilibrium force that can do work, the probability that the active system is found in a given configuration j after a finite time τ will depend on two key aspects: (1) the initial configuration k of the system and (2) the actual path of intermediate configurations l that the system takes under the drive of the non-equilibrium force. The latter is given by the actual work W done on the system by the non-equilibrium force during the transition between the two configurations rather than by thermal fluctuations alone. The work W , the heat dissipated by the system ΔQ , and the change in its internal energy ΔE during the transition are connected by energy conservation so that

$$\Delta Q = W - \Delta E . \quad (1.2)$$

For reference, if the system is at equilibrium (i.e. in the absence of the driving force but still subject to thermal fluctuations), $W = 0$ and the relative probabilities of the two configurations in steady state are a simple

exponential function of their respective energies according to the Boltzmann distribution, namely:

$$\frac{p(k)}{p(j)} = e^{\frac{E_j - E_k}{k_B T}}. \quad (1.3)$$

By combining energy conservation and time-reversal symmetry, we can obtain an expression for the relative probabilities π associated to γ and its time-reversal equivalent γ^* :

$$\frac{\pi(\gamma)}{\pi(\gamma^*)} = e^{\frac{\Delta Q(\gamma)}{k_B T}}. \quad (1.4)$$

This equation states that the likelihood of a forward path γ from k to j over is time-reversed version γ^* is higher by an exponential factor of the amount of heat dissipated when the system follows the forward path γ . This expression might sound abstract but has some important consequences for active matter systems and, in general, for far-from-equilibrium systems. In fact, if we consider all possible paths and their time reverse, we can come to the following conclusions: [3]

- First of all, the previous expression highlights the importance of *dissipation* in active matter. Even in the absence of external potentials, the probability of a change in configuration (powered by a change in internal energy or by an external energy source) is intrinsically connected to the amount of heat released by the system in this process (i.e., its dissipation).
- This change in configuration is *intrinsically* irreversible due to dissipation. This means that different configurations of the system are not going to be equally efficient in utilizing the work from the non-equilibrium force. In particular, the more irreversible a path is, the more work a system must do to follow it and the higher the dissipation.
- Because of this, while any change of configuration might (appear to) be random, the most durable and irreversible of these changes occur under the conditions for which the system shows optimal absorption of energy and dissipation of work.
- In particular, the absorption of energy from the environment allows the system to jump over potential energy barriers that are too high to be crossed by thermal fluctuation alone (i.e., in the equilibrium case). This energy is typically dissipated during the cross-over, and it is no longer available for the system to reverse the process unless more energy is provided and absorbed.
- With time, the accumulation of these preferential changes will allow the active matter system to evolve towards a collective state of self-organization that appears in adaptation to the environmental conditions.

1.3 Overview of the Book

This book explores how these key physical concepts drive the mesmerizing collective behaviours typically observed in nature (Figure 1.2) and discusses their translation to synthetic materials that aim to harness them for human activities. This book therefore provides a comprehensive exploration of active matter, covering fundamental principles, theoretical frameworks, experimental approaches, and applications. It is structured into several parts, each focusing on a specific aspect of the field. This book is a direct outcome of the ActiveMatter ITN, a European research network financed by the European Commission that has brought together scientists from diverse backgrounds to explore active matter across multiple scales and disciplines. The contributions from researchers within this network have shaped the content of this book, and we hope it will serve as a valuable resource for students, academics, and professionals interested in the field. Below, we provide an overview of the book's parts and chapters.

1.3.1 Part I: Theory

This part establishes the theoretical foundation necessary for understanding active matter systems and serves as a basis for the experimental and applied topics explored in later parts of the book. It includes:

- **Chapter 2: What is Soft Matter?** — This chapter provides a foundational introduction to soft matter physics, emphasizing the distinct properties of polymers, colloids, liquid crystals, and other soft materials. These materials serve as the building blocks of active matter systems, offering a bridge between classical condensed matter and the dynamic behaviors characteristic of active systems.
- **Chapter 3: Nonequilibrium Physics of Small Systems** — Exploring how active matter systems operate far from equilibrium, this chapter explores fundamental principles like energy dissipation, thermodynamic fluctuation theorems, and the interplay between forces and self-organization in microscopic systems. It highlights the contrast between equilibrium and nonequilibrium dynamics.
- **Chapter 4: From Passive to Active Brownian Motion** — This chapter examines the transition from passive Brownian motion to active Brownian dynamics, exploring how self-propulsion mechanisms and interactions with the environment drive particle motion. The concepts presented here are crucial for understanding biological and synthetic active systems.
- **Chapter 5: Stochastic Differential Equations and Active Matter** — Introducing stochastic differential equations as powerful tools for modeling active matter systems, this chapter covers key mathematical approaches to describe the noisy, fluctuating behaviors inherent in active dynamics.
- **Chapter 6: Anomalous Diffusion and Ergodicity Breaking** — This chapter investigates the phenomena of *anomalous diffusion* and *ergodicity breaking*, where classical diffusion laws fail, revealing the complexity of transport processes in active matter systems. It emphasizes both theoretical models and experimental implications.
- **Chapter 7: Low Reynolds Number Hydrodynamics** — The unique fluid dynamics of active matter at *low Reynolds numbers* is explored in this chapter. It discusses the Stokes regime, hydrodynamic interactions, and their relevance to microswimmers and other active systems.
- **Chapter 8: Phoretic and Self-Phoretic Transport** — This chapter deals with *phoretic mechanisms*, highlighting how gradients in temperature, concentration, or electric fields drive motion. It explains the underlying physics and explores *self-phoretic motion* in synthetic and biological systems.
- **Chapter 9: Dynamical Density Functional Theory** — This chapter provides a detailed overview of *dynamical density functional theory* (DDFT) as a framework to model collective behaviors and phase transitions in active matter, bridging microscopic and macroscopic descriptions.
- **Chapter 10: Motility-Induced Phase Separation** — Focusing on the phenomenon of *motility-induced phase separation* (MIPS), this chapter explains how active particles segregate into dense and dilute regions due to self-propulsion, offering insights into both theoretical predictions and experimental validations.
- **Chapter 11: Liquid Crystals** — Liquid crystals form an important class of soft matter with unique phase behaviors. This chapter explores their relevance to active matter, particularly in systems where orientation and flow couple dynamically, such as *active nematics*.
- **Chapter 12: Active Nematics** — Extending the discussion of liquid crystals, this chapter introduces active nematic systems where self-driven units generate long-range orientational order. Theoretical models and experimental realizations are discussed in detail.
- **Chapter 13: Activity in the Cytoskeleton** — The cytoskeleton, a dynamic protein network within cells, serves as a biological example of active matter. This chapter explores molecular motors, filamentous proteins, and their roles in cellular mechanics and motility.
- **Chapter 14: From Cells to Tissues** — Transitioning from individual cell behavior to multicellular systems, this chapter examines tissue mechanics, morphogenesis, and collective cell migration through the lens of active matter principles.

1.3.2 Part II: Simulation

This part highlights the computational frameworks essential for understanding and predicting active matter behaviors, bridging theoretical concepts with numerical simulations. It includes:

- **Chapter 15: Numerical Models for Active Matter** — This chapter introduces key numerical approaches for simulating active matter, ranging from *agent-based simulations* that capture individual particle behaviors to *continuum models* that address macroscopic properties. By bridging these scales, it provides a holistic view of active system modeling.
- **Chapter 16: Molecular Dynamics** — This chapter explores molecular dynamics simulations, focusing on their application to active systems. Topics include the implementation of inter-particle forces, self-propulsion algorithms, and thermodynamic conditions relevant to active particles.
- **Chapter 17: Brownian Dynamics** — This chapter deals with Brownian dynamics, a simulation method suited for modeling active particles in fluctuating environments. It covers stochastic motion, interaction potentials, and methods for analyzing particle trajectories and ensemble behaviors.
- **Chapter 18: Accelerating Molecular Dynamics** — Expanding on traditional molecular dynamics, this chapter introduces advanced techniques such as coarse-grained modeling, adaptive resolution schemes, and parallelization strategies. These methods enable the study of complex, large-scale active matter systems.
- **Chapter 19: Implementation with LAMMPS** — This chapter serves as a practical guide to implementing active matter simulations using the LAMMPS simulation package. It provides step-by-step instructions for setting up simulations, integrating custom force fields, and analyzing output data.
- **Chapter 20: Interaction Rules in Models of Active Matter** — This chapter focuses on the rules governing particle interactions in computational models of active matter. It explores key interaction types such as alignment, attraction–repulsion, and quorum sensing, as well as their implications for emergent collective behavior.
- **Chapter 21: Mesoscopic Modelling of Epithelial Tissues** — Active matter principles are applied to biological systems in this chapter that introduces mesoscopic models of epithelial tissues. Topics include vertex and continuum models, tissue boundary conditions, and mechanical properties influencing collective cell behavior.
- **Chapter 22: Active Fluids** — This chapter examines active fluid systems, focusing on hydrodynamic interactions and on the interplay between activity and fluid mechanics. It includes discussions on continuum descriptions, stability analysis, and simulations of active turbulence.
- **Chapter 23: Multiparticle Collision Dynamics** — This chapter introduces multiparticle collision dynamics, a computational approach combining particle-based and continuum mechanics. It highlights its versatility in simulating solvent-mediated interactions in active matter.
- **Chapter 24: Lattice Boltzmann Method** — This chapter discusses the lattice Boltzmann method, a numerical technique for modeling fluid dynamics in active systems. It covers its theoretical foundations, implementation strategies, and applications to active suspensions and phase separations.
- **Chapter 25: Lattice Boltzmann for Active Fluids** — Building on the previous chapter, this chapter deals with specific applications of the lattice Boltzmann method for studying active fluids, including active nematics, polar fluids, and binary mixtures. Advanced implementations and challenges are also discussed.
- **Chapter 26: Phase Field Models of Active Systems** — This chapter presents phase field models as tools for studying diffuse interfaces in active matter. Applications include biphasic active systems, active deformable surfaces, and the dynamics of cell monolayers, bridging the gap between theory and experiment.

1.3.3 Part III: Experiments

This part highlights cutting-edge experimental techniques, bridging theoretical models and computational simulations with real-world observations of active matter systems. It includes:

- **Chapter 27: Chemistry of Micromotors** — This chapter explores the chemical mechanisms that drive the self-propulsion of micromotors. It discusses the types of catalytic reactions involved, energy conversion processes, and the influence of chemical gradients on motion, offering insights into both natural and synthetic micromotor systems.
- **Chapter 28: Experiments with Artificial Active Particles** — This chapter introduces laboratory techniques for fabricating and characterizing artificial active particles. Topics include propulsion mechanisms, material choices, and analysis methods for studying their motion and interactions.
- **Chapter 29: Active Particles with Feedback Interactions** — Feedback mechanisms allow active particles to dynamically respond to their environment. This chapter discusses experimental setups, theoretical models, and applications of feedback-controlled active systems, with an emphasis on self-regulating behaviors.
- **Chapter 30: Active Colloids at Fluid Interfaces** — Focusing on the unique behaviors of active colloids at fluid interfaces, this chapter examines their self-assembly, propulsion mechanisms, and interactions with surface forces, with implications for applications in materials science.
- **Chapter 31: Optical Tweezers** — Optical tweezers are powerful tools for manipulating and studying active particles. This chapter describes the principles of optical trapping, experimental setups, and their applications in probing active matter properties at the microscale.
- **Chapter 32: Atomic Force Microscopy** — Atomic force microscopy (AFM) is presented as a tool for probing the mechanical and structural properties of active matter systems. This chapter covers AFM techniques, force measurements, and their applications in characterizing soft and active materials.
- **Chapter 33: Spectroscopy** — This chapter explores spectroscopic techniques for investigating the chemical and physical properties of active matter. It includes Raman and fluorescence spectroscopy, offering insights into molecular interactions and dynamics.
- **Chapter 34: Experimental Active Nematics** — This chapter discusses active nematics, characterized by long-range orientational order. It covers experimental methods for preparing and observing nematic systems, including defect dynamics and collective behaviors.
- **Chapter 35: Planktonic Active Matter** — This chapter explores biological active matter in marine environments. Topics include the swimming behavior of plankton, their interactions with turbulent flows, and the ecological implications of their collective dynamics.
- **Chapter 36: Microfluidics** — This chapter examines the role of microfluidic devices in active matter research. It discusses fabrication techniques, experimental setups, and their use in studying confined and flow-driven active systems.
- **Chapter 37: 3D Printing** — This chapter deals with the use of 3D printing to fabricate active matter components and devices. It highlights printing techniques, material choices, and applications in creating programmable active structures.
- **Chapter 38: Active Mechanical Metamaterials** — This chapter introduces mechanical metamaterials with active properties, focusing on their design, fabrication, and potential applications in robotics and adaptive systems.
- **Chapter 39: Living Robotics** — Combining active matter principles with robotics, this chapter explores *bio-hybrid systems* that integrate living cells or tissues into robotic platforms. Applications in medicine and environmental sensing are highlighted.

- **Chapter 40: Locomotor Transitions in Complex 3D Terrain** — The ability of active matter to adapt its locomotion in complex environments is explored in this chapter. Examples include transitions between locomotor modes to traverse a variety of large obstacles, with insights into biological and robotic systems.
- **Chapter 41: Experiments with Robot Swarms** — This chapter investigates the collective behaviors of robot swarms as active matter systems. Topics include coordination algorithms, self-organization, and applications in exploration and task optimization.

1.3.4 ActiveMatter Game

The final chapter presents the **ActiveMatter Game**, a novel educational initiative developed by the PhD students of the ActiveMatter ITN (<http://active-matter.eu/>). This interactive game introduces key concepts of active matter to a broad audience, blending scientific principles with engaging gameplay. The game reflects the collaborative and interdisciplinary nature of the ActiveMatter ITN doctoral network financed by the European Commission, which has played a crucial role in fostering new research and educational activities in the field of active matter.

Bibliography

- [1] F. Cichos, K. Gustavsson, B. Mehlig, and G. Volpe. Machine learning for active matter. *Nature Machine Intelligence*, 2:94–103, 2020.
- [2] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Review of Modern Physics*, 88:045006, 2016.
- [3] J. L. England. Dissipative adaptation in driven self-assembly. *Nature Nanotechnology*, 10:919–923, 2015.

Part I

Theory

Chapter 2

What is Soft Matter?

MARGARIDA MARIA TELO DA GAMA, JOSÉ MARIA TAVARES, MYKOLA TASINKEVYCH

Condensed matter physics, with its focus on the macroscopic properties of materials originating from their atomic building blocks, emphasizes the role of structural correlations in shaping these properties. Unlike hard matter, soft matter consists of building blocks that are substantially larger than atoms, such as polymers, which can contain chains of up to millions of atoms. Thus, soft matter, a distinct subtype of condensed matter, has unique characteristics such as its inherent softness or flexibility. Its dramatic response to minor external perturbations goes beyond the linear regime defining hard matter, necessitating the development of novel theories, including classical field theories encompassing non-linear elasticity and time-dependent, out-of-equilibrium terms. Soft matter physics as a discipline can trace its lineage to the foundational work in polymer physics done by pioneers such as Sir Sam Edwards of the University of Cambridge and Pierre Gilles de Gennes of Collège de France, a Nobel Laureate in Physics in 1991 (Fig. 2.1). Soft matter is ubiquitously present, even constituting the majority of our own bodies. This chapter serves as a preliminary exploration of condensed matter's structure and organization, emphasizing soft matter, and presenting the basic theoretical tools used for its study.

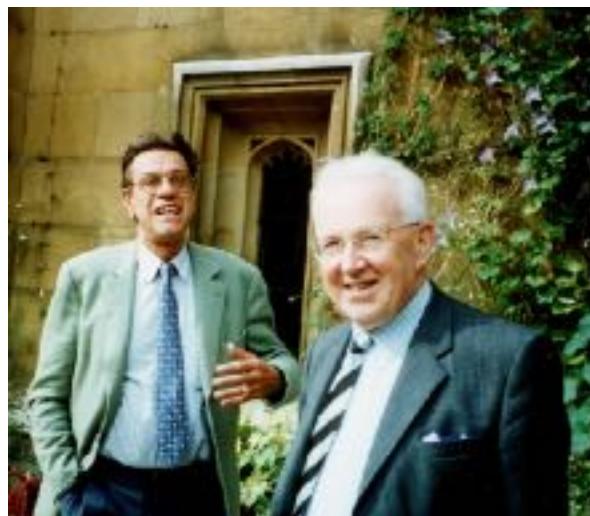


Figure 2.1: **Pierre Gilles de Gennes and Sir Sam Edwards.** The Nobel-Prize winners Pierre Gilles de Gennes and Sir Sam Edwards are widely considered the founders of the soft matter discipline. Here, they are photographed at Sir Sam Edwards' retirement meeting in Cambridge, in September 1995. Reproduced from Ref. [1].

2.1 Condensed matter

Richard Feynman astutely observed that the role of a physicist is to observe and develop theories to explain phenomena across all length scales [2]. Both the infinitesimally small and the infinitely large, elusive to casual observation, require significant collaborative efforts and robust equipment to facilitate progress. Indeed, significant strides have been made, both as a response to and a driver of new technological advancements, fueled by expansive international collaborations at institutions like the European Organization for Nuclear Research (CERN) and the European Space Agency (ESA), to cite two Europe-based examples. Conversely, the quest to comprehend phenomena at the human scale is the primary task of condensed matter physics. Here, the interplay of order and disorder governs the structure and properties of various matter phases, extending to the intricate complexity of living matter.

The study of the macroscopic properties of condensed matter involves analyzing its microscopic components, their interactions, and how these interactions manifest as observable phenomena at larger length scales. The challenge lies in addressing the emerging many-body problems, where numerous particles (e.g., electrons, atoms, or colloids) are continuously interacting and moving. Applying the appropriate dynamics, whether quantum or classical, along with the relevant interactions (e.g., Coulomb interactions between electrons in solids or effective interactions between colloids in soft matter) would theoretically yield an “exact” system description. However, it is widely acknowledged that these equations are generally too complex and impractical for delineating macroscopic properties. Consequently, condensed matter physics relies on statistical physics methodologies for a more manageable and applicable system description.

Condensed matter is characterized by a large number of interacting particles and macroscopic emerging properties. For example, the rigidity of solids and the fluidity of liquids are emerging properties. While the exact values of these properties may differ, solids are always rigid and fluids always flow. Thus, their rigidity is a robust property of solid. It depends on the order of the particles in the solid rather than on their specific microscopic interactions. The same could be said about the fluidity of liquids. Solids and liquids are different states or phases of matter, corresponding to different types of order of their constituents. In this sense, the states or phases of condensed matter are also emerging properties. The interactions are responsible for the emergence of new states, which are different from combinations of states of the individual components.

2.1.1 Phases of matter

The quintessential example of long-range order can be observed in crystalline structures. Crystals, with their extensive structures, possess discrete translation symmetry, a microscopic periodicity that was unveiled through X-ray diffraction by W. Bragg and L. Bragg in the early 20th century, a discovery that won them the Nobel Prize in Physics in 1915 [3]. Within a crystal, the pertinent degrees of freedom pertain to the positions of its constituent atoms or molecules. These atomic positions replicate to form a periodic structure, defined by three translation vectors. Notably, crystal symmetry is not a property of individual atoms or molecules, but rather of the repeating “structural units” comprising interacting atoms.

While the interactions between atoms and molecules, predominantly electrostatic in nature, can be intricate, their effects on the organization of a vast number of particles are remarkably straightforward. In isolated systems, many interacting atoms or molecules will display equilibrium states that exhibit increased order as temperature lowers.

The crystalline state emerges from a *spontaneous symmetry breaking*, upsetting the balance between order and disorder, resulting in characteristics that are not inherent to the individual constituents. The resulting order may be conserved across different systems, leading to some universally shared properties in ordered systems. The global order and extended system properties, such as the rigidity of solids, remain robust, unaffected by individual defects.

Consider a macroscopic system comprising around 10^{23} particles (the approximate number of atoms or molecules in a tablespoon of water). These particles are constantly interacting, bonding, and colliding, leading to an astronomical number of simultaneous interactions. So, the question arises: how can we predict the structure and properties of such a complex system?

An answer at the macroscopic scale was given more than two hundred years ago [4] Macroscopic systems in thermal equilibrium have an extremely simple behavior described by the laws of thermodynamics. For example, the *second law* (or the *principle of maximum entropy*) tells us that the stable state of a system at

temperature T with constant volume is determined by the state that minimizes Helmholtz free energy F :

$$F = E - TS, \quad (2.1)$$

where S is the entropy of the system and E is its internal energy (sum of the kinetic and potential energy).

The stability of a solid or gas derives from this principle and the fact that matter is made up of atoms. At low temperatures and high densities, the crystalline solid state becomes stable, as molecular order minimizes energy. This is due to the fact that the interatomic potential reaches a minimum at a specific distance, and the crystal aligns with the periodic spatial arrangement that brings the system's potential energy to its lowest point. In contrast, at high temperatures and low densities, the gas state is stable. In this scenario, interaction energy becomes negligible, and molecular disorder maximizes the system's entropy.

This argument is general and can be used to show the existence of phase transitions. The two phases, in this case the crystalline solid phase and the gas phase, have different symmetries. At the macroscopic scale, the gas is characterized by continuous translation invariance (all points are equivalent from the point of view of their physical properties), while the crystalline solid has discrete translation symmetry (the equivalent points are those related by a discrete translation group, namely the *Bravais lattice*). Since the solid is stable at low temperatures and the gas is stable at high temperatures, the system must necessarily exhibit at least one phase transition.

There are essentially two types of phase transitions: *continuous phase transitions* and *discontinuous phase transitions*. The transition between the solid and the gas is discontinuous because it involves *latent heat*. When a substance changes from a disordered phase (the gas) to an ordered phase (the solid), an amount of heat (the latent heat) is released in the process. This release of heat reveals that the material's structure is radically altered. This happens, in particular, along the lines of sublimation, melting and condensation of most known substances.

Recently manufactured colloidal systems (dispersions of spherical particles, with radii of the order of $1\text{ }\mu\text{m}$, in a solvent) exhibit only one gas (or fluid) and another solid phase, separated by a sublimation line. However, most natural substances are characterized by slightly more complex phase diagrams and also have a liquid phase.

Unlike the solid and the gas, the liquid is stable over a temperature range limited by the *triple point* (where the solid, the liquid and the gas coexist simultaneously) and the *critical point*, where the *condensation line*, along which they coexist, ends. The liquid–gas critical point is an example of a continuous phase transition, that is, a transition that does not involve latent heat.

In a liquid, the total interaction energy E is balanced by the entropic term TS , i.e., $E \approx TS$, and there is no simple analytical argument that predicts the stability of the liquid phase. In fact, liquids do not have a simple reference system, like the perfect lattice associated with crystalline solids. Many aspects of the structure and dynamics of liquids are now well understood, with most of progress arising from the results of computer simulations. In fact, simulations are now an extremely important tool, along with theory and experiment, in the study of liquids and of soft matter.

2.1.2 Thermal fluctuations and criticality

Thermal fluctuations are random deviations of a system from its average state that occur in a system at equilibrium. All thermal fluctuations increase as the temperature increases, and decrease as the temperature approaches absolute zero. Thermal fluctuations are a manifestation of the temperature of thermodynamic systems: a system at nonzero temperature does not stay in a single microscopic state, but randomly samples all possible states (compatible with the external constraints) with probabilities given by the *Boltzmann distribution*.

Thermal fluctuations generally affect all the *degrees of freedom* of a system. Let us go back to the periodic solid: there will be, for example, random vibrations that increase as the temperature increases. Thermodynamic variables, such as pressure, temperature, or entropy, likewise undergo thermal fluctuations. For example, for a system that has an equilibrium pressure, the system pressure fluctuates to some extent about the equilibrium value. Only the *control variables* of statistical ensembles (such as the number of particles N , the volume V , and the internal energy E in the microcanonical ensemble) do not fluctuate.

For large systems at equilibrium, the *central limit theorem* applies and the distribution of the thermal fluctuations is a sharply peaked Gaussian. Fluctuations about the equilibrium value do occur but large fluctuations (compared to the standard deviation) are extremely rare and negligible in the thermodynamic limit.

By contrast, thermal fluctuations play a major role in continuous phase transitions. The understanding that the fluctuations become non-Gaussian is relatively recent and has had an enormous impact in condensed matter physics and as a result new paradigms were introduced. Indeed, in the second half of the 20th century, the understanding of fluctuations (at all length scales) in critical phenomena led to the development of a profound and sophisticated physical theory, i.e., the *renormalization group theory*. This theory provides a microscopic understanding of *scaling* and *universality*, observed experimentally, and established the concepts of broken symmetry and order parameters as unifying theoretical concepts to describe condensed matter phases ranging from periodic solids to disordered systems, including simple fluids, liquid crystals and polymers.¹

2.1.3 Broken symmetry and phase transitions

Spontaneous symmetry breaking requires the existence of a symmetric probability distribution, i.e., a probability distribution where any of the possible outcomes has the same probability. For instance, with a non-biased coin toss, the outcomes of landing on heads or tails are equally probable. More broadly, this principle implies that the fundamental rules remain invariant under a symmetry transformation. Phases of matter, including crystals and magnets, and simple phase transitions can be characterized through *spontaneous symmetry breaking*.

For *ferromagnetic materials*, the underlying laws are invariant under spatial rotations. Here, the *order parameter* (see Section 2.3) is the *magnetization*, which measures the total magnetization or the magnetic density. Above the *Curie temperature*, the order parameter is zero, which is spatially invariant, and there is no symmetry breaking. Below the Curie temperature, however, the magnetization acquires a constant nonvanishing value, which points in a certain direction. The rotational symmetries that leave the orientation of this vector invariant remain unbroken, unlike the rotations that do not and are spontaneously broken. Likewise, the laws describing a solid are invariant under the full Euclidean group, but the solid itself spontaneously breaks this group down to a space group (Bravais lattice). The displacement and the orientation of the atoms or molecules are the order parameters.

Most phases of matter can be understood through the concept of spontaneous symmetry breaking. For example, crystals are periodic arrays of atoms that are not invariant under all translations (only under a subset of discrete translations). Magnets have north and south poles that are oriented in a specific direction, breaking rotational symmetry. In addition to these examples, there are a whole host of other symmetry-breaking phases of matter — including the nematic phase of liquid crystals, and many others.

Lev Landau introduced a framework in an attempt to formulate a general theory of continuous (i.e., second-order) phase transitions (see Section 2.3.1). This theory can also be extended to systems under externally-applied fields, and used as a quantitative model for discontinuous (i.e., first-order) transitions. Other generalizations include vector and tensor order parameters, which are appropriate to describe polar and nematic ordered phases.

2.2 Soft condensed matter

As its name implies, soft matter is concerned with materials that are soft, i.e., they are easily deformed. These materials, which include polymers, gels, colloids, emulsions, foams, surfactant assemblies, liquid crystals, granular materials, and many biological materials, have in common that they are organized on mesoscopic length scales, with structural units or features that are much larger than an atom, but much smaller than the overall size of the material.

The large size of the basic structural units and the relatively weak interactions that hold them together are responsible for the characteristic softness of these materials, but they also lead to many other distinct features of soft materials, such as sensitivity to thermal fluctuations and external forces and slow responses with long relaxation times, often resulting in kinetic arrest in non-equilibrium states. These features make soft matter problems not only interesting but challenging. The realization that liquid crystals and polymers exhibit symmetry breaking and many fluctuating degrees of freedom has on the one hand revived classical fields of physics, such as the study of elasticity and (non-Newtonian) fluid flows, and on the other hand helped to establish the field of soft condensed matter physics.

¹A spectacular illustration of critical fluctuations may be visualized at <https://www.doitpoms.ac.uk/tplib/solid-solutions/demo.php> for a binary mixture of cyclohexane and aniline, close to the liquid-liquid critical point.

Although condensed matter physics began as the study of matter in its solid state, dominated by electronic interactions, an increasing number of physicists are currently working on soft condensed matter systems. There are many reasons for this but an important one is that there is a wide variety of old and new soft matter materials. One of the most recent is *active matter*. By contrast, assemblies of colloids that are often disordered may be used to study the glassy state and the glass transition, probably the oldest open problem in classical statistical mechanics. In general terms, in soft matter physics the aim is to understand and predict the properties of the systems given the structure of the extended systems and to control them through the design of the particles or the building blocks.

2.2.1 Building blocks and interactions

Atoms and small molecules, found in nature or synthesized in laboratories, constitute the building blocks of solid-state physics at the (sub)nanometer scale. However, the building blocks of soft matter, namely colloids, polymers, and other macromolecules, fall between 5 nanometers and 5 micrometers in size.

Interactions between particles at the nanometer and micrometer scales vary dramatically in both nature and strength. At the nanoscale, interactions are predominantly electronic (on the order of electronvolts) and governed by quantum mechanics, while interactions between larger particles are considerably weaker and classical. These interactions are typically entropically driven and on the order of room temperature thermal energy ($k_B T$, where k_B is the Boltzmann constant or 0.025 eV at room temperature).

To illustrate, let's consider interactions in colloidal systems. Model colloidal particles, such as monodisperse spheres made of polystyrene or silica, usually aggregate due to attractive London dispersion or van der Waals forces. However, these particles can also be charged, causing them to repel each other, with the repulsion often screened by mobile ions in the solvent. Deryaguin-Landau-Verwey-Overbeek theory (DLVO theory) combines these attractive and repulsive forces, resulting in an interaction potential as illustrated by the green curve in Fig. 2.2a.

A defining characteristic of soft matter is the weak interaction between its particles, dominated by energies of the order of room temperature thermal energy. As a result, soft matter exhibits a range of states, sensitivity to external conditions, metastable states, and macroscopic softness. This variety of states, along with their responsiveness to external conditions, makes soft materials valuable in a broad spectrum of applications, from packaging materials and adhesives to cosmetics and detergents. Moreover, several biological materials, such as blood, muscle, and even jello, fall under the category of soft matter.

Another defining characteristic is the versatility in shaping and sculpting particles at the micrometer scale, allowing for a much broader range of forms than their nanoscale counterparts. Despite these differences, many

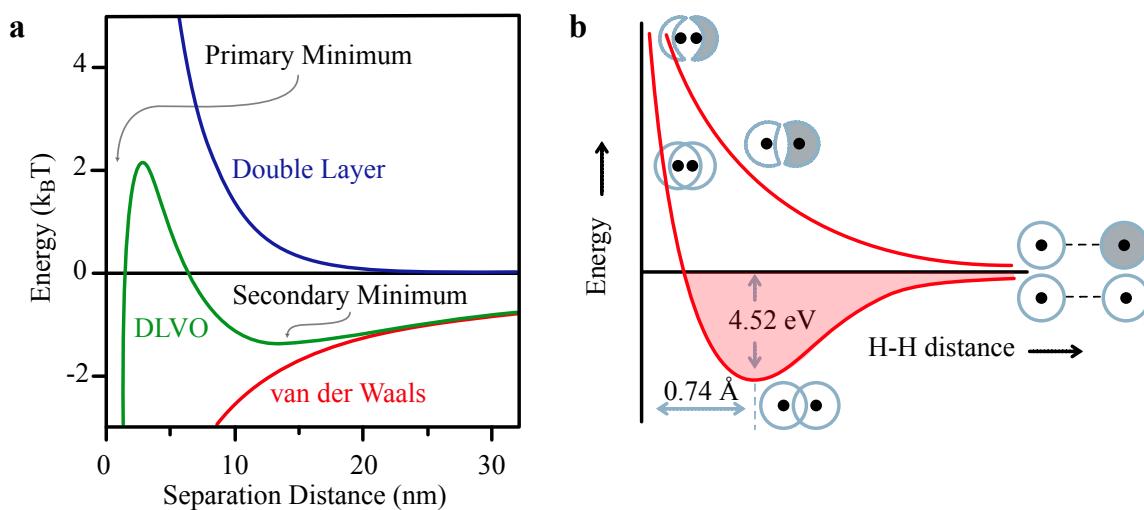


Figure 2.2: **Energy of interaction.** (a) DLVO interaction between two colloidal particles. (b) Covalent bond between two hydrogen atoms.

properties of soft matter result directly from the weak interactions between structural elements and a delicate equilibrium between entropic and enthalpic contributions to the free energy. At a coarse-grained level, colloids and liquid crystals can be described by a generalized Landau framework, which will be discussed quantitatively later.

2.2.2 Some examples

The macroscopic properties of soft matter are contingent upon its mesoscale structure. The challenge lies in understanding this structure, the stability of the extended materials, and their respective properties. An even more complex problem is designing building blocks that assemble into materials with pre-specified macroscopic properties. Investigating soft matter systems and bridging the relevant length and time scales requires a combination of computer simulations, theoretical concepts, and a variety of experimental tools. Effective interaction potentials and coarse-grained statistical field theories also play pivotal roles. These concepts, among others, will be discussed in various chapters of this book, beginning with Section 2.3 of this Chapter.

In the rest of this section, we offer a qualitative description of some of the most common forms of soft matter and their building blocks, emphasizing shared aspects and some associated challenges. Note that this is a selection based on personal preference, not a comprehensive review of the field's state of the art. Most soft materials are composed of linear or branched polymers, colloids, and liquid crystals. In some instances, the building blocks are not particularly large, but they self-assemble into units or phases that react to external forces and exhibit collective movement, as in liquid crystals. In other cases, the building blocks comprise millions or billions of atoms, as seen in polymers and colloids. Regardless of the system, the response is substantial and the dynamics slow, leading to non-equilibrium and non-linear responses, which contribute to the unique properties of soft materials.

Polymers

A polymer is a substance composed of macromolecules, often simply referred to as polymers. These macromolecules are characterized by their unique chain-like structure, constructed from repeating units known as monomers. While the number of monomers in a polymer commonly numbers in the thousands, it can extend considerably further, reaching into the millions or even tens of millions.

A polymeric material's structure can be described across a range of scales, from the sub-nanometer to the macroscopic level. At the sub-nanometer scale, the structure is defined by the individual constituent monomers. Moving up the scale, the structure is determined by how these monomers are arranged within the polymer chain, which could be either linear or branched. Subsequently, the focus shifts to the structure or conformation of the polymer chain itself, which is influenced by factors such as chain rigidity and the surrounding solvent.

In his book "Scaling Concepts in Polymer Physics" [5], de Gennes describes some of the developments that contributed to major advances in the study of flexible polymers in solutions and melts. These are neutron-scattering experiments on selectively deuterated molecules and inelastic scattering of laser light; last but not least was the discovery of a relationship between polymer statistics and critical phenomena, which led to simple scaling laws, for which de Gennes was ultimately awarded the Nobel Prize.

Both the chain structure and conformation determine the type of extended structure or polymer phase, for example semi-crystalline, glassy or microphase separated. These phases play a major role in determining the physical properties of the polymeric material. Polymers have become indispensable in modern technology; they are used as plastics, rubbers and textiles. Polymers are also the basic molecules of life, as proteins or DNA.

Colloids

A colloid is a phase-separated mixture where one substance of microscopically dispersed particles is suspended throughout another substance. Sometimes the dispersed substance alone is called the colloid. Unlike a solution, whose solute and solvent constitute a single phase, a colloid has a dispersed phase (the suspended particles) and a continuous phase (the medium of suspension).

Typically, colloids are subject to Brownian motion. Colloids are an interesting model system for atoms. Many of the forces that govern hard matter, such as excluded volume interactions or electrostatic forces, govern

the structure and behavior of colloidal suspensions. For example, the same equations used to model ideal gases can be applied to model the behavior of a hard sphere colloidal suspension.

A strategy to control material properties from the bottom up is by starting with colloidal building blocks. Colloidal particles can be synthesized in many sizes and shapes, and the interactions between them can be tuned by varying the surface charge, by adding polymeric additives, or by decorating the particles with complementary DNA strands. This can lead to self-organization into a large variety of structures and phases, such as colloidal liquids and crystals mimicking and going beyond the phase behavior of atoms and molecules.

Since colloids are big and slow enough to be followed individually using an optical microscope, they can be used to study fundamental physics problems at the individual-particle level, such as nucleation, melting, crystal defects, and glasses. Non-equilibrium states are more prominent for colloids than for atoms or small molecules. For example, gelation can arrest the particles in a non-equilibrium aggregated state, suppressing crystallization.

Liquid crystals

Liquid crystals are a state of matter that has properties between those of conventional liquids and those of solid crystals. For instance, a liquid crystal may flow like a liquid, but its molecules may be oriented in a crystal-like way more akin to a solid.

If a material is made of rod-like molecules, in the crystalline state the molecules are regularly placed on lattice sites and are also aligned. There is order in position and in orientation. In the liquid state, the position and orientation of the molecules are both random. In the liquid crystalline state, molecules have orientational order and lack full positional order. In *nematics*, the molecules retain the orientational order, while there is no positional order (Fig. 2.3a). In *smectics*, in addition to the orientational order, there is partial positional order: the molecules are regularly placed along one direction, but are random on the perpendicular plane (one-dimensional positional order). If the direction of the positional and the orientational order is the same, this phase is called *smectic A* (Fig. 2.3b), while, if these directions do not coincide, this phase is called *smectic C* (Fig. 2.3c). Among many other liquid crystalline phases, there is the *cholesteric*, or twisted nematic, where the direction of the orientational alignment twists with a characteristic period, or pitch (Fig. 2.3d).

Liquid crystals consist mostly of anisotropic organic molecules. *Thermotropic liquid crystals* exhibit a phase transition into the liquid-crystal phase as temperature is changed. *Lyotropic liquid crystals* exhibit phase transitions as a function of both temperature and molecular concentration in a solvent (typically water). There are also *colloidal* and *amphiphilic liquid crystals*, where the particles are colloids or assemblies of amphiphiles. Liquid crystal phases are distinguished by their different optical properties (textures). The contrasting areas in the textures correspond to domains where the liquid crystal molecules are oriented in different directions. Within a domain, however, the molecules are ordered (see Chapter 11 for more details).

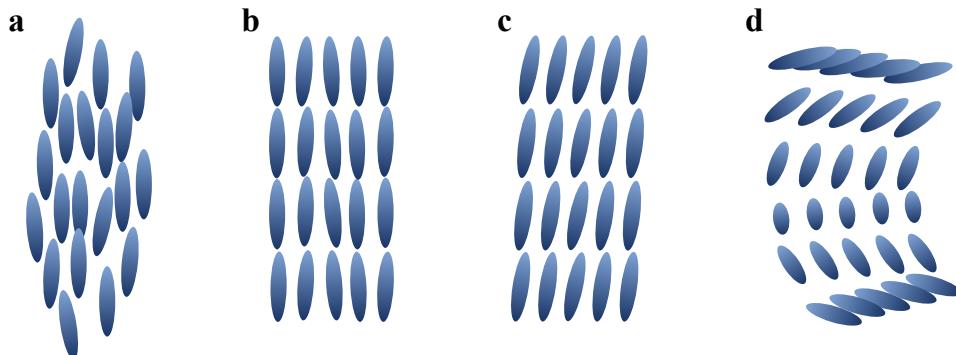


Figure 2.3: **Typical phases of thermotropic liquid crystals.** As their temperature is lowered, thermotropic liquid crystals explore the (a) nematic, (b) smectic A, (c) smectic C, and (d) cholesteric phases. This sequence of phases is not universal and depends on the details of the liquid crystal particles.

Examples of liquid crystals can be found both in the natural world and in technological applications. Widespread liquid crystal displays (LCD) use liquid crystals, since their optical properties are easily controlled by electric fields. Lyotropic liquid-crystalline phases are abundant in living systems. For example, many proteins and cell membranes are liquid crystals. The classic book “The Physics of Liquid Crystals” [6] by de Gennes (now in its second edition co-authored by Jacques Prost) extensively reviews recent advances in the field of liquid crystals.

2.3 Basic theoretical tools

As mentioned in Section 2.1, Landau theory is a powerful theoretical tool to study equilibrium soft matter systems that exhibit several phases. The concepts of order parameter, coarse graining, symmetry, and symmetry breaking are at the core of this approach and become clearer through the study of some examples. The first to be considered here is Landau theory for the ferromagnetic transition, while the second is de Gennes generalization to describe the nematic-isotropic transition. There are numerous derivations of Landau theory in the literature, ranging from very simple to very sophisticated. The present one follows the notes of D. Tong [7].

2.3.1 Landau theory for the ferromagnetic–paramagnetic transition

Landau theory is phenomenological. It requires a priori knowledge of the order parameter. The order parameter of the paramagnetic–ferromagnetic transition is the magnetization, m , a macroscopic property that measures the ability of the material to create static magnetic fields and to interact with other magnetized materials. The magnetization can be varied by applying a magnetic field h and by changing its temperature T . A permanent magnet is a system that exhibits a non-zero magnetization when the applied field is zero. In a ferromagnet (like nickel or iron), the features of the phase transition may be summarized as follows:

- (a) if a permanent magnet ($m \neq 0$) is heated at zero applied field ($h = 0$), it becomes non-magnetized ($m = 0$) above a well-defined temperature (the *Curie temperature* or *critical temperature* T_c ; Fig. 2.4a); when approaching T_c from below, the magnetization varies abruptly as a power law: $m \sim (T_c - T)^\beta$, with $\beta \approx 1/3$.
- (b) if a magnet is subjected to a decreasing applied field h while the temperature is constant at $T < T_c$, it remains magnetized down to $h = 0$ in the direction of the applied field; if the field h is then increased infinitesimally in a different direction, the magnetization keeps its magnitude but discontinuously changes to the new direction of the field (the blue line in Fig. 2.4b).
- (c) if the experiment described in (b) is repeated at $T > T_c$, then, when $h \rightarrow 0$, m is proportional to h and vanishes (green line in Fig. 2.4b).

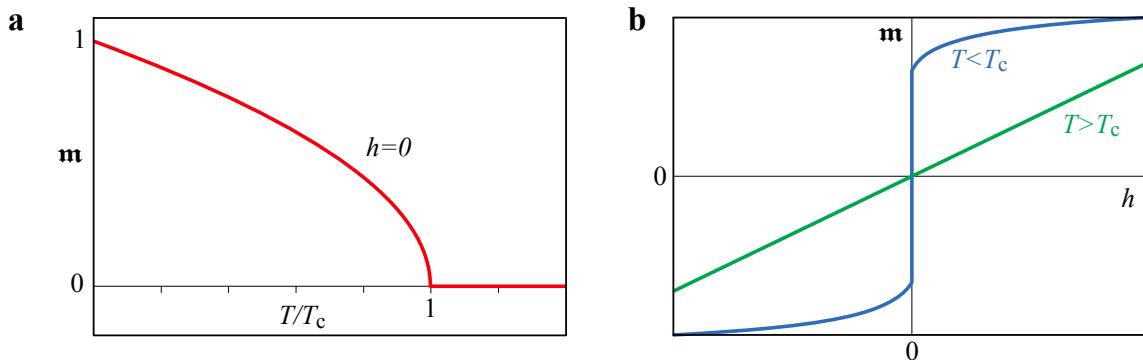


Figure 2.4: **Magnetization as a function of temperature.** (a) Schematic representation of the magnetization m as a function of temperature for $h = 0$. (b) Two isotherms at temperatures above (green line) and below (blue line) T_c ; notice the first-order phase transition for $h = 0$ at $T < T_c$.

The transition described in (a) is a second-order phase transition from a ferromagnetic ($m \neq 0$) to a paramagnetic ($m = 0$) state that occurs at $h = 0$ and $T = T_c$; the phase transition described in (b) is a first-order phase transition between two ferromagnetic states that exists for $T < T_c$ only.

The microscopic origin of magnetism in materials is the spin of the electrons. As a consequence, ferromagnetic systems are modeled by considering only the spin degrees of freedom s_i placed at the atomic positions on a lattice. For a uniaxial ferromagnet, we assume that each spin takes one of two values: $s_i = \pm 1$. The partition function of the ferromagnet is then:

$$Z(N, T, h) = \sum_{\{s_i\}} \exp \left(-\beta \mathcal{H}(\{s_i\}) + \beta h \sum_i s_i \right), \quad (2.2)$$

where $\beta^{-1} \equiv k_B T$, $\sum_{\{s_i\}}$ is a sum over all possible configurations of the N spins, and $\mathcal{H}(\{s_i\})$ is the Hamiltonian of the system. The Hamiltonian represents the interactions between the spins. Thermodynamics is obtained from Eq. (2.2) since the (equilibrium) free energy is $F_{\text{eq}} = -\beta^{-1} \ln Z$.

Phase diagram

The first step of Landau theory is to define an order parameter and to approximate the partition function by an explicit function of it. The order parameter is simply the magnetization m , the macroscopic quantity that changes at the phase transition described above. The magnetization of a given configuration of the spins is $m = \sum_i s_i / N$. The sum in the partition function of Eq. (2.2) may be rearranged so that m appears explicitly:

$$Z(N, T, h) = \sum_m \sum_{\{s_i\}|m} \exp(-\beta \mathcal{H}(\{s_i\}) + \beta N h m), \quad (2.3)$$

where \sum_m represents the sum over all possible values of the magnetization, and $\sum_{\{s_i\}|m}$ is the sum over all the configurations with magnetization m . The second sum in Eq. (2.3) is interpreted as the partition function of a system with magnetization m so that:

$$\sum_{\{s_i\}|m} \exp(-\beta \mathcal{H}(\{s_i\})) = \exp(-\beta F(N, m, T)). \quad (2.4)$$

The function $F(N, m, T)$ is a generalized free energy (not the equilibrium free energy) of a system with N spins, at temperature T , $h = 0$ with magnetization m (not the equilibrium magnetization, m_{eq}). It is further assumed that F is an extensive function of N , so that the relation between F and the equilibrium free energy of the system Eq. (2.2) is:

$$\beta F_{\text{eq}}(N, T, h) = -\ln \left(\sum_m \exp(-\beta N(f(m, T) - h m)) \right). \quad (2.5)$$

The equilibrium magnetization $m_{\text{eq}}(T, h)$ is related to Eq. (2.5) through the thermodynamic relation $m_{\text{eq}} \equiv -\partial(F_{\text{eq}}/N)/\partial h$.

Let us note that the procedure that leads from Eq. (2.2) to Eq. (2.5) is an example of coarse-graining: a system that was described by a set of *microscopic* variables (the individual spins) is now described by a related *mesoscopic* quantity, the magnetization.

The Landau approach proceeds by noting that for large values of N and low temperatures, the dominant term in the sum in Eq. (2.5) is obtained for the value of m that minimizes the argument of the exponential. That value, m_0 , is determined, as a function of h and T by:

$$h = \left(\frac{\partial f}{\partial m} \right)_{T, m=m_0}, \quad (2.6)$$

together with the condition:

$$\left(\frac{\partial^2 f}{\partial m^2} \right)_{T, m=m_0} > 0. \quad (2.7)$$

If the sum in Eq. (2.5) is approximated by its dominant term (saddle point approximation), the equilibrium free energy becomes:

$$\frac{F_{\text{eq}}(N, T, h)}{N} \equiv f_{\text{eq}}(T, h) = f(m_0(T, h), T) - h m_0(T, h). \quad (2.8)$$

It is also possible to show that $m_{\text{eq}}(T, h) = m_0(T, h)$ within this approximation. Therefore an approximation for the thermodynamics of the system is obtained when an approximation for $f(m, T)$ is adopted.

Sufficiently close to the critical point, the magnetization is not very large, so $f(m, T)$ can be expanded in powers of m . In the next step, symmetry makes its appearance: the symmetries of the microscopic description (i.e. of the Hamiltonian) have to be reflected in the expansion. The Hamiltonian is invariant when all the spins are reversed $\{s_i\} \rightarrow \{-s_i\}$. This symmetry implies that $f(m, T) = f(-m, T)$ and thus the free energy $f(m, T)$ is expanded in even powers of m :

$$f(m, T) = f_0(T) + a_2(T)m^2 + a_4(T)m^4 + \dots \quad (2.9)$$

In order to ensure that $f(m, T)$ has at least one minimum, Eq. (2.9) is cut off beyond the 4th order with $a_4(T)$ constant and > 0 ($a_4(T) = b$). Moreover, in the vicinity of the critical point, $a_2(T)$ may be approximated by $a_2(T) = a(T - T_c)$, with $a > 0$. Therefore, the function $f(m, T)$ is finally given by:

$$f(m, T) = f_0(T) + a(T - T_c)m^2 + bm^4. \quad (2.10)$$

In Fig. 2.5, $f(m, T) - f_0(T) - mh$ is plotted as a function of m for several values of h and T . Clearly, all the curves have at least one minimum, as required.

It is now useful to separate the results for $h = 0$ and $h \neq 0$.

Zero field, $h = 0$

The equilibrium magnetization as a function of temperature for $h = 0$ is easily obtained using Eq. (2.10) in Eq. (2.6):

$$m_{\text{eq}}(T) = \begin{cases} 0 & T > T_c, \\ \pm \sqrt{\frac{-a(T_c - T)}{2b}} & T \leq T_c. \end{cases} \quad (2.11)$$

This solution (Fig. 2.5a) corresponds to two symmetric minima when $T < T_c$, and one minimum at $m = 0$ when $T > T_c$. This is in line with a second-order phase transition as observed in experiments. The sign \pm signals the spontaneous symmetry breaking nature of the transition: While the Hamiltonian and the free energy $f(m, T)$ are invariant under the transformation $\{s_i\} \rightarrow \{-s_i\}$ and $m \rightarrow -m$, respectively, the system breaks this symmetry by aligning the spins in one of the two equivalent directions.

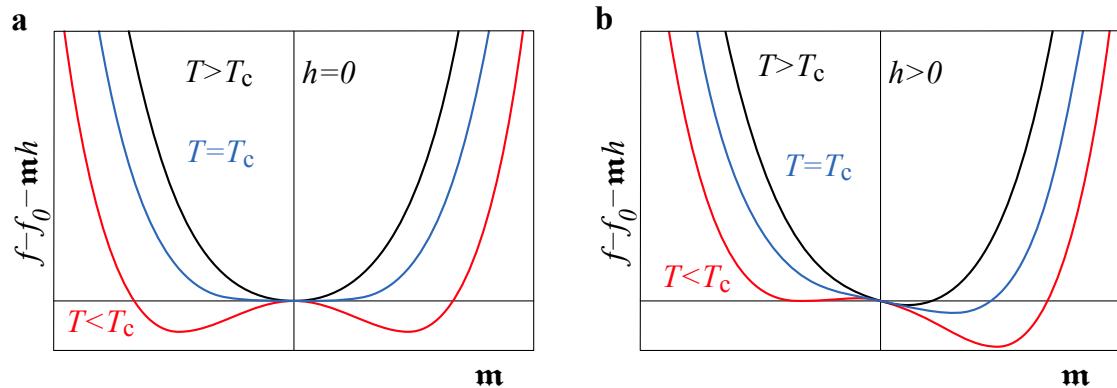


Figure 2.5: Landau free energy as a function of magnetization for different temperatures and magnetic field intensities. Schematic representation of $f(m, T) - f_0(T) - mh$ as a function of m for several values of h and T (Eq. (2.10)) for (a) $h = 0$ and (b) $h \neq 0$. Black, blue and red curves correspond to $T > T_c$, $T = T_c$ and $T < T_c$, respectively. The minima are the equilibrium magnetizations for each case.

The equilibrium magnetization, Eq. (2.11), is introduced in Eq. (2.9) to obtain the equilibrium free energy per particle (or spin), $f_{\text{eq}}(T)$, from which the specific heat $C(T)$ can be calculated:

$$C(T) \equiv -T \frac{\partial^2 f_{\text{eq}}}{\partial T^2} = \begin{cases} -T \frac{\partial^2 f_0}{\partial T^2} & T > T_c, \\ -T \frac{\partial^2 f_0}{\partial T^2} + T \frac{a^2}{b} & T \leq T_c. \end{cases} \quad (2.12)$$

As expected, the specific heat is singular, reflected by the discontinuity at $T = T_c$.

Non-zero field, $h \neq 0$

When $h \neq 0$, the equilibrium magnetization is obtained from Eq. (2.6) as an implicit function of h and T at all temperatures:

$$2a(T - T_c)\mathbf{m}_{\text{eq}} + 4b\mathbf{m}_{\text{eq}}^3 = h. \quad (2.13)$$

The numerical solution for $\mathbf{m}_{\text{eq}}(h, t)$ reveals that:

- (a) for $T > T_c$, Eq. (2.13) has a unique solution for \mathbf{m}_{eq} ;
- (b) for $T < T_c$, Eq. (2.13) has one or two solutions for \mathbf{m} with opposite signs, i.e., $f(\mathbf{m}, T) - h\mathbf{m}$ has one or two minima (Fig. 2.5b).

Note that the solutions of Eq. (2.13) show the tendency of the magnetization to align with the field h .

The analysis of equation Eq. (2.13) in the limit $h \rightarrow 0$ indicates the existence of a first-order phase transition for $T < T_c$, in line with experiments. In fact, when $T > T_c$, both terms on the left-hand side of Eq. (2.13) have the same sign and therefore, when $h \rightarrow 0$, the equilibrium magnetization $\mathbf{m}_{\text{eq}} \rightarrow \frac{h}{2a(T - T_c)}$. On the other hand, when $T < T_c$, this limit is more subtle:

$$\mathbf{m}_{\text{eq}} \rightarrow \begin{cases} \sqrt{\frac{a(T_c - T)}{2b}} & \text{for } h \rightarrow 0^+, \\ -\sqrt{\frac{a(T_c - T)}{2b}} & \text{for } h \rightarrow 0^-, \end{cases} \quad (2.14)$$

which confirms that Landau theory predicts a discontinuous jump (a first-order phase transition) of the magnetization when (at fixed $T < T_c$) the magnetic field h changes direction.

Landau theory correctly predicts the existence of the ferromagnetic phase transition but the behaviour of the binodal near $T = T_c$ is characterized by the mean-field exponent $\beta = 1/2$, which differs from that observed experimentally.

2.3.2 Landau-de Gennes theory for the nematic-isotropic transition

We now consider a phenomenological approach to orientational uniaxial nematic ordering based on a tensor order parameter $\mathbf{Q}(\mathbf{r})$. This is a traceless second rank tensor with the symmetry of the order in a uniaxial nematic (see the book by de Gennes in Ref. [6] or Chapter 11 for a microscopic justification of this choice of order parameter).

The phase diagram

We factor out the scalar order parameter $S(\mathbf{r})$, which is the eigenvalue of $\mathbf{Q}(\mathbf{r})$ corresponding to the nematic director $\mathbf{n}(\mathbf{r})$:

$$Q_{ij}(\mathbf{r}) = \frac{S(\mathbf{r})}{2} (3n_i(\mathbf{r})n_j(\mathbf{r}) - \delta_{ij}). \quad (2.15)$$

In the spirit of the phenomenological Landau approach to phase transitions, a free energy functional of a uniaxial nematic is postulated in the form

$$F[\mathbf{Q}] = \int_V [f_{\text{LdG}}(\mathbf{Q}(\mathbf{r})) + f_{\text{el}}(\nabla \mathbf{Q}(\mathbf{r})) + f_{\text{H}}(\mathbf{r}, \mathbf{Q}(\mathbf{r}))] d^3 r, \quad (2.16)$$

where $f_{\text{LdG}}(\mathbf{Q}(\mathbf{r}))$ (with the subscript LdG standing for ‘‘Landau-de Gennes’’) is a bulk contribution, $f_{\text{el}}(\nabla \mathbf{Q}(\mathbf{r}))$ is an elastic contribution, and $f_{\text{H}}(\mathbf{r}, \mathbf{Q}(\mathbf{r}))$ is the contribution due to the presence of an external magnetic (or electric) field. The equilibrium $\mathbf{Q}(\mathbf{r})$ minimizes $F[\mathbf{Q}]$, and the minimum of F is the free energy of the system.

The contribution $f_{\text{LdG}}(\mathbf{Q}(\mathbf{r}))$ in Eq. (2.16) describes the nematic–isotropic phase transition in a spatially uniform system without external fields. Following the Landau phenomenological approach, f_{LdG} is expanded in a Taylor series in the scalar order parameter S . The series is truncated at the quartic term in S , which is sufficiently close to the transition (but of course there are higher order terms in the expansion). Since f_{LdG} is a scalar quantity, and \mathbf{Q} is a second rank tensor with $\text{Tr } \mathbf{Q} = 0$, f_{LdG} must contain only scalar combinations of the tensor order parameter: $\text{Tr } \mathbf{Q}^2 \sim S^2$ and $\text{Tr } \mathbf{Q}^3 \sim S^3$. The general form of f_{LdG} is then given by:

$$f_{\text{LdG}} = f_I + a(T) \text{Tr } \mathbf{Q}^2 - b \text{Tr } \mathbf{Q}^3 + c (\text{Tr } \mathbf{Q}^2)^2, \quad (2.17)$$

where f_I is the free energy density of the isotropic phase. The presence of the term $\sim S^3$ reflects the fact that the nematic states described by S and $-S$ are distinct, and therefore the free energy is not symmetric with respect to the transformation $S \rightarrow -S$. In general, the coefficients a, b , and c in the Landau-de Gennes expansion, Eq. 2.17, are temperature dependent. To simplify, we assume that b and c are temperature independent positive constants. For spatially uniform systems Eq. (2.17) describes the coexistence between the nematic and isotropic phases. To this end, we substitute Eq. (2.15) into expansion (2.17) and obtain:

$$f_{\text{LdG}} = f_I + \frac{3}{2} a(T) S^2 - \frac{3}{4} b S^3 + \frac{9}{4} c S^4, \quad (2.18)$$

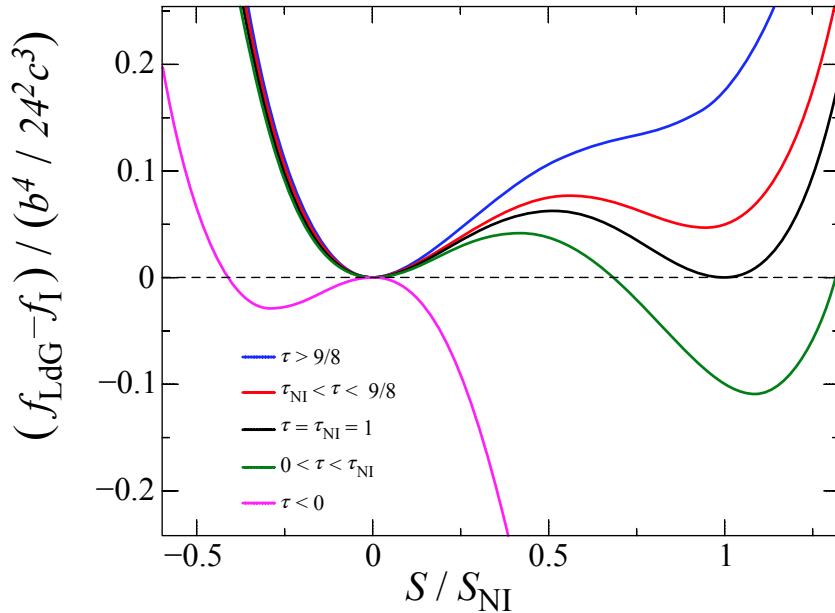


Figure 2.6: **Landau-de Gennes free energy density.** Landau-de Gennes free energy density, Eq. (2.18), for a uniaxial nematic as a function of the scalar order parameter S for several values of the reduced temperature $\tau = 24a(T)c/b^2$. At $\tau = \tau_{\text{NI}} = 1$ (black curve), the isotropic $S = 0$ and the nematic $S > 0$ phases coexist. For $1 < \tau < 9/8$ (red curve), the nematic phase is metastable, while for $\tau > 9/8$ (blue curve) the nematic phase is unstable. For $0 < \tau < 1$ (green curve), the isotropic phase is metastable, while for $\tau < 0$ the isotropic phase is unstable.

which should be minimized with respect to S . It is convenient to define the dimensionless temperature $\tau = 24a(T)c/b^2$. The equation $\partial f_{\text{LdG}}/\partial S = 0$ has three solutions: the isotropic phase $S_I = 0$; the nematic phase $S_N = \frac{b}{8c} \left(1 + \sqrt{1 - \frac{8\tau}{9}} \right) > 0$; and a third solution $S_3 = \frac{b}{8c} \left(1 - \sqrt{1 - \frac{8\tau}{9}} \right)$. This third solution should be discarded as it corresponds either to a free energy maximum, with $S_3 > 0$, or to a metastable minimum ($f_{\text{LdG}}(S_3) > f_{\text{LdG}}(S_N)$) with $S_3 < 0$ (Fig. 2.6). The transition temperature T_{NI} and the corresponding value S_{NI} of the order parameter are defined from the condition that the free energy densities of the two phases are equal $f_{\text{LdG}}(S_{\text{NI}}) = f_0$ which gives

$$\begin{cases} \tau_{\text{NI}} &= 1, \\ S_{\text{NI}} &= \frac{b}{6c}. \end{cases} \quad (2.19)$$

The isotropic phase is unstable for $\tau < 0$, while the nematic becomes unstable for $\tau > 9/8$.

The elastic free energy

The tensor order parameter \mathbf{Q} may depend on the spatial coordinates, which means that either the director \mathbf{n} or the scalar order parameter S (or both) vary from place to place. This variation may be due to external forces imposed on the system, thermal fluctuations, or boundary conditions. Assuming these deformations to vary slowly in space relative to the molecular scale, it is possible to describe the response of the nematic liquid crystal using continuum elastic theory. The elastic free energy density can be written as

$$f_{\text{el}} = L_1 \frac{\partial Q_{ij}}{\partial x_k} \frac{\partial Q_{ij}}{\partial x_k} + L_2 \frac{\partial Q_{ij}}{\partial x_j} \frac{\partial Q_{ik}}{\partial x_k} + L_3 \frac{\partial Q_{ij}}{\partial x_k} \frac{\partial Q_{ik}}{\partial x_j}, \quad (2.20)$$

where L_1 , L_2 , and L_3 are phenomenological constant elastic parameters and the summation convention is assumed. Deep in the nematic phase, when the scalar order parameter S is constant, an expansion in terms of the director \mathbf{n} is normally used to calculate the elastic free energy density. Substituting Eq. (2.15) into Eq. (2.20) and using the condition $n_i n_i = 1$, f_{el} may be written in this form (modulo full divergences):

$$\begin{aligned} f_{\text{el}} &= \frac{9S^2}{8} \left((2L_1 + L_2 + L_3)(\nabla \cdot \mathbf{n})^2 + 2L_1(\mathbf{n} \cdot [\nabla \times \mathbf{n}])^2 + (2L_1 + L_2 + L_3)[\mathbf{n} \times [\nabla \times \mathbf{n}]]^2 \right) \\ &= \frac{1}{2} \left(K_1 (\nabla \cdot \mathbf{n})^2 + K_2 (\mathbf{n} \cdot [\nabla \times \mathbf{n}])^2 + K_3 [\mathbf{n} \times [\nabla \times \mathbf{n}]]^2 \right). \end{aligned} \quad (2.21)$$

The second line represents the famous Frank-Oseen elastic free energy density for nematics with the splay K_1 , twist K_2 , and bend K_3 elastic constants:

$$K_1 = K_3 = \frac{9S^2}{4}(2L_1 + L_2 + L_3), \quad K_2 = \frac{9S^2}{4}2L_1. \quad (2.22)$$

For the purpose of qualitative calculations it is sometimes useful to assume that $K_1 = K_2 = K_3 = K$ (*one elastic constant approximation*). The elastic free energy density in this case reduces to

$$f_{\text{el}} = \frac{1}{2}K((\nabla \cdot \mathbf{n})^2 + (\nabla \times \mathbf{n})^2). \quad (2.23)$$

The Frank-Oseen elastic free energy of nematic liquid crystals is the starting point to discuss nematic textures and topological defects, which will be done in Chapter 11.

We close by noting that the tools discussed above are limited to equilibrium Landau-like theories for the order-disorder transitions. Even at equilibrium, a complete description of these phases requires the calculation of correlation and response functions, which may be done within the same framework. The most comprehensive reference on these topics is the classic book by Chaikin and Lubensky in Ref. [8]. Other books, on specific systems or on more advanced techniques, are included in the bibliography and should satisfy the needs of most readers. Last but not least, although active matter is generally soft, it is always out of equilibrium. The chapters that follow are the most up to date, pedagogical, introduction to the vast number of theoretical concepts, computational and experimental techniques used in current research in this very rapidly expanding field.

2.4 Further readings

There are numerous books, reviews, and articles about the topics discussed in this Chapter. For an introduction to Soft Matter, we recommend Refs. [9, 10, 8]. For some on polymers, we recommend Refs. [5, 11, 12]. For details on the theoretical tools, we recommend Refs. [6, 8, 10, 13, 14, 7].

Bibliography

- [1] J.-F. Joanny and M. Cates. Pierre-Gilles de Gennes. 24 October 1932 – 18 May 2007. *Biographical Memoirs of Fellows of the Royal Society*, 66:143–158, 2019.
- [2] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman lectures on physics; New millennium ed.* Basic Books, New York, NY, 2010. Originally published 1963-1965.
- [3] The Nobel Foundation. The Nobel Prize in Physics 1915, 1915. Accessed on Feb 8, 2025.
- [4] I. Müller. *A History of Thermodynamics: The Doctrine of Energy and Entropy*. Springer-Verlag, Berlin Heidelberg, 2007.
- [5] P. G. de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.
- [6] P. G. de Gennes and J. Prost. *The Physics of Liquid Crystals*. Oxford University Press, 1993.
- [7] D. Tong. Lectures on statistical field theory. <http://www.damtp.cam.ac.uk/user/tong/sft.html>, 2017.
- [8] P. M. Chaikin and T. C. Lubensky. *Principles of condensed matter physics*. Cambridge University Press, 2000.
- [9] M. Doi. *Soft Matter Physics*. Oxford University Press, 2013.
- [10] M. Kleman and O. D. Lavrentovich. *Soft Matter Physics: An Introduction*. Springer, 2003.
- [11] M. Doi. *Introduction to Polymer Physics*. Oxford University Press, 1996.
- [12] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Oxford University Press, 1986.
- [13] M. Kardar. *Statistical Physics of Particles*. Cambridge University Press, 2007.
- [14] M. Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007.

Chapter 3

Nonequilibrium Physics of Small Systems

FELIX RITORT

Non-equilibrium systems are ubiquitous in Nature, encompassing diverse entities from stars to microbes. A vivid representation of this is depicted in Fig. 3.1, showcasing the high-resolution structure of the 70S ribosomal unit from *E. coli*, a complex subject to non-equilibrium dynamics at a molecular scale. When systems deviate from equilibrium, they exhibit non-zero currents of conserved quantities, leading to entropy production. This Chapter delves into the intricacies of non-equilibrium phenomena, especially in small systems, where the traditional understanding of entropy can be challenged — a field of research often called *nonequilibrium thermodynamics of small systems* [2] or *stochastic thermodynamics* [3, 4]. From the erratic Brownian motion observed by Robert Brown to advanced nonequilibrium thermodynamics involving small systems, we will explore stochastic thermodynamics and its applications. With advancements in technology, our ability to observe and measure these phenomena has surged, shedding light on the nuanced behaviors of small systems. This Chapter presents an overview of these concepts, complemented by selected examples and problems to provide

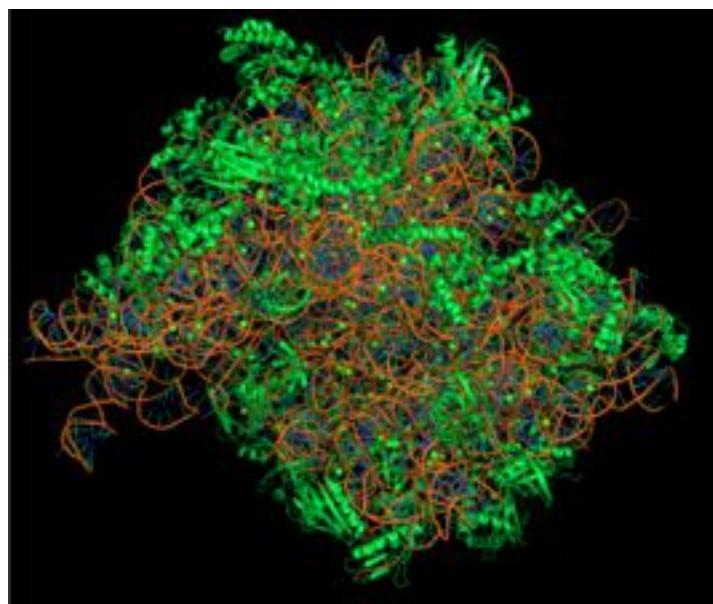


Figure 3.1: **High-Resolution Structure of the 70S Ribosomal Unit from *E. coli*.** The ribosome, a central player in protein synthesis, is conserved across prokaryotes. This complex, weighing between 2.3 and 4 Mega Daltons, consists of both RNA and protein. Given that one nucleotide or amino acid is approximately 100 Daltons, this intricate structure comprises tens of thousands of nucleotides and amino acids. The image showcases the ribosome obtained using cryo-electron microscopy at 2 Å resolution. Reproduced from Ref. [1].

a comprehensive understanding of the nonequilibrium physics of small systems.

3.1 Nonequilibrium phenomena

Natural systems, ranging from the grandeur of stars to the intricacy of microbes, typically exist in a state of nonequilibrium. Such a state is characterized by the presence of non-zero currents of any conserved quantity, be it mass, charge, momentum, or energy. Consider a simple example: a block resting on the floor. When subjected to a pulling force, friction at the block's contact point produces heat, resulting in a flow of heat towards the floor. The exerted mechanical power is irreversibly transformed into heat dQ during a time dt , producing entropy at a rate $\sigma = \frac{1}{T}dQ/dt \geq 0$, where T is the temperature of the environment. Nonequilibrium systems have a positive entropy production rate σ , whereas in equilibrium $\sigma = 0$. If the opposite is observed, i.e., heat spontaneously flowing from the floor to the block, the entropy production is negative during dt , $\sigma < 0$. Such events are rare. If observed, they are evidence that entropy must have been produced elsewhere to compensate for such a decrease. According to the second law of thermodynamics, the global balance of entropy production in the universe is always positive.

Furthermore, when the manipulated object is sufficiently small, such as a charged dust particle influenced by an electric field and exposed to friction, or when the time interval dt is exceptionally brief, the rate of entropy production can occasionally become negative. This implies that heat can spontaneously transfer from the environment to the object. As the size diminishes and the duration shortens, these anomalous events increase in frequency, leading to greater fluctuations in entropy production. In the extreme, when the timescale drops below that of molecular collisions, irreversibility diminishes and motion starts to resemble a reversible process.

These behaviors emerge from the so-called Brownian motion [5]. In 1827, Robert Brown, a botanist well known for his detailed descriptions of the cell and contributions to plants' taxonomy, made an important discovery during his pollination studies. While examining the motion of the grains of pollen suspended in water through the microscope, he observed that motion to be erratic, as if the grains were alive. Instead, Brown observed the effect of the random collisions of the water molecules against the pollen grains. Being kicked from all directions, the pollen grains jiggle erratically in the water solution. Statistical physics builds on the atomistic nature of matter and the motion-like origin of heat and work. The equipartition law sets $k_B T$ (k_B being Boltzmann's constant) as the energy scale where molecular motion and fluctuations are observed [6]. The branch of physics that studies nonequilibrium processes in systems with a few degrees of freedom is called *nonequilibrium thermodynamics of small systems* [2] or *stochastic thermodynamics* [3, 4]. Prominent examples are colloidal particles captured in optical traps, biomolecules pulled under mechanical forces, and single-electron transistors. With the development of high-temporal resolution cameras, photomultipliers, and photodetectors for light detection, our ability to measure energy fluctuations has recently expanded, as we will see in the rest of this Chapter.

3.2 Small systems and single-molecule experiments

Single-molecule biophysics has become a fabulous playground to investigate the nonequilibrium processes in small systems in physics and biology [7, 8]. Several techniques permit us to manipulate individual biological molecules one at a time, such as nucleic acids and proteins [9, 10, 11, 12, 13, 14]. Living matter is soft and stabilized by weak molecular forces of entropic origin, such as electrostatic and hydrophobic, of typical energies on the order of $k_B T$. Since these energies are comparable to the thermal forces, biological matter is suitable for observing large fluctuations and negative entropy events. Laser optical tweezers for single-particle and single-molecule manipulation have become a crucial technique in the field [15]. Invented by Arthur Ashkin in Bell Labs in 1970 (who was awarded the Nobel Prize in Physics in 2018 for it), optical tweezers have revolutionized physics, chemistry, and biology research.

Optical traps for single-molecule manipulation are produced by focusing an infrared beam inside a fluidics chamber, optically trapping a micrometer-sized bead, and measuring the force from the deflected light using position-sensitive detectors. In particular, counter-propagating traps measure forces by directly measuring the change in light momentum [16].

Take a single DNA molecule and pull from its extremities while recording the force-extension curve until it gets fully straightened. This thought experiment, a dream a few decades ago, has become standard in many

research institutes worldwide. Labeling the ends of a DNA molecule with specific chemical groups (e.g., biotin, avidin, digoxigenin) makes it possible to tether a single DNA between two beads [17]. Moving one bead relative to the other and using it as a force sensor makes it possible to measure a single biopolymer's *force-extension curve*, from DNA to RNA and proteins. In single-trap setups, one bead is immobilized in a pipette by air suction, while the other is captured in an optical trap and is used to measure the force exerted on the molecule as a function of the trap position. Fig. 3.2 shows the force-extension curve obtained by pulling a 24 kb fragment of the DNA of bacteriophage lambda, a virus that infects bacteria, and a model in molecular biology for decades. The measured elastic response is described by semiflexible polymer models, such as the *worm-like chain* (WLC) model, a simplified version of the more general *elastic rod model* [18]. The key ingredient of the WLC is the bending stiffness A , or the energy cost to locally bend the polymer, which is proportional to the persistence length ℓ^* , $A = k_B T \ell^*$. The energy of a configuration in the WLC is given by

$$E = \frac{A}{2} \int_0^L ds \left(\frac{\partial \hat{\mathbf{t}}}{\partial s} \right)^2, \quad (3.1)$$

where L is the contour length, s is the arc coordinate, and $\hat{\mathbf{t}}(s)$ is the tangent unit vector field.

Example 3.1: Persistence length ℓ^* in the WLC model. Let us consider Eq. (3.1).

a. Compute analytically the partition function $\mathcal{Z} = \int \mathcal{D}\hat{\mathbf{t}}(s) \exp(-\beta E)$ either by Fourier transforming the tangent vector field or by discretizing the polymer chain and mapping the resulting discrete model to a classical 1D Heisenberg model without external field.

b. Compute the tangent unit vector correlation function in the thermodynamic limit $L \rightarrow \infty$, $C(s) = \langle \hat{\mathbf{t}}(0)\hat{\mathbf{t}}(s) \rangle$, demonstrating that $C(s) = \exp(-s/\ell^*)$. This shows that the persistence length is the distance along the contour length over which directionality is lost.

c. Finally, demonstrate that the polymer end-to-end distance, defined as $\mathbf{R} = \int_0^L ds \hat{\mathbf{t}}(s)$ satisfies, $\langle \mathbf{R} \rangle = 0$ and $\langle \mathbf{R}^2 \rangle = 2L\ell^*$. [Hint: Show first that $\langle \mathbf{R}^2 \rangle = 2L \int_0^L ds C(s)$.]

Measurements of ℓ^* in biopolymers are done with electrophoresis, light diffraction, atomic force microscopy, and fluorescence imaging. Accurate estimations of ℓ^* are obtained by measuring the force-extension

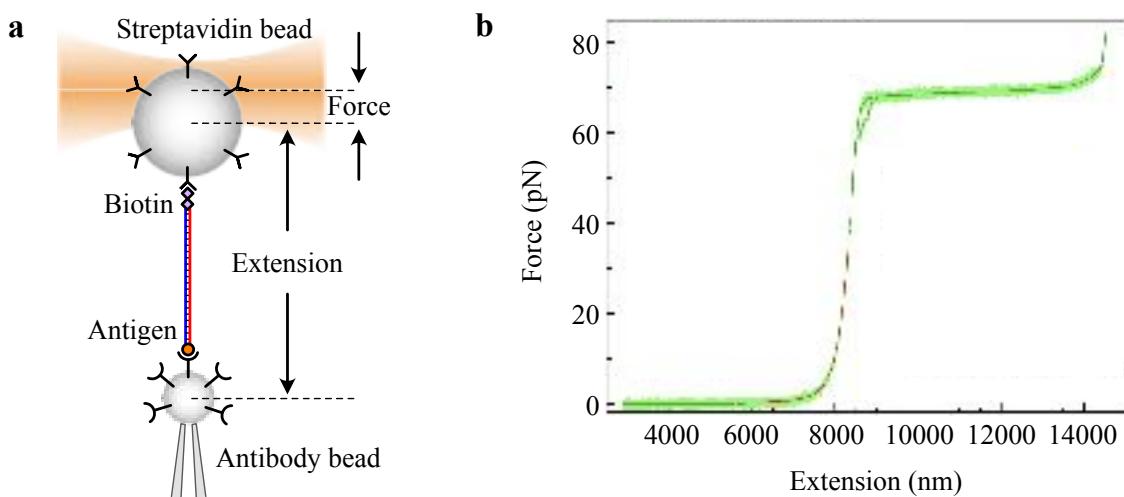


Figure 3.2: **Force-extension curve for DNA.** (a) Experimental setup in a DNA pulling experiment. (b) The force-extension curve was measured by pulling a 24kb DNA molecule (half λ -DNA) at standard conditions ($T = 298\text{K}$ and $1\text{M NaCl Tris buffer}$). Three force regimes are identified. At forces below 10pN , the DNA random coil is extended against thermal fluctuations (*entropic regime*). At forces between 10 and 60pN , the DNA is stretched above its contour length ($8.3\text{ }\mu\text{m}$) (*enthalpic regime*). Above 65pN , the DNA is overextended by $\sim 70\%$ of its contour length.

curve in single-molecule pulling experiments and fitting it to the WLC model. Marko and Siggia derived a formula interpolating low and high forces two decades ago [19] (Problem 3.1). It is given by

$$f(x) = \frac{k_B T}{\ell^*} \left(\frac{1}{4(1-x/L)^2} - \frac{1}{4} + \frac{x}{L} \right). \quad (3.2)$$

It is common to include the extensibility of the polymer's contour length by introducing a stretching term in Eq. (3.1). This leads to the *extensible WLC*, described by Eq. (3.2) with the rescaling $L \rightarrow L(1+f/Y)$, where Y is the Young modulus [20]. The resulting implicit equation can be inverted to obtain $x(f)$ by mapping it to a third-degree equation [21]. Fig. 3.2 shows a WLC fit (red line) to the experimental data (green line) between 0 and 40 pN that reproduces the data pretty well. Values at standard conditions (298 K, 1 M NaCl) are, e.g., $\ell^* = 50$ nm, $Y = 1000$ pN for double-stranded DNA (dsDNA), $\ell^* = 0.75$ nm for single-stranded DNA (ssDNA) [22, 23] and single-stranded RNA (ssRNA) [24], $\ell^* = 0.6$ nm for polypeptide chains [Provide value of Y for all cases]. The WLC model does not consider self-exclusion effects, which are essential at low salts [25, 26], and other models are needed, such as the *thick-chain model* [27, 28]. Elastic measurements have also been performed at different temperatures using a temperature-controlled optical trap [29]. While for dsDNA P decreases linearly with T , it increases linearly with T for ssDNA [30]

3.3 Nonequilibrium force spectroscopy

Besides measuring polymer elasticity, single-molecule experiments provide a tool to investigate molecular folding of nucleic acids and proteins, DNA–protein and DNA–peptide interactions (e.g., intercalation, condensation, aggregation phenomena), protein–protein interactions (e.g., ligand–receptor binding), molecular motors (e.g., cellular transport, DNA–RNA polymerases, ATPases and proton pumps, viral packaging motors, topoisomerase, helicases). For example, folding thermodynamics and kinetics are investigated by unzipping molecular structures: Molecular ends are pulled apart to break the bonds that hold the native structure [31, 32]. Unzipping experiments on DNA hairpins of a few kilobases have permitted determining the hybridization energies of DNA [33, 34] and RNA [35, 36], as well as the binding sites of ligands [37] at single base pair resolution with 0.1 kcal mol⁻¹ accuracy. Mechanical unzipping has provided a single-molecule test of the functional renormalization group approach for disordered elastic systems [38, 39]. In Fig. 3.3, we show a typical unzipping

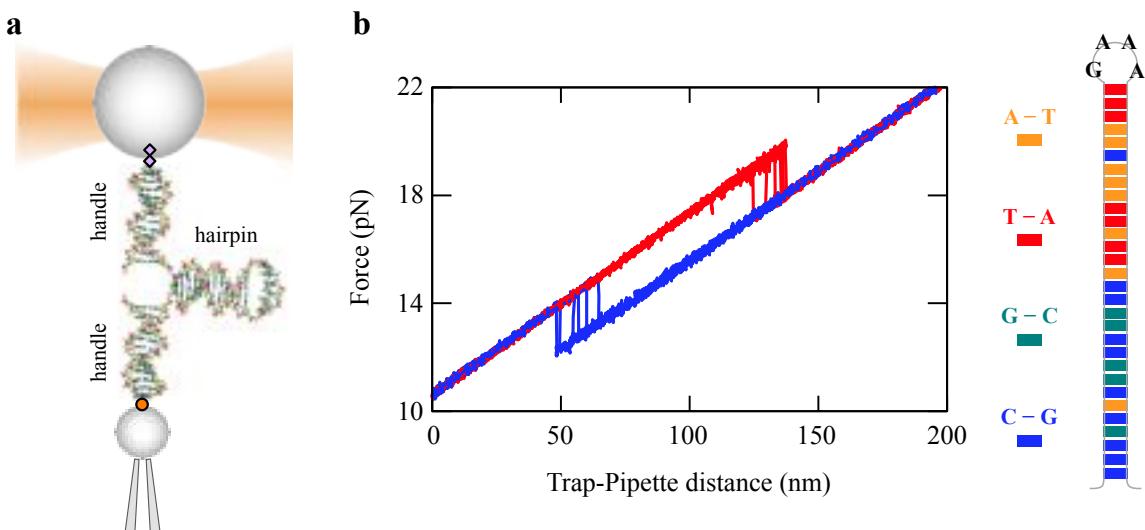


Figure 3.3: **Mechanical unzipping experiments.** (a) Experimental setup showing the optical trap, the flanking DNA handles, and the DNA hairpin (not to scale). (b) Force–distance curves measured during the mechanical unfolding and folding of DNA hairpins. The force rips correspond to unfolding (red) and folding (blue) transitions. (c) Sequence logo of the DNA hairpin.

experiment of a 20-bp-DNA hairpin (a stem-loop structure made of two complementary DNA strands ending in a loop). A molecular construct of the hairpin linked to dsDNA handles is tethered between two beads for unzipping [40]. Moving the optical trap relative to the pipette, we measure force and work that give valuable information about the folding process, such as free energy, kinetic rates, and folding pathways [41, 42].

To characterize mechanical unfolding and folding stability, it is customary to measure the rupture force distributions where the molecule unfolds and refolds for the first time. These distributions give information about the folding free-energy landscape of the molecule, such as the height of the kinetic barrier B , the folding free energy at zero force ΔG_0 , and the distances between the folded state and the transition state, x^\dagger , and the unfolded state and the transition state, x^* . A successful model that fits the experimental data is the *Bell-Evans model with force-dependent kinetic rates*,

$$k_{F \rightarrow U}(f) = k_0 \exp\left(\frac{fx^\dagger}{k_B T}\right), \quad (3.3)$$

$$k_{U \rightarrow F}(f) = k_0 \exp\left(\frac{\Delta G_0 - fx^*}{k_B T}\right), \quad (3.4)$$

where k_0 is the unfolding rate at zero force. These rates fulfill detailed balance, i.e., $k_{F \rightarrow U}(f)/k_{U \rightarrow F}(f) = \exp[(fx_m - \Delta G_0)/k_B T]$ with $x_m = x^\dagger + x^*$.

A first-order Markov process yields the first rupture force distributions and the survival probabilities of the folded and unfolded states during unfolding and refolding, respectively. Particularly useful are the mean and variance of the rupture forces that can be derived within a Gaussian approximation. By detecting cooperative changes in force and extension, force spectroscopy has become a powerful approach to identify intermediate states and pathways in molecular folding [43, 44].

Example 3.2: Analytical rupture forces in unzipping experiments. Consider a molecule that can be in two states (folded and unfolded) under an applied force. Unfolding and folding kinetic rates are given by Eqs. (3.3) and (3.4). Initially, at $f = f_{\min}$, the molecule is folded and the force is ramped at the pulling rate $r = \dot{f}$.

a. Write the master equation for the survival probability of the folded state during unfolding and derive the first rupture force distribution $\rho_{F \rightarrow U}(f)$,

$$\rho_{F \rightarrow U}(f) = \frac{k_0}{r} \exp\left(\beta x^\dagger f + \frac{k_0}{\beta r x^\dagger} [\exp(\beta x^\dagger f_{\min}) - \exp(\beta x^\dagger f)]\right) \quad (3.5)$$

[Hint: Express the master equation in terms of force rather than time.]

b. Use symmetry ($F \rightarrow U$) arguments to derive the corresponding expression for folding $\rho_{U \rightarrow F}(f)$, starting from a maximum force, f_{\max} , where the molecule is unfolded with probability one. Plot the two functions, $\rho_{F \rightarrow U}(f)$ and $\rho_{U \rightarrow F}(f)$. [Suggestion: To simplify the plots, take $f_{\min} = -\infty$ for $\rho_{F \rightarrow U}(f)$ and $f_{\max} = \infty$ for $\rho_{U \rightarrow F}(f)$.]

c. From Eq. (3.5), find the most probable rupture force f^* and the standard deviation σ_f of the rupture force distribution using a Gaussian approximation (i.e., expanding Eq. (3.5) to second order around its maximum). Demonstrate that

$$f^* = \frac{k_B T}{x^\dagger} \log\left(\frac{x^\dagger r}{k_0 k_B T}\right), \quad (3.6)$$

$$\sigma_f = \log\left(\frac{3 + \sqrt{5}}{2}\right) \frac{k_B T}{x^\dagger}. \quad (3.7)$$

d. Estimate the values of f^* and σ_f for the DNA hairpin of 20 bp shown in Fig. 3.3 and check they agree with what you see in the figure (take $T = 298$ K, $x^\dagger = 9$ nm, $k_0 = 10^{-16}$ s $^{-1}$, $r = 20$ pN s $^{-1}$).

3.4 Fluctuation theorems

Fluctuation theorems quantify the probability P of observing the rare negative entropy production events $P(-S_t)$ relative to the most probable positive ones $P(S_t)$. In their simplest form, they obey a simple mathematical relation [45],

$$\frac{P(S_t)}{P(-S_t)} = \exp\left(\frac{S_t}{k_B}\right), \quad (3.8)$$

where S_t is the entropy produced during time t . Eq. (3.8) shows that trajectories with $S_t < 0$ are exponentially suppressed relative to those with $S_t > 0$. Originally proven by Giovanni Gallavotti and Eddy Cohen for nonequilibrium steady states in the large t limit [46], Eq. (3.8) is also valid for transient states at any t . Transient states are generated by driving out of equilibrium a system initially in equilibrium at $t = 0$. Dynamics is modeled by the action of time-dependent external forces parametrized by $\lambda(s)$ with $0 \leq s \leq t$. For a given forward process (F), it is customary to define its time-reversed (R). In the reverse process, the system starts in equilibrium at $\lambda(t)$ and the control parameter evolves backward in time relative to the forward process, i.e., $\lambda_R(s) = \lambda(t-s)$. The *transient fluctuation theorem*, Eq. (3.8), takes the form

$$\frac{P_F(W_t)}{P_R(-W_t)} = \exp\left(\frac{W_t - \Delta G}{k_B T}\right), \quad (3.9)$$

where W_t equals the work exerted by the external driving forces during time t , and ΔG is the system's free energy difference between equilibrium states at $\lambda(0)$ and $\lambda(t)$. ΔG also equals the reversible work between the initial and final states. Eq. (3.9) is known as *Crooks fluctuation theorem* for work [47]. Eq. (3.9) can be related to Eq. (3.8) taking $S_t = W_d(t)/T$ with $W_d = W - \Delta G$ the difference between the actual work and the reversible work, also called dissipated work. Note the difference between Eq. (3.9) and Eq. (3.8): in the latter case, there is no distinction between forward and reverse. Another difference between Eq. (3.9) and Eq. (3.8) is the role of the so-called boundary terms. While Eq. (3.9) is exact for all times, the more general Eq. (3.8) is only valid in the large-time limit. For nonequilibrium steady-states, S_t grows linearly with t on average, defining the entropy production rate $\sigma = \langle S_t \rangle/t$. Eq. (3.8) takes the form

$$\sigma = k_B \lim_{t \rightarrow \infty} \frac{1}{t} \log \left(\frac{P(S_t)}{P(-S_t)} \right), \quad (3.10)$$

with finite-time corrections to σ on the order of $1/t$. To understand the origin of such corrections, we observe that the heat Q_t and the work W_t are related by energy conservation, i.e., $Q_t + W_t = \Delta U_t$, with the internal energy difference, $\Delta U_t = U_t - U_0$, a boundary term between the initial and final time. While Q_t and W_t grow on average linearly with t , ΔU_t does not, making the difference between Q_t and W_t finite in the large t limit, and of magnitude $1/t$ relative to either Q_t or W_t . Therefore, while S_t in Eq. (3.8) stands for the heat produced Q_t , W_t in Eq. (3.9) stands for the exerted work W_t . The difference between both quantities is a boundary term ΔU_t , mostly contributing to the tails of the distributions. These tails can be described in the framework of the *large deviations theory*. We are not going to address them here, but more about this in problem can be found in Ref. 3.2.

Eqs. (3.8) and (3.9) imply two results. First, as we said, trajectories with negative S_t are exponentially suppressed relative to the positive S_t ones. Suppression exponentially grows with time and size, due to the extensivity property of S_t with time and size. Second, a corollary of Eq. (3.9) is the *Jarzynski equality* [48]. Rewritten as $\exp(-W_t/k_B T)P_F(W_t) = P_R(-W_t)\exp(-\Delta G/k_B T)$ and integrating over W_t , we get $\langle \exp(-W_t/k_B T) \rangle = \exp(-\Delta G/k_B T)$ or $\langle \exp(-W_d(t)/k_B T) \rangle = 1$, where $\langle \dots \rangle$ denotes an average over many experimental realizations of the forward process. From Jensen's inequality ($\langle \exp(x) \rangle \geq \exp(\langle x \rangle)$, being x a random variable, it follows that $\langle W_d(t) \rangle \geq 0$, in agreement with the second law of thermodynamics. Similarly, one can prove that $\sigma \geq 0$ from Eq. (3.10). The beauty of Eqs. (3.8) and (3.9) lies in the fact that the second law is derived from a mathematical equality, rather than from the less informative Clausius inequality.

The first experimental test of Eqs. (3.8) and (3.9) was made in 2002 by Denis Evans and collaborators [49]. They measured the entropy production in a microsphere optically trapped and dragged through water. Initially equilibrated in a resting optical trap, the bead is driven out of equilibrium by instantaneously moving the optical trap at a constant velocity v . The work exerted on the bead up to time t is given by $W_t = W_d(t) = v \int_0^t f(s) ds$, where $\Delta G = 0$ because the bead-trap potential is translationally invariant. Note that the reverse

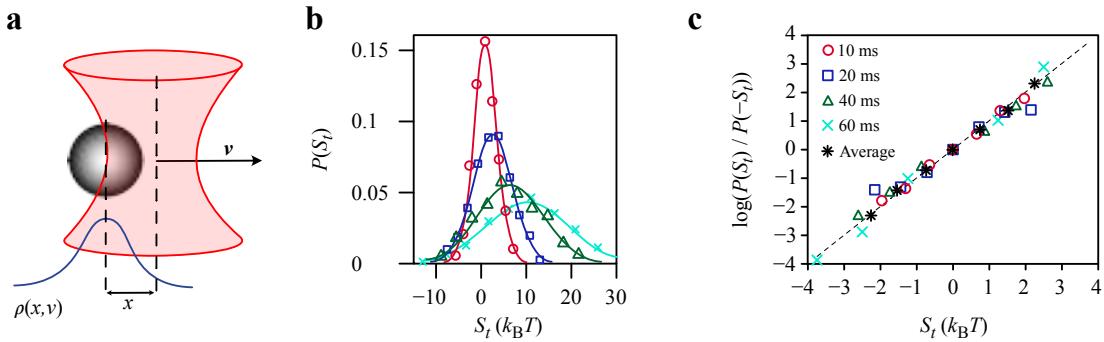


Figure 3.4: Bead in a trap dragged through water. (a) Experimental setup showing a bead captured in an optical trap moving at a speed $v \sim 1 \mu\text{m s}^{-1}$. (b) Entropy production distributions are measured at different times. (c) Experimental test of Eq. (3.9).

process requires to reverse the motion of the optical trap. For even traps, $V(x) = V(-x)$ and $P_F = P_R$. Evans and collaborators measured the entropy production distribution from the particle's trajectories finding S_t negative events as predicted by Eq. (3.9). An analytical proof of Eq. (3.9) can be obtained from the Langevin equation describing the overdamped motion of a bead captured in a harmonic potential moving in the lab frame at speed v :

$$\gamma \dot{x}(t) = -k(x(t) - x_0(t)) + \eta(t), \quad (3.11)$$

where η is a white noise of zero mean and correlation $\langle \eta(t) \eta(s) \rangle = 2k_B T \gamma \delta(t-s)$, k is the trap stiffness, and $x_0(t) = vt$ is the trap position. Similar results have been obtained for non-Markovian systems with memory effects [50]. Eq. (3.11) is linear and can be readily solved for an arbitrary initial condition finding $P(S_t)$.

Example 3.3: Entropy production of an optically trapped bead dragged through water. Consider a microsphere of radius R immersed in water and optically trapped in the focus of a laser (typically, in the near infrared to avoid light absorption by water). The trapping potential can be considered quadratic, i.e., $V(x) = \frac{1}{2}k(x - x_0)^2$, where x_0 is the trap center and k is the trap stiffness. The resulting Langevin equation of motion is Eq. (3.11).

a. Demonstrate that, in equilibrium (x_0 fixed), the autocorrelation function of the force, $C(t, s) = \langle f(t) f(s) \rangle$ with $f(t) = -k(x(t) - x_0)$, is given by

$$C(t, s) = C(t-s) = k_B T k \exp(-\omega_c(t-s)), \quad (3.12)$$

where $\omega_c = k/\gamma$ is the corner frequency of the bead that sets the relaxation timescale of the bead in the trap $\tau_r = 1/\omega_c = \gamma/k$.

b. Calculate the average and variance of the entropy production, $\langle S_t \rangle$ and $\sigma_{S_t}^2 = \langle S_t^2 \rangle - \langle S_t \rangle^2$, and show that both grow linearly with time. Check that they fulfill the fluctuation-dissipation relation at any t , i.e.,

$$\frac{\sigma_{S_t}^2}{2k_B \langle S_t \rangle} = 1. \quad (3.13)$$

[Hint: Use $f(t) = -k(x(t) - x_0)$ and $S_t = \frac{v}{T} \int_0^t f(s) ds$.]

c. Demonstrate that $P(S_t)$ is a Gaussian distribution and that this implies Eq. (3.13). [Hint: To proof Gaussianity note that Eq. (3.11) is linear in x .]

We have remarked that Eq. (3.8) is asymptotically valid in nonequilibrium steady states in the long-time limit. Problem 3.2 analyzes $P(S_t)$ for work W_t and heat Q_t in a steady state where the trap is moved at constant speed v . The main difference concerning the previous transient case is the initial condition. In the steady-state fluctuation theorem, the bead is already dragged by the moving trap at speed v . Instead, in the transient fluctuation theorem, the bead is initially in equilibrium in a trap at rest, which is suddenly put in motion

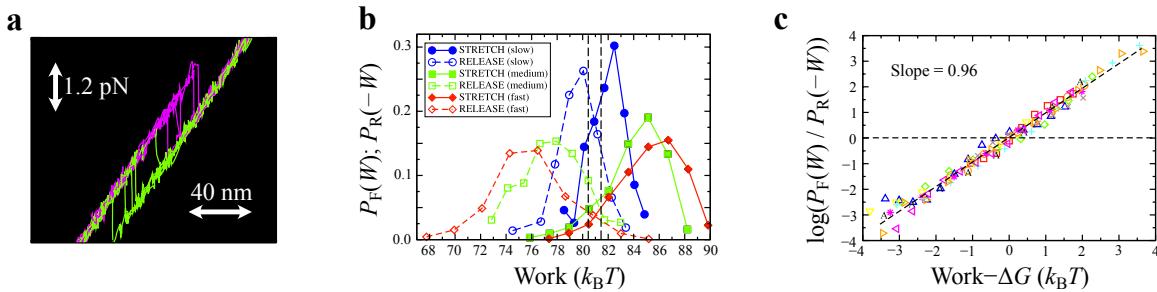


Figure 3.5: Free energy inference from nucleic acid hairpin pulling experiments. (a) Experimental force–distance curve of pulling a DNA hairpin with optical tweezers. The molecule opens around 16 pN with unfolding (magenta) and refolding (green) trajectories exhibiting hysteresis. (b) Unfolding or forward (continuous lines) and refolding or reverse (dashed lines) work distributions at three different pulling speeds. Hysteresis grows with the pulling speed. As predicted by Eq. (3.9), all work distributions cross at $\Delta G \sim 81 k_B T$. (c) Experimental test of Eq. (3.9). Data for different molecules and speeds collapse into a straight line of slope ~ 1 . Reproduced from Ref. [42]. Add references to all three panels of this figure in the text

at the constant speed v . In the steady-state fluctuation theorem, $P(S_t = W_t/T)$ is a Gaussian at all times, however Eqs. (3.8) and (3.13) only hold in the limit $t/\tau_r \gg 1$. In contrast, in the transient fluctuation theorem $P(S_t = W_t/T)$ is also a Gaussian and Eqs. (3.9) and (3.13) are exact at all times.

In general, for transient states (Eq. (3.9)), Jarzynski’s equality implies

$$\Delta G = -k_B T \log \left\langle \exp \left(-\frac{W}{k_B T} \right) \right\rangle, \quad (3.14)$$

showing that we can recover ΔG by exponentially averaging the irreversible work [51]. The average must be taken over an infinite number of experimental repetitions, the exponential average being strongly biased for a finite number of experiments [52]. Eq. (3.9) permits us to extract less biased free energies of native structures from bi-directional pulling experiments, i.e., by combining unfolding (forward) and folding (reverse) work measurements [53]. In particular, Eq. (3.9) predicts that $P_F(W)$ and $P_R(-W)$ cross at $W = \Delta G$. An extended version of Eq. (3.9) [54] has found important applications in molecular folding of nucleic acids [55, 56], proteins [57, 58], and ligand binding [59]. Fluctuation theorems have also been used to reconstruct folding free energy landscapes [60]. They have also been used to infer free energy differences from partial work measurements, i.e., when full work cannot be measured, in yet another application of fluctuation theorems to extract information about irreversible processes in small systems [61].

The experimental measurement of energy fluctuations in molecular systems will open new perspectives to explore living matter’s marvelous performance. The successful exploitation of negative entropy production events during evolution has led to molecular structures of astonishing efficiency [62]. By rectifying thermal fluctuations [63, 64], many enzymatic reactions (e.g., molecular motors powered by ATP hydrolysis [65] and light-harvesting complexes in photosynthetic reactions) have reached nearly 100% efficiencies. Such astonishing high efficiencies are not paralleled by man-built devices. The continuous remodeling of living matter by active weak molecular forces (operating in the $k_B T$ scale), with their large intrinsic deviations and negative entropy production events, may have shaped biological matter over millions of years of evolution. Such questions touch the core of biology and remain open for future research.

3.5 Energy, information and the Maxwell demon

In 1867, the renowned Scottish scientist James Clerk Maxwell, known for unifying the theories of electricity and magnetism, postulated a thought experiment challenging the second law of thermodynamics [67]. He envisioned a minuscule, intelligent entity — endowed with free will and an exceptional ability to perceive and manipulate individual molecules — capable of influencing the behavior of matter [68] — the eponymous *Maxwell demon*. In his thought experiment (Fig. 3.6a), two compartments of gas are separated by an adiabatic wall and connected with a small hole and a gate that can be opened and closed by the demon. By measuring

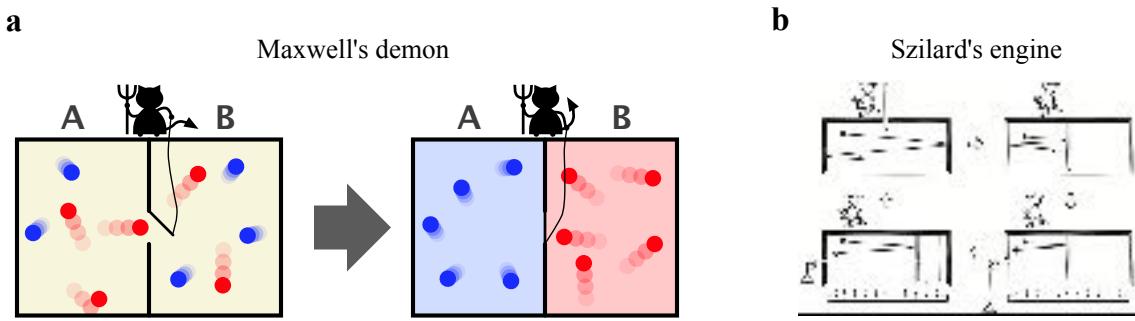


Figure 3.6: **Information in thermodynamics.** (a) Maxwell demon schematics. Slow-moving molecules (blue) are separated from fast-moving ones (red) into the left and right compartments, creating a temperature difference without any work expenditure. Image from Wikipedia in the CC BY-SA 3.0 license. (b) Szilard's engine schematics. The trajectory of a single particle in a two-compartment vessel is monitored. Work is extracted by inserting a wall and a pulley and isothermally expanding the chamber containing the particle. Reproduced from Ref. [66].

the speed of the individual molecules, the demon selectively opens and closes the gate to separate the fast and slow molecules, thus creating a temperature gradient between the two compartments. The demon can do this effortlessly without the expenditure of work, thereby violating the second law.

The solution to the paradox came in the 1960's from the theory of computing developed by Landauer and others [70]. They showed that recording information does not bring the system to its original state unless the information is erased. It is widely accepted that erasing bits produces heat that increases the entropy of the universe and restores the second law (see however Ref. [71]). Maxwell demon can be experimentally realized in the *Szilard engine* (Fig. 3.6b). A two-compartment box containing a single particle is in contact with a

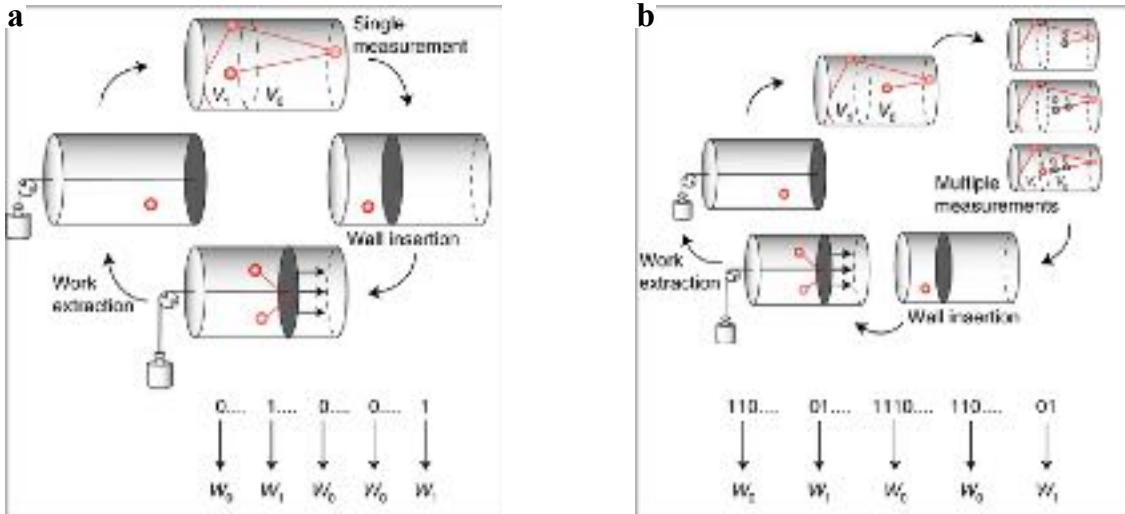


Figure 3.7: **Single-measurement versus multiple-measurements Maxwell demons.** (a) Stored sequences in the single-measurement Maxwell demons consist of one bit ($\sigma = 0, 1$) depending on whether the molecule is folded or unfolded. The average work per cycle is given by $W_\sigma = -k_B T \log p_\sigma$ where p_σ is the equilibrium probability of the measurement outcome σ . (b) Stored sequences in the multiple-measurement continuous Maxwell demon consist of strings of 2 or more bits such that all bits are identical until the last one which is different (e.g., 0001, 10, 11110, 00000001). Reproduced from Ref. [69].

thermal bath at temperature T . An observation is made, and depending on which compartment the particle is in, a movable wall and a pulley mechanism that lifts a weight are implemented, fully converting heat from the bath into work. As before, the maximum extracted work per cycle equals the cost of erasing one bit (also called the *Landauer limit*), restoring the validity of the second law. The Szilard engine has been experimentally realized in small systems such as colloidal particles [72, 73], electronic devices [74], and single molecules [69]. Crucial in the Szilard engine is the information content of the stored sequences \mathcal{I} [75], which must be larger than the average extracted work according to the second law.

A continuous version of the Szilard engine, dubbed *continuous Maxwell demon*, has been recently introduced where the particle's position is repeatedly monitored every time τ and work extracted only when the particle changes compartment [69, 76]. The average work per cycle has been shown to be larger than $k_B T \log 2$. Because the stored sequences contain two or more bits, their average information content \mathcal{I} is always larger than the Landauer limit $\log 2$, and the second law inequality, $W_{CMD} \leq k_B T I$, preserved (see Problem 3.3). The distinction between the single-measurement Maxwell demon and the multiple-measurement continuous Maxwell's demon is shown in Fig. 3.7.

Example 3.4: Information-to-energy conversion in the Maxwell demon. Consider the single-measurement and the multiple-measurement Maxwell demons depicted in Fig. 3.7.

- a. Show that the maximum average work that can be extracted when the particle is in compartment $\sigma = 0, 1$ (left or right) is given by $W_\sigma = -k_B T \log p_\sigma$, p_σ being the probability to observe the particle in state σ ($p_0 + p_1 = 1$).
- b. Demonstrate that the maximum average work per cycle that can be extracted in the single-measurement Maxwell demon is given by

$$W_{MD} = -k_B T (p_0 \log p_0 + p_1 \log p_1), \quad (3.15)$$

whereas for the continuous Maxwell demon with repeated measurements every time τ it is given by

$$W_{CMD} = -k_B T (p_0 \log p_1 + p_1 \log p_0). \quad (3.16)$$

Explain why W_{CMD} is independent of τ .

b. Demonstrate that $W_{MD} \leq W_L \leq W_{CMD}$ where $W_L = k_B T \log 2$ is the Landauer limit. Show that, while W_{MD} vanishes in the limit $p_0 \rightarrow 0, 1$, W_{CMD} diverges in the same limit. Interpret this result.

c. Fig. 3.9 shows the experimentally measured work per cycle distributions in the Maxwell demon and continuous Maxwell demon. Can you explain the origin of the negative work events observed for the latter? Note that a negative extracted work event means that the extraction process is inefficient, i.e., heat is dissipated rather than work being extracted.

The continuous Maxwell demon has been experimentally implemented in single DNA hairpins mechanically unfolded and folded with optical tweezers (Fig. 3.8) and the information content of the stored sequences analytically calculated for arbitrary τ . In the limit of uncorrelated measurements $\tau \rightarrow \infty$, the average information content has been demonstrated to be minimum and larger than the extracted work, in agreement with the second law. Problem 3.3 asks to calculate the average information content per cycle of the stored sequences in the continuous Maxwell demon in the limit $\tau \rightarrow \infty$, showing that it is larger than the average work per cycle, in agreement with the second law. A new class of generalized continuous Maxwell demon models that combine Szilard and continuous-Maxwell-demon-type protocols has been recently introduced demonstrating that measurement correlations enhance information to energy conversion efficiency [77, 78]. The continuous Maxwell demon has been also extended to N states [79].

Example 3.5: Work distributions in Maxwell demons: insights into negative work events. Fig. 3.9 shows the experimentally measured work per cycle distributions in the Maxwell demon and continuous Maxwell demon. Can you explain the origin of the negative work events observed for the latter? Note that a negative extracted work event means that the extraction process is inefficient, i.e., heat is dissipated rather than work being extracted.

The Szilard engine has endowed information of a thermodynamic meaning, extending fluctuation theorems to the case of feedback information [80, 81, 82]. Feedback finds multiple applications in physics [83, 84] and

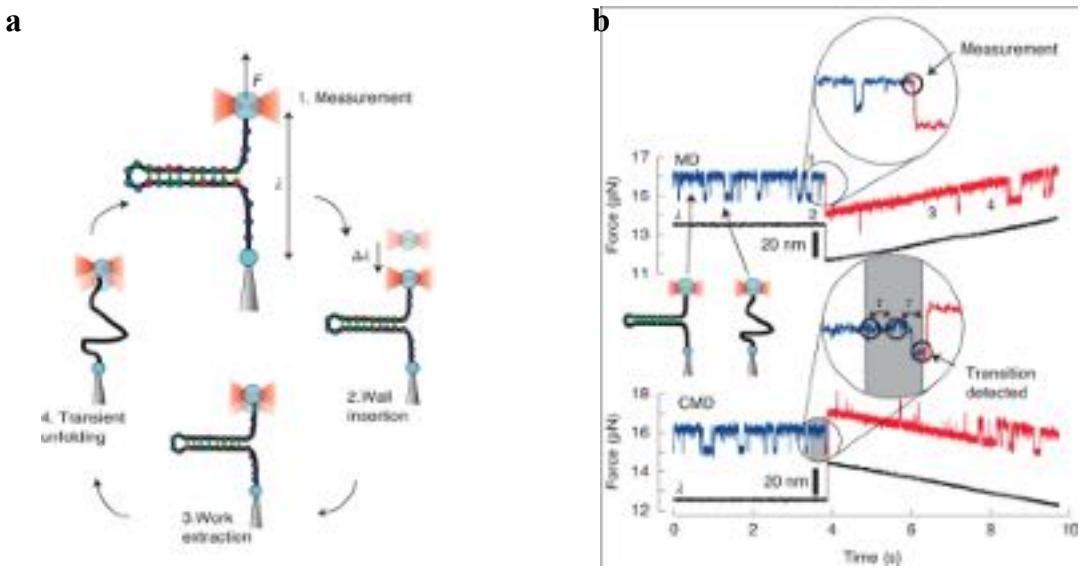


Figure 3.8: Experimental realization of Szilard’s engine in the folding–unfolding of a DNA hairpin. (a) Schematics of the work extraction procedure. A DNA hairpin pulled by an optical trap hops between the folded and unfolded states. The molecular state is determined at a given time (step 1) and, depending on the outcome (folded or unfolded), the optical trap is instantaneously moved (folded, downwards; unfolded, upwards) a distance $\Delta\lambda$ to stabilize that state (step 2). By adiabatically re-positioning the original trap position (step 3), work is extracted from the transition events (step 4). The schematics are particularized for the case where the molecule is observed in the folded state. (b) Force-versus-time trace for the single-measurement Szilard’s engine (dubbed MD for Maxwell demon, top) and for the multiple-measurements Szilard’s engine (dubbed CMD for continuous Maxwell demon, bottom). In the latter case, measurements of the molecular state are made every τ , and a work extraction protocol is implemented the first time the molecule is observed to change state (blue to red transition, circled zoomed regions). Reproduced from Ref. [69].

quantum systems [85], it can be used to reduce dissipation [86] and to increase thermodynamic efficiency. Yet, we do not understand how feedback thermodynamics fits in a biological context. We distinguish discrete-time feedback from continuous-time feedback. A system is driven out of equilibrium by the action of an external agent. In *discrete-time feedback*, the pulling protocol is changed depending on the outcome of one (or more) measurements taken at specific times [87, 88, 89]. In *continuous-time feedback*, an observable is continuously monitored, and the pulling protocol is changed when the observable fulfills a specific condition [90, 91]. For transient nonequilibrium states with feedback, the Jarzynski equality reads $\langle \exp(-W_d/k_B T) \rangle = \gamma$, where γ is called the efficacy and its logarithm $\Upsilon = \log \gamma$ thermodynamic information [91]. The convex property of the exponential function leads to $\langle W_d \rangle \geq -k_B T \Upsilon$. For $\Upsilon \geq 0$ (information-to-work conversion), the dissipated work with feedback can be negative but larger than $-k_B T \Upsilon$ [91].

3.6 Outlook

In 1944, Erwin Schrödinger published an enlightening monography titled “What is Life?”, where he wrote [92]:

The large and important and very much discussed question is: *How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?* [emphasis ours]

We accept that living beings do not violate fundamental laws of physics, yet they are of a particular kind. They seem to circumvent or mock the laws of physics as we understand them: a stone will fall if let fall due to gravity; however, birds fly whenever they decide so. Biologists have invented a term for this (teleonomy) to express

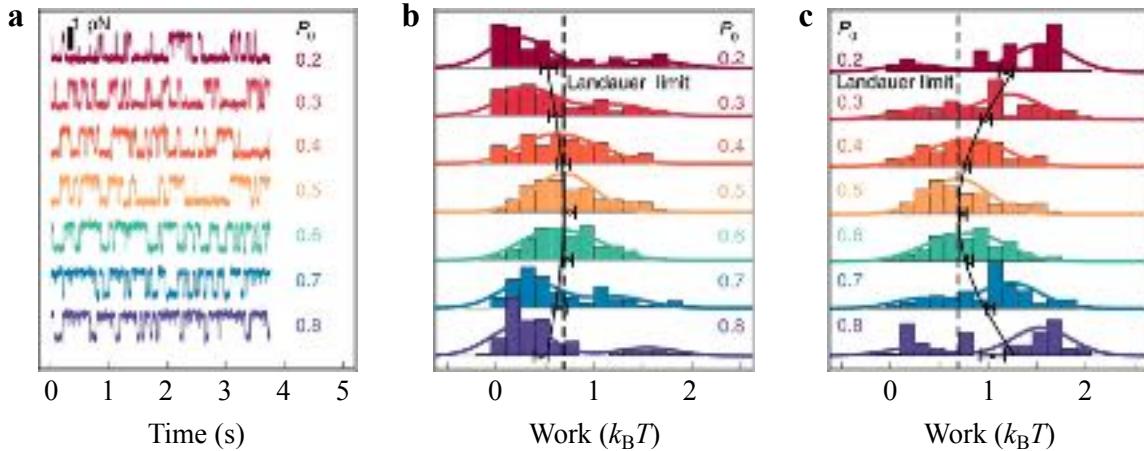


Figure 3.9: Work distributions in Maxwell demons. Work distributions in the Maxwell demon and the continuous Maxwell demon for the single DNA hairpin unfolding experiments of Fig. 3.8. **(a)** Experimental hopping traces at different p_σ conditions. **(b)** Distributions of the extracted work per cycle for the single-measurement Maxwell demon. Symbols are the average values and the dashed line is the theoretical prediction, Eq. (3.15). **(c)** Distributions of the extracted work per cycle for the multiple-measurement continuous Maxwell demon. Symbols are the average values and the dashed line is the theoretical prediction, Eq. (3.16). Reproduced from Ref. [69].

the fundamental fact that living beings have their agenda, meaning that living beings move, eat, reproduce, play, do business, and so on [93, 94]. Measuring entropy production rates has become a major challenge and a fundamental thermodynamic probe for life [95, 96]. Yet, we remain ignorant about which features of entropy production are intrinsic to life.

Living matter is heterogeneous and soft [97], sharing features with active [98] and disordered systems [99]. Cell populations are intrinsically heterogeneous: the same strain, environmental conditions, and *everything* often produce different outcomes. Often, there is no magic bullet to defeat dysfunctional masses of cells [100]. At the molecular level, heterogeneity is widespread, as many RNAs and proteins fold into a heterogeneous set of functional native structures [101, 102]. Living matter is also soft, being actively remodeled by many forces. These are strong enough to maintain stability, but weak enough to allow for remodeling and adaptation. It is not accidental that the energy scale of thermal forces equals that of biochemical reactions ($1k_B T = 0.6 \text{ kcal mol}^{-1}$ at 298 K), reflecting that biological forces operate at the thermal noise level [103]. Most free energies of molecular structures are a few tens of kcal/mol or $k_B T$. As enthalpy (ΔH) and entropy ($T\Delta S$) contributions are typically one order of magnitude higher than free energy (ΔG) values, a compensation between enthalpy and entropy drives most molecular reactions (e.g., folding, binding, self-assembly). Indeed, the enthalpy of a single hydrogen bond contributes with $\sim 7 \text{ kcal mol}^{-1}$, while the total enthalpy ΔH due to all bonds in an RNA or a protein can easily reach several hundreds of kcal/mol. Yet, the folding ΔG is typically ten times smaller due to a similar contribution by the entropy, $T\Delta S$. This scenario describes protein folding, where, under appropriate conditions, the unfolded polypeptide chain collapses to the native state previous formation of a molten globule, a precursor of the native state in a funneled folding energy landscape [104, 105, 106, 107, 58]. Evolutionary forces have built biological systems over eons, from proteins to macromolecular complexes and beyond, driven by endogenous (internal) and exogenous (environmental) factors. Operating over millions of years, these forces have modeled things how we know.

It is too soon to tell whether and how nonequilibrium physics will contribute to understanding the marvelous features of living matter [108, 109]. The prominent role of information undeniably emerges from the experiments and theories developed over the past decades, a quantity that physicists identify with entropy but that calls for a broader ground to explain the complexity of biological matter [110]. One of the most appealing features of nonequilibrium work relations and fluctuation theorems is that they allow us to recover the second law as a particular case of more general mathematical equality [51]. This fact raises the intriguing question of

whether the second law of thermodynamics is a disguised inequality of a yet unknown fundamental conservation law about energy and information [111]. In the best tradition of natural philosophy, experiment hand in hand with theory offers the best chances to unravel the secrets of life. The nonequilibrium physics of small systems is just a milestone in this exciting adventure.

3.7 Problems

Problem 3.1: Inextensible worm-like-chain. We will reproduce the force–extension formula (Eq. (3.2)) originally derived by John Marko and Eric Siggia in 1997 [19]. The formula interpolates the low and high force regimes of the elastic rod model (Eq. (3.1)) when a force is applied to the polymer ends along the x -axis

$$E = \frac{A}{2} \int_0^L ds \left(\frac{\partial \hat{\mathbf{t}}}{\partial s} \right)^2 - f \hat{\mathbf{x}} \int_0^L ds \hat{\mathbf{t}}(s). \quad (3.17)$$

The average extension is defined as

$$\langle x \rangle = -\frac{\partial \log \mathcal{Z}}{\partial f} = L \hat{\mathbf{x}} \frac{\int \mathcal{D}\hat{\mathbf{t}}(s) \hat{\mathbf{t}}(s) \exp(-\beta E)}{\mathcal{Z}}, \quad (3.18)$$

where L is the contour length and $\mathcal{Z} = \int \mathcal{D}\hat{\mathbf{t}}(s) \exp(-\beta E)$ is the partition function. To derive Eq. (3.2), we calculate the leading behavior of Eq. (3.18) in the low and high f regimes, and interpolate between them. The low and high f expansions are similar to the low and high temperature expansions commonly employed in statistical mechanics (e.g., in the Ising model).

- a. Expand Eq. (3.18) around $f = 0$ up to linear order in f finding at low forces

$$\langle x \rangle = \frac{2L\ell^*}{3k_B T} f + \mathcal{O}(f^2), \quad (3.19)$$

where $\ell^* = A/k_B T$ is the persistence length. [Hint: Expand $\log \mathcal{Z}$ to first order in f in Eq. (3.18). Evaluate the second cumulant of E at $f = 0$, $\langle E^2 \rangle_{f=0} - \langle E \rangle_{f=0}^2$, using the zero force results of Example 3.1.]

- b. Expand Eq. (3.18) around $\langle x \rangle = L$ and $f = \infty$ and find for the leading correction at high forces

$$\langle x \rangle = L \left[1 - \sqrt{\frac{k_B T}{4\ell f}} \right] + \mathcal{O}\left(\frac{1}{f}\right). \quad (3.20)$$

[Hint: This derivation is more elaborated. It requires decomposing $\hat{\mathbf{t}} = \hat{\mathbf{t}}_{||} + \hat{\mathbf{t}}_{\perp}$ in a longitudinal component, $\hat{\mathbf{t}}_{||} = (\hat{\mathbf{t}} \cdot \hat{\mathbf{x}}) \hat{\mathbf{x}}$, and a transverse component, $\hat{\mathbf{t}}_{\perp} = (\hat{\mathbf{t}} \cdot \hat{\mathbf{y}}) \hat{\mathbf{y}} + (\hat{\mathbf{t}} \cdot \hat{\mathbf{z}}) \hat{\mathbf{z}}$. At high forces $|\hat{\mathbf{t}}_{\perp}| \ll |\hat{\mathbf{t}}_{||}|$, we can express $\hat{\mathbf{t}}$ as a function of the transverse modes only, $\hat{\mathbf{t}} = (1 - |\hat{\mathbf{t}}_{\perp}|^2/2) \hat{\mathbf{x}} + \hat{\mathbf{t}}_{\perp}$. By rewriting Eq. (3.17) in terms of the Fourier transverse modes or discretizing the chain as suggested in Example 3.1 and using periodic Born-Von Karman conditions we can integrate over the Gaussian transverse modes finding the above result.]

- c. From Eqs. (3.19),(3.20) derive the interpolation formula Eq. (3.2).

Problem 3.2: Steady-state work fluctuation theorem. We will derive Eq. (3.8) for the steady state of a bead dragged through water by an optical trap moving at speed v . We will consider Eq. (3.11) and derive the steady-state work distribution for $W_t = \frac{v}{T} \int_0^t f(s) ds$ using $f(t) = -k_B(x(t) - x_0)$.

a. Derive the autocorrelation function of the force in the steady state, $C(t) = \langle f(0)f(t) \rangle$, and compare the result with the equilibrium case, Eq. (3.12). Show also that the stationary distribution for the energy of the particle, $U(x) = (1/2)k_B(x - x_0)^2$, is given by the Boltzmann distribution

$$P(U) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{U}{k_B T}\right), \quad (3.21)$$

with $\mathcal{Z} = \sqrt{\frac{2\pi k_B T}{k_B}}$.

b. Demonstrate that $P(W_t)$ is a Gaussian distribution that fulfills

$$\frac{P(W_t)}{P(-W_t)} = \exp\left(\frac{W_t}{k_B T \left[1 - \frac{\tau_r}{t} \left(1 - e^{-\frac{t}{\tau_r}}\right)\right]}\right). \quad (3.22)$$

Show that Eq. (3.13) does not hold for finite t , but, in the limit $\frac{t}{\tau_r} \gg 1$, the work fluctuation theorem and Eq. (3.13) are recovered.

c. From the first law, $Q_t + W_t = \Delta U$, and Eqs. (3.21) and (3.22) explain what should be the form of the heat distribution $P(Q_t)$. How do the negative tails of $P(Q_t)$ decay? Do you think there is a corresponding theorem in the large t limit ($\frac{t}{\tau_r} \gg 1$)? [Hint: The problem for heat is not as simple as for the work. Keep your discussion at a qualitative level.]

Problem 3.3: Information-to-energy conversion in the continuous Maxwell demon. In the continuous Maxwell demon, a work extraction cycle consists of repeated measurements at every time τ until the particle changes compartment (Fig. 3.7) or the molecule changes state (Fig. 3.8b). Stored sequences contain $n \geq 2$ bits, being of the type $\mathcal{S}_n = \{\sigma, \sigma, \dots, \sigma, 1 - \sigma\}$ with $\sigma = 0, 1$. Let p_σ be the probability of a measurement outcome, σ . In the limit of uncorrelated measurements ($\tau \rightarrow \infty$), the probability of a sequence \mathcal{S}_n is given by, $P(\mathcal{S}_n) = \prod_{k=1}^n p_{\sigma_k}$.

a. Calculate the average information content per cycle of the stored sequences

$$\mathcal{I} = - \sum_{n \geq 2} \sum_{\mathcal{S}_n} P(\mathcal{S}_n) \log P(\mathcal{S}_n) \quad (3.23)$$

and demonstrate it is given by

$$\mathcal{I} = - \sum_{\sigma=0,1} \left(\frac{p_\sigma}{p_{1-\sigma}} + p_{1-\sigma} \right) \log p_\sigma. \quad (3.24)$$

b. Demonstrate the validity of the second law

$$W_{CMD} \leq k_B T \mathcal{I}, \quad (3.25)$$

where W_{CMD} and \mathcal{I} are given in Eqs. (3.16) and (3.24), respectively.

c. The information-to-energy thermodynamic cycle efficiency is defined by

$$\eta_I = \frac{W_{CMD}}{k_B T \mathcal{I}}. \quad (3.26)$$

Show that the minimum efficiency ($\eta_I = 1/3$) is obtained for $p_0 = p_1 = 1/2$. Demonstrate that maximum efficiency ($\eta_I = 1$) is asymptotically obtained in the limit of rare events ($p_0 \rightarrow 0, 1$). Show that in this limit both W_{CMD} and \mathcal{I} diverge. Interpret this result. How does Eq. (3.26) compare with the efficiency of the standard MD, Eq. (3.15)?

Bibliography

- [1] Z. L. Watson, F. R. Ward, R. Méheust, O. Ad, A. Schepartz, J. F. Banfield, and J. H. D. Cate. Structure of the bacterial ribosome at 2 Å resolution. *elife*, page e60482, 2020.
- [2] C. Bustamante, J. Liphardt, and F. Ritort. The nonequilibrium thermodynamics of small systems. *Physics Today*, 58:43–48, 2005.
- [3] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75:126001, 2012.
- [4] S. Ciliberto. Experiments in stochastic thermodynamics: Short history and perspectives. *Physical Review X*, 7:021051, 2017.

- [5] R. Zwanzig. *Nonequilibrium statistical mechanics*. Oxford university press, 2001.
- [6] D. Chandler. *Introduction to modern statistical mechanics*. Oxford University Press, Oxford, UK, 1987.
- [7] F. Ritort. Nonequilibrium fluctuations in small systems: From physics to biology. *Advances in Chemical Physics*, 137:31, 2008.
- [8] E. Dieterich, J. Camunas-Soler, M. Ribezzi-Crivellari, U. Seifert, and F. Ritort. Single-molecule measurement of the effective temperature in non-equilibrium steady states. *Nature Physics*, 11:971–977, 2015.
- [9] F. Ritort. Single-molecule experiments in biological physics: methods and applications. *Journal of Physics: Condensed Matter*, 18:R531, 2006.
- [10] K. C. Neuman, T. Lionnet, and J.-F. Allemand. Single-molecule micromanipulation techniques. *Annual Review of Materials Research*, 37:33–67, 2007.
- [11] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante. Recent advances in optical tweezers. *Annual Review of Biochemistry*, 77:205–228, 2008.
- [12] D. B. Ritchie and M. T. Woodside. Probing the structural dynamics of proteins and nucleic acids with optical tweezers. *Current Opinion in Structural Biology*, 34:43–51, 2015.
- [13] H. Miller, Z. Zhou, J. Shepherd, A. J. M. Wollman, and M. C. Leake. Single-molecule techniques in biophysics: a review of the progress in methods and applications. *Reports on Progress in Physics*, 81:024601, 2017.
- [14] C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang. Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers*, 1:25, 2021.
- [15] J. Gieseler, J. R. Gomez-Solano, A. Magazzù, I. Pérez Castillo, L. Pérez García, M. Gironella-Torrent, X. Viader-Godoy, F. Ritort, G. Pesce, A. V. Arzola, K. Volke-Sepúlveda, and G. Volpe. Optical tweezers – from calibration to applications: a tutorial. *Advances in Optics and Photonics*, 13:74–241, 2021.
- [16] S. B. Smith, Y. Cui, and C. Bustamante. Optical-trap force transducer that operates by direct measurement of light momentum. In *Methods in enzymology*, volume 361, pages 134–162. Elsevier, 2003.
- [17] S. B. Smith, Y. Cui, and C. Bustamante. Overstretching b-dna: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271:795–799, 1996.
- [18] P. C. Nelson, M. Radosavljević, S. Bromberg, and D. S Goodsell. *Biological physics: energy, information, life*. WH Freeman New York, 2008.
- [19] J. F. Marko and E. D. Siggia. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.
- [20] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of λ -phage DNA. *Science*, 265:1599–1600, 1994.
- [21] A. Severino, A. M. Monge, P. Rissone, and F. Ritort. Efficient methods for determining folding free energies in single-molecule pulling experiments. *Journal of Statistical Mechanics: Theory and Experiment*, 2019:124001, 2019.
- [22] A. Bosco, J. Camunas-Soler, and F. Ritort. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Research*, 42:2064–2074, 2014.
- [23] D. R. Jacobson, D. B. McIntosh, M. J. Stevens, M. Rubinstein, and O. A. Saleh. Single-stranded nucleic acid elasticity arises from internal electrostatic tension. *Proceedings of the National Academy of Sciences*, 114:5095–5100, 2017.

- [24] C. V. Bizarro, A. Alemany, and F. Ritort. Non-specific binding of Na⁺ and Mg²⁺ to RNA determined by force spectroscopy methods. *Nucleic Acids Research*, 40:6922–6935, 2012.
- [25] O. A. Saleh, D. B. McIntosh, P. Pincus, and N. Ribeck. Nonlinear low-force elasticity of single-stranded DNA molecules. *Physical Review Letters*, 102:068301, 2009.
- [26] X. Viader-Godoy, C. R. Pulido, B. Ibarra, M. Manosas, and F. Ritort. Cooperativity-dependent folding of single-stranded DNA. *Physical Review X*, 11:031037, 2021.
- [27] N. M. Toan, D. Marenduzzo, and C. Micheletti. Inferring the diameter of a biopolymer from its stretching response. *Biophysical Journal*, 89:80–86, 2005.
- [28] N. M. Toan and C. Micheletti. Inferring the effective thickness of polyelectrolytes from stretching measurements at various ionic strengths: applications to DNA and RNA. *Journal of Physics: Condensed Matter*, 18:S269, 2006.
- [29] S. de Lorenzo, M. Ribezzi-Crivellari, J. R. Arias-Gonzalez, S. B. Smith, and F. Ritort. A temperature-jump optical trap for single-molecule manipulation. *Biophysical Journal*, 108:2854–2864, 2015.
- [30] M. Rico-Pasto and F. Ritort. Temperature-dependent elastic properties of DNA. *Biophysical Reports*, 2(3), 2022.
- [31] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot. Mechanical separation of the complementary strands of DNA. *Proceedings of the National Academy of Sciences*, 94(22):11935–11940, 1997.
- [32] M. Rief, H. Clausen-Schaumann, and H. E. Gaub. Sequence-dependent mechanics of single DNA molecules. *Nature Structural Biology*, 6:346–349, 1999.
- [33] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proceedings of the National Academy of Sciences*, 107:15431–15436, 2010.
- [34] J. M. Huguet, M. Ribezzi-Crivellari, C. V. Bizarro, and F. Ritort. Derivation of nearest-neighbor DNA parameters in magnesium from single molecule experiments. *Nucleic Acids Research*, 45:12921–12931, 2017.
- [35] P. Rissone, C. V. Bizarro, and F. Ritort. Stem-loop formation drives RNA folding in mechanical unzipping experiments. *Proceedings of the National Academy of Sciences*, 119:e2025575119, 2022.
- [36] P. Rissone and F. Ritort. Nucleic acid thermodynamics derived from mechanical unzipping experiments. *Life*, 12:1089, 2022.
- [37] M. Manosas, J. Camunas-Soler, V. Croquette, and F. Ritort. Single molecule high-throughput footprinting of small and large DNA ligands. *Nature Communications*, 8:304, 2017.
- [38] J. M. Huguet, N. Forns, and F. Ritort. Statistical properties of metastable intermediates in DNA unzipping. *Physical Review Letters*, 103:248106, 2009.
- [39] C. Ter Burg, P. Rissone, M. Rico-Pasto, F. Ritort, and K. J. Wiese. Experimental test of Sinai's model in DNA unzipping. *Physical Review Letters*, 130:208401, 2023.
- [40] N. Forns, S. de Lorenzo, M. Manosas, K. Hayashi, J. M. Huguet, and F. Ritort. Improving signal/noise resolution in single-molecule experiments using molecular constructs with short handles. *Biophysical Journal*, 100:1765–1774, 2011.
- [41] M. Manosas, A. Mossa, N. Forns, J. M. Huguet, and F. Ritort. Dynamic force spectroscopy of DNA hairpins: II. Irreversibility and dissipation. *Journal of Statistical Mechanics: Theory and Experiment*, 2009:P02061, 2009.

- [42] A. Mossa, M. Manosas, N. Forns, J. M. Huguet, and F. Ritort. Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, 2009:P02060, 2009.
- [43] A. Alemany and F. Ritort. Force-dependent folding and unfolding kinetics in DNA hairpins reveals transition-state displacements along a single pathway. *The Journal of Physical Chemistry Letters*, 8:895–900, 2017.
- [44] M. Rico-Pasto, A. Alemany, and F. Ritort. Force-dependent folding kinetics of single molecules with multiple intermediates and pathways. *The Journal of Physical Chemistry Letters*, 13:1025–1032, 2022.
- [45] D. J. Evans and D. J. Searles. The fluctuation theorem. *Advances in Physics*, 51:1529–1585, 2002.
- [46] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Physical Review Letters*, 74:2694, 1995.
- [47] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60:2721, 1999.
- [48] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690, 1997.
- [49] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Physical Review Letters*, 89(5):050601, 2002.
- [50] I. Di Terlizzi, F. Ritort, and M. Baiesi. Explicit solution of the generalised langevin equation. *Journal of Statistical Physics*, 181:1609–1635, 2020.
- [51] C. Jarzynski. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. *Annual Review of Condensed Matter Physics*, 2:329–351, 2011.
- [52] M. Palassini and F. Ritort. Improving free-energy estimates from unidirectional work measurements: theory and experiment. *Physical Review Letters*, 107:060601, 2011.
- [53] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr, and C. Bustamante. Verification of the crooks fluctuation theorem and recovery of rna folding free energies. *Nature*, 437:231–234, 2005.
- [54] I. Junier, A. Mossa, M. Manosas, and F. Ritort. Recovery of free energy branches in single molecule experiments. *Physical Review Letters*, 102:070602, 2009.
- [55] A. Alemany, A. Mossa, I. Junier, and F. Ritort. Experimental free-energy measurements of kinetic molecular states using fluctuation theorems. *Nature Physics*, 8:688–694, 2012.
- [56] A. Alemany, M. Ribetti-Crivellari, and F. Ritort. From free energy measurements to thermodynamic inference in nonequilibrium small systems. *New Journal of Physics*, 17:075009, 2015.
- [57] A. Alemany, B. Rey-Serra, S. Frutos, C. Cecconi, and F. Ritort. Mechanical folding and unfolding of protein barnase at the single-molecule level. *Biophysical Journal*, 110:63–74, 2016.
- [58] M. Rico-Pasto, A. M. Zaltron, S. J. Davis, S. Frutos, and F. Ritort. Molten globule–like transition state of protein barnase measured with calorimetric force spectroscopy. *Proceedings of the National Academy of Sciences*, 119:e2112382119, 2022.
- [59] J. Camunas-Soler, A. Alemany, and F. Ritort. Experimental measurement of binding energy, selectivity, and allostery using fluctuation theorems. *Science*, 355:412–415, 2017.
- [60] M. T. Woodside, P. C. Anthony, W. M. Behnke-Parks, K. Larizadeh, D. Herschlag, and S. M Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314:1001–1004, 2006.

- [61] M. Ribezzi-Crivellari and F. Ritort. Free-energy inference from partial work measurements in small systems. *Proceedings of the National Academy of Sciences*, 111:E3386–E3394, 2014.
- [62] N. Goldenfeld and C. Woese. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annual Review of Condensed Matter Physics*, 2:375–399, 2011.
- [63] G. Verley, M. Esposito, T. Willaert, and C. Van den Broeck. The unlikely carnot efficiency. *Nature Communications*, 5:4721, 2014.
- [64] R. K. Schmitt, J. M. R. Parrondo, H. Linke, and J. Johansson. Molecular motor efficiency is maximized in the presence of both power-stroke and rectification through feedback. *New Journal of Physics*, 17:065011, 2015.
- [65] M. Manosas, X. G. Xi, D. Bensimon, and V. Croquette. Active and passive mechanisms of helicases. *Nucleic Acids Research*, 38:5518–5526, 2010.
- [66] J. D. Norton. All shook up: Fluctuations, Maxwell’s demon and the thermodynamics of computation. *Entropy*, 15:4432–4483, 2013.
- [67] K. Maruyama, F. Nori, and V. Vedral. Colloquium: The physics of Maxwell’s demon and information. *Reviews of Modern Physics*, 81(1):1, 2009.
- [68] C. H. Bennett. Demons, engines and the second law. *Scientific American*, 257:108–117, 1987.
- [69] M. Ribezzi-Crivellari and F. Ritort. Large work extraction and the Landauer limit in a continuous Maxwell demon. *Nature Physics*, 15:660–664, 2019.
- [70] H. Leff and A. F. Rex. *Maxwell’s Demon 2 Entropy, Classical and Quantum Information, Computing*. CRC Press, 2002.
- [71] I. J. Ford. Maxwell’s demon and the management of ignorance in stochastic thermodynamics. *Contemporary Physics*, 57:309–330, 2016.
- [72] A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, and E. Lutz. Experimental verification of Landauer’s principle linking information and thermodynamics. *Nature*, 483:187–189, 2012.
- [73] É. Roldán, I. A. Martínez, J. M. R. Parrondo, and D. Petrov. Universal features in the energetics of symmetry breaking. *Nature Physics*, 10:457–461, 2014.
- [74] J. V. Koski, V. F. Maisi, J. P. Pekola, and D. V. Averin. Experimental realization of a Szilard engine with a single electron. *Proceedings of the National Academy of Sciences*, 111:13786–13789, 2014.
- [75] T. M. Cover. *Elements of information theory*. John Wiley & Sons, 1999.
- [76] M. Ribezzi-Crivellari and F. Ritort. Work extraction, information-content and the landauer bound in the continuous Maxwell demon. *Journal of Statistical Mechanics: Theory and Experiment*, 2019:084013, 2019.
- [77] T. Admon, S. Rahav, and Y. Roichman. Experimental realization of an information machine with tunable temporal correlations. *Physical Review Letters*, 121:180601, 2018.
- [78] J. P. Garrahan and F. Ritort. Generalized continuous Maxwell demons. *Physical Review E*, 107:034101, 2023.
- [79] P. Raux and F. Ritort. N-states continuous Maxwell demon. *Entropy*, 25:321, 2023.
- [80] T. Sagawa. Thermodynamic and logical reversibilities revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2014:P03025, 2014.
- [81] J. M. R. Parrondo, J. M. Horowitz, and T. Sagawa. Thermodynamics of information. *Nature Physics*, 11:131–139, 2015.

- [82] R. K. Schmitt, P. P. Potts, H. Linke, J. Johansson, P. Samuelsson, M. Rico-Pasto, and F. Ritort. Information-to-work conversion in single-molecule experiments: From discrete to continuous feedback. *Physical Review E*, 107:L052104, 2023.
- [83] J. Bechhoefer. Feedback for physicists: A tutorial essay on control. *Reviews of modern physics*, 77:783, 2005.
- [84] E. Dieterich, J. Camunas-Soler, M. Ribezzi-Crivellari, U Seifert, and F. Ritort. Control of force through feedback in small driven systems. *Physical Review E*, 94:012107, 2016.
- [85] P. Strasberg. *Quantum Stochastic Thermodynamics: Foundations and Selected Applications*. Oxford University Press, 2022.
- [86] S. Tafoya, S. Large, S. J. and Liu, C. Bustamante, and D. A. Sivak. Using a system’s equilibrium behavior to reduce its energy dissipation in nonequilibrium processes. *Proceedings of the National Academy of Sciences*, 116:5920–5924, 2019.
- [87] T. Sagawa and M. Ueda. Generalized Jarzynski equality under nonequilibrium feedback control. *Physical Review Letters*, 104:090602, 2010.
- [88] S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, and M. Sano. Experimental demonstration of information-to-energy conversion and validation of the generalized jarzynski equality. *Nature Physics*, 6:988–992, 2010.
- [89] P. P. Potts and P. Samuelsson. Detailed fluctuation relation for arbitrary measurement and feedback schemes. *Physical Review Letters*, 121:210603, 2018.
- [90] J. M. Horowitz and S. Vaikuntanathan. Nonequilibrium detailed fluctuation theorem for repeated discrete feedback. *Physical Review E*, 82:061120, 2010.
- [91] M. Rico-Pasto, R. K. Schmitt, M. Ribezzi-Crivellari, J. M. R. Parrondo, H. Linke, J. Johansson, and F. Ritort. Dissipation reduction and information-to-measurement conversion in DNA pulling experiments with feedback protocols. *Physical Review X*, 11:031052, 2021.
- [92] E. Schrödinger. *What is life? The physical aspect of the living cell and mind*. Cambridge University Press, 1944.
- [93] M. Kirschner, J. Gerhart, and T. Mitchison. Molecular “vitalism”. *Cell*, 100:79–88, 2000.
- [94] A. Pross. *What is life?: How chemistry becomes biology*. Oxford University Press, 2016.
- [95] I. Di Terlizzi, M. Gironella, D. Herráez-Aguilar, T. Betz, F. Monroy, M. Baiesi, and F. Ritort. Variance sum rule for entropy production. *Science*, 383:971–976, 2024.
- [96] É. Roldán. Thermodynamic probes of life. *Science*, 383:952–953, 2024.
- [97] T. R. Kirkpatrick and D. Thirumalai. Colloquium: Random first order transition theory concepts in biology and physics. *Reviews of Modern Physics*, 87:183, 2015.
- [98] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [99] S. Cocco, A. De Martino, A. Pagnani, M. Weigt, and F. Ritort. Statistical physics of biological molecules. *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking After 40 Years*, pages 523–559, 2023.
- [100] Q. Zhang and R. H. Austin. Physics of cancer: the impact of heterogeneity. *Annual Review of Condensed Matter Physics*, 3:363–382, 2012.
- [101] S. V. Solomatin, M. Greenfeld, S. Chu, and D. Herschlag. Multiple native states reveal persistent ruggedness of an rna folding landscape. *Nature*, 463:681–684, 2010.

- [102] B. Liu, R. J. Baskin, and S. C. Kowalczykowski. DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate. *Nature*, 500:482–485, 2013.
- [103] R Dean Astumian and Peter Hänggi. Brownian motors. *Physics today*, 55(11):33–39, 2002.
- [104] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21:167–195, 1995.
- [105] W. A. Eaton. Searching for “downhill scenarios” in protein folding. *Proceedings of the National Academy of Sciences*, 96:5897–5899, 1999.
- [106] D. Thirumalai and C. Hyeon. Rna and protein folding: common themes and variations. *Biochemistry*, 44:4957–4970, 2005.
- [107] H. Maity, M. Maity, M. M. G. Krishna, L. Mayne, and S. W. Englander. Protein folding: the stepwise assembly of foldon units. *Proceedings of the National Academy of Sciences*, 102:4741–4746, 2005.
- [108] F. Ritort and B. van Tiggelen. Physics for understanding life. *Europhysics News*, 53:24–27, 2022.
- [109] P. Bassereau, A. Cangelosi, J. Čejková, C. Gershenson, R. Goldstein, Z. Martins, S. Matthews, F. Ritort, S. Seager, B. Van Tiggelen, et al. Physics for understanding life. In *EPS Grand Challenges: Physics for Society in the Horizon 2050*. IOP Publishing, 2024.
- [110] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
- [111] F. Ritort. The physics of small systems: From energy to information. *Contributions to science*, 11:137–146, 2016.

Chapter 4

From Passive to Active Brownian Motion

DEMIAN LEVIS, NUNO A. M. ARAÚJO, IGNACIO PAGONABARRAGA

The erratic motion of active particles, such as that of the bacteria shown in Fig. 6.1, resembles the Brownian motion observed by Robert Brown for passive systems. The latter refers to fluctuations in thermal equilibrium, while active systems are out of equilibrium. Nevertheless, the ideas and methods of kinetic theory and stochastic processes can be extended to study active systems. In this Chapter, we will start with an introduction to the concept of Brownian motion and discuss the relevant time scales. We will introduce the Langevin approach where this many-body problem is reduced to a one-body problem with a stochastic force. Then, we will present a probabilistic description of the Brownian motion. Finally, we will discuss the extension to active systems. Our goal here is not to discuss these topics in detail but rather to provide an overview of the key aspects that are relevant to understand the next chapters. For the passive systems, there are many textbooks that address this topic from different perspectives. For example, for further reading in the context of Soft Matter systems, we suggest Refs. [2, 3, 4]. For more extended and general discussion on active matter models, we suggest the following reviews [5, 6, 7, 8].

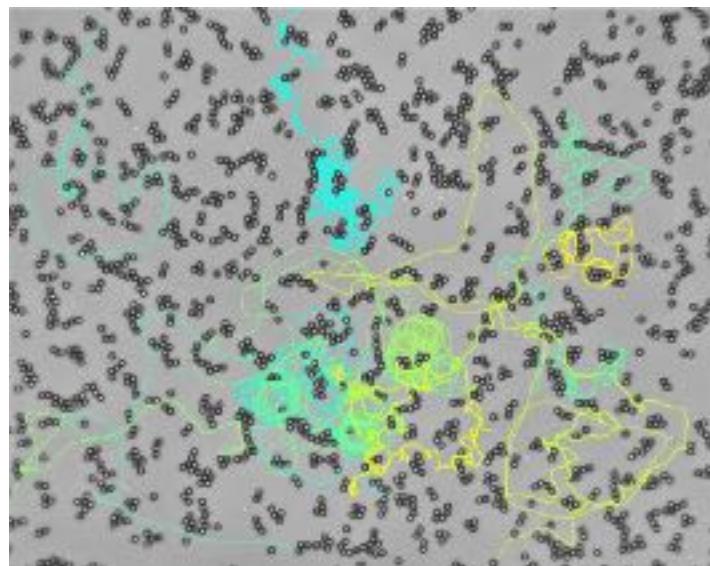


Figure 4.1: **Motile bacteria swimming in a landscape of passive colloidal particles.** Trajectories of *E. Coli* bacteria swimming near a glass surface with colloidal obstacles attached to the surface [1]. Credit: Stanislaw Makarchuk and Giorgio Volpe.

4.1 Brownian motion: degrees of freedom, length and time scales

Particles in active matter systems self-propel in a preferential direction. Through particle–particle and particle–medium interactions, the direction of motion may change in time, and the trajectory of an active particle depends strongly on the nature and time scale of these changes. To resolve these trajectories, an explicit description of the self-propulsion, alignment interactions, and coupling with the medium is an impossible endeavor for two main reasons. First, the fine details of the interactions typically involve non-trivial (bio)mechanical processes, which are system dependent, complicated, and to a large extent unknown [9]. Second, the number of microscopic degrees of freedom is very large and impossible to consider, even numerically. For example, for a single bacterium in a droplet of water of 1 cm^3 , one would need 10^5 times the actual overall Google storage capacity to save the initial position and velocity of all the $\sim 10^{22}$ molecules of water (see Example 4.1). Thus, to obtain theoretical insight into the single and collective dynamics of active matter systems, one needs to focus on general features and develop coarse-grained models to access the relevant length and time scales.

Instead of considering all degrees of freedom, a useful strategy is to focus on a subset of variables, which are relevant to describe the active matter system (e.g., spatial coordinates and swimming direction of the active particles), and describe the effect of all the other degrees of freedom in a statistical way. This is particularly useful when the selected variables change slowly in time (*slow variables*) when compared to the remaining ones (*fast variables*). In this limit, we take advantage of this separation of time scales and describe the effect of the slow variables based on average (coarse-grained) quantities and fluctuations around those averages. In this Chapter, we will discuss this approach starting with Brownian motion as an example of motion generated by random fluctuations. Brownian motion refers to fluctuations in thermal equilibrium. Nevertheless, the ideas and methods of kinetic theory and stochastic processes discussed here will pave the way for the discussion of intrinsically non-equilibrium active systems in the following Chapters.

Example 4.1: Computational resources to integrate the equations of motion for all molecules in a droplet. Let us consider a droplet of water of 1 cm^3 . Since the density of water is $\approx 1\text{ g cm}^{-3}$ and its molar mass is $\approx 18\text{ g mol}^{-1}$. There are ≈ 0.06 moles of water in the droplet, which corresponds to $\approx 4 \times 10^{22}$ water molecules. Even if we assume that all molecules are spheres and neglect their vibrational and rotational degrees of freedom, there are of the order of 10^{23} degrees of freedom. To resolve the trajectory of each molecule, one needs to integrate Newton equations of motion for each particle, which implies solving 10^{23} coupled second-order differential equations. To get started, one needs to know the initial position and velocity of each molecule. Each real number requires eight bytes of memory to be saved with 64-bit precision, which adds up to 10^{23} bytes (10^{11} TB) of memory to store this information for all the molecules. To put it into perspective, the current Google storage capacity is 10^6 TB . Clearly, even saving the initial position of all the molecules is currently impossible.

The accessible time scales are also very short. Among the fastest supercomputers in the TOP 500 world rank of 2018 is the Sunway TaihuLight in China, with a rating of 93 PFLOPS. This means that it performs 93×10^{15} operations per second. Assuming that each iteration step of the integration scheme is one operation, it would take 10^5 s ($\approx 27\text{ hours}$) to perform a single iteration step for all the particles. Since the relevant time scale for the vibration of a water molecule is 1 fs , it would take 10^{12} days to simulate 10^{-3} s , which is the minimum time resolution for the human eye. Note that this simulation time corresponds to 100 times the estimated age of the Universe.

Let us consider a passive spherical particle of radius $R = 1\text{ }\mu\text{m}$ and mass m suspended in water at a temperature $T = 293.15\text{ K}$. Figure 4.2 shows a typical trajectory of the particle over a period of 60 seconds. The density of the particle equals the one of water and so gravity is balanced by the buoyant force. Nevertheless, one clearly observes a non-negligible erratic motion of the particle which, within 60 seconds, is displaced by a distance that is roughly ten times its radius. This erratic motion was first reported by the Scottish botanist Robert Brown in the XIX century for microscopic particles of pollen grains in water — the reason why it is widely known as Brownian motion. This motion cannot be justified based on the action of forces such as gravity and buoyancy. It requires a microscopic description.

The suspending fluid consists of several water molecules ($\sim 10^{22}\text{ cm}^{-3}$, see Example 4.1) which wiggle around colliding with each other and with the particle. On average (over time and independent samples), there is no preferential direction for such collisions as their spatial distribution is uniform and the velocity

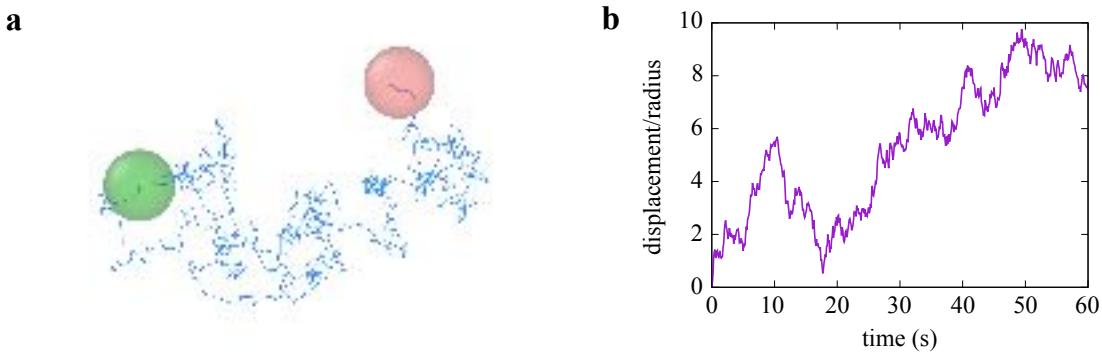


Figure 4.2: Brownian motion. (a) Stochastic trajectory of a spherical Brownian particle of radius $R = 1 \mu\text{m}$ in water over 60 seconds. The green (red) spherical particle corresponds to the initial (final) position. (b) Time dependence of the displacement from the original position in units of the particle radius. The trajectory was obtained numerically as described in Chapter ??, where we set the viscosity of water $\eta = 10^{-3} \text{ Pas}$ and the thermostat temperature $T = 293.15 \text{ K}$.

distribution isotropic. However, the velocity of each molecule follows an equilibrium distribution and thus the impulse on the spherical particle, resulting from all collisions with the molecules in any time interval is non-negligible. Asymptotically, the entire system should be in thermal equilibrium and so the velocity distribution of the particle should follow the Maxwell–Boltzmann distribution for a particle of mass m at temperature T , which corresponds to an expected root-mean-square speed of the order of 1 mm s^{-1} (see Example 4.2).

Example 4.2: Average speed for a suspended particle. From the equipartition theorem, a particle in thermal equilibrium with a heat reservoir at temperature T has an average kinetic energy per degree of freedom of $k_B T/2$, where $k_B \approx 1.38 \times 10^{-23} \text{ J K}^{-1}$ is the Boltzmann constant. Thus, in d dimensions, the root mean square speed v_{rms} of a particle of mass m is

$$v_{\text{rms}} = \sqrt{\frac{dk_B T}{m}}.$$

For a spherical particle of radius $R = 1 \mu\text{m}$ of the same density as water and $T = 293.15 \text{ K}$, $m \approx 10^{-15} \text{ kg}$ and $v_{\text{rms}} \approx 3 \text{ mm s}^{-1}$ in 3D.

As the particle collides with the water molecules, energy is also dissipated. If we assume that the macroscopic mechanism of dissipation is described by Stokes' law, a dissipative drag force

$$\mathbf{F}_d = -\gamma \mathbf{v} \quad (4.1)$$

is applied by the fluid on the particle. This force is proportional to the particle velocity \mathbf{v} , with a Stokes' coefficient γ that depends on the particle shape and viscosity of the fluid. For a spherical particle in water, $\gamma = 6\pi\eta R$, where $\eta \approx 10^{-3} \text{ Pas}$ is the viscosity of water. The Newton equation of motion is then

$$m\dot{\mathbf{v}} = \mathbf{F}_d = -\gamma \mathbf{v}, \quad (4.2)$$

where $\dot{\mathbf{v}}$ is the time derivative of the velocity. Thus,

$$\mathbf{v}(t) = \mathbf{v}(0) \exp\left(-\frac{\gamma}{m} t\right). \quad (4.3)$$

Energy is dissipated on a time scale $t_d = m/\gamma$, which in the case of a spherical particle of $R = 1 \mu\text{m}$ suspended in water is $t_d \approx 10^{-6} \text{ s}$. Thus, in the absence of any external force, the Brownian motion for time scales larger than t_d should be independent of the initial conditions and emerging from the random nature of collisions with the water molecules. Note that Stokes' law is a macroscopic description of the fluid–particle interaction and it is only valid for time scales much larger than the typical time between collisions t_w with the water molecules, which in this case is $t_w \approx 10^{-11} \text{ s}$ (see Example 4.3).

Example 4.3: Inter-collision time for water molecules. As discussed in Example 4.1, there are $\sim 10^{22}$ molecules per cubic centimeter of water. This corresponds to $\approx 10^{-22} \text{ cm}^3$ per molecule. We estimate the mean distance ℓ between molecules as the length of a cube with the same volume. Thus, $\ell \approx 10^{-7} \text{ cm}$. As in example 4.2, one can estimate the root mean square speed for water molecules v_w from the equipartition theorem, $v_w = \sqrt{dk_B T/m} \approx 10^2 \text{ m s}^{-1}$. So, the inter-collision time for water molecules is of the order of 10^{-11} s .

4.2 Langevin approach

Paul Langevin proposed a mechanical approach to describe the effect of random collisions [10, 11]. The idea is to coarse grain the interaction between the fluid molecules and the particle, and replace them by a combination of two forces: a Stokes friction force given by Eq. (4.1) and a stochastic force $\xi(t)$, which reduces the many-body problem to a much simpler one-body problem. The stochastic force is a random time series that has the following properties:

- $\xi(t)$ is independent of the spatial position.
- $\xi(t)$ fluctuates at any time scale larger than t_w to describe the effect of multiple collisions.
- Since the velocity distribution of water molecules is isotropic, there is no preferential collision direction and so the time and ensemble averages of the stochastic force are zero ($\langle \xi(t) \rangle = 0$).
- The random noise is considered uncorrelated in space and time and so $\langle \xi_i(t) \xi_j(t') \rangle = \Gamma \delta(t - t') \delta_{ij}$, where ξ_i and ξ_j are components of ξ , t and t' are instants of time, $\delta(t - t')$ is the Dirac delta function, δ_{ij} the Kronecker delta, which is one when $i = j$ and zero otherwise, and Γ sets the amplitude of the noise.

The resulting Newton equation of motion for the suspended particle in the Langevin approach is

$$m\dot{\mathbf{v}} = -\gamma\mathbf{v} + \xi. \quad (4.4)$$

The first term on the right-hand side describes the slowly varying contribution of the interaction with the fluid (in the time scale of t_d) and is parameterized by γ , while the second term describes the fast fluctuating contribution due to thermal noise and is parameterized by the amplitude of the noise Γ . The two parameters γ and Γ are related by a fluctuation–dissipation relation, as we show next.

The differential equation (4.4) is a first-order non-homogeneous differential equation. The solution for the time dependence of the velocity is

$$\mathbf{v}(t) = \mathbf{v}(0) \exp\left(-\frac{\gamma}{m}t\right) + \frac{\exp\left(-\frac{\gamma}{m}t\right)}{m} \int_0^t dt' \xi(t') \exp\left(\frac{\gamma}{m}t'\right), \quad (4.5)$$

where $\xi(t')$ is a stochastic time series and its values are different for different realizations. Thus, a study of the details of a single stochastic trajectory is meaningless. Instead, we focus on the time dependence of average quantities. If we average over different samples, since $\langle \xi(t) \rangle = 0$, we obtain,

$$\langle \mathbf{v}(t) \rangle = \mathbf{v}(0) \exp\left(-\frac{\gamma}{m}t\right), \quad (4.6)$$

which is equivalent to the solution in the absence of the noise term (see Eq. (4.3)). For simplicity, we have assumed that the initial velocity is the same for all samples ($\langle \mathbf{v}(0) \rangle \equiv \mathbf{v}(0)$). Again, we expect that any dependence on the initial conditions should be dissipated on a time scale $t_d = m/\gamma$. The mean-square velocity is,

$$\langle \mathbf{v}^2(t) \rangle = \mathbf{v}^2(0) \exp\left(-2\frac{\gamma}{m}t\right) + \frac{\exp\left(-2\frac{\gamma}{m}t\right)}{m^2} \int_0^t dt' \int_0^t dt'' \langle \xi(t') \xi(t'') \rangle \exp\left(-\frac{\gamma}{m}(t' + t'')\right). \quad (4.7)$$

Since $\langle \xi_i(t') \xi_j(t'') \rangle = \Gamma \delta(t' - t'') \delta_{ij}$,

$$\langle \mathbf{v}^2(t) \rangle = \mathbf{v}^2(0) \exp\left(-2\frac{\gamma}{m}t\right) + \frac{d\Gamma}{2m\gamma} \left(1 - e^{-2\frac{\gamma}{m}t}\right), \quad (4.8)$$

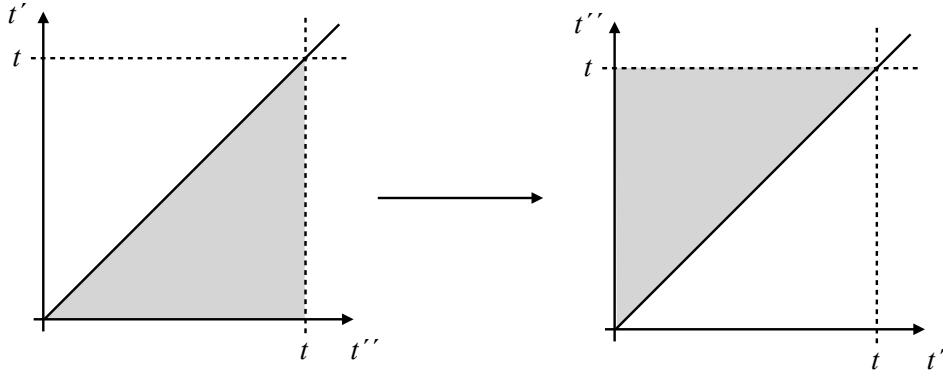


Figure 4.3: **Limits of integration.** Schematic representation of the domain of integration for the double integral in Eq. (4.9). To derive Eq. (4.11), one swaps the order of integration, which implies in turn changing the limits of integration (see Eq. (4.10)). The gray regions represent the region of integration, before (left) and after (right) after swapping the order of integration.

where d is the spatial dimension. From the equipartition theorem, the equilibrium kinetic energy is $dk_B T/2$, which is only obtained asymptotically ($t \rightarrow \infty$) if $\Gamma = 2\gamma k_B T$, where k_B is the Boltzmann constant.

Let us consider now the position $\mathbf{r}(t)$ of the particle at time t . Since $\dot{\mathbf{r}}(t) = \mathbf{v}(t)$, from Eq. (4.5) we obtain,

$$\mathbf{r}(t) = \mathbf{r}(0) + \mathbf{v}(0) \frac{m}{\gamma} \left(1 - e^{-\frac{\gamma}{m}t} \right) + \int_0^t dt'' \int_0^{t''} dt' \exp\left(\frac{\gamma}{m}(t' - t'')\right) \xi(t'). \quad (4.9)$$

As $\xi(t')$ does not depend on t'' , one can integrate over t'' the third term on the right-hand side. However, to swap the order of integration, one also needs to change the limits of integration. Figure 4.3 shows a schematic representation of the region over which one wants to integrate. It is clear that

$$\int_0^t dt'' \int_0^{t''} dt' \exp\left(\frac{\gamma}{m}(t' - t'')\right) \xi(t') = \int_0^t dt' \int_{t'}^t dt'' \exp\left(\frac{\gamma}{m}(t' - t'')\right) \xi(t'), \quad (4.10)$$

and so Eq. (4.9) simplifies to

$$\mathbf{r}(t) = \mathbf{r}(0) + \mathbf{v}(0) \frac{m}{\gamma} \left(1 - e^{-\frac{\gamma}{m}t} \right) + \frac{m}{\gamma} \int_0^t dt' \xi(t') \left(1 - e^{\frac{\gamma}{m}(t'-t)} \right). \quad (4.11)$$

If the initial position $\langle \mathbf{r}(0) \rangle$ is the same for all samples,

$$\langle \mathbf{r}(t) \rangle = \mathbf{r}(0) + \mathbf{v}(0) \frac{m}{\gamma} \left(1 - e^{-\frac{\gamma}{m}t} \right). \quad (4.12)$$

Squaring Eq. (4.11), we obtain

$$\begin{aligned} \langle \mathbf{r}^2(t) \rangle &= \mathbf{r}^2(0) + 2\mathbf{r}(0)\mathbf{v}(0) \frac{m}{\gamma} \left(1 - e^{-\frac{\gamma}{m}t} \right) + \mathbf{v}^2(0) \left(\frac{m}{\gamma} \right)^2 \left(1 - e^{-\frac{\gamma}{m}t} \right)^2 + \frac{dm\Gamma}{2\gamma^3} g(t), \\ g(t) &= 2\frac{\gamma}{m}t - 3 + 4e^{-\frac{\gamma}{m}t} - e^{-2\frac{\gamma}{m}t}. \end{aligned} \quad (4.13)$$

Thus, for $t \ll t_d = m/\gamma$ the motion is ballistic, while for $t \gg t_d$ the motion is diffusive, with $\langle \mathbf{r}^2(t) \rangle \sim d\Gamma t/\gamma^2$ (see Problem 4.1). Since $\Gamma = 2\gamma k_B T$, then

$$\langle \mathbf{r}^2(t) \rangle \sim 2d \left(\frac{k_B T}{\gamma} \right) t. \quad (4.14)$$

By comparison to the solution of the diffusion equation (see next Section), the diffusion constant D should equal the term inside the brackets, i.e.,

$$D = \frac{k_B T}{\gamma}. \quad (4.15)$$

4.3 Fokker–Planck equation

The explicit solution of Eq. (4.4) depends strongly on the specific realization of the noise term ξ . Thus, the discussion of the properties of an individual trajectory is meaningless. In the previous section, to draw general conclusions about the statistical properties of an ensemble of trajectories, we have considered averages $\langle \dots \rangle$ over a large number of realizations of the noise. In this section, we discuss the *Fokker–Planck approach*, named after Adriaan Fokker and Max Planck [12, 13]. It consists in deriving a partial differential equation to describe the time evolution of the distribution function $p(\mathbf{r}, t)$, defined such that $p(\mathbf{r}, t)dV$ is the probability of finding a Brownian particle in a volume dV around \mathbf{r} at time t .

Let us define $W(\mathbf{r}, \mathbf{r}')$ as the probability per unit time per unit volume that the Brownian particle moves from \mathbf{r} to \mathbf{r}' . The master equation for the time evolution of $p(\mathbf{r}, t)$ is then

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = \int d\mathbf{r}' [p(\mathbf{r}', t)W(\mathbf{r}', \mathbf{r}) - p(\mathbf{r}, t)W(\mathbf{r}, \mathbf{r}')], \quad (4.16)$$

where the gain factor is due to the motion of the particle from the neighborhood of \mathbf{r}' to \mathbf{r} and the lost term due to the motion in the opposite direction. In this coarse-grained approach, we assume that the relevant time scale is much larger than t_d and thus inertia effects are negligible (limit known as overdamped regime). The equation of motion is then

$$\gamma \dot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) + \xi, \quad (4.17)$$

where $\mathbf{F}(\mathbf{r})$ is an external force (see Problem 4.2).

In the overdamped regime, large displacements of the Brownian particle are negligible and $W(\mathbf{r}, \mathbf{r}')$ should decay rapidly with the absolute value of the distance $\chi = \mathbf{r}' - \mathbf{r}$. If we write Eq. (4.16) in terms of \mathbf{r} and χ and perform a Taylor expansion to second order around χ , we obtain

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = \int d^d \chi \left[-\nabla(p(\mathbf{r}, t)W(\mathbf{r}, \chi)) \cdot \chi + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial r_i \partial r_j} (p(\mathbf{r}, t)W(\mathbf{r}, \chi)) \chi_i \chi_j \right], \quad (4.18)$$

where r_i and r_j (χ_i and χ_j) are components of \mathbf{r} (χ) (see Problem 4.3).

Defining the quantities

$$\alpha_i(\mathbf{r}) = \int d^d \chi W(\mathbf{r}, \chi) \chi_i \quad (4.19)$$

and

$$\beta_{ij}(\mathbf{r}) = \int d^d \chi W(\mathbf{r}, \chi) \chi_i \chi_j, \quad (4.20)$$

we obtain the Fokker–Planck equation,

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = - \sum_i \frac{\partial}{\partial r_i} (\alpha_i(\mathbf{r})p(\mathbf{r}, t)) + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial r_i \partial r_j} (\beta_{ij}(\mathbf{r})p(\mathbf{r}, t)). \quad (4.21)$$

Here, we considered the position \mathbf{r} , but the same equation can be derived for any vectorial quantity defining the state of a system [4].

For a Brownian particle, $W(\mathbf{r}, \chi)$ is the probability per unit volume and time that a particle in the neighborhood of \mathbf{r} is displaced by χ . Thus, the quantities α_i and β_{ij} are related to average quantities. In particular,

$$\alpha_i = \frac{\langle \Delta r_i \rangle}{\Delta t} \quad (4.22)$$

is the average displacement per unit time along the i -th direction and

$$\beta_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\Delta t} \quad (4.23)$$

is the average square displacement per unit time. For a stochastic motion given by Eq. (4.17), with a constant force \mathbf{F}_0 , $\alpha_i = F_i/\gamma$, where F_i is the i -th component of the force \mathbf{F}_0 and $\beta_{ij} = 2\delta_{ij}D$ (see previous section). Thus,

$$\frac{\partial p(\mathbf{r}, t)}{\partial t} = -\frac{\mathbf{F}_0}{\gamma} \cdot \nabla p(\mathbf{r}, t) + D \nabla^2 p(\mathbf{r}, t). \quad (4.24)$$

In the absence of an external force and no boundaries, the diffusion equation is recovered, whose solution for $p(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)$ is

$$p(\mathbf{r}, t) = \left[\frac{1}{4\pi D t} \right]^{\frac{d}{2}} \exp \left[-\frac{(\mathbf{r} - \mathbf{r}_0)^2}{4 D t} \right], \quad (4.25)$$

which corresponds to a Gaussian distribution in dimension d (see Problem 4.4).

4.4 Active Brownian motion

Active systems are made of interacting elementary constituents, or particles, able to convert some source of energy from the environment to self-propel. Self-propelled particles can be found across different scales, from molecular motors inside the cell, to large animal groups. At the micro-scale, examples include both living and non-living systems, such as flagellated bacteria or self-catalytic colloidal suspensions. The constant energy input needed to sustain the self-propulsion of each constituent drives these systems out of equilibrium. It is worth noticing that active matter breaks detailed balance at the level of a single particle, in contrast to driven passive matter, for which the departure from equilibrium is due to a global external stimuli. As a consequence, active systems exhibit a rich variety of emergent states where their elementary units self-organize in ways that are hard (if not impossible) to attain in passive systems.

Despite such phenomenological richness, many activity-induced collective phenomena can be observed in a small number of paradigmatic minimal models. Surely inspired by the success of the theory of critical phenomena, quite some efforts have been devoted to the study of these simplified models, trying to identify some sort of *universality classes*. Here we shall just present a sub-set of these, focusing on salient physical ingredients generically encountered in active systems and from which one can rationalize widespread collective phenomena, such as the emergence of collective motion, or flocking, and, on the other hand, particle clustering in the absence of attractive interactions.

We now introduce and review the main features of simple models of self-propelled particles. We mainly focus on overdamped Langevin formulations of model self-propelled particles, based on *active Brownian particles* (ABP) and on *active Ornstein–Uhlenbeck particles* (AOUP), as they likely are the simplest and more studied active particles models. We also show how to include different ingredients that might be relevant in many different situations, such as alignment interactions or inertia. As for passive particles, we move to the Fokker–Planck representation of these particle-based stochastic models to present a general route to consistently derive continuum descriptions of these model systems.

As we shall see, self-propulsion can be easily included by introducing persistence, i.e., memory, in an otherwise Brownian walker. This can be done in different ways, giving rise to a variety of models. The key point is to introduce non-thermal fluctuations that account, in a somehow coarse-grained scale, for the energy conversion mechanisms involved in self-propulsion. Other models have tried to consider the details of such mechanisms at smaller scales for specific systems. For instance, to describe a self-propelled colloid, swimming in a solvent at low Reynolds number, one might need to consider also interactions mediated by the environment, such as hydrodynamic and/or phoretic interactions. But these are not what we consider *simple models* anymore, and will not discuss them in this Chapter. We mainly focus on single particle properties and only very briefly discuss collective effects.

4.4.1 Active Brownian particles

We start by considering a particularly simple and widely studied stochastic model of self-propelled particles called, active Brownian particles (ABP) [14]. Such model has been (and still is) commonly used as a first simplified description of the persistent motion of a self-phoretic active colloid [15, 16, 17], where a force mimicks the effect of the local gradients propelling the particle. At this scale, inertia effects can be neglected. However, at larger scales, if one considers macroscopic objects such as robots or swimmers [18], inertia has to be considered. The ABP model can be readily extended to account for inertia, as we shall see below. On top of such self-propulsion force, one also has to consider dissipation and inter-particle interactions. In the following, we shall restrict ourselves to pairwise mutual interactions, usually deriving from a potential. The interest of the ABP model presented here goes well beyond the description of self-propelled colloids as it constitutes a central paradigm to investigate how self-propulsion affects the behavior of interacting Brownian particles. As

we will show, this model allows us to incorporate subsequently different key physical ingredients and explore situations of increasing complexity within a unified framework.

Non-interacting case

An Active Brownian Particle, located in $\mathbf{r}(t)$ at time t , and moving in a 2D plane, is governed by the following overdamped Langevin equations (see Eq. (4.2) and discussion around for passive particles)

$$\dot{\mathbf{r}}(t) = v_0 \mathbf{e} + \sqrt{2D_0} \boldsymbol{\xi}(t), \quad \dot{\phi}(t) = \sqrt{2D_r} \eta(t), \quad (4.26)$$

where (each component of) $\boldsymbol{\xi}(t)$ and $\eta(t)$ are Gaussian white noises with zero mean and unit variance (as in the case of passive particles):

$$\langle \boldsymbol{\xi}(t) \boldsymbol{\xi}(t') \rangle = \delta(t - t') \mathbf{1}, \quad \langle \eta(t) \eta(t') \rangle = \delta(t - t'), \quad \langle \boldsymbol{\xi}(t) \rangle = \langle \eta(t) \rangle = 0. \quad (4.27)$$

It follows that¹ (see Problem 4.5.1)

$$\langle \mathbf{e}(t) \cdot \mathbf{e}(t') \rangle = e^{-D_r(t-t')}, \quad t > t'. \quad (4.28)$$

The coefficients D_0 and D_r thus set the amplitude of the two noise terms. Particles self-propel at constant *self-propulsion speed* v_0 in the direction given by $\mathbf{e} = (\cos \varphi, \sin \varphi)$ — which in two dimensions can be parameterized by a single scalar φ . The orientation of the particle follows itself a diffusive process, its self-propulsion direction smoothly de-correlates in time. As such rotational diffusion is coupled to translational motion via v_0 , it introduces a new characteristic time scale and length scale in the dynamics: the *persistence time* $\tau = D_r^{-1}$ and the *persistence length* $\ell = v_0 \tau$, respectively. In the absence of activity, say $v_0 = 0$, we recover Brownian motion of particles coupled to an equilibrium thermal bath at temperature $k_B T = D_0 \gamma_0$, where γ_0 is the damping coefficient, which at the single particle level can be absorbed in the definition of the time unit.

In order to quantify the activity, it is convenient to define a non-dimensional Péclet number, Pe , comparing advection of the active particle against its diffusion due to thermal fluctuations. Popular choices are the rotational and the translational Péclet numbers, defined as

$$\text{Pe}_r = \frac{\ell}{\sigma} = \frac{v_0}{\sigma D_r}, \quad \text{Pe}_t = \frac{\gamma_0 v_0 \sigma}{k_B T} = \frac{v_0}{D_0/\sigma}, \quad (4.29)$$

where $\sigma = 2R$ is the diameter of the particle, which from now on we take as our unit of length. Although most studies focus on the 2D case, such model can be easily extended to higher dimensions [19, 20, 21, 22, 23]:

$$\dot{\mathbf{r}}(t) = v_0 \mathbf{e} + \sqrt{2D_0} \boldsymbol{\xi}(t), \quad \dot{\mathbf{e}}(t) = \sqrt{2D_r} \mathbf{e}(t) \times \boldsymbol{\eta}(t), \quad (4.30)$$

where now both the position and orientation vector are defined in d -dimensions (different from $d = 1$). Such dynamics introduces persistence in the random motion of the particle over a time scale $\tau = D_r^{-1}$.

Like in the passive case, in order to understand the random motion performed by an ABP, we look into the *mean-squared displacement* (MSD)

$$\Delta^2(t - t_0) = \langle (\mathbf{r}(t) - \mathbf{r}(t_0))^2 \rangle, \quad (4.31)$$

where the brackets denote averages over the noise(s). This quantity can be computed exactly, and reads [24] (see Problem 4.5)

$$\Delta^2(t - t_0) = 2dD_0(t - t_0) + 2v_0^2 \tau^2 \left[e^{-(t-t_0)/\tau} + (t - t_0)/\tau - 1 \right]. \quad (4.32)$$

From this MSD, it is straightforward to identify a short time regime, always with respect to the characteristic time scale in the problem, i.e. $t \ll \tau$, in which $\Delta^2(t) \rightarrow (v_0 t)^2$, meaning that ABP move ballistically, with a velocity v_0 . At long times though, $t \gg \tau$, the particle eventually loses memory of its orientation as angular noise randomizes it, and its motion becomes diffusive $\Delta^2(t) \rightarrow 2dD_{at}$, with an enhanced effective diffusivity

$$D_a = D_0 + \frac{v_0^2 \tau}{d} \propto \ell^2 / \tau. \quad (4.33)$$

¹In d dimensions, $\langle \mathbf{e}(t) \cdot \mathbf{e}(t') \rangle = e^{-(d-1)D_r(t-t')}$.

The crossover time between the ballistic and diffusive regime is given by τ . The persistent random walk described by a single ABP is depicted in Fig. 4.4a.

Active particles, ranging from microorganisms to self-propelled colloids, exhibit asymptotic diffusive behavior. ABP recovers this diffusive behavior at long times, which has been experimentally measured, and can be orders of magnitude larger than the thermal diffusion of the corresponding passive particles. For example, a passive micron size particle at room temperature has $D_{\text{th}} \sim 10^{-9} \text{ cm}^2 \text{s}^{-1}$, while an active Janus colloid of similar size can have $D \sim 10^{-8} \text{ cm}^2 \text{s}^{-1}$ or for *E. coli* $D \sim 10^{-6} \text{ cm}^2 \text{s}^{-1}$.

Active matter is grossly defined as collections of interacting units that evolve out-of-equilibrium at the level of their elementary constituents. Does a single ABP evolve out-of-equilibrium? and if so, how can we quantify its departure from equilibrium? In order to tackle these questions, one can quantify the violations, if any, of the *fluctuation-dissipation theorem* (FDT) (see Chapter 3). The latter stands that, for any equilibrium system gently driven out-of-equilibrium by a small perturbation h , conjugated to an observable B , the response of a given observable, say A , is given by the conjugated equilibrium correlation function, namely

$$R_A(t, t') = \frac{\delta \langle A \rangle_t}{\delta h(t')} = \frac{1}{k_B T} \frac{\partial}{\partial t'} \langle A(t) B(t') \rangle_{\text{eq}}, \quad t > t', \quad (4.34)$$

where $\langle \cdot \rangle_{\text{eq}}$ denotes the average over the equilibrium distribution and $\langle \cdot \rangle_t$ over different realizations of the dynamics up to a time t . Any violation of this relation thus means the dynamics is out-of-equilibrium. If we perturb a passive Brownian particle by applying a constant force $\mathbf{f} = f \mathbf{u}_x$, the FDT implies the following equality

$$\Delta^2(t, t') = 2k_B T \chi(t, t'), \quad \chi(t, t') = (t - t')/\gamma_0. \quad (4.35)$$

For an ABP, this relation still holds, but with an *effective temperature* replacing the thermal bath one [25, 26]:

$$\Delta^2(t, t') = 2k_B T_{\text{eff}}(t, t') \chi(t, t'). \quad (4.36)$$

In the ballistic regime, $\tau \gg t - t'$, $k_B T_{\text{eff}}(t, t') = k_B T$, while in the diffusive, long-time regime:

$$k_B T_{\text{eff}}(t, t') = \gamma_0 D_a = k_B T + \gamma_0 \frac{v_0^2 \tau}{d(d-1)}, \quad (4.37)$$

leading to an extended Stokes-Einstein relation. Overall, a single ABP moves asymptotically as a classical Brownian walker, but at a higher effective temperature $T_{\text{eff}} \propto \ell^2/\tau \propto \text{Pe}^2$. Similarly, without any equilibrium assumption, one can define an active contribution to the mechanical pressure and compute the resulting pressure of a non-interacting system of ABP at density ρ , resulting into an effective ideal gas equation of state $P_0 = \rho k_B T_{\text{eff}}$ [27, 28, 21, 29]. So, although the individual dynamics breaks detailed balance (see Problem 4.5) [26], a collection of non-interacting ABPs can be understood as an ideal gas at a higher effective temperature. As we shall see below, the non-equilibrium nature of ABPs manifests in the presence of interactions.

(Isotropic) excluded volume interactions

We consider now a 2D system of N interacting ABPs: self-propelled disks of diameter $2R$, located at \mathbf{r}_i with orientations \mathbf{e}_i , interacting via pairwise, isotropic potential. Following Eq. (4.26), the equations of motion now read

$$\dot{\mathbf{r}}_i(t) = v_0 \mathbf{e}_i + \gamma_0^{-1} \mathbf{F}_i(\{\mathbf{r}_j(t)\}) + \sqrt{2D_0} \boldsymbol{\xi}_i(t), \quad \dot{\phi}_i(t) = \sqrt{2D_r} \eta_i(t), \quad (4.38)$$

where the sub-index i runs from 1 to N . All particles self-propel at the same constant speed v_0 in the direction given by \mathbf{e}_i , which diffuses independently for each particle, and $\mathbf{F}_i = -\nabla_i \sum_{j \neq i} u(r_{ij})$. Most studies have focused on purely repulsive interactions, trying to understand the collective behavior arising from the competition between self-propulsion and crowding effects. Indeed, as we shall see, these two ingredients are sufficient to trigger a non-equilibrium phase transition.

The simplest way to introduce excluded volume interactions is to consider the hard sphere model (or disk in 2D), as it is fully controlled by a single parameter, the packing fraction $\phi = \pi N/(4L^2)$ (the temperature in this system is irrelevant). To capture the impact of activity on the fluid and solid phases of matter, it would be desirable to incorporate hard-core interactions in the ABP dynamics. Although it is conceptually appealing to use a hard-core potential (as the resulting model would be completely controlled by only two parameters, ϕ and

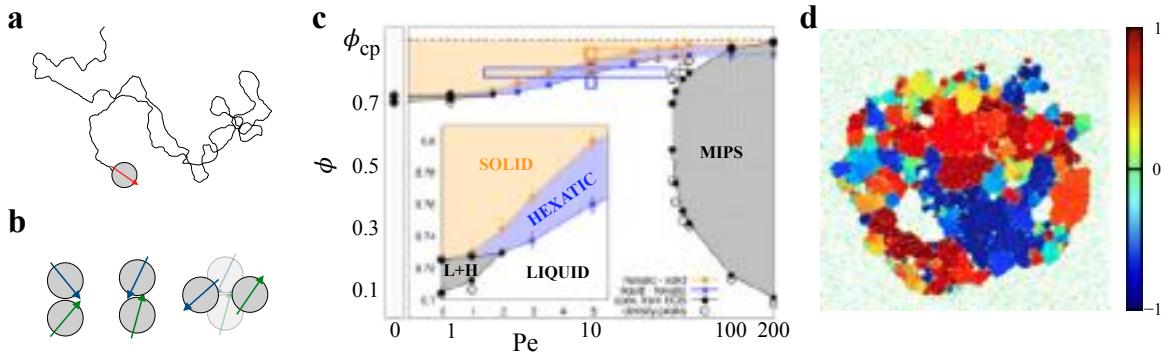


Figure 4.4: Active Brownian particles with excluded volume interactions (in 2D). (a) Typical random trajectory of a single ABP with a self-propulsion velocity $v_0 \mathbf{e}$. (b) Sketch illustrating the kinetic trapping of particles arising from the competition between excluded volume and persistence. (c) Phase diagram of ABP in the activity (Pe) surface fraction (ϕ) plane (adapted from Ref. [30]). The inset shows a detailed view of the low activity region of the phase diagram indicated by a blue ellipsis. (d) Typical configuration in the motility-induced phase separation (MIPS) coexistence regime (adapted from Ref. [31]). Parameters: $N = 512^2$, $\phi = 0.25$, $Pe = 200$. Particles are colored according to the projection of their local hexatic order on the average orientation of the system.

Pe), its discontinuity poses practical inconveniences, and one usually integrates Eq. (4.38) using a continuous potential. Different approaches have been introduced to deal with hard-core interactions, such as event-driven [32] or Monte Carlo approaches (see Section below) [33, 34, 35, 36].

The phase diagram obtained from extensive numerical simulations of ABPs with a short-range repulsive potential ($u(r) = 4\epsilon[(\sigma/r)^{64} - (\sigma/r)^{32}] + \epsilon$ if $r < 2^{1/32}\sigma$ and 0 otherwise) is shown in Fig. 4.4 (discussed in more detail in Ref. [30]). Note that it only has two axes, corresponding to Pe (activity) and the packing fraction ϕ (crowding): the potential is so stiff that, roughly speaking, we can consider the particles as being effectively hard disks. At $Pe = 0$, the system exhibits the equilibrium 2D melting scenario of hard disks described in Ref. [37] that we quickly summarize here. The melting of passive hard disks takes place in two steps: as ϕ is increased from a disordered homogeneous fluid (or liquid²), a first-order transition between the fluid and hexatic phases occurs, followed by a continuous Berezinskii–Kosterlitz–Thouless (BKT) transition between the hexatic and the solid at higher densities. The hexatic phase exhibits quasi-long-range orientational order and short-range positional one, while the solid phase has quasi-long-range positional and long-range orientational order. Fluid and hexatic phases coexist close to the fluid phase, within a narrow interval of packing fractions. In order to characterize orientational order, one defines the (local) hexatic order parameter $\psi_6(\mathbf{r}_j) = n_j^{-1} \sum_{k=1}^{n_j} e^{i6\theta_{jk}}$, where θ_{jk} is the angle between the segment that connects the center of the j -th disk and the one of its k -th nearest neighbor (among n_j neighbors). In practice, this quantity is computed using a Voronoi tessellation to define the neighborhood of a given particle. Positional order is characterized with density correlations in Fourier space, $C(r) = \langle e^{i\mathbf{q}_0 \cdot (\mathbf{r}_j - \mathbf{r}_k)} \rangle$, where the average is taken over all the pairs of particles at a distance $r = |\mathbf{r}_j - \mathbf{r}_k|$ and \mathbf{q}_0 is the wave vector corresponding to the first peak of the structure factor. We refer the reader to Ref. [38] for more details.

At finite Pe values, an active hexatic phase exists for all the explored activities (blue sector in Fig. 4.4c). Active disks arrange in a hexatic phase across a larger density range than passive ones. At even higher densities, orientational long-range and positional quasi-long-range order emerge for any activity, signaling the presence of an active solid phase (orange region). As activity increases, the liquid–hexatic and hexatic–solid transitions shift towards higher densities, meaning that one can melt the hexatic and the solid phase by increasing Pe .

In the high Pe regime, another phenomenon sets in: *motility-induced phase separation* (MIPS). Chapter 10 is fully devoted to this phenomenon, and we shall here just very briefly describe it. As sketched in Fig. 4.4b, when two persistent particles collide, they block each other until their orientation changes enough to be able to escape and stop colliding. Such kinetic arrest is responsible for the aggregation of active particles: if the time needed to reorient, controlled by τ , is longer than the time it takes for another particle to join the first two, a

²In the absence of attractive forces, there is no liquid–gas transition and both phases are indistinguishable.

cluster will grow. Thus at high enough density (faster collision rate) and activity, one can eventually observe the system phase separating into a macroscopic dense cluster surrounded by a dilute gas-like phase, a phenomenon known as MIPS [39]. MIPS arises from the mere competition between volume-exclusion and persistence, and as such it occurs in any model of (overdamped) active particles based on the same ingredients, such the AOUP, run-and-tumble, Monte Carlo and lattice models (see Refs. [40, 41, 34, 42, 43, 44]) . A typical snapshot of the system in this regime is shown in Fig. 4.4d, where, on top of showing the aggregation phenomenon, we also show the hexatic structure of the dense drop generated through MIPS.

From the analysis of the local density distribution and the equations of state (pressure as a function of density), one can construct the coexistence region associated to MIPS shown in Fig. 4.4c. From $\text{Pe} \geq 35$, the coexistence region broadens as activity increases, and it eventually crosses the hexatic and solid transition lines. Activity can thus trigger a phase separation between a dilute and a high-density phase, which can either be liquid, hexatic, or solid.

We refrain from discussing more subtle aspects that have been studied in more recent works, such as the nature of the phase transitions described in this section, the nature of the MIPS critical point, or the phase separation kinetics across MIPS [31, 45, 46, 47].

Alignment interactions

If ABP with isotropic repulsive interactions constitute the paradigm for MIPS, self-propelled particles with velocity alignment interactions constitute the paradigm for flocking. Flocking stands for the emergence of collective motion in systems of self-propelled agents. It is arguably the most studied and earliest introduced *universality class* in active matter [48], popularized by the celebrated Vicsek model [49, 50] and the murmuration of starling flocks [51]. The ABP model can be extended to include aligning interactions in the following way [52, 53, 54]:

$$\dot{\mathbf{r}}_i(t) = v_0 \mathbf{e}_i + \gamma_0^{-1} \mathbf{F}_i(\{\mathbf{r}_j(t)\}) + \sqrt{2D_0} \boldsymbol{\xi}_i(t), \quad (4.39)$$

$$\dot{\varphi}_i(t) = -\gamma_r^{-1} \frac{\partial H}{\partial \varphi_i} + \sqrt{2D_r} v_i(t). \quad (4.40)$$

Now, the time evolution of the orientation, or self-propulsion direction, is advected by a local torque $T_i = -\gamma_r^{-1} \frac{\partial H}{\partial \varphi_i}$ that one can, for instance, define as a Vicsek-like rule. It is worth noting that, in the classical Vicsek model self-propelled particles are point-like, they do not interact through a repulsive potential but only via the alignment rule. Thus, the present model allows to explore both the role of alignment and excluded volume interactions.

For the sake of generality, we derive local torques from the following potential

$$H = - \sum_{i,j} v(r_{ij}) w(\varphi_{ij}), \quad (4.41)$$

where $\varphi_{ij} = \varphi_i - \varphi_j$ and

$$v(r_{ij}) = \begin{cases} \frac{2}{\pi R_\varphi^2} (r_{ij} - R_\varphi)^2 & r_{ij} < R_\varphi \\ 0 & r_{ij} > R_\varphi \end{cases} \quad (4.42)$$

is a spatially decaying function with a cutoff distance R_φ , setting the interaction range beyond which particles do not align. The spatial dependency of $v(r_{ij})$ ensures that there are no discontinuities in the resulting torque. The angular dependency of the alignment potential, $w(\varphi_{ij})$, is chosen in different ways in order to study the interactions sketched in Fig. 4.5a: ferromagnetic and nematic alignment.

Ferromagnetic torques lead to a Langevin variant (in continuous time) of the Vicsek model, which, in the limit of vanishing velocities, reduces to the equilibrium 2D XY model of a ferromagnet. Indeed, a ferromagnetic coupling in Eq. (4.40) can be derived from the classical 2D XY Hamiltonian

$$H = -J \sum_i \sum_{j \in \omega_i} v(r_{ij}) \mathbf{e}_i \cdot \mathbf{e}_j, \quad (4.43)$$

where $w(\varphi_{ij}) = J \cos(\varphi_{ij})$, ω_i is the vicinity of particle i , defined by R_φ , and $J > 0$ is the coupling strength. This interaction describes particles, or agents, that tend to move along the same direction, just as spins in a ferromagnetic material tend to align. As the structure of this interaction is inherited from a spin Hamiltonian,

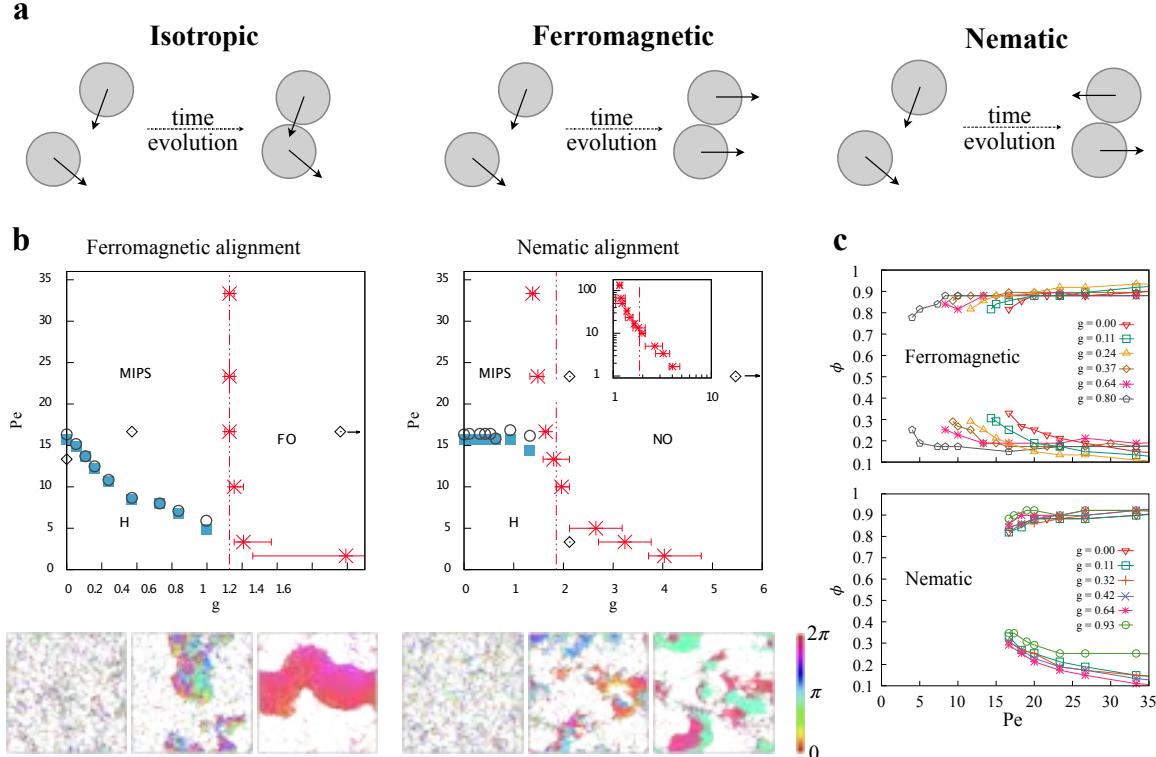


Figure 4.5: Active Brownian particles with excluded volume and velocity alignment interactions. (a) Schematic representation of the model described by Eqs. (4.39) and (4.40), illustrating APB with (i) only isotropic excluded volume interactions, (ii) ferromagnetic alignment, and (iii) nematic alignment (third column). (b) Phase diagrams in the (g, Pe) plane for a system at fixed packing fraction $\phi = 0.4$ for ferromagnetic (left) and nematic (right) alignment. The phase separation from a homogeneous (H) to a MIPS state is indicated by blue squares (grey circles indicate the prediction from a continuum theory derived as explained in the next section). The onset of ferromagnetic order (FO) and nematic order (NO) is marked by red symbols. The inset in (b) shows the onset of nematic order up to a value of $Pe = 133.4$ in log-log scale. Broken lines correspond to $g = g_f = 0.93$ (left) and $g = 2g_f$ (right). Below we show snapshots of the system at different points of the phase diagrams (indicated by symbols): $(0, 13.3)$; $(0.37, 16.6)$; $(7.96, 16.6)$ in the ferromagnetic case; $(2.12, 3.3)$; $(2.12, 23.4)$; $(13.23, 23.4)$ in the nematic case. Particles are colored according to their self-propulsion direction, using a cyclic color code. (c) Phase coexistence regions for several values of the ferromagnetic and nematic coupling. The coexistence region shifts to lower values of Pe as the ferromagnetic coupling is increased, while increasing the nematic coupling does not significantly affect it. Adapted from Ref. [54].

it is additive, as opposed to the agent-based alignment formulated by Vicsek where the individual interaction term is normalized by its number of neighbors, which depends on time (see Ref. [55] for a discussion).

Particles that align along a given axis but with head-tail symmetry are also typically encountered, see Fig. 4.5a. Such nematic alignment is relevant to model systems of self-propelled elongated objects, such as bacteria or microtubules and constitutes a whole active matter class called self-propelled rods [56]. Nematic alignment can be modeled by the following Hamiltonian

$$H = -J \sum_i \sum_{j \in \omega_i} v(r_{ij}) \mathbf{q}_i \cdot \mathbf{q}_j, \quad (4.44)$$

where $\mathbf{q}_i = \mathbf{e}_i \otimes \mathbf{e}_i - \frac{1}{2}\mathbf{1}$ is the nematic director [57]. Note that Eq. (4.44) has the same structure as the 2D XY model interaction, but now the nematic tensor plays the role of the orientation of the particles (spins) [58]. Then, one gets $w(\varphi_{ij}) = J \cos(2\varphi_{ij})$. Alignment, both in the ferromagnetic and nematic case, introduces an extra controlled parameter, the coupling strength J . As usual, it is convenient to manipulate dimensionless parameters, such as the reduced coupling

$$g = \frac{2J}{\pi R_\phi^2 \gamma D_r}. \quad (4.45)$$

The different collective states exhibited by this model system are summarized in the phase diagrams in Fig. 4.5b. At low couplings, the orientation of the particles is dominated by noise, and the system is in a disordered state. Upon increasing g , the competition between alignment and noise triggers a phase transition towards a globally oriented state, whose orientation depends on the symmetry of the interaction.

For ferromagnetic alignment, the ordered phase displays polar order and strong density heterogeneities in the form of lanes, bands and clusters, as illustrated in the snapshots Fig. 4.5b, along which a macroscopic fraction of the system moves (roughly) along the same direction. The formation of dense structures such as traveling bands are typical in ferromagnetic Vicsek-style models [59]. The extensions of such models including volume interactions display richer structures, and among them, it is typical to find lanes [60, 61, 53]. The onset of ferromagnetic order. The value of the critical coupling beyond which ferromagnetic order sets in, is largely independent of the value of Pe.

For nematic alignment, the ordered state is characterized by a macroscopic fraction of particles self-propelling along the same axis but not the same directions. We observe again the formation of dense structures, although of different nature than in the ferromagnetic case. Nematic particles aggregate into coherently-moving structures, with a high degree of local polar order. These domains collide and interpenetrate, giving rise to nematic bands, see Fig. 4.5b. In systems of elongated self-propelled particles that align through collisions, one generically observes the formation of polar lanes, similar to the ones we observe here in the ferromagnetic case at high enough densities and shape anisotropy [62, 63, 64, 65]. We also find a tendency towards polar ordering in this model of self-propelled disks, although only locally. The onset of order has a stronger dependence on the value of Pe in the nematic than in the ferromagnetic case. The higher the Pe is, the lower the alignment strength has to be to trigger nematic ordering [66].

For $g = 0$, we are back to the ABP model with only excluded volume interactions which exhibits MIPS for high enough Pe and ϕ . At finite g below the emergence of orientational order, there is a region of the phase diagram where we still find MIPS. As shown in Figs 4.5b-c, ferromagnetic alignment shifts MIPS to lower values of Pe, as the coupling parameter g is increased. Ferromagnetic alignment enhances the aggregation of particles and the eventual phase separation of the system. In contrast, for nematic alignment, the critical Pe at which MIPS takes place remains fairly unchanged (as also does the coexistence region) as the coupling parameter g is increased.

4.4.2 Underdamped active Brownian particles

So far, we have only considered stochastic dynamics in the overdamped limit. This is a reasonable simplification if we have in mind microswimmers moving at low Reynolds number. At larger scales though, inertia is likely to be relevant. An overdamped extension of the ABP model in Eq. (4.38) is given by the following Langevin equations [67, 68, 69]:

$$m\ddot{\mathbf{r}}(t) = -\gamma_0 \dot{\mathbf{r}}(t) + \mathbf{F}_i + F_a \mathbf{e}_i + \gamma_0 \sqrt{2D_0} \boldsymbol{\xi}(t), \quad (4.46)$$

$$I\ddot{\phi}_i(t) = -\gamma_r \dot{\phi}_i(t) + \gamma_r \sqrt{2D_r} \eta_i(t). \quad (4.47)$$

Self-propulsion is modelled as a constant magnitude active force active $\mathbf{F}_a = F_a \mathbf{e}_i$ pointing along the orientation of the particles, $\mathbf{e}_i = (\cos \phi_i, \sin \phi_i)$. The moment of inertia I and the rotational damping γ_r have to be introduced in order to have a consistent underdamped picture. In the overdamped limit ($m, I \rightarrow 0$), we recover the ABP model discussed above with $F_a = \gamma_0 v_0$. We will not discuss here the phenomenology that inertia brings into the physics of ABPs. We just want to mention that, recent investigations have shown that, while translational inertia suppresses MIPS [67], rotational inertia favors it [69, 70].

4.4.3 Run and tumble motion

Run-and-tumble motion refers to a simplified description of the dynamics of flagellated bacteria such as *Escherichia coli* [71]. These microorganisms, as they swim at low Reynolds numbers, follow straight lines for some period of time, called a *run*, and then reorient at a much shorter time scale, picking up a new direction of motion at random, this is called a *tumble*. The run-and-tumble model considers ballistic runs at constant velocity v_0 and instantaneous, Poisson distributed, tumbles occurring at a fixed rate α . The tumbles decorrelate the self-propulsion direction at long time and length scales, larger than the persistence time $\tau = \alpha^{-1}$ and the persistence length $\ell = v_0 \tau$ [72]. Such persistent random walk is reminiscent of ABP: it is characterized by a persistence time associated to the particle's reorientation. Indeed, there is an equivalence between reorientations induced by run-and-tumble and rotational Brownian motion as in the ABP model [73], and in both cases the long-time diffusion coefficient scales as $D_a \propto \ell^2/\tau$. MIPS was originally identified in systems of run-and-tumble particles [74].

4.4.4 Active Ornstein–Uhlenbeck particles

We now consider a similar model of self-propelled particles, where memory is introduced through the famous Ornstein–Uhlenbeck process, the stochastic process describing the velocity of an inertial Brownian particle, relaxing over a time scale $\tau = m/\gamma$ (see discussion on the passive case in Sec. 4.2). This model was originally introduced in the context of active matter in Ref. [75] as a continuous time version of the kinetic Monte Carlo model introduced in [33, 34] (described below). The Langevin equations governing the motion of AOUP read

$$\dot{\mathbf{r}}_i(t) = \gamma_0^{-1} \mathbf{F}_i + \mathbf{v}_i, \quad \dot{\mathbf{v}}_i(t) = -\frac{\mathbf{v}_i}{\tau} + \sqrt{\frac{2D_0}{\tau^2}} \boldsymbol{\eta}_i(t), \quad (4.48)$$

where \mathbf{F}_i is a conservative force acting on particle i (can be due to interparticle interactions or an external field) and \mathbf{v}_i is the self-propulsion velocity. Note that, contrary to ABP, the self-propulsion velocity in this case fluctuates both in orientation and amplitude. Like in the ABP model, self-propulsion introduces persistence in the spatio-temporal dynamics of the active particles via the autocorrelation function of the self-propulsion velocity

$$\langle \mathbf{v}_i(t) \mathbf{v}_j(t') \rangle = \frac{D_0}{\tau} e^{-|t-t'|/\tau} \delta_{ij} \mathbf{1}, \quad (4.49)$$

which reduces to passive (equilibrium) Brownian motion in the limit $\tau \rightarrow 0$, for which

$$\langle \mathbf{v}_i(t) \mathbf{v}_j(t') \rangle \rightarrow 2D_0 \delta(t-t') \delta_{ij} \mathbf{1}, \quad (4.50)$$

where $\mathbf{1}$ stands for the unit matrix. Although a standard thermal noise could be added into the Langevin equation of AOUP, such contribution is assumed to be small with respect to the active noise \mathbf{v} and might be considered redundant, as it is not needed to recover equilibrium. This model is convenient for analytical treatments, allowing for a perturbative expression of the stationary distribution in the limit of small relaxation time [76, 77, 26]. The AOUP model can be thought of as an approximate treatment of the ABP dynamics [78]. Indeed, the reduced ABP dynamics obtained from the integration of the angular variables in Eq. (4.38) can be related, in the absence of translational noise, to the AOUP dynamics by setting $v_0^2/(2\gamma_0)$ (in ABP) to D_0/τ (in AOUP). See [79] for an underdamped version of the model.

4.4.5 Active particles with an energy depot

We now present one of the earliest models of active particles where the energy intake of each particle from the environment in order to self-propel is made explicit. It is somehow more complex than ABP-like models, where self propulsion is just directly added via a persistent process with typical time τ . However, as in this model the energy exchanges between the particles and the environment have a clearer thermodynamic interpretation, it might be more suitable than ABP-like models to construct a stochastic thermodynamics for active matter [80, 81] (although most works deal with ABP [82, 83] and AOUP [76, 84] anyway). In the energy depot model, each particle stocks energy extracted from the environment, and converts it into kinetic energy, to self-propel. The dynamics of such internal energy ε depot is given by [85, 86]

$$\dot{\varepsilon}_i(t) = q - c\varepsilon_i(t) - d\mathbf{v}_i^2\varepsilon_i(t), \quad (4.51)$$

where $\mathbf{v}_i = \dot{\mathbf{r}}_i$ is the velocity of a particle located in \mathbf{r}_i . The parameters q , c and d quantify the rate of energy intake, energy loss (in the form of dissipation) and conversion into useful (here kinetic) energy, respectively. This latter point becomes more concrete once the dynamics of the particles in real space are specified. The equation of motion of the particles are governed by the following Langevin equation

$$m\ddot{\mathbf{r}}_i = -(\gamma_0 - d\varepsilon_i(t))\dot{\mathbf{r}}_i + \mathbf{F}_i + \sqrt{2D_0}\boldsymbol{\xi}_i(t), \quad (4.52)$$

where $\boldsymbol{\xi}$ is a Gaussian white noise of zero mean and unit variance, mimicking a thermal bath at temperature T and \mathbf{F}_i an external force. Particles are subjected to a damping force, characterized by γ_0 , verifying the Stokes-Eistein relation $\gamma_0 D = k_B T$. The energy depot adds a new, self-propulsion-like, force. It appears as an advective term, which can be cast as an effective velocity-dependent damping coefficient $\gamma(\mathbf{v}) = \gamma_0 - d\varepsilon(t)$. If $\gamma(\mathbf{v}) < 0$ the particle is accelerated by the effect of the energy depot, while in the opposite case it is still damped but with a smaller damping coefficient. In the adiabatic limit, once we consider that the energy depot is filled at much shorter time scale than the relaxation of the velocity, one can further simplify the dynamics and write

$$\gamma(\mathbf{v}) = \gamma_0 - \frac{q d}{c + d\mathbf{v}^2}. \quad (4.53)$$

For $v > v_c = \frac{q}{\gamma_0} - \frac{c}{d}$ the motion of the particle is damped, while for smaller velocities it is accelerated. In this latter regime

$$\Delta^2(t) = 4D^{\text{eff}}t + 4D^{\text{eff}}\tau \left[\exp\left(-\frac{t}{\tau}\right) - 1 \right] \quad (4.54)$$

with

$$D^{\text{eff}} = v_0^4/2k_B T \gamma_0 \propto q^2, \quad \tau = v_0^2/2k_B T \gamma_0 \propto q. \quad (4.55)$$

4.4.6 Monte Carlo and lattice gas models of self-propelled particles

With the sake of capturing at the most basic level how self-propulsion affects the structure and dynamics of soft matter, one can extend the Monte Carlo algorithm used to simulate particle systems [33, 34, 35, 36]. Self-propulsion is then modelled by a non-Markovian stochastic drive, implemented by the following Monte-Carlo rule. At time t , a particle i is chosen at random and a small displacement $\boldsymbol{\delta}_i(t)$ of amplitude $|\boldsymbol{\delta}_i(t)| < \delta_0$ is proposed. Then, just as for equilibrium hard spheres [87], the particle position is updated according to

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \boldsymbol{\delta}_i(t) P_a, \quad (4.56)$$

where $P_a = \min\{1, e^{-\beta U}\}$. For hard spheres, $P_a = 1$ if the move does not generate any overlap with a neighbouring disk, and $P_a = 0$ otherwise. Self-propulsion is introduced via the rules to construct $\boldsymbol{\delta}_i(t)$ at successive time steps. We choose

$$\boldsymbol{\delta}_i(t + \Delta t) = \boldsymbol{\delta}_i(t) + \boldsymbol{\eta}_i(t),$$

where $\boldsymbol{\eta}_i(t)$ is a uniformly distributed random shift of amplitude $|\boldsymbol{\eta}_i(t)| < \delta_1$, chosen independently at each step and for each particle. As a consequence, successive displacements $\boldsymbol{\delta}_i$ decorrelate after a persistence time $\tau = (\delta_0/\delta_1)^2$. Such discrete dynamics generates persistent random walks for isolated particles. (See Ref. [34] for more details about the model.) This model is appealing because it has only two control parameters:

The persistence time τ controlling the self-propulsion of the particles, and the packing fraction ϕ quantifying crowding effects. The equilibrium hard disk model is restored in the limit $\tau \rightarrow 0$. As anticipated, this model also exhibits motility-induced aggregation and phase separation.

Lattice gases constitute another class of minimal models of self-propelled particles. These consider particles carrying an internal state, sitting on the sites of a lattice, and jumping from site to site with a given rate that depends on its internal state. For instance, to study the competition between excluded volume and self-propulsion, one can consider particles with an orientation (a spin) that randomly changes with some rate, and which dictates the preferred direction of the jump, generating a persistent random walk on the lattice. Then, excluded volume interactions can be introduced by imposing that two particles cannot sit on the same site simultaneously [42]. Alignment can then be easily included by adding interactions between the internal *spins* [88, 89]. The simplicity of these models allows to make analytical progress that would be out-of-reach in off-lattice particle systems [90, 91, 92].

4.5 Derivation of continuum descriptions

The dynamics of a system of self-propelled particles can also be described in terms of the joint probability distribution function $P_N(\Gamma \equiv \{\mathbf{r}_i, \mathbf{e}_i\}, i = 1 \dots N, t)$, which gives the probability to find N particles at positions \mathbf{r}_i with orientations \mathbf{e}_i for $i = 1, \dots, N$ at a given time t . For concreteness, we will consider 2D ABPs that self-propel at constant speed, v_0 , in the direction given by $\mathbf{e}_i = (\cos \theta_i, \sin \theta_i)$, and subject to thermal and rotational noise, characterized by the diffusion constants D_0 and D_r , respectively. From now on, we set $D_0 = 1$ without loss of generality.

The N -body probability distribution evolves according to the Smoluchowski equation

$$\frac{\partial P_n}{\partial t} = \sum_{i=1}^N \nabla_i \cdot [(\nabla_i \mathcal{U} P_N - v_0 \mathbf{e}_i P_N + D_0 \nabla_i P_N] + \frac{\partial}{\partial \theta_i} \left[\frac{\partial \mathcal{U}}{\partial \theta_i} P_N + D_r \frac{\partial P_N}{\partial \theta_i} \right]. \quad (4.57)$$

ABPs interact through the general pairwise interaction potential

$$\mathcal{U}(\{\mathbf{r}_i\}, \{\theta_i\}) = \sum_{i=1}^N \sum_{i < j} u(|\mathbf{r}_i - \mathbf{r}_j|, \theta_i, \theta_j). \quad (4.58)$$

For the sake of generality, we do not specify the functional form of \mathcal{U} . Consequently, we are deriving a general framework to describe systems of self-propelled particles whose interactions depend only on the center-to-center distance between pairs. Integrating out the degrees of freedom of $N - 1$ particles, one can obtain the 1-body Smoluchowski equation, $P_1(\mathbf{r}_1, \theta_1, t) \equiv \int d\mathbf{r}_2 \dots d\theta_1 \dots \theta_N P_N(\{\mathbf{r}_i, \theta_i\}, t)$,

$$\frac{\partial P_1}{\partial t} = -\nabla \cdot [\mathbf{F}(\mathbf{r}, \theta, t) - \nabla P_1] + D_r \frac{\partial^2 P_1}{\partial \theta^2}, \quad (4.59)$$

which constitutes the first equation of a BBGKY-like hierarchy of equations, coupled to two-body terms through $\mathbf{F}(\mathbf{r}, \theta, t)$, which is the effective force exerted by the surrounding particles into the tagged particle (labeled 1). The effective force reads

$$\mathbf{F}(\mathbf{r}_1, \theta_1, t) = -N \int d\mathbf{r}_2 \dots d\theta_N \int_0^{2\pi} d\theta_2 \dots d\theta_N P_N \nabla_1 \mathcal{U} = - \int d\mathbf{r}_2 \int_0^{2\pi} d\theta_2 P_2 \nabla_1 u(|\mathbf{r}_2 - \mathbf{r}_1|, \theta_1, \theta_2), \quad (4.60)$$

where $P_2(\mathbf{r}_1, \mathbf{r}_2, \theta_1, \theta_2, t)$ is the two-body probability distribution.

Forces come from the spatial dependence of the pair potential. Typically, one considers excluded-volume interactions, which set the particles characteristic finite size. To proceed, we introduce a change of variables. Since the interaction is pairwise, it can be fully defined by the vector distance $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1 = r_{12}(\cos \omega, \sin \omega)$ and the orientations θ_1 and θ_2 . Thus, in the lab frame of reference, the system can be described by the relevant set of variables $(r_{12}, \omega, \theta_1, \theta_2)$. Since the integrals in Eq. (4.60) are over \mathbf{r}_2 and θ_2 , while \mathbf{r}_1 and θ_1 are kept fixed, one can express the orientations as a function of θ_1 , and therefore introduce $\theta_{12} = \theta_2 - \theta_1$. Employing a body-fixed frame, one can express the directions along the plane relative to \mathbf{e}_1 , and hence use r_{12} , θ_1 and θ_{12} as the reference variables without loss of generality.

It is possible to decompose P_2 in terms of the new set of variables using the identity

$$P_2(\mathbf{r}_1, \mathbf{r}_2, \theta_1, \theta_2, t) = \bar{\rho} P_1(\mathbf{r}_1, \theta_1, t) \mathcal{G}(r_{12}, \theta_2, \theta_{12}, t), \quad (4.61)$$

where $\bar{\rho}$ is the average density and $\mathcal{G}(r_{12}, \theta_{12}, t)$ the pair correlation function that encodes the microscopic structure of the system through higher order correlation functions. It corresponds to the probability of finding a particle with orientation θ_2 in the plane-direction φ_1 , at a distance r_{12} from the tagged particle (located at \mathbf{r}_1 with orientation θ_1).

Using this set of variables, we can rewrite Eq. (4.60) as

$$\mathbf{F}(\mathbf{r}_1, \theta_1, t) = \bar{\rho} P_1(\mathbf{r}_1, \theta_1, t) \int d\mathbf{r}_{12} \int_0^{2\pi} d\theta_{12} \frac{\partial u(\mathbf{r}_{12}, \theta_{12})}{\partial r_{12}} \frac{\mathbf{r}_{12}}{r_{12}} \mathcal{G}(r_{12}, \theta_{12}, t). \quad (4.62)$$

In the remainder, we will drop the subscripts for clarity. In order to group the two-body terms in terms of scalar coefficients, we decompose $\mathbf{F}(\mathbf{r}_1, \theta_1, t)$ in the vector basis spanned by the direction of self-propulsion and the gradient of the probability density, $(\mathbf{e}, \nabla P_1)$. We follow a Gram-Schmidt orthonormalization scheme, which in this case is an approximation, due to the fact that it is not guaranteed that \mathbf{e} and ∇P_1 remain a vector basis, since they are time dependent and could become, at some point, parallel throughout the system evolution. The decomposition of $\mathbf{F}(\mathbf{r}_1, \theta_1, t)$ reads

$$\mathbf{F} \simeq -\bar{\rho} P_1 \zeta \mathbf{e} + (1 - \mathcal{D}) \nabla P_1, \quad (4.63)$$

where we have introduced the scalar coefficients

$$\zeta = - \int dr r \int_0^{2\pi} d\varphi \cos \varphi \int_0^{2\pi} d\theta \frac{\partial u(\mathbf{r})}{\partial r} \mathcal{G}(r, \varphi, \theta, t), \quad (4.64)$$

$$\mathcal{D} = 1 - \frac{|\nabla P_1 - \mathbf{e} \mathbf{e} \cdot \nabla P_1| \cdot \mathbf{F}}{|\nabla P_1|^2}. \quad (4.65)$$

The first term in the right-hand side of Eq. (4.63) is the component of the force along the self-propulsion direction. This component quantifies the imbalance between the self-propulsion of the tagged particle and its arrest induced by collisions with neighboring particles.

In a system of passive colloids, $\zeta = 0$, which can be understood because the distribution of particles around a tagged one is isotropic. This means that, within the approximations carried out, we do not capture the fact that the effective diffusion, \mathcal{D} , in general will depend on the interparticle interactions, \mathcal{U} , and hence, this theoretical description does not capture the possibility that a suspension of attractive diffusing particles in equilibrium can phase separate. We will show, however, that it is enough to capture the structural modifications induced by activity of ABPs (exploiting the preferred direction of motion \mathbf{e}) to decompose the effective force (as done in Eq. (4.63)), and describe activity-induced phase separation. In order to account for equilibrium phase separation we need to include the evolution of the pair correlation function \mathcal{G} approximately and take into account the changes in the microstructure due to particle interactions. This approach has been carried out in detail for colloidal suspensions, and the methods developed for those systems can be easily adapted for the active systems analyzed in this chapter. For example, one can use a mean-field approximation and split the two-body distribution function as a product of one-body ones. This will lead to an effective diffusivity at the level of the one-body Smoluchowski equation (see below) that will change sign when the homogenous state becomes unstable, signalling a spinodal long-wavelength instability.

As soon as activity enters the systems, the $\varphi \rightarrow \varphi + \pi$ symmetry is broken and thus $\zeta \neq 0$, due to the $\cos \varphi$ term stemming from the projection of the force. Due to activity, ζ will have additional contributions if the conservative potential has an aligning contribution and depends not only on the interparticle distance, but on the relative orientation of the self-propelling directions at which active particles move. We do not consider this general case here, but it can be derived using the same framework. This implies that an aligning potential modifies the force imbalance arising from the collision persistence and is captured by the magnitude of ζ .

The second term in the right-hand side of Eq. (4.63) can be interpreted as an effective diffusion acting along the gradient of the one-body probability distribution. Introducing the expressions for the force in the one-body Smoluchowski equation leads to

$$\frac{\partial P_1}{\partial t} = -\nabla \cdot [\nu_{\bar{\rho}} \mathbf{e} P_1 - \mathcal{D} \nabla P_1] + D_r \frac{\partial^2 P_1}{\partial \theta^2}, \quad (4.66)$$

where

$$v_{\bar{\rho}} = v_0 - \bar{\rho} \zeta \quad (4.67)$$

accounts for the fact that the interactions between particles affect the magnitude at which they self-propel.

The two terms on the right-hand side of Eq. (4.66) correspond to advection and diffusion of the active particles, respectively. Advection sets an effective self-propulsion speed, $v_{\bar{\rho}}$, which decays with the mean density $\bar{\rho}$ at a rate proportional to ζ , and which can thus be interpreted as a translational friction coefficient, accounting for the arrest of particles in crowded environments. \mathcal{D} can be interpreted as an effective diffusivity.

The microscopic information of the system is captured by ζ and \mathcal{D} . They depend functionally on the two-body correlation function, \mathcal{G} . Therefore, Eq. (4.66) is not closed, but is the first of a hierarchy of equations for the active particles' probability distribution. We will assume that these two coefficients are constant, and will disregard both their dependence on the tagged particle's position, which is valid as long as the system is in (close to) a homogeneous state, and on the state of the system and relative particle distribution (quantified by the system microscopic correlations).

It is worth pointing out that, opposed to top-down approaches that base the derivation of the effective hydrodynamic equations on symmetry arguments and conservation laws [93], the current derivation directly coarse-grains the ABP microscopic dynamics. Thus, the coefficients obtained are not phenomenological but stem from interparticle interactions. They indeed take specific numerical values in particle-based models, whose calculation allows for a direct mapping between the microscopic model and the coarse-grained theory. The decay of the effective self-propelling speed, v_{ρ} , at increasing density (a signature of MIPS), emerges as a consequence of the dynamics of the system, and it is not introduced as an ad-hoc hypothesis [94].

We close the hierarchy of coupled equations by considering these parameters as constants, and hence Eq. (4.66) becomes the one-body Smoluchowski equation for ABPs.

The hydrodynamic equations for ABPs can be obtained from the one-body Smoluchowski equation, because the relevant hydrodynamic fields can be expressed as moments of the one-body probability distributions. Specifically, the first two moments

$$\begin{aligned} \rho(\mathbf{r}, t) &\equiv \int_0^{2\pi} d\varphi \psi_1(\mathbf{r}, \varphi, t) \\ \mathbf{p}(\mathbf{r}, t) &\equiv \int_0^{2\pi} d\varphi \mathbf{e} \psi_1(\mathbf{r}, \varphi, t) \end{aligned} \quad (4.68)$$

correspond to the ABP density and polarization, respectively. Integrating Eq. (4.66) with constant coefficients, we arrive at

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = -\nabla \cdot [v_{\bar{\rho}} \mathbf{p} - \mathcal{D} \nabla \rho], \quad (4.69)$$

$$\frac{\partial \mathbf{p}(\mathbf{r}, t)}{\partial t} = -\nabla \cdot \left[v_{\bar{\rho}} \left(\frac{\rho}{2} \mathbf{1} + \mathbf{Q} \right) - \mathcal{D} \nabla \mathbf{p} \right] - D_r \mathbf{p}, \quad (4.70)$$

where $\mathbf{1}$ stands for the unit matrix and \mathbf{Q} is the nematic tensor field, which is expressed as a higher-order moment of the one-body probability distribution function,

$$\mathbf{Q}(\mathbf{r}, t) \equiv \int_0^{2\pi} d\varphi \left(\mathbf{e} \mathbf{e} - \frac{1}{2} p^2 \right) \psi_1(\mathbf{r}, \varphi, t). \quad (4.71)$$

Therefore, the time evolution equation of each moment is linearly coupled to the next order moment, leading to a hierarchy of evolution equations for the moments of ψ_1 . To close the hierarchy of hydrodynamic equations, different strategies can be followed. A standard approach is to express the nematic tensor in terms of the polarization, e.g., $\mathbf{Q} \simeq \mathbf{p} \mathbf{p} - \frac{p^2}{2} \mathbf{1}$. We will neglect the dependence of \mathbf{p} on \mathbf{Q} . In fact, if the evolution equation of \mathbf{Q} is derived and the dependence on higher order hydrodynamic fields is disregarded, it is possible to show that \mathbf{Q} does not affect qualitatively the dynamics of ρ and \mathbf{p} to linear order [95].

The hydrodynamic equations derived above have an isotropic homogenous steady solution $\rho(\mathbf{r}, t) = \bar{\rho}$ and $\mathbf{p}(\mathbf{r}, t) = 0$, but do not admit a polar steady solution, such as that obtained in the continuum theories of flocking. The present approach does not allow for symmetry breaking in the polarization field because it requires an interaction term that either tends to align interacting particles, e.g., through a torque or, phenomenologically, through a symmetry breaking term a la Landau, as in the Toner-Tu theory [96]. Therefore, the hydrodynamic equations for ABPs are limited to the description of non-polar states.

4.5.1 Linear stability analysis

We can carry out a linear stability analysis of Eqs. (4.68) to identify the stability of the homogeneous, steady state, and the physical mechanisms that can lead to its instability into a phase with different symmetries. To this end, we follow the dynamic evolution of a perturbation of the density and polarization fields with respect to their values in the homogeneous state, namely, $\rho(\mathbf{r}, t) = \bar{\rho} + \delta\rho$ and $\mathbf{p}(\mathbf{r}, t) = \delta\mathbf{p}$ with $\delta\rho/\bar{\rho} \ll 1$ and $\delta\mathbf{p} \ll 1$. Accordingly, Eqs. (4.68) read

$$\frac{\partial \delta\rho}{\partial t} = -\nabla \cdot [(\nu_0 - \bar{\rho}\zeta)\delta\mathbf{p} - \mathcal{D}\nabla\delta\rho], \quad (4.72)$$

$$\frac{\partial \delta\mathbf{p}}{\partial t} = -\nabla \cdot \left[\frac{1}{2}(\nu_0 - 2\bar{\rho}\zeta)\delta\rho - \mathcal{D}\nabla\delta\mathbf{p} \right] - D_r\delta\mathbf{p}. \quad (4.73)$$

Since the equations are linear in the relevant hydrodynamic fields, it is convenient to write the hydrodynamic equations for the perturbation in Fourier space, $\mathbf{u} \sim \tilde{\mathbf{u}}e^{i\mathbf{k}\cdot\mathbf{r}}$ where $\mathbf{u} = (\delta\rho, \delta\mathbf{p})$, which leads to

$$\frac{\partial \delta\hat{\rho}}{\partial \tilde{t}} = -i\mathbf{q} \cdot \left[4\left(\frac{\nu_0}{\nu^*} - \zeta\right)\delta\hat{\mathbf{p}} - i\mathbf{q} \cdot \delta\hat{\rho} \right], \quad (4.74)$$

$$\frac{\partial \delta\hat{\mathbf{p}}}{\partial \tilde{t}} = -i\mathbf{q} \left[2\left(\frac{\nu}{\nu^*} - 2\zeta\right)\delta\hat{\rho} - i\mathbf{q} \cdot \delta\hat{\mathbf{p}} \right] - \delta\hat{\mathbf{p}}, \quad (4.75)$$

where we have introduced the dimensionless quantities

$$\tilde{t} = D_r t, \quad \tilde{\mathbf{q}} = \sqrt{\frac{\mathcal{D}}{D_r}}\mathbf{q}, \quad \frac{\nu_0}{\nu^*} = \frac{\nu_0}{4\sqrt{\mathcal{D}D_r}}. \quad (4.76)$$

Although we will work with dimensionless quantities, we drop the tilde (hence, e.g., $\tilde{\mathbf{q}} \equiv \mathbf{q}$). We write the previous system of equation in matrix form, $\partial_t(\delta\hat{\rho}, \delta\hat{\mathbf{p}}) = M(\delta\hat{\rho}, \delta\hat{\mathbf{p}})^T$, with

$$M = \begin{bmatrix} -q^2 & -4i\left(\frac{\nu_0}{\nu^*} - \zeta\right)q_x & -4i\left(\frac{\nu_0}{\nu^*} - \zeta\right)q_y \\ -2i\left(\frac{\nu_0}{\nu^*} - 2\zeta\right)q_x & -1 - q^2 & 0 \\ -2i\left(\frac{\nu_0}{\nu^*} - 2\zeta\right)q_y & 0 & -1 - q^2 \end{bmatrix}. \quad (4.77)$$

The eigenvalues of M correspond to the dispersion relations quantifying the growth of a perturbation with dimensionless wave vector \mathbf{q} , which allow us to explore the onset of linear instabilities. The parameter space of the model is conformed by ν_0/ν^* and ζ , the competition between activity and interparticle collisions. The eigenvalues of M can be expressed as

$$\lambda_1 = -\frac{1}{2}(2q^2 + 1) + \sqrt{1 - 32q^2\left(\frac{\nu_0}{\nu^*} - \zeta\right)\left(\frac{\nu_0}{\nu^*} - 2\zeta\right)}, \quad (4.78)$$

$$\lambda_2 = -1 - q^2, \quad (4.79)$$

$$\lambda_3 = -\frac{1}{2}(2q^2 + 1) - \sqrt{1 - 32q^2\left(\frac{\nu_0}{\nu^*} - \zeta\right)\left(\frac{\nu_0}{\nu^*} - 2\zeta\right)}, \quad (4.80)$$

At low distances, expanding the eigenvalues for low \mathbf{q} magnitudes, we get

$$\lambda_1 = -\left[1 + 8\left(\frac{\nu_0}{\nu^*} - \zeta\right)\left(\frac{\nu_0}{\nu^*} - 2\zeta\right)\right]q^2 + \mathcal{O}(q^3), \quad (4.81)$$

$$\lambda_2 = -1 - q^2, \quad (4.82)$$

$$\lambda_3 = -1 + \left[-1 + 8\left(\frac{\nu_0}{\nu^*} - \zeta\right)\left(\frac{\nu_0}{\nu^*} - 2\zeta\right)\right]q^2 + \mathcal{O}(q^3). \quad (4.83)$$

The eigenvalues λ_2 and λ_3 are negative for any wave vector \mathbf{q} , indicating that the homogeneous state is stable upon a perturbation along these two modes, as depicted in Fig. 4.6a. On the contrary, λ_1 can become positive at sufficiently small wave vector and triggers the growth of a long-wavelength instability. The instability region at $q \rightarrow 0$ as a function of $(\nu_0/\nu^*, \zeta)$ is shown in Fig. 4.6b in blue. One can obtain the limits of

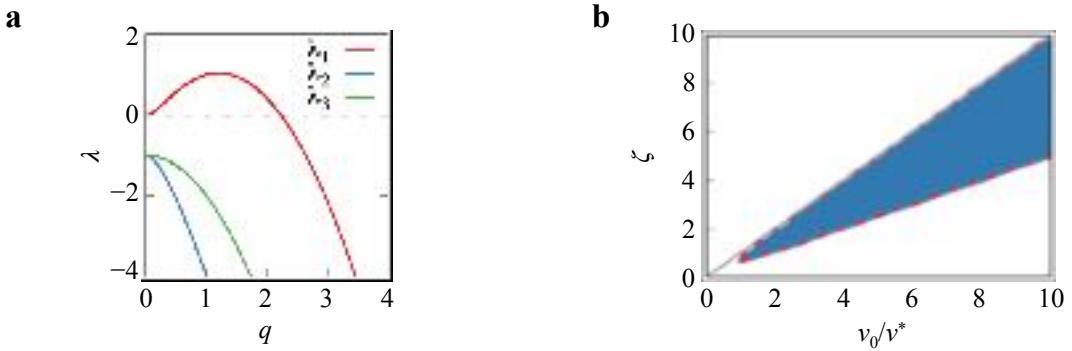


Figure 4.6: Linear stability of the homogeneous phase. (a) Dispersion relations as a function of the dimensionless wave vector $q = |\mathbf{q}|$ for $v_0/v^* = 2.5$, and $\zeta = 2$: $\lambda_2 < 0$ and $\lambda_3 < 0$, for all q , while $\lambda_1 > 0$ in a range of small wave vectors. (b) Long wavelength instability region in the (v_0, ζ) plane. The limit of stability is illustrated by the red dashed line, and corresponds to Eq. (4.84).

stability analytically by taking the second order term in the Taylor expansion in Eq. (4.83), and setting it to zero, leading to

$$\zeta\left(\frac{v_0}{v^*}, q \rightarrow 0\right) = \zeta = \frac{3}{4} \frac{v_0}{v^*} \pm \frac{1}{4} \sqrt{\left(\frac{v_0}{v^*}\right)^2 - 1} \quad (4.84)$$

and represented by two broken lines in Fig. 4.6b. The dispersion relations do not have any complex term, implying that no oscillating instabilities take place. This long wavelength instability is associated to MIPS, and comes from an increase of the effective friction ζ along the direction of self-propulsion as the self-propulsion speed increases.

Problems

Problem 4.1: Short and long time regimes. The mean squared displacement for a Brownian particles is given by Eq. (4.13).

- a. Show that, for $t \ll t_d = m/\gamma$, the motion is ballistic. [Hint: Perform a Taylor expansion to first order in t .]
- b. Take the limit $t \gg t_d$ and show that the motion is diffusive, as described by Eq. (4.14).

Problem 4.2: Overdamped regime. A spherical Brownian particle of mass m and radius R is suspended in a viscous fluid of viscosity η at temperature T . Consider the limit $t_d \approx 0$ (overdamped regime) in one dimension.

- a. Show that the equation of motion is

$$\frac{dx}{dt} = u(x) + g(t),$$

where $u(x)$ and $g(t)$ are functions of x and t , respectively.

- b. Show that $u(x) = F(x)/\gamma$, $\langle g(t) \rangle = 0$, and $\langle g(t)g(t') \rangle = 2k_B T \delta(t-t')/\gamma$.
- c. Assume that $u(x) = u_0$ and derive the mean squared displacement and obtain the dependence of the diffusion coefficient on the temperature T .
- d. Obtain the Fokker–Planck equation.

Problem 4.3: Master equation. Complete the derivation of Eq. (4.18) from the master equation Eq. (4.16).

Problem 4.4: Solution of the diffusion equation. Consider the solution of the diffusion equation given by Eq. (4.25).

- a. Re-write Eq. (4.25) as a product of d Gaussian distributions, one for each coordinate r_i .
- b. Obtain the first and second moment of the distribution and show that, for $r_{i,0} = 0$, is

$$\langle r_i^2 \rangle - \langle r_i \rangle^2 = 2Dt . \quad (4.85)$$

- c. Derive the mean squared displacement in d dimensions.

Problem 4.5: (Overdamped) active Brownian particle.

- a. Compute the auto-correlation of an ABP self-propulsion direction as given in Eq. (4.28). Use this result to compute the mean-squared displacement of a single ABP in Eq. (4.26).
- b. Show that at time shorter than the persistence time τ , $t \ll \tau$, the motion is ballistic.
- c. Show that in the long time regime $t \gg \tau$ and show that the motion is diffusive. Provide the corresponding diffusion coefficient.
- d. Compute the response function associated to a constant force perturbation. Show that the mobility is independent of activity. Does the Stokes-Einstein relation still hold?
- e. The detailed balance condition for an ABP reads:

$$v_0 \mathbf{e} \Psi(\Gamma) = \gamma_0^{-1} k_B T \nabla \Psi(\Gamma) \text{ and } \partial_\theta \Psi(\Gamma) = 0 .$$

Show that this condition can also be met if $v_0 = 0$.

Problem 4.6: Active Ornstein Uhlenbeck Particles.

- a. Map an AOUP into an ABP in the asymptotic long-time regime.
- b. Derive under which conditions an AOUP fulfills detailed balance.

Bibliography

- [1] S. Makarchuck, V. C. Braz, N. A. M. Araujo, L. Ciric, and G. Volpe. Enhanced propagation of motile bacteria on surfaces due to forward scattering. *Nature Communications*, 10:4110, 2019.
- [2] P. L. Krapivsky, S. Redner, and E. Ben-Naim. *A Kinetic View of Statistical Physics*. Cambridge University Press, Cambridge, United Kingdom, 2010.
- [3] M. Doi. *Soft Matter Physics*. Oxford University Press, Oxford, United Kingdom, 2013.
- [4] R. Livi and P. Politi. *Nonequilibrium Statistical Physics: a Modern Perspective*. Cambridge University Press, Cambridge, United Kingdom, 2017.
- [5] M. C. Marchetti, J.-F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85:1143, 2013.
- [6] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [7] É. Fodor and M. C. Marchetti. The statistical physics of active matter: From self-catalytic colloids to living cells. *Physica A: Statistical Mechanics and its Applications*, 504:106–120, 2018.
- [8] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger. Computational models for active matter. *Nature Reviews Physics*, 2:181–199, 2020.
- [9] E. Lauga. *The Fluid Dynamics of Cell Motility*. Cambridge University Press, Cambridge, United Kingdom, 2020.
- [10] P. Langevin. Sur la théorie du mouvement brownien. *Comptes Rendus de l'Académie des Sciences - Series I*, 146:530–533, 1908.
- [11] D. S. Lemons and A. Gythiel. Paul Langevin's 1908 paper “on the theory of Brownian motion”. *American Journal of Physics*, 65:1079–1081, 1997.

- [12] A. D. Fokker. Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld. *Annalen der Physik*, 348:810–820, 1914.
- [13] M. Planck. Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie. *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*, XXIV:324–341, 1917.
- [14] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier. Active Brownian particles. *The European Physical Journal Special Topics*, 202:1–162, 2012.
- [15] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. Self-motile colloidal particles: from directed propulsion to random walk. *Physical Review Letters*, 99:048102, 2007.
- [16] J. Palacci, C. Cottin-Bizonne, C. Ybert, and L. Bocquet. Sedimentation and effective temperature of active colloidal suspensions. *Physical Review Letters*, 105:088304, 2010.
- [17] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck. Dynamical clustering and phase separation in suspensions of self-propelled colloidal particles. *Physical Review Letters*, 110:238301, 2013.
- [18] D. Klotsa. As above, so below, and also in between: mesoscale active matter in fluids. *Soft matter*, 15:8946–8950, 2019.
- [19] A. Wysocki, R. G. Winkler, and G. Gompper. Cooperative motion of active Brownian spheres in three-dimensional dense suspensions. *Europhysics Letters*, 105:48004, 2014.
- [20] J. Stenhammar, D. Marenduzzo, R. J. Allen, and M. E. Cates. Phase behaviour of active Brownian particles: the role of dimensionality. *Soft matter*, 10:1489–1499, 2014.
- [21] R. G. Winkler, A. Wysocki, and G. Gompper. Virial pressure in systems of spherical active Brownian particles. *Soft matter*, 11:6680–6691, 2015.
- [22] F. Turci and N. B. Wilding. Phase separation and multibody effects in three-dimensional active Brownian particles. *Physical Review Letters*, 126:038002, 2021.
- [23] A. K. Omar, K. Klymko, T. GrandPre, and P. L. Geissler. Phase diagram of active Brownian spheres: Crystallization and the metastability of motility-induced phase separation. *Physical Review Letters*, 126:188002, 2021.
- [24] J. Elgeti, R. G. Winkler, and G. Gompper. Physics of microswimmers – single particle motion and collective behavior: a review. *Reports on progress in physics*, 78:056601, 2015.
- [25] I. Petrelli, L. F. Cugliandolo, G. Gonnella, and A. Suma. Effective temperatures in inhomogeneous passive and active bidimensional Brownian particle systems. *Physical Review E*, 102:012609, 2020.
- [26] S. Dal Cengio, D. Levis, and I. Pagonabarraga. Fluctuation–dissipation relations in the absence of detailed balance: formalism and applications to active matter. *Journal of Statistical Mechanics: Theory and Experiment*, 2021:043201, 2021.
- [27] S. C. Takatori, W. Yan, and J. F. Brady. Swim pressure: stress generation in active matter. *Physical Review Letters*, 113:028103, 2014.
- [28] A. P. Solon, Y. Fily, A. Baskaran, M. E. Cates, Y. Kafri, M. Kardar, and J. Tailleur. Pressure is not a state function for generic active fluids. *Nature Physics*, 11:673–678, 2015.
- [29] T. Speck and R. L. Jack. Ideal bulk pressure of active Brownian particles. *Physical Review E*, 93:062605, 2016.
- [30] P. Digregorio, D. Levis, A. Suma, L. F. Cugliandolo, G. Gonnella, and I. Pagonabarraga. Full Phase Diagram of Active Brownian Disks: From Melting to Motility-Induced Phase Separation. *Physical Review Letters*, 121:98003, 2018.

- [31] C. B. Caporusso, P. Digregorio, D. Levis, L. F. Cugliandolo, and G. Gonnella. Motility-induced microphase and macrophase separation in a two-dimensional active Brownian particle system. *Physical Review Letters*, 125:178004, 2020.
- [32] D. Levis, J. Codina, and I. Pagonabarraga. Active Brownian equation of state: metastability and phase coexistence. *Soft Matter*, 13:8113, 2017.
- [33] L. Berthier. Nonequilibrium glassy dynamics of self-propelled hard disks. *Physical Review Letters*, 112:220602, 2014.
- [34] D. Levis and L. Berthier. Clustering and heterogeneous dynamics in a kinetic Monte Carlo model of self-propelled hard disks. *Physical Review E*, 89:062301, 2014.
- [35] J. U. Klamser, S. C. Kapfer, and W. Krauth. Thermodynamic phases in two-dimensional active matter. *Nature communications*, 9:1–8, 2018.
- [36] J. U. Klamser, S. C. Kapfer, and W. Krauth. A kinetic-Monte Carlo perspective on active matter. *The Journal of Chemical Physics*, 150:144113, 2019.
- [37] E. P. Bernard and W. Krauth. Two-step melting in two dimensions: first-order liquid-hexatic transition. *Physical Review Letters*, 107:155704, 2011.
- [38] L. F. Cugliandolo and G. Gonnella. Phases of active matter in two dimensions. *arXiv preprint arXiv:1810.11833*, 2018.
- [39] J. Tailleur and M. E. Cates. Statistical mechanics of interacting run-and-tumble bacteria. *Physical Review Letters*, 100:218103, 2008.
- [40] Y. Fily and M. C. Marchetti. Athermal phase separation of self-propelled particles with no alignment. *Physical Review Letters*, 108:235702, 2012.
- [41] G. S. Redner, M. F. Hagan, and A. Baskaran. Structure and dynamics of a phase-separating active colloidal fluid. *Physical Review Letters*, 110:055701, 2013.
- [42] S. Whitelam, K. Klymko, and D. Mandal. Phase separation and large deviations of lattice active matter. *The Journal of Chemical Physics*, 148:154902, 2018.
- [43] B. Partridge and C. F. Lee. Critical motility-induced phase separation belongs to the ising universality class. *Physical Review Letters*, 123:068002, 2019.
- [44] C. Maggi, M. Paoluzzi, A. Crisanti, E. Zaccarelli, and N. Gnan. Universality class of the motility-induced critical point in large scale off-lattice simulations of active particles. *Soft Matter*, 17:3807–3812, 2021.
- [45] P. Digregorio, D. Levis, L. F. Cugliandolo, G. Gonnella, and I. Pagonabarraga. Unified analysis of topological defects in 2d systems of active and passive disks. *Soft Matter*, 18:566–591, 2022.
- [46] C. B. Caporusso, L. F. Cugliandolo, P. Digregorio, G. Gonnella, D. Levis, and A. Suma. Dynamics of motility-induced clusters: coarsening beyond Ostwald ripening. *Physical Review Letters (in press)* *arXiv:2211.12361*, 2022.
- [47] T. Speck. Critical behavior of active Brownian particles: Connection to field theories. *Physical Review E*, 105:064601, 2022.
- [48] H. Chaté. Dry aligning dilute active matter. *Annual Review of Condensed Matter Physics*, 11:189–212, 2020.
- [49] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75:1226, 1995.
- [50] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517:71–140, 2012.

- [51] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107:11865–11870, 2010.
- [52] F. Peruani, A. Deutsch, and M. Bär. A mean-field theory for self-propelled particles interacting by velocity alignment mechanisms. *The European Physical Journal Special Topics*, 157:111–122, 2008.
- [53] A. Martín-Gómez, D. Levis, A. Díaz-Guilera, and I. Pagonabarraga. Collective motion of active Brownian particles with polar alignment. *Soft Matter*, 14:2610–2618, 2018.
- [54] E. Sesé-Sansa, D. Levis, and I. Pagonabarraga. Phase separation of self-propelled disks with ferromagnetic and nematic alignment. *Physical Review E*, 104:054611, 2021.
- [55] O. Chepizhko, D. Saintillan, and F. Peruani. Revisiting the emergence of order in active matter. *Soft Matter*, 17:3113–3120, 2021.
- [56] M. Bär, R. Großmann, S. Heidenreich, and F. Peruani. Self-propelled rods: Insights and perspectives for active matter. *Annual Review of Condensed Matter Physics*, 11:441–466, 2020.
- [57] P.-G. De Gennes and J. Prost. *The physics of liquid crystals*, volume 83. Oxford university press, 1993.
- [58] F. C. Poderoso, J. J. Arenzon, and Y. Levin. New ordered phases in a class of generalized X Y models. *Physical Review Letters*, 106:067202, 2011.
- [59] H. Chaté. Dry aligning dilute active matter. *Annual Review of Condensed Matter Physics*, 11:189–212, 2020.
- [60] F. Peruani, T. Klauss, A. Deutsch, and A. Voss-Boehme. Traffic jams, gliders, and bands in the quest for collective motion of self-propelled particles. *Physical Review Letters*, 106:128101, 2011.
- [61] F. D. C. Farrell, M. C. Marchetti, D. Marenduzzo, and J. Tailleur. Pattern formation in self-propelled particles with density-dependent motility. *Physical Review Letters*, 108:248101, 2012.
- [62] M. Abkenar, K. Marx, T. Auth, and G. Gompper. Collective behavior of penetrable self-propelled rods in two dimensions. *Physical Review E*, 88:062314, 2013.
- [63] X. Q. Shi and H. Chate. Self-propelled rods: Linking alignment-dominated and repulsion-dominated active matter. *arXiv:1807.00294v2*, pages 2–6, 2018.
- [64] A. Jayaram, A. Fischer, and T. Speck. From scalar to polar active matter: Connecting simulations with mean-field theory. *Physical Review E*, 101:22602, 2020.
- [65] R. Großmann, Igor S. Aranson, and F. Peruani. A particle-field approach bridges phase separation and collective motion in active matter. *Nature Communications*, 11:5365, 2020.
- [66] B. Bhattacherjee and D. Chaudhuri. Re-entrant phase separation in nematicallly aligning active polar particles. *Soft Matter*, 15:8483, 2019.
- [67] S. Mandal, B. Liebchen, and H. Löwen. Motility-induced temperature difference in coexisting phases. *Physical Review Letters*, 123:228001, 2019.
- [68] H. Löwen. Inertial effects of self-propelled particles: From active Brownian to active Langevin motion. *The Journal of Chemical Physics*, 152:040901, 2020.
- [69] L. Caprini, R. K. Gupta, and H. Löwen. Role of rotational inertia for collective phenomena in active matter. *Physical Chemistry Chemical Physics*, 24:24910–24916, 2022.
- [70] A. R. Sprenger, L. Caprini, H. Löwen, and R. Wittmann. Dynamics of active particles with translational and rotational inertia. *Journal of Physics: Condensed Matter*, 35:305101, 2023.
- [71] H. C. Berg. *E. coli in Motion*. Springer, 2004.

- [72] M. J. Schnitzer. Theory of continuum random walks and application to chemotaxis. *Physical Review E*, 48:2553–2568, 1993.
- [73] M. E. Cates and J. Tailleur. When are active Brownian particles and run-and-tumble particles equivalent? consequences for motility-induced phase separation. *Europhysics Letters*, 101:20010, 2013.
- [74] J. Tailleur and M. E. Cates. Statistical mechanics of interacting run-and-tumble bacteria. *Physical Review Letters*, 100:218103, 2008.
- [75] G. Szamel. Self-propelled particle in an external potential: Existence of an effective temperature. *Physical Review E*, 90:012111, 2014.
- [76] É. Fodor, C. Nardini, M. E. Cates, J. Tailleur, P. Visco, and F. Van Wijland. How far from equilibrium is active matter? *Physical Review Letters*, 117:038103, 2016.
- [77] L. L. Bonilla. Active Ornstein-Uhlenbeck particles. *Physical Review E*, 100:022601, 2019.
- [78] T. F. F. Farage, P. Krininger, and J. M. Brader. Effective interactions in active Brownian suspensions. *Physical Review E*, 91:042310, 2015.
- [79] L. Caprini and U. Marini Bettolo Marconi. Inertial self-propelled particles. *The Journal of Chemical Physics*, 154:024902, 2021.
- [80] D. Chaudhuri. Active Brownian particles: Entropy production and fluctuation response. *Physical Review E*, 90:022131, 2014.
- [81] U. Seifert. From stochastic thermodynamics to thermodynamic inference. *Annual Review of Condensed Matter Physics*, 10:171–192, 2019.
- [82] G. Szamel. Stochastic thermodynamics for self-propelled particles. *Physical Review E*, 100:050603, 2019.
- [83] S. Shankar and M. C. Marchetti. Hidden entropy production and work fluctuations in an ideal active gas. *Physical Review E*, 98:020604, 2018.
- [84] L. Caprini, U. Marini Bettolo Marconi, A. Puglisi, and A. Vulpiani. The entropy production of ornstein–uhlenbeck active particles: a path integral method for correlations. *Journal of Statistical Mechanics: Theory and Experiment*, 2019:053203, 2019.
- [85] F. Schweitzer, W. Ebeling, and B. Tilch. Complex motion of Brownian particles with energy depots. *Physical Review Letters*, 80:5044, 1998.
- [86] F. Schweitzer and J. D. Farmer. *Brownian agents and active particles: collective dynamics in the natural and social sciences*, volume 1. Springer, 2003.
- [87] W. Krauth. *Statistical mechanics: algorithms and computations*, volume 13. OUP Oxford, 2006.
- [88] F. Peruani, T. Klauss, A. Deutsch, and A. Voss-Boehme. Traffic jams, gliders, and bands in the quest for collective motion of self-propelled particles. *Physical Review Letters*, 106:128101, 2011.
- [89] A. P. Solon and J. Tailleur. Revisiting the flocking transition using active spins. *Physical Review Letters*, 111:078101, 2013.
- [90] M. Kourbane-Houssene, C. Erignoux, T. Bodineau, and J. Tailleur. Exact hydrodynamic description of active lattice gases. *Physical Review Letters*, 120:268003, 2018.
- [91] T. Agranov, S. Ro, Y. Kafri, and V. Lecomte. Exact fluctuating hydrodynamics of active lattice gases – typical fluctuations. *Journal of Statistical Mechanics: Theory and Experiment*, 2021:083208, 2021.
- [92] C. Erignoux. On the hydrodynamics of active matter models on a lattice. *arXiv preprint arXiv:2108.04003*, 2021.

- [93] R. Wittkowski, A. Tiribocchi, J. Stenhammar, R. J. Allen, D. Marenduzzo, and M. E. Cates. Scalar ϕ field theory for active-particle phase separation. *Nature Communications*, 5:4351, 2014.
- [94] J. Bialké, H. Löwen, and T. Speck. Microscopic theory for the phase separation of self-propelled repulsive disks. *Europhysics Letters*, 103:30008, 2013.
- [95] E. Sesé-Sansa, D. Levis, and I. Pagonabarraga. Microscopic field theory for structure formation in systems of self-propelled particles with generic torques. *The Journal of Chemical Physics*, 157:224905, 2022.
- [96] J. Toner, Y. Tu, and S. Ramaswamy. Hydrodynamics and phases of flocks. *Annalen der Physik*, 318:170–244, 2005.

Chapter 5

Stochastic Differential Equations and Active Matter

JAN WEHR

Virtually every physical system interacts with its environment. This interaction can sometimes be neglected, in which case one calls the system *isolated*. When it has to be taken into account, it is often impossible, or at least impractical, to do it in full detail, including all energy exchanges occurring between the environment and the system. Perhaps surprisingly, as discussed in Chapter 5 [Should it be ‘this chapter’ instead? Or is a reference to a previous chapter?], modeling them as random events, leads to a theory agreeing very well with experiment. This is the theory of *Langevin equations* or, more generally, *stochastic differential equations* (SDE). This Chapter introduces SDEs, together with some applications to physics, including active matter, e.g., the motion of light-activated robots with sensorial delay (Fig. 5). The style is mathematical, but, in the interest

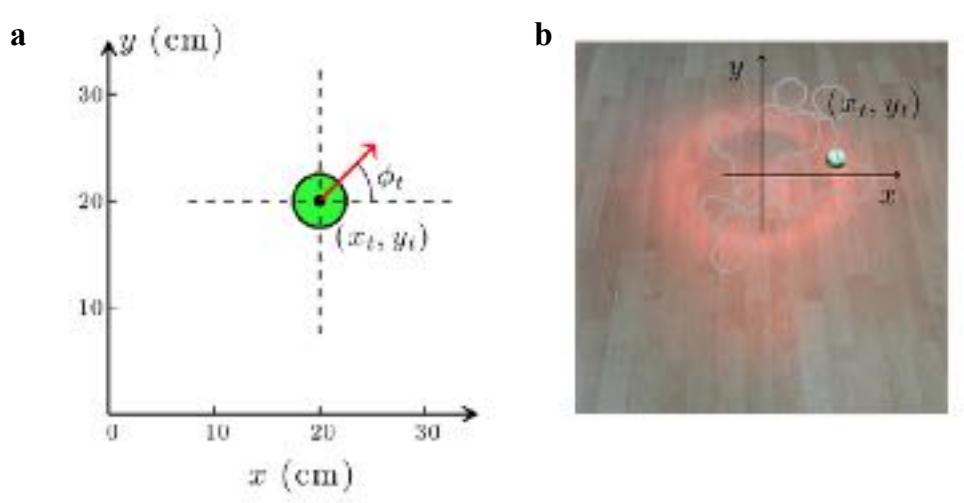


Figure 5.1: **Motion of light-sensitive robots in an inhomogeneous environment.** (a) An autonomous agent is described by its position at time t , (x_t, y_t) , its instantaneous speed v , and its direction of motion (arrow with angle ϕ_t with the reference x -axis), which varies randomly with a characteristic time τ . (b) Experimental realization of an autonomous agent sensitive to light: picture of a phototactic robot in a light gradient generated by an infrared lamp. The robot is programmed such that its propulsion speed depends on the light intensity it instantaneously measures in its surroundings, and its orientation changes randomly. A sample trajectory is shown by the gray solid line. Reproduced from Ref. [1].

of brevity, some technical details are left out. The reader is encouraged to consult the references listed at the end, starting from Ref. [2].

5.1 Stochastic differential equations

A stochastic differential equation (SDE) has the form

$$\frac{dx_t}{dt} = b(x_t) + \sigma(x_t)\eta_t, \quad (5.1)$$

where x_t denotes the state of the system at time t , $b(x_t)$ is the *drift*, η_t is a unit *Gaussian white noise*, and $\sigma(x_t)$ is the noise intensity. In general, both the intensity and the drift b depend on the state of the system at time t . When σ vanishes, Eq. (5.1) becomes an ordinary differential equation (ODE). When $\sigma \neq 0$, the ODE is perturbed by a random term. We think of η_t as a family of random variables, indexed by t , which are independent. This is not quite precise — for the realizations of such process, i.e., functions $t \mapsto \eta_t$ to be actual functions (rather than distributions), η_t would have to be correlated. Instead of introducing such correlations, we will express everything in terms of the integrated process

$$W_t = \int_0^t \eta_s ds. \quad (5.2)$$

As η_t can be thought of as steps of a random walk, W_t is the continuous-time analog of the position of the walk at time t . The process W_t , known as a Wiener process, can be defined as a random process with the properties:

- $W_0 = 0$.
- The increment $W_t - W_s$ has a Gaussian distribution with mean zero and variance $t - s$.
- The increments corresponding to disjoint time intervals are independent.

Equivalently, W_t , $t \geq 0$ is a Gaussian process with zero mean and covariance

$$\mathbf{E}[W_s W_t] = \min(s, t). \quad (5.3)$$

Example 5.1: Covariance of the Wiener process. Prove the above formula, Eq. (5.3)

[Hint: If $s < t$, the expected values of W_s^2 , W_t^2 , and $(W_t - W_s)^2$ are equal to s , t , and $t - s$ respectively.]

5.1.1 Itô integral

In order to endow equation (5.1) with a precise meaning, we have to define what it means that for every $t \geq 0$

$$x_t = x_0 + \int_0^t b(x_s) ds + \int_0^t \sigma(x_s) dW_s. \quad (5.4)$$

In particular, we need to define the last term. More generally, we will now define the *stochastic integral*

$$\int_0^t h_s dW_s \quad (5.5)$$

for a class of stochastic processes h_t . We will indicate the main steps of the construction and study it using simple examples, referring the reader to Ref. [2] for all details. The most important feature that distinguishes the stochastic integral from the standard one is the magnitude of the fluctuations of the Wiener process. The next fact makes it precise.

Lemma: Fix $T > 0$ and divide the interval $[0, T]$ into n equal parts, introducing $t_j = \frac{j}{n}T$. Strictly speaking, the value of t_j depends on n as well, but we suppress this dependence in the notation. Then the sum

$$\sum \left(W_{t_{j+1}} - W_{t_j} \right)^2 \quad (5.6)$$

converges in mean square to T , i.e.,

$$E \left[\left(\sum \Delta_j^2 - T \right)^2 \right] \rightarrow 0. \quad (5.7)$$

[Define Δ]

Example 5.2: Quadratic variation of the Wiener process. Prove the above statement.

[Hint: In calculations, use the fact that if X is a Gaussian random variable with mean zero and variance $E[X^2] = a^2$, then its fourth moment equals $E[X^4] = 3a^4$.]

Equipped with the above lemma, we are now ready to study an example that, in a nutshell, contains the main subtlety behind the definition of a stochastic integral. Suppose we want to define the integral $\int_0^T W_t dW_t$. For a smooth function ϕ , the integral $\int_0^T \phi(t) d\phi(t)$ may be obtained as the limit of a sequence of integral sums

$$\sum \phi(t_j^*) [\phi(t_{j+1}) - \phi(t_j)], \quad (5.8)$$

where, as earlier, $t_j = \frac{j}{n}T$ and t_j^* can be chosen in an arbitrary way between t_j and t_{j+1} . However, in the case of the integral sums

$$\sum W(t_j^*) [W(t_{j+1}) - W(t_j)], \quad (5.9)$$

the limit depends on the choice of t_j^* (for convenience, we are temporarily writing the value of the Wiener process as $W(t)$ rather than more common W_t). To see this, let

$$I_n = \sum W(t_j) [W(t_{j+1}) - W(t_j)] \quad (5.10)$$

and

$$J_n = \sum W(t_{j+1}) [W(t_{j+1}) - W(t_j)]. \quad (5.11)$$

Example 5.3: Different integral sums have different limits. Show that

$$\lim_{n \rightarrow \infty} I_n = \frac{1}{2} W_T^2 - \frac{1}{2} T \quad (5.12)$$

and

$$\lim_{n \rightarrow \infty} J_n = \frac{1}{2} W_T^2 + \frac{1}{2} T. \quad (5.13)$$

[Hint: Study the sum and the difference of I_n and J_n . Use Eq. (5.7).]

The above result shows that, when defining the stochastic integral as a limit of integral sums, one has to carefully and consistently choose the points at which the integrand is evaluated. For many purposes it is most convenient to evaluate it at the left endpoints of the intervals defining the integral sum. This leads to the fundamental definition of the Itô integral.

Definition (Itô integral): Let h_t be a stochastic process, *adapted* to the Wiener process W , i.e., such that for any t , h_t depends only on *the past*, i.e., on the Wiener variables W_s with $s \leq t$. We then define

$$\int_0^T h_s dW_s = \lim_{n \rightarrow \infty} \sum h(t_j) (W_{t_{j+1}} - W_{t_j}). \quad (5.14)$$

This limit can be shown to exist in L^2 , which is enough to identify it uniquely. We will not discuss all technical details here, referring the reader to mathematical literature on stochastic analysis, e.g., Ref. [2]. Instead, we will stress the most important points.

The role of the adaptedness assumption is that h_t is independent of the Wiener increments $W_s - W_r$ over time intervals in the future of t , i.e., with $r, s \geq t$. In particular, h_{t_j} is independent of $W_{t_{j+1}} - W_{t_j}$. The mean value of the product of independent random variables is the product of their mean values, which allows us to prove that the Itô integral sum

$$I_n = \sum h(t_j) (W_{t_{j+1}} - W_{t_j}) \quad (5.15)$$

has mean zero.

Example 5.4: Using the adaptedness assumption. Prove that indeed $\mathbf{E}[I_n] = 0$.

This relation is preserved in the limit and we obtain the following important property of Itô integrals:

$$\mathbf{E} \left[\int_0^T h_t dW_t \right] = 0. \quad (5.16)$$

Example 5.5: Variance of the Itô integral sum. Prove that

$$\mathbf{E}[I_n^2] = \sum \mathbf{E}[h(t_j)^2] (t_{j+1} - t_j). \quad (5.17)$$

[Hint: This example requires a little more complicated calculation, but uses the same facts about the Wiener process and adaptedness of h .]

Taking the limit $n \rightarrow \infty$ we obtain one of the most important facts of the Itô calculus, which is the Itô isometry.

Theorem (Itô isometry):

$$\mathbf{E} \left[\left(\int_0^T h_t dW_t \right)^2 \right] = \int_0^T \mathbf{E}[h_t^2] dt. \quad (5.18)$$

Example 5.6: Covariance of two Itô integrals. Extend the Itô isometry to show that for two adapted processes h_t and k_t

$$\mathbf{E} \left[\int_0^T h_t dW_t \cdot \int_0^T k_t dW_t \right] = \int_0^T \mathbf{E}[h_t k_t] dt. \quad (5.19)$$

[Hint: Use the method of Example 1.]

Now, that we have defined the stochastic integral, let us go back to the Eq. (5.1) which we can finally endow with a definite meaning.

Definition: A stochastic process x_t is a solution of Eq. (5.1) with the initial condition ξ_0 if $x_0 = \xi_0$ with probability one and for every t Eq. (5.4) holds.

5.1.2 The Itô formula

We emphasize that, while a solution to an ordinary differential equation is simply a function of time t , a solution to a stochastic one is a stochastic process which can be thought of as a random function of t . Stochastic differential equations are thus written using differentials, but interpreted using integrals — regular and stochastic. The following definition is a special case of a more general one.

Definition: A stochastic process y_t has a stochastic differential $u_t dt + v_t dW_t$ if for any t

$$y_t = y_0 + \int_0^t u_s ds + \int_0^t v_s dW_s. \quad (5.20)$$

In this case we write

$$dx_t = u_t dt + v_t dW_t. \quad (5.21)$$

It is clear that a process x_t solves an SDE if and only if its stochastic differential can be expressed in terms of two functions (b and σ) of x_t , according to Eq. (5.1), rewritten in terms of differentials:

$$dx_t = b(x_t) dt + \sigma(x_t) dW_t. \quad (5.22)$$

The differential relation

$$dy_t = u_t dt + v_t dW_t \quad (5.23)$$

is a stochastic analog of the familiar calculus relation

$$dy_t = u_t dt, \quad (5.24)$$

which is a way of expressing the fact that u is the derivative of y . The chain rule, which is the basic tool of calculus, asserts that if Eq. (5.24) holds, then for any continuously differentiable function g ,

$$dg(y_t) = g'(y_t) dy_t = g'(y_t) u_t dt. \quad (5.25)$$

The next theorem is the stochastic calculus version of the chain rule.

Theorem (Itô rule): If Eq. (5.21) holds, then for every twice continuously differentiable function g

$$dg(y_t) = g'(y_t) dy_t + \frac{1}{2} g''(y_t) v(y_t)^2 dt = \left[g'(y_t) u(t) + \frac{1}{2} g'(y_t) v(t)^2 \right] dt + g'(y_t) v_t dW_t. \quad (5.26)$$

We are soon going to introduce a simple way to remember this formula, as well as its multivariable generalization, which will be discussed later. First, let us apply the Itô rule to the case when $y_t = W_t$, i.e., $u_t \equiv 0$ and $v_t \equiv 1$. We obtain:

$$d(W_t^2) = dt + 2W_t dW_t, \quad (5.27)$$

which means that

$$W_t^2 = t + 2 \int_0^t W_s dW_s \quad (5.28)$$

recovering Eq. (5.12).

We are not going to prove Itô rule here. While the proof is more complicated than the derivation of Eq. (5.12), it uses similar ideas and estimates. For more details, we again refer the reader to Ref. [2].

To state the multidimensional version of the Itô formula, we consider stochastic differentials with several random terms:

$$dy_t^{(i)} = u_t^{(i)} dt + \sum_{\alpha=1}^m v_{\alpha}^{(i)} dW_t^{(\alpha)}, \quad (5.29)$$

where $W_t^{(\alpha)}$, $\alpha = 1, \dots, m$ are independent Wiener processes. We want a formula for the stochastic differential of $g(y^{(1)}, \dots, y^{(n)})$. To state it, we first define the product of stochastic differentials, postulating that:

$$dt \cdot dt = dt \cdot dW_t^{(\alpha)} = dW_t^{(\alpha)} \cdot dt = 0 \quad (5.30)$$

$$dW_t^{(\alpha)} \cdot dW_t^{(\beta)} = \delta_{\alpha\beta}. \quad (5.31)$$

The product of two arbitrary stochastic differentials of the form (5.29) is now defined using the distributive law. It allows to state the next theorem in a way that is easy to remember.

Theorem (multidimensional Itô formula): Let $y^{(i)}$, $i = 1, \dots, n$ be processes with stochastic differentials (5.29). If g is a twice continuously differentiable function of n variables, then the process $g(y^{(1)}, \dots, y^{(n)})$ has the stochastic differential

$$dg\left(y^{(1)}, \dots, y^{(n)}\right) = \sum_i \frac{\partial g}{\partial y^{(i)}} dy_t^{(i)} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 g}{\partial y^{(i)} \partial y^{(j)}} dy_t^{(i)} \cdot dy_t^{(j)}. \quad (5.32)$$

Example 5.7: Explicit form of the Itô formula. Rewrite the Itô formula explicitly in terms of the functions $u^{(i)}$ and $v_\alpha^{(i)}$, entering the expressions for the stochastic differentials $dy_t^{(i)}$.

Example 5.8: Itô product rule. Use the Itô formula to derive the expression for the differential of a product of two Itô processes:

$$d(x_t y_t) = x_t dy_t + y_t dx_t + dx_t \cdot dy_t. \quad (5.33)$$

[Hint: Use the multidimensional Itô formula with $g(x, y) = xy$.]

Note that the Itô formula reduces to the familiar product rule for regular differentials unless *both* stochastic differentials dx_t and dy_t contain nonzero stochastic terms.

5.1.3 The Ornstein-Uhlenbeck equation

We now have the tools to study a simple (and very important) SDE, called the *Ornstein-Uhlenbeck equation*.

$$dx_t = -kx_t dt + \sigma dW_t \quad (5.34)$$

with the initial condition x_0 at time 0. To this end, we apply the Leibniz rule to the product $e^{kt}x_t$, obtaining:

$$d(e^{kt}x_t) = ke^{kt}x_t dt + e^{kt}dx_t. \quad (5.35)$$

Substituting the expression Eq. (5.33) for dx_t and performing a cancellation, we get

$$d(e^{kt}x_t) = \sigma e^{kt} dW_t \quad (5.36)$$

so that

$$x_t = e^{-kt}x_0 + \sigma \int_0^t e^{-k(t-s)} dW_s. \quad (5.37)$$

From this representation of the solution as a stochastic integral, we obtain important information about its behavior.

Example 5.9: Moments of the solution to the Ornstein-Uhlenbeck equation. Find the first two moments of the solution, i.e., $\mathbf{E}[x_t]$ and $\mathbf{E}[x_t^2]$. Use the form Eq. (5.19) of Itô isometry to calculate its covariance function $\mathbf{E}[x_t x_u]$.

From the result of the last example, we obtain in the limit of $t, u \rightarrow \infty$

$$\mathbf{E}[x_t] \rightarrow 0 \quad (5.38)$$

and

$$\mathbf{E}[x_t x_u] \rightarrow \frac{\sigma^2}{2k} \exp(-k|u-t|). \quad (5.39)$$

In the next example, we show how to obtain a solution for which these relations hold exactly, not only in the limit, by admitting an initial condition which is random.

Example 5.10: Stationary solution of the Ornstein-Uhlenbeck equation. Suppose that x_0 is a Gaussian random variable with mean zero and variance $\frac{\sigma^2}{2k}$. Crucially, assume that x_0 is independent of the Wiener process W_t . The solution of the initial value problem for the Ornstein-Uhlenbeck equation with the initial condition x_0 is given by the same formula Eq. (5.37). Show that this solution satisfies

$$\mathbf{E}[x_t] = 0 \quad (5.40)$$

and

$$\mathbf{E}[x_t x_u] = \frac{\sigma^2}{2k} \exp(-k|u-t|). \quad (5.41)$$

This solution is a Gaussian process, i.e., the random variables x_t have a jointly Gaussian distribution. Such a distribution is characterized by the mean and the covariance (see literature). Since, for any a the expressions calculated above are invariant under the translation $t \mapsto t+a$, $u \mapsto u+a$, it follows that the distribution of the whole process x_t is invariant under translations, or *stationary*. The stationary solution of the Ornstein-Uhlenbeck equation is often used to model correlated (*colored*) noise.

Example 5.11: Noisy exponential growth. Solve the equation

$$dx_t = kx_t dt + \sigma x_t dW_t \quad (5.42)$$

with the initial condition x_0 at time $t = 0$. This equation can be rewritten as

$$dx_t = (k + \sigma \eta_t) dt \quad (5.43)$$

where η_t is a white noise, which can be interpreted as an exponential growth equation with a fluctuating growth constant $k + \sigma \eta_t$.

[Hint: Divide both sides of the equation by x_t ; use the Itô formula for $d \log x_t$.]

The answer is

$$x_t = x_0 \exp \left[\left(k - \frac{\sigma^2}{2} \right) t + \sigma W_t \right]. \quad (5.44)$$

As a continuous-time analog of a random walk, W_t is of order \sqrt{t} . It follows that the term $\left(k - \frac{\sigma^2}{2} \right) t$ dominates (as long as it is nonzero). This implies that if $k < \frac{\sigma^2}{2}$, the solution converges to zero as $t \rightarrow \infty$, with the interpretation that despite the growth constant being positive on average, its fluctuations lead the population to extinction.

5.2 Stratonovich integral

An alternative definition of the stochastic integral evaluates the integrand at midpoints of the intervals contributing to the integral sum. That is, we consider the sums

$$S_n = \sum h \left(\frac{t_j + t_{j+1}}{2} \right) [W_{t_{j+1}} - W_{t_j}]. \quad (5.45)$$

These sums no longer satisfy an isometry property of the type given by Eq. (5.17), but for a large class of integrals it can be shown that they also converge in the L^2 sense to a limit, thus defining another stochastic integral — the *Stratonovich integral*, which is usually denoted

$$\lim S_n = \int_0^t h_s \circ dW_s. \quad (5.46)$$

In general, the Stratonovich integral is different from the Itô integral, i.e.,

$$\int_0^t h_s \circ dW_s \neq \int_0^t h_s dW_s. \quad (5.47)$$

Example 5.12: A simple Stratonovich integral. Calculate the Stratonovich integral of a Wiener process, i.e., $\int_0^T W_t \circ dW_t$. While the calculation is not completely straightforward, but the answer is simple: $\frac{1}{2}W_T^2$.

The previous example is a case of a more general rule, which makes the Stratonovich integral convenient: unlike in the Itô calculus, the corresponding formulas for Stratonovich calculus (i.e., the chain rule, the product rule and others) look formally just like the formulas of the regular calculus, not containing any second-order corrections. It is important that a Stratonovich SDE can be rewritten as an equivalent Itô one at the price of changing the drift term.

Theorem (Stratonovich correction, one dimension): The Stratonovich equation

$$dx_t = b(x_t) dt + \sigma(x_t) \circ dW_t \quad (5.48)$$

is equivalent to the Itô equation

$$dx_t = b(x_t) dt + \frac{1}{2} \sigma(x_t) \sigma'(x_t) dt + \sigma(x_t) dW_t. \quad (5.49)$$

We leave out a proof, which is related to the calculation in Example 5.12.

5.3 General SDE systems

Let us now generalize the single SDE (5.1), considering a system of equations

$$dx_t^i = b^i(x_t) dt + \sum_{\alpha} \sigma_{\alpha}^i(x_t) dW_t^{\alpha}, \quad (5.50)$$

where $x_t = (x_t^1, \dots, x_t^m)$ and $W_t^{(\alpha)}$ with $\alpha = 1, \dots, n$ are independent Wiener processes, modeling different *noise sources*, which introduce randomness into the equations. We are going to associate with the system (5.50) two partial differential equations (PDE): the *forward* and *backward Kolmogorov equations*. In physics, the forward Kolmogorov equation is usually called *Fokker-Planck equation*. Both are equations for functions of t and x where $x = (x^1, \dots, x^m)$. The backward equation reads:

$$\partial_t f = \sum_i b^i(x) \partial_{x_i} f + \frac{1}{2} \sum_{ij} i j a^{ij}(x) \partial_{x_i x_j}^2 f, \quad (5.51)$$

where the matrix a^{ij} is obtained from the coefficients σ_{α}^i in (5.50) by the formula

$$a^{ij} = \sum_{\alpha} \sigma_{\alpha}^i \sigma_{\alpha}^j \quad (5.52)$$

or, in matrix form,

$$a = \sigma \sigma^T. \quad (5.53)$$

The backward Kolmogorov equation can be written as

$$\partial_t f = L f, \quad (5.54)$$

where the differential operator

$$L = \sum_i b^i(x) \partial_{x_i} + \frac{1}{2} \sum_{ij} a^{ij}(x) \partial_{x_i x_j}^2 \quad (5.55)$$

is called the *infinitesimal generator* of the diffusion process corresponding to the SDE system (5.50). The formal adjoint of f is the operator obtained from L by integration by parts:

$$L^* = - \sum_i \partial_{x_i} [b^i(x)f(x)] + \sum_{ij} \partial_{x^i x^j} [a^{ij}(x)f(x)]. \quad (5.56)$$

Using this notation, we define the forward Kolmogorov equation associated with the system (5.54) as

$$\partial_t f = L^* f. \quad (5.57)$$

We now explain the significance of these equations for the properties of solutions of the SDE system.

Proposition: Let ϕ be a function of $y \in \mathbf{R}^n$. Consider the solution x_t of Eq. (5.50) with the initial condition x at time $t = 0$. Define

$$f(t, x) = \mathbf{E}[\phi(x_t)]. \quad (5.58)$$

Then, f satisfies the backward Kolmogorov equation (5.54).

Proposition: Suppose a solution x_t of Eq. (5.50) is a random variable distributed in \mathbf{R}^m with the density $\rho(t, x)$. Then, ρ satisfies the forward Kolmogorov equation (5.57).

Proofs of the above statements can be found in texts on stochastic analysis, see, e.g., Ref. [2].

A probability distribution on \mathbf{R}^n is called a stationary distribution for the system (5.50) if it does not change under the evolution defined by it. Assuming that it has a density $\rho(x)$, it follows from the above proposition that ρ satisfies the stationary forward equation:

$$L^* \rho = 0. \quad (5.59)$$

In one dimension, this equation can be solved explicitly. Consider the equation

$$dx_t = b(x_t) dt + \sigma(x_t) dW_t. \quad (5.60)$$

The stationary forward equation reads:

$$-\partial_x(b\rho) + \partial_{xx}^2(\sigma\rho) = 0 \quad (5.61)$$

or

$$-b\rho + \partial_x(\sigma\rho) = C \quad (5.62)$$

where C is a constant.

Example 5.13: Explicit solution of the forward Kolmogorov equation in one dimension. Solve Eq. (5.62), choosing $C = 0$. This choice can be justified, see Ref. [3].

Example 5.14: A particular case of the Ornstein-Uhlenbeck equation. Show that the Ornstein-Uhlenbeck equation can be seen as a particular case of Eq. (5.62).

Example 5.15: Stationary distribution of the Ornstein-Uhlenbeck dynamics. Apply the expression obtained in Exercise 5.14 to the Ornstein-Uhlenbeck equation, recovering an earlier result about stationarity of a Gaussian distribution with an appropriate variance under the Ornstein-Uhlenbeck dynamics.

For many SDE systems, it can be shown that a stationary density ρ is unique and that with any initial condition (including random initial conditions) the density of the distribution of x_t converges to ρ [4].

5.4 Small mass limit

Getting closer to physics, let us study Newton's equation in the presence of noise, as in Chapter ?? [Check missing reference]. That is, we consider a particle in a force field F , moving in a medium, which causes random perturbations. Such a medium also inevitably leads to a drag force, slowing the particle down. Thus, in addition to F , the equation of motion contains two other terms, one of which is stochastic. For simplicity, we study the one-dimensional case

$$\begin{cases} dx_t &= v_t \\ m dv_t &= F(x_t) dt - \gamma(x_t) v_t dt + \sigma(x_t) dW_t \end{cases} \quad (5.63)$$

Since the drag term $-\gamma(x_t) dt$ and the noise term $\sigma(x_t) dW_t$ have the same physical source, they are not independent, but satisfy the *fluctuation-dissipation relation*

$$\sigma^2 = 2k_B T \gamma, \quad (5.64)$$

where T denotes absolute temperature and k_B is the Boltzmann constant[5].

Example 5.16: Deriving the canonical Gibbs distribution from stochastic dynamics. Find the stationary distribution of the stochastic process defined by Eq. (5.63).

[Hint: Solve the forward Kolmogorov equation. Remember that, even though the motion is one-dimensional, the stationary density is a function of two variables: x and v .]

5.4.1 Noise-induced drift

As a simplest case of small-mass limit, we now study equations describing the motion of a massive Brownian particle. We are interested in the regime when the particle's mass is small. The SDE describing the motion is Eq. (5.63) with the coefficients γ and σ satisfying the fluctuation-dissipation relation (5.64). Suppose that the diffusion is taking place at the temperature T . It is convenient to introduce a function $D(x)$ such that

$$\gamma(x) = \frac{k_B T}{D(x)}, \quad (5.65)$$

where k_B is the Boltzmann constant. Eq. (5.64) then implies that

$$\sigma(x) = \frac{k_B T \sqrt{2}}{\sqrt{D(x)}}. \quad (5.66)$$

Let us denote the solution of the resulting SDE system with initial condition (x_0, v_0) by $(x_t^{(m)}, v_t^{(m)})$. We are interested in the asymptotic behavior of this solution when $m \rightarrow 0$. It can be shown that in this limit $v^{(m)}$ diverges, so, more precisely, we want to study existence of the limit of $x_t^{(m)}$ and, whether this limit satisfies some SDE. Both questions have a positive answer.

Theorem: As $m \rightarrow 0$, $x_t^{(m)}$ converges to the solution of

$$dx_t = \left[\frac{F(x_t) D(x_t)}{k_B T} + D'(x_t) \right] dt + \sqrt{2D(x_t)} dW_t. \quad (5.67)$$

The proof is outlined in Ref. [6] and details can be found in Ref. [7]. The interest of the result is the presence of the term $D'(x_t)$. The two other terms can be obtained by a naïve calculation, setting $m = 0$ in Eq. (5.63). The additional term is an example of a *noise-induced drift*. See Ref. [6] for more details and context.

5.4.2 Generalizations

It is possible and important to generalize the above result. First, a general equation (5.63) can be considered, assuming no relation between γ and σ . In the limit $m \rightarrow 0$ one obtains

$$dx_t = \left(\frac{F(x_t)}{\gamma(x_t)} - \frac{1}{2} \frac{\gamma'(x_t)}{\gamma^3(x_t)} \sigma^2(x_t) \right) dt + \frac{\sigma(x_t)}{\gamma(x_t)} dW_t. \quad (5.68)$$

The full treatment of the more complicated multidimensional case can be found in Ref. [7].

5.5 Motion of light-sensitive robots with sensorial delay

We now study a mathematical model of an experiment involving light-sensitive robots and modeling an exploration of an inhomogeneous environment. After some approximations, we are going to cast it in the form of an SDE system, involving a small parameter. We will then conduct an asymptotic analysis of the system, finding the effective limiting equations and drawing from them conclusions about the dynamics of the system, comparing them with the experimental data.

To start with, consider a particle (a robot), exploring a planar region D [This can be confused with the diffusion coefficient]. The robot randomly changes the direction of its motion, according to a rescaled Wiener process. Suppose the speed of its motion at a point with coordinates x and y is $v(x, y)$, where v is a given scalar function on D [This can be confused with the diffusion coefficient]. More precisely, the robot adjusts its speed to the perceived location, but the perception and the adjustment occur with a delay for which we will account in the equations. They are thus as follows:

$$\begin{cases} dx_t &= \frac{1}{\varepsilon} v(x_{t-\varepsilon^2}, y_{t-\varepsilon^2}) \cos \phi_t dt \\ dy_t &= \frac{1}{\varepsilon} v(x_{t-\varepsilon^2}, y_{t-\varepsilon^2}) \sin \phi_t dt \\ d\phi_t &= \frac{1}{\varepsilon} dW_t \end{cases} \quad (5.69)$$

For small ε , the $\frac{1}{\varepsilon}$ prefactor in the third equation accounts for fast direction changes. The same prefactors in the first two equations scale the speed to keep the mean square distance covered by the particle of order 1 as ε changes. As we will see, scaling the sensorial delay as the square of ε ensures the existence of a nontrivial limit as $\varepsilon \rightarrow 0$.

The above equations are *stochastic delay equations*. They are very hard to study directly, so we will rewrite them in an approximate form. First, we linearize the functions x_t and y_t :

$$\begin{cases} x_{t-\varepsilon^2} &\approx x_t - c\varepsilon^2 \dot{x}_t \\ y_{t-\varepsilon^2} &\approx y_t - c\varepsilon^2 \dot{y}_t \end{cases} \quad (5.70)$$

Next, we apply a linearization to the function v :

$$v(x_t - c\varepsilon^2 \dot{x}_t, y_t - c\varepsilon^2 \dot{y}_t) \approx v(x_t, y_t) - c\varepsilon^2 v_x \dot{x}_t - c\varepsilon^2 v_y \dot{y}_t. \quad (5.71)$$

Substituting into Eq. (5.69), rewriting the equations in terms of derivatives rather than the differentials and rearranging, we obtain:

$$\begin{cases} (1 + c\varepsilon v_x \cos \phi_t) \dot{x}_t + c\varepsilon v_y \cos \phi_t \dot{y}_t &= \frac{1}{\varepsilon} v \cos \phi_t \\ c\varepsilon v_x \sin \phi_t \dot{x}_t + (1 + c\varepsilon v_y \sin \phi_t) \dot{y}_t &= \frac{1}{\varepsilon} v \sin \phi_t \end{cases} \quad (5.72)$$

This is a system of linear equations for \dot{x}_t and \dot{y}_t with the determinant [D below can be confused with the diffusion coefficient. Is it the same D that was mentioned above or does it have a different meaning?] This is a **determinant**, before it was a **domain**. I suggest that we use \mathcal{D} for one and \mathbb{D} for the other. Or some other symbol. We must ask Janek.

$$D = 1 + c\varepsilon v_x \cos \phi_t + c\varepsilon v_y \sin \phi_t. \quad (5.73)$$

Thus,

$$D^{-1} \approx 1 - c\varepsilon v_x \cos \phi_t - c\varepsilon v_y \sin \phi_t \quad (5.74)$$

and we obtain an approximate solution

$$\begin{cases} \dot{x}_t &= \frac{1}{\varepsilon} v \cos \phi_t - cvv_x \cos^2 \phi_t - cvv_y \cos \phi_t \sin \phi_t \\ \dot{y}_t &= \frac{1}{\varepsilon} v \sin \phi_t - cvv_x \cos \phi_t \sin \phi_t - cvv_y \sin^2 \phi_t \end{cases} \quad (5.75)$$

We may rewrite the above equations using differentials, and add an equation for ϕ_t , to obtain a system of three SDE driven by a single Wiener process:

$$\begin{cases} dx_t &= \left(\frac{1}{\varepsilon} v \cos \phi_t - cvv_x \cos^2 \phi_t - cvv_y \cos \phi_t \sin \phi_t \right) dt \\ dy_t &= \left(\frac{1}{\varepsilon} v \sin \phi_t - cvv_x \cos \phi_t \sin \phi_t - cvv_y \sin^2 \phi_t \right) dt \\ d\phi_t &= \frac{1}{\varepsilon} dW_t \end{cases} \quad (5.76)$$

We will study the system (5.76) in the limit $\varepsilon \rightarrow 0$. The main idea is to pass to the backward Kolmogorov equation, carry out the asymptotic analysis to obtain its limiting form, and reconstruct from it the limiting SDE. The Kolmogorov equation for this system reads:

$$\partial_t f = L^{(\varepsilon)} f, \quad (5.77)$$

where

$$L^{(\varepsilon)} = \varepsilon^{-2} L_{-2} + \varepsilon^{-1} L_{-1} + L_0 \quad (5.78)$$

with

$$\begin{cases} L_{-2} &= \frac{1}{2} \partial_{\phi\phi}^2 \\ L_{-1} &= v \cos \phi \partial_x + v \sin \phi \partial_y \\ L_0 &= -c (vv_x \cos^2 \phi + vv_y \cos \phi \sin \phi) \partial_x - c (vv_x \cos \phi \sin \phi + vv_y \sin^2 \phi) \partial_y \end{cases} \quad (5.79)$$

We write the solution to the Kolmogorov equation as a formal series in powers of ε :

$$f = f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \dots \quad (5.80)$$

Substituting this expression into Eq. (5.77) and equating the coefficients of consecutive powers of ε , we obtain:

$$\begin{cases} L_{-2} f_0 &= 0 \\ L_{-2} f_1 + L_{-1} f_0 &= 0 \\ L_{-2} f_2 + L_{-1} f_1 + L_0 f_0 &= \partial_t f_0 \end{cases} \quad (5.81)$$

It is natural to choose a function $f_0(x, y)$ as a solution of the first equation. The second equation becomes

$$\partial_{\phi\phi} f_1 = -2v \cos \phi \partial_x f_0 - 2v \sin \phi \partial_y f_0, \quad (5.82)$$

whence

$$f_1 = 2v \cos \phi \partial_x f_0 + 2v \sin \phi \partial_y f_0. \quad (5.83)$$

The third equation can be written as

$$\partial_t f_0 - L_0 f_0 - L_{-1} f_1 = L_{-2} f_2. \quad (5.84)$$

The left-hand side thus belongs to the range of the operator L_{-2} . As is well known, this range is equal to the orthogonal complement of the null space of the adjoint operator L_{-2}^* , which, in this case, is the same as L_{-2} . Note that ϕ is a periodic variable (with the period 2π). The null space of L_{-2} consists of the constants, so the orthogonality condition can be written as

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} d\phi [\partial_t f_0 - L_0 f_0 - L_{-1} f_1] = 0. \quad (5.85)$$

Executing the simple trigonometric integration and noting that f_0 does not depend on ϕ , we obtain the limiting Kolmogorov equation

$$\partial_t \rho = v^2 (\partial_{xx}^2 \rho + \partial_{yy}^2 \rho) + \left(1 - \frac{c}{2}\right) (vv_x \partial_x \rho + vv_y \partial_y \rho) = v^2 \Delta \rho + \left(1 - \frac{c}{2}\right) v \nabla v \cdot \nabla \rho, \quad (5.86)$$

which corresponds to the SDE system

$$\begin{cases} dx_t &= \left(1 - \frac{c}{2}\right) vv_x dt + \sqrt{2}v dW_t^{(1)} \\ dy_t &= \left(1 - \frac{c}{2}\right) vv_y dt + \sqrt{2}v dW_t^{(2)} \end{cases} \quad (5.87)$$

We see that, in contrast to the system (5.76), the limiting system is driven by two independent noises, $W_t^{(1)}$ and $W_t^{(2)}$. To explain this, we note that the distribution of the two-dimensional vector

$$\frac{1}{\sqrt{2}} \left(\frac{1}{\varepsilon} \int_0^T \cos \frac{W_t}{\varepsilon} dt, \frac{1}{\varepsilon} \int_0^T \sin \frac{W_t}{\varepsilon} dt \right) \quad (5.88)$$

converges to the distribution of a two-dimensional Wiener process $(W_T^{(1)}, W_T^{(2)})$ as $\varepsilon \rightarrow 0$. This approach to taking the limit was developed in detail in [8].

The obtained result is a basis for drawing interesting conclusions about the behavior of the robot. In particular, one can calculate the stationary probability distribution of the position of the robot, by solving the stationary Fokker-Planck equation

$$\nabla(v^2\rho) = \left(1 - \frac{c}{2}\right)\rho\nabla v. \quad (5.89)$$

Example 5.17: Stationary distribution of the robot's position. Show that the solution of Eq. (5.89) is

$$\rho(x, y) = Bv(x, y)^{1+\frac{c}{2}}. \quad (5.90)$$

where B is a constant. Normalizing the integral of ρ over x and y in the region where the motion takes place:

$$\int dx dy \rho(x, y) = 1. \quad (5.91)$$

The experiment conducted by M. Mijalkov *et al.* [1] to realize this system corresponds not to the original SDE (5.69) but to their approximate version (5.76). Because of this, the parameter c can have any sign either positive or negative. The formula (5.90) tells us that as the value of c crosses -2 , the behavior of the particle changes qualitatively: for $c > -2$, it spends more time where its speed is slower, but, for $c < -2$, it has a preference for the region where its motion is faster. This surprising effect has been confirmed experimentally and is a point of departure for designing robot swarms with different behaviors. The idea is for each robot to emit light and react to the light emitted by the other robots, adjusting its speed. The speed is a decreasing function of the light intensity, i.e., it is smaller where there is more light. The role of the robots in the experiment was played by robots programmed to satisfy equations of the form of SDE (5.76), except now v also depended on time, since at every time t it was determined by the positions of all other robots. Applying heuristically what we learned about one-particle motion, one is led to a conjecture that, for $c > -2$, the robots spend more time where they move slower, i.e., close to other robots (since the light intensity is greater there). For $c < -2$, this tendency should be reversed. The predicted behavior is actually observed in the experiment in Ref. [1].

It is worth mentioning that one can study (both experimentally and theoretically) a more complicated situation where the rate of the direction change also varies with the position of the robot. In this situation, the motivating system (5.69) is replaced by

$$\begin{cases} dx_t &= \frac{1}{\varepsilon}v(x_{t-\varepsilon^2}, y_{t-\varepsilon^2}) \cos \phi_t dt \\ dy_t &= \frac{1}{\varepsilon}v(x_{t-\varepsilon^2}, y_{t-\varepsilon^2}) \sin \phi_t dt \\ d\phi_t &= \frac{1}{s(x_{t-\varepsilon^2}, y_{t-\varepsilon^2})} dW_t \end{cases} \quad (5.92)$$

Depending on the relation between parameters c and k one can engineer several types of behavior, including new phenomena, absent in the previous, simpler situation. The details can be found in Ref. [9].

Let us also mention that another class of systems with time delay can be analyzed by methods similar to those used to study small mass limits. A prototypical situation of this type was discussed in Ref. [10]. Its physical realization was an RC circuit with a nonlinear modulation. The results were vastly generalized in Ref. [11], which discusses how effects of colored (i.e., time-correlated) noise can be accounted for.

Bibliography

- [1] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe. Engineering sensorial delay to control phototaxis and emergent collective behaviors. *Physical Review X*, 6:011008, 2016.
- [2] B. Øksendal. *Stochastic Differential Equations*. Springer, Springer Heidelberg New York Dordrecht London, 2014.
- [3] J. Birrell and J. Wehr. A homogenization theorem for Langevin systems with an application to Hamiltonian dynamics. In Vladas Sidoravicius, editor, *Sojourns in Probability Theory and Statistical Physics - I*, pages 89–122, Singapore, 2019. Springer Singapore.
- [4] L. R. Bellet. Ergodic properties of Markov processes. In Stéphane Attal, Alain Joye, and Claude-Alain Pillet, editors, *Open Quantum Systems II: The Markovian Approach*, pages 1–39, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [5] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, 2001.
- [6] G. Volpe and J. Wehr. Effective drifts in dynamical systems with multiplicative noise: a review of recent progress. *Reports on Progress in Physics*, 79(5):053901, 2016.
- [7] S. Hottovy, A. McDaniel, G. Volpe, and J. Wehr. The Smoluchowski-Kramers limit of stochastic differential equations with arbitrary state-dependent friction. *Communications in Mathematical Physics*, 336(3):1259–1283, 2015.
- [8] J. Birrell, Latifi, T.J. Reese, and J. Wehr. in preparation.
- [9] M. Leyman, F. Ogemark, J. Wehr, and G. Volpe. Tuning phototactic robots with sensorial delays. *Physical Review E*, 98:052606, 2018.
- [10] G. Pesce, A. McDaniel, S. Hottovy, J. Wehr, and G. Volpe. Stratonovich-to-Itô transition in noisy systems with multiplicative feedback. *Nature Communications*, 4:2733, 2013.
- [11] S. H. Lim and J. Wehr. Homogenization for a class of generalized Langevin equations with an application to thermophoresis. *Journal of Statistical Physics*, 174:656–691, 2019.

Chapter 6

Anomalous Diffusion and Ergodicity Breaking

CARLO MANZO, GORKA MUÑOZ-GIL

Understanding the dynamic behavior of individual particles is of paramount importance across a wide range of disciplines, encompassing biophysics, materials science, and beyond. The intimate relationship between diffusion and single-particle motion becomes invaluable for unraveling stochastic movement and extracting profound insights into the underlying processes. This chapter will focus on the quantification of individual particle trajectories and the extraction of distinctive parameters, enabling us to assess dynamic interactions that govern intriguing phenomena in heterogeneous environments (Fig. 6.1). This chapter provides some basic

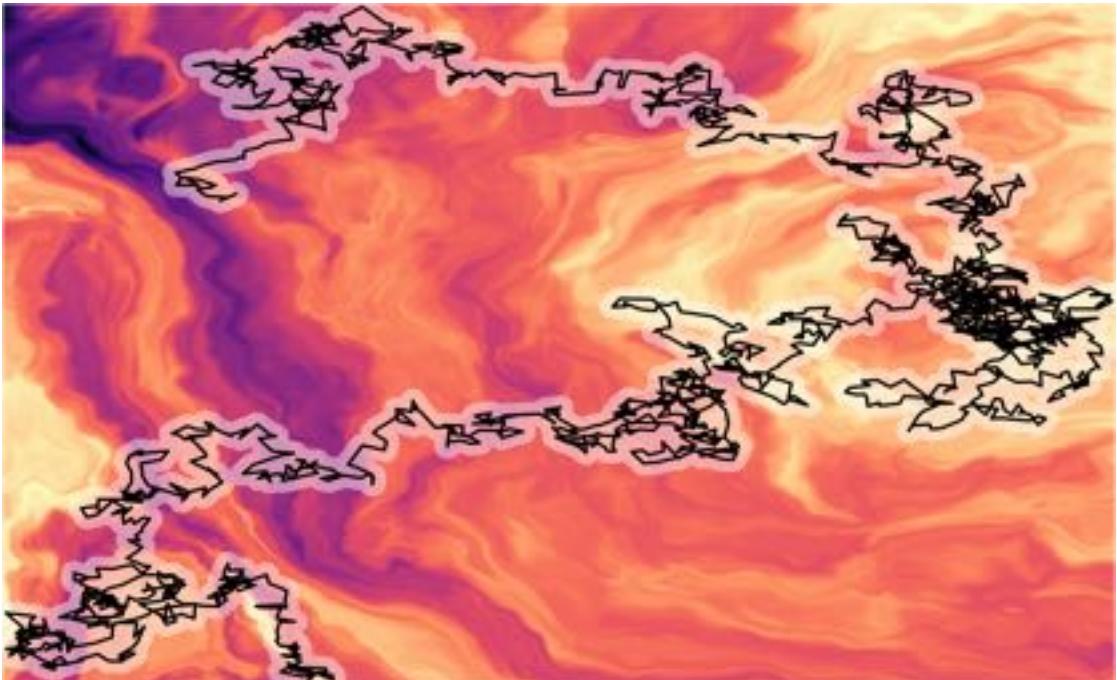


Figure 6.1: **Single-particle trajectory in a heterogeneous medium.** The quantification of individual particle trajectories and extraction of physical parameters provide insights into the dynamic interactions governing transport processes in heterogeneous environments. The image illustrates the trajectory of a Brownian particle moving through an idealized complex environment. The background was created using domain warping of fractal noise, as introduced in Ref. [1]. Image created by G. Muñoz-Gil.

knowledge about the principle of diffusion and Brownian motion, and further details can be found in Chapter ???. In the following, we will describe the conditions that lead to deviations from Brownian motion and produce the emergence of anomalous diffusion and ergodicity breaking. We will briefly introduce two popular theoretical models of anomalous diffusion. We will review some of the classical estimators used to quantify anomalous diffusion and ergodicity breaking, both at the ensemble and single-trajectory levels. Finally, we will describe how machine-learning-based approaches can help estimate anomalous diffusion and weak ergodicity breaking from individual trajectories.

6.1 Introduction

The word diffusion derives from the Latin “diffundere”, which means “to spread out”. The meaning of this word is strictly connected to the idealized experiment typically used to represent diffusion, depicting a drop of ink progressively increasing its area in water while losing its original color. Based on Fick’s laws, the *phenomenological approach* to diffusion describes it as the movement of a substance from a region of high concentration to a region of low concentration [2]. The *atomistic description* provides an alternative way to introduce the same phenomenon. According to this view, particles composing a substance perform random walks, propelled by the thermal energy transferred through collisions with surrounding particles. The atomistic framework was developed simultaneously by Albert Einstein [3] and Marian Smoluchovsky [4]. This frame-

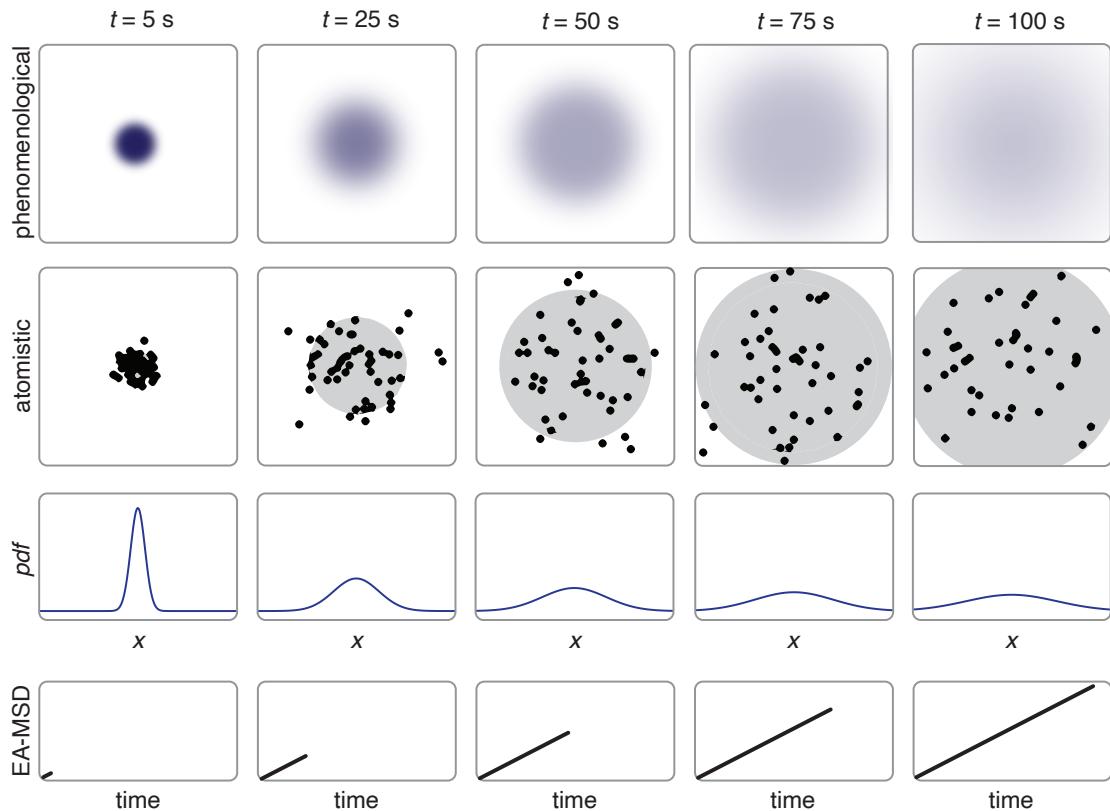


Figure 6.2: Schematic view of the diffusion process and ensemble-averaged mean squared displacement. The phenomenological approach describes diffusion in terms of changes in the concentration of a substance, whereas the atomistic approach describes it in terms of a probability density of finding a particle in space and time. For an amount of substance released in water at $x = 0$ for time $t = 0$, both the concentration and the probability density spread over time. The Gaussian distribution with linearly increasing variance depicts the time evolution of both quantities. The variance is proportional to the average area explored by particles in time, leading to the ensemble-averaged mean squared displacement (EA-MSD).

work managed to explain the famous observation by Robert Brown, who described particles suspended in a liquid as undergoing a rapid and erratic movement, now known as *Brownian motion* [5].

As shown in Fig. 6.2, the two approaches describe the same phenomenon either in terms of a concentration of a substance (phenomenological) or in terms of a probability density of finding a particle in space and time (atomistic). If a given amount of substance is released at $x = 0$ for $t = 0$, the solution of Fick's law is a Gaussian distribution, depicting how both the concentration and the probability density spread. In 1D, the time evolution of the variance of this Gaussian is given by

$$\langle x^2(t) \rangle = 2Dt, \quad (6.1)$$

where D is the diffusion coefficient. Eq. (6.1) represents the mean squared displacement (MSD), a measure of the area explored by the particles in the time t . Because of its relation with Fick's law, this behavior is typically referred to as *Fickian diffusion*.

Einstein's theory [3] considers three fundamental assumptions:

- (i) The particles are independent, which always holds true at low concentrations.
- (ii) There exists a timescale τ such that consecutive particle displacements performed in intervals of duration τ are statistically independent and uncorrelated.
- (iii) The distribution of displacements performed in intervals of duration τ has finite variance.

These assumptions reflect the hypothesis that particle displacements correspond to random kicks by molecules of the surrounding fluid with no preferential direction [2].

This picture can be further extended to cases where particles are sampled at variable time intervals. In these cases, displacements and waiting times are decoupled, providing two independent random variables. Thus, assumption (iii) also requires the distribution of waiting times to have a finite mean [6].

The same assumptions lead to another derivation of Brownian motion through the overdamped Langevin equation [7], obtained by neglecting inertia effects in the equation of motion (see Chapter ??). The motion of a single particle can be described as:

$$\frac{dx(t)}{dt} = \xi(t), \quad (6.2)$$

where the term on the right-hand side corresponds to a stochastic force ξ , often modeled as a Gaussian process with zero mean and autocorrelation, i.e.,

$$\langle \xi(t) \xi(t') \rangle = 2D\delta(t - t'). \quad (6.3)$$

Equation (6.2) can be integrated through a finite-difference approach using the Euler-Mayurama method to obtain

$$x(t + \Delta t) - x(t) = \sqrt{2D\Delta t}\eta(t), \quad (6.4)$$

where η is a standard normal random number. This derivation is relevant for the simulation of Brownian dynamics and for the description of single-particle experiments, where the position of a particle is recorded at L discrete times $t_i = i\Delta t$. In this context, one can obtain a discretized version of the MSD by calculating the *time-averaged mean squared displacement* (TA-MSD) of a single particle,

$$\text{TA-MSD}(t_{\text{lag}} = m\Delta t) = \frac{1}{L-m} \sum_{i=1}^{L-m} [x(t_i + m\Delta t) - x(t_i)]^2 = 2Dt_{\text{lag}}. \quad (6.5)$$

An intuitive picture of the calculation of the TA-MSD is given in Fig. 6.3. For a fixed timelag t_{lag} , it corresponds to the average of the squared displacements of segments of trajectory with duration t_{lag} obtained by sliding the initial time t_i . Similarly, starting from trajectory coordinates and having access to an ensemble of N trajectories, one can calculate the *ensemble-averaged mean squared displacement* (EA-MSD) as

$$\text{EA-MSD}(t) = \frac{1}{N} \sum_{i=1}^N (x(t) - x(0))^2 = 2Dt. \quad (6.6)$$

By identifying the timelag t_{lag} in Eq. (6.5) with the regular time t in Eq. (6.6), we find the equivalence between time and ensemble averages in the limit of long measurement times, which holds true for *ergodic processes* [2].

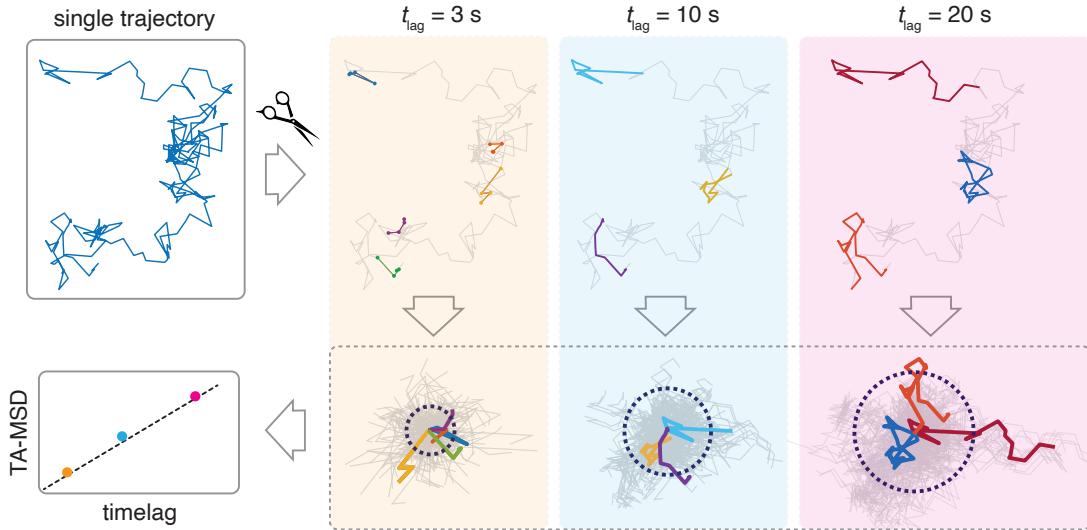


Figure 6.3: Time-averaged mean squared displacement. The calculation of the time-averaged mean squared displacement (TA-MSD) can help to understand its meaning in analogy to the ensemble-averaged mean squared displacement (EA-MSD). For a given timelag t_{lag} , the TA-MSD is obtained as described in Eq. (6.5). The calculation corresponds to chopping the trajectory in segments of the same duration t_{lag} but with different starting times, and then averaging the squared displacements across each segment. The latter operation is equivalent to calculating the average area explored by all the segments during t_{lag} , as they would correspond to different trajectories.

Example 6.1: Simulations of Brownian motion

- Implement a code to simulate trajectories performing Brownian motion using Eq. (6.4). Generate 1000 trajectories of length $L = 100$ and $D = 1$.
- Calculate the TA-MSD for each trajectory and then the EA-MSD for the ensemble.
- Estimate the diffusion coefficient by performing a linear fit of the TA-MSD and the EA-MSD and compare the results. For the TA-MSD, calculate the diffusion coefficient for each trajectory and then average over all the trajectories. Plot a histogram of the values obtained for all trajectories.
- Repeat the calculation by changing the initial point and the number of points used for the fit. This corresponds to restricting the fitting to a range of t_{lag} for the TA-MSD or t for the EA-MSD, respectively. Is there an optimal choice?
- Corrupt the trajectories by adding Gaussian noise with standard deviation $\sigma_n = 0.1, 0.5$ and 1 . This mimics the limited localization precision obtained in experiments. Does the level of noise affect the estimation of D ?

[Hint: Example code for the calculation of the TA-MSD, the EA-MSD, and the TEA-MSD is provided in Ref. [8]. You can compare your results to those obtained with functions contained in Ref. [9].]

6.2 Anomalous diffusion and ergodicity breaking

Anomalous diffusion typically refers to deviations from the linear time dependence of the MSD in the form of a power law:

$$\text{MSD} \sim t^\alpha, \quad (6.7)$$

where we use the notation MSD when we refer generically to any of the forms defined in Eqs. (6.5) and (6.6). As shown in Fig. 6.4, both averages can feature deviations from the linear behavior. The regime $\alpha < 1$ is typically referred to as *subdiffusion*, whereas $\alpha > 1$ corresponds to *superdiffusion* [10]. Anomalous diffusion can emerge from the violation of any of the assumptions described in the previous section for Einstein's diffusion. If the

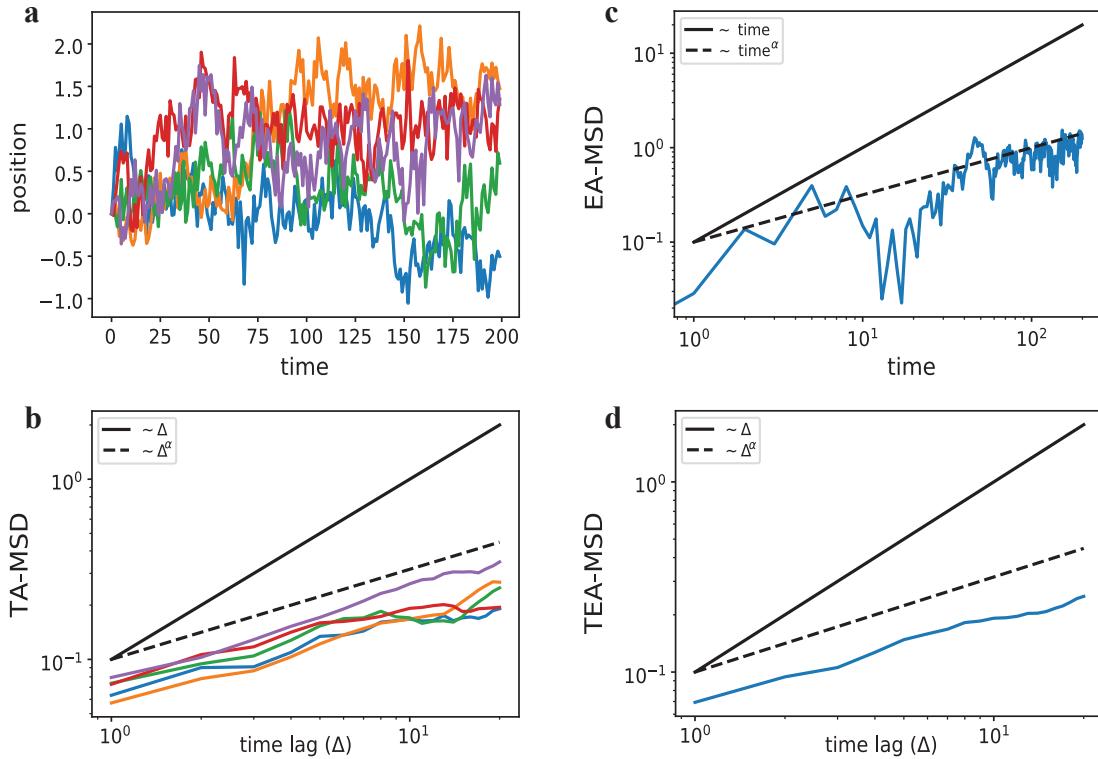


Figure 6.4: The mean squared displacement in an anomalous diffusion process. (a) Examples of simulated trajectories in 1D undergoing anomalous diffusion with $\alpha = 0.5$. (b) Log–log plot of the TA-MSD for each trajectory as a function of the timelag. (c) Log–log plot of the EA-MSD as a function of time. (d) Log–log plot of the TEA-MSD as a function of the timelag.

central limit theorem does not hold anymore, the process does not converge to the Fickian/Brownian behavior predicted by Einstein. The non-linear scaling of the MSD has been widely observed in a variety of fields [2]. For example, in biological systems, it has been generally associated to the effect of molecular crowding [11], traps with a wide distribution of binding energies or escape times [12], heterogeneity in the interaction with the environment [13], or viscoelastic effects [14].

We have already anticipated that, for Brownian diffusion, the equivalence of the scaling of the MSD obtained for an ensemble of particles (EA-MSD) and for an individual trajectory (TA-MSD) suggests that the process is ergodic. Therefore, following a single representative particle over a long time provides the same information as monitoring the ensemble. However, for some anomalous stochastic processes, such equivalence is broken, giving rise to an exotic effect called *weak ergodicity breaking*. The concept of weak ergodicity breaking was originally introduced for physical glasses [15]. In these systems, the phase space is not separated into mutually inaccessible domains, as for *strong ergodicity breaking*. However, the time required for a single particle to explore the phase space is infinite. The system is thus non-ergodic from the single particle point-of-view. Weak non-ergodic systems display non-reproducible TA-MSD curves calculated for different trajectories of identical particles (Fig. 6.5d).

6.3 Anomalous diffusion models

A large variety of stochastic processes can yield anomalous diffusion in the power-law form of the MSD expressed in Eq. (6.7) [2]. In this section, we will briefly review two of the most popular theoretical models used to describe anomalous diffusion, as examples of ergodic and non-ergodic behavior.

6.3.1 Fractional Brownian motion

Fractional Brownian motion (FBM) is a random walk with correlated steps, used to describe, for example, viscoelastic effects [16]. FBM arises from Eq. (6.2) but with the stochastic force ξ_{FBM} modeled as a fractional Gaussian process [16]. The latter has a normal distribution with zero mean but power-law autocorrelation:

$$\langle \xi_{\text{FBM}}(t)\xi_{\text{FBM}}(t') \rangle = \alpha(\alpha-1)K_\alpha|t-t'|^{\alpha-2}, \quad (6.8)$$

for $t, t' > 0$ and $t \neq t'$. Anomalous diffusion thus arises here from breaking Einstein's assumption (ii) and the MSD scales as t^α . FBM features a superdiffusive regime for positively-correlated noise ($1 < \alpha < 2$), and a subdiffusive regime for negatively-correlated noise ($0 < \alpha < 1$). For $\alpha = 1$, the noise is uncorrelated: assumption (ii) is verified and hence Brownian motion is recovered. FBM has a Gaussian distribution of displacement and an ergodic behavior [2]. The TA-MSD obtained from different trajectories shows the same scaling as the EA-MSD (Figs. 6.5a-b).

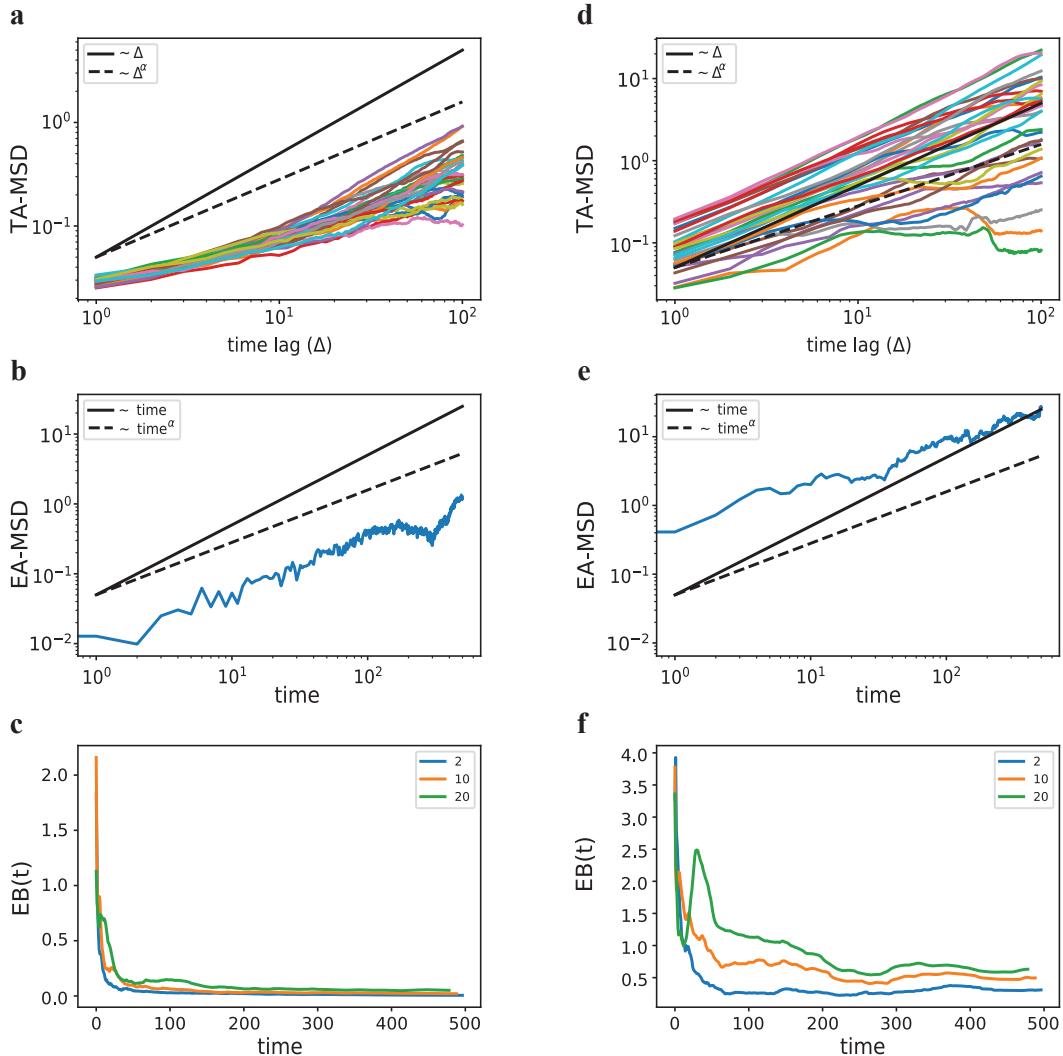


Figure 6.5: Ergodic and non-ergodic theoretical models of anomalous diffusion. (a, d) Log–log plot of the TA-MSD for simulated trajectories in 1D undergoing either FBM (a) or CTRW (d) with $\alpha = 0.75$. (b, e) EA-MSD for simulated trajectories in 1D undergoing either FBM (b) or CTRW (e) with $\alpha = 0.75$. (c, f) Ergodicity breaking parameter calculated at three different timelags for simulated trajectories in 1D undergoing either FBM (c) or CTRW (f) with $\alpha = 0.75$. [Define legend parameter in c, f]

6.3.2 Continuous time random walk

The *continuous time random walk* (CTRW) defines a family of random walks for which the waiting time between consecutive steps is a stochastic variable [17]. It is often used to describe walkers undergoing local trapping with a pathological distribution of waiting times [17]; as an example, we will consider waiting times sampled from a power-law distribution $\psi(t) \sim t^{-\sigma}$ and displacements sampled from a Gaussian distribution. In such a case, the anomalous diffusion exponent is given by $\alpha = \sigma - 1$. Since trajectories are sampled at irregular times, to obtain Brownian diffusion one should fulfill the extended version of assumption (iii), requiring the finite mean of the waiting time distribution. Thus, this assumption is not verified for $1 < \sigma < 2$, producing a subdiffusive behavior. Subdiffusive CTRW displays weak ergodicity breaking. TA-MSD curves show a large scattering of values for the same timelag and nearly linear scaling, in contrast to the power-law scaling of EA-MSD (Fig. 6.5d-e).

6.4 Estimators

In this section, we review some of the classical estimators used to quantify anomalous diffusion and ergodicity breaking, both at the ensemble and single-trajectory levels.

6.4.1 Anomalous diffusion exponent

The most direct and straightforward way to determine anomalous diffusion from a trajectory relies on the fit of either the TA-MSD or the EA-MSD. This procedure allows the estimation of the *anomalous diffusion exponent* α . Besides providing a way to discriminate between Brownian ($\alpha = 1$) and anomalous diffusion ($\alpha \neq 1$), as we have seen above, it provides a parameter that is strictly related to the fundamental characteristics of the underlying stochastic process, such as its temporal correlations and the distribution of its waiting times.

Despite its apparent simplicity, the fit of the MSD presents several drawbacks [18, 19]. First, the accuracy of the estimation depends on fluctuations, which can be reduced only by increasing the number of trajectories (in the case of EA-MSD) or the length of the trajectory (in the case of TA-MSD), which is often not possible in practice. For TA-MSD, the larger the timelag, the less the number of squared displacements. Therefore, TA-MSD curves feature a larger variance at large timelags [20]. One might thus suggest using only the first points of the MSD to reduce the effect of these points. However, for several stochastic processes, the anomalous regime is reached only asymptotically. Fitting too few points might completely miss the anomalous diffusion regime. Moreover, for finite trajectories, the MSD also features fluctuations at small timelags. A first problem to face is thus the identification of the fitting timescale [19].

Another factor that must be taken into account is that the value of α obtained through fitting is biased by noise and measurement error. In fact, the MSD is typically fitted by performing a linear regression of its logarithm with respect to the logarithm of time. As such, α is obtained as the slope of the resulting straight line. However, in the presence of a normally-distributed noise with variance σ_n (e.g., due to the finite localization precision to determine particle position), the MSD in one dimension can be written as:

$$\text{MSD} = 2D_\alpha t^\alpha + 2\sigma_n^2, \quad (6.9)$$

where D_α is a generalized diffusion coefficient with dimensions $[D_\alpha] = L^2 T^{-\alpha}$. The logarithmic transformation does not simplify the calculation and the slope provides an apparent exponent α_{ap} that underestimates the true α value, even for Fickian/Brownian behavior ($\alpha = 1$) [18]:

$$\alpha_{ap} = \alpha \left[\frac{1}{1 + \sigma_n^2 / D_\alpha t^\alpha} \right]. \quad (6.10)$$

The correlation between α_{ap} and σ_n can lead to a significant error in the calculation of the anomalous diffusion exponent unless σ_n can be independently estimated [19].

Last, as described above, for non-ergodic systems EA-MSD and TA-MSD might have different scaling, a characteristic that invalidates the use of TA-MSD to determine the anomalous diffusion exponent.

Example 6.2: Simulations of Anomalous Diffusion and the TA-MSD

a. Using the `andi_datasets` package [8], simulate $N = 100$ trajectories of FBM with exponent $\alpha \in (0, 1]$ and length $L = 200$ (see [this tutorial](#) for details).

b. Calculate the TA-MSD for each trajectory considering timelags from 1 to 100 and display it in a plot.

[Hint: Example code is provided [here](#).]

6.4.2 Ergodicity breaking

The fingerprint of weak ergodicity breaking is represented by the different scaling of the EA-MSD and the TA-MSD in the long-time limit. Therefore, a first evidence of non-ergodic behavior can be obtained by comparing the anomalous diffusion exponents of the two estimators, following the warnings provided for the fitting procedures. It must be noticed that, while the EA-MSD produces a unique curve (Fig. 6.4c) and thus only one exponent, TA-MSD produces as many curves as the number of trajectories (Fig. 6.4b), hence being able to extract the same number of exponents by their fitting. For practical use, one might either average the exponents obtained from TA-MSD [21] (as we did in Problem 6.1) or calculate the time-ensemble-averaged mean squared displacement:

$$\text{TEA-MSD} = \frac{1}{N} \sum_{i=1}^N \text{TA-MSD}_i \quad (6.11)$$

where N is the number of trajectories. By fitting this curve, one can estimate its anomalous diffusion exponent (Fig. 6.4d).

A striking feature of non-ergodic behavior is that TA-MSD curves are not identical (Fig. 6.5d). Thus, TA-MSD remains a random variable even for long trajectories. In contrast, for ergodic diffusion, TA-MSD will converge for long measurement times and will show the same exponent as that of the ensemble average (Fig. 6.5a). The *ergodicity breaking parameter* [22],

$$\text{EB}(t_{\text{lag}}) = \frac{\langle \text{TA-MSD}(t_{\text{lag}})^2 \rangle}{\text{TEA-MSD}(t_{\text{lag}})^2} - 1, \quad (6.12)$$

measures the variance of the fluctuations of TA-MSD at a given timelag as a function of the measurement time and has thus been proposed as an estimator of ergodicity (Fig. 6.5c,f). Its asymptotic value is independent of the timelag. For a CTRW, the long time limit of EB monotonically varies from 1 for $\alpha = 0$ to 0 for $\alpha = 1$ (ergodic).

However, both methods described so far for assessing the ergodicity of a given process require the evaluation of an ensemble of trajectories. For a single trajectory, there exist ergodicity estimators based on functionals of the trajectory position or displacement [23, 24]. In particular, the ergodicity estimator [24]

$$\hat{F}_{\omega}(n, N) = \frac{1}{n} \sum_{k=1}^n \hat{E}_{\omega}(k, N) \quad (6.13)$$

based on the mixing functional

$$\begin{aligned} \hat{E}_{\omega}(n, N) &= \frac{1}{N-n+1} \sum_{k=0}^{N-n} e^{i\omega[x(k+n)-x(k)]} \\ &\quad - \frac{1}{N(N+1)} \left| \sum_{k=0}^N e^{i\omega[x(k)-x(0)]} \right| + \frac{1}{N} \end{aligned} \quad (6.14)$$

vanishes for ergodic models of anomalous diffusion, whereas it remains nonzero for non-ergodic models at long-time increments or positions of a single trajectory, removing the need to perform ensemble averages over many trajectories.

Example 6.3: Ergodicity breaking

- a.** Compare the anomalous exponents obtained via TA-MDS and EA-MSD. Follow the steps:
 1. Using the `andi_datasets` package [8], simulate $N = 100$ trajectories of CTRW with exponent $\alpha = 0.75$, $\sigma_n = 0.5$ and length $L = 200$.
 2. Calculate the TA-MSD for each trajectory.
 3. Calculate the EA-MSD for the ensemble of trajectories.
 4. Calculate the TEA-MSD, averaging the TA-MSD curves.
 5. Compare the average anomalous diffusion exponent obtained for TA-MSD with those one obtained from EA-MSD and TEA-MSD.
- b.** Calculate the ergodicity breaking parameter for the ensemble of trajectories.
- c.** Assess ergodicity breaking at the single-trajectory level: calculate the ergodicity estimator for each trajectory.
- d.** Ergodicity breaking dependence on the diffusion model. Repeat the steps given in points **a**, **b**, **c** for trajectories undergoing FBM.

6.5 Machine learning

Recently, the characterization of anomalous diffusion has received large attention in the scientific literature. Several methods, based both on classical statistics and on machine learning approaches, have been developed to quantify anomalous diffusion from single trajectories. A non-exhaustive list of articles about this subject includes Refs. [19, 25, 26, 27, 28, 29, 30, 31, 30, 32, 33, 34, 35].

Mainly, these methods aim at characterizing single trajectories according to two different schemes: the discrimination among diffusion models and the prediction of the anomalous diffusion exponent, which inherently implies classification as normal or anomalous diffusion. To perform an objective comparison of these methods, we launched a joint effort in the form of a competition, the Anomalous Diffusion (AnDi) Challenge (www.andi-challenge.org). Methods were thus benchmarked on the same reference dataset. Interestingly, the challenge revealed that all of the top-scoring methods used, at certain steps, machine learning approaches. The results of the competition are described in Ref. [36].

6.5.1 Convolutional neural network for trajectory analysis

To provide an example of a machine-learning approach for the characterization of individual trajectories, we will describe a simple architecture based on a convolutional neural network (CNN). The architecture will be applied to the two problems: the determination of the anomalous diffusion exponent and the classification of ergodic behavior at the single trajectory level.

CNNs are neural networks composed of convolutional filters of various sizes. They were initially developed for the analysis of images, as they are particularly suitable to exploit correlations between values of adjacent pixels. Nevertheless, they have been applied to a plethora of different problems. For instance, the analysis of short-time correlations may also be helpful in time series analysis, and more precisely, in single trajectory characterization.

The CNN that we will describe (Fig. 6.6) utilizes 1D convolutions for processing time series data. We provide the code for its implementation as a sequential model using the Keras library in Ref. [37]. The CNN takes as input a time series with length T and is composed of three convolutional layers that extract features while progressively extending the receptive field (the first with 3 filters and a kernel size of 3; the second with 8 filters and a kernel size of 5; the third, with 3 filters and a kernel size of 2). These layers use a ReLU activation function and apply L2 regularization with a parameter of 0.001. A batch normalization layer is added after each convolutional layer to normalize the activations of the previous layer using a momentum value of 0.99 and a small epsilon value of 0.001 to avoid division by zero. The model then flattens the output of the previous

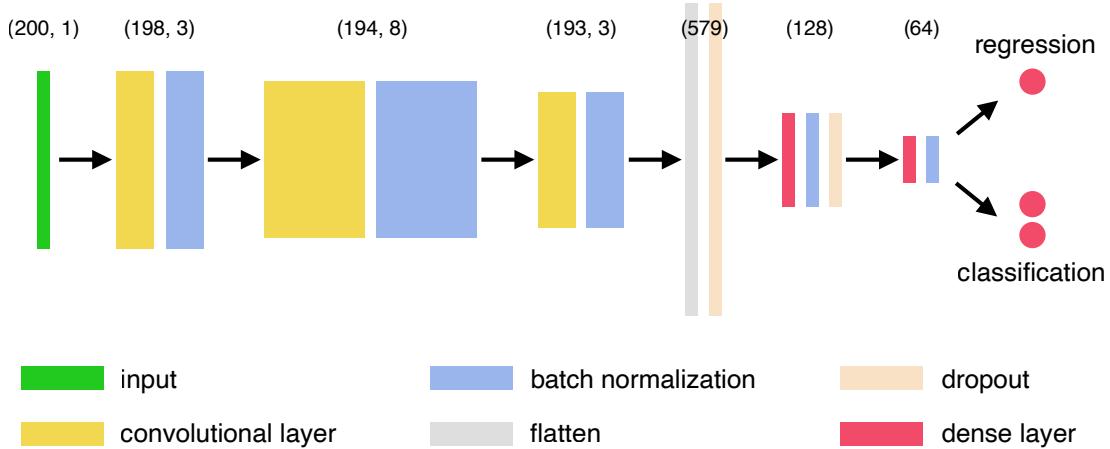


Figure 6.6: **Architecture of the convolutional neural network.** Schematic architecture of the CNN used in the examples for predicting the anomalous exponent (regression) and classifying trajectories as ergodic or non-ergodic (classification). The dimensions of the features at every block are indicated on the top.

layer into a 1D vector. A dropout layer with a rate of 0.5 is applied to reduce overfitting. The convolutional part is followed by a dense top, composed of two fully-connected dense layers with 128 and 64 units, with sigmoid activation function and batch normalization. This architecture aims to capture temporal patterns in the input time series data through the use of convolutional layers. The batch normalization and dropout layers help regularize the model and prevent overfitting. The sigmoid activation functions in the dense layers are used for binary classification tasks, where the output values are interpreted as probabilities. The output layer's size depends on the problem one wants to solve. For predicting the anomalous exponent, it is sufficient to provide a single neuron with a linear activation that will output the estimation of α . For classifying a trajectory between two classes, e.g., ergodic and non-ergodic, one will instead need to use a two-neuron output layer with a softmax activation that will output the probability of belonging to each class.

As shown in the following examples, this architecture described so far allows us to obtain respectable performance for the two tasks. However, to create more robust neural networks and enhance their performance, it is recommendable to explore how performance is affected by adding other layers and fine-tuning the hyperparameters [38].

6.5.2 Application of a CNN for trajectory analysis

Through the following examples, we will guide the reader through the simulation of trajectories and the analysis of the features described above. We will explore the effect of the noise level and the choice of the fitting timescale on the calculation of the anomalous diffusion exponent. We will further compare results obtained with classical and machine-learning approaches. However, these examples are only intended as guidelines, as a starting point for further exploration. Example code can be downloaded at Ref. [37].

Example 6.4: Prediction of the anomalous diffusion exponent

- Using the `andi_datasets` package, simulate $N = 5000$ trajectories of FBM with exponent $\alpha \in (0, 1]$, $\sigma_n = 0.1$, and length $L = 200$.
- Use the dataset to train the convolutional neural network to predict the anomalous diffusion exponent.
- Use the trained model to predict the exponent of the trajectories simulated in Example 6.2.
- Examine the anomalous diffusion exponent predicted by the neural network in comparison to the outcomes obtained through the fitting the TA-MSD on the same set of trajectories. In the absence of localization error, both approaches are expected to yield similar results. Conversely, when trajectories are

corrupted with localization error, the convolutional neural network trained on noisy trajectories is expected to provide superior performance compared to the TA-MSD fitting. Repeat the aforementioned procedures employing CTRW trajectories. The neural network is expected to reliably predict the anomalous diffusion exponent, which for CTRW is associated with ensemble behavior, from individual trajectories.

Example 6.5: Classification of ergodicity

- a. Using the `andi_datasets` package, simulate $N = 2000$ trajectories of FBM and $N = 2000$ trajectories of CTRW with exponent $\alpha = 0.75$, $\sigma_n = 0.5$, and length $L = 200$.
- b. Use the dataset to train the convolutional neural network to classify trajectories between two output classes: ergodic and non-ergodic.
- c. Use the trajectory simulated in Examples 6.3a and 6.3d as a test set. Calculate the accuracy of the model on these trajectories.
- d. Compare the classification results obtained with the convolutional neural network with those obtained through the calculation of the ergodicity estimator. The convolutional neural network is expected to correctly classify relatively short trajectories, while the ergodicity estimator is accurate at long times.

6.6 Problems

Problem 6.1: Regression of the TA-MSD

With the results obtained in Example 6.2, calculate the anomalous diffusion exponent by performing a linear regression of the logarithm of the TA-MSD with respect to the logarithm of the timelag. Compare the results obtained when considering the following conditions:

- a. Using only the first 10% of the timelags of the TA-MSD.
- b. Using the second 10% of the timelags of the TA-MSD.
- c. After corrupting the trajectories with localization noise, e.g., $\sigma_n = 1$.
- d. After subtracting the noise term $2\sigma_n^2$ from the TA-MSD.
- e. By fitting the TA-MSD using one free parameter, explicitly including the information about the noise as a fixed parameter.
- f. Repeat the steps of Example 6.2 and Problem 6.1a-e for trajectories undergoing CTRW.

Problem 6.2: Improving the CNN performance.

Modify the convolutional neural network to improve the network performance.

- a. Try to change some of the hyperparameters (e.g., the number of convolutional filters, the kernel size, or the dropout rate) to improve your results.
- b. Modify the architecture by adding extra convolutional layers.

Bibliography

- [1] Inigo Quilez. Domain warping. <https://iquilezles.org/articles/warp/>. Accessed: 2023-12-11.
- [2] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and aging at the centenary of single particle tracking. *Physical Chemistry Chemical Physics*, 16:24128–24164, 2014.
- [3] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322:549–560, 1905.
- [4] M. Smoluchowski. Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der Physik*, 326:756–780, 1906.

- [5] R. Brown. A brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The Philosophical Magazine*, 4:161–173, 1828.
- [6] I. M. Sokolov and J. Klafter. From diffusion to anomalous diffusion: a century after Einstein's Brownian motion. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15:026103, 2005.
- [7] P. Langevin. Sur la théorie du mouvement brownien. *Comptes rendus de l'Académie des Sciences*, 146:530–533, 1908.
- [8] G. Muñoz-Gil and et al. andi_dataset - the anomalous diffusion library. https://github.com/AnDiChallenge/ANDI_datasets. doi:10.5281/zenodo.4775310.
- [9] MSD-based analysis. https://andichallenge.github.io/andi_datasets/lib_nbs/analysis.html#msd-based-analysis.
- [10] C. Manzo and M. F. Garcia-Parajo. A review of progress in single particle tracking: from methods to biophysical insights. *Reports on Progress in Physics*, 78:124601, 2015.
- [11] F. Höfling and T. Franosch. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76:046602, 2013.
- [12] A. V. Weigel, B. Simon, M. M. Tamkun, and D. Krapf. Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proceedings of the National Academy of Sciences*, 108:6438–6443, 2011.
- [13] C. Manzo, J. A. Torreno-Pina, P. Massignan, G. J. Lapeyre Jr, M. Lewenstein, and M. F. Garcia-Parajo. Weak ergodicity breaking of receptor motion in living cells stemming from random diffusivity. *Physical Review X*, 5:011021, 2015.
- [14] L. Stadler and M. Weiss. Non-equilibrium forces drive the anomalous diffusion of telomeres in the nucleus of mammalian cells. *New Journal of Physics*, 19:113048, 2017.
- [15] J.-P. Bouchaud. Weak ergodicity breaking and ageing in disordered systems. *Journal de Physique I*, 2:1705–1713, 1992.
- [16] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.
- [17] H. Scher and E. W. Montroll. Anomalous transit-time dispersion in amorphous solids. *Physical Review B*, 12:2455, 1975.
- [18] D. S. Martin, M. B. Forstner, and J. A. Käs. Apparent subdiffusion inherent to single particle tracking. *Biophysical Journal*, 83:2109–2117, 2002.
- [19] E. Kepten, A. Weron, G. Sikora, K. Burnecki, and Y. Garini. Guidelines for the fitting of anomalous diffusion mean square displacement graphs from single particle tracking experiments. *PLoS One*, 10:e0117722, 2015.
- [20] X. Michalet. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Physical Review E*, 82:041914, 2010.
- [21] A. Sabri, X. Xu, D. Krapf, and M. Weiss. Elucidating the origin of heterogeneous anomalous diffusion in the cytoplasm of mammalian cells. *Physical Review Letters*, 125:058101, 2020.
- [22] Y. He, S. Burov, R. Metzler, and E. Barkai. Random time-scale invariant diffusion and transport coefficients. *Physical Review Letters*, 101:058101, 2008.
- [23] M. Magdziarz and A. Weron. Anomalous diffusion: testing ergodicity breaking in experimental data. *Physical Review E*, 84:051138, 2011.

- [24] Y. Lanoiselée and D. S. Grebenkov. Revealing nonergodic dynamics in living cells from a single particle trajectory. *Physical Review E*, 93:052146, 2016.
- [25] N. Granik, L. E. Weiss, E. Nehme, M. Levin, M. Chein, E. Perlson, Y. Roichman, and Y. Shechtman. Single-particle diffusion characterization by deep learning. *Biophysical Journal*, 117:185–192, 2019.
- [26] F. Cichos, K. Gustavsson, B. Mehlig, and G. Volpe. Machine learning for active matter. *Nature Machine Intelligence*, 2:94–103, 2020.
- [27] G. Muñoz-Gil, M. A. Garcia-March, C. Manzo, J. D. Martín-Guerrero, and M. Lewenstein. Single trajectory characterization via machine learning. *New Journal of Physics*, 22:013010, 2020.
- [28] S. Bo, F. Schmidt, R. Eichhorn, and G. Volpe. Measurement of anomalous diffusion using recurrent neural networks. *Physical Review E*, 100:010102, 2019.
- [29] P. Kowalek, H. Loch-Olszewska, and J. Szwabiński. Classification of diffusion modes in single-particle tracking data: Feature-based versus deep-learning approach. *Physical Review E*, 100:032410, 2019.
- [30] S. Thapa, M. A. Lomholt, J. Krog, A. G. Cherstvy, and R. Metzler. Bayesian analysis of single-particle tracking data using the nested-sampling algorithm: maximum-likelihood model selection applied to stochastic-diffusivity data. *Physical Chemistry Chemical Physics*, 20:29018–29037, 2018.
- [31] D. Krapf, E. Marinari, R. Metzler, G. Oshanin, X. Xu, and A. Squarcini. Power spectral density of a single brownian trajectory: what one can and cannot learn from it. *New Journal of Physics*, 20:023029, 2018.
- [32] M. Magdziarz, A. Weron, K. Burnecki, and J. Klafter. Fractional Brownian motion versus the continuous-time random walk: A simple test for subdiffusive dynamics. *Physical Review Letters*, 103:180602, 2009.
- [33] K. Burnecki, E. Kepten, Y. Garini, G. Sikora, and A. Weron. Estimating the anomalous diffusion exponent for single particle tracking data with measurement errors—an alternative approach. *Scientific Reports*, 5:1–11, 2015.
- [34] A. G. Cherstvy, S. Thapa, C. E. Wagner, and R. Metzler. Non-Gaussian, non-ergodic, and non-Fickian diffusion of tracers in mucin hydrogels. *Soft Matter*, 15:2526–2551, 2019.
- [35] A. S. Serov, F. Laurent, C. Floderer, K. Perronet, C. Favard, D. Muriaux, N. Westbrook, C. L. Vestergaard, and J.-B. Masson. Statistical tests for force inference in heterogeneous environments. *Scientific Reports*, 10:1–12, 2020.
- [36] G. Muñoz-Gil and et al. Objective comparison of methods to decode anomalous diffusion. *Nature Communications*, 12:6253, 2021.
- [37] CNN for processing time series data. https://github.com/AnDiChallenge/andi_datasets/blob/master/source_nbs/tutorials/andi_and_ergodicity.ipynb.
- [38] F. Chollet. *Deep learning with Python*. Manning New York, 2018.

Chapter 7

Low Reynolds Number Hydrodynamics

HOLGER STARK

The realm of *low Reynolds number hydrodynamics*, where fluid inertia is negligible, unveils a world remarkably dissimilar from our intuitive understanding of fluid dynamics. Contrary to the ubiquity of turbulence in macroscopic flows, microscopic environments are domains of friction-dominated, *laminar fluid flows* (Fig. 7.1). This characteristic deeply impacts swimming microorganisms and processes such as fluid mixing, a non-trivial task on microscales yet pivotal for microfluidics and lab-on-a-chip applications. An equally significant concern is the understanding of colloidal suspensions, where moving particles create long-range flow fields influencing neighboring particle motion and the collective dynamics of microswimmers. This Chapter delves into the guiding principles of low Reynolds number hydrodynamics, empowering the reader to comprehend how fluid environments dictate microswimmer propulsion modes and the role of hydrodynamic flow fields in governing their collective behaviors. This narrative introduces key concepts like the *Stokes equations*, *kinematic reversibility*, and the *Oseen tensor*. Furthermore, it explores *friction coefficients* for translating/rotating spheres, particles of arbitrary shapes, and slender rods. Extensions of these principles to *resistive force theory* and *Faxén's theorem* precede a deep dive into *hydrodynamic interactions*, *multipole expansion*, and *Lorentz reciprocal theorem*. The discussion concludes with insights into the *Blake tensor* and *time-dependent Stokes equations*, encapsulating the essentials of low Reynolds number hydrodynamics. This Chapter is based on the classic text books on low-Reynolds number hydrodynamics by Happel and Brenner [1] as well as Kim and Karrila [2], the text book of Dhont on colloidal dynamics [3] and an own review on emergent behavior in active colloids [4]. Further references are given in the single sections.



Figure 7.1: **Transition from laminar to turbulent flow.** The smoke from the cigarette in the hand of Humphrey Bogart rises and accelerates due to buoyancy. It shows a transition from *laminar* to *turbulent flow*. In this Chapter, we concentrate on *low Reynolds numbers*, where flow is laminar. Image credit: Yousuf Karsh. Library and Archives Canada, PA-212506.

7.1 Stokes equations and low Reynolds numbers

The flow of a *viscous* or *Newtonian fluid* is described by the *Navier-Stokes equations* for the fluid velocity field $\mathbf{v}(\mathbf{r}, t)$,

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla p + \eta \nabla^2 \mathbf{v} + \rho \mathbf{b} \quad \text{and} \quad \operatorname{div} \mathbf{v} = 0, \quad (7.1)$$

with density ρ and shear viscosity η [Define p . I assume it's pressure?][Define \mathbf{b}]. The second equation means that the fluid is incompressible. The Navier-Stokes equations are Newton's equation of motion applied to a continuous medium, where forces per unit volume or force densities are balanced. In the inertial term on the left, $d\mathbf{v}/dt = \partial\mathbf{v}/\partial t + \mathbf{v} \cdot \nabla \mathbf{v}$ is the total time derivative, which gives the acceleration of a fluid volume moving with the fluid. Therefore, the convective derivative $\mathbf{v} \cdot \nabla \mathbf{v}$ has to be used in addition to the partial derivative $\partial\mathbf{v}/\partial t$. On the right-hand side, the force densities describe pressure forces and viscous shear forces, where the latter are due to adjacent fluid layers flowing along each other. These are intrinsic forces acting within the fluid, while the last term addresses volume or body forces applied from the outside (e.g., gravitational forces). Since the fluid is assumed to be incompressible, we left out viscous compressional forces, which are responsible for the damping of acoustic waves.

On the microscopic scale, inertial forces are typically negligible. This is quantified by the *Reynolds number*, which compares inertial to viscous forces, and is usually much smaller than one for microscopic systems:

$$\operatorname{Re} = \frac{\text{inertial force}}{\text{viscous drag}} = \frac{\rho v^2/a}{\eta v/a^2} = \frac{\rho va}{\eta} \ll 1. \quad (7.2)$$

In the third expression, we introduced a characteristic velocity v and length scale a , on which the flow field changes in space, in order to estimate the different forces in the Navier-Stokes equations. Strictly speaking, very small Reynolds number means that only the convective derivative or non-linear term, $\mathbf{v} \cdot \nabla \mathbf{v}$, can be neglected. For being able to also omit $\partial\mathbf{v}/\partial t$, one has to restrict oneself to times larger than the vorticity diffusion time as explained in Section 7.10, then the Navier-Stokes equations turn into the Stokes equations together with the incompressibility condition:

$$\mathbf{0} = -\nabla p + \eta \nabla^2 \mathbf{v} + \rho \mathbf{b} \quad \text{and} \quad \operatorname{div} \mathbf{v} = 0. \quad (7.3)$$

While the full Navier-Stokes equations describe turbulent flow due to the non-linear term, solutions of the Stokes equations give laminar flow. The transition between both flow types is governed by the Reynolds number and nicely illustrated in Fig. 7.1 by the smoke from the cigarette in the hand of Humphrey Bogart. Initially the smoke rises with a laminar flow, but when it further accelerates due to buoyancy it becomes turbulent.

The Navier-Stokes equations for incompressible fluids are governed by two parameters: the shear viscosity $\eta [\text{kg m}^{-1}\text{s}^{-1}]$ and density $\rho [\text{kg m}^{-3}]$, where in the square brackets we show the dimensions. From these parameters one cannot construct a characteristic length; therefore, the Navier-Stokes equations are valid on all length scales with a lower bound when atomic dimensions are reached. However, one can formulate an intrinsic force scale $F_{\text{crit}} = \eta^2/\rho$. For characteristic forces F_{char} in the fluid one can show the equivalence

$$\frac{F_{\text{char}}}{F_{\text{crit}}} \leq 1 \quad \longleftrightarrow \quad \operatorname{Re} \leq 1. \quad (7.4)$$

Example 7.1: Intrinsic force scale. Show that low Reynolds number hydrodynamics ($\operatorname{Re} < 1$) is equivalent to the case that the characteristic forces in the fluid are smaller than F_{crit} , i.e., show the equivalence of Eq. (7.4) by choosing appropriate characteristic forces F_{char} .

For water with $\eta = 1000 \text{ kg m}^{-3}$ and $\eta = 0.9 \cdot 10^{-3} \text{ kg m}^{-1}\text{s}^{-1}$ one has $F_{\text{crit}} = 0.8 \cdot 10^{-9} \text{ N} \approx 1 \text{ nN}$. Typical forces in the nano-/microworld are smaller than F_{crit} . For example, in a biological cell $F_{\text{char}} \approx 1 \text{ pN}$, meaning inertia is negligible.

The consequences of the negligible inertia can be illustrated by the following situation. Suppose a colloidal particle with radius $a = 1 \mu\text{m}$ is given an initial velocity of $u = 10 \mu\text{m/s}$. Then, only the Stokes friction force $6\pi\eta au$ (which we derive in Section 7.3.1) acts on the colloidal particle and slows it down to $u \approx 0$ within the time of $\approx 1 \mu\text{s}$ while the colloid only moves ≈ 0.1 [units?] (less than the radius of an atom). Thus, the colloidal particle hardly moves and stops *instantaneously*. This is the limit of *Aristotelian mechanics*, according to which a body only moves while a force is acting on it.

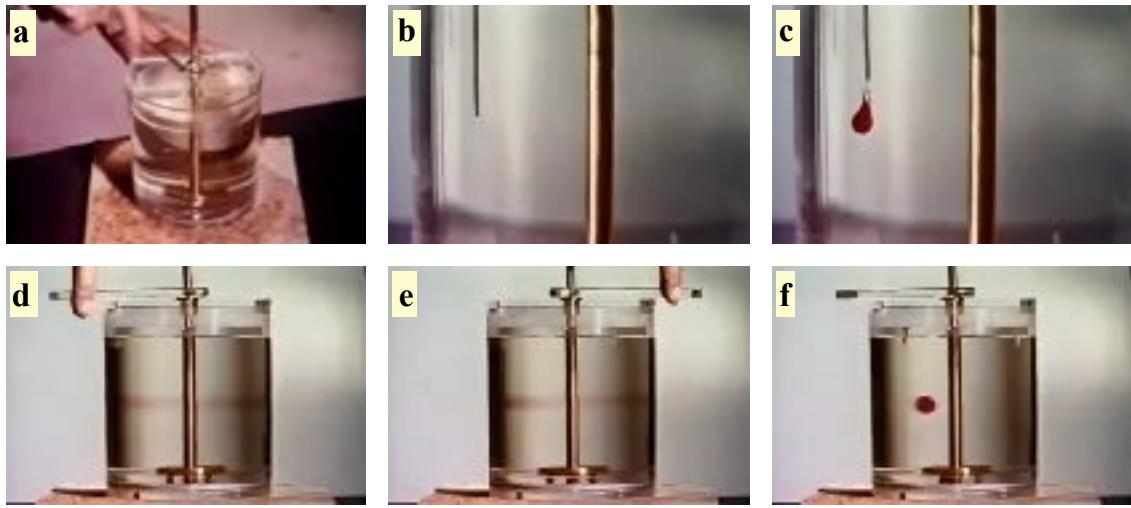


Figure 7.2: **Kinematic reversibility.** Experiment demonstrating the kinematic reversibility of the Stokes equations. (a) Taylor-Couette cell with a highly viscous fluid. (b) Injection needle. (c) The dye is injected. (d) The inner cylinder is turned four times: the blob of dye has been smeared out completely. (e) The rotation is reversed: the force acting on the fluid from the surface of the inner cylinder is reversed. (f) After rotating back the inner cylinder four times, the blob of dye is recovered. Reproduced from: G.I. Taylor: Low Reynolds number flows (1967), YouTube video under <https://www.youtube.com/watch?v=zeUzhy7Fgml>.

Example 7.2: Aristotelian mechanics. Verify the numbers obtained in the previous paragraph for the duration and displacement of the motion of the particle solving Newton's law.

Another very relevant property is the kinematic reversibility of the Stokes equations. The time-inverted flow field $-\mathbf{v}(\mathbf{r}, -t)$ is also a solution of the Stokes equations if pressure and external forces are reversed; $\nabla p \rightarrow -\nabla p$ and $\rho \mathbf{b} \rightarrow -\rho \mathbf{b}$. This is instructively demonstrated by an experiment popularized by G.I. Taylor in a video¹. He uses a Taylor-Couette cell (Fig. 7.2a) with a highly viscous fluid and locally stains the fluid with a dye (Fig. 7.2bc). After turning the inner cylinder four times, the blob of dye has been smeared out completely (Fig. 7.2d). Turning the inner cylinder in the opposite direction (Fig. 7.2e), which means reversing the force acting on the fluid from the surface of the inner cylinder, the blob of dye is recovered (Fig. 7.2f). It only does not look as sharp as before due to the very slow diffusion of the dye molecules. The kinematic reversibility is directly related to Purcell's scallop theorem, according to which a microswimmer has to perform a non-reciprocal shape deformation in order to move forward.

7.2 Oseen tensor

The Stokes equations are linear in the velocity and pressure field. Therefore, the method of *Green's functions* is applicable, which allows to calculate solutions for velocity and pressure fields for arbitrary volume forces $\rho \mathbf{b}$ according to the convolution integrals

$$\mathbf{v}(\mathbf{r}) = \int d^3 r' \mathbf{O}(\mathbf{r} - \mathbf{r}') \rho \mathbf{b}(\mathbf{r}') \quad \text{and} \quad p(\mathbf{r}) = \int d^3 r' \mathbf{g}(\mathbf{r} - \mathbf{r}') \cdot \rho \mathbf{b}(\mathbf{r}'). \quad (7.5)$$

¹G.I. Taylor: Low Reynolds number flows (1967), YouTube video under <https://www.youtube.com/watch?v=zeUzhy7Fgml>. The reported experiment starts at 12:55 min. See Fig. 7.2 for the relevant screenshots.

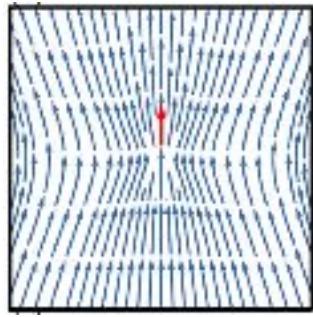


Figure 7.3: **Flow field of a Stokeslet.** The point force is indicated by a red arrow, while the generated flow field is indicated by the blue arrows.

Here, Green's functions for the Stokes equations are called *Oseen tensor* and *pressure vector*, respectively. They have the form

$$\mathbf{O}(\mathbf{r}) = \frac{1}{8\pi\eta r} (\mathbf{1} + \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) \propto \frac{1}{r} \quad \text{and} \quad \mathbf{g}(\mathbf{r}) = -\frac{1}{4\pi} \nabla \frac{1}{r} = \frac{1}{4\pi} \frac{\hat{\mathbf{r}}}{r^2}, \quad \hat{\mathbf{r}} = \frac{\mathbf{r}}{r}. \quad (7.6)$$

Note that \otimes means tensor product with $[\mathbf{a} \otimes \mathbf{b}]_{ij} = a_i b_j$. In the following we will always concentrate on the velocity fields.

The velocity field of a point force f with density $\rho \mathbf{b}(\mathbf{r}) = f \mathbf{e} \delta(\mathbf{r} - \mathbf{r}_0)$ is called *stokeslet* and is readily obtained from Eqs. (7.5) and (7.6) as

$$\mathbf{v}_s(\mathbf{r}) = \frac{f}{8\pi\eta|\mathbf{r} - \mathbf{r}_0|} [\mathbf{e} + (\hat{\mathbf{r}} \cdot \mathbf{e}) \hat{\mathbf{r}}] \quad \text{with} \quad \hat{\mathbf{r}} = \frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|}. \quad (7.7)$$

This shows that the velocity field due to an external force is long-ranged and decays like $1/r$. It is illustrated in Fig. 7.3.

Finally, using $\rho \mathbf{b}(\mathbf{r}) = \mathbf{1} \delta(\mathbf{r})$ in Eqs. (7.5), were $\mathbf{1}$ is the unit tensor, one recognizes that Green's functions obey the Stokes equations with a delta-shaped inhomogeneity, as usual,

$$\nabla \otimes \mathbf{g}(\mathbf{r}) - \eta \nabla^2 \mathbf{O}(\mathbf{r}) = \mathbf{1} \delta(\mathbf{r}), \quad \text{div } \mathbf{O}(\mathbf{r}) = \mathbf{0} \quad \text{and} \quad \nabla^2 \mathbf{g}(\mathbf{r}) = \nabla \delta(\mathbf{r}). \quad (7.8)$$

The third equation for the pressure vector follows by taking the divergence of the first equation and using $\text{div } \mathbf{O} = \mathbf{0}$. Knowing Green's function for the Poisson equation, one immediately verifies the form of $\mathbf{g}(\mathbf{r})$ in Eqs. (7.6). The derivation of the explicit form of the Oseen tensor is more evolved and can be done following, for example, the standard way of Fourier transformation.

Example 7.3: Derivation of the Oseen tensor. Derive the Oseen tensor from Eqs. (7.8) using the known expression for $\mathbf{g}(\mathbf{r})$ in Eqs. (7.6).

[Hints: Write and solve the first and second equations of Eqs. (7.8) in Fourier space also using the Fourier transform of $\mathbf{g}(\mathbf{r})$. From the Fourier transform $\mathbf{O}(\mathbf{k})$ determine $\mathbf{O}(\mathbf{r})$ in real space. The derivation can be found in Ref. [3].]

7.3 Friction coefficients

When bodies move through a fluid, they experience a friction force, which acts against their motion. In this Section, we derive the *Stokes friction coefficients* of a translating/rotating sphere, introduce the *friction tensor* of a particle with arbitrary shape, and exemplify it for a slender rod.

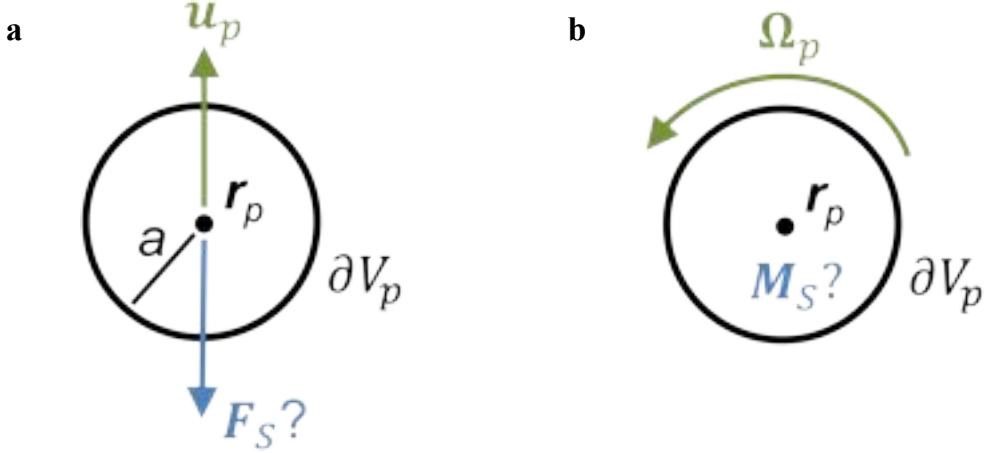


Figure 7.4: **Geometry for Stokes friction coefficients.** Geometry for Stokes friction coefficients for (a) translating and (b) rotating sphere of radius a and position \mathbf{r}_p .

7.3.1 Stokes friction

The friction force \mathbf{F}_S for a sphere translating with velocity \mathbf{u}_p and the friction torque \mathbf{M}_S for a sphere rotating with angular velocity Ω_p , as sketched in Fig. 7.4, are standard situations in colloidal physics. To determine them, the conventional method is to solve Stokes equations for the two situations and then calculate the stress tensor \mathbf{T} using the velocity and pressure fields. By integrating the stress tensor over the surface of the sphere, one ultimately arrives at \mathbf{F}_S and \mathbf{M}_S . We present here an alternative method using the Oseen tensor $\mathbf{O}(\mathbf{r} - \mathbf{r}')$ and illustrate the solutions for both situations.

Translating sphere

The core idea is to make an ansatz for the surface force density, with which the sphere acts on the fluid via its surface. Because the Stokes equations are linear, we will have $\mathbf{F}_S \propto \mathbf{u}_p$. Therefore, we choose the following force density on the sphere surface and insert it into Eq. (7.5),

$$\rho \mathbf{b}(\mathbf{r}') |_{\mathbf{r}' \in \partial V_p} = \frac{c}{4\pi a^2} \mathbf{u}_p \quad \longrightarrow \quad \mathbf{v}(\mathbf{r}) = \frac{c}{4\pi a^2} \oint_{\partial V_p} \mathbf{O}(\mathbf{r} - \mathbf{r}') d\mathbf{f}' \mathbf{u}_p, \quad (7.9)$$

where c is an unknown constant and ∂V_p is the surface of the particle with volume V_p . Note that we only integrate here over the surface and not a volume. To evaluate the integral over the Oseen tensor, we refer the surface point \mathbf{r}' to the center of the sphere \mathbf{r}_p , $\mathbf{r}' = \mathbf{r}_p + \Delta \mathbf{r}$, and perform a Taylor expansion in $\Delta \mathbf{r}$ around \mathbf{r}_p . The terms odd in $\Delta \mathbf{r}$ vanish by symmetry, while the even terms of order four and higher vanish since the Oseen tensor, as any solution of the Stokes equations, satisfies the biharmonic equation $\nabla^2 \nabla^2 \mathbf{O}(\mathbf{r}) = \mathbf{0}$. The second-order term in $\Delta \mathbf{r}$ can be calculated using symmetry arguments and once the velocity field is known, the missing constant $c = 6\pi\eta a$ is determined from the boundary condition $\mathbf{v}(\mathbf{r})|_{\mathbf{r} \in \partial V_p} = \mathbf{u}_p$. Thus, the flow field generated by a sphere when dragged through the fluid with velocity \mathbf{u}_p becomes

$$\mathbf{v}(\mathbf{r}) = \mathbf{S}(\mathbf{r} - \mathbf{r}_p) \mathbf{u}_p \quad (7.10)$$

with

$$\mathbf{S}(\mathbf{r}) = 6\pi\eta a \left(\mathbf{1} + \frac{1}{6} a^2 \nabla^2 \right) \mathbf{O}(\mathbf{r}) = \frac{3}{4} \frac{a}{r} \left(\mathbf{1} + \frac{\mathbf{r} \otimes \mathbf{r}}{r^2} \right) + \frac{1}{4} \left(\frac{a}{r} \right)^3 \left(\mathbf{1} - 3 \frac{\mathbf{r} \otimes \mathbf{r}}{r^2} \right). \quad (7.11)$$

To calculate the Stokes friction force on the sphere, we need to simply integrate $-\rho \mathbf{b}(\mathbf{r})$ over its surface and arrive at the final result,

$$\mathbf{F}_S = - \oint_{\partial V_p} \rho \mathbf{b}(\mathbf{r}) d\mathbf{f} \quad \Longrightarrow \quad \mathbf{F}_S = -6\pi\eta a \mathbf{u}_p. \quad (7.12)$$

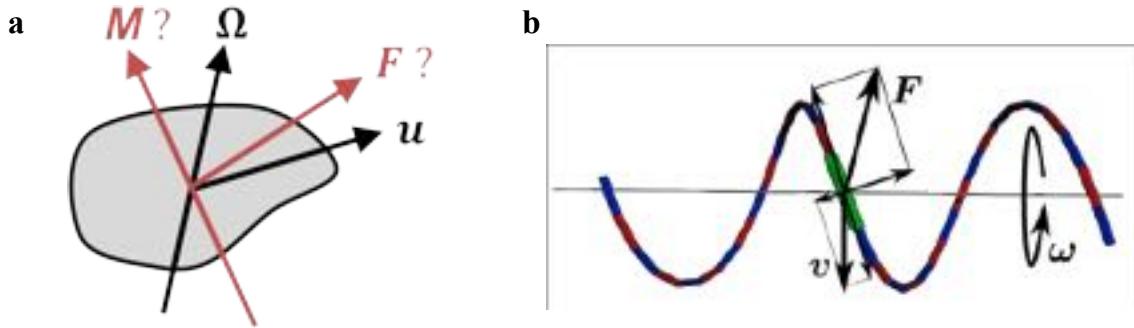


Figure 7.5: **Friction of non-spherical objects.** (a) A particle of arbitrary shape needs an external force \mathbf{F} and an external torque \mathbf{M} to translate with velocity \mathbf{u} and to rotate with angular velocity $\boldsymbol{\Omega}$, which are not necessarily parallel to \mathbf{F} and \mathbf{M} , respectively. (b) A filament with helical shape rotating about the helix axis generates a thrust force pushing the helix forward.

Rotating sphere

Considering that $\mathbf{M}_S \propto \boldsymbol{\Omega}_p$, the ansatz for the surface force density is

$$\rho \mathbf{b}(\mathbf{r}')|_{\mathbf{r}' \in \partial V_p} = \frac{c}{4\pi a^2} \boldsymbol{\Omega}_p \times (\mathbf{r}' - \mathbf{r}_p). \quad (7.13)$$

Again, using the convolution with the Oseen tensor according to the first formula of Eqs. (7.5) and performing the same steps as before, one finds $c = 12\pi\eta a$ and finally arrives at the vorticity flow field of a rotating sphere,

$$\mathbf{v}(\mathbf{r}) = \left(\frac{a}{|\mathbf{r} - \mathbf{r}_p|} \right)^3 \boldsymbol{\Omega}_p \times (\mathbf{r} - \mathbf{r}_p). \quad (7.14)$$

Integrating the surface torque generated by $\rho \mathbf{b}(\mathbf{r}')|_{\mathbf{r}' \in \partial V_p}$ over the whole surface gives the Stokes friction torque,

$$\mathbf{M}_S = - \int_{\partial V_p} (\mathbf{r} - \mathbf{r}_p) \times \rho \mathbf{b}(\mathbf{r}) d\mathbf{f} \implies \mathbf{M}_S = -8\pi\eta a^3 \boldsymbol{\Omega}_p. \quad (7.15)$$

Example 7.4: Stokes friction. Calculate the flow fields of a translating and rotating sphere following the steps outline above and determine the Stokes friction force/torque.

[Hints: (a) For the friction force perform the Taylor expansion of $\mathbf{O}(\mathbf{r} - (\mathbf{r}_p + \Delta\mathbf{r}))$ in $\Delta\mathbf{r}$, calculate the integral in Eq. (7.9), and explain why the terms odd in $\Delta\mathbf{r}$ vanish. The integral over the second-order term $\Delta\mathbf{r} \otimes \Delta\mathbf{r}$ is proportional to the unit tensor $\mathbf{1}$; why? Use this to calculate the integral and to argue together with $\nabla^2 \nabla^2 \mathbf{O}(\mathbf{r}) = \mathbf{0}$ that the even terms of order four and higher vanish. Verify that the unit tensor of the second-order term gives $\nabla^2 \mathbf{O}(\mathbf{r})$ to ultimately arrive at the explicit expression for $\mathbf{v}(\mathbf{r})$ in Eqs. (7.10) and (7.11), where the constant c follows from the boundary condition $\mathbf{v}(\mathbf{r})|_{\mathbf{r} \in \partial V_p} = \mathbf{u}_p$. Finally, the evaluation of \mathbf{F}_S in Eq. (7.12) is straightforward.

(b) The calculation of the friction torque follows similar steps. However, since the surface force density in Eq. (7.13) is already linear in $\Delta\mathbf{r}$, only the first-order term in the Taylor expansion of $\mathbf{O}(\mathbf{r} - (\mathbf{r}_p + \Delta\mathbf{r}))$ contributes to $\mathbf{v}(\mathbf{r})$. The constant c follows from the boundary condition $\mathbf{v}(\mathbf{r})|_{\mathbf{r} \in \partial V_p} = \boldsymbol{\Omega}_p \times (\mathbf{r} - \mathbf{r}_p)$. The evaluation of the integral in Eq. (7.15) is a bit more involved but also straightforward.]

7.3.2 Friction of a particle with arbitrary shape

Particles are not always spherical (Fig. 7.5a) and there are also complex objects such as the bundle of helical flagella with which a bacterium propels itself forward (Fig. 7.5b). Thus, an object translating with velocity \mathbf{u} can also have an external torque \mathbf{M} besides the force \mathbf{F} to keep the object moving (the same applies to the

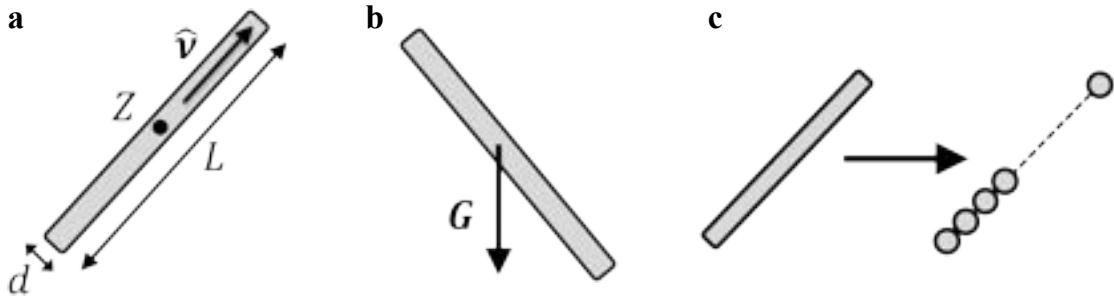


Figure 7.6: **Friction of a slender rod.** (a) A slender rod oriented along unit vector \hat{v} with length L , diameter d , and aspect ratio L/d . The reaction center Z is indicated. (b) A tilted rod sediments under the gravitational force \mathbf{G} . (c) To derive its friction coefficients in the limit $L/d \rightarrow \infty$, the rod is decomposed in touching spheres of diameter d . Each sphere is the source of a stokeslet with the total force acting on the rod divided equally on all spheres. The stokeslet acts with its velocity field on all other spheres, while the generating sphere has the friction coefficient $3\pi\eta d$ [3].

rotational velocity $\boldsymbol{\Omega}$). The relation between the external force/torque and the velocities should be linear and the most general form is

$$\begin{pmatrix} \mathbf{F} \\ \mathbf{M} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma} & \mathbf{C} \\ \mathbf{C}^t & \boldsymbol{\beta} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\Omega} \end{pmatrix}. \quad (7.16)$$

A few comments are in order. First, we have introduced friction tensors for translation ($\boldsymbol{\gamma}$) and rotation ($\boldsymbol{\beta}$) so that, in general, $\mathbf{F} \nparallel \mathbf{u}$ and $\mathbf{M} \nparallel \boldsymbol{\Omega}$. Second, the friction matrix as a whole is symmetric, which can be derived from the Lorentz reciprocal theorem (see Section 7.8). Therefore, also the friction tensors are symmetric, $\boldsymbol{\gamma} = \boldsymbol{\gamma}'$ and $\boldsymbol{\beta} = \boldsymbol{\beta}'$, where the superscript t means transposed tensor. Third, the coupling tensor \mathbf{C} couples translation to rotation, which then enables bacterial locomotion as we explain further below. It is a pseudo tensor meaning under spatial inversion it does change its sign. Thus, it can only exist for chiral objects, which do not coincide with their mirror picture. There exists a reaction center for which the coupling tensor becomes symmetric, $\mathbf{C} = \mathbf{C}^t$ [1].

7.3.3 Friction of a slender rod

We now consider a long slender rod, as represented in Fig. 7.6a. Since the rod is a non-chiral object, the coupling tensor \mathbf{C} vanishes relative to the reaction center Z , which is situated at the center of the rod. Then, the general linear relation of Eq. (7.16) reduces to $\mathbf{F} = \boldsymbol{\gamma}\mathbf{u}$ and $\mathbf{M} = \boldsymbol{\beta}\boldsymbol{\Omega}$.

The uniaxial symmetry of the rod determines its translational friction tensor, which can be written in coordinate-free representation and as a matrix with rod axis chosen along the z -axis, $\hat{v} = \hat{z}$,

$$\boldsymbol{\gamma} = \gamma_{\parallel} \hat{v} \otimes \hat{v} + \gamma_{\perp} (\mathbf{1} - \hat{v} \otimes \hat{v}) \implies \begin{pmatrix} \gamma_{\perp} & 0 & 0 \\ 0 & \gamma_{\perp} & 0 \\ 0 & 0 & \gamma_{\parallel} \end{pmatrix}, \quad (7.17)$$

where \otimes means the tensor product. Here, γ_{\parallel} and γ_{\perp} are the friction coefficients for moving the rod parallel and perpendicular to its axis, respectively. For large aspect ratios, $L \gg d$, they can be approximated as [5]

$$\gamma_{\perp} \approx \frac{4\pi\eta}{\ln(L/d)} L \approx 2\gamma_{\parallel}. \quad (7.18)$$

Note that for aspect ratios approaching infinity, $\gamma_{\perp} = 2\gamma_{\parallel}$.

Example 7.5: Sedimentation of tilted rod. As a simple application of the different friction coefficients of a slender rod, show that the tilted rod in Fig. 7.6b drifts to the right when sedimenting under gravity. [Hint: Start by decomposing the gravitational force \mathbf{G} into its components parallel and perpendicular to the rod.]

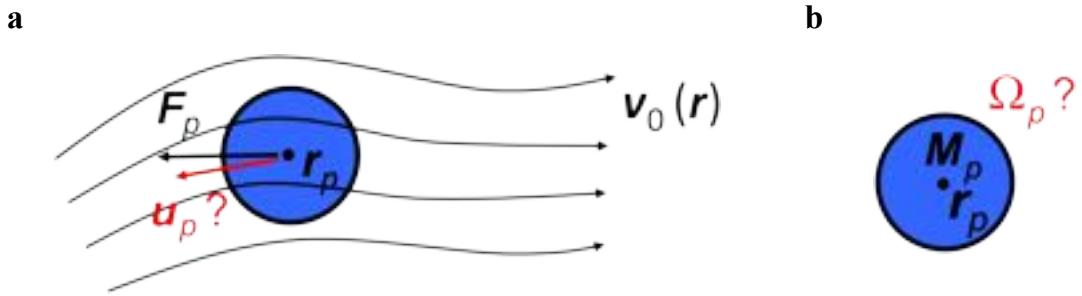


Figure 7.7: **Faxén theorem.** A particle is placed in a velocity field $\mathbf{v}_0(\mathbf{r})$, which is a solution of the Stokes equations. (a) When a force \mathbf{F}_p acts on the particle, the Faxén theorem in Eq. (7.21) provides its velocity \mathbf{u}_p . (b) When a torque \mathbf{M}_p acts on the particle, the Faxén theorem in Eq. (7.22) provides its angular velocity Ω_p .

For rods with finite aspect ratio, boundary effects become important and $\gamma_{\perp} \neq 2\gamma_{\parallel}$. This already becomes significant for large aspect ratios such as $L/D \approx 20$, where one finds $\gamma_{\perp} \approx 1.5\gamma_{\parallel}$. The friction coefficients in the limit of infinite aspect ratio, $L/D \rightarrow \infty$, can be calculated by decomposing the rod into touching spheres, which are sources of stokeslets as explained in the caption of Fig. 7.6c and in Ref. [3].

Finally, the rotational friction coefficient for rotating the rod about an axis perpendicular to the rod is

$$\mathbf{M} = \beta \boldsymbol{\Omega} \perp \hat{\mathbf{v}} \quad \text{with} \quad \beta = \frac{\pi \eta L^3}{3 \ln(L/d)}. \quad (7.19)$$

For a thin rod, the rotational friction coefficient for rotations about the rod is negligible against β .

7.4 Resistive force theory

The resistive force theory is a simplified approach to calculate the friction tensor of a thin slender body, which is not necessarily straight, but follows a specific curve in space with local tangent vector \mathbf{t} . The basic idea is to choose at each location of the slender body the friction tensor per length of a rigid rod with infinite aspect ratio. Then, this friction tensor is integrated along the whole space curve, which is parametrized in arc length s :

$$\boldsymbol{\zeta} = \zeta_{\parallel} \mathbf{t} \otimes \mathbf{t} + \zeta_{\perp} (\mathbf{1} - \mathbf{t} \otimes \mathbf{t}) \quad \text{with} \quad \zeta_{\perp} = 2\zeta_{\parallel} \quad \Rightarrow \quad \boldsymbol{\gamma} = \int \boldsymbol{\zeta} ds. \quad (7.20)$$

Note that $\boldsymbol{\gamma}$ can only be an approximation of the friction tensor since hydrodynamic interactions (see Section 7.6) between different parts of the slender body are neglected. Nevertheless, using resistive force theory one can calculate friction and coupling coefficients for a helical flagellum of a bacterium, which is sketched in Fig. 7.5b. For example, this was done in Ref. [6].

7.5 Faxén theorem

The Faxén theorem is a very deep consequence of the Stokes equations. It gives the velocity \mathbf{u}_p of a spherical particle when placed in a velocity field $\mathbf{v}_0(\mathbf{r})$ that solves the Stokes equations. In addition, an external force \mathbf{F}_p acts on the particle, as shown in Fig. 7.7a. The proof relies again on an expansion of the Oseen tensor, similarly to what we sketched in Section 7.3.1 [3]. Although the particle disturbs the initial flow field, the final result for its velocity \mathbf{u}_p is remarkably simple,

$$\mathbf{u}_p = \frac{1}{6\pi\eta a} \mathbf{F}_p + \left(\mathbf{1} + \frac{1}{6} a^2 \nabla^2 \right) \mathbf{v}_0(\mathbf{r}_p). \quad (7.21)$$

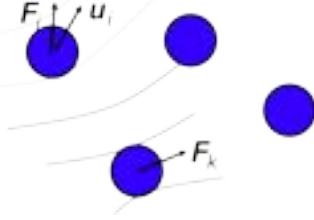


Figure 7.8: **Hydrodynamic interactions.** A colloidal particle i moves under the influence of an external force \mathbf{F}_i . However, its velocity \mathbf{u}_i also depends on the flow fields generated by other particles, which experience forces \mathbf{F}_k .

For $\mathbf{v}_0(\mathbf{r}) = \mathbf{0}$, one just recovers the Stokes friction law. Without an external force, i.e., with $\mathbf{F}_p = \mathbf{0}$, the particle drifts with the velocity of the unperturbed flow field at the position of the particle, $\mathbf{v}_0(\mathbf{r}_p)$, plus a correction from the curvature, $\nabla^2 \mathbf{v}_0(\mathbf{r}_p)$. In particular, one does not need to know how the particle disturbs the initial flow field. The Faxén theorem is also used to develop a perturbation theory for hydrodynamic interactions [3]. We will explain the initial step in the following Section 7.6.

There also exists the Faxén theorem for the angular velocity $\boldsymbol{\Omega}_p$ of the particle (Fig. 7.7b):

$$\boldsymbol{\Omega}_p = \frac{1}{8\pi\eta a^3} \mathbf{M}_p + \frac{1}{2} \nabla_p \times \mathbf{v}_0(\mathbf{r}_p). \quad (7.22)$$

Again, the final result is simple. The angular velocity is determined by the external torque \mathbf{M}_p divided by the rotational friction coefficient plus the vorticity of the unperturbed flow field at the particle position, which is the meaning of the second term on the right-hand side.

7.6 Hydrodynamic interactions

Suspensions of N micron-sized colloidal particles are omnipresent in daily life. When the particles move under the influence of external forces, they generate flow fields that influence the motion of neighboring particles. Since the Stokes equations are linear, the velocity \mathbf{u}_i of particle i should depend linearly on the forces acting on all the particles (Fig. 7.8):

$$\mathbf{u}_i = \sum_k \boldsymbol{\mu}_{ik} \mathbf{F}_k \quad \text{with} \quad \boldsymbol{\mu}_{ik} = \boldsymbol{\mu}_{ik}(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_k, \dots, \mathbf{r}_N), \quad (7.23)$$

where the *mobility tensors* $\boldsymbol{\mu}_{ik}$ are complicated functions of the particle positions \mathbf{r}_i that fully determine the dynamics of a colloidal suspension. For isolated particles, the *self mobilities* $\boldsymbol{\mu}_{ii}$ are the inverse of the Stokes friction coefficients, i.e., $\boldsymbol{\mu}_{ii} = (6\pi\eta a)^{-1} \mathbf{1}$. In general, the presence of other particles modifies $\boldsymbol{\mu}_{ii}$. For $i \neq k$, the $\boldsymbol{\mu}_{ik}$ are called *cross mobilities*. The mobility tensors obey the non-trivial symmetry conditions $\boldsymbol{\mu}_{ik} = \boldsymbol{\mu}_{ki}^t$, which follow from Lorentz reciprocal theorem (see Section 7.8) and can be viewed as Onsager relations. In the following, we consider spherical particles with the same radius a and do not apply any torque so that we do not need to consider the angular velocities of the particles, i.e., they are just freely rotating.

7.6.1 Point particles

There exist approximations for the mobility tensors, in particular, when they are not too close to each other. We start with the simplest approximation. In the limiting case that distances between the particles are much larger than the particle radius a , $r_{ik} = |\mathbf{r}_k - \mathbf{r}_i| \gg a$, we can consider them all as point particles. The self mobilities are not changed by the other particles, thus $\boldsymbol{\mu}_{ii} = (6\pi\eta a)^{-1} \mathbf{1}$. To determine the cross mobilities, we consider force \mathbf{F}_1 acting on particle 1. Neglecting higher orders, it just generates a stokeslet. A forceless particle 2 drifts with the stokeslet velocity of particle 1 at its position \mathbf{r}_2 , $\mathbf{u}_2 = \mathbf{v}(\mathbf{r}_2)$. So we can immediately identify the cross-mobility tensor:

$$\mathbf{u}_2 = \mathbf{v}(\mathbf{r}_2) = \mathbf{O}(\mathbf{r}_2 - \mathbf{r}_1) \mathbf{F}_1 \implies \boldsymbol{\mu}_{ik} = \mathbf{O}(\mathbf{r}_i - \mathbf{r}_k). \quad (7.24)$$

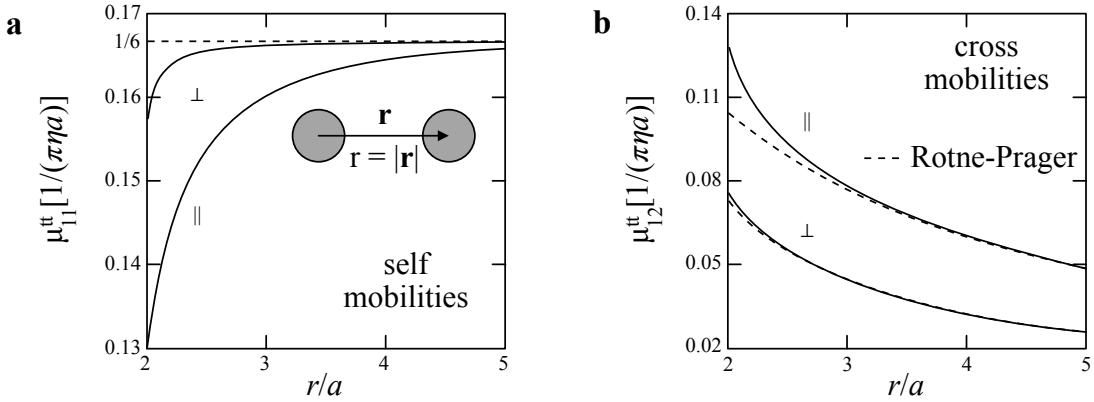


Figure 7.9: **Self and cross mobilities.** Translational (a) self and (b) cross mobilities for motion parallel and perpendicular to the line connecting particle 1 and 2 (inset in a). Solid lines: exact mobilities. Dashed lines: Rotne-Prager approximations.

Thus, in leading order of the inverse particle distances, the cross-mobility tensors are *Oseen tensors*. Therefore, hydrodynamic interactions are long-ranged since they decay algebraically with $\mu_{ik} \propto 1/r_{ik}$. Since we do not apply torques to the particles, they rotate freely and their angular velocities do not have to be considered explicitly.

7.6.2 Rotne-Prager approximation

When the particles come closer to each other, one also has to consider their extension, which in the next higher order provides the *Rotne-Prager approximation* for the mobility tensors.

In leading order, the velocity of particle k is given by $\mathbf{u}_k = \mu_0 \mathbf{F}_k = \mathbf{F}_k / (6\pi\eta a)$. So that with Eqs. (7.10) and (7.11), one can write the Stokes velocity field of particle k as

$$\mathbf{v}(\mathbf{r}) = \left(\mathbf{1} + \frac{1}{6} a^2 \nabla_k^2 \right) \mathbf{O}(\mathbf{r} - \mathbf{r}_k) \mathbf{F}_k. \quad (7.25)$$

Now, we place particle i in this velocity field and use the Faxén theorem to calculate its velocity:

$$\mathbf{u}_i = \mu_0 \mathbf{F}_i + \left(\mathbf{1} + \frac{1}{6} a^2 \nabla_i^2 \right) \left(\mathbf{1} + \frac{1}{6} a^2 \nabla_k^2 \right) \mathbf{O}(\mathbf{r}_i - \mathbf{r}_k) \mathbf{F}_k. \quad (7.26)$$

Thus, from this equation we can directly obtain the self and cross mobilities using $\mathbf{r}_{ik} = \mathbf{r}_k - \mathbf{r}_i$ as:

$$\boldsymbol{\mu}_{ii} = \mu_0 \mathbf{1} = (6\pi\eta a)^{-1} \mathbf{1} \quad (7.27)$$

$$\boldsymbol{\mu}_{ik} = \left(\mathbf{1} + \frac{1}{6} a^2 \nabla_i^2 \right) \left(\mathbf{1} + \frac{1}{6} a^2 \nabla_k^2 \right) \mathbf{O}(\mathbf{r}_i - \mathbf{r}_k) \quad (7.28)$$

$$= \mu_0 \left[\frac{3}{4} \frac{a}{r_{ik}} \left(\mathbf{1} + \frac{\mathbf{r}_{ik} \otimes \mathbf{r}_{ik}}{r_{ik}^2} \right) + \frac{1}{2} \left(\frac{a}{r_{ik}} \right)^3 \left(\mathbf{1} - 3 \frac{\mathbf{r} \otimes \mathbf{r}}{r_{ik}^2} \right) \right]. \quad (7.29)$$

The mobilities are expanded to third order in $1/r_{ik}$. While the first term in the cross mobilities is the Oseen tensor as found for point particles, the second term results from using the full Stokes velocity field of particle k and the curvature term of the Faxén theorem. The self mobility still coincides with the mobility of an isolated particle since within the Rotne-Prager approximation we do not consider how the Stokes flow field is modified due to the presence of other particles. Such effects of higher order can also be taken into account in a systematic way [3]. Thus, the Rotne-Prager approximation only includes two-particle interactions and does not take into account that the presence of a third particle l influences the interactions between particles i and k .

The self and cross mobilities can be decomposed into mobilities for moving along the difference vector \mathbf{r}_{ik} and perpendicular to it:

$$\boldsymbol{\mu}_{ik} = \mu_{ik}^{\parallel}(r_{ik}) \frac{\mathbf{r}_{ik} \otimes \mathbf{r}_{ik}}{r_{ik}^2} + \mu_{ik}^{\perp}(r_{ik}) \left(\mathbf{1} - \frac{\mathbf{r}_{ik} \otimes \mathbf{r}_{ik}}{r_{ik}^2} \right). \quad (7.30)$$

The mobilities for motion parallel and perpendicular to \mathbf{r}_{ik} in Rotne-Prager approximation are compared to exact values in Fig. 7.9. For particles distances $r_{ik} \geq 3a$, the deviations are below 3%.

The mobility tensors of Eq. (7.30) can be used to study the interesting consequences of hydrodynamic interactions. For example, two identical particles sedimenting next to each other under the influence of gravity sink faster than a single particle since they drag along each other due to the initiated flow fields. In particular, two particles behind each other sediment the fastest since they push and pull each other the strongest. This can qualitatively be understood from the stokeslet in Eq. (7.7). [Check this sentence, not clear] The initiated flow velocities at the positions of the neighboring particles for the two cases, sedimentation direction \mathbf{e} either perpendicular or parallel to the separation vector $\mathbf{r}_1 - \mathbf{r}_2$, differ by a factor of two. In Problem 7.2 the sedimentation velocities of two spheres sedimenting either next to each other at the same height or behind each other should be determined using the Rotne-Prager approximation.

Sedimenting clusters of three and more particles can even show oscillatory dynamics [7, 8] and transient chaotic motion [9]. Furthermore, particles circling in a toroidal trap move faster due to hydrodynamic interactions reminiscent but different in explanation to cyclists driving in the wind shadow [10]. Finally, circling particles perform caterpillar-like motion when *helping* each other to cross potential barriers [11].

7.6.3 Application to microswimmers

The method of mobilities can be extended to active particles that move in direction \mathbf{e} with a fixed speed v_0 . The velocity of particle i then becomes

$$\mathbf{u}_i - v_0 \mathbf{e}_i = \sum_k \boldsymbol{\mu}_{ik} \mathbf{F}_k + \sum_{k \neq i} \mathbf{v}_k^A(\mathbf{r}_i), \quad (7.31)$$

where $\mathbf{v}_k^A(\mathbf{r}_i)$ is the velocity field of particle k due to its active motion at the position of particle i [12]. Supplemented by equations for the rotational velocities $\boldsymbol{\omega}_i$, which also depend on the velocity fields of the other particles and contain the important Brownian noise, we have studied active Brownian particles in a three-dimensional harmonic trap potential [12]. Without any hydrodynamic interactions the particles get stuck at a horizon radius, where the radially outwards swimming particles are just stopped by inwards directed trap forces. When the hydrodynamic interactions are switched on and for increasing v_0 , the radially symmetric configuration becomes unstable and a symmetry breaking transition to a fluid pump is observed. The particles form a more tightly packed cluster, where they align their swimming directions and thus produce a macroscopic fluid flow.

7.7 Hydrodynamic singularities (multipoles)

Besides the stokeslet of Eq. (7.7), there are further hydrodynamic multipoles that are singular at their positions \mathbf{r}_0 . Taking the gradient of the determining equation of the Oseen tensor (Eq. (7.8)), one realizes that $\nabla \otimes \mathbf{O}(\mathbf{r})$ solves the Stokes equations with the derivative of the delta function as inhomogeneity, $\nabla \otimes \mathbf{1}\delta(\mathbf{r})$. Differently speaking, using derivates of $\delta(\mathbf{r})$ as force densities in Eq. (7.5), one generates further hydrodynamic multipoles. Below, we will introduce a further type of multipoles, which are generated from derivatives of hydrodynamic source fields.

7.7.1 Force dipoles, pushers, and pullers

With this in mind, we introduce the *force dipole* at position \mathbf{r}_0 , which consists of two point forces of strength f with a distance l , and pointing into opposite directions $\pm \mathbf{e}$, where \mathbf{e} is a unit vector, as shown in Fig. 7.10a. Thus, the corresponding force density is

$$\rho \mathbf{b}(\mathbf{r}) = f \mathbf{e} \delta(\mathbf{r} - \mathbf{r}_0 - (l/2)\mathbf{e}) - f \mathbf{e} \delta(\mathbf{r} - \mathbf{r}_0 + (l/2)\mathbf{e}) \quad (7.32)$$

$$\approx -f l \mathbf{e} \cdot \nabla \delta(\mathbf{r} - \mathbf{r}_0), \quad (7.33)$$

where in the second line we have used the Taylor expansion of both delta functions up to the first order. Using this force density in Eq. (7.5) or adding up the two stokeslets and expanding around $\mathbf{r} - \mathbf{r}_0$, we obtain the flow

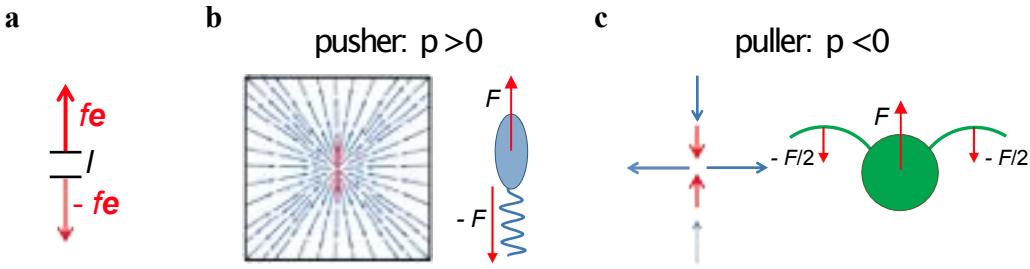


Figure 7.10: **Force-dipole microswimmers.** (a) Definition of a force dipole as combination of two point forces, which consists of two point forces of strength f with a distance l pointing into opposite directions $\pm\mathbf{e}$, where \mathbf{e} is a unit vector. (b) Flow field of a pusher, exemplified by the *Escherichia coli* bacterium. (c) Sketch of the flow field of a puller, exemplified by the *Chlamydomonas reinhardtii* alga.

field of a force dipole or *stresslet*,

$$\mathbf{v}_{fd}(\mathbf{r}) = \frac{p}{8\pi\eta r^2} [-1 + 3(\mathbf{e} \cdot \hat{\mathbf{r}})^2] \hat{\mathbf{r}} \quad \text{with} \quad r = |\mathbf{r} - \mathbf{r}_0| \quad \text{and} \quad \hat{\mathbf{r}} = \frac{\mathbf{r} - \mathbf{r}_0}{r}, \quad (7.34)$$

where $p = fl$ gives the strength of the force dipole.

Example 7.6: Stresslet. Calculate the flow field of the stresslet.

[Hint: Insert $\rho\mathbf{b}(\mathbf{r})$ from either Eq. (7.32) or (7.33) into Eq. (7.5), perform the integral, and expand around $\mathbf{r} - \mathbf{r}_0$.]

Microswimmers are force-free. So, the force dipole is the leading hydrodynamic multipole to describe the flow field of a microswimmer. It decays like $1/r^2$ and, depending on the sign of the dipole strength p , it distinguishes two swimmer types: *pushers* ($p > 0$), such as the bacterium *Escherichia coli*, push fluid out along their main axis and pull fluid in from the side (Fig. 7.10b), while *pullers* ($p < 0$), such as the alga *Chlamydomonas reinhardtii*, do the reverse (Fig. 7.10c).

7.7.2 Fundamental singular solutions

We now shortly summarize the two types of fundamental solutions of the Stokes equations.

Gradients of unit stokeslet

As explained before, the first type of solutions are created from gradients or directional derivatives of the unit stokeslet $\mathbf{v}_s(\mathbf{r}) = \mathbf{O}(\mathbf{r} - \mathbf{r}_0)\mathbf{e}$ [13]. The unit stresslet with strength $p = 1$ follows from the operation

$$\mathbf{v}_{fd}(\mathbf{r}) = \mathbf{e} \cdot \nabla_0 \mathbf{O}(\mathbf{r} - \mathbf{r}_0) \mathbf{e} \propto \frac{1}{r^2}, \quad (7.35)$$

which gives exactly the flow field of Eq. (7.7). One can also perform the directional derivative using a general displacement vector \mathbf{d} between the two point forces as illustrated in Fig. 7.11a:

$$\mathbf{v}_{fd \& rot}(\mathbf{r}) = \mathbf{d} \cdot \nabla_0 \mathbf{O}(\mathbf{r} - \mathbf{r}_0) \mathbf{e} \propto \frac{1}{r^2} \quad \text{with} \quad \mathbf{v}_{rot}(\mathbf{r}) = \boldsymbol{\Omega} \times \frac{\hat{\mathbf{r}}}{r^2}, \quad \boldsymbol{\Omega} = \mathbf{e} \times \mathbf{d}. \quad (7.36)$$

The result is a combination of a stresslet $\mathbf{v}_{fd}(\mathbf{r})$ (with an orientation vector different from \mathbf{e} and \mathbf{d}) and $\mathbf{v}_{rot}(\mathbf{r})$, which describes a rotlet or vortex field. It is determined by the component \mathbf{d} perpendicular to \mathbf{e} and thus creates a swirling flow. One can continue this scheme and calculate higher directional derivatives with further displacement vectors and thereby introduce higher multipoles such as force quadrupoles, force octupoles, and so forth.

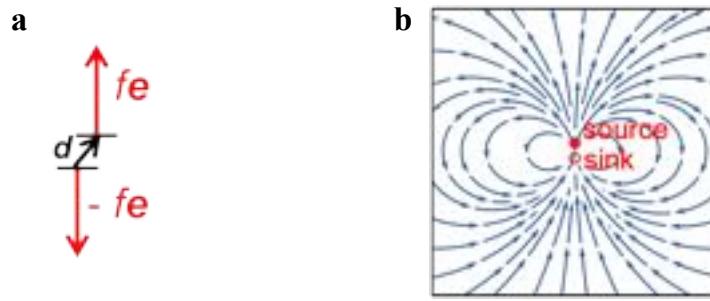


Figure 7.11: **General force dipole and source dipole.** (a) A general force dipole with $\mathbf{d} \nparallel \mathbf{e}$ contains a rotlet. (b) Two nearby potential flow fields of a sink and a source generate the flow field of a source dipole.

Potential flows from sources

Potential flows derive from a potential according to $\mathbf{v}_p = \nabla\phi$. They solve the Stokes equations for constant pressure since taking the gradient of $\nabla \cdot \mathbf{v}_p = \nabla^2\phi = 0$ gives $\nabla^2\mathbf{v}_p = \mathbf{0}$. Source and sink fields of a fluid are potential flows. The unit source monopole with a point source at \mathbf{r}_0 is given by

$$\phi(\mathbf{r}) = \frac{1}{r} \longrightarrow \mathbf{v}_{sm}(\mathbf{r}) = \frac{\mathbf{r} - \mathbf{r}_0}{|\mathbf{r} - \mathbf{r}_0|^3} \propto \frac{1}{r^2}. \quad (7.37)$$

Now, taking the directional derivative with displacement unit vector \mathbf{d} generates the flow field of a unit source dipole, i.e., the field of a source and sink close to each other,

$$\mathbf{v}_{sd}(\mathbf{r}) = \mathbf{d} \cdot \nabla_0 \mathbf{v}_{sm}(\mathbf{r}) \propto \frac{1}{r^3}, \quad (7.38)$$

which is illustrated in Fig. 7.11b. Flow fields of source dipoles are also relevant for microswimmers since they describe neutral swimmers, such as the *Volvox* alga, where the force dipole is absent. Higher source multipoles are generated by further directional derivatives of the source monopole, which introduce a combination of point sources and sinks, the flow field of which decays faster than the dipole.

Expansion into hydrodynamic multipoles

Any solution of the Stokes equations can be expanded into hydrodynamic force and source multipoles. To see this, we note that one can perform a Taylor expansion of the Oseen tensor $\mathbf{O}(\mathbf{r} - \mathbf{r}')$ in Eq. (7.5) around a reference point \mathbf{r}_0 . This generates derivatives of $\mathbf{O}(\mathbf{r} - \mathbf{r}_0)$, which can be pulled out of the integral and the remaining integral gives the different force multipoles in analogy to the multipole expansion in electrostatics. The analogy to a static electric field is even more complete for the source/sink fields that generate flow.

An example for an expansion into hydrodynamics multipoles is the model microswimmer called squirmer [14, 4]. This is a spherical particle with an intrinsic direction given by the unit vector \mathbf{e} . The squirmer has a tangential velocity field on its surface, which pushes the neighboring fluid along the surface and thereby the squirmer swims forward. The surface velocity field mimics, for example, the flow generated by beating cilia at the surface of a microswimmer such as the paramecium. Typically, one assumes an axial symmetry around \mathbf{e} , introduces the polar angle θ with respect to \mathbf{e} , and expands the tangential surface flow field $v_s \mathbf{e}_\theta$, where \mathbf{e}_θ is the local tangential unit vector along the longitudes, into Fourier modes $\sin n\theta$:

$$v_s = B_1 \sin \theta (1 + \beta \cos \theta) + \dots \quad (7.39)$$

The coefficients B_1 , $B_2 = B_1\beta$, etc. determine which hydrodynamic multipoles contribute to the flow field around the squirmer. We show here the most important ones. The coefficient B_1 belongs to a source dipole field and determines the swimming speed of the squirmer, as we show in Section 7.8. The coefficient β generates a force dipole field and distinguishes between pushers ($\beta < 0$) and pullers ($\beta > 0$).

7.8 Lorentz reciprocal theorem

Connected to the Stokes equation is the Lorentz reciprocal theorem, which helps to solve aspects of a flow problem using the known solution of an auxiliary flow problem [2]. To start, we introduce the stress tensor of a Newtonian fluid,

$$\boldsymbol{\sigma} = -p\mathbf{1} + 2\eta\mathbf{A} \quad \text{with} \quad \mathbf{A} = \frac{1}{2} [\nabla \otimes \mathbf{v} + (\nabla \otimes \mathbf{v})^t], \quad (7.40)$$

where \mathbf{A} is the strain-rate tensor and $(\dots)^t$ means the transposed tensor. Given a fluid volume with a local outward normal \mathbf{n} on the surface, $\boldsymbol{\sigma}\mathbf{n}$ gives the surface force density acting locally on the fluid surface. Using $\boldsymbol{\sigma}$, the Stokes equations can be rewritten as

$$-\nabla p + \eta\nabla^2\mathbf{v} = \rho\mathbf{b} \quad \Rightarrow \quad \operatorname{div}\boldsymbol{\sigma} = \nabla\boldsymbol{\sigma} = \rho\mathbf{b}. \quad (7.41)$$

Now, consider a fluid volume V with surface ∂V , where due to different settings two different solutions, \mathbf{v}_1 and \mathbf{v}_2 , of the Stokes equations exist with respective stress tensors $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$. Using the definition of the stress tensor, Eq. (7.40), one can show that $\boldsymbol{\sigma}_1 \cdot \mathbf{A}_2 = \boldsymbol{\sigma}_2 \cdot \mathbf{A}_1$, where we use the notation $\boldsymbol{\sigma}_1 \cdot \mathbf{A}_2 = [\boldsymbol{\sigma}_1]_{ij}[\mathbf{A}_2]_{ij}$. Integrating over the fluid volume and using Gauss's theorem, one then arrives at the Lorentz reciprocal theorem:

$$\oint_{\partial V} \mathbf{v}_1 \cdot \boldsymbol{\sigma}_2 \mathbf{n} d^2 r - \int_V \mathbf{v}_1 \cdot \nabla \boldsymbol{\sigma}_2 d^3 r = \oint_{\partial V} \mathbf{v}_2 \cdot \boldsymbol{\sigma}_1 \mathbf{n} d^2 r - \int_V \mathbf{v}_2 \cdot \nabla \boldsymbol{\sigma}_1 d^3 r. \quad (7.42)$$

It connects the two flow problems $\mathbf{v}_1, \boldsymbol{\sigma}_1$ and $\mathbf{v}_2, \boldsymbol{\sigma}_2$ to each other.

Example 7.7: Lorentz reciprocal theorem. Perform the derivation of the theorem.

[*Hints: Using the stress tensor of Eq. (7.40) and $\mathbf{1} \cdot \mathbf{A} = \delta_{ij}A_{ij} = 0$ (why?), show $\boldsymbol{\sigma}_1 \cdot \mathbf{A}_2 = \boldsymbol{\sigma}_2 \cdot \mathbf{A}_1$. Integrate over the fluid volume V , use the product rule to let the nabla operator ∇ in the definition of \mathbf{A} act on $\boldsymbol{\sigma}$, and apply Gauss's theorem.]*

For zero volume forces in V , $\rho\mathbf{b}_1 = \rho\mathbf{b}_2 = \mathbf{0}$, one has $\nabla\boldsymbol{\sigma}_1 = \nabla\boldsymbol{\sigma}_2 = \mathbf{0}$ and the theorem becomes

$$\oint_{\partial V} \mathbf{v}_1 \cdot \boldsymbol{\sigma}_2 \mathbf{n} d^2 r = \oint_{\partial V} \mathbf{v}_2 \cdot \boldsymbol{\sigma}_1 \mathbf{n} d^2 r. \quad (7.43)$$

We use the theorem to calculate the swimming velocity of a squirmer, which we introduced with its tangential surface velocity field in Eq. (7.39). Taking into account the directed swimming of the squirmer with velocity $v_0\mathbf{e}$, the surface velocity field in the laboratory frame becomes $\mathbf{v}_1|_{\partial V} = v_0\mathbf{e} + v_s\mathbf{e}_\theta$. Now, we introduce the auxiliary problem, where a passive particle is pulled with the velocity $v_0\mathbf{e}$ through the fluid. Thus, we have

$$\mathbf{v}_2|_{\partial V} = v_0\mathbf{e} \quad \text{and} \quad \boldsymbol{\sigma}_2 \mathbf{n} = \frac{6\pi\eta a}{4\pi a^2} \mathbf{u}_p \quad \text{with} \quad \mathbf{u}_p = v_0\mathbf{e}, \quad (7.44)$$

where the middle equation is the surface force density introduced in Eq. (7.9). Since the squirmer is force-free, we have

$$\oint_{\partial V} \boldsymbol{\sigma}_1 \mathbf{n} d^2 r = \mathbf{0} \quad \Rightarrow \quad \oint_{\partial V} \mathbf{v}_1 \cdot \boldsymbol{\sigma}_2 \mathbf{n} d^2 r = \mathbf{0} \quad \Rightarrow \quad \oint_{\partial V} v_0\mathbf{e} \cdot \boldsymbol{\sigma}_2 \mathbf{n} d^2 r = - \oint_{\partial V} v_s\mathbf{e}_\theta \cdot \boldsymbol{\sigma}_2 \mathbf{n} d^2 r. \quad (7.45)$$

To arrive at the second equation, we used the auxiliary problem defined in Eq. (7.44) and the Lorentz reciprocal theorem of Eq. (7.43). To arrive at the third equation, we inserted the surface velocity \mathbf{v}_1 in the laboratory frame. Using $\boldsymbol{\sigma}_2 \mathbf{n}$ of the auxiliary problem from Eq. (7.44) and the surface velocity v_s of a squirmer from Eq. (7.39), we are able to calculate the swimming velocity as

$$v_0 = \frac{2}{3}B_1. \quad (7.46)$$

It is only determined by the source dipole.

Example 7.8: Swimming velocity of a squirmer. Follow the outlined steps to determine v_0 .

[Hint: Insert $\sigma_2 \mathbf{n}$ from Eq. (7.44) and v_s from Eq. (7.39) into the third expression of Eqs. (7.45) and calculate the integrals.]

7.9 Blake tensor

Often one has bounding walls and the question arises of what Green's function should be used for a special geometry. Blake in his famous paper determined Green's function for a no-slip planar wall using the method of images [15]. The two important cases are illustrated in Fig. 7.12. The flow field for a point force parallel to the bounding wall (illustrated in Fig. 7.12a) is constructed by adding an image system to the bulk Stokeslet, which consists of the image Stokeslet with the force pointing in opposite direction, a force dipole consisting of a stresslet and a rotlet, and a source dipole. This is nicely illustrated in Ref. [16]. The flow field for a point force perpendicular to the bounding wall (illustrated in Fig. 7.12b) contains an image system with the image Stokeslet, a pure stresslet, a source dipole and a force quadrupole.

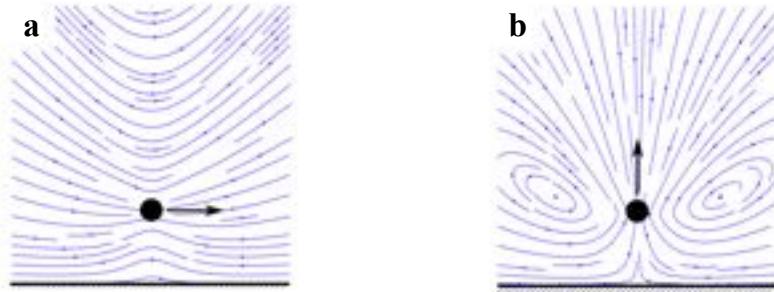


Figure 7.12: **Blake tensor.** Flow fields of a point force close to a bounding wall oriented (a) parallel or (b) perpendicular to the wall.

7.10 Time-dependent Stokes equations

Finally, we introduce the time-dependent Stokes equations neglecting the non-linear term or convective derivative:

$$\frac{\partial \mathbf{v}}{\partial t} = -\frac{1}{\rho} \nabla p + \frac{\eta}{\rho} \nabla^2 \mathbf{v} + \mathbf{b}. \quad (7.47)$$

Taking the curl of this equation, we obtain a diffusion equation for the vorticity $\boldsymbol{\omega} = \text{curl } \mathbf{v}/2$:

$$\left(\frac{\partial}{\partial t} - \frac{\eta}{\rho} \nabla^2 \right) \boldsymbol{\omega} = \text{curl } \mathbf{b}/2. \quad (7.48)$$

The vorticity of any disturbance in the fluid spreads with the diffusion constant η/ρ . For a characteristic length a the diffusion time becomes $\tau_H = a^2/(6\eta/\rho)$, which is the time the disturbance needs to diffuse the length a . This helps us to make a statement about the validity of the time-independent Stokes equations. They are valid for times $t \gg \tau_H$, which assures that the flow field has developed a stationary state on the length scale a .

We also shortly introduce Stokes second problem as sketched in Fig. 7.13. A no-slip wall oscillates with frequency ω and one wants to know the velocity field in the lower half-space. Due to the setting it makes sense to assume the velocity vector parallel to the wall and only dependent on z . Since no pressure gradient is applied, we set pressure constant and from Eq. (7.47) arrive at

$$p = \text{constant} \implies \left(\frac{\partial}{\partial t} - \frac{\eta}{\rho} \frac{\partial^2}{\partial z^2} \right) v(z, t) = 0. \quad (7.49)$$

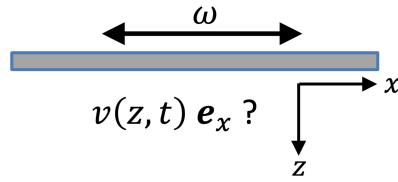


Figure 7.13: **Stokes second problem.** A no-slip wall oscillates with frequency ω . The generated oscillating flow field is described in Eq. (7.50). It is parallel to the wall and penetrates with the depth δ into the lower half-space.

This can be solved by the plane-wave ansatz,

$$v(z, t) = v(\omega) \exp^{i\omega t} \exp^{-(1+i)z/\delta} \quad (7.50)$$

and the penetration depth amounts to

$$\delta = \sqrt{\frac{2\eta}{\omega\rho}}. \quad (7.51)$$

Thus, the imposed oscillation penetrates into the fluid on the length δ . It diverges for $\omega \rightarrow 0$ and becomes zero for $\omega \rightarrow \infty$. The length δ also determines the wavelength $\lambda = 2\pi\delta$ on which the flow field is modulated.

7.11 Problems

Problem 7.1: Self-propulsion of an *E. coli* bacterium. An *E. coli* propels itself by rotating a bundle of helical flagella, which we approximate here as single helical filament.

a. Use resistive force theory to calculate the friction and coupling coefficients in Eq. (7.16) for a helix with pitch p and radius R , when the forces/torques and velocities are aligned with the helical axis.

[Hint: To determine the single coefficients, consider a translating and rotating helix.]

b. The helical flagella in the bundle are driven by rotary motors anchored in the wall of the cell body. Let a spherical cell body act with a torque on the the single helix along the helical axis and determine the velocity with which the whole object moves. Compare with typical values of an *E. coli* bacterium.

Problem 7.2: Two particles sedimenting under gravity. Use the Rotne-Prager approximation to calculate the sedimentation velocities of two spheres sedimenting either next to each other at the same height or behind each other. Why do they keep their distance?

Problem 7.3: Fundamental singular solutions.

a. Calculate the stresslet and rotlet of Eq. (7.36) for a force dipole with orientation unit vector \mathbf{e} and displacement vector \mathbf{d} .

b. Combine four forces such that stokeslet and stresslet as well as any rotlets are zero and determine the flow field of a force quadrupole.

Bibliography

- [1] J. Happel and H. Brenner. *Low-Reynolds Number Hydrodynamics*. Kluwer, The Hague, 1983.
- [2] S. Kim and S. J. Karrila. *Microhydrodynamics: Principles and Selected Applications*. Dover Publications, 2013.
- [3] J.K.G. Dhont. *An Introduction to Dynamics of Colloids*. Elsevier, Amsterdam, 1996.

- [4] A. Zöttl and H. Stark. Topical review: Emergent behavior in active colloids. *Journal of Physics: Condensed Matter*, 28:253001, 2016.
- [5] E. Lauga and T. R. Powers. The hydrodynamics of swimming microorganisms. *Reports on Progress in Physics*, 72:096601, 2009.
- [6] R. Vogel and H. Stark. Motor-driven bacterial flagella and buckling instabilities. *The European Physical Journal E: Soft Matter and Biological Physics*, 35:15, 2012.
- [7] R. E. Caflisch, C. Lim, J. H. C. Luke, and A. S. Sangani. Periodic solutions for three sedimenting spheres. *Physics of Fluids*, 31:3175, 1988.
- [8] I. K. Snook, K. M. Briggs, and E. R. Smith. Hydrodynamic interactions and some new periodic structures in three particle sediments. *Physica A*, 240:547–559, 1997.
- [9] I. M. Jánosi, T. Tél, D. E. Wolf, and J. A. C. Gallas. Chaotic particle dynamics in viscous flows: The three-particle stokeslet problem. *Physical Review E*, 56:2858, 1997.
- [10] M. Reichert and H. Stark. Circling particles and drafting in optical vortices. *Journal of Physics: Condensed Matter*, 16:S4085–S4094, 2004.
- [11] C. Lutz, M. Reichert, H. Stark, and C. Bechinger. Surmounting barriers: The benefit of hydrodynamic interactions. *Europhysics Letters*, 74:719–725, 2006.
- [12] M. Hennes, K. Wolff, and H. Stark. Self-induced polar order of active Brownian particles in a harmonic trap. *Physical Review Letters*, 112:238204, 2014.
- [13] A. T. Chwang and T. Y.-T. Wu. Hydromechanics of low-Reynolds-number flow. Part 2. Singularity method for Stokes flows. *Journal of Fluid Mechanics*, 67:787, 1975.
- [14] J. R. Blake. A spherical envelope approach to ciliary propulsion. *Journal of Fluid Mechanics*, 46:199–208, 1971.
- [15] J. R. Blake. A note on the image system for a Stokeslet in a no-slip boundary. *Mathematical Proceedings of the Cambridge Philosophical Society*, 70:303–310, 1971.
- [16] J. R. Blake and A. T. Chwang. Fundamental singularities of viscous flow. Part I: The image systems in the vicinity of a stationary no-slip boundary. *Journal of Engineering Mathematics*, 8:23–29, 1974.

Chapter 8

Phoretic and Self-phoretic Transport in Colloidal Science

IVO BUTTINONI

Anybody working with colloids soon comes to term with the fact that directed motion is only achieved by continuously applying a net force. The force either acts directly on the particles or is used to impose a flow field in the sample cell. *Phoretic transport* and *osmotic transport* appear to defy this rule: Directed motion stems from the fact that solid–liquid boundaries, such as particle–liquid interfaces, are layers of finite thickness where the solvent molecules are coupled (e.g., electrostatically) to the solid surface. If these interactions are not homogeneous around freely suspended particles, effective slip velocities boost their net transport properties. Local asymmetries can be also generated by the colloids themselves, provided that their surface is equipped with reactive patches that continuously exchange energy with the surrounding fluid. The discovery of these *self-phoretic microswimmers* (a.k.a. *autophoretic microswimmers*) dates back to the beginning of the century and paved the way for new colloidal materials with self-sustaining properties. This Chapter introduces the main features of interfacial transport as a tool to engineer self-propelling colloidal particles. First, it will showcase the importance of interfacial layers and slip boundary conditions by describing the simple problem of fluid flow in small pipes. Then, it will describe various mechanisms of interfacial transport: *electrophoresis*, *diffusiophoresis*, and *thermophoresis*, where the solid surface is free to move, as in the case of particles; and *electroosmosis*, *diffusioosmosis*, and *thermoosmosis*, where the solid surface is fixed, as in the case of walls. Finally, it will discuss how the same transport is achieved in an autonomous fashion by presenting several examples of self-phoretic microswimmers.

8.1 Fluid transport at small length scales

Let us begin our journey through colloid transport with a very simple experiment (involving no colloidal particles at all). Let us consider the laminar flow of an ionic solution (e.g., tap water) in a long cylindrical pipe (Fig. 8.2a). If we apply a pressure difference ΔP and assume no slip at the pipe's walls, the flow rate Q_{HP} is given by the *Hagen-Poiseuille equation*,

$$Q_{\text{HP}} = \frac{\pi R^4}{8\eta L} \Delta P, \quad (8.1)$$

where R and L are, respectively, the pipe's radius and length, and η is the fluid viscosity. The hydrodynamic resistance scales as R^{-4} , suggesting that pumping fluids through micro- and nanochannels becomes quickly a hard task. This problem affects not only lab-on-a-chip experiments, but also the permeability of porous biological and synthetic materials [3, 4, 5, 6]. However, when we compare experimental measurements of flow

Figure 8.1: **Introductory Figure Title.** Introductory Figure Caption add figure + refer to it in the intro paragraph

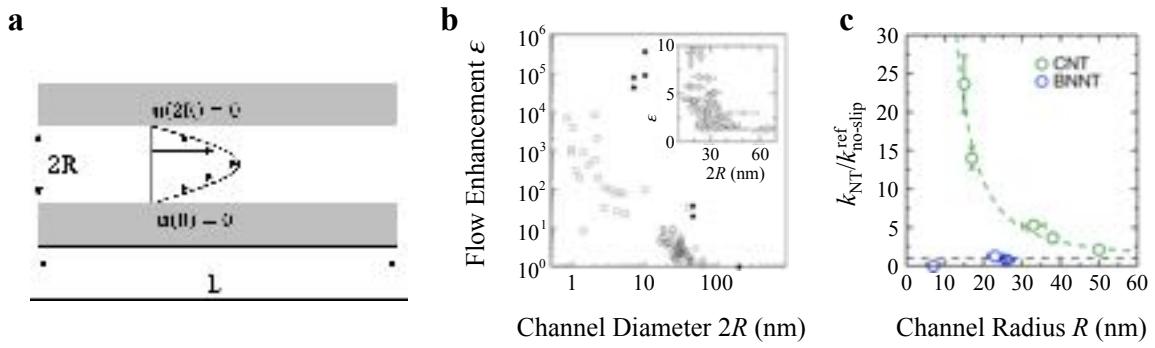


Figure 8.2: **Flow in a long cylindrical pipe.** (a) Pipe flow schematics. (b) Flow enhancement $\epsilon = Q_{\text{measured}}/Q_{\text{HP}}$ in nanotubes of different diameters under various experimental conditions. The inset represents a zoom of the corresponding region in the plot. Reproduced from Ref. [1]. (c) Permeability of carbon nanotubes (CNT) and boron nitride nanotubes (CNNT) as a function of their radius. Reproduced from Ref. [2].

rates in nanoconfinement with predictions from Eq. (8.1) (or its geometrical corrections due to, for instance, the presence of channel mouths), we find flow enhancements, $\epsilon = Q_{\text{measured}}/Q_{\text{HP}}$ (Fig. 8.2b), of up to $\approx 10^5$ [7, 1], meaning that small pipes can actually boost fluid transport. The plot in Fig. 8.2b also highlights massive deviations depending on the experimental conditions, such as the properties of the channel's walls. Discrepancies exist, for example, when we compare the permeability of carbon (CNT) and boron nitride nanotubes of the same size (see Fig. 8.2c) [2]. Where is the catch?

To understand the physics of fluid transport in small confinement, one must carefully appreciate what a solid–liquid interface is. According to the Oxford dictionary an interface is “a common boundary between two portions of matter or space”. Even though these boundaries are unfavorable environments that try to be as thin as possible, solid objects including particles and walls are separated from any surrounding liquid by a layer of small, but finite, thickness where the atoms and molecules interact anisotropically with the two phases. Fig. 8.3a shows that, although a continuum hydrodynamic description is valid down to approximately 1 nm (the lower bound for the definition of fluid viscosity), interfacial interactions become paramount to describe fluid transport in the sub-micrometer range. At these length scales, assuming a vanishing fluid velocity (i.e., no slip conditions) at the solid–liquid interface is no longer appropriate.

Unfortunately, the thickness of a solid–fluid boundary is not uniquely defined, because interfacial interactions are due to a large number of physical phenomena, including adsorption, electrostatic forces, friction, and entropy, just to mention but a few [7]. Let us first introduce the *dynamic slip length*, which is only determined by the motion of the fluid near a solid substrate, i.e., solid–liquid friction. Within linear response theory, the force balance for a fluid flowing in the x-direction with slip velocity \mathbf{u}_S (Fig. 8.3b) is $\sigma_{xz} = \mu \mathbf{u}_S$, where σ and μ are the stress tensor and the friction coefficient, respectively [8]. Since for Newtonian fluids the stress is also $\sigma_{xz} = \eta \partial \mathbf{u}_S / \partial z$, the slip boundary condition can be written as

$$\mathbf{u}_S = b \left[\frac{\partial \mathbf{u}_S}{\partial z} \right]_{\text{wall}} \quad (8.2)$$

where $b = \eta/\mu$ is the dynamic slip length, which can be interpreted as a virtual depth inside the solid at which the linearly extrapolated fluid velocity profile vanishes (Fig. 8.3b). If the fluid does not exert any stress on the substrates, i.e., the two phases do not interact, \mathbf{u}_S and b are zero. While the characterization of b goes beyond the scope of this Chapter (further details can be found in Ref. [7]), it is important to emphasize that b is given by the magnitude and range of typical fluid–solid molecular energies. In particular, the slip lengths of hydrophobic surfaces wetted by water can be as large as 20 nm [9], leading to pronounced flow enhancements in sub-micrometric confinements including nanotubes and porous media.

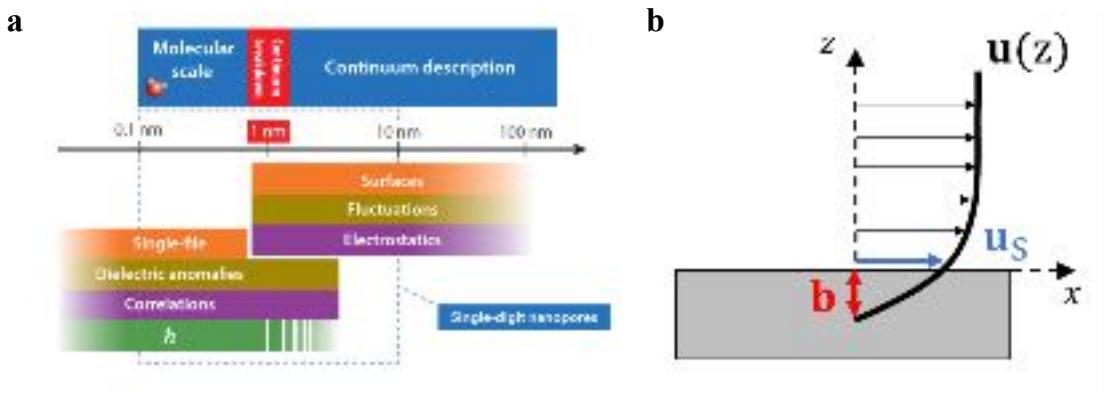


Figure 8.3: **Fluid transport in small confinement.** (a) Critical length scales appearing when we downsize a channel. Reproduced from Ref. [8]. (b) Sketch showing the definition of the dynamic slip length, b . $\mathbf{u}(z)$ is the fluid velocity in bulk and \mathbf{u}_S the slip velocity near the substrate. [Check last sentence is correct]

8.2 Interfacial transport

We now describe three different types of transport which are often realized in colloidal systems, namely transport in an electric field, in a chemical gradient, and in a temperature gradient.

8.2.1 Transport in electric fields

A charged particle suspended in an aqueous solution undergoes directed motion under an external electric field. This is surprising because the particle's charge is neutralized by a cloud of counterions. To understand the mechanism of transport, we consider a flat charged substrate immersed in a solution of monovalent ions. No external flow is applied. Even though the bulk liquid remains electrically neutral, the equilibrium balance between electrostatic interactions and thermal agitation generates a charge density profile adjacent to the solid boundary. For example, a common glass slide wetted by tap water ionizes to produce SiO^- surface groups so that positive counterions accumulate in its vicinity. This layer of fluid where electro-neutrality is lost is now our *interface* and the corresponding slip velocity is defined at its outer limit.

The electrostatic coupling between the fluid and the charged substrate includes a rich library of length scales (all linked to each other), the most important of which is the *Debye screening length*,

$$\lambda_D = \sqrt{\frac{\epsilon\epsilon_0 k_B T}{2e^2 c}}, \quad (8.3)$$

where $k_B T$ is the thermal energy, ϵ_0 and ϵ are the vacuum and relative permittivity, e is the electron charge, and c is the bulk number density of ions. The value of λ_D ranges between Ångströms and tens of nanometers, and is derived in almost any book of soft condensed matter (see, for example, Ref. [10]). In a first approximation, λ_D is the characteristic decay length of the electrostatic potential $\phi_e(z)$, i.e., $\phi_e(z) = \phi_e(0) \exp(-z/\lambda_D)$, and thus reveals the range of the diffuse cloud of counterions.

Let us now hold the substrate and apply an electric field \mathbf{E} in the x -direction as shown in Fig. 8.4a. \mathbf{E} produces a body force on the fluid equal to $(c_+ - c_-)\mathbf{E}$, where c_+ and c_- are the densities of positive and negative ions. The Stokes equation can be written as

$$\eta \frac{\partial^2 \mathbf{u}(z)}{\partial z^2} + (c_+ - c_-)\mathbf{E} = 0, \quad (8.4)$$

where the net charge density is given by Poisson's equation,

$$(c_+ - c_-) = -(\epsilon_0 \epsilon) \frac{\partial^2 \phi_e(z)}{\partial z^2}. \quad (8.5)$$

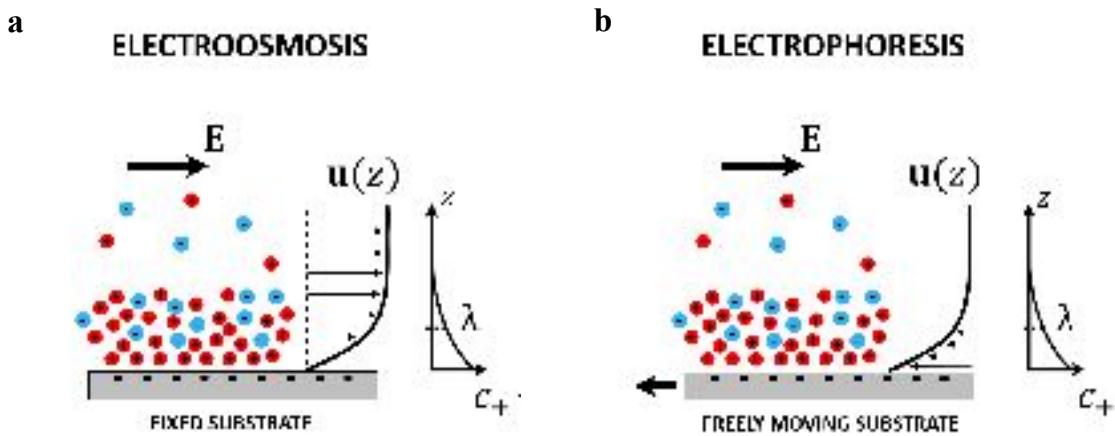


Figure 8.4: **Electroosmotic vs electrophoretic transport.** Schematic illustrations of (a) electroosmotic and (b) electrophoretic transport phenomena. $\mathbf{u}(z)$ is the fluid velocity in an interfacial layer of characteristic thickness λ_D (Debye screening length). The physics is identical, but the boundary conditions change: in (a) the substrate is held, but in (b) it is the fluid that is held far away.

Assuming no dynamic slip at the solid surface ($\mathbf{u}(0) = 0$, do not confuse the dynamic slippage with the electrostatic slippage we are after) and guaranteeing the continuity of stress between the interface and the bulk ($\partial\mathbf{u}(\infty)/\partial z = 0$), the integration of Eq. (8.4) gives the velocity profile $\mathbf{u}(z)$ of the interfacial fluid [11]:

$$\mathbf{u}(z) = (\epsilon_0 \epsilon) \frac{[\phi_e(z) - \zeta]}{\eta} \mathbf{E}, \quad (8.6)$$

where $\zeta = \phi_e(0)$ to satisfy the boundary condition $\mathbf{u}(0) = 0$, with $\zeta < 0$ for negatively charged surfaces, and vice versa. Eq. (8.6) describes the *electroosmotic flow* near a fixed charged surface under an applied electric field; its physical meaning is quite clear: The fluid velocity is determined by a balance between the driving electric force and the viscous friction force on the surface. The slip velocity $\mathbf{u}_{S,EO}$ is defined at the outer limit of the interfacial layer — strictly speaking, at $z = \infty$ because the diffuse layer extends to infinity, but in practice somewhere after the Debye screening length — as [11]

$$\mathbf{u}_{S,EO} = \lim_{z \rightarrow \infty} \mathbf{u}(z) = -(\epsilon_0 \epsilon) \frac{\zeta}{\eta} \mathbf{E}. \quad (8.7)$$

Thus, $\mathbf{u}_{S,EO}$ increases linearly with the electric field and the sign of ζ determines whether $\mathbf{u}_{S,EO}$ and \mathbf{E} are aligned or counter-aligned.

The leap from electroosmosis to electrophoresis is straightforward (Fig. 8.4b): we no longer fix the substrate, but let it move freely and hold the fluid far away ($\mathbf{u}(\infty) = 0$), instead. This would be the case where the *substrate* is the surface of a colloidal particle suspended in an electrolyte solution. Due to symmetry, the balance between Stokes friction and electric forces is identical. However, the new boundary condition leads to fluid velocities pointing in the opposite direction. Hence, the electrophoretic slip velocity, defined just outside the interface, is

$$\mathbf{u}_{S,EP} = \lim_{z \rightarrow \infty} \mathbf{u}(z) = (\epsilon_0 \epsilon) \frac{\zeta}{\eta} \mathbf{E}, \quad (8.8)$$

where $\mu_{EP} = (\epsilon_0 \epsilon \zeta)/\eta$ is the *electrophoretic mobility*. The sign of μ_{EP} is given by $\zeta = \phi_e(\infty)$, which is now the *Zeta potential* of the particle, i.e., the electrostatic potential in the slip plane located somewhere after the Debye length. Negatively charged particles migrate against the electric field, whereas positively charged particles follow its direction.

What is typically measured by electrokinetic analyzers (e.g., Malvern Zetasizers or Anton Paar SurPASS instruments) is ζ : An electric field is applied in one direction and the velocity of micro- or nanoparticles is measured to deduce ζ , according to Eq. (8.8). Typical values of $\mathbf{u}_{S,EP}$ and ζ for colloidal particles are 1 to

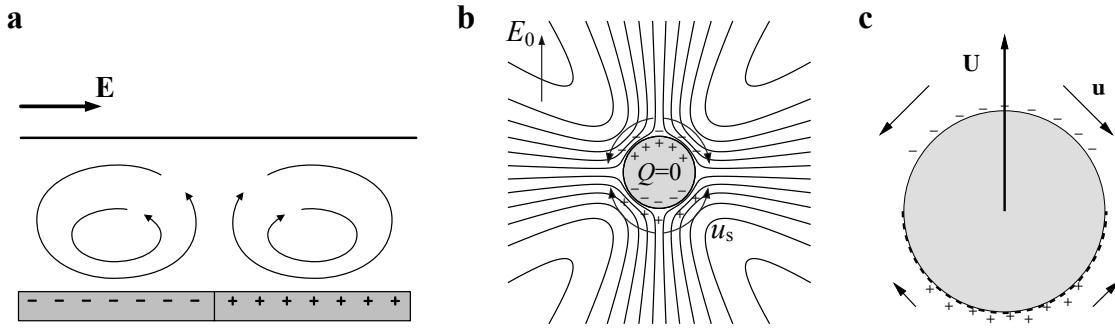


Figure 8.5: Electroosmotic flow in different charge and geometry conditions. Electroosmotic flow near (a) an heterogeneous surface with two stripes of opposite charge, (b) a conducting cylinder with zero net charge, and (c) a positively charged conducting cylinder. (b) is reproduced from Ref. [12] and (c) from Ref. [13].

$10 \mu\text{ms}^{-1}$ and 0 to $\pm 100 \text{ mV}$, respectively. It is tempting to claim that the surface charge of colloidal particles can be extracted from Zeta-potential measurements. However, the fluid layer near the charged solid surface is much more complex than what is captured by the simple picture presented above. First, corrections need to be introduced to take into account non-zero dynamic slip (see Section 8.1); this means that the fluid at $z = 0$ no longer has the same velocity of the solid surface. Second, the electrical double layer is affected by the density distributions of the ions under consideration, c_+ and c_- . The exponential decay, $\phi_e(z) = \phi_e(0) \exp(-z/\lambda_D)$, is only valid for weak potentials and sufficiently far away from the solid surface. If these conditions are not satisfied, the ionic cloud becomes increasingly complex involving ions that are bound to the solid surface. All in all, ζ does not have a physical meaning living up to its popularity: It is *simply* the potential in the region where the diffuse layer meets the bulk fluid.

So far, we have assumed that both the electric field and the Zeta potential are constant in the x-direction. In practice, solid surfaces are often littered with electrical heterogeneities. To understand the impact that these *defects* have on interfacial transport, let us consider a surface that is patterned with two bands of positive and negative surface charge density and apply an electric field parallel to the substrate (Fig. 8.5a). Even assuming that the external electric field does modify the surface charge distribution, the electrical mobility, and thus the effective slip velocity, varies in the x-direction, leading to pressure gradients and recirculating rolls, as reported by Ajdari and co-workers in the 1990s [14, 15]. Rolls can also be observed near uncharged conducting objects (e.g., an inert metal wire) under external DC fields [12, 13]. Here, the electric field induces a non-uniform cloud of counterions, i.e., regions where the current leaves and enters the surface. This effect is known as *induced-charge electro-osmosis* (ICEO); the resulting flow field is quadrupolar (Fig. 8.5b) and the slip velocity (defined just outside the induced screening layer) is proportional to the electric field squared, $\mathbf{u}_{S,\text{ICEO}} \propto E^2$. The reason for the latter scaling, which is in stark contrast to classical electroosmosis ($\mathbf{u}_{S,\text{EO}} \propto E$), lies in the fact that a conductive object of size R subjected to an external electric field E acquires a Zeta potential of the order of ER . Hence, ICEO dominates over standard EO in strong fields. The same arguments hold for *induced-charge electrophoresis* (ICEP), provided that we take into account the different boundary conditions [13].

According to the arguments given in this Section, the configurations in Fig. 8.4 and Fig. 8.5 should produce zero time-averaged flow whenever AC fields are used. However, this would be only true if the build-up of the diffuse double layers were instantaneous. If a uniform electric field is suddenly turned on at $t = 0$, the characteristic time for the formation of an induced screening cloud around an object of size R is the charging time $\tau_C = R\lambda_D/D_{\text{ions}}$, where λ_D and D_{ions} are the Debye screening length and the ion diffusivity, respectively; thicker interfaces made of slow ions take longer time to form. It is therefore possible to model the build-up of an induced-charge slip velocity using an RC-circuit of double-layer capacitors of thickness λ_D ,

$$\frac{\mathbf{u}_{S,\text{ICEO}}(t)}{\mathbf{u}_{S,\text{ICEO}}^\infty} \sim \left(1 - e^{-\frac{t}{\tau_C}}\right), \quad (8.9)$$

where $\mathbf{u}_{S,\text{ICEO}}^\infty$ is the steady-state slip velocity at times $t \gg \tau_C$ [13, 12]. Net flows are achieved in oscillating electric fields if the AC frequency, ω , is high enough so that the formation of the screening cloud is never fully

complete ($\omega\tau_C \gg 1$). On the other hand, ω must not be so large that the double layer does not have the time to build up at all; this requires that $\omega\tau_D \ll 1$, where $\tau_D = \lambda_D^2/D_{\text{ions}}$ is the diffusion time of ions across the Debye layer thickness.

8.2.2 Transport in chemical gradients

The coupling between electrolytes and solid surface is often not limited to the electrostatic interactions that occur within the diffuse double layer. Even neutral solute molecules show a degree of affinity with a neighboring substrate as a result, for instance, of excluded volume effects and van der Waals forces. These interactions are short-ranged and occur within a tiny, but finite, solid-liquid interface known as the *adsorption layer*.

The easiest way to understand interfacial transport caused by adsorption is to consider a liquid that contains two neutral solutes, say 1 and 2, with binding energies E_1 and E_2 with respect to a nearby substrate. The concentration of the two species is not uniform but decreases/increases in the direction parallel to the surface, as illustrated in Fig. 8.6. In this simple model, adsorption takes place within a distance λ_a from the solid surface. Far away from the surface the osmotic pressure, Π , has to equilibrate, i.e., $\Pi = (c_1 + c_2)k_B T = ck_B T = \text{const}$, with $c = c_1 + c_2$. Instead, at distances smaller than λ_a , the pressure is a function of the local position x ,

$$\Pi(x) = \left(c_1(x)e^{E_1/k_B T} + c_2(x)e^{E_2/k_B T} \right) k_B T, \quad (8.10)$$

or, equivalently,

$$\Pi(x) = c_1(x) \left(e^{E_1/k_B T} - e^{E_2/k_B T} \right) k_B T + c e^{E_2/k_B T} k_B T. \quad (8.11)$$

The interfacial layer is therefore characterized by a gradient (in the x direction) of osmotic pressure

$$\nabla\Pi = \nabla c_1 \left(e^{E_1/k_B T} - e^{E_2/k_B T} \right) k_B T, \quad (8.12)$$

whose direction depends on the magnitude of E_1 and E_2 . If $E_1 > E_2$ (the surface prefers 1 over 2), $\nabla\Pi$ follows the direction of ∇c_1 . The slip velocity is obtained by integrating the Stokes equation, $\partial^2 \mathbf{u}(z)/\partial z^2 = \nabla P/\eta$, and imposing the same boundary conditions we used for electrophoresis and electroosmosis. *Diffusioosmosis* — we hold the substrate, $\mathbf{u}(0) = 0$, see Fig. 8.6a — promotes fluid flow in the direction of the least favorite species,

$$\mathbf{u}_{\text{S,DO}} = -(\lambda_a^2/2\eta)\nabla\Pi \propto -(\lambda_a^2/2\eta)\nabla c_1. \quad (8.13)$$

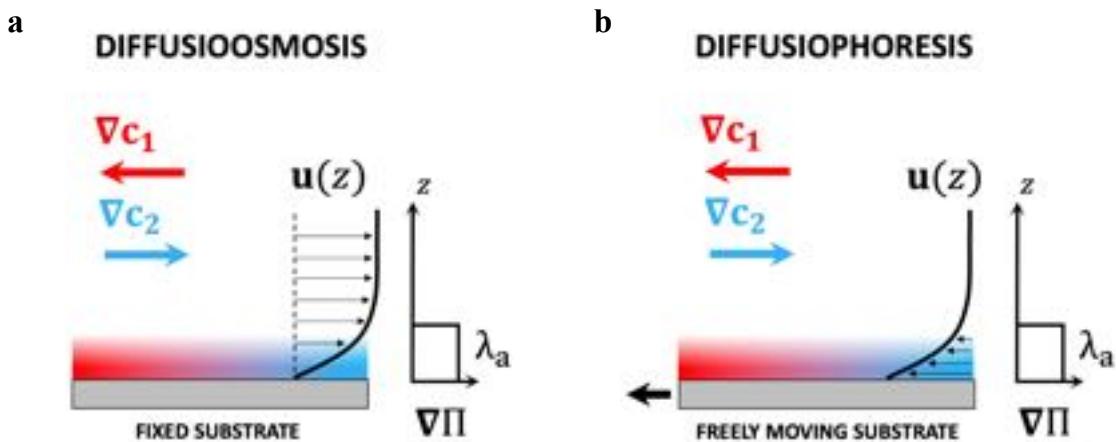


Figure 8.6: Diffusioosmotic vs diffusiophoretic transport. Schematic illustrations of (a) diffusioosmotic and (b) diffusiophoretic transport phenomena. $\mathbf{u}(z)$ is the fluid velocity in an interfacial layer of characteristic thickness λ_a (adsorption length). The physics is identical but the boundary conditions change: in (a) the substrate is held, but in (b) it is the fluid that is held far away.

Diffusiophoresis — we hold the bulk fluid, $\mathbf{u}(\infty) = 0$, see Fig. 8.6b — drives the surface/particle towards the preferred solute,

$$\mathbf{u}_{S,DP} = (\lambda_a^2/2\eta)\nabla\Pi \propto (\lambda_a^2/2\eta)\nabla c_1, \quad (8.14)$$

where $\mu_{DP} = (\lambda_a^2)/(2\eta)$ is the *diffusiophoretic mobility*.

Nonetheless, adsorption layers are more complex than the simplified model in Fig. 8.6. A more general expression for the slip velocity generated by a concentration gradient, ∇c_i , of neutral solute i was given by Anderson and Prieve in the 1990s [16]:

$$\mathbf{u}_{S,DO-DP}^i = \mp \left(\frac{k_B T}{\eta} \right) \Gamma L \nabla c_i, \quad (8.15)$$

where $-$ and $+$ refer, respectively, to DO and DP, Γ is the excess ($\Gamma > 0$) or depletion ($\Gamma < 0$) of solute induced by the solid surface, and L is the range of interaction.

Further complications stem from the fact that neutral solutes are rare, whereas situations in which surfaces (or particles) are immersed in electrolyte solutions are much more common. Since gradients of ions generate an electric field to keep a zero electrical current, the effective interfacial transport becomes the sum of electrophoretic (electroosmotic) and neutral diffusiophoretic (diffusioosmotic) slippage. Importantly — and in clear contrast to neutral diffusiophoresis (diffusioosmosis) — both terms are proportional to $\nabla \log(c)$, where c is the electrolyte concentration in bulk. This dependence can be qualitatively justified by noting that the interfacial thickness λ_a in Eqs. (8.13) and (8.14) is now the Debye length, $\lambda_D \sim 1/\sqrt{c}$, so that $\mathbf{u}_{S,DP}/DO \sim \nabla \log(c)$. Likewise, the electric field, $\mathbf{E} = -\nabla\phi_e$, in Eqs. (8.7) and (8.8) is linked to c via the Boltzmann distributions, $c_{\pm} = c \exp(\mp\phi_e/k_B T)$, which leads to $\mathbf{E} \sim \nabla \log(c)$.

The absolute value and direction of the phoretic mobility of particles surrounded by gradients of electrolytes are determined by the sign and magnitude of μ_{EP} and μ_{DP} . Consider for instance the phoretic motion of negatively charged particles ($\zeta < 0$) in a concentration gradient of salt, ∇c . Diffusiophoretic slippage always

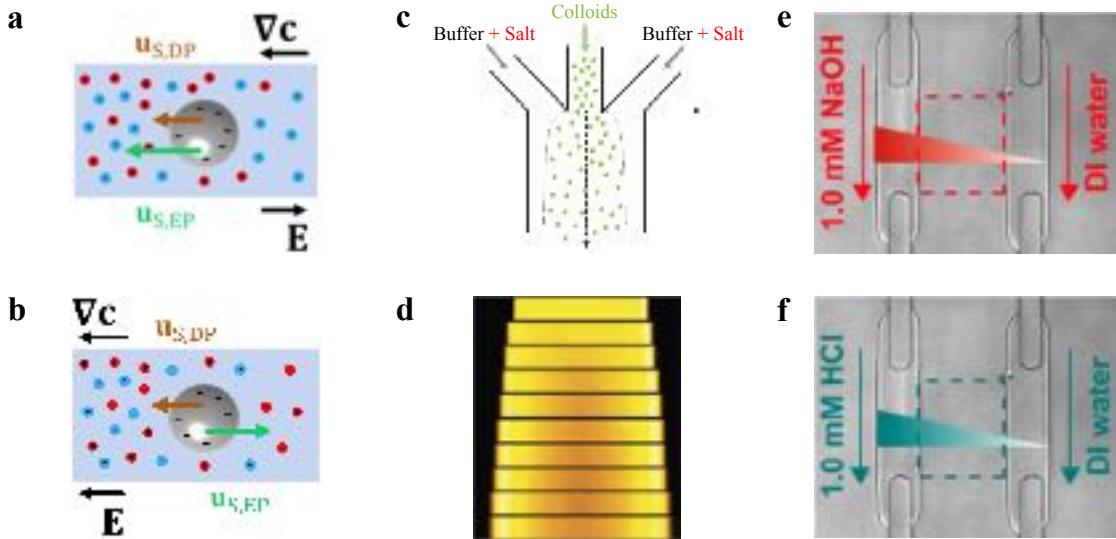


Figure 8.7: **Phoretic transport in a concentration gradient of a salt.** Schematic illustration of phoretic transport in a concentration gradient of a salt made of ions of diffusivity D_+ and D_- . (a) The particle is driven up the gradient if $D_- > D_+$, (b) whereas migration can occur in either directions if $D_+ > D_-$. Experimental technique to measure the phoretic transport in concentration gradients. (c) Schematics of a microfluidic device where salt is injected using two side inlets. (d) The colloids spread due to positive diffusiophoresis. (e-f) Diffusiophoresis of colloids in (c) NaOH and (d) HCl concentration gradients. The colloids migrate up and down the gradient, respectively. (c) and (d) are reproduced from Ref. [17]. (e) and (f) are reproduced from Ref. [18].

drives the particle up the concentration gradient (Eq. (8.14)), whereas the sign of μ_{EP} depends on which ion diffuses faster. If the negative ions are more mobile (Fig. 8.7a), the electric field is counter-aligned with ∇c and $\mathbf{u}_{S,EP}$ and $\mathbf{u}_{S,DP}$ point in the same direction. In contrast, faster positive ions (Fig. 8.7b) lead to electrophoretic velocities driving the particle down the gradient and thus counteracting the diffusiophoretic drift; in the latter case, the direction of transport is dictated by the magnitude of μ_{EP} and μ_{DP} .

In experiments, the phoretic transport in concentration gradients is usually measured by injecting particles in microchannels and applying ∇c in the direction perpendicular to the channel's walls. The colloids migrate up or down the gradient — ∇c can be calibrated by means of fluorescent labels [19] — and the phoretic mobility is inferred from the magnitude and direction of their velocity. Typical examples are illustrated in Fig. 8.7c-f. Figs. 8.7c and 8.7d show an experiment in which ions are pumped using two side inlets, whereas the colloids flow in the centre of the microchannel [17]; if the phoretic mobility of the particles is *positive*, the particle density profile quickly spreads out towards the boundaries. The experiments in Figs. 8.7e and 8.7f take advantage of hydrogel walls to achieve controllable diffusion of electrolytes from the left side only, while setting to zero the concentration at the opposite wall [18]; the colloids still flow in the central channel. Changing electrolyte solution leads to ions of different diffusivity and the phoretic slip velocity can even swap sign as illustrated in Fig. 8.7a-b and discussed therein.

8.2.3 Transport in temperature gradients

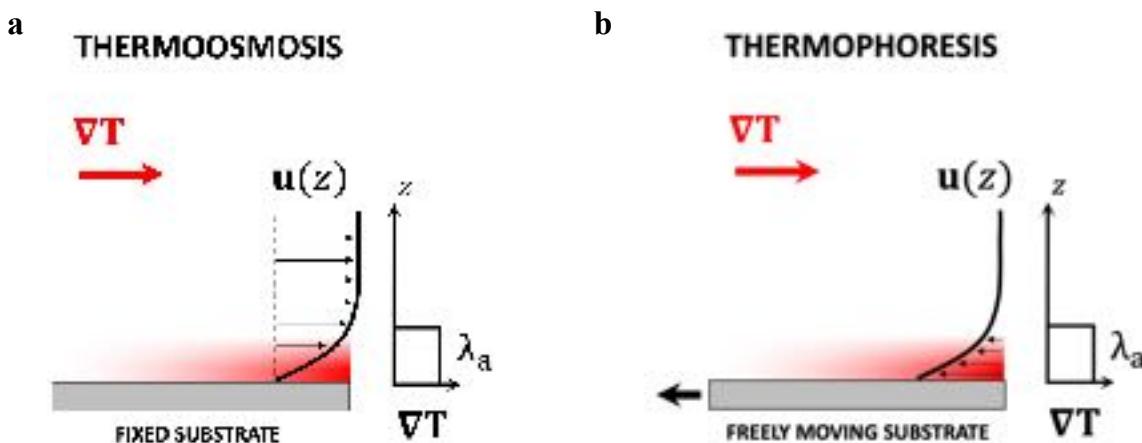


Figure 8.8: **Thermoosmotic vs thermophoretic transport.** Schematic illustrations of (a) thermoosmotic and (b) thermophoretic transport phenomena. $\mathbf{u}(z)$ is the fluid velocity in an interfacial layer of characteristic thickness λ_a (adsorption length). The physics is identical, but the boundary conditions change: in (a) the substrate is held, but in (b) it is the fluid that is held far away. [Is the notation λ_a correct for thermal effects or should it be different from the notation for chemical effects??]

Thermoosmosis and thermophoresis refer to the interfacial transport in the presence of thermal gradients (Fig. 8.8). The transport of mass caused by temperature differences is closely related to the fluid flow induced by osmotic pressure gradients; in fact, the Onsager's reciprocal theorem states that the heat flow per unit of pressure difference and the density matter flow per unit of temperature difference are equal. By computing the flux of energy (enthalpy) carried by pressure-driven convection of fluid, one can derive the momentum flux that would result from an applied temperature gradient and, therefore, the slip velocity just outside the interfacial layer [We could add a reference to Frank's experimental chapter here] [11, 20]. Due to reciprocity, the thickness of the interface, i.e., the region where the interaction solvent-solid surface plays a role, is also reasonably assumed to be the same adsorption layer introduced in Section 8.2.2. The slip velocity is [11]:

$$\mathbf{u}_{S,TO-TP} = \mp \frac{2H}{\eta} \nabla \log T, \quad (8.16)$$

[I assume the log here is wrong? Please check] where $-$ and $+$ refer, respectively, to thermoosmosis and

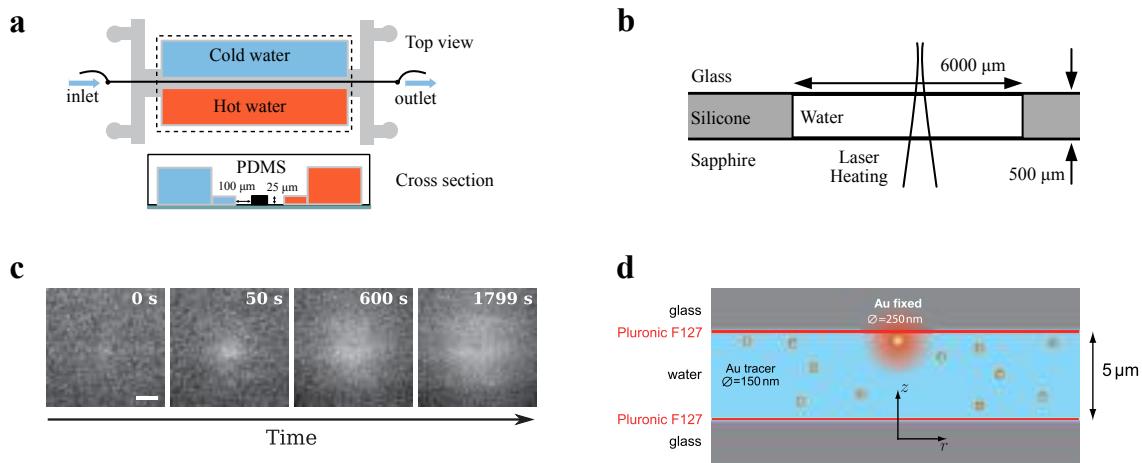


Figure 8.9: Microfluidic devices used to impose a temperature gradient. The temperature gradient ∇T is established (a) by flowing hot and cold water in side channels [21] or (b) focusing an infrared laser on a light-absorbing material [22]. (c) Thermophoretic accumulation of colloids [23] and (d) thermoosmotic flow [24] caused by point-like heat sources.

thermophoresis, ∇T is the temperature gradient, and H is the first moment for the excess enthalpy at the interface (calculated integrating over the whole interfacial layer). If the solvent molecules are attracted by the solid surface (most common situation), the thermoosmotic slip velocity is directed towards regions of higher temperature, whereas thermophoretic slippage leads to colloids migrating towards the cold side.

The problem of Eq. (8.16) consists in the fact that it is hard to measure, or even estimate, H . For this reason, the thermophoretic drift of microparticles is often characterized using a different framework. Piazza and co-workers [25] considered a suspension of colloidal particles of mean number density n and monitored the concentration profile when a temperature gradient was applied along one direction (say the x -direction). The number of particles that flow through a slice dx per unit of area and time is the sum of a diffusion flux, $D(dn/dx)$, and a thermophoretic one, $n\mu_{TP}(dT/dx)$, where D is the diffusion coefficient of the particles and μ_{TP} is their *thermophoretic mobility*. By imposing a steady-state zero flux and defining the *Soret coefficient* as $S_T = \mu_{TP}/D$, we obtain:

$$\frac{\partial n}{\partial x} = -nS_T \frac{\partial T}{\partial x}. \quad (8.17)$$

Eq. (8.17) tells us that it is possible to understand the phoretic transport of microparticles inside weak temperature gradients by looking at the density profile, in analogy to colloidal suspension driven by external fields. S_T is positive when the particles move to the cold and negative otherwise.

In experiments, colloid transport under controlled thermal gradients can be achieved using microfluidic devices (Fig. 8.9a) consisting of two side channels that contain hot and cold water and one central chamber that is filled with the colloidal suspension [21]. The particles migrate towards either the hot or the cold wall depending on the sign of their Soret coefficient, which is in turn a function of the background temperature, the particles' size and surface chemistry, and the properties of the surrounding fluid. It is also possible to establish temperature differences using thermal lenses [26] or focused laser beams. For example, a sample cell illuminated by an infrared laser (Fig. 8.9b) can induce thermophoretic aggregation of colloidal particles at the cold wall [22]. Instead, focusing the beam on an absorbing particle leads to a point-like heat source triggering the accumulation of colloids with negative Soret coefficient (Fig. 8.9c [23]) or the formation of thermoosmotic flows (Fig. 8.9d [24], note that the absorbing particle is anchored to the substrate).

8.3 Self-phoretic colloids

This section explores self-phoretic colloids, beginning with the concept of phoresis where microparticles move due to external fields. We then delve into autophoresis, where the colloids themselves create the gradients

influenced by surface differences. The following parts discuss the various ways these colloids move, from electric fields to chemical and thermophoretic interactions. Detailed illustrations and explanations provide a thorough insight into self-phoretic colloids.

8.3.1 From phoresis to autophoresis

Phoresis describes the migration of microparticles in electric fields (electrophoresis, Subsection 8.2.1), up or down concentration gradients of neutral and/or ionic chemical species (diffusiophoresis, Subsection 8.2.2), and in environments characterized by temperature differences (thermophoresis, Subsection 8.2.3). Homogeneous colloids experience a zero net force, yet move due to interfacial slippage. The step from phoresis to autophoresis (or self-phoresis) is simple (see Fig. 8.10): Gradients are no longer applied by the user, but established by the colloidal particles themselves via surface heterogeneities. As such, surface patches determine not only the strength of the self-generated gradient but also the direction of the corresponding slip velocity; if the microparticle rotates due to Brownian motion, the velocity vector will do the same.

There exist several methods to prepare colloidal particles with patchy surfaces. The list includes physical vapor deposition, colloidal fusion, electrodeposition, surface modification at interfaces, colloidal lithography, and bulk synthesis. Discussing these fabrication techniques is beyond the scope of this Chapter and we refer you to Refs. [27, 28, 29] for comprehensive reviews of the topic [Isn't there an experimental chapter we can refer to?]. Hereafter, we assume that *patchy colloids* (also known as *Janus particles* when the patch covers one hemisphere) are ready for use and describe how similar transport phenomena can be achieved in an autonomous fashion. We should be aware, however, that self-phoretic transport is far more complex than its phoretic counterpart, not only because the particles no longer have a homogeneous surface, but also due to multiple mechanisms acting simultaneously.

8.3.2 Self-electrophoretic colloids

Self-electrophoretic colloids were the first experimental realization of self-phoretic active particles. At the beginning of the century, Sen and co-workers pioneered an experimental system consisting in negatively charged bimetallic (e.g., gold–platinum) rods immersed in a solution of hydrogen peroxide (H_2O_2) [31]. The Janus rods behave as batteries (Fig. 8.11), with a platinum anode and a gold cathode. An oxidation reaction decomposes the peroxide near the platinum end into protons, electrons and oxygen molecules,



the electrons then travel through the rod to the gold side, where a reduction reaction combines them with protons and peroxide to form water,

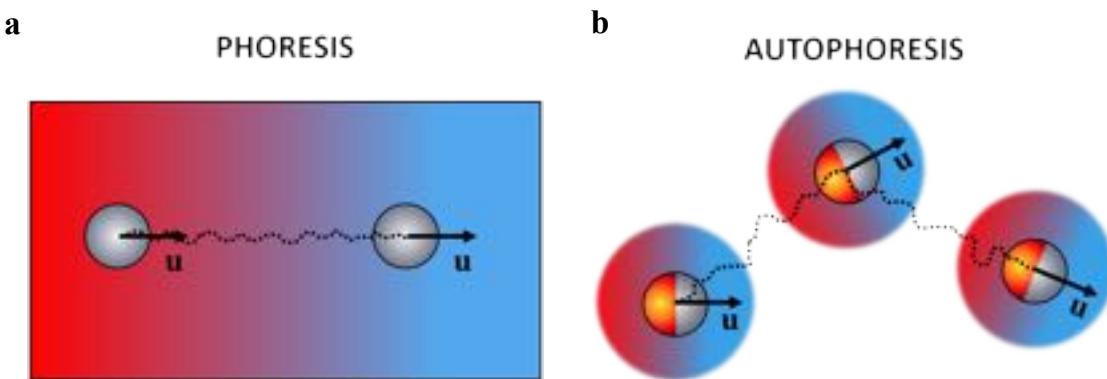


Figure 8.10: **Phoretic vs self-phoretic colloid transport.** Sketch of (a) phoretic transport, where the gradient is applied externally, and (b) phoretic transport, where the gradient is generated by the colloid itself.

Thus, the excess and depletion of protons generate an electric field, \mathbf{E} , which points from the Pt-anode to the Au-cathode and leads to electric body forces on the surrounding fluid.

In contrast to classical electrophoresis, the strategy shown in Fig. 8.11 includes two sources of net charge density: the *standard* diffuse cloud and the reaction-driven flux. Since experimental data suggest that the first contribution is dominant [30], the self-phoretic motion is described by the same arguments introduced in Subsection 8.2.1, i.e., the slip velocity is

$$\mathbf{u}_{S,EP} = (\epsilon_0 \epsilon) \frac{\zeta}{\eta} \mathbf{E}, \quad (8.20)$$

where $\mu_{EP} = (\epsilon_0 \epsilon \zeta) / \eta$ is the self-electrophoretic mobility. Since the rods carry a negative surface charge ($\zeta < 0$), self-phoretic motion and electric field are counter-aligned, and the particle moves with the platinum heading. Importantly, \mathbf{E} is internally generated; it increases with the distance between the charge density clouds (roughly equal to the length of the rod), their thickness (the Debye length), and the reaction proton flux, which is in turn given by the bulk fuel concentration $[H_2O_2]_0$. [30]. The propulsion ceases whenever the circuit is interrupted, e.g., by adding an insulating layer between platinum and gold. Reported self-electrophoretic velocities range from few $\mu\text{m s}^{-1}$ to fractions of mm s^{-1} . This broad spectrum is not only due to the parameters mentioned above, but also to other experimental details including the curvature of the rods or the incorporation of carbon nanotubes within the Pt-segment [32]. Many aspects of the propulsion are still a matter of debate.

8.3.3 Self-phoretic transport in AC fields

Active colloids can be also engineered by taking advantage of *induced-charge electroosmosis* (ICEO) and *induced-charge electrophoresis* (ICEP). While external AC electric fields are needed to pursue these strategies, \mathbf{E} is applied in the direction normal to the plane of motion, as opposed to classical electrophoresis and electroosmosis. The typical experiment involves Janus microparticles (or dumbbells) with hemispheres (or lobes) made of materials with different dielectric properties; often, but not always, one side is coated with a metal, whereas the bare hemisphere is dielectric. The colloids are dispersed in aqueous solution and confined between two planar electrodes (e.g., gold or indium tin oxide substrates) connected to an AC power source. The frequency and peak-to-peak voltage are in the range of 1 to 500 V and 10^2 to 10^7 Hz, respectively. The largest induced dipole moment aligns with the electric field and the patchy particles move parallel to the electrodes following the axis that links the poles of the two hemispheres (Fig. 8.12).

The mechanism of propulsion is subtle and depends on the constituent materials of the particle, its distance from the electrodes and the frequency of the applied electric field. At frequencies between $\approx 10^2$ and $\approx 10^4$ Hz, self-transport of colloids with dielectric asymmetry is usually due to imbalanced induced-charge electroosmotic flows (ICEO) near the electrodes [34, 35, 36]. Consider first an homogenous spherical particle sitting on a conductive surface. An electric field applied in the vertical direction sets in motion the electrical double layer of the substrate and, simultaneously, polarises the particle. The electric field increases (decreases) beneath a particle of high (low) surface charge and fluid is transported away from (toward) it accordingly. The net

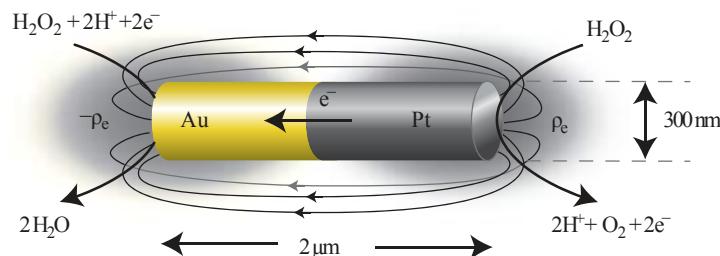


Figure 8.11: **Self-electrophoretic bimetallic rods in hydrogen peroxide solution.** The particle acts as a short-circuited galvanic cell producing a flux of protons at the Pt-anode and consuming protons at the Au-cathode. The grey areas schematically show the charge density clouds and the arrows describe the electric field. Reproduced from Ref. [30].

result is recirculating electroosmotic rolls similar to those illustrated in Fig. 8.5 [37, 38], which however do not lead to net motion due to symmetry. If the particle is Janus, the magnitude, or even the direction, of the rolls near the two hemispheres is different, causing net propulsion as shown in the left sketch of Fig. 8.12. The minimum operating frequency (≈ 10 to 100Hz) stems from the need to impose sufficiently large electric fields (\mathbf{E} significantly decreases at low frequencies ω [37]) in the sample cell. Instead, the upper bound (≈ 10 to 100kHz) is determined by the inverse of the charging time of the underlying conductive substrate, $\tau_{C,E}^{-1} = D_{\text{ions}}/(\lambda_D L)$, where λ_D is the Debye screening length, D_{ions} is the ions diffusivity, and L is the cell thickness. When $\omega \gg \tau_{C,E}^{-1}$, the diffuse cloud of counterions does not have time to build up near the electrode during the AC period. In practice, experiments work in a regime where the electrical double layer of the electrodes is only partially formed ($\omega > \tau_{C,E}^{-1}$). At these operating frequencies, the velocity of the particles decreases with ω and is proportional to the electric field squared, E^2 , as predicted for ICEO (see Section 8.2.1).

Even though the electroosmotic slippage produced by the electrodes becomes very small above $\approx 10\text{kHz}$, metalodielectric microparticles can still self-propel via induced-charge electrophoresis (ICEP) [33, 39, 40]. Here, the presence of electrodes just underneath the colloids is no longer required (although metal-coated particles unavoidably sink onto the bottom substrate) since the induced-charge slip flow around the conducting coating is strong enough to cause migration as illustrated in the right sketch of Fig. 8.12. Since the dominant slippage now occurs at the surface of the colloids rather than at the electrode–fluid interface, the relevant charging frequency is $\tau_{C,P}^{-1} = D_{\text{ions}}/(\lambda_D R)$, where R is the particle's radius. If $\omega < \tau_{C,P}^{-1}$, the electrical double layer of the colloids develops entirely and the particle moves with the less polarisable hemisphere forward at velocities proportional to E^2 . For $\omega > \tau_{C,P}^{-1}$, the screening layer struggles to form and the velocity gradually drops. If ω is increased even further, the direction of transport changes, i.e., the colloids move with the metal heading, until no motion is observed above approximately 10MHz . The frequency of arrest is ascribed to the inverse of the time it takes to the ions to diffuse the Debye layer thickness, $\tau_D^{-1} = \lambda_D^2/D_{\text{ions}}$. On the other hand, the reason for the change of direction is controversial; the reversal is somewhat related to the time it takes to the ions to diffuse across the particle's perimeter but the underlying mechanism remains elusive.

8.3.4 Self-diffusiophoretic colloids

The most paradigmatic example of self-diffusiophoretic colloids is the design proposed by Ramin Golestanian and co-workers in 2005 [41] and shown in Fig. 8.13a. A model spherical colloid of radius R is immersed in water and equipped with an enzymatic site which produces neutral product particles of diffusivity D_{pp} at an uniform rate k (products per unit of time). The self-generated concentration gradient of products drives the model colloid via diffusiophoresis (Eq. (8.14)) and depends on k and D_{pp} . If the enzymatic reactions are much

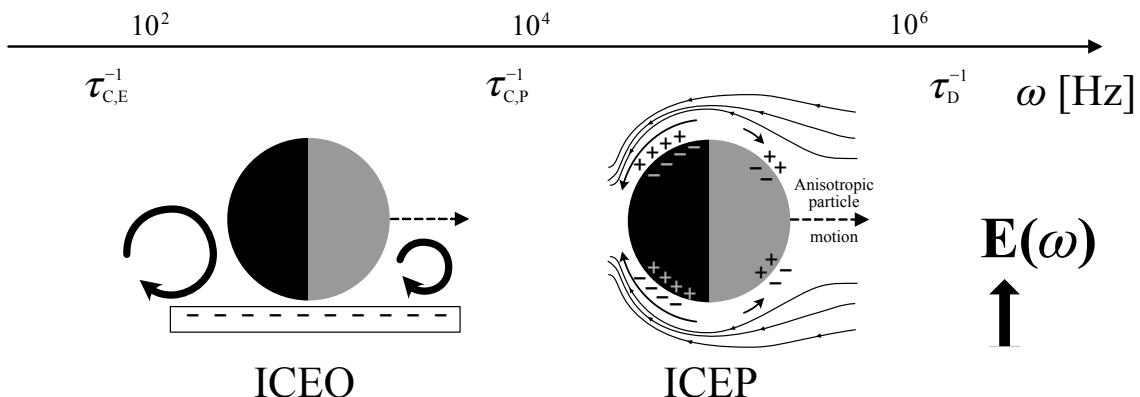


Figure 8.12: **Schematic illustration of self-propulsion under AC electric fields.** The active motion is caused by (a) induced-charge electroosmotic flows or (b) induced-charge electrophoresis depending on the AC frequency. The horizontal axis roughly marks the operating frequencies: $\tau_{C,E}^{-1}$, $\tau_{C,P}^{-1}$ and τ_D^{-1} (see text for details). (b) is reproduced from Ref. [33].

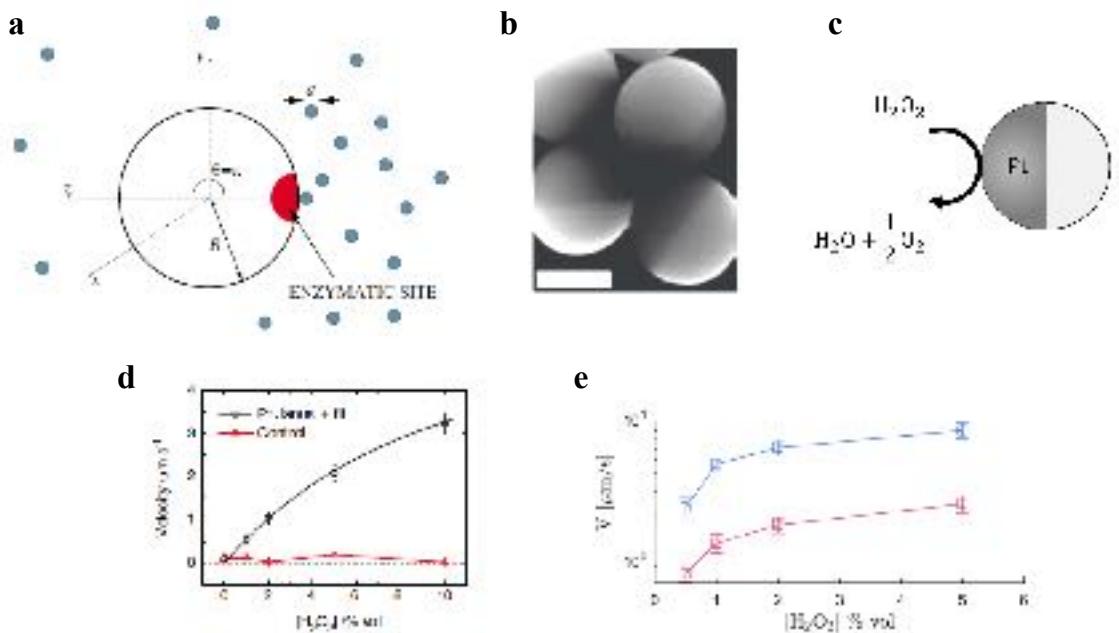


Figure 8.13: Catalytic Janus particles. (a) Model self-diffusiophoretic particles driven by a local enzymatic reaction [41]. (b) SEM picture reproduced from [42] of polymer colloids half-coated with 2 nm of platinum (scale bar, 500 nm). (c) Sketch of the hydrogen peroxide decomposition boosted by a Pt-coating. (d) Velocity of uncoated (red) and Pt-coated particles as a function of the hydrogen peroxide concentration in bulk. Reproduced from Ref. [43]. (e) Velocity of catalytic colloids coated with 2 nm of platinum via sputtering (blue) and e-beam evaporation (red). Reproduced from Ref. [44].

faster than the diffusion of product particles across the colloid (reaction limit), the slip velocity, $\mathbf{u}_{S,DP}$, increases with k and does not depend on the particles size (unless the size of the enzymatic patch changes with R). In contrast, if products diffuse fast compared to the reaction rate (diffusion limit), $\mathbf{u}_{S,DP}$ is no longer a function of k and decreases with R (the gradient is smoother if the products diffuse larger distances).

To date, the synthetic self-propelling particles that, at first glance, better resemble the design proposed in Ref. [41] are catalytic colloids similar to those illustrated in Fig. 8.11, but with an inert material in place of the Au-anode. This system is widely used among the synthetic active matter community since experiments are often as simple as sputtering a thin layer of platinum onto one hemisphere of polystyrene or silica microparticles (Fig. 8.13b), dispersing them into a container filled with a hydrogen peroxide solution ($[\text{H}_2\text{O}_2]_{\infty,\text{vol}}$ between 0.1% and 10%), and observing their motion under a commercial microscope. The platinum cap locally decomposes H_2O_2 and, as such, imposes concentration gradients of reactants and reaction products in the vicinity of the particle. The same decomposition can be activated by UV light if the patch is made of hematite rather than platinum [45]. Despite the seeming simplicity, understanding the mechanism of self-propulsion of Pt-coated catalytic colloids is a taunting task. Many explanations have been proposed, ranging from neutral diffusiophoresis to combinations of diffusiophoresis and electrophoresis. In the following, we qualitatively discuss them in increasing order of complexity.

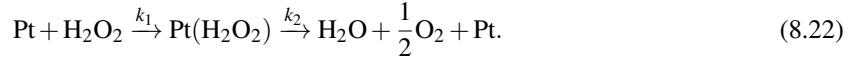
Neutral diffusiophoresis. First, we consider only the decomposition of hydrogen peroxide into water and oxygen promoted by the platinum coating (Fig. 8.13c),



As opposed to bimetallic swimmers, neither ionic species nor electron fluxes feature in the decomposition. Hence, the slip velocity is only determined by the concentration gradients of O_2 and H_2O_2 ; the adsorption length, λ_{eff} (Eq. (8.14)), is an effective thickness which takes into account the interaction of both species with the surface of the colloid. The particles typically move in the direction of the uncoated hemisphere

[42, 43, 46, 47, 45, 48], suggesting that the reactants (H_2O_2) interact more favorably with the surface than the products (O_2).

Recent experiments also demonstrated that the reaction rate is such that neutral diffusiophoretic transport occurs in the reaction regime [47], with slipping velocities proportional to both the bulk peroxide concentration $[\text{H}_2\text{O}_2]_\infty$ and reaction rate k , $\mathbf{u}_{\text{S,DP}} \propto k(\lambda_{\text{eff}}^2/\eta)[\text{H}_2\text{O}_2]_\infty$. Nonetheless, this description fails to capture the fact that the linear increase as a function of $[\text{H}_2\text{O}_2]_\infty$ does not hold at high H_2O_2 concentrations (see black data in Fig. 8.13d [43]). To understand the upward curvature, one must mind that the decomposition occurs in two stages via the formation of a bound state $\text{Pt}(\text{H}_2\text{O}_2)$ followed by its dissociation into water and oxygen, i.e.,



This leads to a *Michaelis-Menten behavior* and an effective reaction rate [43],

$$k_{\text{eff}} = k_2 \frac{[\text{H}_2\text{O}_2]_\infty}{[\text{H}_2\text{O}_2]_\infty + k_2/k_1}. \quad (8.23)$$

Diffusiophoresis + electrophoresis. Arguments uniquely based on neutral diffusiophoresis leave a number of issues unanswered. Experimental data agrees with the model only if the lengthscale of the interactions between the solutes and the particle's surface is about 1 nm, which rules out classical van der Waals or excluded volume interactions (unless a large dynamic slip length is simultaneously taken into account). Excluding the presence of ion gradients also overrules the fact that H_2O_2 is de-protonated and protonated in aqueous solutions.

The excess and depletion of ions near the colloid's surface add an electrophoretic term to the slip velocity. This complementary scenario is discussed in great details in Ref. [47], where the authors conclude that high dynamic slippage or very high Zeta potentials would be needed to justify the swimming velocities of colloidal microparticles. Ref. [47] also proposes that internal electric fields, due to oxidation/reduction reactions similar to those reported for bimetallic particle (Fig. 8.11), are unavoidably present in experiments because of variations in Pt thickness and O_2 gradients between the equator and the poles of the particle. The latter mechanism would explain the large increase of swimming speed (greater than one order of magnitude) reported when Pt-layers of the same thickness are deposited on the colloids via sputtering rather than e-beam evaporation (Fig. 8.13e) [44]; coatings obtained through sputtering are in fact known to be rougher than those made using evaporation techniques.

8.3.5 Self-thermophoretic colloids

Catalytic self-propulsion is easy to implement in experiments but hard to fully understand. In contrast, the physics behind self-thermophoretic active particles trivially follows from Section 8.2.3, but experiments that yield self-propulsion velocities larger than a few microns per seconds are very challenging.

Colloidal particles produce local temperature gradients if they are partially coated with a light-absorbing material (e.g., a metal) and illuminated by light, as schematically shown in Figs. 8.14a and 8.14b. If the light beam is sufficiently powerful, the temperature of the metallic layer increases and the heat is transferred to the surrounding fluid. The temperature gradient, ∇T , depends on the illumination intensity, I , the heat-absorption cross-section of the cap, σ_{abs} , and the thermal conductivity inside and outside the particle, $k_{\text{th,in}}$ and $k_{\text{th,out}}$. Assuming $k_{\text{th,in}} = k_{\text{th,out}} = k_{\text{th}}$,

$$\nabla T \propto \frac{\sigma_{\text{abs}} I}{R^2 k_{\text{th}}}, \quad (8.24)$$

where σ_{abs} is proportional to R^2 in the thin-cap limit [50]. If ∇T is large enough, the particles self-propel with slip velocity $\mathbf{u}_{\text{S,TP}} = -DS_{\text{T}}\nabla T$ (see also Section 8.2.3), where D is the diffusivity and S_{T} is an effective Soret coefficient (unfortunately, S_{T} is not as well-defined as in Section 8.2.3 because the particles are patchy). If S_{T} is positive/negative, the particle moves with the uncoated/coated hemisphere heading (Fig. 8.14c). Interestingly, the sign of the Soret coefficient can be altered, from positive (propulsion towards the cold region) to negative (propulsion towards the hot region) by adding surfactants to the aqueous solution [49].

Self-thermophoretic transport attracted the interest of experimental researchers because light is a minimally invasive way to control in-situ the strength of self-propulsion. The drawback is that demanding illumination intensities are required to achieve suitable temperature differences between the bare and coated poles and

therefore endow the colloids with sufficiently large slip velocities. Laser intensities as high as tens of $\text{mW } \mu\text{m}^{-2}$ limit the illumination area (i.e., the region where self-propulsion takes place) to a few microns [49] or call for advanced feedback techniques to guide the beam along the particle's trajectory [50].

Alternatively, moderate local heating upon laser intensities of the order of few $\mu\text{W } \mu\text{m}^{-2}$ can be used to trigger asymmetries in the vicinity of the coated colloids (note, however, that self-thermophoresis is no longer the dominant mechanism of self-propulsion). For example, local heating can decompose a binary mixture (e.g., water-2,6-lutidine) with lower critical temperature leading to a demixed phase near the light-absorbing cap (Fig. 8.14d, left panel) [51]; the asymmetry becomes more and more pronounced as ∇T increases, i.e., at larger laser intensities, and the particles self-propel faster and faster (Fig. 8.14d, right panel) [52]. Finally, swimming speeds greater than 10 mm s^{-1} can be achieved if light-absorbing Janus particles are confined at flat fluid–fluid interfaces, where thermal gradients promote surface tension differences and tangential Marangoni stresses (Fig. 8.14e) [53].

8.4 Experiments with active Brownian motion

In this last Section, we describe a simple experiment¹ where the motion of catalytic microswimmers on a glass substrate is observed using fluorescence microscopy and characterized by extracting the mean squared displacement (MSD) as defined and proposed, for example, in Chapter ??.

8.4.1 Methods

The first step consists in preparing microparticles half-coated with a thin layer of platinum. To this end, microscope slides are used as platforms to make monolayers of commercial polystyrene beads of radius $R = 1.4 \mu\text{m}$ (PS-FluoGreen, Microparticle GmbH). The slides are carefully wiped with an optical tissue soaked in isopropanol (Fig. 8.15a) and exposed to plasma for approximately two minutes (Fig. 8.15b) in order to increase their water wettability. Approximately $400 \mu\text{L}$ of aqueous colloidal suspension is then uniformly spread onto the glass substrate using a pipette (Fig. 8.15c) and let dry completely. The top surface of the slide is now

¹The full video of the experiment can be found at [XXX](#).

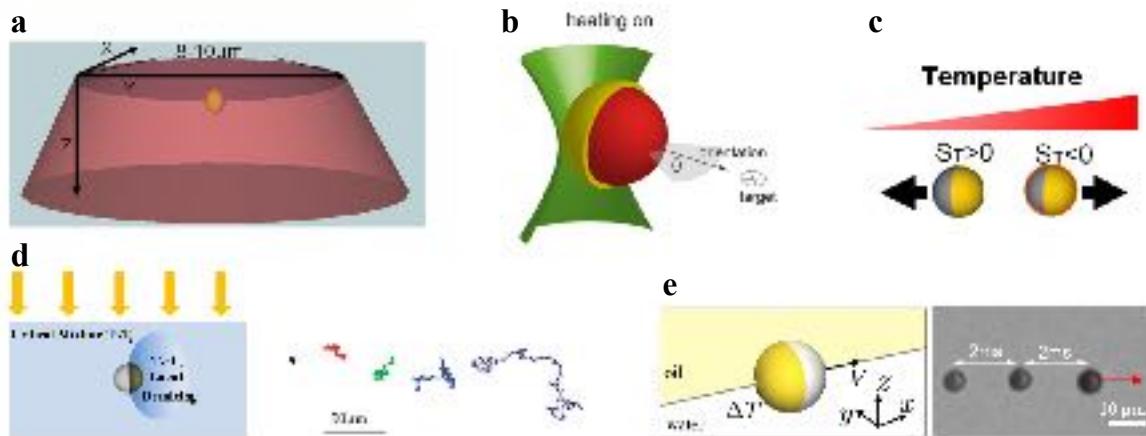


Figure 8.14: **Active particles driven by light.** (a-c) Self-thermophoretic active colloids. (a,b) Particles partly coated with a layer of gold are illuminated using a laser beam and (c) swim towards the cold or hot side depending on the sign of the Soret coefficient. (a) and (c) are reproduced from Ref. [49]. (b) is reproduced from Ref. [50]. (d,e) Other examples of active particles driven by light, but not self-thermophoretic. (d) Colloids driven by local demixing of a binary solvent with lower critical temperature. Left: sketch of the mechanism [51]. Right: particle trajectories for increasing illumination intensities [52]. (e) Marangoni surfers driven by surface tension gradients [53]. Left: sketch of the mechanism. Right: image overlay of a swimming particle. (d) is reproduced from Ref. [51] and [52], (e) is reproduced from Ref. [53].

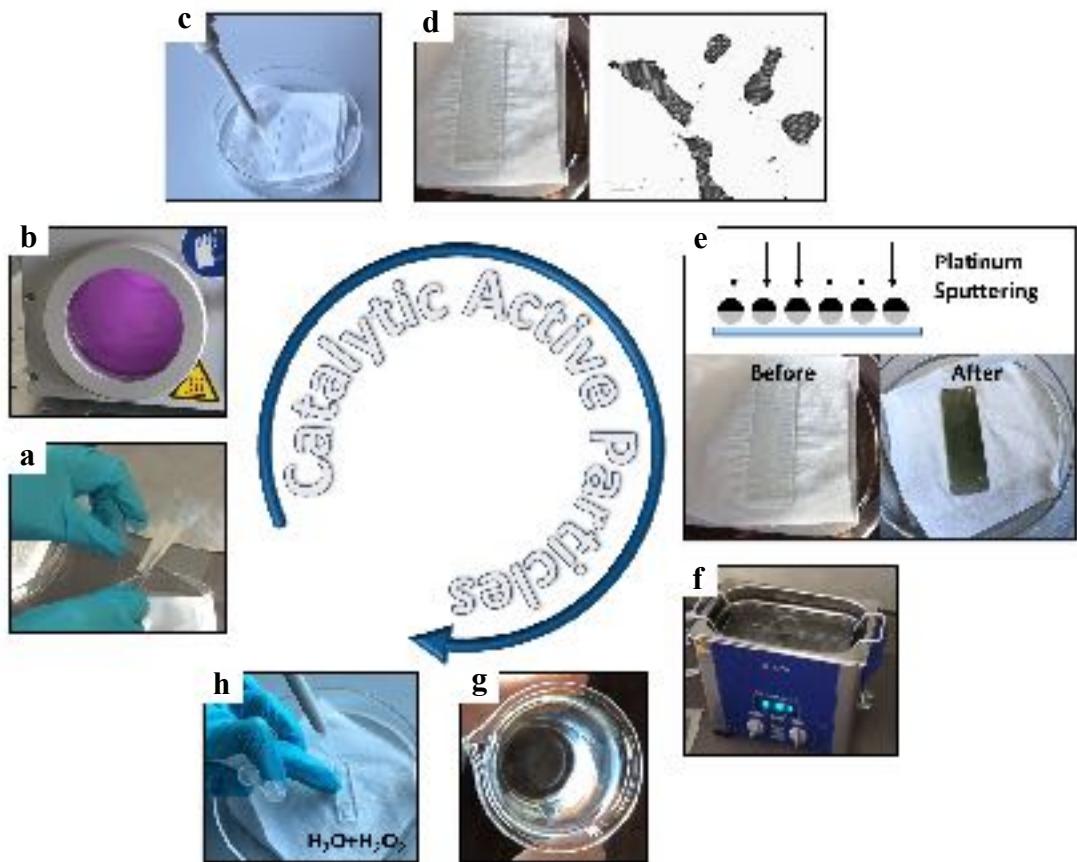


Figure 8.15: Main experimental steps during the preparation of catalytic active particles. A glass slide is (a) cleaned with isopropanol and (b) treated with plasma. (c) An aqueous suspension of polystyrene particles is spread on the slide and let dry, leading to a monolayer made of crystalline colloidal domains. (d) Two nanometres of platinum are sputtered onto the top hemispheres of the particles, which (e) are later re-dispersed in water via sonication. (f) After sedimentation, the coated colloids are (g) collected from the bottom of the container, dispersed in a $\text{H}_2\text{O}/\text{H}_2\text{O}_2$ solution and (h) pipetted into a quartz cell.

covered by *islands* of microparticles, as illustrated in Fig. 8.15d. Note that it is important that the colloidal suspension is sufficiently diluted to avoid the formation of double layers during the drying process; for this, we typically use solid contents between 0.1% and 0.5% w/v.

The monolayer is laid flat in the chamber of a sputter coater so that 2 nm of platinum can be deposited onto the exposed hemispheres of the particles, as sketched in Fig. 8.15e. The coated colloids are then re-dispersed in water by placing the microscope slide in a beaker (with the particles facing down) and sonicating for up to one minute (Fig. 8.15f). The particles come off the microscope slide and slowly sink to the bottom surface of the beaker (Fig. 8.15g); the excess of water is removed and the patchy beads are stored in a small container (e.g., an Eppendorf capsule).

Just before an experiment, a small amount of particles is added to a aqueous solution containing hydrogen peroxide at concentration, $[\text{H}_2\text{O}_2]_{\text{vol}}$, between 0.001 and 5 %. v/v. At this stage, the colloids become active due to the decomposition of H_2O_2 across the platinum-coated hemispheres and the formation of concentration gradients of reactants and reaction products nearby the microparticle (self-diffusiophoresis, see Subsection 8.3.4). The colloidal suspension is finally pipetted into a 100 μm thick quartz cell (Hellma cell, see Fig. 8.15h); the Pt-coated particles sediment onto the bottom substrate and their active motion – both translational and rotational motion occur in 2D [46] – is observed using an inverted fluorescence microscope (Olympus IX73, 20 \times magnification).

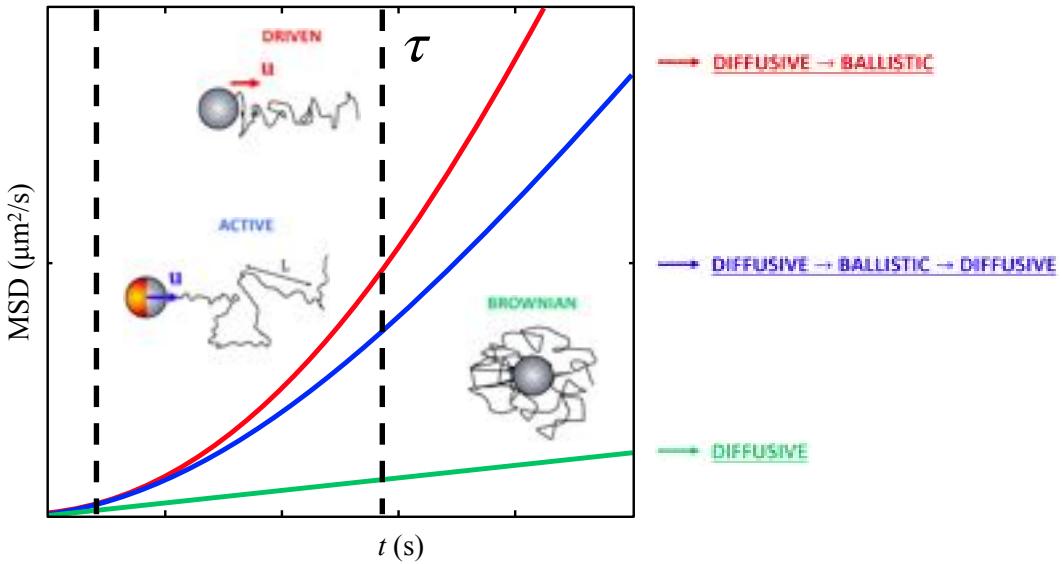


Figure 8.16: **Mean square displacements.** Sketches of Brownian, driven, and active colloidal particles, and their corresponding mean squared displacements (MSD) as a function of time.

8.4.2 MSD analysis

As described in Chapter ??, active microparticles swim at constant speed, u , and reorient at characteristic timescales, τ . The magnitude of u depends on the specific propulsion methods; in the case of catalytic colloids, it increases with the bulk concentration of *fuel* ($[\text{H}_2\text{O}_2]_{\text{vol}}$). On the other hand, τ is determined by the rotational diffusivity of the particles because the velocity vector follows the axis that links the poles of the coated and bare hemispheres (see sketch in Fig. 8.10). The product $(u \cdot \tau)$ Nuno: check notation of this paragrph and consistency between vectorial/scalar notation defines the *persistence length*, L , of the trajectories, i.e., the average length of the straight paths (analogous to the persistence length of polymer chains); fast but slowly rotating particles undergo long linear displacements (large L), whereas the active trajectories of colloids with low activity and large rotational diffusivity are more erratic (small L).

In the analysis of the current experiments, we extract u and τ from the mean squared displacement (MSD) of the particles. In order to briefly illustrate why this is possible, let us consider three types of microparticles in two dimensions (insets in Fig. 8.16): a Brownian particle, a particle driven by a constant external force (e.g., a magnetic field), and an active particle. The MSD of a *passive* Brownian particle grows linearly with time (Fig. 8.16, green curve), i.e.,

$$\text{MSD}_{\text{BP}}(t) = 4D_0 t, \quad (8.25)$$

where $D_0 = k_B T / (6\pi\eta R)$ is the translational diffusion coefficient. Instead, a driven particle undergoes not only diffusion but also ballistic motion at constant speed, u , in the direction of the applied force. The corresponding MSD is (Fig. 8.16, red curve), i.e.,

$$\text{MSD}_{\text{DP}}(t) = 4D_0 t + u^2 t^2, \quad (8.26)$$

where the quadratic terms dominates after a characteristic time (left vertical dashed line in Fig. 8.16) related to the Péclet number of the system. An active particle behaves as a driven colloid (Eq. (8.26)) for $t \ll \tau$, although the transition from diffusive to ballistic motion is hard (but not impossible [54]) to measure since translational diffusion is nearly negligible for micrometric particles swimming at few microns per second. What distinguishes an active particle from a driven particle is that, at times $t \gg \tau$, the directionality is lost due to rotational diffusion. Thus, in the long-time limit the motion becomes diffusive again, with an effective

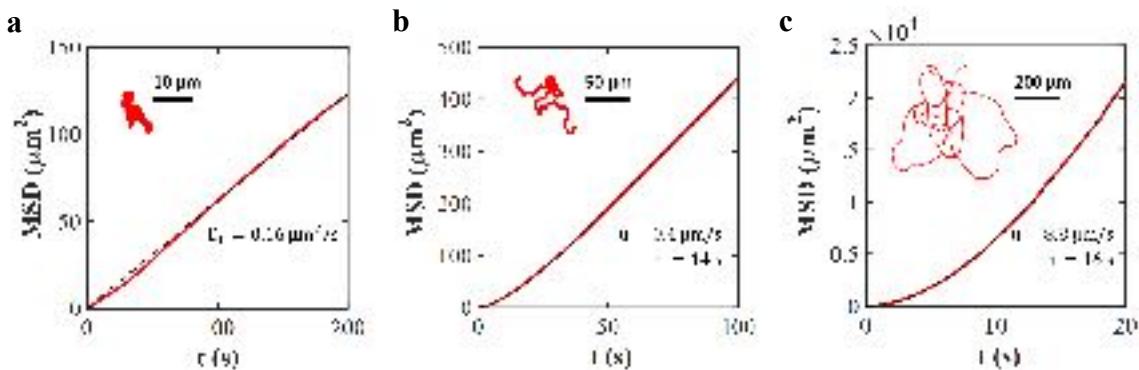


Figure 8.17: Mean square displacement vs lag time for different hydrogen peroxide concentrations. MSD vs lag time t for patchy particles immersed in a $\text{H}_2\text{O}/\text{H}_2\text{O}_2$ solution at (a) $[\text{H}_2\text{O}_2]_{\text{vol}} = 0\%$, (b) $[\text{H}_2\text{O}_2]_{\text{vol}} = 0.1\%$, and (c) $[\text{H}_2\text{O}_2]_{\text{vol}} = 2\%$. The insets show three typical trajectories after 450 s.

diffusivity, $D_{\text{eff}} > D_0$, that is a function of both u and τ . All in all, the MSD of an active particle is (Fig. 8.16b, blue curve) [55]²

$$\text{MSD}_{\text{AP}}(t) = 4D_0t + 2u^2\tau^2 \left(\frac{t}{\tau} - 1 + e^{-t/\tau} \right). \quad (8.27)$$

The three regimes (diffusive \rightarrow ballistic \rightarrow diffusive) become evident from the limits of Eq. (8.27). For $t \rightarrow 0$, Eq. (8.27) becomes equal to Eq. (8.25). For $t \ll \tau$, we can expand the exponential term of Eq. (8.27) in Taylor series ($e^{-t/\tau} \sim 1 - t/\tau + t^2/(2\tau^2)$) and obtain that $\text{MSD}_{\text{AP}}(t \ll \tau) = \text{MSD}_{\text{DP}}(t)$. For $t \gg \tau$, the exponential vanishes and $\text{MSD}_{\text{AP}}(t \gg \tau) = 4D_{\text{eff}}t$, where $D_{\text{eff}} = D_0 + (1/2)u^2\tau$.

8.4.3 Experimental data analysis

We can now apply the concepts illustrated in Section 8.4.2 to the analysis of experimental data. We discuss the following example.

Example 8.1: Pt-coated Janus particles in a hydrogen peroxide solution. Videos of experiments with Pt-coated Janus particles in a hydrogen peroxide solution can be found at [put here the link](#). The experimental acquisitions are performed at peroxide concentrations (i) $[\text{H}_2\text{O}_2]_{\text{vol}} = 0\%$, (ii) $[\text{H}_2\text{O}_2]_{\text{vol}} = 0.1\%$ and (iii) $[\text{H}_2\text{O}_2]_{\text{vol}} = 2\%$. Images were acquired at one frame per second and one pixel corresponds to $0.2925 \mu\text{m}$.

- a. Track and trace the particles for each of the cases.
- b. Extract the Brownian diffusion coefficient (D_0) using the data at $[\text{H}_2\text{O}_2]_{\text{vol}} = 0\%$.
- c. Calculate the MSD at all concentrations.
- d. Extract u and τ . How do they change with $[\text{H}_2\text{O}_2]_{\text{vol}}$?
- e. Compare τ to the free rotational diffusivity of the particles and comment on the result.
- f. Determine L and D_{eff} . How “hot” are the active particles?
- g. Discuss why active colloids are different from “hot” colloids, i.e., colloids in a bath that is at high temperature.

For each of the former points, we provide here a guide.

²Beware that few typos in the expression of the MSD of self-propelling particles have regrettably propagated in the active matter community.

$[H_2O_2]_{vol}$	$u (\mu m s^{-1})$	$\tau (s)$	$L (\mu m)$	$D_{eff} (\mu m^2 s^{-1})$	$T_{eff} (K)$
0.1 %	0.4	14	5.6	1.28	≈ 2000
2.0 %	8.8	16	140	619	$\approx 10^6$

Table 8.1: **Measured parameters.** Swimming velocity, reorientation time, persistence length, effective diffusion, and effective temperature at different concentrations for the active Brownian motion of catalytic particles at $[H_2O_2]_{vol} = 0.1\%$ and $[H_2O_2]_{vol} = 2\%$.

a. Track and trace the particles

Particles can be tracked using the methods described in Ref. [56] and routines that are freely available online³. Alternatively, it is sufficient to convert each frame into a binary image by applying a brightness threshold so that the bright pixels occupied by the particles take a value 1 (white), whereas the dark background pixels become 0 (black); the centers of mass of the white areas correspond to the positions of the particles. Trajectories are then reconstructed by connecting the closest positions in all consecutive frames. Three examples of particle trajectories are shown as insets in Fig. 8.17.

b. Extract the Brownian diffusion coefficient (D_0) using the data at $[H_2O_2]_{vol} = 0\%$

In the absence of “fuel”, D_0 is measured from the two-dimensional MSD (Fig. 8.17a), i.e.,

$$MSD(t) = \langle (x(t_0 + t) - x(t_0))^2 + (y(t_0 + t) - y(t_0))^2 \rangle, \quad (8.28)$$

where $x(t_0)$ and $y(t_0)$ are the coordinates of the particle at a given initial time, t_0 . The average takes into account all values of t_0 and trajectories available. Fitting Eq. (8.25) (Fig. 8.17a, dashed line) to the experimental data yields $D_0 \approx 0.16 \mu m^2 s^{-1}$.

c. Calculate the MSD at all concentrations

Figs. 8.17b and 8.17c show the MSD of the coated particles for $[H_2O_2]_{vol} = 0.1\%$ and $[H_2O_2]_{vol} = 2\%$, respectively. The data exhibit a quadratic increase for short time lags and grow linearly in the long-time limit, in agreement with the discussion in Section 8.4.2. The colloids become more mobile at higher peroxide concentrations. Note that, at the largest $[H_2O_2]_{vol}$, the particles do not stay in the field of view for the whole duration of the video, so the long-time statistics is limited.

d. Extract u and τ

The swimming velocity and the characteristic reorientation time are measured by fitting Eq. (8.27) to the MSD data (Figs. 8.17bc, dashed lines), where D_0 is the diffusion coefficient obtained without fuel. The resulting values of u and τ are shown in Figs. 8.17b ($[H_2O_2]_{vol} = 0.1\%$) and 8.17c ($[H_2O_2]_{vol} = 2\%$), and in Table 8.1. Alternatively, u and τ can be extracted separately from the short-time and long-time approximation of the MSD, respectively. While u increases with $[H_2O_2]_{vol}$, τ remains almost constant. This confirms that the reorientation time is not determined by the specific propulsion method (see Section 8.4.2).

e. Compare τ to the free rotational diffusivity of the particles and comment on the result

Instead, the free rotational diffusion time of particles of radius $R = 1.4 \mu m$ in water ($\eta = 1 \text{ mPa} \cdot \text{s}$) is

$$\tau_R = D_R^{-1} = \frac{8\pi\eta R^3}{k_B T} \approx 17 \text{ s}. \quad (8.29)$$

This value is slightly larger than the measured reorientation time. We should first notice that the particles rotate only in the plane parallel to the substrate. A full 3D rotation would in fact lead to a non-constant

³For example, Matlab codes and tutorials can be found at <http://site.physics.georgetown.edu/matlab/code.html>, IDL routines are available at <http://www.physics.emory.edu/faculty/weeks/idl/>, and a Python package can be accessed at <http://soft-matter.github.io/trackpy/v0.4.2/>.

swimming speed and/or particles moving out of focus whenever the cap is pointing upwards or downwards. In addition, the velocity vector might slightly tilted towards the underlying surface causing effective 2D rotations along circles of radii smaller than R [46].

f. Determine L and D_{eff}

The persistence length, $L = u \cdot \tau$, and the effective diffusion coefficient, $D_{\text{eff}} = D_0 + (1/2)u^2\tau$, are reported in the fourth and fifth column of Table 8.1. It is possible to associate to each effective diffusion coefficient, D_{eff} , an effective temperature, T_{eff} (see Table 8.1, last column), via the relation $D_{\text{eff}} = k_B T_{\text{eff}} / (6\pi\eta R)$. This description has been used to explain phenomena such as sedimentation profiles [42] or nonequilibrium equation of states [57], but becomes misleading if not handled with care (see below).

g. Discuss why active colloids are different from hot colloids

The analogy between active particles and *hot* colloids is tempting but fails on many level. First, the temperature appears in both the translational ($D = k_B T / (6\pi\eta R)$) and rotational ($D_R = k_B T / (8\pi\eta R^3)$) diffusion coefficients, but only the translational motion is effectively hot; rotation remains governed by Brownian diffusion at room temperature. This leads to the paradox that two degrees of freedom belonging to the same particle are at different temperatures. Second, consider a hot colloid and a self-propelling particle in the proximity of a physical confinement such as a circular well. While the Brownian particle explores more space as the temperature increases, the active colloid bumps against the wall and remains there for times of the order of τ ; hence, active particles with a large effective temperature (i.e., large τ) becomes more confined. Similarly, the formation of clusters of active particles observed at large effective temperature defeats the understanding that a colloidal system melts as the temperature increases.

Bibliography

- [1] D. Mattia, H. Leese, and K. P. Lee. Carbon nanotube membranes: from flow enhancement to permeability. *Journal of Membrane Science*, 475:266–272, 2015.
- [2] E. Secchi, S. Marbach, A. Niguès, D. Stein, A. Siria, and L. Bocquet. Massive radius-dependent flow slippage in carbon nanotubes. *Nature*, 537:210–213, 2016.
- [3] J. K. Holt, H. G. Park, Y. Wang, M. Stadermann, A. B. Artyukhin, C. P. Grigoropoulos, A. Noy, and O. Bakajin. Fast mass transport through sub-2-nanometer carbon nanotubes. *Science*, 312:1034–1037, 2006.
- [4] T. Jovanovic-Talisman, J. Tetenbaum-Novatt, A. S. McKenney, A. Zilman, R. Peters, M. P. Rout, and B. T. Chait. Artificial nanopores that mimic the transport selectivity of the nuclear pore complex. *Nature*, 457:1023–1027, 2009.
- [5] K. Murata, K. Mitsuoka, T. Hirai, T. Walz, P. Agre, J. B. Heymann, A. Engel, and Y. Fujiyoshi. Structural determinants of water permeation through aquaporin-1. *Nature*, 407:599–605, 2000.
- [6] H. Sui, B.-G. Han, J. K. Lee, P. Walian, and B. K. Jap. Structural basis of water-specific transport through the aqp1 water channel. *Nature*, 414:872–878, 2001.
- [7] L. Bocquet and E. Charlaix. Nanofluidics, from bulk to interfaces. *Chemical Society Reviews*, 39:1073–1095, 2010.
- [8] N. Kavokine, R. R. Netz, and L. Bocquet. Fluids at the nanoscale: From continuum to subcontinuum transport. *Annual Review of Fluid Mechanics*, 53, 2020.
- [9] D. M. Huang, C. Sendner, D. Horinek, R. R. Netz, and L. Bocquet. Water slippage versus contact angle: A quasiuniversal relationship. *Physical Review Letters*, 101:226101, 2008.
- [10] I. W. Hamley. *Introduction to soft matter: synthetic and biological self-assembling materials*. John Wiley & Sons, 2013.

- [11] J. L. Anderson. Colloid transport by interfacial forces. *Annual Review of Fluid Mechanics*, 21:61–99, 1989.
- [12] T. M. Squires and M. Z. Bazant. Induced-charge electro-osmosis. *Journal of Fluid Mechanics*, 509:217–252, 2004.
- [13] M. Z. Bazant and T. M. Squires. Induced-charge electrokinetic phenomena: theory and microfluidic applications. *Physical Review Letters*, 92:066101, 2004.
- [14] H. A. Stone, A. D. Stroock, and A. Ajdari. Engineering flows in small devices: microfluidics toward a lab-on-a-chip. *Annual Review of Fluid Mechanics*, 36:381–411, 2004.
- [15] A. Ajdari. Electro-osmosis on inhomogeneously charged surfaces. *Physical Review Letters*, 75:755, 1995.
- [16] J. L. Anderson and D. C. Prieve. Diffusiophoresis caused by gradients of strongly adsorbing solutes. *Langmuir*, 7:403–406, 1991.
- [17] B. Abécassis, C. Cottin-Bizonne, C. Ybert, A. Ajdari, and L. Bocquet. Boosting migration of large particles by solute contrasts. *Nature Materials*, 7:785–789, 2008.
- [18] N. Shi, R. Nery-Azevedo, A. I Abdel-Fattah, and T. M. Squires. Diffusiophoretic focusing of suspended colloids. *Physical Review Letters*, 117:258001, 2016.
- [19] J. Palacci, C. Cottin-Bizonne, C. Ybert, and L. Bocquet. Osmotic traps for colloids and macromolecules based on logarithmic sensing in salt taxes. *Soft Matter*, 8:980–994, 2012.
- [20] J. Burelbach and H. Stark. Determining phoretic mobilities with Onsager’s reciprocal relations: Electro- and thermophoresis revisited. *The European Physical Journal E*, 42:1–9, 2019.
- [21] D. Vigolo, R. Rusconi, H. A. Stone, and R. Piazza. Thermophoresis: microfluidics characterization and separation. *Soft Matter*, 6:3489–3493, 2010.
- [22] S. Duhr and D. Braun. Two-dimensional colloidal crystals formed by thermophoresis and convection. *Applied Physics Letters*, 86:131921, 2005.
- [23] N. Bruot and H. Tanaka. Externally driven local colloidal ordering induced by a pointlike heat source. *Physical Review Research*, 1:033200, 2019.
- [24] A. P. Bregulla, A. Würger, K. Günther, M. Mertig, and F. Cichos. Thermo-osmotic flow in thin films. *Physical Review Letters*, 116:188303, 2016.
- [25] R. Piazza and A. Parola. Thermophoresis in colloidal suspensions. *Journal of Physics: Condensed Matter*, 20:153102, 2008.
- [26] R. Rusconi, L. Isa, and R. Piazza. Thermal-lensing measurement of particle thermophoresis in aqueous dispersions. *Journal of the Optical Society of America B*, 21:605–616, 2004.
- [27] G.-R. Yi, D. J. Pine, and S. Sacanna. Recent progress on patchy colloids and their self-assembly. *Journal of Physics: Condensed Matter*, 25:193101, 2013.
- [28] S. Jiang, Q. Chen, M. Tripathy, E. Luijten, K. S. Schweizer, and S. Granick. Janus particle synthesis and assembly. *Advanced Materials*, 22:1060–1071, 2010.
- [29] J. Zhang, B. A. Grzybowski, and S. Granick. Janus particle synthesis, assembly, and application. *Langmuir*, 33:6964–6977, 2017.
- [30] J. L. Moran and J. D. Posner. Electrokinetic locomotion due to reaction-induced charge auto-electrophoresis. *Journal of Fluid Mechanics*, 680:31, 2011.
- [31] T. R. Kline, W. F. Paxton, T. E. Mallouk, and A. Sen. Catalytic nanomotors: Remote-controlled autonomous movement of striped metallic nanorods. *Angewandte Chemie International Edition*, 44:744–746, 2005.

- [32] A. Aubret, S. Ramananarivo, and J. Palacci. Eppur si muove, and yet it moves: Patchy (phoretic) swimmers. *Current opinion in colloid & interface science*, 30:81–89, 2017.
- [33] S. Gangwal, O. J. Cayre, M. Z. Bazant, and O. D. Velev. Induced-charge electrophoresis of metallodielectric particles. *Physical Review Letters*, 100:058302, 2008.
- [34] F. Ma, X. Yang, H. Zhao, and N. Wu. Inducing propulsion of colloidal dimers by breaking the symmetry in electrohydrodynamic flow. *Physical Review Letters*, 115:208302, 2015.
- [35] S. Ni, E. Marini, I. Buttinoni, H. Wolf, and L. Isa. Hybrid colloidal microswimmers through sequential capillary assembly. *Soft Matter*, 13:4252–4259, 2017.
- [36] M. A. Fernandez-Rodriguez, F. Grillo, L. Alvarez, M. Rathlef, I. Buttinoni, G. Volpe, and L. Isa. Feedback-controlled active Brownian colloids with space-dependent rotational dynamics. *Nature Communications*, 11:1–10, 2020.
- [37] W. D. Ristenpart, I. A. Aksay, and D. A. Saville. Electrohydrodynamic flow around a colloidal particle near an electrode with an oscillating potential. *Journal of Fluid Mechanics*, 575:83, 2007.
- [38] W. D. Ristenpart, I. A. Aksay, and D. A. Saville. Assembly of colloidal aggregates by electrohydrodynamic flow: Kinetic experiments and scaling analysis. *Physical Review E*, 69:021405, 2004.
- [39] J. Yan, M. Han, J. Zhang, C. Xu, E. Luijten, and S. Granick. Reconfiguring active particles by electrostatic imbalance. *Nature materials*, 15:1095–1099, 2016.
- [40] M. N. van der Linden, L. C. Alexander, D. G. A. L. Aarts, and O. Dauchot. Interrupted motility induced phase separation in aligning active colloids. *Physical Review Letters*, 123:098001, 2019.
- [41] R. Golestanian, T. B. Liverpool, and A. Ajdari. Propulsion of a molecular machine by asymmetric distribution of reaction products. *Physical Review Letters*, 94:220801, 2005.
- [42] J. Palacci, C. Cottin-Bizonne, C. Ybert, and L. Bocquet. Sedimentation and effective temperature of active colloidal suspensions. *Physical Review Letters*, 105:088304, 2010.
- [43] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. Self-motile colloidal particles: from directed propulsion to random walk. *Physical Review Letters*, 99:048102, 2007.
- [44] K. Dietrich, G. Volpe, M. N. Sulaiman, D. Renggli, I. Buttinoni, and L. Isa. Active atoms and interstitials in two-dimensional colloidal crystals. *Physical Review Letters*, 120:268004, 2018.
- [45] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339:936–940, 2013.
- [46] K. Dietrich, D. Renggli, M. Zanini, G. Volpe, I. Buttinoni, and L. Isa. Two-dimensional nature of the active Brownian motion of catalytic microswimmers at solid and liquid interfaces. *New Journal of Physics*, 19:065008, 2017.
- [47] A. Brown and W. Poon. Ionic effects in self-propelled Pt-coated Janus swimmers. *Soft matter*, 10:4016–4027, 2014.
- [48] S. Ketzeltzi, J. de Graaf, R. P. Doherty, and D. J. Kraft. Slip length dependent propulsion speed of catalytic colloidal swimmers near walls. *Physical Review Letters*, 124:048002, 2020.
- [49] H.-R. Jiang, N. Yoshinaga, and M. Sano. Active motion of a Janus particle by self-thermophoresis in a defocused laser beam. *Physical Review Letters*, 105:268302, 2010.
- [50] A. P. Bregulla, H. Yang, and F. Cichos. Stochastic localization of microswimmers by photon nudging. *ACS Nano*, 8:6542–6550, 2014.
- [51] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger. Active Brownian motion tunable by light. *Journal of Physics: Condensed Matter*, 24:284129, 2012.

- [52] G. Volpe, I. Buttinoni, D. Vogt, H.-J. Kümmerer, and C. Bechinger. Microswimmers in patterned environments. *Soft Matter*, 7:8810–8815, 2011.
- [53] K. Dietrich, N. Jaensson, I. Buttinoni, G. Volpe, and L. Isa. Microscale marangoni surfers. *Physical Review Letters*, 125:098001, 2020.
- [54] X. Zheng, B. ten Hagen, A. Kaiser, M. Wu, H. Cui, Z. Silber-Li, and H. Löwen. Non-Gaussian statistics for the motion of self-propelled Janus particles: Experiment versus theory. *Physical Review E*, 88:032304, 2013.
- [55] H. Löwen. Inertial effects of self-propelled particles: From active Brownian to active Langevin motion. *The Journal of chemical physics*, 152:040901, 2020.
- [56] J. C. Crocker and D. G. Grier. Methods of digital video microscopy for colloidal studies. *Journal of Colloid and Interface Science*, 179:298–310, 1996.
- [57] F. Ginot, I. Theurkauff, D. Levis, C. Ybert, L. Bocquet, L. Berthier, and C. Cottin-Bizonne. Nonequilibrium equation of state in suspensions of active colloids. *Physical Review X*, 5:011004, 2015.

Chapter 9

Dynamical Density Functional Theory

HARTMUT LÖWEN

In the last 50 years, equilibrium *density functional theory* (DFT) has been proven to be a powerful, versatile, and predictive approach for the statics and structure of classical particles. DFT relies on the existence of a functional of the one-particle density that gives access to the equilibrium thermodynamics when it is minimized with respect to this density. This theory can be extended to the nonequilibrium dynamics of completely overdamped Brownian colloidal particles in the so-called *dynamical density functional theory* (DDFT). The success of DDFT makes it a promising candidate for a first-principle description of active matter. In this Chapter, we shall first recapitulate classical DDFT for passive colloidal particles typically described by Smoluchowski equation. After a basic derivation of DDFT from the Smoluchowski equation, we discuss orientational degrees of freedom and the effect of hydrodynamic interactions for passive particles. This brings us into an ideal position to generalize DDFT to active matter. In particular we distinguish between *dry active matter*, which is composed of self-propelled particles that contain no hydrodynamic flow effects of a surrounding solvent, and *wet active matter*, where the hydrodynamic flow fields generated by the microswimmers are taken into account. For the latter, DDFT is a tool that unifies thermal fluctuations, direct particle interactions, external driving fields, and hydrodynamic effects arising from internal self-propulsion discriminating between *pushers*

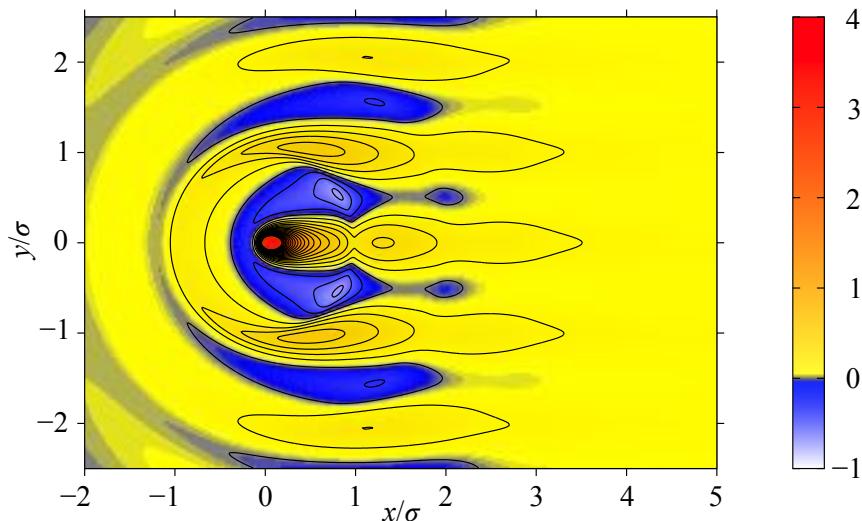


Figure 9.1: **Inhomogenous density field obtained with DFT.** Example for a two-dimensional inhomogenous density field as a contour plot in the xy -plane, see color bar. An initial strongly correlated density peak is exposed to a tilting external potential and therefore flows from left to right making the initial density profile anisotropic. σ is an arbitrary length scale. Reproduced from [1].

and *pullers*. Several recent examples relevant for microswimmers have been explored with DDFT, ranging from the formation of a hydrodynamic pump in confined systems, to collections of circle swimmers and binary mixtures of pushers and pullers. For parts of this Chapter, we follow the ideas outlined in Ref. [2]. For more technical aspects, we refer to the recent review in Ref. [3].

9.1 Density functional theory in equilibrium

Density functional theory is a cornerstone in classical statistical mechanics of inhomogeneous systems. We first briefly present the basics of the traditional theory at equilibrium.

9.1.1 Basics

We shall now consider DFT for classical systems at finite temperature interacting via a radially-symmetric pair-wise potential $v(r)$. The basic variational principle of DFT establishes the existence of a unique grand canonical free-energy-density functional $\Omega(T, \mu_p, [\rho])$, which gets minimal for the equilibrium density $\rho_0(\mathbf{r})$. The latter measures the probability density to find any particle at position \mathbf{r} in equilibrium. Here the equilibrium condition is denoted with the subscript 0. The notation $[\rho]$ means a *functional*. A functional takes a function, such as a general density field $\rho(\mathbf{r})$, as an input and produces a real number as an output. According to the basic variational theorem, at the minimum, the free-energy functional coincides with the real equilibrium. The minimum condition reads as follows:

$$\frac{\delta\Omega(T, \mu_p, [\rho])}{\delta\rho(\mathbf{r})} \Big|_{\rho(\mathbf{r})=\rho_0(\mathbf{r})} = 0, \quad (9.1)$$

where T is the imposed temperature and μ_p the prescribed chemical potential of the system. However, the functional $\Omega(T, \mu_p, [\rho])$ is not known explicitly, in general. One can split the functional $\Omega(T, \mu_p, [\rho])$ as

$$\Omega(T, \mu_p, [\rho]) = \mathcal{F}(T, [\rho]) + \int_V d^3\mathbf{r} \rho(\mathbf{r}) (V_{\text{ext}}(\mathbf{r}) - \mu_p), \quad (9.2)$$

where $\mathcal{F}(T, [\rho])$ is a Helmholtz free energy functional and V denotes the system volume. The knowledge of the functional $\mathcal{F}(T, [\rho])$ for a given pair potential $v(r)$ provides a lot of information (much more than just a bulk equation of state, for instance) since it can be applied to any inhomogeneous system in an external potential $V_{\text{ext}}(\mathbf{r})$. For example, the second functional derivative taken in the homogeneous bulk limit is proportional to the direct fluid pair correlation function.

9.1.2 Approximations for the density functional

Let us first recall the exact functional for the ideal gas where the pair interaction $v(r)$ between the particles vanishes, i.e., $v(r) = 0$. In three spatial dimensions, this functional reads as

$$\mathcal{F}(T, [\rho]) = \mathcal{F}_{\text{id}}(T, [\rho]) = k_B T \int_V d^3\mathbf{r} \rho(\mathbf{r}) [\ln(\rho(\mathbf{r})\Lambda^3) - 1], \quad (9.3)$$

where Λ is the thermal wavelength and k_B is the Boltzmann constant. Actually the value of the thermal wavelength is irrelevant since it only provides a trivial shift in the chemical potential which does not change the physics. In this case, the minimization condition

$$0 = \frac{\delta\Omega}{\delta\rho(\mathbf{r})} \Big|_0 = k_B T \ln(\rho(\mathbf{r})\Lambda^3) + V_{\text{ext}}(\mathbf{r}) - \mu_p \quad (9.4)$$

leads to the generalized barometric law

$$\rho_0(\mathbf{r}) = \frac{1}{\Lambda^3} \exp\left(-\frac{V_{\text{ext}}(\mathbf{r}) - \mu_p}{k_B T}\right) \quad (9.5)$$

for the inhomogeneous density.

For non-vanishing pair interactions $v(r)$, one can split

$$\mathcal{F}(T, [\rho]) = \mathcal{F}_{\text{id}}(T, [\rho]) + \mathcal{F}_{\text{exc}}(T, [\rho]), \quad (9.6)$$

which defines a so-called *excess* free energy density functional $\mathcal{F}_{\text{exc}}(T, [\rho])$ that typically needs to be approximated. One important approximation is the *mean-field approximation*, where

$$\mathcal{F}_{\text{exc}}(T, [\rho]) \approx \frac{1}{2} \int d^3r \int d^3r' v(|\mathbf{r} - \mathbf{r}'|) \rho(\mathbf{r}) \rho(\mathbf{r}'). \quad (9.7)$$

Other approximations are the perturbative Ramakrishnan-Yussouff approach and the non-perturbative fundamental measure theory for steric interactions [4, 5]. Similar approximations can be formulated in two spatial dimensions, e.g., for hard disks [6]. An example for an inhomogeneous density field as obtained with DFT is shown in Fig. 9.1.

9.2 Classical dynamical density functional theory

To introduce the main concepts and methods, we consider nonequilibrium situations for the dynamics of passive Brownian particles.

9.2.1 Brownian dynamics and Smoluchowski equation

DFT can be made time-dependent for passive overdamped Brownian particles [3] leading to *dynamical density functional theory* (DDFT), where the time-dependent density field is the central quantity. It will follow a deterministic diffusion-like equation.

Non-interacting Brownian particles

For non-interacting particles with an inhomogeneous time-dependent particle density $\rho(\mathbf{r}, t)$, Fick's law for the current density $\mathbf{j}(\mathbf{r}, t)$ states

$$\mathbf{j}(\mathbf{r}, t) = -D \nabla \rho(\mathbf{r}, t), \quad (9.8)$$

where D is a phenomenological diffusion coefficient.

The continuity equation of particle number conservation,

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{j}(\mathbf{r}, t) = 0, \quad (9.9)$$

then leads to the well-known diffusion equation for $\rho(\mathbf{r}, t)$:

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = D \Delta \rho(\mathbf{r}, t). \quad (9.10)$$

In the presence of an external potential $V_{\text{ext}}(\mathbf{r}, t)$, the force $\mathbf{F} = -\nabla V_{\text{ext}}(\mathbf{r}, t)$ acts on the particles and will induce a drift velocity \mathbf{v}_D giving rise to the additional current density $\mathbf{j}_D = \rho \mathbf{v}_D$ with the drift velocity $\mathbf{v}_D = \frac{\mathbf{F}}{\mu} = -\frac{1}{\mu} \nabla V_{\text{ext}}(\mathbf{r}, t)$. Here, μ denotes the friction coefficient (for a sphere of radius R in a viscous solvent of viscosity η , Stokes law tells us that $\mu = 6\pi\eta R$). With the Stokes-Einstein relation $D = \frac{k_B T}{\mu}$, we get $\mathbf{j} = -\frac{1}{\mu} (k_B T \nabla \rho + \rho \nabla V_{\text{ext}})$ and the continuity equation yields

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = \frac{1}{\mu} (k_B T \Delta \rho(\mathbf{r}, t) + \nabla \cdot (\rho(\mathbf{r}, t) \nabla V_{\text{ext}}(\mathbf{r}, t))), \quad (9.11)$$

which is the *Smoluchowski equation* for non-interacting particles. Note that the external force can even be time-dependent.

Interacting Brownian particles

Now, we consider N interacting particles at positions \mathbf{r}_i ($i = 1, \dots, N$). The total potential energy is

$$U_{\text{tot}}(\mathbf{r}^N, t) = \sum_{i=1}^N V_{\text{ext}}(\mathbf{r}_i, t) + \sum_{\substack{i,j=1 \\ i < j}}^N v(|\mathbf{r}_i - \mathbf{r}_j|). \quad (9.12)$$

With $\mathbf{r}^N = \{\mathbf{r}_i \mid i = 1, \dots, N\}$, the generalization of the Smoluchowski equation [7] for the joint probability density $p(\mathbf{r}^N, t)$ reads as

$$\frac{\partial p}{\partial t} = \hat{\mathcal{O}} p = \frac{1}{\mu} \sum_{i=1}^N \nabla_i \cdot [k_B T \nabla_i + \nabla_i U_{\text{tot}}(\mathbf{r}^N, t)] p, \quad (9.13)$$

where the operator $\hat{\mathcal{O}}$ is called *Smoluchowski operator*.

9.2.2 Derivation

In the following we shall derive DDFT by using two possible routes: first a phenomenological route, and second a microscopic route based on the Smoluchowski equation.

Phenomenological derivation of DDFT

The general Fick's law assumes that the particle current density is proportional to the gradient of the chemical potential [7] and proportional to the time-dependent density $\rho(\mathbf{r}, t)$:

$$\mathbf{j} = \mu \rho(\mathbf{r}, t) \nabla \mu_p. \quad (9.14)$$

In equilibrium, when the chemical potential is constant, there is no such current. We now take a functional derivative with respect to the density in Eq. (9.2) and obtain in the absence of an external potential

$$\left. \frac{\delta \mathcal{F}(T, [\rho])}{\delta \rho(\mathbf{r})} \right|_{\rho(\mathbf{r})=\rho_0(\mathbf{r})} = \mu_p. \quad (9.15)$$

When combining this with the continuity equation of particle number conservation, we get the important DDFT equation:

$$\mu \frac{\partial \rho(\mathbf{r}, t)}{\partial t} = \nabla \rho(\mathbf{r}, t) \nabla \frac{\delta \mathcal{F}[\rho]}{\delta \rho(\mathbf{r}, t)}, \quad (9.16)$$

which is obviously generalized to the presence of an external potential $V_{\text{ext}}(\mathbf{r}, t)$ by replacing $\mathcal{F}[\rho]$ with $\Omega[\rho]$. This is a deterministic time evolution equation for $\rho(\mathbf{r}, t)$. For an ideal gas, it reduces to the exact Smoluchowski equation which is the standard diffusion equation (see Problem 9.2). For an interacting system, the DDFT equation is approximative.

Derivation of DDFT from the Smoluchowski equation

The DDFT equation can be derived from the Smoluchowski equation [8], but one essential additional approximation, the so-called *adiabatic approximation*, needs to be performed here as well. In more detail, one integrates out degrees of freedom from the Smoluchowski equation to obtain the following exact equation

$$\mu \frac{\partial}{\partial t} \rho(\mathbf{r}_1, t) = k_B T \Delta_1 \rho(\mathbf{r}_1, t) + \nabla_1 [\rho(\mathbf{r}_1, t) \nabla_1 V_{\text{ext}}(\mathbf{r}_1, t)] + \nabla_1 \int d^3 \mathbf{r}_2 \rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2, t) \nabla_1 v(|\mathbf{r}_1 - \mathbf{r}_2|). \quad (9.17)$$

In equilibrium, necessarily $\frac{\partial \rho(\mathbf{r}_1, t)}{\partial t} = 0$, which implies

$$0 = \nabla \left(k_B T \nabla \rho(\mathbf{r}) + \rho(\mathbf{r}) \nabla V_{\text{ext}}(\mathbf{r}) + \int d^3 \mathbf{r}' \rho^{(2)}(\mathbf{r}, \mathbf{r}') \nabla v(|\mathbf{r} - \mathbf{r}'|) \right), \quad (9.18)$$

meaning that the divergence of a current density must vanish. The current density itself is imposed to vanish for $r \rightarrow \infty$ in equilibrium and thus the current density is identical to zero everywhere. Therefore:

$$0 = k_B T \nabla \rho(\mathbf{r}) + \rho(\mathbf{r}) \nabla V_{\text{ext}}(\mathbf{r}) + \int d^3 \mathbf{r}' \rho^{(2)}(\mathbf{r}, \mathbf{r}') \nabla v(|\mathbf{r} - \mathbf{r}'|), \quad (9.19)$$

which is also known as *Yvon-Born-Green-relation*. Here, $\rho^{(2)}(\mathbf{r}, \mathbf{r}')$ is the two-body joint probability density in nonequilibrium. We now take a gradient of the density functional derivative of Eq. (9.2) and combine it with the Yvon-Born-Green relation. Then, we obtain:

$$\int d^3 \mathbf{r}' \rho^{(2)}(\mathbf{r}, \mathbf{r}') \nabla V(|\mathbf{r} - \mathbf{r}'|) = \rho(\mathbf{r}) \nabla \frac{\delta \mathcal{F}_{\text{exc}}[\rho]}{\delta \rho(\mathbf{r})}. \quad (9.20)$$

We postulate that this argument holds also in nonequilibrium. In doing so, nonequilibrium correlations are approximated by equilibrium ones at the same $\rho(\mathbf{r}, t)$ (identified via a suitable time-independent $V_{\text{ext}}(\mathbf{r})$ in equilibrium). Equivalently one can say the following [9]: the one-body density is a very slow variable as compared to all higher order correlations such as pair correlations. At a given one-body density, which can be viewed as an effective instantaneous equilibrium density profile via the suitable time-independent potential, all these higher-order correlations decay very quickly to equilibrium such that they follow *adiabatically* the one-body density as a function of time. This assumption is called *adiabatic approximation*. This results finally in the DDFT equation:

$$\mu \frac{\partial \rho(\mathbf{r}, t)}{\partial t} = \nabla \rho(\mathbf{r}, t) \nabla \frac{\delta \Omega[\rho]}{\delta \rho(\mathbf{r}, t)}. \quad (9.21)$$

For further alternate derivations of the DDFT equation, see Refs. [10, 11, 9].

9.3 Polar particles

We now turn to particles with an additional orientational degree of freedom (such as rods, platelets and Janus particles) and will recapitulate the equilibrium and dynamical versions of DFT for this case.

9.3.1 Density functional theory

We consider polar particles that possess an additional rotational degree of freedom in the two-dimensional plane, which can be described by a simple angle θ or a unit vector $\hat{\mathbf{n}} = (\cos \theta, \sin \theta)$ relative to a prescribed axis. Having applications to swimmers on a substrate in mind, we consider motion in two-dimensions only. Equilibrium DFT can readily be extended to polar particles. A configuration of N particles is now fully specified by the set of positions of their centers of mass and their corresponding orientations $\{\mathbf{r}_i, \hat{\mathbf{n}}_i, i = 1, \dots, N\}$. Pairwise interactions are described by a pair-potential $v(\mathbf{r}_i - \mathbf{r}_j, \hat{\mathbf{n}}_i, \hat{\mathbf{n}}_j)$ that depends on the difference vector $\mathbf{r}_i - \mathbf{r}_j$ between the centers of the particle i and another particle j plus their two orientations $\hat{\mathbf{n}}_i$ and $\hat{\mathbf{n}}_j$. In the general context of active matter, if this function *only* depends on $\mathbf{r}_i - \mathbf{r}_j$, the interactions are called *non-aligning*. An example is provided by spherical self-propelled Janus particles, which do not change their orientation when bouncing onto each other. If it is energetically favorable to have parallel orientations, the interactions are called *aligning*. In the rare case that neighboring particle tend to stay anti-parallel, these interactions are called *anti-aligning*. Clearly the external potential $V_{\text{ext}}(\mathbf{r}, \hat{\mathbf{n}}, t)$ can also depend on the particle's orientation.

As in the case of spherical particles, DFT establishes the existence of a functional of the one-particle density $\rho(\mathbf{r}, \hat{\mathbf{n}})$, which gets minimal in equilibrium

$$\left. \frac{\delta \Omega(T, \mu_p, [\rho])}{\delta \rho(\mathbf{r}, \hat{\mathbf{n}})} \right|_{\rho=\rho(\mathbf{r}, \hat{\mathbf{n}})} = 0. \quad (9.22)$$

Again, the functional can be decomposed as follows

$$\begin{aligned}\Omega(T, \mu_p, [\rho]) &= \underbrace{k_B T \int d^2 \mathbf{r} \int_0^{2\pi} d\theta \rho(\mathbf{r}, \hat{\mathbf{n}}) [\ln(\Lambda^2 \rho(\mathbf{r}, \hat{\mathbf{n}})) - 1]}_{\mathcal{F}_{id}[\rho^{(1)}]} \\ &\quad + \underbrace{\int d^2 \mathbf{r} \int_0^{2\pi} d\theta (V_{ext}(\mathbf{r}, \hat{\mathbf{n}}) - \mu_p) \rho(\mathbf{r}, \hat{\mathbf{n}}) + \mathcal{F}_{exc}(T, [\rho])}_{\mathcal{F}_{exc}(T, [\rho^{(1)}])}.\end{aligned}\quad (9.23)$$

The first term on the right hand side of Eq. (9.23) is the functional $\mathcal{F}_{id}[\rho^{(1)}]$ for ideal rotators. The excess part $\mathcal{F}_{exc}(T, [\rho^{(1)}])$ is in general unknown and requires approximative treatments. Again nonperturbative fundamental measure theory for hard cylinders is available [12].

9.3.2 Dynamic density functional theory

The Smoluchowski equation for the joint probability density $p(\mathbf{r}_1, \dots, \mathbf{r}_N; \hat{\mathbf{n}}_1, \dots, \hat{\mathbf{n}}_N, t) = p(\mathbf{r}^N, \hat{\mathbf{n}}^N, t)$ is

$$\frac{\partial p}{\partial t} = \hat{\mathcal{O}}_S p, \quad (9.24)$$

with the Smoluchowski operator

$$\hat{\mathcal{O}}_S = \sum_{i=1}^N \left[\nabla_{\mathbf{r}_i} \cdot \bar{\tilde{D}}(\hat{\mathbf{n}}_i) \cdot \left(\nabla_{\mathbf{r}_i} + \frac{1}{k_B T} \nabla_{\mathbf{r}_i} U(\mathbf{r}^N, \hat{\mathbf{n}}^N, t) \right) + D_R \hat{\mathbf{R}}_i \cdot \left(\hat{\mathbf{R}}_i + \frac{1}{k_B T} \hat{\mathbf{R}}_i U(\mathbf{r}^N, \hat{\mathbf{n}}^N, t) \right) \right], \quad (9.25)$$

where $U(\mathbf{r}^N, \hat{\mathbf{n}}^N, t)$ is the total potential energy. Here, the rotation operator $\hat{\mathbf{R}}_i$ is defined as $\hat{\mathbf{R}}_i = \partial / \partial \theta$ and the anisotropic translational diffusion tensor is given by

$$\bar{\tilde{D}}(\hat{\mathbf{n}}_i) = D^\parallel \hat{\mathbf{n}}_i \otimes \hat{\mathbf{n}}_i + D^\perp (\mathbf{1} - \hat{\mathbf{n}}_i \otimes \hat{\mathbf{n}}_i). \quad (9.26)$$

The two diffusion constants D^\parallel and D^\perp , parallel and perpendicular to the orientations, reflect the fact that the translational diffusion is anisotropic. The quantity D_R is called rotational diffusion constant and sets the Brownian dynamics of the orientations. Integrating the Smoluchowski equation yields the following DDFT equation for the time-dependent $\rho(\mathbf{r}, \theta, t)$ [13]:

$$\frac{\partial \rho(\mathbf{r}, \theta, t)}{\partial t} = \nabla_{\mathbf{r}} \cdot \bar{\tilde{D}}(\theta) \cdot \left[\rho(\mathbf{r}, \theta, t) \nabla_{\mathbf{r}} \frac{\delta \Omega[\rho(\mathbf{r}, \theta, t)]}{\delta \rho(\mathbf{r}, \theta, t)} \right] + D_R \frac{\partial}{\partial \theta} \left[\rho(\mathbf{r}, \theta, t) \frac{\partial}{\partial \theta} \frac{\delta \Omega[\rho(\mathbf{r}, \theta, t)]}{\delta \rho(\mathbf{r}, \theta, t)} \right]. \quad (9.27)$$

9.4 Dynamical density functional theory for dry active matter

Ignoring hydrodynamic interactions, micrswimmers can simply be modeled by polar particles driven by a constant internal effective force along their orientations [14]; this force corresponds to an effective drift velocity and mimics the actual propulsion mechanism. On top of this intrinsic propulsion, the particles are affected by the Brownian noise of the solvent. The corresponding motion is intrinsically a nonequilibrium one and even the dynamics of a single Brownian swimmer was solved only in this century [15, 16, 17].

For dry active matter, the many-body Smoluchowski equation now reads

$$\frac{\partial p}{\partial t} = \hat{\mathcal{O}}_a p, \quad (9.28)$$

with the *active* Smoluchowski operator

$$\hat{\mathcal{O}}_a = \hat{\mathcal{O}}_S + \sum_{i=1}^N \left[\nabla_{\mathbf{r}_i} \cdot \bar{\tilde{D}}(\hat{\mathbf{n}}_i) \cdot \left(\frac{1}{k_B T} v_0 \hat{\mathbf{n}}_i \right) \right]. \quad (9.29)$$

The active part involves a particle current along its orientation with a strength proportional to v_0 , which is the self-propulsion velocity a single particle assumes. This Smoluchowski equation is stochastically equivalent to the Langevin equations of active Brownian motion [18]. For ideal particles ($U(\mathbf{r}^N, \hat{\mathbf{n}}^N, t) = 0$), the active Smoluchowski equation has been the starting point to calculate the intermediate scattering function of an active Brownian particle [19].

DDFT for dry active matter can be derived using the same adiabatic approximation, Eq. (9.20), as in the passive case. The resulting equation of motion for the one-particle density $\rho(\mathbf{r}, \theta, t)$ then reads [20]:

$$\begin{aligned} k_B T \frac{\partial \rho(\mathbf{r}, \theta, t)}{\partial t} &= \nabla_{\mathbf{r}} \cdot \bar{\mathcal{D}}(\theta) \cdot \left[k_B T v_0 \hat{\mathbf{n}} \rho(\mathbf{r}, \theta, t) + \rho(\mathbf{r}, \theta, t) \nabla_{\mathbf{r}} \frac{\delta \Omega[\rho(\mathbf{r}, \theta, t)]}{\delta \rho(\mathbf{r}, \theta, t)} \right] + \\ &\quad + D_R \frac{\partial}{\partial \theta} \left[\rho(\mathbf{r}, \theta, t) \frac{\partial}{\partial \theta} \frac{\delta \Omega[\rho(\mathbf{r}, \theta, t)]}{\delta \rho(\mathbf{r}, \theta, t)} \right]. \end{aligned} \quad (9.30)$$

For a non-interacting system, it is important to note here that this equation is exact under self-propulsion and any external forces. It is therefore an ideal starting point to study a single active Brownian particle under gravity [21, 22, 23, 24]. As a result, polar order was discovered for an ideal gas of sedimentating particles in the steady state even if the particles are not bottom-heavy.

As an application, for active particles in a channel with aligning interactions, the time-dependent density profiles were found to be in agreement with Brownian dynamics computer simulations [20]. In Ref. [20], a crude mean-field Onsager-like density functional approximation [25] was used. Qualitatively, a transient formation of hedgehog-like clusters near the channel boundaries was found in simulations and reproduced by DDFT.

Finally we remark that DDFT was generalized towards three spatial dimensions for swimmers of arbitrary shape with complicated friction tensors [26]. Moreover, superadiabatic DDFT, which goes beyond the adiabatic approximation, has been applied to active Brownian systems with non-aligning interactions [27, 28]. A special application was performed for motility-induced phase separations of active particles [27, 29].

9.5 Dynamical density functional theory for wet active matter

The most general DDFT framework for microswimmers can be found in Ref. [30], which we closely follow here. This approach includes simultaneously thermal fluctuations, external forces, interparticle interactions by body forces and hydrodynamic interactions, and self-propulsion effects. In principle, it includes all previous cases in the special limits of vanishing self-propulsion (passive particle) and dry active matter (neglected hydrodynamic interactions).

9.5.1 The swimmer model

To derive the DDFT, we consider a dilute suspension of N identical self-propelled microswimmers at low Reynolds number [31] in two dimensions in an unbounded three-dimensional fluid. Following Ref. [30], the self-propulsion of a microswimmer is concatenated to self-induced fluid flows in the surrounding medium. This then represents a major source of hydrodynamic interaction between different microswimmers. To proceed we consider a minimal model for an individual microswimmer, as depicted in Fig. 9.2.

Each microswimmer consists of a spherical body of hydrodynamic radius a . The swimmer body is subjected to hydrodynamic drag with respect to surrounding fluid flows including self-convection. The latter is generated by two active force centers located at a distance L from each other (Fig. 9.2), and exert two antiparallel forces, $+f$ and $-f$ respectively, onto the surrounding fluid setting it into motion. Summing up the two forces, we find that the microswimmer exerts a vanishing net force onto the fluid. Moreover, since $f \parallel \hat{\mathbf{n}}$, there is no net active torque [32]. Self-propulsion is now achieved by shifting the swimmer body along $\hat{\mathbf{n}}$ out of the geometric center. We introduce a parameter α to quantify this shift (Fig. 9.2). The distances between the body center and the force centers are now αL and $(1 - \alpha)L$, respectively. We confine α to the interval $]0, 0.5]$. For $\alpha = 0.5$, no net self-induced motion occurs by symmetry. For $\alpha \neq 0.5$, the swimmer body feels a net self-induced fluid flow due to the proximity to one of the two force centers. Due to the resulting self-induced hydrodynamic drag on the swimmer body, the swimmer self-propels. It is important to note that the two force centers are completely fixed or attached to the particle center. So they propel the particle and are moving with the particle at the same

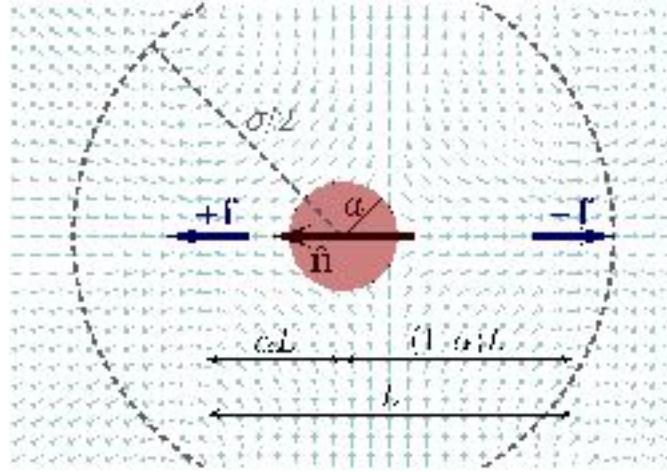


Figure 9.2: **Individual model microswimmer.** The spherical swimmer body of hydrodynamic radius a is subjected to hydrodynamic drag. Two active point-like force centers exert active forces $+f$ and $-f$ onto the surrounding fluid. This results in a self-induced fluid flow indicated by the small light arrows. L is the distance between the two force centers. The whole set-up is axially symmetric with respect to the axis $\hat{\mathbf{n}}$. If the swimmer body is shifted along $\hat{\mathbf{n}}$ out of the geometric center, leading to distances αL and $(1 - \alpha)L$ to the two force centers, it feels a net self-induced hydrodynamic drag. The microswimmer then self-propels. In the depicted state (*pusher*), fluid is pushed outward. Upon inversion of the two forces, fluid is pulled inward (*puller*). We consider soft isotropic steric interactions between the swimmer bodies of typical interaction range σ , implying an effective steric swimmer radius of $\sigma/2$. Reproduced from Ref. [30].

time. In the state of outward oriented forces (depicted in Fig. 9.2), the swimmer pushes the fluid outward and is called a *pusher* [33]. Inverting the forces, the swimmer pulls fluid inward and is termed a *puller* [33].

We now consider an assembly of N interacting identical self-propelled model microswimmers, suspended in a viscous, incompressible fluid at low Reynolds number [31]. The flow profile within the system then follows Stokes' equation [34],

$$\eta \nabla^2 \mathbf{v}(\mathbf{r}, t) + \nabla p(\mathbf{r}, t) = \sum_{i=1}^N \mathbf{f}_i(\mathbf{r}_i, \hat{\mathbf{n}}_i, t), \quad (9.31)$$

where t denotes time and \mathbf{r} any spatial position in the suspension, $\mathbf{v}(\mathbf{r}, t)$ gives the corresponding fluid flow velocity field, η is the viscosity of the fluid, and $p(\mathbf{r}, t)$ is the pressure field. On the right-hand side, \mathbf{f}_i denotes the total force density field exerted by the i -th microswimmer onto the fluid, \mathbf{r}_i and $\hat{\mathbf{n}}_i$ mark the current position and orientation of the i -th swimmer at time t , respectively. Thus, each microswimmer contributes to the overall fluid flow in the system by the force density it exerts on the fluid. In this way, each swimmer can transport itself via active self-propulsion since the point force centers are firmly attached to the swimmer body. Moreover, all swimmers hydrodynamically interact with each other via their induced flow fields.

The linearity of Eq. (9.31) and the incompressibility of the fluid, i.e., $\nabla \cdot \mathbf{v}(\mathbf{r}, t) = 0$, imply a linear relation between velocities (angular velocities) and forces (torques). We denote by \mathbf{F}_j and \mathbf{T}_j the forces and torques, respectively, acting directly on the swimmer bodies ($j = 1, \dots, N$), except for frictional forces and torques resulting from the surrounding fluid. The non-hydrodynamic body forces and torques may for example result from external potentials or steric interactions and will be specified below. From them, in the passive case, i.e., for $\mathbf{f} = \mathbf{0}$, the instantly resulting velocity \mathbf{v}_i and angular velocity $\boldsymbol{\omega}_i$ of the i -th swimmer body follows as

$$\begin{bmatrix} \mathbf{v}_i \\ \boldsymbol{\omega}_i \end{bmatrix} = \sum_{j=1}^N \mathbf{M}_{ij} \cdot \begin{bmatrix} \mathbf{F}_j \\ \mathbf{T}_j \end{bmatrix} = \sum_{j=1}^N \begin{bmatrix} \boldsymbol{\mu}_{ij}^{tt} & \boldsymbol{\mu}_{ij}^{tr} \\ \boldsymbol{\mu}_{ij}^{rt} & \boldsymbol{\mu}_{ij}^{rr} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{F}_j \\ \mathbf{T}_j \end{bmatrix}, \quad (9.32)$$

where \mathbf{M}_{ij} are the mobility matrices, the components of which $(\boldsymbol{\mu}_{ij}^{tt}, \boldsymbol{\mu}_{ij}^{tr}, \boldsymbol{\mu}_{ij}^{rt}, \boldsymbol{\mu}_{ij}^{rr})$ likewise form matrices. They describe hydrodynamic translation–translation, translation–rotation, rotation–translation, and rotation–rotation

coupling, respectively. The mobility matrices can approximately be calculated as

$$\boldsymbol{\mu}_{ii}^{tt} = \mu^t \mathbf{1}, \quad \boldsymbol{\mu}_{ii}^{rr} = \mu^r \mathbf{1}, \quad \boldsymbol{\mu}_{ii}^{tr} = \boldsymbol{\mu}_{ii}^{rt} = \mathbf{0} \quad (9.33)$$

for entries $i = j$ (no summation over i in these expressions), and

$$\boldsymbol{\mu}_{ij}^{tt} = \mu^t \left(\frac{3a}{4r_{ij}} (\mathbf{1} + \hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ij}) + \frac{1}{2} \left(\frac{a}{r_{ij}} \right)^3 (\mathbf{1} - 3\hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ij}) \right), \quad (9.34)$$

$$\boldsymbol{\mu}_{ij}^{rr} = -\mu^r \frac{1}{2} \left(\frac{a}{r_{ij}} \right)^3 (\mathbf{1} - 3\hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ij}), \quad (9.35)$$

$$\boldsymbol{\mu}_{ij}^{tr} = \boldsymbol{\mu}_{ij}^{rt} = \mu^r \left(\frac{a}{r_{ij}} \right)^3 \mathbf{r}_{ij} \times, \quad (9.36)$$

for entries $i \neq j$; this should be understood in an operator sense where the cross-product acts on the subsequent vector. Here, we have introduced the abbreviations

$$\mu^t = \frac{1}{6\pi\eta a}, \quad \mu^r = \frac{1}{8\pi\eta a^3}. \quad (9.37)$$

Because of the linearity of Eq. (9.31), the effect of the active forces can be added to the swimmer velocities and angular velocities on the right-hand side of Eq. (9.32). Thus,

$$\begin{bmatrix} \mathbf{v}_i \\ \boldsymbol{\omega}_i \end{bmatrix} = \sum_{j=1}^N \left(\begin{bmatrix} \boldsymbol{\mu}_{ij}^{tt} & \boldsymbol{\mu}_{ij}^{tr} \\ \boldsymbol{\mu}_{ij}^{rt} & \boldsymbol{\mu}_{ij}^{rr} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{F}_j \\ \mathbf{T}_j \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_{ij}^{tt} & \mathbf{0} \\ \boldsymbol{\Lambda}_{ij}^{rt} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} f\hat{\mathbf{n}}_j \\ \mathbf{0} \end{bmatrix} \right). \quad (9.38)$$

Note that there are no active torques here, i.e., we are considering a linear swimmer.

For circle swimmers, a constant torque must be included to describe the circling. Moreover, $\boldsymbol{\Lambda}_{ij}^{tt}$ and $\boldsymbol{\Lambda}_{ij}^{rt}$ summarize effect of both $+f\hat{\mathbf{n}}_j$ and $-f\hat{\mathbf{n}}_j$ such that the total swimmer is force-free. In detail,

$$\boldsymbol{\Lambda}_{ij}^{tt} = \boldsymbol{\mu}_{ij}^{tt+} - \boldsymbol{\mu}_{ij}^{tt-}, \quad \boldsymbol{\Lambda}_{ij}^{rt} = \boldsymbol{\mu}_{ij}^{rt+} - \boldsymbol{\mu}_{ij}^{rt-}.$$

For $i = j$, the term $\boldsymbol{\Lambda}_{ii}^{tt}$ contains the self-propulsion of the particles.

9.5.2 Derivation

We now aim at a statistical description for the full joint probability density $p = p(\mathbf{r}^N, \hat{\mathbf{n}}^N, t)$ and start from the dynamical Smoluchowski equation

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^N \left\{ \nabla_{\mathbf{r}_i} \cdot (\mathbf{v}_i p) + (\hat{\mathbf{n}}_i \times \nabla_{\hat{\mathbf{n}}_i}) \cdot (\boldsymbol{\omega}_i p) \right\}. \quad (9.39)$$

Integrating out all degrees of freedom except for one swimmer, we get the following exact relation for the dynamics of the swimmer one-body density

$$\frac{\partial \rho^{(1)}(\mathbf{X}, t)}{\partial t} = -\nabla_{\mathbf{r}} \cdot (\mathcal{J}^{tt} + \mathcal{J}^{tr} + \mathcal{J}^{ta}) - (\hat{\mathbf{n}} \times \nabla_{\hat{\mathbf{n}}}) \cdot (\mathcal{J}^{rt} + \mathcal{J}^{rr} + \mathcal{J}^{ra}), \quad (9.40)$$

where $\mathbf{X} = (\mathbf{r}, \hat{\mathbf{n}})$ is a compact notation for both translational and orientational degrees of freedom. The six current densities are given by

$$\begin{aligned}\mathcal{J}^{\text{tt}} &= -\mu^{\text{t}} \left(k_B T \nabla_{\mathbf{r}} \rho^{(1)}(\mathbf{X}, t) + \rho^{(1)}(\mathbf{X}, t) \nabla_{\mathbf{r}} V_{\text{ext}}(\mathbf{r}) + \int d\mathbf{X}' \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) \nabla_{\mathbf{r}'} v(|\mathbf{r} - \mathbf{r}'|) \right) \\ &\quad - \int d\mathbf{X}' \boldsymbol{\mu}_{\mathbf{r}, \mathbf{r}'}^{\text{tt}} \cdot \left(k_B T \nabla_{\mathbf{r}'} \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) + \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) \nabla_{\mathbf{r}'} V_{\text{ext}}(\mathbf{r}') \right. \\ &\quad \left. + \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) \nabla_{\mathbf{r}'} v(|\mathbf{r} - \mathbf{r}'|) + \int d\mathbf{X}'' \rho^{(3)}(\mathbf{X}, \mathbf{X}', \mathbf{X}'', t) \nabla_{\mathbf{r}''} v(|\mathbf{r}' - \mathbf{r}''|) \right), \\ \mathcal{J}^{\text{tr}} &= - \int d\mathbf{X}' k_B T \boldsymbol{\mu}_{\mathbf{r}, \mathbf{r}'}^{\text{tr}} \cdot (\hat{\mathbf{n}}' \times \nabla_{\hat{\mathbf{n}}'}) \rho^{(2)}(\mathbf{X}, \mathbf{X}', t), \\ \mathcal{J}^{\text{ta}} &= f \left(\boldsymbol{\Lambda}_{\mathbf{r}, \mathbf{r}}^{\text{tt}} \cdot \hat{\mathbf{n}} \rho^{(1)}(\mathbf{X}, t) + \int d\mathbf{X}' \boldsymbol{\Lambda}_{\mathbf{r}, \mathbf{X}'}^{\text{tt}} \cdot \hat{\mathbf{n}}' \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) \right), \\ \mathcal{J}^{\text{rt}} &= - \int d\mathbf{X}' \boldsymbol{\mu}_{\mathbf{r}, \mathbf{r}'}^{\text{rt}} \left(k_B T \nabla_{\mathbf{r}'} \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) + \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) \nabla_{\mathbf{r}'} V_{\text{ext}}(\mathbf{r}') \right. \\ &\quad \left. + \rho^{(2)}(\mathbf{X}, \mathbf{X}', t) \nabla_{\mathbf{r}'} v(|\mathbf{r} - \mathbf{r}'|) + \int d\mathbf{X}'' \rho^{(3)}(\mathbf{X}, \mathbf{X}', \mathbf{X}'', t) \nabla_{\mathbf{r}''} v(|\mathbf{r}' - \mathbf{r}''|) \right), \\ \mathcal{J}^{\text{rr}} &= -k_B T \boldsymbol{\mu}^{\text{r}} \hat{\mathbf{n}} \times \nabla_{\hat{\mathbf{n}}} \rho^{(1)}(\mathbf{X}, t) - \int d\mathbf{X}' k_B T \boldsymbol{\mu}_{\mathbf{r}, \mathbf{r}'}^{\text{rr}} \cdot (\hat{\mathbf{n}}' \times \nabla_{\hat{\mathbf{n}}'}) \rho^{(2)}(\mathbf{X}, \mathbf{X}', t), \\ \mathcal{J}^{\text{ra}} &= f \int d\mathbf{X}' \boldsymbol{\Lambda}_{\mathbf{r}, \mathbf{X}'}^{\text{rt}} \hat{\mathbf{n}}' \rho^{(2)}(\mathbf{X}, \mathbf{X}', t),\end{aligned}\tag{9.41}$$

where $\rho^{(3)}(\mathbf{X}, \mathbf{X}', \mathbf{X}'', t)$ is the nonequilibrium triplet density.

We close these exact equations approximatively by using the DDFT relations on the pair and triplet level, namely,

$$\int d\mathbf{r}' d\hat{\mathbf{n}}' \rho^{(2)}(\mathbf{r}, \mathbf{r}', \hat{\mathbf{n}}, \hat{\mathbf{n}}', t) \nabla_{\mathbf{r}'} v(|\mathbf{r} - \mathbf{r}'|) = \rho^{(1)}(\mathbf{r}, \hat{\mathbf{n}}, t) \nabla_{\mathbf{r}} \frac{\delta \mathcal{F}_{\text{exc}}}{\delta \rho^{(1)}(\mathbf{r}, \hat{\mathbf{n}}, t)}\tag{9.42}$$

and

$$\begin{aligned}\nabla_{\mathbf{r}'} \rho^{(2)}(\mathbf{r}, \mathbf{r}' \hat{\mathbf{n}}, \hat{\mathbf{n}}', t) + \rho^{(2)}(\mathbf{r}, \mathbf{r}' \hat{\mathbf{n}}, \hat{\mathbf{n}}', t) \nabla_{\mathbf{r}'} v(|\mathbf{r} - \mathbf{r}'|) + \int d\mathbf{r}' d\hat{\mathbf{n}}' \rho^{(3)}(\mathbf{r}, \mathbf{r}', \mathbf{r}'', \hat{\mathbf{n}}, \hat{\mathbf{n}}', \hat{\mathbf{n}}'', t) \nabla_{\mathbf{r}''} u(\mathbf{r}' \mathbf{r}'') = \\ = \rho^{(2)}(\mathbf{r}, \mathbf{r}' \hat{\mathbf{n}}, \hat{\mathbf{n}}', t) \left(\nabla_{\mathbf{r}'} \ln \left(\lambda^3 \rho^{(1)}(\mathbf{r}', \hat{\mathbf{n}}', t) \right) + \nabla_{\mathbf{r}'} \frac{\delta \mathcal{F}_{\text{exc}}}{\delta \rho^{(1)}(\mathbf{r}', \hat{\mathbf{n}}', t)} \right).\end{aligned}$$

The remaining input is a concrete approximation for the equilibrium density functional where we adopt the mean-field approximation

$$\mathcal{F}_{\text{exc}} = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' d\hat{\mathbf{n}} d\hat{\mathbf{n}}' \rho^{(1)}(\mathbf{r}, \hat{\mathbf{n}}, t) \rho^{(1)}(\mathbf{r}', \hat{\mathbf{n}}', t) v(|\mathbf{r} - \mathbf{r}'|)\tag{9.43}$$

and a high-temperature factorization approximation for the remaining nonequilibrium pair correlation

$$\rho^{(2)}(\mathbf{r}, \mathbf{r}', \hat{\mathbf{n}}, \hat{\mathbf{n}}', t) = \rho^{(1)}(\mathbf{r}, \hat{\mathbf{n}}, t) \rho^{(1)}(\mathbf{r}', \hat{\mathbf{n}}', t) \exp(-\beta v(|\mathbf{r} - \mathbf{r}'|)).\tag{9.44}$$

Then, the full set of equations is closed: They only involve the dynamical one-body density field and can be solved numerically.

9.5.3 Applications

As a first application, we consider the motion of microswimmers which are moving in two spatial dimensions surrounded by a three-dimensional bulk fluid [30]. They are confined to a quartic external potential

$$V_{\text{ext}}(\mathbf{r}) = k |\mathbf{r}|^4,\tag{9.45}$$

with k defining a confinement strength, and exhibit non-aligning interactions embodied in the steric pair potential

$$v(r) = \epsilon_0 \exp \left(-\frac{r^4}{\sigma^4} \right),\tag{9.46}$$

where ε_0 sets the strength of this potential and σ an effective interaction range.

The calculation protocol is the following. First, turn the activity f off and equilibrate the particles in the quartic potential. The parameters are chosen in such a way that the equilibrium system is in the fluid phase but exhibits density peaks due to the steric potential. Then, the self-propulsion f is switched on and the density evolution is followed by solving the DDFT equations numerically. For small self-propulsion strengths $|f|$, a stationary high-density ring is formed both for pushers ($f > 0$) and pullers ($f < 0$), which is extended relative to the typical extension of the equilibrated system. If the self-propulsion strength is getting larger, a tangential instability occurs and the system forms spontaneously a hydrodynamic pump as predicted earlier by simulations [35, 36]. For even higher $|f|$ the pump gets unstable resulting in a continuous dynamic *swashing* of the density cloud. The behavior is similar for pushers and pullers but details are different, as illustrated in Fig. 9.3.

The swimmer model can be generalized to force centers that are not collinear with the swimmer. This results in circle-swimming, which was further analyzed within DDFT in Ref. [37]. Moreover the DDFT approach was applied to global polar ordering in pure pusher or puller suspensions [38]. As a result, at sufficiently high concentrations, polar ordering was found for pullers, but not for pushers. Finally, the versatility of the DDFT is documented by its generalization to binary mixtures and to dynamics in a sheared fluid, which were considered and elaborated on in Ref. [39].

As a final remark, if one has dry active matter in mind from the very beginning, it is better to start with the approach described in Section 9.4. The limit of vanishing hydrodynamic interactions is not a simple one if one uses the swimmer model of this section, because hydrodynamics and self-propulsion are intrinsically coupled here.

9.6 Final remarks

In conclusion, DDFT, which has been known as a successful theory for interacting Brownian colloidal particles, can be applied to active matter as well. In particular, both *dry active matter* and *wet active matter* (microswimmers) can be treated at different levels of complexity.

We believe that future research will be directed along the following topics:

- DFT provides an ideal framework to tackle *aligning interactions*. This strength needs to be exploited further to establish a first-principle approach to the Vicsek-model of swarming [40] and to the impact of alignment on motility-induced phase separation [41, 42, 43].

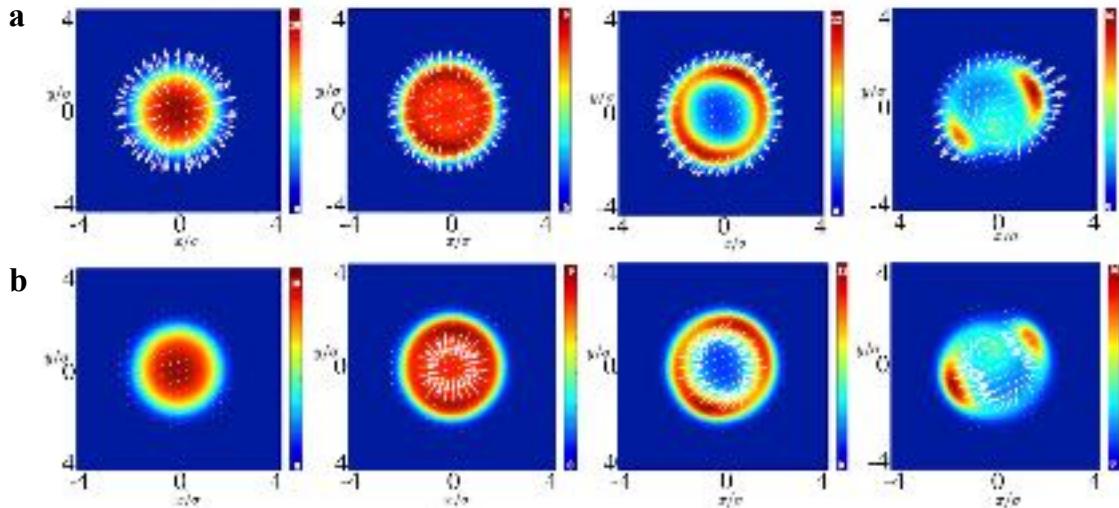


Figure 9.3: Time evolution of the density and orientation profiles for pushers and pullers. Time evolution of the density profiles (color maps) and orientation profiles (white arrows) (a) for pushers and (b) for pullers. The snapshots were obtained at times $t = 0.05, 0.1, 0.2, 0.8$ in units of the Brownian time $\sigma^2/k_B T \mu'$. Reproduced from Ref. [30]

- So far, we discussed swimmers in a viscous Newtonian fluid, but in many situations there is a *viscoelastic solvent*. Then, memory effects of the solvent play a role, modifying and affecting the swimming process. One basic example for a viscoelastic medium is a polymer solution. For colloids swimming in a polymer solution, a strongly enhanced rotational diffusion was found in experiment [44] and simulation [45]. It is challenging to treat these memory effects of the viscoelastic solvent within DDFT.
- DFT is ideal for the calculation of *interfaces and wetting effects* [46]. So, it should be applied to study interfaces between coexisting states for active particles. These can be both fluid–fluid and fluid–solid interfaces. For an effective equilibrium model for dry active particles, this was done by Wittmann and Brader [47] but an extension towards the full DDFT is still needed.
- As DFT describes freezing and crystallization in equilibrium, DDFT should be applied to *freezing* of active particles [48]. A simplified approach based on the phase-field-crystal model has been proposed [49], but this needs extension towards a theory that includes microscopic correlations.
- Particles with *time-dependent pair interactions*, such as breathing particle whose interaction diameter changes periodically as a function of time, are another interesting field of research. In principle, there should be no obstacle to formulate a DDFT for these nonequilibrium systems, which play an important role for modelling dense biological tissue [50].
- Bacteria subjected to simultaneous *growth and division* establish a complex dynamical phenomenon when strongly interacting [51]. A DDFT approach seems to be particularly promising to describe nematic order on the particle scale for growing bacterial colonies [52].

9.7 Problems

Problem 9.1: Explicit calculation of relevant functional derivatives. Calculate the first two functional derivatives $\frac{\delta \mathcal{F}[\rho]}{\delta \rho(\mathbf{r})}$ and $\frac{\delta^2 \mathcal{F}[\rho]}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')}$ for

1. $\mathcal{F}[\rho] = \frac{1}{2} \int d^3\mathbf{r}_1 \int d^3\mathbf{r}_2 w(|\mathbf{r}_1 - \mathbf{r}_2|) \rho(\mathbf{r}_1) \rho(\mathbf{r}_2),$
2. $\mathcal{F}[\rho] = \int d^3\mathbf{r}_1 \Psi(\rho(\mathbf{r}_1))$

in three spatial dimensions. Here, $w(r)$ and $\Psi(\rho)$ are prescribed given functions.

Problem 9.2: DDFT for the ideal gas in an external potential. Show that for the ideal gas in an external potential the DDFT reduces to the exact Smoluchowski equation.

Problem 9.3: DDFT for dry active matter and Smoluchowski equation. Show that the DDFT for dry active matter is equivalent to the underlying Smoluchowski equation if the particles are non-interacting ($v(r) = 0$) but exposed to a general external potential $V_{\text{ext}}(\mathbf{r}, \hat{\mathbf{n}}, t)$.

Problem 9.4: DDFT equations for dry active matter, mean-field approximation. Give the DDFT equations for the dynamics of the density field $\rho(\mathbf{r}, \theta, t)$ for dry active matter with the mean-field approximation of the density functional and aligning interactions.

Bibliography

- [1] M. Heinen and H. Löwen. Dynamical density functional theory for the diffusion of injected Brownian particles. *European Physical Journal - Special Topics*, 223:3113–3127, 2014.
- [2] H. Löwen. Dynamical density functional theory for Brownian dynamics of colloidal particles. In *Variational Methods in Molecular Modeling*, pages 255–284. Springer, 2017.

- [3] M. te Vrugt, H. Löwen, and R. Wittkowski. Classical dynamical density functional theory: from fundamentals to applications. *Advances in Physics*, 69:121–247, 2020.
- [4] H. Löwen. Melting, freezing and colloidal suspensions. *Physics Reports*, 237:249–324, 1994.
- [5] R. Roth. Fundamental measure theory for hard-sphere mixtures: a review. *Journal of Physics: Condensed Matter*, 22:063102, 2010.
- [6] R Roth, K Mecke, and M Oettel. Communication: Fundamental measure theory for hard disks: fluid and solid. *Journal of Chemical Physics*, 136:081101, 2012.
- [7] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*, volume 73. Oxford University Press, 1988.
- [8] A. J. Archer and R. Evans. Dynamical density functional theory and its application to spinodal decomposition. *Journal of Chemical Physics*, 121:4246–4254, 2004.
- [9] P. Espa  ol and H. Löwen. Derivation of dynamical density functional theory using the projection operator technique. *Journal of Chemical Physics*, 131:244101, 2009.
- [10] U. Marini Bettolo Marconi and P. Tarazona. Dynamic density functional theory of fluids. *Journal of Chemical Physics*, 110:8032–8044, 1999.
- [11] U. Marini Bettolo Marconi and P. Tarazona. Dynamic density functional theory of fluids. *Journal of Physics: Condensed Matter*, 12:A413, 2000.
- [12] R. Wittmann, C. E. Sitta, F. Smallenburg, and H. Löwen. Phase diagram of two-dimensional hard rods from fundamental mixed measure density functional theory. *Journal of Chemical Physics*, 147:134908, 2017.
- [13] M. Rex, H. H. Wensink, and H. Löwen. Dynamical density functional theory for anisotropic colloidal particles. *Physical Review E*, 76:021403, 2007.
- [14] B. ten Hagen, R. Wittkowski, D. Takagi, K. K  mmel, C. Bechinger, and H. Löwen. Can the self-propulsion of anisotropic microswimmers be described by using forces and torques? *Journal of Physics: Condensed Matter*, 27:194110, 2015.
- [15] J. R Howse, R. A.L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. Self-motile colloidal particles: from directed propulsion to random walk. *Physical Review Letters*, 99:048102, 2007.
- [16] S. van Teeffelen and H. Löwen. Dynamics of a Brownian circle swimmer. *Physical Review E*, 78:020101, 2008.
- [17] B. ten Hagen, S. van Teeffelen, and H. Löwen. Brownian motion of a self-propelled particle. *Journal of Physics: Condensed Matter*, 23:194119, 2011.
- [18] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [19] C. Kurzthaler, S. Leitmann, and T. Franosch. Intermediate scattering function of an anisotropic active Brownian particle. *Scientific Reports*, 6:36702, 2016.
- [20] H. H. Wensink and H. Löwen. Aggregation of self-propelled colloidal rods near confining walls. *Physical Review E*, 78:031409, 2008.
- [21] M. Enculescu and H. Stark. Active colloidal suspensions exhibit polar order under gravity. *Physical Review Letters*, 107:058301, 2011.
- [22] B. ten Hagen, F. K  mmel, R. Wittkowski, D. Takagi, H. Löwen, and C. Bechinger. Gravitaxis of asymmetric self-propelled colloidal particles. *Nature Communications*, 5:4829, 2014.
- [23] S. Hermann and M. Schmidt. Active ideal sedimentation: exact two-dimensional steady states. *Soft Matter*, 14:1614–1621, 2018.

- [24] J. Vachier and M. G. Mazza. Dynamics of sedimenting active Brownian particles. *The European Physical Journal E: Soft Matter and Biological Physics*, 42:11, 2019.
- [25] L. Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51:627–659, 1949.
- [26] R. Wittkowski and H. Löwen. Dynamical density functional theory for colloidal particles with arbitrary shape. *Molecular Physics*, 109:2935–2943, 2011.
- [27] P. Krininger, M. Schmidt, and J. M. Brader. Nonequilibrium phase behavior from minimization of free power dissipation. *Physical Review Letters*, 117:208003, 2016.
- [28] P. Krininger and M. Schmidt. Power functional theory for active Brownian particles: General formulation and power sum rules. *Journal of Chemical Physics*, 150:074112, 2019.
- [29] S. Hermann, P. Krininger, D. de las Heras, and M. Schmidt. Phase coexistence of active Brownian particles. *Physical Review E*, 100:052604, 2019.
- [30] A. M. Menzel, A. Saha, C. Hoell, and H. Löwen. Dynamical density functional theory for microswimmers. *Journal of Chemical Physics*, 144:024115, 2016.
- [31] E. M. Purcell. Life at low Reynolds number. *American Journal of Physics*, 45:3–11, 1977.
- [32] Y. Fily, A. Baskaran, and M. C. Marchetti. Cooperative self-propulsion of active and passive rotors. *Soft Matter*, 8:3002–3009, 2012.
- [33] A. Baskaran and M. C. Marchetti. Statistical mechanics and hydrodynamics of bacterial suspensions. *The Proceedings of the National Academy of Sciences*, 106:15567–15572, 2009.
- [34] J. Happel and H. Brenner. *Low Reynolds number hydrodynamics*. Springer Netherlands, 1981.
- [35] R.W. Nash, R. Adhikari, J. Tailleur, and M.E. Cates. Run-and-tumble particles with hydrodynamics: Sedimentation, trapping, and upstream swimming. *Physical Review Letters*, 104:258101, 2010.
- [36] M. Hennes, K. Wolff, and H. Stark. Self-induced polar order of active Brownian particles in a harmonic trap. *Physical Review Letters*, 112:238104, 2014.
- [37] C. Hoell, H. Löwen, and A. M. Menzel. Dynamical density functional theory for circle swimmers. *New Journal of Physics*, 19:125004, 2017.
- [38] C. Hoell, H. Löwen, and A. M. Menzel. Particle-scale statistical theory for hydrodynamically induced polar ordering in microswimmer suspensions. *Journal of Chemical Physics*, 149:144902, 2018.
- [39] C. Hoell, H. Löwen, and A. M. Menzel. Multi-species dynamical density functional theory for microswimmers: Derivation, orientational ordering, trapping potentials, and shear cells. *Journal of Chemical Physics*, 151:064902, 2019.
- [40] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75:1226, 1995.
- [41] S. Nilsson and G. Volpe. Metastable clusters and channels formed by active particles with aligning interactions. *New Journal of Physics*, 19:115008, 2017.
- [42] E. Sese-Sansa, I. Pagonabarraga, and D. Levis. Velocity alignment promotes motility-induced phase separation. *Europhysics Letters*, 124:30004, 2018.
- [43] M. N. van der Linden, L. C. Alexander, D. G. A. L. Aarts, and O. Dauchot. Interrupted motility induced phase separation in aligning active colloids. *Physical Review Letters*, 123:098001, 2019.
- [44] J. R. Gomez-Solano, A. Blokhuis, and C. Bechinger. Dynamics of self-propelled Janus particles in viscoelastic fluids. *Physical Review Letters*, 116:138301, 2016.

- [45] K. Qi, E. Westphal, G. Gompper, and R. G. Winkler. Enhanced rotational motion of spherical squirmer in polymer solutions. *Physical Review Letters*, 124:068001, 2020.
- [46] R. Evans. The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. *Advances in Physics*, 28:143–200, 1979.
- [47] R. Wittmann and J. M. Brader. Active Brownian particles at interfaces: An effective equilibrium approach. *Europhysics Letters*, 114:68004, 2016.
- [48] J. Bialké, T. Speck, and H. Löwen. Crystallization in a dense suspension of self-propelled particles. *Physical Review Letters*, 108:168301, 2012.
- [49] A. M. Menzel and H. Löwen. Traveling and resting crystals in active systems. *Physical Review Letters*, 110:055702, 2013.
- [50] E. Tjhung and L. Berthier. Discontinuous fluidization transition in time-correlated assemblies of actively deforming particles. *Physical Review E*, 96:050601, 2017.
- [51] R. Los, D. van Holthe tot Echten, G. Nordemann, M. Wehrens, S. J. Tans, and T. Idema. Defect dynamics in growing bacterial colonies. *arXiv:2003.10509*, 2022.
- [52] D. Dell’Arciprete, M. L. Blow, A. T. Brown, F. D. C. Farrell, J. S. Lintuvuori, A. F. McVey, D. Marenduzzo, and W. C. K. Poon. A growing bacterial colony in two dimensions as an active nematic. *Nature Communications*, 9:1–9, 2018.

Chapter 10

Motility-Induced Phase Separation

JOAKIM STENHAMMAR

The phenomenon of *motility-induced phase separation* (MIPS) is one of the most well-studied collective behaviors in numerical and theoretical models of active matter. As the name indicates, MIPS involves a phase coexistence between two phases of different densities, similar to what happens in a binary fluid mixture as the temperature is decreased below the critical temperature. The difference between MIPS and an equilibrium phase separation is that the driving force is completely non-thermodynamic: While phase separation in equilibrium fluids is driven by attractive inter-particle interactions that overcome the entropy of mixing as quantified by the thermal energy $k_B T$ (a *static* property), MIPS is instead induced by the non-equilibrium character of the single-particle motion (a *dynamic* property). MIPS has turned out to be a robust phenomenon occurring in a range of active matter models without alignment interactions [2, 3] (*scalar* active matter) in contact with a solid substrate or other momentum sink (*dry* active matter). Due to its conceptual similarity to equilibrium phase separation, much effort has been made to understand MIPS using tools and concepts from classical

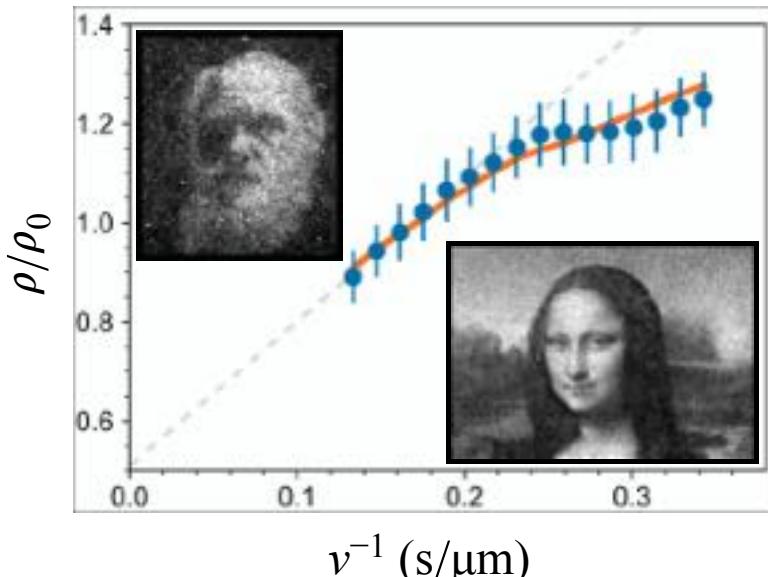


Figure 10.1: **Accumulation of swimming bacteria using spatially varying light intensities.** The plot shows the measured local density ρ of *E. coli* bacteria as a function of their swimming speed v , with the predicted $\rho \sim v^{-1}$ behavior shown by the dashed line. The deviations at low speeds is primarily due to interactions between bacteria that limit their local density. The insets show *density-painted* pictures of the Mona Lisa and Charles Darwin, achieved by adaptively varying the local light intensity. Reproduced from Ref. [1].

(equilibrium) thermodynamics [4, 5, 6, 7, 8, 9, 10, 11, 12, 13], recast to take into account the non-equilibrium character of the system. This Chapter will address some of these developments, focusing on computational and theoretical descriptions of MIPS in microscopic active particle models. Due to the very extensive literature on the topic, the aim is not to give a comprehensive review of the results in the field, but rather to focus on a few key concepts. We will first illustrate the mechanistic origin of pattern formation and MIPS in active particles that interact via a position- or density-dependent propulsion speed. We will then switch to active particles that interact through pair-wise forces. In particular, we will address the concept of *pressure* in suspensions of active Brownian particles, and to what extent this concept can be used as a thermodynamic variable to describe phase coexistence. Finally, we will briefly discuss some of the open research problems within the MIPS field.

10.1 Pattern formation in active particles with spatially varying speeds

Arguably, the simplest example of an active system is a single self-propelled particle that moves with constant speed v_0 in a direction \mathbf{n} that relaxes either continuously through rotational Brownian motion (*active Brownian particles*) or via discontinuous tumbling events (*run-and-tumble particles*, see Chapter ??). This directed motion requires either the consumption of chemical fuel, or some sort of external driving (e.g., by light), both of which drive the particle out of thermodynamic equilibrium. Now, we divide our physical system into two compartments where the particle has two distinct swimming speeds v_1 and v_2 with $v_2 > v_1$. This setup can be most easily imagined for the case of a particle driven by an external light field [14, 1, 15] of varying intensity, but can also be imagined to arise through a non-uniform distribution of nutrients in a suspension of swimming bacteria. Everything else alike, the active particle will spend more time in the region where it moves slower, so that, over long times, the particle density ρ obeys $\rho_1 > \rho_2$; in other words, active particles accumulate where they move slowly. This is easily imagined in everyday settings — think about traffic congestion near a roadwork area with reduced speed — but it is important to remember that this is an intrinsically non-equilibrium phenomenon: In equilibrium, the mean-square velocity $\langle v^2 \rangle$ of a particle is related to the temperature T by the equipartition theorem $\langle v^2 \rangle = 3k_B T/m$, where m is the particle's mass; thus, the only way to create a non-uniform velocity for a Brownian particle is to have a non-uniform temperature, destroying the equilibrium nature of the system.

To analyze this accumulation effect in quantitative terms, we follow Ref. [16] and consider run-and-tumble particles in $d = 1$ dimension: this system is composed of right- and left-moving particles that move in a motility landscape $v(x)$ and tumble with spatially varying frequency $\alpha(x)$. The particle current $J(x)$ in this system can be expressed as:

$$J(x) = -\frac{v^2(x)}{\alpha(x)} \frac{d\rho}{dx} - \frac{v(x)\rho(x)}{\alpha(x)} \frac{dv}{dx}. \quad (10.1)$$

Assuming the existence of a flux-free steady state ($J = 0$), this equation can be solved for the density distribution $\rho(x)$, yielding [16, 2]

$$\rho(x) = \rho_0 \frac{v_0}{v(x)}, \quad (10.2)$$

where ρ_0 is the reference density in a system with particle speed v_0 . For our specific case of two different swim speeds, this means that $\rho_1 v_1 = \rho_2 v_2$, i.e., the local particle density is inversely proportional to the local particle speed but independent of the tumbling rate. This simple yet powerful result has been elegantly demonstrated in experiments on light-powered swimming bacteria [1, 15]: if a properly shaped *light pattern* of varying intensity is used, such bacteria can be directed to dynamically self-assemble into intricate (and artistic) patterns (see Fig. 10.1).

Example 10.1: Particle current for run-and-tumble particles. Derive Eq. (10.1) and then verify that the criterion in Eq. (10.2) implies $J = 0$. [Hint: Start from the expressions of the density of right- and

left-moving particles $R(x, t)$ and $L(x, t)$, respectively [16]:

$$\begin{aligned}\frac{\partial R}{\partial t} &= -\frac{\partial(vR)}{\partial x} - \frac{\alpha R}{2} + \frac{\alpha L}{2} \\ \frac{\partial L}{\partial t} &= +\frac{\partial(vL)}{\partial x} + \frac{\alpha R}{2} - \frac{\alpha L}{2}.\end{aligned}\quad (10.3)$$

First, combine the above equations to yield PDEs for the auxiliary fields $\rho \equiv R + L$ and $\sigma \equiv R - L$, and note that the particle current is given by $J = v\sigma$. Then, combine these equations to an expression for $\partial_t^2 \rho$ and assume that the density ρ relaxes slowly enough that we can set $\partial_t^2 \rho = 0$ to yield Eq. (10.1) for J . See also Ref. [16], but beware of a typo in the last term of Eq. (2.4).]

10.2 MIPS in quorum-sensing active particles

We now consider the closely related case where the active particle speed depends on position implicitly through a dependence on the local particle density, i.e., $v(x) = v[\rho(x)]$ [5, 2]. From a microbiological viewpoint, this constitutes a minimal model of *quorum sensing*, whereby bacteria chemically sense and respond to their local density by a change of gene expression (phenotype). The phenotypic change can constitute deflagellation or a decrease in swimming speed, for example during the initial stages of biofilm formation [17].

Following [2], we consider a small perturbation $\delta\rho$ around a homogeneous state of particle density $\rho_0 \equiv c/v(\rho_0)$ with c some constant. The speed at the perturbed density is given by

$$v[\rho_0 + \delta\rho(x)] = v(\rho_0) + v'(\rho_0)\delta\rho(x), \quad (10.4)$$

where $v'(\rho) \equiv \frac{dv}{d\rho}$. Thus, if $v' < 0$, a positive density perturbation ($\delta\rho > 0$) leads to a velocity decrease ($\delta v < 0$). For steeply enough decreasing $v(\rho)$, this leads to a further increase in the local density to $\rho_0 + \delta\hat{\rho}$, triggering a positive feedback loop that leads to MIPS. More precisely, to linear order in $\delta\rho$, we have

$$\rho_0 + \delta\hat{\rho} = \frac{c}{v(\rho_0) + v'(\rho_0)\delta\rho} \approx \rho_0 \left[1 - \frac{v'(\rho_0)}{v(\rho_0)}\delta\rho \right]. \quad (10.5)$$

The condition for the onset of the MIPS feedback loop is that $\delta\hat{\rho} > \delta\rho$, leading to

$$\frac{v'(\rho_0)}{v(\rho_0)} < -\frac{1}{\rho_0}. \quad (10.6)$$

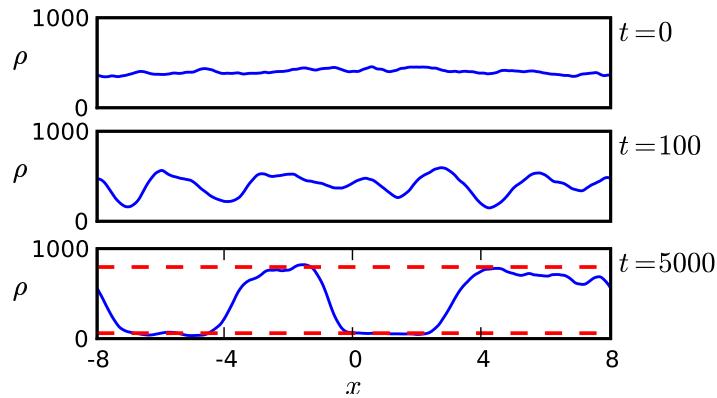


Figure 10.2: MIPS in quorum-sensing particles. For densities where the instability criterion given by Eq. (10.6) is fulfilled, an initially homogeneous system ($t = 0$) will phase separate into coexisting regions of high and low densities. In one dimension ($d = 1$), the domains maintain a finite size at steady state ($t = 5000$) due to the finite energy cost of forming an interface, while for $d > 1$ the domains will coarsen until bulk phase separation is reached. Image reproduced from [5].

This constitutes a linear instability criterion on the homogeneous density ρ_0 : For $\rho > \rho_0$, the homogeneous state will destabilize, leading to a spinodal decomposition into coexisting dense and dilute regions, as shown in Fig. 10.2 for the choice of an exponentially decreasing $v(\rho)$ [5].

Interestingly, as you will show in Ex. 10.2, the MIPS criterion (10.6) can be reformulated as an equilibrium spinodal condition $d^2 f_0/d\rho^2 < 0$ on an effective bulk free energy density $f_0(\rho)$ given by [5]

$$f_0(\rho) = \rho(\ln \rho - 1) + \int^\rho \ln[v(u)]du, \quad (10.7)$$

where the first term can be identified as the ideal gas entropy, which favors a homogeneous particle distribution, and the second term encodes an effective attraction due to the decreasing $v(\rho)$ responsible for MIPS. This thermodynamic mapping of MIPS to phase coexistence in equilibrium fluids breaks down at higher orders in a gradient expansion of the free energy, i.e., for interfacial free energy terms [5, 18, 19]. Nevertheless, it points towards strong analogies between MIPS and equilibrium phase coexistence at the mesoscopic level, in spite of the strongly nonequilibrium dynamics at the microscopic level.

Example 10.2: Onset of MIPS.

- a. Show that Eq. (10.5) implies the MIPS criterion in Eq. (10.6).
- b. Show that the latter is equivalent to a spinodal criterion ($\partial_\rho^2 f_0 < 0$) on the bulk free energy density given by Eq. (10.7).

10.3 Microscopic particle models for MIPS

In this Section, we will briefly review the three main particle-based models used for theoretical and computational studies of MIPS, namely *quorum-sensing active particles* (QSAPs), *active Brownian particles* (ABPs), and *active Ornstein-Uhlenbeck particles* (AOUPs). While the qualitative properties of MIPS can be readily demonstrated in one dimension (see Fig. 10.2), studying two- and three-dimensional systems is necessary to gain a quantitative understanding of the phase diagram and coarsening dynamics. Since the known differences between MIPS in 2D and 3D are small [20, 21], we will restrict ourselves to $d = 2$ in the following.

The one-dimensional, isotropic QSAP model introduced in the previous section is readily extended to $d = 2$, yielding the following equation of motion for the position \mathbf{r}_i of particle i with orientation \mathbf{n}_i :

$$\partial_t \mathbf{r}_i = v[\tilde{\rho}(\mathbf{r}_i)]\mathbf{n}_i + \sqrt{2D}\boldsymbol{\xi}_i, \quad (\text{QSAPs}) \quad (10.8)$$

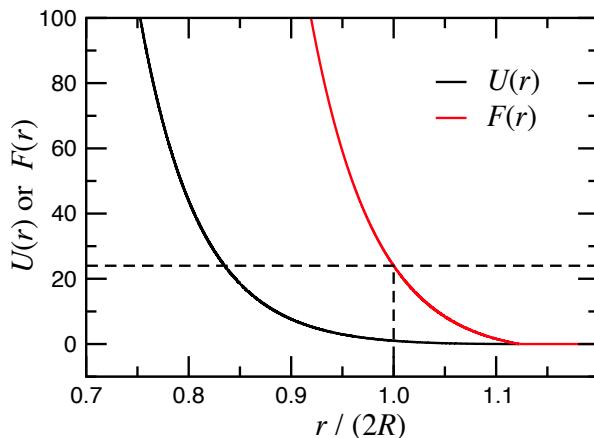


Figure 10.3: **Weeks-Chandler-Andersen (WCA) potential.** Inter-particle potential $U(r)$ and force $F(r) = -dU/dr$ for the WCA potential in Eq. (10.15), expressed in units of ϵ and $\epsilon/(2R)$. Using the constant self-propulsion force $F_p = v_0 = 24$ (dashed line) is common, as it leads to an *effective diameter* of $\sim 2R$ for two colliding particles [20].

where the second term represents Brownian diffusion with translational diffusion coefficient D and ξ_i represents a zero-mean, unit variance Gaussian white noise process that obeys $\langle \xi_{i\alpha}(t)\xi_{j\beta}(t') \rangle = \delta_{ij}\delta_{\alpha\beta}\delta(t-t')$. In computational implementations, the choice of functional form for $v(\rho)$ needs to be complemented by a recipe for measuring the local density field $\tilde{\rho}$. This is typically achieved by coarse-graining the instantaneous particle density $\hat{\rho}(\mathbf{r}) = \sum_i \delta(\mathbf{r} - \mathbf{r}_i)$ in the neighborhood of each particle by convoluting it with an isotropic coarse-graining kernel K_{cg} [22]:

$$\tilde{\rho}(\mathbf{r}) = \int K_{cg}(\mathbf{r} - \mathbf{r}') \hat{\rho}(\mathbf{r}') d\mathbf{r}'. \quad (10.9)$$

A typical choice for K_{cg} is the short-ranged, bell-shaped kernel [22]

$$K_{cg}(r) = \frac{\Theta(r_0 - r)}{Z} \exp\left(-\frac{r_0^2}{r_0^2 - r^2}\right), \quad (10.10)$$

where Θ is the Heaviside function, Z a normalization constant, and r_0 the cutoff radius. The dynamics of the particle orientation \mathbf{n}_i can furthermore be implemented either as a run-and-tumble process with frequency α as in one dimension, or as Brownian rotational diffusion with diffusion constant D_R (which can only be realized for $d \geq 2$), as for ABPs (see Eq. (10.12)); from the perspective of MIPS, these two choices are essentially equivalent upon the mapping $D_R \longleftrightarrow 2(d-1)\alpha$ [23].

A more computationally straightforward way to treat inter-particle interactions is through a constant swimming speed v_0 combined with pair-wise inter-particle forces $\mathbf{F}_i(\{\mathbf{r}_j\})$:

$$\partial_t \mathbf{r}_i = v_0 \mathbf{n}_i + \mathbf{F}_i(\{\mathbf{r}_j\}) + \sqrt{2D} \boldsymbol{\xi}_i, \quad (\text{ABPs}) \quad (10.11)$$

where we have implicitly set the single-particle mobility to unity, which, together with the overdamped dynamics, enables the implementation of self-propulsion as an effective *self-propulsion force* $F_p = v_0$. As discussed in previous chapters, for ABPs, Eq. (10.11) is combined with a diffusional relaxation of the swimming direction $\mathbf{n}_i \equiv (\cos \theta_i, \sin \theta_i)$:

$$\partial_t \theta_i = \sqrt{2D_R} \xi_{r,i}, \quad (\text{ABPs}) \quad (10.12)$$

where D_R is the rotational diffusion constant and $\xi_{r,i}$ is a white-noise process. Eqs. (10.11)–(10.12) together constitute the active Brownian particle (ABP) model [24], which has become a mainstay model for computational studies of MIPS.

A third well-studied model for active particles undergoing MIPS is the active Ornstein-Uhlenbeck particle (AOUP) model [25, 26, 27, 28], which replaces the constant self-propulsion term with a translational noise with memory (colored noise), and therefore does not require a separate description of the orientational dynamics. The equation of motion for an AOUP interacting through pair-wise forces is given by

$$\partial_t \mathbf{r}_i = \mathbf{v}_i + \mathbf{F}_i(\{\mathbf{r}_j\}), \quad (\text{AOUPs}) \quad (10.13)$$

with the velocity \mathbf{v}_i obtained as the solution of

$$\tau \partial_t \mathbf{v}_i = -\mathbf{v}_i + \sqrt{2D_A} \boldsymbol{\xi}_i, \quad (\text{AOUPs}) \quad (10.14)$$

where τ is the persistence time of the propulsion direction, D_A is an active (effective) diffusion coefficient describing the persistent random walk process, and $\boldsymbol{\xi}_i$ is a white noise with the same properties as $\boldsymbol{\xi}_i$ in Eq. (10.11). The solution of Eq. (10.14) yields an exponentially-correlated velocity, $\langle v_{i\alpha} v_{j\beta} \rangle = \delta_{ij} \delta_{\alpha\beta} \frac{D_A}{\tau} e^{-|t|/\tau}$. The AOUP model simplifies analytical progress, while sharing the qualitative features of MIPS in ABPs. [25]

While all three models described above exhibit MIPS, the field-theoretical description of QSAPs is qualitatively different from that of ABPs and AOUPs [22], since QSAPs interact in a non-pair-wise manner. Nevertheless, as we will see below, the $v(\rho)$ dynamics in QSAPs can to some extent be viewed as a mean-field description of the repulsive pair-wise forces in ABPs, allowing for easier analytical progress.

Due to their popularity, we will mainly focus on MIPS in ABP suspensions. The computational treatment of ABP suspensions is straightforward, and the dynamics of Eqs. (10.11) and (10.12) can be solved using standard techniques from molecular dynamics simulations. Since the ABP dynamics are overdamped (i.e., the particle velocity $\partial_t \mathbf{r}_i(t)$ is directly proportional to the force $\mathbf{F}_i(t)$), they can be integrated using a simple Euler scheme. The inter-particle force $\mathbf{F}_i(\{\mathbf{r}_j\})$ in Eq. (10.12) is typically pair-wise additive and isotropic, i.e., $\mathbf{F}_i = \sum_{j \neq i} \mathbf{F}_{ij}$,

where $\mathbf{F}_{ij} = -\nabla U(r_{ij})$ is the force derived from the interaction potential U and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. A particularly common choice for $U(r)$ is the short-ranged, repulsive *Weeks-Chandler-Andersen potential* (WCA potential) depicted in Fig. 10.3, given by

$$U(r) = \begin{cases} 4\epsilon \left[\left(\frac{2R}{r}\right)^{12} - \left(\frac{2R}{r}\right)^6 \right] + \epsilon & \text{for } r < 2^{1/6} \times (2R) \\ 0 & \text{otherwise} \end{cases} \quad (10.15)$$

where R is the effective particle radius and ϵ controls the steepness of the potential. Due to the short range of the WCA potential, the computation time can be made linear in the number of particles N by using standard tools from molecular simulations, such as neighbor lists (see Chapter 18). Thus, simulating systems of $N > 10^4$ ABPs is relatively straightforward on standard computers, while system sizes of $N > 10^7$ have been simulated on supercomputers [20], enabling the resolution of mesoscopic structure and dynamics of systems undergoing MIPS in both two and three dimensions.

10.4 MIPS in active Brownian particles

MIPS in two-dimensional suspensions of ABPs was first realized in 2012 by Fily and coauthors [29], who observed that, starting from a disordered initial configuration, dense and dilute phases would quickly form for high enough particle swimming speeds and densities. These results were expanded by Redner and coauthors [30], who characterized MIPS through a phase diagram (Fig. 10.4a), defined by

$$\text{Pe} = \frac{3v_0}{2RD_R}, \quad (10.16)$$

where the factor 3 comes from the connection between the translational and rotational diffusion constants in equilibrium, $D_R = 3D/4R^2$. Apart from this numerical factor, Pe measures the persistence length $l_p = v_0 \tau_R$ of the single-particle motion normalized by the particle diameter $2R$, where $\tau_R = D_R^{-1}$ is the rotational relaxation time. Pe plays the role of an inverse temperature in determining the phase diagram shown in Fig. 10.4b: For $\text{Pe} > \text{Pe}_c \approx 60$, the homogeneous system phase separates into dense and dilute phases with densities ρ_l and ρ_g , independent of the overall packing fraction ρ .

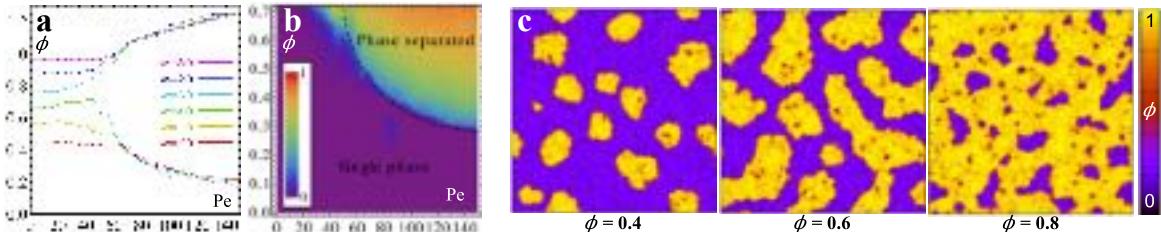


Figure 10.4: Phase coexistence in ABPs. (a) Snapshots of the local packing fraction $\phi = \rho\pi R^2$ in ABP suspensions of three different overall densities undergoing MIPS. Reproduced from Ref. [20]. (b) Measured coexistence densities as a function of Pe and ϕ for varying overall densities, as indicated. For $\text{Pe} > \text{Pe}_c \approx 60$, the suspension separates into two coexisting phases, whose densities are independent of the overall density. (c) Cluster fraction ϕ_c (Eq. (10.18)) measured from simulations in the low- and intermediate-density region. The dashed line shows the analytical prediction for $\phi_c = 0$ from Eq. (10.19) with $q = 4.5$. Reproduced from Ref. [30].

Phenomenologically, the mechanism for MIPS in QSAPs discussed in Section 10.2 can be applied also to the case of ABP suspensions: Collisions between particles slow them down in areas of high density, leading to a decreasing $v(\rho)$, which causes MIPS for large enough particle densities. For low and intermediate densities, the collision frequency between ABPs is proportional to density, while the average duration of a collision event is density-independent, leading to a linearly decreasing effective $v(\rho)$:

$$v(\rho) \simeq v_0(1 - \rho/\rho_*), \quad (10.17)$$

where ρ_* is a near close-packed density above which $v = 0$. ABP simulations for $\text{Pe} < \text{Pe}_c$ confirm Eq. (10.17) up to densities very close to ρ_* , where $v(\rho)$ goes smoothly to zero and higher-order terms in ρ become important [20, 22] (see Ex. 10.4).

Furthermore, $v(\rho)$ has a negligible dependence on Pe for $\text{Pe} < \text{Pe}_c$. Inserting Eq. (10.17) into the instability criterion given by Eq. (10.6) leads to the simple condition $\rho > \rho_*/2$ for MIPS. While reasonably close to the observed critical density in the high- Pe limit, this condition lacks a dependence on both v_0 and D_R (or α), in qualitative contrast to the computational observation of a $\text{Pe}_c \gg 1$. This is an effect of the isotropic nature of the coarse-grained density $\tilde{\rho}$ in Eq. (10.9), as discussed further in the next section. Instead, using the coarse-grained density $\tilde{\rho}(\mathbf{r}_i + \xi \mathbf{n}_i)$ a distance $\xi > 0$ in front of the particle, we can introduce an anisotropy into the dynamics which yields a dependence on v_0 and D_R that qualitatively captures the Pe dependence in the ABP phase diagram [22]. Rather than giving a full account of the anisotropic QSAP description, we will in the next section instead review the kinetic route developed in Ref. [30] to explain the strong Pe -dependence of the ABP coexistence densities.

10.4.1 Kinetic model of MIPS in ABPs

We start by assuming a dense phase with a (Pe -independent) close-packed density $\rho_l = 1/(2\sqrt{3}R^2)$ surrounded by an ABP gas at (Pe -dependent) density ρ_g . We furthermore assume that the area of the dense cluster is large enough to yield an effectively flat vapor–liquid interface. The incoming flux of particles per unit length from the gas to the liquid is then $k_{\text{in}} \sim \rho_g v_0$. The outgoing flux is incorporated by noticing that a particle will exit the cluster into the gas phase whenever its orientation \mathbf{n} has diffused above the *horizon*, i.e., $\mathbf{n} \cdot \hat{\mathbf{s}} > 0$, where $\hat{\mathbf{s}}$ is the outward surface normal [30]. Since \mathbf{n} evolves through rotational diffusion, the outward flux depends on D_R , yielding $k_{\text{out}} \sim D_R/(2R)$. Since particles are assumed to be completely trapped as long as they are oriented below the horizon, k_{out} is independent of v_0 . Crucially, this D_R dependence is absent in the effective $v(\rho)$ picture, since the speed of a QSAP is independent of its orientation as long as $\tilde{\rho}$ in Eq. (10.8) is measured isotropically.

At steady state, $k_{\text{in}} = k_{\text{out}}$, which leads to an expression for ρ_g in terms of v_0 and D_R . Taking into account prefactors, this leads to $\rho_g = \pi q D_R / (2R v_0)$ [30], where the free parameter $q \geq 1$ incorporates the fact that, when a particle escapes the cluster, it will lead to the escape of additional sub-surface particles that had previously been trapped by it. Using the expressions for ρ_g and ρ_l , we can now calculate the fraction φ_c of particles in the cluster phase, defined as

$$\varphi_c \equiv \frac{\rho_l V_l}{\rho_l V_l + \rho_g V_g} = \frac{\rho \rho_l - \rho_l \rho_g}{\rho \rho_l - \rho \rho_g}, \quad (10.18)$$

where V_l and V_g are the respective volumes of the two phases, and ρ is the overall density of the system. Inserting the expressions for ρ_l and ρ_g and using the definition of Pe (Eq. (10.16)) yields the following expression for $\varphi_c(\text{Pe}, \rho)$ [30]:

$$\varphi_c(\text{Pe}, \rho) = \frac{4R^2 \rho \text{Pe} - 3\pi q}{4R^2 \rho \text{Pe} - 6\sqrt{3}\pi q R^2 \rho}. \quad (10.19)$$

In this kinetic picture, the Pe dependence thus arises as a direct consequence of the balance between the v_0 -dependent inward flux and the D_R -dependent outward flux. The line $\varphi_c = 0$ corresponds to the stability of the dense phase in the thermodynamic limit, equivalent to the binodal line, yielding $\rho \sim 1/\text{Pe}$ (this should be contrasted with the QSAP criterion in Eq. (10.6) which describes the stability of the *homogeneous* phase, corresponding to the *spinodal* line). Restoring dimensional quantities, the binodal criterion becomes $2R v_0 \rho \gtrsim D_R$, where the left-hand side can be identified as the average collision frequency in the homogeneous phase. This criterion shows that the reorientation time τR , which sets the typical time needed for two colliding particles to separate, needs to be longer than the time between particle collisions for MIPS to occur. In other words, two particles need to stay together long enough after a collision that a third particle can collide with them in order for a cluster to be stable. In Fig. 10.4b, the binodal line $\varphi_c = 0$ from Eq. (10.19) is plotted together with φ_c measured from ABP simulations [30], showing that the kinetic picture sketched above indeed describes the gas binodal well at moderate densities.

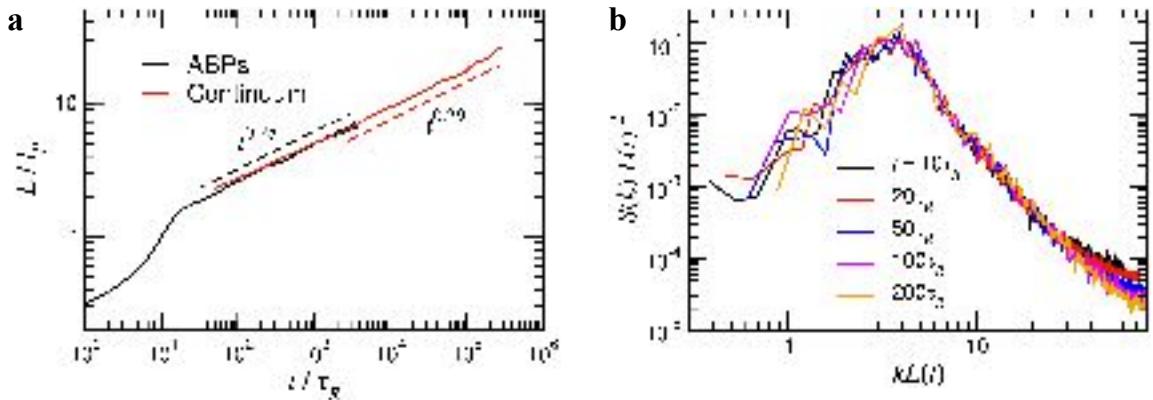


Figure 10.5: Coarsening kinetics of ABPs. (a) Growth of the characteristic cluster size $L(t)$ with time in an ABP suspension undergoing MIPS (black) and in a field theory based on the effective free energy density of Eq. (10.7) with a linearly decaying $\nu(\rho)$ measured from simulations (red). $L(t)$ is normalized by the persistence length $l_p = v_0 \tau_R$. The dashed lines indicate approximate growth exponents; the small deviations from the predicted $t^{1/3}$ scaling is an intermittent effect due to the high noise level. (b) Structure factor $S(k)$ measured at different times during coarsening. When rescaled properly by $L(t)$, the curves collapse onto each other due to self-similarity of the domain morphology. Reproduced from Ref. [20].

10.4.2 Similarities between MIPS and equilibrium phase coexistence

In spite of its intrinsically non-equilibrium origin, MIPS shares a number of similarities with simple equilibrium fluids phase separating due to attractive interactions, with Pe playing the role of an inverse temperature:

- When changing the average density ρ inside the two-phase region, ρ_g and ρ_l remain constant (Figs. 10.4c and 10.4a), while the relative volumes of the two phases change continuously according to the lever rule $N_g/N_l = (\rho_l - \rho)/(\rho - \rho_g)$, where N_g and N_l are the number of particles in the dilute and dense phases, respectively [30].
- The ABP phase diagram has both binodal and spinodal lines where the region between them is characterized by a metastable homogeneous phase, which eventually undergoes phase separation through nucleation and growth, while systems near and inside the spinodal region decompose instantaneously [7, 22].
- The phase separation kinetics are consistent with the scaling law $L(t) \sim t^{1/3}$, where $L(t)$ measures the typical size of phase-separated domains as a function of time (Fig. 10.5a) [18, 20]. In equilibrium, this scaling is expected in systems where coarsening is driven by surface tension leading to diffusive transport of material from small droplets to large ones, so-called *Ostwald ripening*.
- The evolving domain pattern after a quench is self-similar, in the sense that the domain patterns, as measured through the structure factor $S(q)$ at different times, are statistically identical when rescaled by $L(t)$ (Fig. 10.5b) [18, 20].

These observed similarities have inspired a number of studies connecting MIPS to equilibrium thermodynamic concepts such as chemical potential [18, 5, 9], pressure [31, 12, 32, 33, 34] and temperature [10, 35, 36], re-defined to take into account the intrinsic non-equilibrium character of ABPs. For example, using the QSAP model as a mean-field description for ABPs, the bulk free energy in Eq. (10.7) can be combined with a suitable interfacial-tension-like term (which in contrast does not correspond to any equilibrium counterpart) to build a continuum model of the phase separation kinetics in MIPS (Fig. 10.5a). Such field theories accurately reproduce the coarsening behavior (the last two points above) in ABPs undergoing MIPS for deep quenches [18, 20]. However, to capture the coexistence densities, it turns out that a description based on pressure in coexisting phases is more fruitful, which we will address in the next section.

10.5 Mechanical pressure in ABPs

In a phase-separating equilibrium fluid, the coexistence densities are determined by the equality between the two phases of two bulk quantities: the chemical potential μ_p and the pressure P , defined respectively as derivatives of the free energy with respect to density and volume — which provides a *thermodynamic definition* of pressure. Additionally, the pressure also has a *mechanical definition*, as the force per unit area measured either on the container walls or transmitted across an imaginary plane inside the bulk fluid, the so-called *virial pressure*. In equilibrium, these definitions must be equivalent and, since the free energy depends on bulk properties only, the mechanical pressure must be independent of the particle–wall potential: The pressure is an *equation of state* (EOS) in equilibrium. The EOS is usually expressed as pressure P as a function of either the density ρ or the volume per particle $v \equiv \rho^{-1}$ at constant temperature. As long as $P(v)$ is a monotonically decreasing function, the homogeneous system is stable for all compositions. For parameters where $P(v)$ is non-monotonic, the system becomes mechanically (and thermodynamically) unstable for some compositions and separates into coexisting phases of densities $\rho_g = v_g^{-1}$ and $\rho_l = v_l^{-1}$, whose values can be determined through the *Maxwell equal-area construction* on $P(v)$ [37], as demonstrated in Fig. 10.6.

10.5.1 Pressure in non-interacting ABPs

Since ABP suspensions are far from equilibrium, we cannot *a priori* assume them to have any mapping to thermodynamic quantities such as pressure or chemical potential, although the mechanical pressure is still a well-defined quantity. For simplicity, we start by considering the pressure in a suspension of *non-interacting* ABPs, confined by a flat wall acting on the particle x -coordinate via a force $\mathbf{F}_w(x) = \hat{\mathbf{x}} F_w(x)$; since the system is translation invariant along the y -direction, we consider only the x -dependence of the relevant quantities. Following Ref. [38], we write the master equation for the probability $\psi(x, \theta, t)$ of finding a particle at position x pointing in the direction θ at time t as

$$\partial_t \psi(x, \theta, t) = -\partial_x [v_0 \cos \theta + F_w(x)] \psi + D \partial_x^2 \psi + D_R \partial_\theta^2 \psi. \quad (10.20)$$

We now define the n^{th} moment of ψ as $m_n = \int_0^{2\pi} \cos(n\theta) \psi(x, \theta) d\theta$, where $m_0 = \rho(x)$ corresponds to the density, $m_1 = \mathcal{P}(x)$ to the polarization, and $m_2 = \mathcal{Q}(x)$ to the nematic order, where the two latter are measured normal to the interface. Taking the first two moments of Eq. (10.20) (see Ex. 10.3) and setting $\partial_t \rho = \partial_t \mathcal{P} = 0$, we obtain

$$\partial_x [v_0 \mathcal{P}(x) + F_w(x) \rho(x)] = D \partial_x^2 \rho(x), \quad (10.21)$$

$$D_R \mathcal{P}(x) = -\partial_x \left[\frac{v_0}{2} (\mathcal{Q}(x) + \rho(x)) + F_w(x) \mathcal{P}(x) - D \partial_x \mathcal{P}(x) \right]. \quad (10.22)$$

The mechanical pressure P exerted on a wall at $x \gg 0$ can now be calculated via the relation [38]

$$P = - \int_0^\infty \rho(x) F_w(x) dx. \quad (10.23)$$

Inserting Eqs. (10.21) and (10.22) and integrating yields the following, remarkably simple expression for the *ideal pressure* P_I of a suspension of non-interacting ABPs at density ρ [38, 31, 33, 34]:

$$P_I(\rho) = \left[D + \frac{v_0^2}{2D_R} \right] \rho, \quad (10.24)$$

where we can identify the first term as the ideal gas pressure $\rho k_B T$ (note that $D = k_B T$, since we have set the single-particle mobility to unity), while the second term equals ρ times the *active diffusion constant* $v_0^2/(2D_R)$. Crucially, this expression is independent of F_w ; thus, mechanical pressure is an EOS for ABPs, and can potentially be used as an effective thermodynamic quantity (Fig. 10.7). As shown in Ref. [38], this property is not true for general self-propelled particle systems, but only as long as the dynamics of \mathbf{n} are decoupled from the translational dynamics. Any torque between the particles and the wall introduces a dependence on F_w in the expression for P_I , and destroys the EOS property of the mechanical pressure, as illustrated in Fig. 10.7a. This constraint is clearly not present in equilibrium fluids, where any nonspherical molecule will experience a torque close to a surface, yet pressure remains an EOS in such systems.

Example 10.3: Mechanical pressure in ABPs.

- a. Derive the moment equations Eq. (10.21) and Eq. (10.22) by integrating the master equation Eq. (10.20), in the latter case after multiplication with $\cos \theta$, and subsequently setting $\partial_t \rho = \partial_t \mathcal{P} = 0$.
- b. Use the pressure definition in Eq. (10.23) and integrate to give the expression (10.24) for the ideal ABP pressure. [Hint: For the integration limits note that, for $x = 0$ (in the disordered bulk) and $x \rightarrow \infty$ (inside the wall), we have $\mathcal{P} = \mathcal{Q} = 0$, while $\rho(0) = \rho_0$ and $\rho(x \rightarrow \infty) = 0$.]

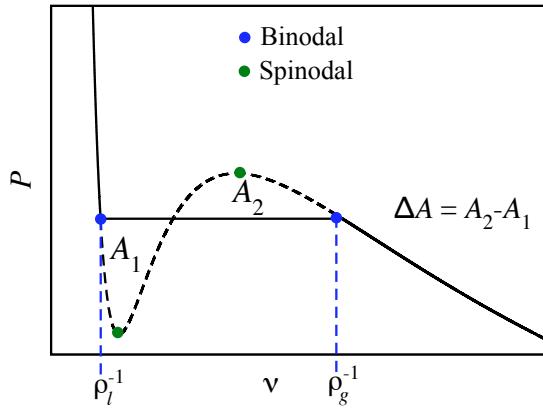


Figure 10.6: **Maxwell equal-area construction.** In equilibrium fluids, the coexistence densities ρ_g and ρ_l are determined by the equality of pressures and chemical potentials between the two bulk phases. The first condition implies a horizontal line connecting the two phases, while the latter geometrically corresponds to $\Delta A = A_2 - A_1 = 0$, leading to the binodal densities indicated by blue circles. The spinodal region, corresponding to the linear instability of the homogeneous phase, is given by the condition $dP/dv \geq 0$ (green circles). In the two-phase region, the system follows the horizontal line, corresponding to a constant pressure, while the dashed line corresponds to the underlying (unphysical) EOS of the one-phase system. For ABPs, the pressure remains constant in the two phases, while the equal-area construction on P fails, leading to $\Delta A \neq 0$ (Eq. (10.27) and Fig. 10.8a).

10.5.2 Pressure of interacting ABPs

The case of ABPs interacting both with each other and with the walls can be handled using a similar approach as for non-interacting particles (Section 10.5.1), with the master equation (10.20) extended to include the effects of inter-particle interactions. These not only lead to an additional term containing the inter-particle force \mathbf{F} , but also to a description in terms of the *fluctuating hydrodynamic fields* $\hat{\psi}$ and $\hat{\rho}$ [32]:

$$\begin{aligned} \partial_t \hat{\psi} = & -\nabla \cdot \left[(v_0 \mathbf{n} + \mathbf{F}_w) \hat{\psi} + \int \mathbf{F}(\mathbf{r}' - \mathbf{r}) \hat{\rho}(\mathbf{r}') \hat{\psi} d\mathbf{r}' \right] + D \nabla^2 \hat{\psi} + D_R \partial_\theta^2 \hat{\psi} + \\ & + \nabla \cdot \left(\sqrt{2D} \hat{\psi} \boldsymbol{\xi}_t \right) + \partial_\theta \left(\sqrt{2D_R} \hat{\psi} \boldsymbol{\xi}_r \right), \end{aligned} \quad (10.25)$$

where $\boldsymbol{\xi}_t$ and $\boldsymbol{\xi}_r$ are δ -correlated, unit-variance noise fields. While the coarse-graining process leading to Eq. (10.25) in the interacting case is more demanding due to the multiplicative noise [39], a similar overall strategy as in the non-interacting case can be used to derive the following expression for the mechanical pressure

$P_0(\rho)$ in a suspension of interacting ABPs at homogeneous density ρ [32, 31, 33]:

$$\begin{aligned} P_0 &= D\rho + P_0^D + P_0^A, \\ P_0^D &\equiv \int_{x>\Lambda} dx \int_{x'<\Lambda} d\mathbf{r}' F_x(\mathbf{r} - \mathbf{r}') \langle \hat{\rho}(\mathbf{r}) \hat{\rho}(\mathbf{r}') \rangle, \\ P_0^A &\equiv \frac{v_0 v(\rho)}{2D_R} \rho, \end{aligned} \quad (10.26)$$

where we call P_0^D the *direct pressure* and P_0^A the *active pressure*. The integral in the definition of P_0^D measures the average inter-particle force F_x transmitted across a virtual plane Λ perpendicular to the x direction in the bulk fluid. This expression is equivalent to that of the standard virial pressure in equilibrium fluids [11, 32], and can be more succinctly expressed as $P_0^D = (2V)^{-1} \langle \mathbf{F}_i \cdot \mathbf{r}_i \rangle$, with V the total system volume. While equivalent to its equilibrium definition, P_0^D implicitly depends on activity through the density correlator $\langle \hat{\rho}(\mathbf{r}) \hat{\rho}(\mathbf{r}') \rangle$. P_0^A , on the other hand, is an intrinsically active contribution that vanishes in the passive limit $v_0 \rightarrow 0$. It differs from the active term in the ideal pressure in Eq. (10.24) in that one of the v_0 factors has been changed into the average, density-dependent particle speed $v(\rho)$, measured along the particle orientation vector \mathbf{n}_i , i.e., $v(\rho) \equiv v_0 + \langle \mathbf{F}_i \cdot \mathbf{n}_i \rangle$. For purely repulsive forces, \mathbf{F}_i is typically antiparallel to \mathbf{n}_i , since head-on collisions are more frequent than rear collisions; thus, $v(\rho) \leq v_0$, in accordance with the mean-field QSAPs mapping. Just as for non-interacting ABPs, neither of the terms in Eq. (10.26) depends on \mathbf{F}_w , meaning that the mechanical pressure remains an EOS (Fig. 10.7a); crucially, however, the EOS property breaks down for QSAPs, which interact through non-pair-wise forces [38]. In addition to the requirement of no particle–wall torques, the EOS property for interacting ABPs holds only as long as there are no inter-particle alignment interactions. These two requirements are clearly violated in most (if not all) experimental examples of ABPs, where particles will reorient either due to their anisotropic shapes (such as for swimming bacteria) or due to long-ranged hydrodynamic or chemotactic interactions. Nevertheless, analyzing pressure as an effective thermodynamic variable in the idealized, torque-free ABP limit can give important fundamental knowledge about the possibilities (and

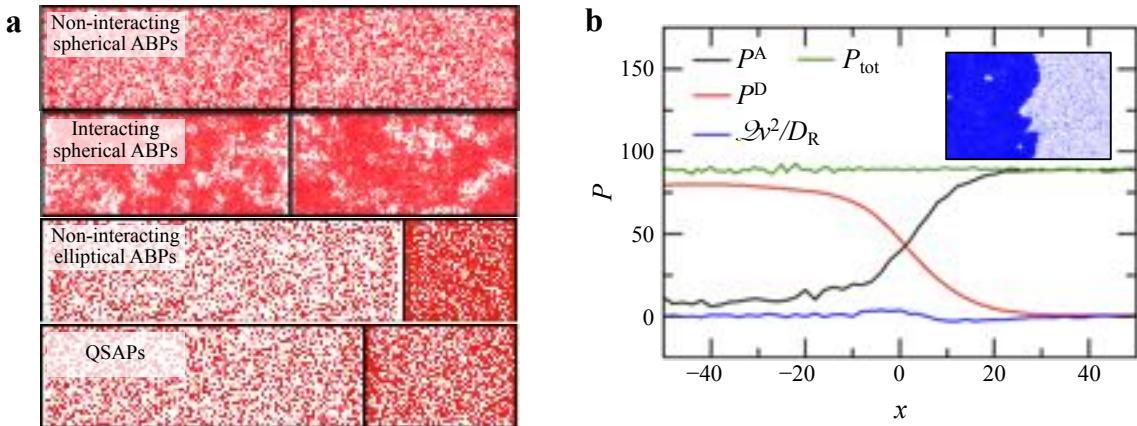


Figure 10.7: **Pressure is an equation of state for spherical ABPs.** (a) Snapshots of steady-state configurations in two sub-systems separated by a mobile, asymmetric wall with different F_w on its two sides. For non-interacting ABPs and for ABPs interacting via the spherical WCA potential Eq. (10.15), the density, and thus the bulk pressure, is independent of the wall potential, showing that mechanical pressure is an EOS in these systems. For elliptical ABPs that experience particle–wall torques and for QSAPs, the bulk density depends on the wall potential, showing that pressure is not an EOS. Reproduced from Ref. [38]. (b) Components of the pressure measured across a liquid–gas interface (inset). The active pressure $P^A = P_0^A + P_1^A$ and direct pressure $P^D = P_0^D + P_1^D$ contain both bulk and interfacial contributions, while the nematic contribution $\mathcal{Q}v_0^2/D_R$ is zero away from the interface. The sum P_{tot} of the three terms is constant across the interface, verifying that mechanical pressure is equal in coexisting phases. Reproduced from Ref. [22].

limitations) of equilibrium mappings in active matter systems. In the next section, we will thus connect the EOS property of pressure in interacting ABP suspensions with their MIPS coexistence densities.

10.5.3 Pressure and phase coexistence in ABPs

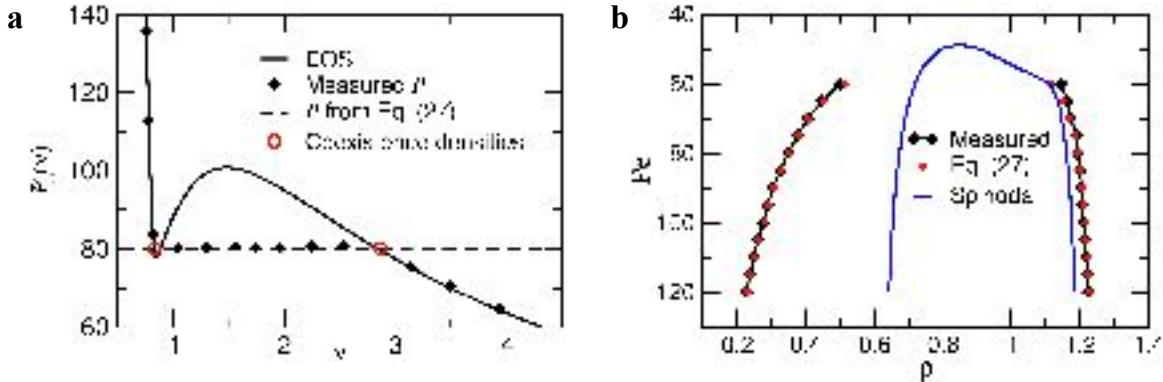


Figure 10.8: **Unequal area construction in ABPs.** (a) $P_0(v)$ for ABPs at $\text{Pe} = 80$, obtained from fitting the pressure contributions in Eq. (10.26) in the low-Pe regime together with scaling arguments extending it into the two-phase region. Dashed line shows predicted coexistence pressure \bar{P} according to Eq. (10.27) using values of $P_1(x)$ and $v(x)$ measured from simulations, diamonds the numerically measured pressures across the tieline, and red circles the measured coexistence densities, in good accordance with the theoretical prediction. (b) Measured (black symbols) and predicted (red symbols) coexistence densities at various Pe from Eq. (10.27). The solid line shows the spinodal determined by the condition $dP_0/dv > 0$. Reproduced from Ref. [22].

An important corollary of the EOS nature of the mechanical pressure in ABPs is that the pressure in coexisting phases must be identical. For non-interacting ABPs, P_1 of Eq. (10.24) is a monotonically increasing function of ρ , implying that phase coexistence is not possible. For interacting ABPs, the pressure is composed of one term, P_0^D , that increases monotonically with density (assuming that \mathbf{F} is purely repulsive), and another, P_0^A , that depends on the product $v(\rho)\rho$. Since $v(\rho)$ decreases with ρ , $P_0^A(\rho)$ is generally non-monotonic, implying that a liquid phase at density ρ_l and a gas phase at density ρ_g can coexist at the same total pressure, $P_0(\rho_l) = P_0(\rho_g) = \bar{P}$ (Fig. 10.7b).

To determine the coexistence densities, it is not sufficient to consider the bulk pressure, but we rather need to analyze the full pressure profile across the interface of the coexisting gas and liquid phases. To this end, we formally decompose the total pressure P as $P = P_0(\rho(x)) + P_1([\rho],x)$, where P_1 contains all *non-local contributions* to the pressure, which depend on gradients of ρ of all orders. Since mechanical stability requires the pressure to be constant throughout the whole system, we have that $P_0(\rho(x)) - P = P_1([\rho],x)$ at phase coexistence. We now view the bulk pressure P_0 as a function of the volume per particle v and integrate both sides to yield

$$\int_{v_l}^{v_g} [P_0(v) - \bar{P}] dv = \int_{x_\ell}^{x_g} P_1(x) \partial_x v dx \equiv \Delta A, \quad (10.27)$$

where the $\partial_x v$ factor on the right-hand side comes from the variable substitution, and the integral runs from a point x_ℓ in the bulk liquid to a point x_g in the gas, where P_1 vanishes. For phase separation driven by a free energy density, as described through the standard equilibrium *Model B* [40], the integrand on the right-hand side is proportional to the interfacial tension κ , and can be written as the total derivative [22]

$$P_1(x) \partial_x v = \partial_x \left(\frac{\kappa (\partial_x \rho)^2}{2\rho} \right). \quad (10.28)$$

Since $\partial_x \rho = 0$ in both bulk phases, in the equilibrium case we get $\Delta A = 0$, so that the coexistence condition Eq. (10.27) becomes independent of κ . Eq. (10.27) is then identical to the equal area construction expressed geometrically in Fig. 10.6; thermodynamically, setting $\Delta A = 0$ is thus equivalent to imposing the equality of chemical potentials in coexisting bulk phases.

For ABPs, the interfacial contributions to the pressure are given by [32, 22]

$$P_1[\rho] = P_1^D[\rho] + P_1^A[\rho] + \frac{v_0^2}{D_R} \mathcal{Q} - \frac{v_0 D}{D_R} \partial_x \mathcal{P}, \quad (10.29)$$

where P_1^D and P_1^A contain gradient contributions to the direct and active pressure that can be expressed as averages of microscopic correlators [22], and \mathcal{P} and \mathcal{Q} , as before, denote the polar and nematic order normal to the interface, both of which vanish in the disordered bulk phases. Unlike the equilibrium case, P_1 cannot generally be expressed as a total derivative, and we thus expect $\Delta A \neq 0$; thus, ΔA can be said to quantify the *violation* of the equilibrium Maxwell construction in coexisting ABPs. If we know the EOS $P_0(v)$, the interfacial pressure $P_1[v]$, and the (inverse) density profile $v(x)$ for each Pe, we can numerically calculate ΔA on the right-hand side, and then adjust \bar{P} in the left-hand-side integral until its value becomes equal. This value of \bar{P} then corresponds to the coexistence pressure, while the corresponding v_l and v_g give the coexistence densities, in the same way as for the Maxwell construction (*c.f.* Fig. 10.6). An accurate form of the EOS for the full range of Pe can be constructed by fitting $v(\rho)$ and $P_0^D(\rho)$ of Eq. (10.26) in the homogeneous phase ($Pe < Pe_c$), and extrapolating their behavior into the $Pe > Pe_c$ region using straightforward scaling arguments [22] (see Ex. 10.4). However, the interfacial terms in $P_1[\rho]$ and the density profile $\rho(x)$ cannot be straightforwardly expressed by such scaling assumptions, but can be measured from simulations in the phase-separated state set up with a flat gas-liquid interface (Fig. 10.7b). In this way, the binodal densities v_g^{-1} and v_l^{-1} can be predicted and compared to their directly measured values at each Pe. As seen in Fig. 10.8, this procedure confirms the validity of the criterion in Eq. (10.27), although it does not constitute a predictive theory as long as we do not have *a priori* knowledge of $P_1[\rho]$.

According to Eq. (10.27), the violation of the equilibrium equal-area construction is measured by interfacial terms in the pressure, highlighting the strong density–polarization coupling underlying MIPS: Near the gas–liquid interface, particles are on average oriented towards the dense phase, creating a net polarization. Due to the self-propulsion, this polarization is coupled to the flux balance encoded in the pressures P_0 and P_1 , which in turn sets the steady-state coexistence densities. This phenomenology highlights a qualitative difference from the equilibrium case, where, as we showed above, interfacial terms in the free energy strictly do not affect properties of the bulk phases. Taken together, the results discussed here highlight both the surprisingly strong analogies and the subtle but important differences between MIPS and equilibrium phase coexistence at the mesoscopic scale.

Example 10.4: Equation of state for ABP. An accurate EOS for interacting ABPs can be obtained by separately considering P_0^A and P_0^D in Eq. (10.26). Rephrased in terms of Pe, the former reads

$$P_0^A = \frac{\rho v(\rho)}{6} Pe, \quad (10.30)$$

where we use Lennard-Jones units for which $2R = 1$. The effective $v(\rho)$ for interacting ABPs is known to be linear up to high densities, where higher-order terms become important. This is captured by the following form of $v(\rho)$:

$$v(\rho) = v_0 g(\rho)(1 - a_0 \rho + a_1 \rho^2), \quad (10.31)$$

where the regularization

$$g(\rho) = \frac{1}{2} (1 - \tanh[a_2(\rho - \rho_*)]) \quad (10.32)$$

ensures that v smoothly approaches zero as ρ approaches ρ_* , and $a_0 - a_2$ are fitting parameters. For ABPs interacting via the WCA potential (Fig. 10.3), P_0^D is accurately described by a biexponential function:

$$P_0^D = d_0 [\exp(d_1 \rho) - 1] + d_2 [\exp(d_3 \rho) - 1], \quad (10.33)$$

where $d_1 - d_3$ are fitting parameters. Note that, as long as Pe is varied by changing D_R while keeping v_0 constant, P_0^D is to a good approximation Pe-independent.

a. Using the above expressions and the parameter values in Table 10.1, plot the EOS $P(v)$ with $v = \rho^{-1}$ for a few values of Pe below and above the transition to MIPS and approximately locate Pe_c by noticing where the EOS becomes nonmonotonic.

b. Numerically perform the equilibrium Maxwell construction (Eq. (10.27) with $\Delta A = 0$) for $\text{Pe} > \text{Pe}_c$ and show that the predicted coexistence densities fail to reproduce the coexistence densities measured from ABP simulations given in Table 10.2 and in Fig. 10.8.

v_0	a_0	a_1	a_2	ρ_*	d_0	d_1	d_2	d_3
24	0.898	7.03×10^{-3}	13.6	1.19	4.548×10^{-9}	17.79	0.691	3.886

Table 10.1: EOS parameter values for ABPs interacting via the WCA potential [22].

Pe	60	70	80	90	100	110	120
ρ_g	0.500	0.406	0.349	0.304	0.276	0.256	0.228
ρ_l	1.146	1.174	1.193	1.204	1.212	1.217	1.225

Table 10.2: Pe-dependent coexistence densities measured from ABP simulations [22].

10.6 Current research directions

So far we focused on a rather narrow slice of the extensive research literature on MIPS, with an emphasis on the basic features of MIPS in QSAPs and ABPs. We now give a brief overview of three of the most active research directions within the MIPS field and discuss some open problems related to these.

Phenomenological field theories. Field-theoretical descriptions of active matter models have proven to be one

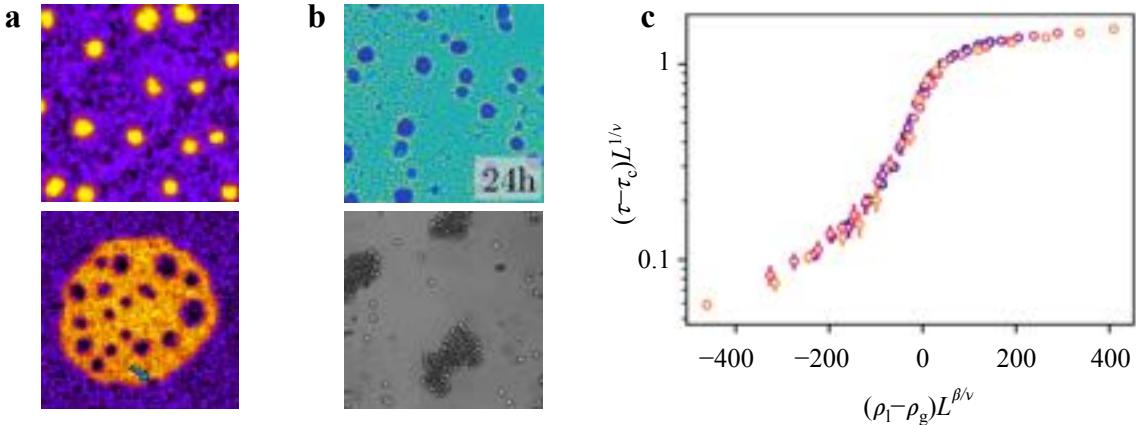


Figure 10.9: **Current topics of MIPS research.** (a) Arrested (top) and bubbly (bottom) phase separation in Active Model B+. In both phases, detailed balance-violating terms in the current induce a reversal of the Ostwald ripening process, leading to steady states with either gas bubbles (black) or clusters (yellow) of a characteristic, finite size. Reproduced from [41] (b) Experimental examples of MIPS in *Myxococcus xanthus* bacteria (top) and self-propelled Janus colloids (bottom). Reproduced respectively from [42] and [43] (c) Critical behavior of AOUPs. The plot shows a collapse of the difference $\rho_l - \rho_g$ for different system sizes L (different colored symbols) when rescaled using the Ising exponents $\nu = 1$ and $\beta = 1/8$. The abscissa shows the rescaled distance from the critical value τ_c of the AOUP persistence time τ (Eq. (10.14)). Reproduced from Ref. [44].

of the most powerful tools to describe their collective behavior [45, 46]. Most of these are *microscopic field theories* that build on an explicit coarse-graining of the equations of motion, such as those outlined in Section 10.5. These have provided, for example, predictions for the pressure [32, 33, 34] and spinodal line [47, 48] in terms of the microscopic parameters, often based on input data from simulations such as pair distribution functions and density profiles. An alternative approach is to use *phenomenological field theories*, where equations of motion for the order parameter fields are instead formulated based on the underlying symmetries and conservation laws. When describing MIPS, the relevant starting point is the equilibrium *Model B* (or *Cahn-Hilliard equation*), which describes phase separation in fluids with local diffusive dynamics and conserved density [40]. This identification relies on the fact that, at bulk level, MIPS can be described by the free energy structure in Eq. (10.7) [5]. Modifications of Model B to take into account the non-equilibrium nature of higher-order gradient terms and the possibility of circulating steady-state currents has respectively lead to the field theories dubbed *Active Model B* [19] and *Active Model B+* [41], which generically describe phase separation in dry, scalar active matter. For certain parameter values, Active Model B+ shows a reversal of the Ostwald ripening process, leading to arrested phase separation with finite-size particle clusters or spontaneously forming gas bubbles in the dense phase [41] (Fig. 10.9a). The latter phenomenology — *bubbly phase separation* — is similar to what is observed in ABP simulations [49, 50] (Fig. 10.4), while arrested particle clustering has been observed in experiments on active Janus colloids [51, 43] and simulations of ABPs with attractions [52]. The fundamental properties of such phenomenological field theories as well as their connection to microscopic active particle models are therefore central open questions in our theoretical understanding of MIPS.

Experimental manifestations of MIPS. Since the QSAP, ABP, and AOUP models are idealized models of biological or synthetic microswimmers, the vast majority of results on MIPS are theoretical or computational. However, the self-trapping mechanism leading to MIPS has also been confirmed in synthetic self-propelled colloids [51, 43] as well as in the bacterium *Myxococcus xanthus* [42] (Fig. 10.9b), which moves by sliding on solid substrates. Since most experimental realizations of active particles at the microscopic level are surrounded by a fluid, the role of hydrodynamic, phoretic and other specific interactions will change the collective behavior even though the basic MIPS mechanism remains robust. One promising avenue to quantitatively test analytical and computational predictions on MIPS is therefore to instead use macroscopic realizations of ABPs, such as grains on a vibrated plate [53], or systems of autonomous robots [54]. Another open problem is to build a deeper knowledge of the importance of the MIPS self-trapping mechanism in microbiology, for example in pattern formation [55] and the initial stages of biofilm formation [17].

Critical behavior in MIPS. Understanding the behavior of phase-separating systems near their critical point is one of the major achievements of modern statistical physics. This understanding is built on the notion that systems can be grouped into a set of universality classes based on their behavior on long length- and timescales where the microscopic details become irrelevant [40]. Since it is not a priori obvious to which universality class(es) ABPs, AOUPs and QSAPs belong, a lot of effort has been put into determining this numerically by measuring critical exponents as the suspensions approach their MIPS critical point [44, 56, 57, 58]. This is highly challenging, due to the diverging correlation lengths near the critical point and the resulting strong finite-size effects. Hence, there is currently a number of conflicting findings for these exponents: Simulations of AOUP suspensions [44] and lattice models of ABPs [56] point towards the same critical exponents as in the Ising universality class (Fig. 10.9c), just as for equilibrium gas-liquid coexistence, while other simulations of ABPs on-lattice [57] and off-lattice [58] show values inconsistent with the Ising exponents. Understanding these inconsistencies, and whether ABPs, AOUPs and QSAPs belong to the same universality class, thus poses a major computational and theoretical challenge.

Bibliography

- [1] G. Frangipane, D. Dell’Arciprete, S. Petracchini, C. Maggi, F. Saglimbeni, S. Bianchi, G. Vizsnyiczai, M. L. Bernardini, and R. Di Leonardo. Dynamic density shaping of photokinetic *E. coli*. *eLife*, 7:e36608, 2018.
- [2] M. E. Cates and J. Tailleur. Motility-induced phase separation. *Annual Review of Condensed Matter Physics*, 6:219, 2015.

- [3] G. Gonnella, D. Marenduzzo, A. Suma, and A. Tiribocchi. Motility-induced phase separation and coarsening in active matter. *Comptes Rendus Physique*, 16:316, 2015.
- [4] J. Palacci, C. Cottin-Bizonne, C. Ybert, and L. Bocquet. Sedimentation and effective temperature of active colloidal suspensions. *Physical Review Letters*, 105:088304, 2010.
- [5] J. Tailleur and M. E. Cates. Statistical mechanics of interacting run-and-tumble bacteria. *Physical Review Letters*, 100:218103, 2008.
- [6] F. Turci and N. B. Wilding. Phase separation and multibody effects in three-dimensional active Brownian particles. *Physical Review Letters*, 126:038002, 2021.
- [7] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur. Generalized thermodynamics of phase equilibria in scalar active matter. *Physical Review E*, 97:020602, 2018.
- [8] S. Hermann, P. Krinnerger, D. de las Heras, and M. Schmidt. Phase coexistence of active Brownian particles. *Physical Review E*, 100:052604, 2019.
- [9] S. Paliwal, J. Rodenburg, R. van Roij, and M. Dijkstra. Chemical potential in active systems: Predicting phase equilibrium from bulk equations of state? *New Journal of Physics*, 20:015003, 2018.
- [10] S. A. Mallory, Šarić A., C. Valeriani, and A. Cacciuto. Anomalous thermomechanical properties of a self-propelled colloidal fluid. *Physical Review E*, 89:052303, 2014.
- [11] S. C. Takatori and J. F. Brady. Towards a thermodynamics of active matter. *Physical Review E*, 91:032117, 2015.
- [12] P. Digregorio, D. Levis, A. Suma, L. F. Cugliandolo, G. Gonnella, and I. Pagonabarraga. Full phase diagram of active Brownian disks: From melting to motility-induced phase separation. *Physical Review Letters*, 121:098003, 2018.
- [13] J.U. Klamser, S.C. Kapfer, and W. Krauth. Thermodynamic phases in two-dimensional active matter. *Nature Communications*, 9:5045, 2018.
- [14] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339:936, 2013.
- [15] J. Arlt, V. A. Martinez, A. Dawson, T. Pilizota, and W. C. K. Poon. Painting with light-powered bacteria. *Nature Communications*, 9:768, 2018.
- [16] M. J. Schnitzer. Theory of continuum random walks and application to chemotaxis. *Physical Review E*, 48:2553, 1993.
- [17] I. Grobas, M. Polin, and M. Asally. Swarming bacteria undergo localized dynamic phase transition to form stress-induced biofilms. *eLife*, 10:e62632, 2021.
- [18] J. Stenhammar, A. Tiribocchi, R. J. Allen, D. Marenduzzo, and M. E. Cates. Continuum theory of phase separation kinetics for active Brownian particles. *Physical Review Letters*, 111:145702, 2013.
- [19] R. Wittkowski, A. Tiribocchi, J. Stenhammar, R. J. Allen, D. Marenduzzo, and M. E. Cates. Scalar φ^4 field theory for active-particle phase separation. *Nature Communications*, 5:4351, 2014.
- [20] J. Stenhammar, D. Marenduzzo, R. J. Allen, and M. E. Cates. Phase behaviour of active Brownian particles: The role of dimensionality. *Soft Matter*, 10:1489, 2014.
- [21] A. Wysocki, R. G. Winkler, and G. Gompper. Cooperative motion of active Brownian spheres in three-dimensional dense suspensions. *Europhysics Letters*, 105:48004, 2014.
- [22] A. P. Solon, J. Stenhammar, M. E. Cates, Y. Kafri, and J. Tailleur. Generalized thermodynamics of motility-induced phase separation: Phase equilibria, Laplace pressure, and change of ensembles. *New Journal of Physics*, 20:075001, 2018.

- [23] M. E. Cates and J. Tailleur. When are active Brownian particles and run-and-tumble particles equivalent? Consequences for motility-induced phase separation. *Europhysics Letters*, 101:20010, 2013.
- [24] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier. Active Brownian particles: From individual to collective stochastic dynamics. *European Physics Journal Special Topics*, 202:1, 2012.
- [25] D. Martin, J. O’Byrne, M. E. Cates, É. Fodor, C. Nardini, J. Tailleur, and F. van Wijland. Statistical mechanics of active ornstein-uhlenbeck particles. *Physical Review E*, 103:032607, 2021.
- [26] G. Szamel. Self-propelled particle in an external potential: Existence of an effective temperature. *Physical Review E*, 90:012111, 2014.
- [27] C. Maggi, U. Marini Bettolo Marconi, N. Gnan, and R. di Leonardo. Multidimensional stationary probability distribution for interacting active particles. *Scientific Reports*, 5:10742, 2015.
- [28] L. Dabelow and R. Eichhorn. Irreversibility in active matter: General framework for active Ornstein-Uhlenbeck particles. *Frontiers in Physics*, 8:516, 2021.
- [29] Y. Fily and M. C. Marchetti. Athermal phase separation of self-propelled particles with no alignment. *Physical Review Letters*, 108:235702, 2012.
- [30] G. S. Redner, M. F. Hagan, and A. Baskaran. Structure and dynamics of a phase-separating active colloidal fluid. *Physical Review Letters*, 110:055701, 2013.
- [31] S. C. Takatori, W. Yan, and J. F. Brady. Swim pressure: Stress generation in active matter. *Physical Review Letters*, 113:028103, 2014.
- [32] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur. Pressure and phase equilibria in interacting active Brownian spheres. *Physical Review Letters*, 114:198301, 2015.
- [33] R. G. Winkler, A. Wysocki, and G. Gompper. Virial pressure in systems of spherical active Brownian particles. *Soft Matter*, 11:6680, 2015.
- [34] T. Speck and R. L. Jack. Ideal bulk pressure of active Brownian particles. *Physical Review E*, 93:062605, 2016.
- [35] F. Ginot, I. Theurkauff, D. Levis, C. Ybert, L. Bocquet, L. Berthier, and C. Cottin-Bizonne. Nonequilibrium equation of state in suspensions of active colloids. *Physical Review X*, 5:011004, 2015.
- [36] I. Petrelli, L. F. Cugliandolo, G. Gonnella, and A. Suma. Effective temperatures in inhomogeneous passive and active bidimensional Brownian particle systems. *Physical Review E*, 102:012609, 2020.
- [37] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. Wiley, 2nd edition, 1985.
- [38] A. P. Solon, Y. Fily, A. Baskaran, M. E. Cates, Y. Kafri, M. Kardar, and J. Tailleur. Pressure is not a state function for generic active fluids. *Nature Physics*, 11:673, 2015.
- [39] D. S. Dean. Langevin equation for the density of a system of interacting Langevin processes. *Journal of Physics A: Mathematical and General*, 29:L613, 1996.
- [40] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49:435, 1977.
- [41] E. Tjhung, C. Nardini, and M. E. Cates. Cluster phases and bubbly phase separation in active fluids: Reversal of the Ostwald process. *Physical Review X*, 8:031080, 2018.
- [42] G. Liu, A. Patch, F. Bahar, D. Yllanes, R. D. Welch, M. C. Marchetti, S. Thutupalli, and J. W. Shaevitz. Self-driven phase transitions drive *myxococcus xanthus* fruiting body formation. *Physical Review Letters*, 122:248102, 2019.

- [43] I. Buttinoni, J. Bialké, F. Kümmel, H. Löwen, C. Bechinger, and T. Speck. Dynamical clustering and phase separation in suspensions of self-propelled colloidal particles. *Physical Review Letters*, 110:238301, 2013.
- [44] C. Maggi, M. Paoluzzi, A. Crisanti, E. Zaccarelli, and N. Gnan. Universality class of the motility-induced critical point in large scale off-lattice simulations of active particles. *Soft Matter*, 17:3807, 2021.
- [45] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85:1143, 2013.
- [46] M. E. Cates. Active field theories. *arXiv:1904.01330*, 2019.
- [47] J. Bialké, H. Löwen, and T. Speck. Microscopic theory for the phase separation of self-propelled repulsive disks. *Europhysics Letters*, 103:30008, 2013.
- [48] J. Bickmann and R. Wittkowski. Predictive local field theory for interacting active Brownian spheres in two spatial dimensions. *Journal of Physics: Condensed Matter*, 32:214001, 2020.
- [49] C. B. Caporusso, P. Digregorio, D. Levis, L. F. Cugliandolo, and G. Gonnella. Motility-induced microphase and macrophase separation in a two-dimensional active Brownian particle system. *Physical Review Letters*, 125:178004, 2020.
- [50] X.-Q. Shi, G. Fausti, H. Chaté, C. Nardini, and A. Solon. Self-organized critical coexistence phase in repulsive active particles. *Physical Review Letters*, 125:168001, 2020.
- [51] I. Theurkauff, C. Cottin-Bizonne, J. Palacci, C. Ybert, and L. Bocquet. Dynamic clustering in active colloidal suspensions with chemical signaling. *Physical Review Letters*, 108:268303, 2012.
- [52] F. Alarcón, C. Valeriani, and I. Pagonabarraga. Morphology of clusters of attractive dry and wet self-propelled spherical particle suspensions. *Soft Matter*, 13:814, 2017.
- [53] J. Deseigne, O. Dauchot, and H. Chaté. Collective motion of vibrated polar disks. *Physical Review Letters*, 105:098001, 2010.
- [54] P. Chvykov, T. A. Berrueta, A. Vardhan, W. Savoie, A. Samland, T. D. Murphey, K. Wiesenfeld, D. I. Goldman, and J. L. England. Low rattling: A predictive principle for self-organization in active collectives. *Science*, 371:90, 2021.
- [55] M. E. Cates, D. Marenduzzo, I. Pagonabarraga, and J. Tailleur. Arrested phase separation in reproducing bacteria creates a generic route to pattern formation. *Proceedings of the National Academy of Sciences*, 107:11715–11720, 2010.
- [56] B. Partridge and C. F. Lee. Critical motility-induced phase separation belongs to the Ising universality class. *Physical Review Letters*, 123:068002, 2019.
- [57] J. T. Siebert, F. Dittrich, F. Schmid, K. Binder, T. Speck, and P. Virnau. Critical behavior of active Brownian particles. *Physical Review E*, 98:030601, 2018.
- [58] F. Dittrich, T. Speck, and P. Virnau. Critical behaviour in active lattice models of motility-induced phase separation. *The European Physical Journal E*, 44:53, 2021.

Chapter 11

Liquid Crystals

GARETH P. ALEXANDER

Liquid crystals are remarkable materials occupying the space between regular solids and simple liquids [1, 2]. The simplest of them is the *nematic phase*, shown in Fig. 11.1, which is a fluid of rod-like molecules with long-range alignment in their orientations, corresponding to a spontaneous breaking of full rotational symmetry. Other phases break other symmetries. For instance, in the *smectic phase*, there is broken translational symmetry along one spatial direction and, in the *columnar phase*, translational symmetry is broken along two directions. In each case, associated to the symmetry breaking are soft degrees of freedom with long-range elasticity and topological defects characteristic of the symmetry breaking. Controlling these soft degrees of freedom (and defects) underpins the uses of liquid crystals in displays, sensors and other optical devices. The traditional liquid crystals have continued to garner new developments, particularly in geometric and topological properties and in the realization of new phases. Alongside this, there has been a rapid growth of interest in *active liquid crystals*, where the constituents continually consume energy to generate work or motion and maintain dynamic, non-equilibrium states [3, 4, 5]. Active liquid crystals bring together the full gamut of liquid crystal physics — order and elasticity, defects, and hydrodynamics —, so it is useful to have a firm background in ordinary liquid crystals to understand their properties. This Chapter is an amuse-bouche for liquid crystals, covering briefly a selection of topics that have been prominent in active liquid crystals. The focus is heavily on nematics as the simplest and prototypical liquid crystal phase, covering aspects of their geometric elasticity, topological defects,

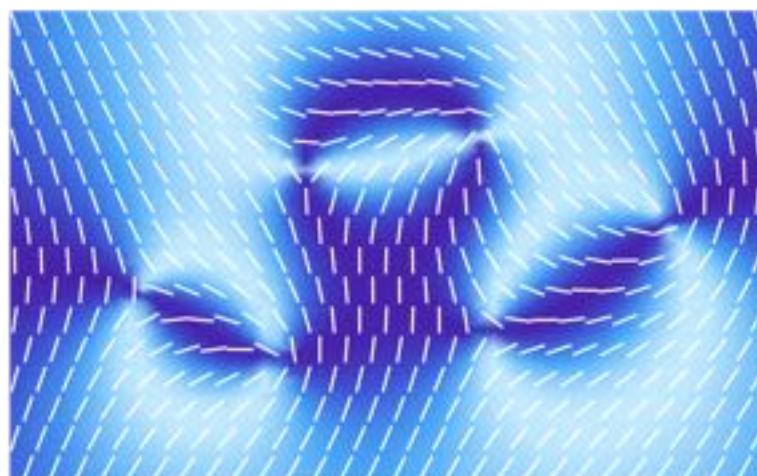


Figure 11.1: **Texture of a liquid crystal.** The classic *Schlieren* texture of a liquid crystal observed under crossed polarisers, with the director field superimposed in white. The dark brushes come together at points which are defects in the orientational order.

and hydrodynamics. As a final topic, we include the elasticity of layered phases (smectics and cholesterics) and their undulational instability, which is predicted to occur in active systems.

11.1 Landau–de Gennes theory of the isotropic–nematic transition

What is the order parameter for the liquid crystal phase? A defining feature of the liquid crystalline state are its optical properties — after all, this is why they find uses in displays. Liquid crystals exhibit different dielectric responses (and consequently different refractive indices) parallel (ϵ_{\parallel}) and perpendicular (ϵ_{\perp}) to a preferred direction called the *optic axis*; this behavior is known as *birefringence*. Therefore, in a nematic, the dielectric tensor ($\boldsymbol{\epsilon}$ or ϵ_{ij} in index notation) is equivalent to the form

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{\perp} & 0 & 0 \\ 0 & \epsilon_{\perp} & 0 \\ 0 & 0 & \epsilon_{\parallel} \end{bmatrix} = \frac{\epsilon_{\parallel} + 2\epsilon_{\perp}}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{\epsilon_{\parallel} - \epsilon_{\perp}}{3} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad (11.1)$$

where the first part represents an isotropic response, so that it is only the second term that is specific for the nematic state. We identify it (or any other physical quantity with equivalent properties [1]) with the order parameter for the nematic phase. This is conventionally denoted by \mathbf{Q} and called the *Q-tensor*; it is a traceless, symmetric second-rank tensor. Having identified the order parameter, the Landau expression for the bulk free energy follows straightforwardly:

$$f = \frac{a(T - T_c)}{2} \text{tr} \mathbf{Q}^2 - \frac{b}{3} \text{tr} \mathbf{Q}^3 + \frac{c}{4} (\text{tr} \mathbf{Q}^2)^2, \quad (11.2)$$

where the presence of a cubic order term indicates that the isotropic–nematic transition is first order, as has indeed been observed experimentally. T is the temperature, T_c is the critical temperature, and a , b , and c are temperature-independent positive quantities (see Chapter 2).

11.1.1 Director field

In coordinate-free form, we can write the Q-tensor for a uniaxial (single optic axis) material as

$$\mathbf{Q} = \frac{3S}{2} \left(\mathbf{n}\mathbf{n} - \frac{1}{3} \mathbf{I} \right), \quad (11.3)$$

where \mathbf{n} is a unit vector along the optic axis, called the *director field*, and S is the *scalar order parameter* (Chapter 2). The choice of normalization makes S the eigenvalue of \mathbf{Q} associated to the director field. An important feature of the director is that it is a line field, as the Q-tensor does not distinguish between \mathbf{n} and $-\mathbf{n}$. The space of unit vectors in three dimensions is the (unit) two-sphere S^2 on which the nematic symmetry $\mathbf{n} \sim -\mathbf{n}$ corresponds to an identification of antipodal points *@Gareth: maybe a figure could help clarifying the identification? Could you provide one?*; the space obtained following this identification is called the *real projective plane* \mathbb{RP}^2 and is the *order parameter space* or *ground state manifold* for nematics. *@Gareth: we feel like a figure is needed here, as it is the first chapter on nematics.*

Accounting for the equivalence between \mathbf{n} and $-\mathbf{n}$ is one of the primary factors in preferring a full Q-tensor description in numerical simulations [6]. The Q-tensor also allows for variations in the magnitude of the order, which allows treatment of phase transitions and can be important when there are defects. However, in analytical work, it is often easier and more insightful to work at the level of the director field, which we shall do for the remainder of this chapter.

Example 11.1: Isotropic–Nematic transition.

- a.** Determine the transition temperature and the magnitude of the discontinuity in the scalar order parameter for the Landau-de Gennes free energy, Eq. (11.2), with a uniaxial Q-tensor, Eq. (11.3).
- b.** Regarding the form of the free energy, Eq. (11.2), why is there not a separate cubic term proportional to $\det \mathbf{Q}$ or a quartic term proportional to $\text{tr} \mathbf{Q}^4$?

11.2 Director gradients and Frank free energy

In the ordered phase, unless we are very close to the transition temperature, S is approximately constant at its preferred value and the energy is dominated by slow variations in the orientation of the optic axis. This is a gradient energy, or *elasticity*, not captured by the bulk part of the free energy, Eq. (11.2). The basic elasticity of nematics was formulated by Oseen [7] and Frank [8] and the presentation here is essentially that of Frank adapted by the perspective introduced in Ref. [9]. The aim is to emphasize an intrinsic, geometric analysis of the director field and its gradients, as well as to connect to recent developments [9, 10, 11, 12, 13].

We characterize the director gradients in terms of the natural splitting that the director field gives between directions parallel and perpendicular to it [9]. Associated to this is a local symmetry group, isomorphic to $\text{SO}(2)$ ¹ [Define $\text{SO}(2)$], consisting of rotations about the director field. Gradients are local quantities and can be analyzed at any (generic) point p . We introduce an adapted local coordinate system such that the director is aligned along z at p . Relative to this adapted basis the matrix of gradients, $\nabla \mathbf{n}|_p$, has the form

$$\nabla \mathbf{n}|_p = \left[\begin{array}{cc|c} \partial_x n_x & \partial_x n_y & 0 \\ \partial_y n_x & \partial_y n_y & 0 \\ \hline \partial_z n_x & \partial_z n_y & 0 \end{array} \right], \quad (11.4)$$

where the zero entries arise because \mathbf{n} has unit magnitude and the lines separate directions parallel and perpendicular to the director. The derivatives along the director field form the components of the *bend* $\mathbf{b}|_p = [\partial_z n_x, \partial_z n_y, 0] = (\mathbf{n} \cdot \nabla) \mathbf{n}|_p$. This is a two-component vector lying, at every point, in the plane perpendicular to the director. Geometrically, it is the *curvature* of the director integral curves. The remaining components, in the xy -block of Eq. (11.4), contain the gradients along orthogonal directions, $\nabla_{\perp} \mathbf{n}$. They decompose under the action of the local $\text{SO}(2)$ group into the three independent pieces [9]

$$\nabla_{\perp} \mathbf{n}|_p = \frac{\partial_x n_x + \partial_y n_y}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\partial_x n_y - \partial_y n_x}{2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} \Delta_1 & \Delta_2 \\ \Delta_2 & -\Delta_1 \end{bmatrix}, \quad (11.5)$$

where $\Delta_1 = \frac{1}{2}(\partial_x n_x - \partial_y n_y)$ and $\Delta_2 = \frac{1}{2}(\partial_x n_y + \partial_y n_x)$. The coefficient of the first term is the *splay* $s = \nabla \cdot \mathbf{n}$ and that of the second is the *twist* $t = \mathbf{n} \cdot \nabla \times \mathbf{n}$. The splay represents the mean curvature of the planes orthogonal to the director, while the twist measures the (local) chirality, or handedness, of the director field². These are both isotropic quantities under rotations about the director; however, the third term, Δ (*Delta*), is a traceless, symmetric object that transforms with spin 2, which has been called the *anisotropic orthogonal gradients* or *deviatoric orthogonal gradients* [9]. Its eigendirections define principal curvature directions for the director field and are closely related to the pitch axis in cholesterics. The geometric decomposition of director gradients is summarized in Fig. 11.2.

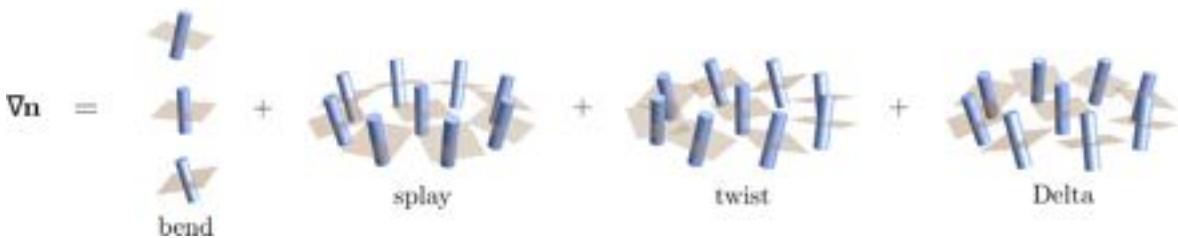


Figure 11.2: **Four different contributions to the director gradients.** Illustration of the breakdown of director gradients into independent parts: bend, splay, twist and Delta. The director field in a neighborhood of a central point p is indicated by the blue cylinders; the planes orthogonal to the director are also shown.

Example 11.2: Director gradients. Determine the structure of the director gradients for:

- a. The planar director field $\mathbf{n} = \cos qx \mathbf{e}_x + \sin qx \mathbf{e}_y$.
- b. The non-planar director field $\mathbf{n} = \cos qz \mathbf{e}_x + \sin qz \mathbf{e}_y$.

¹The full local symmetry group is $D_{\infty h}$, which includes also π rotations about any axis perpendicular to the director.

²The twist is also the obstruction to integrability of the planes orthogonal to the director field.

where q is [Define q]

11.2.1 Frank free energy

We use this geometric decomposition of $\nabla \mathbf{n}$ to derive the *Frank free energy* for nematics [1, 2, 8]. If the four independent parts of the director gradients (the splay, twist, bend, and Δ) contribute separately to the elastic energy, we obtain the free energy density

$$f = \frac{K_s}{2} (\nabla \cdot \mathbf{n})^2 + \frac{K_t}{2} (\mathbf{n} \cdot \nabla \times \mathbf{n} + q_0)^2 + \frac{K_b}{2} |(\mathbf{n} \cdot \nabla) \mathbf{n}|^2 + \frac{K_\Delta}{2} \text{tr} \Delta^2, \quad (11.6)$$

where K_s , K_t , K_b , K_Δ , and q_0 are material-dependent constants; in particular, q_0 is only allowed in chiral materials. There is a general relation³

$$\frac{1}{2} (\nabla \cdot \mathbf{n})^2 + \frac{1}{2} (\mathbf{n} \cdot \nabla \times \mathbf{n})^2 - \text{tr} \Delta^2 + \nabla \cdot [(\mathbf{n} \cdot \nabla) \mathbf{n} - \mathbf{n} (\nabla \cdot \mathbf{n})] = 0, \quad (11.7)$$

which is conventionally used to eliminate the term involving Δ and to discard the total derivative term (called the *saddle-splay*). The free energy density then reduces to the standard form of the Frank free energy

$$f = \frac{K_1}{2} (\nabla \cdot \mathbf{n})^2 + \frac{K_2}{2} (\mathbf{n} \cdot \nabla \times \mathbf{n} + q_0)^2 + \frac{K_3}{2} |(\mathbf{n} \cdot \nabla) \mathbf{n}|^2, \quad (11.8)$$

where K_1 , K_2 and K_3 are the Frank splay, twist and bend elastic constants, respectively. Selinger has given an insightful comparison of the Frank free energy and the form in Eq. (11.6) in how they capture the geometric character of liquid crystal textures [10]. Finally, we mention that a common approximation, known as the *one-elastic-constant approximation*, takes all three Frank elastic constants to be equal and then the free energy density simplifies to the square gradient form (for $q_0 = 0$):

$$f = \frac{K}{2} |\nabla \mathbf{n}|^2. \quad (11.9)$$

In equilibrium, the director field minimizes the free energy $F = \int f dV$. The change in free energy brought about by a small change $\delta \mathbf{n}$ in the director field is (to first order)

$$\delta F = \int_{\Omega} \frac{\partial f}{\partial (\partial_j n_i)} \partial_j \delta n_i dV = \int_{\partial \Omega} \delta n_i \left[\frac{\partial f}{\partial (\partial_j n_i)} \right] dA_j - \int_{\Omega} \delta n_i \left[\partial_j \frac{\partial f}{\partial (\partial_j n_i)} \right] dV. \quad (11.10)$$

The first term represents the contribution from changing the director field at the confining surfaces, for which we assume strong anchoring (Dirichlet) boundary conditions, which eliminate this term. The second contribution gives the response (the term in square brackets) in the bulk to the change in the director; this response is called the *molecular field*, h_i or \mathbf{h} . Since \mathbf{n} is a unit vector, $\delta \mathbf{n}$ has only components in the plane perpendicular to \mathbf{n} and we may take the molecular field to also be confined to this orthogonal plane, i.e., $\mathbf{h} \cdot \mathbf{n} = 0$. The equilibrium condition is then $\mathbf{h} = 0$. For the one elastic constant Frank free energy, we find

$$\mathbf{h} = K [\nabla^2 \mathbf{n} - \mathbf{n} (\mathbf{n} \cdot \nabla^2 \mathbf{n})]. \quad (11.11)$$

Example 11.3: Equilibrium director fields. Within the one-elastic-constant approximation, find the equilibrium director fields for:

- a. A *twisted nematic cell* where the liquid crystal fills the space between two plane surfaces at $z = 0$ and $z = d$ with fixed (*strong anchoring* or *Dirichlet*) tangential boundary conditions $\mathbf{n} = \mathbf{e}_x$ at $z = 0$ and $\mathbf{n} = \mathbf{e}_y$ at $z = d$.
- b. A *hybrid aligned cell* which is as above but with the hybrid (mixed tangential/normal) boundary conditions $\mathbf{n} = \mathbf{e}_x$ at $z = 0$ and $\mathbf{n} = \mathbf{e}_z$ at $z = d$.

³This relation can be derived by a direct calculation starting from $(\nabla \times \mathbf{n})^2$ but its proper significance is that the right-hand-side is the $\mathbf{n} \mathbf{n}$ -component of the Ricci tensor [12, 13]

11.2.2 Electric fields and Frederiks transition

The main ways of controlling the alignment of a liquid crystal are through surface anchoring conditions and applied external electric or magnetic fields. The combination of the two allows for a transition between different states of alignment; this is the *Frederiks transition* that is the basis of operation of liquid crystal displays [1]. The electric field promotes alignment of the director with the field (or perpendicular to it) through the dielectric energy $\frac{1}{2}\mathbf{E} \cdot \boldsymbol{\epsilon} \cdot \mathbf{E}$, which is a contribution $\propto (\mathbf{E} \cdot \mathbf{n})^2$ in terms of the director. The Frank elasticity of the director is therefore

$$F = \int \left\{ \frac{K}{2} |\nabla \mathbf{n}|^2 - \frac{\chi_a}{2} (\mathbf{E} \cdot \mathbf{n})^2 \right\} dV, \quad (11.12)$$

where χ_a is a coupling constant proportional to the dielectric anisotropy. For a cell with tangential surface anchoring where the nematic is initially uniformly aligned along \mathbf{e}_x , a field applied along the cell normal \mathbf{e}_z will cause the director to reorient slightly in the middle of the cell. The threshold field strength at which this happens can be found by taking the linearized perturbation $n_z = \theta_m \sin(\pi z/d)$ to the director field, which gives the critical threshold $E_c = \frac{\pi}{d} \sqrt{K/\chi_a}$, where d is the cell gap. This transition is studied more formally in the problems. **Can we point exactly to which problem? Also, maybe it'd be good to move the relative problems here as examples..**

11.3 Topological defects

Defects are a hallmark of liquid crystal phases that pervade all aspects of their characterization and behavior. They are places of singularity or discontinuity in the order (director field) that imply large distortions and hence increased elastic energies and stresses. The nature of the defects in any phase is closely tied to the broken symmetries and a general classification has been developed in terms of symmetries using methods of homotopy theory [14, 15]. We omit coverage of this and focus instead on aspects of energetics and geometry.

11.3.1 Defects in two dimensions

Defects in two-dimensional nematics are called *disclinations*. We write $\mathbf{n} = \cos \theta \mathbf{e}_x + \sin \theta \mathbf{e}_y$ and describe the director by an angle field θ . Disclinations are windings in this angle field. Some examples are shown in Fig. 11.3a. The nematic symmetry $\mathbf{n} \sim -\mathbf{n}$ implies $\theta \sim \theta + \pi$ and allows for windings by any half-integer multiple of 2π . This half-integer is known as the *winding number, strength, or Frank index* of the disclination, and is a topological invariant for two-dimensional defects. The existence of half-integer winding defects is one of the most distinctive signatures of nematic order and their observation in biological cell cultures, monolayers

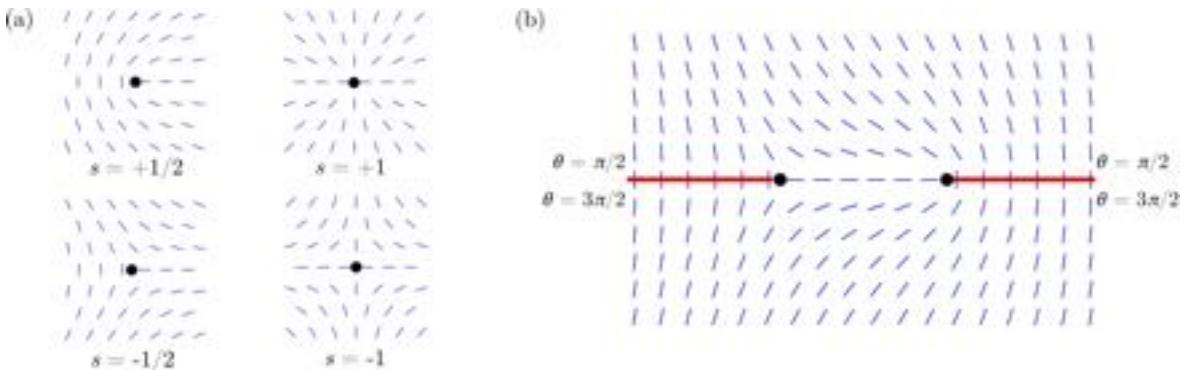


Figure 11.3: **Examples of disclinations.** (a) Local structure of the director field, Eq. (11.13), around two-dimensional point defects with winding numbers $s = \pm 1/2, \pm 1$, and $\theta_0 = 0$. (b) A $\pm 1/2$ pair as described by Eq. (11.14) with $\theta_0 = \pi/2$. A choice of cut surface for the calculation of the energy is indicated in red.

and tissues [16, 17, 18, 19, 20] points to the widespread potential applicability of an (active) nematic formalism in living systems.

With the one-elastic-constant approximation, the Frank free energy becomes $F = \frac{K}{2} \int |\nabla \theta|^2 dA$ and the Euler–Lagrange equation gives that θ is a harmonic function, i.e., $\nabla^2 \theta = 0$. This leads to the basic form

$$\theta(x, y) = s \arctan \frac{y}{x} + \theta_0, \quad (11.13)$$

where θ_0 is a constant and $s \in \frac{1}{2}\mathbb{Z}$ is the winding number. The energy is $F = \pi K s^2 \ln \frac{R}{a}$, where a is a short scale cut-off, the defect core radius, and R is a large-scale cut-off, typically the system size. The defects with the lowest (in magnitude) winding numbers, $s = \pm \frac{1}{2}$, have the lowest energy, while those with higher winding numbers are (usually) unstable, breaking apart into a collection of defects with fundamental windings.

Example 11.4: Defect energy. Verify the result stated for the energy of a disclination.

As Laplace’s equation is linear, multi-defect configurations can also be given easily

$$\theta(x, y) = \sum_a s_a \arctan \frac{y - y_a}{x - x_a} + \theta_0. \quad (11.14)$$

The significance of this representation is that the entire director field is determined by the positions (x_a, y_a) and winding numbers s_a of the defects, which therefore represents an enormous reduction in degrees of freedom; the field theory of the nematic director is reduced to an effective particle theory expressed in terms of the defects, which is frequently called the *Coulomb gas* [2].

From Eq. (11.14) the interaction energy between a pair of $\pm 1/2$ defects can be calculated. One way of doing this is to introduce cut surfaces such that θ is single-valued in the region D obtained by removing the cut surfaces; an example is shown in Fig. 11.3b. Then, we have

$$F = \frac{K}{2} \int_D |\nabla \theta|^2 dA = \frac{K}{2} \int_D [\nabla \cdot (\theta \nabla \theta) - \theta \nabla^2 \theta] dA = \frac{K}{2} \int_{\partial D} \theta \nabla \theta \cdot \mathbf{v} ds, \quad (11.15)$$

where \mathbf{v} is the unit outward normal to D . Carrying out the remaining integral, we obtain an interaction energy

$$F_{\text{int}} = \frac{\pi K}{2} \ln \frac{d}{a}, \quad (11.16)$$

where d is the separation between the two defects. The logarithmic interaction gives rise to an elastic force between the defects that is inversely proportional to their separation. The calculation of the interaction energy is reviewed and expanded upon in the problems. **Can we point specifically to which problems? Also, maybe it'd be good to move the relative problems here as examples.**

11.3.2 Disclination lines and defect loops in three dimensions

In three dimensions, disclinations form extended one-dimensional curves.⁴ They may exist as lines spanning the entire system or as closed loops, knots and links, and the tremendous diversity this affords gives a rich topological classification [21, 22, 23, 24], which has been explored extensively in systems with colloidal inclusions [25, 26, 27]. Recently, disclination loops have gained prominence in three-dimensional active nematics, where they appear as the fundamental excitations with their analysis so far focused on their local structure and geometry [28, 29, 30]. The *local profile* of a disclination line is the structure of the director field on any normal plane. All local profiles are topologically equivalent in that any one of them can be continuously deformed into any other: In the language of homotopy theory, this is the result $\pi_1(\mathbb{RP}^2) = \mathbb{Z}/2$ — there are only two classes, the one corresponding to a disclination line and the one corresponding to a non-singular (defect-free) texture. The global topology of the director field is contained in the variation of the local profile along the contour length of the defect line and the restrictions this imposes on how the director may be extended to the rest of space [21, 22, 23, 24]. Here, we will not discuss the global topology and focus only on the local description.

⁴These are the lines or *threads* from which nematics derive their name — the Greek “*vημα*” meaning “thread”.

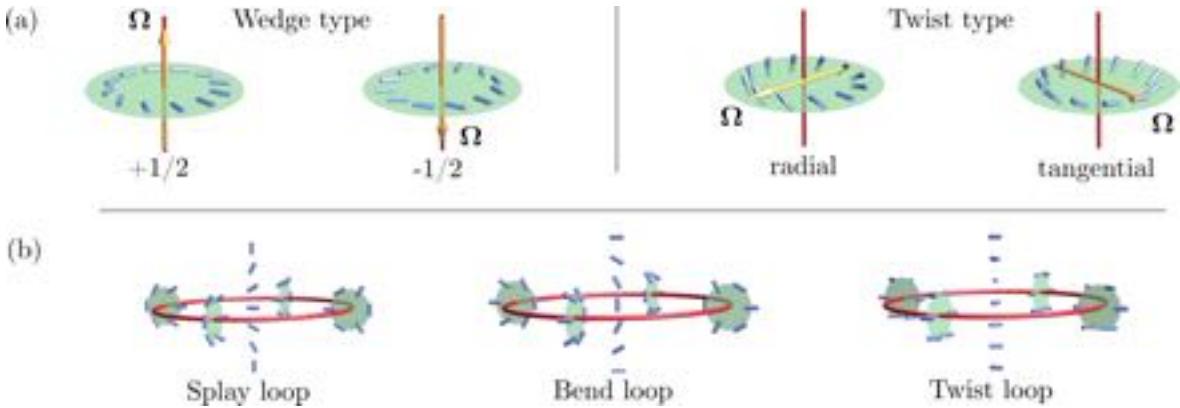


Figure 11.4: Three-dimensional defect lines and loops. (a) Local profiles of defect lines: The blue cylinders show the director in a plane perpendicular to the defect line (red). The rotation vector Ω is shown in orange. The local profile is called *wedge type* when Ω is parallel to the defect line and *twist type* when it is perpendicular. (b) Examples of circular defect loops: They are named *splay*, *bend*, and *twist* according to the predominant director distortion through the center of the loop. The variation in local profile around the loop is indicated at a few points. [The figure looks a bit pixellated]

In the one-constant approximation, the energy minimizing local profiles is still given by director fields that rotate within a single plane; however, this may be any plane in \mathbb{R}^3 and not simply the plane normal to the defect line. This leads to a geometric diversity in three dimensions that is not present in two. We define the plane of rotation of the director by the unit vector Ω normal to it and let t denote the unit vector tangent to the defect line. As Ω can have any orientation relative to t there is a 2-sphere's worth of distinct (energy-minimizing) local profiles. They may be given explicitly by the director field [29, 30]

$$\mathbf{n} = \cos \frac{\phi}{2} \mathbf{m} + \sin \frac{\phi}{2} \Omega \times \mathbf{m}, \quad (11.17)$$

where \mathbf{m} is a unit vector orthogonal to both Ω and t , and ϕ is the azimuthal angle around the line. When Ω is parallel (anti-parallel) to t , this reduces to the $+1/2$ ($-1/2$) profile in two dimensions and the profile is termed *wedge type*; when Ω is perpendicular to t , the profile is called *twist type*. There is a circle's worth of twist type local profiles and we further distinguish them according to whether the vector \mathbf{m} points radially (*radial twist*) or azimuthally (*tangential twist*) away from the line. Examples of these principal types of local profile are given in Fig 11.4a.

A global construction for the director field of a closed defect loop was first introduced by Friedel and de Gennes [31] and is again based on the director rotating within a single plane, orthogonal to the rotation vector Ω . We let $\mathbf{e}_1, \mathbf{e}_2$ be an orthonormal basis for the plane and set

$$\mathbf{n} = \cos \frac{\omega}{4} \mathbf{e}_1 + \sin \frac{\omega}{4} \mathbf{e}_2, \quad (11.18)$$

where ω is the solid angle function [32, 33] for the defect loop. The vector \mathbf{e}_1 gives the asymptotic orientation of the director far from the loop. Although this construction is deceptively simple, it contains enormous generality as it gives a critical point of the one-elastic-constant Frank free energy independently of the shape of the defect loop, or of its orientation relative to the rotation vector Ω and far field director \mathbf{e}_1 . As basic examples of the diversity, we show in Fig. 11.4b three circular defect loops with the same orientation but with different rotation vectors $\Omega = \mathbf{e}_x, \mathbf{e}_y$, and \mathbf{e}_z giving splay, bend, and twist loops, respectively. These different geometries show different responses in active nematics [30].

11.4 Hydrodynamics of nematic liquid crystals

The equations governing the hydrodynamics of nematic liquid crystals were developed by Ericksen [34] and Leslie [35] in the 1960s using the methods of classical continuum mechanics. A separate approach based on the

principles of broken symmetries and conservations laws in condensed matter physics was developed in the early 1970s by the “Harvard group” [36, 37]. The description we give here is a terse summary of the presentation in de Gennes and Prost [1], covering enough of the ideas to follow conceptually but deliberately going light on the details; it is recommended to consult their account for a fuller exposition.

We take the fluid to be incompressible, $\nabla \cdot \mathbf{v} = 0$, and to satisfy the Navier-Stokes equations

$$\rho (\partial_t v_i + v_j \partial_j v_i) = \partial_j \sigma_{ij}. \quad (11.19)$$

The issue then is to determine (i) the form of the stress tensor σ_{ij} accounting for the orientational degrees of freedom, and (ii) the equation for the director. Both of these are obtained by consideration of the processes of dissipation of energy by the fluid flow and director rotation. The total free energy of the flowing nematic is

$$F = \int \left(\frac{1}{2} \rho |\mathbf{v}|^2 + f_{\text{Frank}}(\nabla \mathbf{n}) \right) dV, \quad (11.20)$$

and its rate of dissipation, in an isothermal system, is the rate of entropy production

$$-\frac{dF}{dt} = T \dot{S} = \int (-v_i \partial_j \sigma_{ij} + h_i \dot{n}_i - \sigma_{ij}^E \partial_j v_i) dV. \quad (11.21)$$

The first term comes from the dissipation of kinetic energy, making use of the Navier-Stokes equation, while the second and third terms represent the dissipation associated with the director field. The second term $h_i \dot{n}_i$ is the work done by the rotation of the director within a fluid element; here $\dot{\mathbf{n}} = \partial_t \mathbf{n} + (\mathbf{v} \cdot \nabla) \mathbf{n}$ is the usual convective derivative. The third term $\sigma_{ij}^E \partial_j v_i$ is the change in the elastic distortion energy associated with the motion of material elements; if material elements move relative to each other, even without rotation of the director, this will change the director gradients and hence elastic energy. The term σ_{ij}^E captures this effect and is called the *Ericksen stress*; for the one-elastic-constant free energy $\sigma_{ij}^E = -K \partial_i n_k \partial_j n_k$.

We (i) integrate the first term by parts and then combine it with the Ericksen stress term, (ii) decompose the velocity gradients into symmetric $D_{ij} = \frac{1}{2}(\partial_i v_j + \partial_j v_i)$ and antisymmetric $\Omega_{ij} = \frac{1}{2}(\partial_i v_j - \partial_j v_i)$ parts, and (iii) replace the convective derivative \dot{n}_i with the co-rotational derivative $N_i = \dot{n}_i + \Omega_{ij} n_j$, all of which results in

$$T \dot{S} = \int \left\{ (\sigma_{ij} - \sigma_{ij}^E) D_{ij} + h_i N_i - \left[\sigma_{ij} - \sigma_{ij}^E - \frac{1}{2} (n_i h_j - n_j h_i) \right] \Omega_{ij} \right\} dV. \quad (11.22)$$

Finally, we note that the first term involves only the symmetric part of $\sigma_{ij} - \sigma_{ij}^E$ and the last term only the antisymmetric part. We will call the symmetric part the *viscous stresses* σ_{ij}^{vis} . The antisymmetric part is determined by the observation that pure rotations are a symmetry and must lead to no dissipation. Under a pure rotation only Ω_{ij} is non-zero; both D_{ij} and N_i vanish — this is the physical significance of the co-rotational derivative — and hence the absence of dissipation for pure rotations determines the antisymmetric part of the stress tensor. With this done, the *dissipation* is

$$T \dot{S} = \int (\sigma_{ij}^{\text{vis}} D_{ij} + h_i N_i) dV. \quad (11.23)$$

Example 11.5: Co-rotational derivative. This example illustrates that it is the co-rotational derivative that vanishes for pure rotations. Consider a rigid body rotation of a uniformly aligned nematic sample. Take the rotation to be about the z -axis with angular velocity ω and the director to be aligned along \mathbf{e}_x at $t = 0$.

- a. Write the time-dependent director field $\mathbf{n}(t)$ and compute its time derivative $\partial_t \mathbf{n}$.
- b. Write the velocity field \mathbf{v} for the rotation and compute the velocity gradients. Verify that \mathbf{D} vanishes but $\boldsymbol{\Omega}$ does not. What is the convective derivative $\partial_t \mathbf{n} + (\mathbf{v} \cdot \nabla) \mathbf{n}$?
- c. Combine these calculations to verify that the co-rotational derivative \mathbf{n} is zero.

In linear non-equilibrium thermodynamics, the contributions to the dissipation are grouped as *forces* and *fluxes*, and the forces expanded as a general linear combination of the fluxes. In a simple fluid without director

field (i.e., Eq. (11.23) without the term $h_i N_i$), what this amounts to is an expression for the viscous stress as a linear combination of the shears, which is just the constitutive relation of a Newtonian fluid. Here, we are generalizing that process. If we take $(\sigma_{ij}^{\text{vis}}, h_i)$ as forces and (D_{ijkl}, N_i) as fluxes, then the linear relation is (the equality of the cross terms is known as the *Onsager relation* [38, 39])

$$\sigma_{ij}^{\text{vis}} = A_{ijkl} D_{kl} + B_{ijk} N_k, \quad (11.24)$$

$$h_i = B_{jki} D_{jk} + C_{ij} N_j, \quad (11.25)$$

where the coefficients are the most general expressions compatible with all symmetries. Considering first the molecular field equation, we have $C_{ij} = \gamma_1 \delta_{ij}$ and $B_{ijk} = \gamma_2 (\frac{1}{2} n_i \delta_{jk} + \frac{1}{2} n_j \delta_{ik} - n_i n_j n_k)$, with γ_1 and γ_2 two material-dependent coefficients. The first result is to recognize that the equation for the director dynamics is purely relaxational and therefore given by simply rearranging Eq. (11.25)

$$N_i = \partial_t n_i + v_j \partial_j n_i + \Omega_{ij} n_j = \frac{1}{\gamma_1} h_i - (\gamma_2 / \gamma_1) (D_{ij} n_j - n_i D_{jk} n_j n_k). \quad (11.26)$$

Next, we use this equation to eliminate N_i from the viscous stress in favor of h_i obtaining

$$\sigma_{ij}^{\text{vis}} = A'_{ijkl} D_{kl} + \frac{\gamma_2 / \gamma_1}{2} (n_i h_j + h_i n_j), \quad (11.27)$$

where A'_{ijkl} has the same properties as A_{ijkl} . For our purposes (as in many papers), it will be enough to retain from A'_{ijkl} only the usual viscous stress $2\mu D_{ij}$ of a Newtonian fluid; we leave the determination of its general form as an exercise. Finally, we piece together the full nematic stress tensor

$$\sigma_{ij} = -p \delta_{ij} + 2\mu D_{ij} + \frac{\gamma_2 / \gamma_1}{2} (n_i h_j + h_i n_j) + \sigma_{ij}^E, \quad (11.28)$$

where p is the pressure, which enforces incompressibility. This completes our derivation of the *Ericksen-Leslie equations*: These equations are Eqs. (11.19) and (11.28) for the fluid velocity, and Eq. (11.26) for the director.

For numerical simulations, as discussed in Chapter ??, a version of the hydrodynamic equations in terms of the \mathbf{Q} -tensor is needed, which are known as the *Beris-Edwards equations* [40]. They may be derived by analogy with our presentation for the director field.

Example 11.6: Hydrostatic pressure. Determine the hydrostatic pressure in an equilibrium nematic.
[Hint: In equilibrium $\mathbf{h} = 0$ but the Ericksen stress is non-zero.]

11.4.1 Active nematics

Active nematics are nematic liquid crystals with a sustained supply of free energy at the scale of their constituents [3, 4, 5]. This microscale drive manifests itself macroscopically as nonequilibrium currents and forces. They are described by the same hydrodynamic equations for the director and velocity fields but with an extra contribution, $\sigma_{ij}^a = -\zeta n_i n_j$ (or $-\zeta Q_{ij}$), to the stress tensor. The parameter ζ is a phenomenological quantity, which importantly distinguishes active materials by its sign: When ζ is positive, the activity is *extensile*, while, when it is negative, the activity is *contractile*. This active stress captures at a coarse-grained level the force pairs between neighboring particles associated with the activity, but it is also the simplest term compatible with symmetry that can be added, affording an alternative, model-independent, motivation. Its consequences for material properties and behavior are surveyed in Chapter 12.

11.5 Layered phases including smectics and cholesterics

Smectic liquid crystals are layered phases characterized by a one-dimensional mass density wave [1, 2]. Along one direction, the molecular positions are ordered with a regular spacing a between them, while along the perpendicular directions they are disordered and remain fluid. The ground state is a plane density wave along

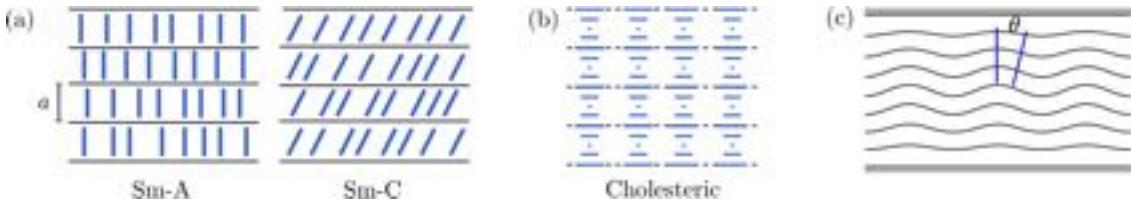


Figure 11.5: **Layered liquid crystal phases.** (a) Smectics: A and C phases. (b) Cholesterics: The director undergoes a helical rotation (between lying in the plane of the page and pointing out of it) along a preferred direction but there is no positional order. (c) Undulational instability, as arises from mechanical strain and also activity.

z (say), $\rho = \rho_0 + \psi_0 \cos q_{\text{sm}} z$ with $q_{\text{sm}} = 2\pi/a$ the ordering wavevector. More generally, the density can be written as

$$\rho = \rho_0 + \text{Re } \psi, \quad (11.29)$$

where ψ is a complex scalar, the order parameter for the smectic transition. Different types of smectic are possible depending on whether the director aligns with the direction of the density modulation, the *layer normal*, or makes an angle with it. In the former case, we have a smectic A and in the latter a smectic C (Fig. 11.5).

Away from the phase transition, the order parameter $\psi = \psi_0 e^{iq_{\text{sm}}\phi}$ [Define q_{sm}] has (approximately) constant magnitude ψ_0 and the smectic can be described just in terms of its phase ϕ . The regular periodicity implies that $\nabla\phi$ should have unit magnitude and, in the smectic A phase, coincide with the director. For instance, in the smectic A ground state $\phi = z$ and $\nabla\phi = \mathbf{e}_z = \mathbf{n}$. The elastic energy describing deformations of this state, and more general smectic textures, comprises two terms: a *compression energy* penalizing deviations from the preferred layer spacing and a *bending energy* penalizing layer curvature. One form of this is

$$F = \int \left\{ \frac{B}{2} (1 - |\nabla\phi|)^2 + \frac{K}{2} (\nabla \cdot \mathbf{n})^2 \right\} dV, \quad (11.30)$$

where B is the bulk (or compression) modulus, K [Can you use a different symbol to avoid confusion with the Frank's coefficients?] is the bending modulus, and $\mathbf{n} = \nabla\phi/|\nabla\phi|$ is the layer normal. In the literature, there are also other forms of the smectic energy that differ in the structure of non-linear terms, but they all reduce to the same linear theory. The linear elasticity is obtained by taking the phase field to be $\phi = z - u$, where u is called the *Eulerian displacement field*, and retaining only terms of quadratic order in gradients of u , which gives

$$F = \int \left\{ \frac{B}{2} (\partial_z u)^2 + \frac{K}{2} (\nabla_{\perp} u)^2 \right\} dV, \quad (11.31)$$

where ∇_{\perp} denotes gradients perpendicular to z .

Example 11.7: Linear elasticity of smectics.

- a. Show that the non-linear energy, Eq. (11.30), reduces to the linear form of Eq. (11.31) when expanded to quadratic order in gradients of the Eulerian displacement field.
- b. Show that within the linear smectic elasticity, Eq. (11.31), the mean square fluctuations in the layer displacement $\langle u^2 \rangle$ diverge logarithmically with sample size (in three dimensions). This divergence is known as the *Landau-Peierls instability*.

Cholesterics are also a *layered* liquid crystal state with several properties analogous to those of smectics, at least at scales large compared to the layer spacing. At a fundamental level, however, they are quite different and, in particular, do not correspond to a density modulation. Cholesterics are chiral nematics: They are locally the same as nematics and are described by a director field and the Frank free energy, Eq. (11.8), but with non-zero q_0 , favoring a natural state of twist. This is minimized by the helical director field

$$\mathbf{n} = \cos q_0 z \mathbf{e}_x + \sin q_0 z \mathbf{e}_y, \quad (11.32)$$

or any equivalent to it by a Euclidean motion. The director lies everywhere in the xy -plane and undergoes a steady rotation along the z -axis. The distance corresponding to a full 2π rotation is called the *pitch*.

The competition between an energetic preference for twist and confinement with frustrating boundary conditions allows cholesterics to display an extremely rich set of geometric and topological states and textures that do not arise in ordinary nematics. These include skyrmions [41], Hopf textures [42] and knotted defect lines [25, 43], among many more. However, we limit our discussion here to the simple helical state in Eq. (11.32) and its analogy with smectics. The connection comes through the phase $q_0 z$ of the helical director rotation, which is the same (just with q_0 in place of q_{sm}) as the phase of the smectic mass density wave. This suggests that on scales large compared to the layer spacing a universal theory of *one-dimensional layers* should apply to both smectics and cholesterics (although the latter may retain chiral couplings that would not be present in achiral smectic A). It is important to keep in mind, however, that the *layer spacing* is very different in cholesterics as compared to smectics; in smectics, the layer spacing is set by the molecular scale and is only a few Ångströms, whereas in cholesterics the pitch is usually several hundred nanometres, or even microns. Nonetheless, a universal description does apply [44] and the elasticity of the cholesteric helix is given by Eq. (11.31).

11.5.1 Undulational instability and active layered phases

The layered structure of a cholesteric or smectic state is unstable to dilatational mechanical strain. This is known as the *Helfrich–Hurault instability*, which was initially introduced as an instability to an applied electric or magnetic field; the mechanical version we present here was first described by Delaye, Ribotta and Durand [45] and, independently, by Clark and Meyer [46]. The reason for the instability is geometric: If the layers tilt locally by an angle θ , then the spacing between them reduces by an amount proportional to θ^2 , which will relieve an extensional strain (Fig. 11.5c). In terms of the Eulerian displacement field $\theta \approx |\nabla_\perp u|$, we get a term $-\frac{1}{2}B\alpha|\nabla_\perp u|^2$ in the energy, where α is the imposed strain. Combining this with the usual layer elasticity of Eq. (11.31) predicts a threshold strain $\alpha_{th} = \frac{2\pi}{d}\sqrt{K/B}$, where d is the cell gap. The details of the instability are covered in a problem [Can we be specific about the problem and/or move it here as an example?](#). The layered state is stable under compression.

This mechanical instability is also the fundamental hydrodynamic instability of active layered systems, for both smectics [47, 48] and cholesterics [49, 50]. The correspondence comes about because the form of the active stress is the same as the mechanical stress induced by an imposed strain. The (linearized) mechanical stress is a *normal–normal* contribution $\sigma_{zz} = B\partial_z u = B\alpha$, for a strain α . Likewise, the active stress, $\sigma^a = -\zeta\nabla\phi\nabla\phi$, is also a normal–normal contribution. Comparing the two suggests that the activity is analogous to a strain $-\zeta/B$ and hence that an active layered phase will be unstable to contractile ($\zeta < 0$) activity, with the same threshold and selected wavevector as in the Helfrich–Hurault instability. This is indeed what has been shown in Refs. [47, 48, 49, 50].

Example 11.8: Extensile cholesterics are contractile smectics. On scales large compared to the pitch, extensile active cholesterics behave like contractile active smectics [49, 50]. Show this by demonstrating that for the director field of the cholesteric ground state, Eq. (11.32), the active stress is

$$\sigma^a = -\zeta \mathbf{n}\mathbf{n} = \frac{\zeta}{2} \mathbf{e}_z \mathbf{e}_z + \dots,$$

where the dots contain terms that vary rapidly on the scale of the pitch, or are isotropic.

11.6 Final remarks

This short account has only managed a superficial survey of the physics of liquid crystals, with central focus on nematics. Notable omissions include the remarkable geometric and topological properties of cholesterics, defects in smectics and cholesterics, droplets, shells and colloidal inclusions, liquid crystal elastomers, and the relatively recently discovered twist-bend and splay-bend nematic phases. Active matter has provided a new direction for liquid crystals, in particular advancing connections with biology. It brings together the full gamut

of liquid crystals: their elasticity, defects and hydrodynamics, which we have tried to cover in the passive setting in this Chapter. The developments in active liquid crystals are described in the next Chapter.

11.7 Problems

Problem 11.1: Biaxial nematics. In ordinary nematics, there is only one optic axis and the index of refraction is the same in all perpendicular directions. Such materials are termed *uniaxial*. If this symmetry is broken, we obtain a *biaxial* material with a second optic axis perpendicular to the director.

- a. Write the general form of the \mathbf{Q} -tensor for a biaxial nematic. [Hint: The \mathbf{Q} -tensor is a traceless, symmetric second-rank tensor.]
- b. Show that the bulk Landau free energy given in Eq. (11.2) is minimized by a uniaxial \mathbf{Q} -tensor.
- c. Would you expect the phase transition between a uniaxial and biaxial nematic to be continuous or first order (in mean field theory)?

Problem 11.2: Frederiks transition. We will now fill in the details of the description of the Frederiks transition. Consider a nematic in a cell with planar surfaces at $z = 0$ and $z = d$, both with tangential anchoring boundary conditions $\mathbf{n} = \mathbf{e}_x$. An electric field of strength E is applied along the cell normal direction, \mathbf{e}_z .

- a. By writing the director field in terms of the angle $\theta(z)$ it makes with the x -axis, $\mathbf{n} = \cos \theta(z) \mathbf{e}_x + \sin \theta(z) \mathbf{e}_z$, show that the free energy given in Eq. (11.12) reduces to

$$F = A \int_0^d \left\{ \frac{K}{2} |\partial_z \theta|^2 - \frac{\chi_a}{2} E^2 \sin^2 \theta \right\} dz,$$

where A is the area of the cell in the xy -plane.

- b. Find a conserved first integral for this functional and determine its value using that θ takes its maximal value θ_m in the cell midplane, i.e., $\theta(d/2) = \theta_m$. [Hint: In classical mechanics this is the Hamiltonian.]
- c. Show that the angle $\theta(z)$ is given by

$$\sin \theta = \sin \theta_m \operatorname{sn} \left(\sqrt{(\chi_a/K)} Ez \mid \sin^2 \theta_m \right),$$

where $\operatorname{sn}(z|m)$ is the Jacobi elliptic function. Show also that θ_m is determined implicitly by the equation

$$F(\sin \theta_m) \equiv \int_0^{\pi/2} \frac{du}{\sqrt{1 - \sin^2 \theta_m \sin^2 u}} = \sqrt{(\chi_a/K)} E \frac{d}{2},$$

where $F(k)$ is the complete elliptic integral of the first kind. By evaluating this in the limit $\theta_m \rightarrow 0$ find the critical field strength E_c for the Frederiks transition.

- d. Plot the director angle $\theta(z)$ through the cell for a field strength just above threshold, say $E = 1.05E_c$, and one considerably above threshold, say $E = 3E_c$.

Problem 11.3: Defect interaction energy. We will now calculate the elastic energy of a pair of defects in terms of their winding numbers and separation.

- a. Making use of Eq. (11.15) and the cut surface shown in Fig. 11.3b, verify the result given in Eq. (11.16) for the energy of a pair of $\pm 1/2$ defects with separation d . How does the answer depend on the choice of cut surface?
- b. Generalize this calculation to a pair of defects with arbitrary winding numbers, s_1 and s_2 , to show that their energy is

$$F = \pi K (s_1 + s_2)^2 \ln \frac{R}{a} - 2\pi K s_1 s_2 \ln \frac{d}{a},$$

where R is a large-scale cut off, a is a short-scale cut off, and d is the separation. [Hint: You can use the same cut surface shown in Fig. 11.3b but need to determine the values of θ on each side of it.]

- c. Show that it is energetically favorable for a single $s = +1$ defect to separate into two $s = +1/2$ defects.

d. Returning to the case of a pair of $\pm 1/2$ defects, let us suppose that they are initially separated along the x -axis with coordinates x_{\pm} and that their motion can be described by

$$\gamma \frac{dx_{\pm}}{dt} = -\frac{dF}{dx_{\pm}},$$

where γ is a frictional drag constant. Determine how their separation evolves with time.

e. How would the motion be changed if the defects were self-propelling (with constant speed)? [Hint: See, for instance, Ref. [51].]

Problem 11.4: Shear flow and Leslie angle. The Ericksen-Leslie equations are complicated and in most cases need to be solved numerically, but there are some situations where the solution can be found directly and which provide general insight into the equations. One is the case of simple shear. When the nematic lies in the plane of the shear flow, it aligns at an angle θ_L relative to the flow direction. This is known as the *Leslie angle*. Consider a simple shear flow $\mathbf{v} = \alpha y \mathbf{e}_x$ and suppose the director aligns uniformly at an angle θ to the x -axis, $\mathbf{n} = \cos \theta \mathbf{e}_x + \sin \theta \mathbf{e}_y$.

a. Determine the form of the stress tensor and verify that the Stokes equation is satisfied with constant pressure.

b. Show that the director equation reduces to

$$\partial_t \theta = -\frac{\alpha}{2} \left(1 + \frac{\gamma_2}{\gamma_1} \cos 2\theta \right),$$

and hence that the Leslie angle is given by $\cos 2\theta_L = -\gamma_1/\gamma_2$. This result provides a direct physical interpretation for the material constant γ_2/γ_1 , which is often simply denoted v and called the *shear alignment parameter*.

c. Solutions exist only for $v < -1$ or $v > +1$. By linearizing around the Leslie angle determine the stability of the solution in each case and hence the stable angle that the director aligns at.

Problem 11.5: Helfrich–Hurault instability. Consider a cholesteric or smectic confined between two glass plates with separation d (along the z -direction) and natural layer spacing a . The plates are moved apart by an amount δ so that the cell gap becomes $d + \delta$. With strong anchoring conditions at both surfaces, the number of layers stays the same and the material is extensionally strained.

a. How many layers are there before the plates are moved apart? What is the new layer spacing after they are moved apart?

b. What is the phase field in the strained state? Linearize your expression in $\delta/d \ll 1$. [Hint: Write this first assuming the layers remain planar and only add a displacement field at the end.]

c. Show that the elastic free energy, Eq. (11.30), becomes

$$F = \int \left\{ \frac{B}{2} \frac{\delta^2}{d^2} + \frac{B}{2} (\partial_z u)^2 - \frac{B\delta}{2d} |\nabla_{\perp} u|^2 + \dots + \frac{K}{2} (\nabla_{\perp}^2 u)^2 \right\} dV. \quad (11.33)$$

The main observation is that the third term favours layer undulation (non-zero $\nabla_{\perp} u$) when δ is positive, i.e., for dilatational strain.

d. Assume an undulation of the form $u = u_0 \sin(\pi z/d) \sin(kx)$ and determine both the threshold for instability and the wavevector k at the threshold.

Bibliography

- [1] P. G. de Gennes and J. Prost. *The Physics of Liquid Crystals*. Oxford University Press, Oxford, 1993.
- [2] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, Cambridge, 1995.
- [3] S. Ramaswamy. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics*, 1(1):323–345, 2010.

- [4] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85:1143–1189, 2013.
- [5] A. Doostmohammadi, J. Ignés-Mullol, J. M. Yeomans, and F. Sagués. Active nematics. *Nature Communications*, 9:3246, 2018.
- [6] M. Ravnik and S. Žumar. Landau-de gennes modelling of nematic liquid crystal colloids. *Liquid Crystals*, 36:1201–1214, 2009.
- [7] C. W. Oseen. The theory of liquid crystals. *Transactions of the Faraday Society*, 29:883–899, 1933.
- [8] F. C. Frank. 1. Liquid crystals. On the theory of liquid crystals. *Discussions of the Faraday Society*, 25:19–28, 1958.
- [9] T. Machon and G. P. Alexander. Umbilic lines in orientational order. *Physical Review X*, 6:011033, 2016.
- [10] J. V. Selinger. Interpretation of saddle-splay and the Oseen-Frank free energy in liquid crystals. *Liquid Crystals Reviews*, 6:129–142, 2018.
- [11] J. Binysh, J. Pollard, and G. P. Alexander. Geometry of bend: Singular lines and defects in twist-bend nematics. *Physical Review Letters*, 125:047801, 2020.
- [12] J. Pollard and G. P. Alexander. Intrinsic geometry and director reconstruction for three-dimensional liquid crystals. *New Journal of Physics*, 23:063006, 2021.
- [13] L. C. B. da Silva and E. Efrati. Moving frames and compatibility conditions for three-dimensional director fields. *New Journal of Physics*, 23:063016, 2021.
- [14] G. Toulouse and M. Kleman. Principles of a classification of defects in ordered media. *Journal de Physique de France Letters*, 37:149–151, 1976.
- [15] N. D. Mermin. The topological theory of defects in ordered media. *Reviews of Modern Physics*, 51:591–648, 1979.
- [16] T. B. Saw, A. Doostmohammadi, V. Nier, L. Kocgozlu, S. Thampi, Y. Toyama, P. Marcq, C. T. Lim, J. M. Yeomans, and B. Ladoix. Topological defects in epithelia govern cell death and extrusion. *Nature*, 544:212–216, 2017.
- [17] G. Duclos, C. Erlenkämper, J-F. Joanny, and P. Silberzan. Topological defects in confined populations of spindle-shaped cells. *Nature Physics*, 13:58–62, 2017.
- [18] O. J. Meacock, A. Doostmohammadi, K. R. Foster, J. M. Yeomans, and W. M. Durham. Bacteria solve the problem of crowding by moving slowly. *Nature Physics*, 17:205–210, 2021.
- [19] K. Copenhagen, R. Alert, N. S. Wingreen, and J. W. Shaevitz. Topological defects promote layer formation in *myxococcus xanthus* colonies. *Nature Physics*, 17:211–215, 2021.
- [20] Y. Maroudas-Sacks, L. Garion, L. Shani-Zerbib, A. Livshits, E. Braun, and K. Keren. Topological defects in the nematic order of actin fibres as organization centres of *hydra* morphogenesis. *Nature Physics*, 17:251–259, 2021.
- [21] S. Čopar and S. Žumer. Nematic braids: Topological invariants and rewiring of disclinations. *Physical Review Letters*, 106:177801, 2011.
- [22] G. P. Alexander, B. G. Chen, E. A. Matsumoto, and R. D. Kamien. Colloquium: Disclination loops, point defects, and all that in nematic liquid crystals. *Reviews of Modern Physics*, 84:497–514, 2012.
- [23] T. Machon and G. P. Alexander. Knotted defects in nematic liquid crystals. *Physical Review Letters*, 113:027801, 2014.
- [24] T. Machon and G. P. Alexander. Global defect topology in nematic liquid crystals. *Proceedings of the Royal Society A*, 472:20160265, 2016.

- [25] U. Tkalec, M. Ravnik, S. Čopar, S. Žumer, and I. Muševič. Reconfigurable knots and links in chiral nematic colloids. *Science*, 333:62–65, 2011.
- [26] S. Čopar, U. Tkalec, I. Muševič, and S. Žumer. Knot theory realizations in nematic colloids. *The Proceedings of the National Academy of Sciences*, 112:1675–1680, 2015.
- [27] I. Muševič. *Liquid Crystal Colloids*. Springer, Cham, 2017.
- [28] S. Čopar, J. Aplinc, Ž. Kos, S. Žumer, and M. Ravnik. Topology of three-dimensional active nematic turbulence confined to droplets. *Physical Review X*, 9:031051, 2019.
- [29] G. Duclos, R. Adkins, D. Banerjee, M. S. E. Peterson, M. Varghese, I. Kolvin, A. Baskaran, R. A. Pelcovits, T. R. Powers, A. Baskaran, F. Toschi, M. F. Hagan, S. J. Streichen, V. Vitelli, D. A. Beller, and Z. Dogic. Topological structure and dynamics of three-dimensional active nematics. *Science*, 367:1120–1124, 2020.
- [30] J. Binysh, Ž. Kos, S. Čopar, M. Ravnik, and G. P. Alexander. Three-dimensional active defect loops. *Physical Review Letters*, 124:088001, 2020.
- [31] J. Friedel and P. G. de Gennes. Boucles de disclinations dans les cristaux liquides. *Comptes rendus de l'Académie des Sciences de Paris B*, 268:257, 1969.
- [32] J. C. Maxwell. *A Treatise on Electricity and Magnetism*. Cambridge University Press, Cambridge, 1873.
- [33] J. Binysh and G. P. Alexander. Maxwell’s theory of solid angle and the construction of knotted fields. *Journal of Physics A: Mathematical and Theoretical*, 51:385202, 2018.
- [34] J. L. Ericksen. Hydrostatic theory of liquid crystals. *Archive for Rational Mechanics and Analysis*, 9:371–378, 1962.
- [35] F. M. Leslie. Some constitutive equations for liquid crystals. *Archive for Rational Mechanics and Analysis*, 28:265–283, 1968.
- [36] D. Forster, T. C. Lubensky, P. C. Martin, J. Swift, and P. S. Pershan. Hydrodynamics of liquid crystals. *Physical Review Letters*, 26:1016–1019, 1971.
- [37] P. C. Martin, O. Parodi, and P. S. Pershan. Unified hydrodynamic theory for crystals, liquid crystals, and normal fluids. *Physical Review A*, 6:2401–2420, 1972.
- [38] L. Onsager. Reciprocal relations in irreversible processes. I. *Physical Review*, 37:405–426, 1931.
- [39] L. Onsager. Reciprocal relations in irreversible processes. II. *Physical Review*, 38:2265–2279, 1931.
- [40] A. N. Beris and B. J. Edwards. *Thermodynamics of Flowing Systems: with Internal Microstructure*. Oxford University Press, Oxford, 1994.
- [41] D. Foster, C. Kind, P. J. Ackerman, J-S. B. Tai, M. R. Dennis, and I. I. Smalyukh. Two-dimensional skyrmion bags in liquid crystals and ferromagnets. *Nature Physics*, 15:655–659, 2019.
- [42] B. G. Chen, P. J. Ackerman, G. P. Alexander, R. D. Kamien, and I. I. Smalyukh. Generating the hopf fibration experimentally in nematic liquid crystals. *Physical Review Letters*, 110:237801, 2013.
- [43] J-S. B. Tai and I. I. Smalyukh. Three-dimensional crystals of adaptive knots. *Science*, 365:1449–1453, 2019.
- [44] L. Radzihovsky and T. C. Lubensky. Nonlinear smectic elasticity of helical state in cholesteric liquid crystals and helimagnets. *Physical Review E*, 83:051701, 2011.
- [45] M. Delaye, R. Ribotta, and G. Durand. Buckling instability of the layers in a smectic-a liquid crystal. *Physics Letters A*, 44:139, 1973.

- [46] N. A. Clark and R. B. Meyer. Strain-induced instability of monodomain smectic a and cholesteric liquid crystals. *Applied Physics Letters*, 22:493, 1973.
- [47] T. C. Adhyapak, S. Ramaswamy, and J. Toner. Live soap: Stability, order, and fluctuations in apolar active smectics. *Physical Review Letters*, 110:118102, 2013.
- [48] L. Chen and J. Toner. Universality for moving stripes: A hydrodynamic theory of polar active smectics. *Physical Review Letters*, 111:088701, 2013.
- [49] C. A. Whitfield, T. C. Adhyapak, A. Tiribocchi, G. P. Alexander, D. Marenduzzo, and S. Ramaswamy. Hydrodynamic instabilities in cholesteric liquid crystals. *European Physical Journal E*, 40:50, 2017.
- [50] S. J. Kole, G. P. Alexander, S. Ramaswamy, and A. Maitra. Layered chiral active matter: Beyond odd elasticity. *Physical Review Letters*, 126:248001, 2021.
- [51] L. Giomi, M. J. Bowick, P. Mishra, R. Sknepnek, and M. C. Marchetti. Defect dynamics in active nematics. *Philosophical Transactions of the Royal Society A*, 372:20130365, 2014.

Chapter 12

Active Nematics

JULIA M YEOMANS

Self-propelled active particles create dipolar flow fields that have *nematic symmetry*. A result of this is that the continuum equations of motion needed to describe dense active particles with hydrodynamic interactions closely resemble those for passive liquid crystals described in Chapter 11, but with an additional active stress. The new term has far reaching consequences. Long-range nematic order is destroyed by the active flow and replaced by active turbulence, a chaotic flow state with strong vorticity (Fig. 12.1). In a two-dimensional, *passive nematic*, defects usually approach each other, driven by elastic interactions, and then annihilate in pairs of topological charge $+1/2$ and $-1/2$. In an *active nematic* [3, 4, 5], the flow of energy means that defects can also be created in pairs. Moreover, a consequence of the active stress is that the defects are self-motile. Hence, *active turbulence* is characterized by a dynamic gas of defects that are continually being created, moving around and annihilating. Active turbulence relies on particles that are elongated in shape or deformable to give local ne-

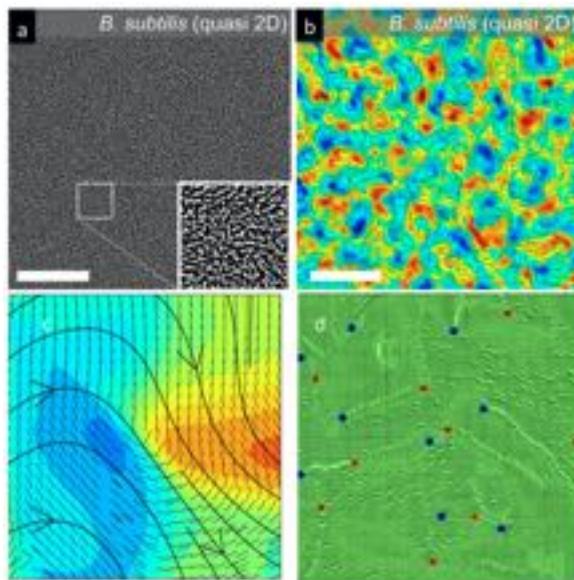


Figure 12.1: **Active turbulence.** (a) Active turbulence in a dense suspension of swimming bacteria. (b) The associated vorticity field. Red (blue) regions correspond to high positive (negative) vorticity. (c) Simulations of active turbulence at a smaller length scale. The bend deformation drives a velocity jet (black flow lines, with arrows). (d) Snapshot from simulations showing $+1/2$ (red) and $-1/2$ (blue) topological defects. Reproduced from Refs. [1, 2]. Agnese: Arrange in a line. Also ensure that the credits are associated to the right panels. Agnese: Reference 2 should be Science Advances 2:1501854, 2016

matic ordering. Examples of active nematics include dense suspensions of microswimmers, bacterial colonies, biological filaments driven by motor proteins, and confluent cell layers and tissues. This Chapter introduces the continuum equations of motion that can be used to describe these materials, and motivates the form for the active stress. It then outlines how active flows lead to instabilities, and thence to active turbulence, and it describes the properties of active turbulence and topological defects. We argue that the short-range nematic ordering required for active turbulence can be initiated by the active flows themselves, and therefore that the turbulence can occur in a system that is isotropic in the passive limit. The interplay of surfaces, interfaces and activity leads to rich physics, and we describe how active nematics behave in confined geometries, and as friction screens the flow. The majority of the experimental and theoretical work so far has been in two dimensions. However active turbulence has now been identified and studied in three dimensions, where the $\pm 1/2$ defects are replaced by active disclination lines and loops. We summarize how active dislocations and active anchoring can drive shape changes in active droplets and shells.

12.1 Active nematohydrodynamics

Chapter 11 introduced the equations governing the hydrodynamics of nematic liquid crystals in terms of the director field \mathbf{n} . Here it will be more convenient to use the tensor order parameter \mathbf{Q} , introduced in Section 11.1. This is because the \mathbf{Q} -tensor carries information about both the direction and the magnitude of the nematic order and can naturally describe topological defects where the magnitude of the ordering falls to zero at the core. After summarising the equation for a passive nematic, we show that a simple additional term in the stress introduces activity and a striking range of dynamical steady states.

12.1.1 Equations of motion

The continuum dynamics of an active nematic can be described by coupled equations for the velocity, \mathbf{u} , and the order parameter, $\mathbf{Q} = \frac{d}{d-1}S(\mathbf{n}\mathbf{n} - \mathbf{I}/d)$, where \mathbf{n} is the director, S is the magnitude of the nematic order, and d is the dimensionality of the space. The quantity \mathbf{Q} evolves according to [6]

$$\partial_t \mathbf{Q} + \mathbf{u} \cdot \nabla \mathbf{Q} - \mathbf{S} = \Gamma \mathbf{h}. \quad (12.1)$$

In addition to advection by the flow, accounted for by the first two terms in this equation, elongated particles respond to flow gradients in a way captured by the co-rotational term,

$$\mathbf{S} = (\lambda \mathbf{E}_s + \boldsymbol{\Omega}) \cdot \left(\mathbf{Q} + \frac{\mathbf{I}}{3} \right) + \left(\mathbf{Q} + \frac{\mathbf{I}}{3} \right) \cdot (\lambda \mathbf{E}_s - \boldsymbol{\Omega}) - 2\lambda \left(\mathbf{Q} + \frac{\mathbf{I}}{3} \right) (\mathbf{Q} : \nabla \mathbf{u}), \quad (12.2)$$

[Is the column in the equation correct? What does it represent?] where $\boldsymbol{\Omega}$ and \mathbf{E}_s are the vorticity and the rate of strain tensors, respectively. The relative dominance of the rate of strain and the vorticity in affecting the alignment of nematogens [Define term] with the flow is characterized by the tumbling parameter λ .

The $\Gamma \mathbf{h}$ term describes relaxational dynamics of the nematic tensor to the minimum of a free energy through a molecular field defined by

$$\mathbf{h} = -\frac{\delta \mathcal{F}}{\delta \mathbf{Q}} + \frac{\mathbf{I}}{3} \text{Tr} \left(\frac{\delta \mathcal{F}}{\delta \mathbf{Q}} \right), \quad (12.3)$$

with Γ the rotational diffusivity. The free energy is typically taken to have the usual Landau-de Gennes form [7] (see Chapters 2 and 11)

$$\mathcal{F}_b = \frac{a}{2} \mathbf{Q}^2 + \frac{b}{3} \mathbf{Q}^3 + \frac{c}{4} \mathbf{Q}^4 + \frac{K}{2} (\nabla \mathbf{Q})^2, \quad (12.4)$$

where the coefficients of the bulk terms a , b , and c are material parameters, and the final term describes the elastic free energy cost of spatial inhomogeneities in the order parameter field, assuming a single elastic constant K .

The velocity field, assuming a constant density ρ , obeys the incompressible Navier-Stokes equations:

$$\nabla \cdot \mathbf{u} = 0, \quad (12.5)$$

$$\rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = \nabla \cdot \boldsymbol{\sigma}. \quad (12.6)$$

Many active particles are sufficiently small that the low Reynolds number limit is appropriate and the left-hand side of Eq. (12.6) can be neglected. The quantity σ is the stress tensor; it includes the pressure, P , the usual viscous stress,

$$\sigma^{\text{viscous}} = 2\eta \mathbf{E}_s, \quad (12.7)$$

where η is the viscosity, and the elastic stress describing the backflow induced by any motion of the nematic particles,

$$\sigma^{\text{elastic}} = 2\lambda(\mathbf{Q} + \mathbf{I}/3)(\mathbf{Q} : \mathbf{h}) - \lambda \mathbf{h} \cdot (\mathbf{Q} + \frac{\mathbf{I}}{3}) - \lambda(\mathbf{Q} + \frac{\mathbf{I}}{3}) \cdot \mathbf{h} - \nabla \mathbf{Q} \frac{\delta \mathcal{F}}{\delta \nabla \mathbf{Q}} + \mathbf{Q} \cdot \mathbf{h} - \mathbf{h} \cdot \mathbf{Q}. \quad (12.8)$$

[Check column symbol in previous equation]

12.1.2 The active stress

Eqs. (12.1)–(12.8) describe a passive nematic. The simplest additional term that can account for the stresses induced by active particles is

$$\sigma^{\text{active}} = -\zeta \mathbf{Q}, \quad (12.9)$$

where ζ is the activity coefficient. An argument to motivate Eq. (12.9) was given by Simha and Ramaswamy [8]. By Newton's third law, the forces on a self-propelled particle must be equal and opposite. Averaging over the details of the swimming stroke, a simple model of a self-propelled particle i that gives the correct far flow field is a rod centered at \mathbf{x}_i along $\hat{\mathbf{n}}_i$ with equal and opposite forces $\pm f\hat{\mathbf{n}}_i$ acting at its ends, at $\mathbf{x}_i + a\hat{\mathbf{n}}_i$ and $\mathbf{x}_i - a'\hat{\mathbf{n}}_i$ [Define a and a']. Hence, the force per unit volume due to the swimmers is

$$\begin{aligned} f_{\alpha}^{\text{active}}(\mathbf{x}) &= \sum_i f \hat{n}_{i,\alpha} \{ \delta[\mathbf{x} - \mathbf{x}_i - a\hat{\mathbf{n}}_i] - \delta([\mathbf{x} - \mathbf{x}_i + a'\hat{\mathbf{n}}_i]) \} \\ &\approx -\frac{(a+a')}{2} \sum_i f \nabla_{\beta} \hat{n}_{i,\beta} \hat{n}_{i,\alpha} \delta[\mathbf{x} - \mathbf{x}_i], \end{aligned} \quad (12.10)$$

where the second line follows from expanding the δ -functions and we have used Greek indices with the usual Einstein summation convention to represent Cartesian directions. Recalling the relation between force and stress,

$$\nabla \cdot \sigma^{\text{active}} \equiv \mathbf{f}^{\text{active}}, \quad (12.11)$$

the leading order contribution to the active stress is

$$\Pi_{\alpha\beta}^{\text{active}} = -(a+a') \sum_i f \hat{n}_{i,\beta} \hat{n}_{i,\alpha} \delta[\mathbf{x} - \mathbf{x}_i]. \quad (12.12)$$

Coarse-graining, we may replace $\hat{\mathbf{n}}_i$ by the director \mathbf{n} and the sum over i by the concentration of active particles $c(\mathbf{x})$, obtaining

$$\Pi_{\alpha\beta}^{\text{active}} = -(a+a') c(\mathbf{x}) n_{\alpha} n_{\beta}. \quad (12.13)$$

From the definition of the \mathbf{Q} -tensor, and noting that constant terms in the stress do not affect the dynamics, Eq. (12.9) follows immediately. The unknown microscopic parameters, which depend on the details of the active forces and the concentration of active particles, determine the activity ζ . The case $\zeta > 0$ corresponds to extensile particles that pump fluid outwards away from their ends, while the case $\zeta < 0$ to contractile particles that draw fluid inwards towards their ends. Note that the form of the active stress remains the same for both polar ($a \neq a'$) and apolar ($a = a'$) active particles.

Example 12.1: Deriving the active stress. Starting from Eq. (12.10), follow through the steps that lead to the expression for the active stress, filling in the necessary details.

12.1.3 Active instabilities

The quantity σ^{active} appears under a derivative in the equations of motion (Eq. (12.11)). Thus, any gradient in the direction or magnitude of the nematic field induces stresses and, hence, flows. Linear stability analysis shows that a far-reaching consequence is that the homogeneous nematic phase is unstable to active stresses [8, 9].

We assume that the director is oriented along the x -direction and calculate how small perturbations to the nematic order evolve with time. The elements of the perturbed nematic tensor are $Q_{xx} = Q_{xx}^0 + \delta Q_{xx}$ and $Q_{xy} = Q_{xy}^0 + \delta Q_{xy}$ where $(Q_{xx}^0, Q_{xy}^0) = (S_0/2, 0)$. Using Eqs. (12.1) and (12.6), representing the Fourier transform of any fluctuating field δf as $\delta f(\mathbf{r}, t) = \int d\mathbf{q} \tilde{f}(\mathbf{q}, t) e^{i\mathbf{q}\cdot\mathbf{r}}$ [Define \mathbf{q}], and setting $\lambda = 0$ for simplicity, the evolution of the perturbations in the low-Reynolds number limit are

$$\partial_t \tilde{Q}_{xx} = -\tilde{Q}_{xx}(Kq^2 + 2S_0^2 A)\Gamma, \quad (12.14)$$

$$\partial_t \tilde{Q}_{xy} = -\tilde{Q}_{xy}\{K\Gamma q^2 - (2\eta)^{-1} S_0 \zeta \cos 2\theta\} - (2\eta)^{-1} \tilde{Q}_{xx} \sin 2\theta S_0 \zeta, \quad (12.15)$$

where θ is the angle between \mathbf{q} and the nematic director. Eq. (12.14) shows that the longitudinal perturbations relax to zero and Eq. (12.15) gives the growth rate of transverse perturbations \tilde{Q}_{xy} as

$$\omega = -\{K\Gamma q^2 - (2\eta)^{-1} S_0 \zeta \cos 2\theta\}. \quad (12.16)$$

For $\omega < 0$, perturbations die out over time, whereas for $\omega > 0$ perturbations grow and long-range nematic order is unstable. Fig. 12.2 indicates pictorially how the instability in an extensile active nematic leads to destabilizing flows.

Example 12.2: Active instabilities.

- a. What type of nematic distortion drives the instability in extensile ($\zeta > 0$) and contractile ($\zeta < 0$) active nematics?
- b. Draw a diagram illustrating the formation of the instability for a contractile active nematic. [Hint: See Fig. 12.2 for the extensile case.]

12.2 Active turbulence

Long-range nematic order is destroyed by active flows, but how do the instabilities grow beyond the linear regime? Numerical solutions of the active nematohydrodynamic equations have helped to answer this question, showing that the system settles into a state, termed active or mesoscale turbulence, which is characterized by strong fluid jets and high vorticity in the flow field, and short-range nematic order and motile topological defects in the director field [10, 11].

Fig. 12.3 illustrates the onset of active turbulence. The distortions created by the hydrodynamic instabilities tend to localize to form *walls*, lines of high distortion separated by nematic regions. Because of their high elastic energy the walls are preferential sites for the formation of $\pm 1/2$ topological defects. Rather than immediately annihilating, local stresses cause the $+1/2$ defect to move away from the $-1/2$ defect. The defects initially tend to move along walls but their dynamics, together with that of the walls, rapidly becomes chaotic. If oppositely charged defects encounter each other, they annihilate [12]. Hence, a dynamical steady state results, with motile defects continuously being created and destroyed. The flow field, driven by stresses due to defects and other director gradients is highly chaotic. Vortices form over a range of length scales with areas that are exponentially distributed [13]. The active length scale, which governs the decay of the vorticity correlations, is $\sqrt{(K/\zeta)}$. Snapshots of active turbulence are shown in Fig. 12.1.

Active turbulence is markedly different from inertial turbulence, as one might expect given the very different Reynolds number. In active turbulence (Reynolds number very small), the energy input is at the scale of the individual particles, and the energy is dissipated on the scale of the vortices. In inertial turbulence (Reynolds number very large), energy is input at large scales and cascades to smaller scales. A detailed statistical analysis of active turbulence indeed shows a clear distinction from inertial turbulence in terms of intermittency, energy spectrum, and flow structure [14].

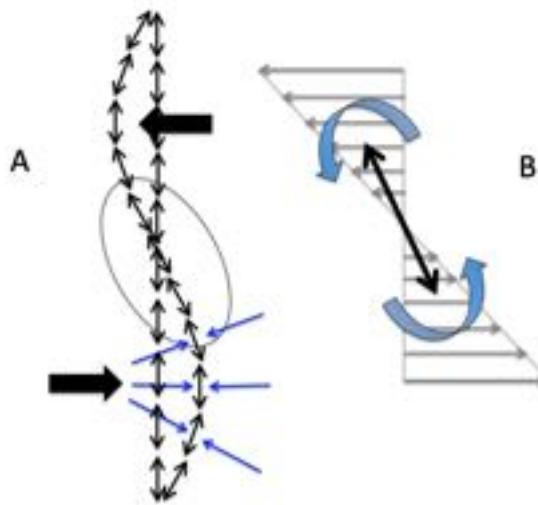


Figure 12.2: **Instability in an extensile active nematic.** (a) Each nematogen in the distorted configuration produces flow shown in blue. The flow is unbalanced giving a resultant vortical flow shown as large black arrows. (b) The shear acts to increase the deformation destabilizing the ordered nematic state (courtesy of S. P. Thampi). [Update panel letters in the figure to be consistent with the caption]

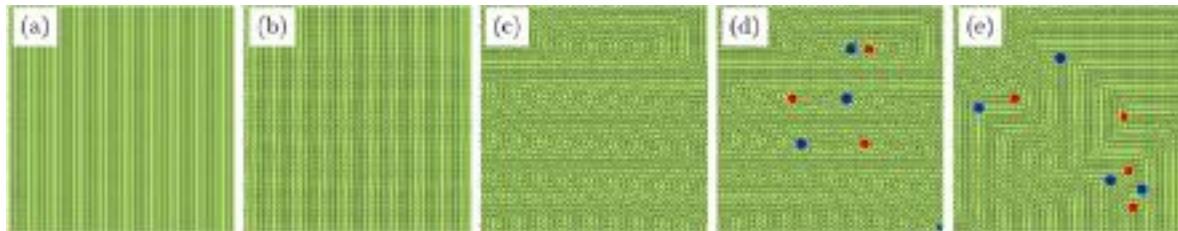


Figure 12.3: **Onset of active turbulence.** (a-c) Snapshots at successive times of the director field (dashed lines) during the development of active turbulence from an ordered nematic state (a) for an extensile system. Walls are formed (b) and sharpen (c). (d) Defect pairs appear at walls ($+1/2$ and $-1/2$ represented by red and blue dots, respectively). (e) Initially, they move along the walls (and restore nematic order), but their motion quickly becomes chaotic. An experiment showing this instability is described in Ref. [16]. Reproduced from Ref. [12]. [Can a higher resolution figure be provided?]

As an alternative to the continuum equations, an approach based on kinetic theory has been used to model the nematohydrodynamics of active, rod-like particles [15]. Each particle is represented as a slender rod with a surface velocity that results in extensile or contractile dipolar flows, and the distribution function for the number density of the particles is described by the Smoluchowski equation. The particle concentration and nematic tensor are then constructed from the first and second moments of the distribution function, and the center of mass position and orientation of the rods are found from slender body theory for zero-Reynolds-number flows. Such a kinetic theory approach is able to reproduce the generation of active turbulence and the dynamics of active defects.

12.2.1 Motile topological defects

Topological defects are regions of high director distortion and hence act as sources of stress and flow. A calculation by Giomi *et al.* [17] has shown that it is possible to obtain the velocity field of defects analytically under certain approximations. The director field of a point defect is $\mathbf{n} = (\cos m\phi, \sin m\phi)$ where ϕ is the polar

angle and m is the topological charge of the defect. The body force due to activity follows immediately as

$$\mathbf{f}^{\text{active}} = \nabla \cdot \boldsymbol{\sigma}^{\text{active}} = \frac{\zeta}{2r} \hat{\mathbf{x}}, \quad m = +1/2, \quad (12.17)$$

$$= \frac{\zeta}{2r} (-\cos 2\phi \hat{\mathbf{x}} + -\sin 2\phi \hat{\mathbf{y}}), \quad m = -1/2. \quad (12.18)$$

Within the Stokes' approximation, the velocity induced by this force field is

$$v_i(\mathbf{r}) = \int dA' G_{ij}(\mathbf{r} - \mathbf{r}') f_j(\mathbf{r}'), \quad (12.19)$$

where G_{ij} is the two-dimensional Oseen tensor

$$G_{ij}(\mathbf{r}) = \frac{1}{2\pi\eta} \left\{ \left(\log \frac{R}{r} - 1 \right) \delta_{ij} + \frac{r_i r_j}{r^2} \right\}. \quad (12.20)$$

[Please, define symbols in the previous equations, particularly, R] Substituting the active force distribution, Eq. (12.18), into Eq. (12.19) and using Eq. (12.20) give (see the appendix of Ref. [17] for details)

$$v_+^a(r, \phi) = \frac{\zeta}{12\eta} \{ (3(R-r) + r \cos 2\phi) \hat{\mathbf{x}} + r \sin 2\phi \hat{\mathbf{y}} \} \quad (12.21)$$

for the velocity field produced by a $+1/2$ defect and

$$v_-^a(r, \phi) = \frac{\zeta r}{12\eta R} \left\{ \left(\left(\frac{3r}{4} - R \right) \cos 2\phi - R \cos \frac{4\phi}{5} \right) \hat{\mathbf{x}} + \left(\left(\frac{3r}{4} - R \right) \sin 2\phi + R \sin \frac{4\phi}{5} \right) \hat{\mathbf{y}} \right\} \quad (12.22)$$

for a $-1/2$ defect.

To obtain these formulas, it is necessary to make assumptions about the extent of the domain and the behavior of the director and velocity fields on its boundaries. The choice used here is that the director field is uniform outside a radius R and that there are no slip conditions on the velocity field at R . In active turbulence, the cut off is most likely to be provided by other defects.

Under the assumption that the convective dynamics of the flow dominates director relaxation, the defect can be considered as a particle moving at the velocity of its core. Putting $r = 0$ in Eqs. (12.21) and (12.22) gives an estimate for the velocity of $+1/2$ defects, $(\zeta R/4\eta) \hat{\mathbf{x}}$, and confirms that $-1/2$ defects are not self-propelled.

Example 12.3: Active topological defects. **a.** Use the expressions (12.21) and (12.22) to plot the flow field around a $+1/2$ and a $-1/2$ active defect. **b.** From the literature, find (at least) five different experimental systems where active topological defects have been identified.

12.2.2 Activity-driven active turbulence

A second, competing instability is also relevant to the behavior of active nematics [18, 19, 20]. If the thermodynamic parameters are chosen so that the passive liquid crystal is in the isotropic phase, short-range nematic order and active turbulence can still be observed. This occurs because local fluctuations in the magnitude of the nematic order set up shear flows, which can, in turn, enhance the ordering. A stability analysis shows that the critical wave number below which the system is unstable is

$$q_c = \sqrt{\frac{\lambda\zeta}{2K(2\eta\Gamma + \lambda^2)}}. \quad (12.23)$$

[Define λ and Γ] Hence, the instability only occurs if $\lambda\zeta > 0$, for extensile, rod-like particles or contractile, disc-shaped particles that align in a shear flow.

A diagram showing the physical mechanism involved in the creation of nematic order is shown in Fig. 12.4. Each active particle creates a dipolar flow field. The dipole lies along the long axis of the particle, and extensile

particles pull fluid in from their sides and push it out from their ends. In a perfectly isotropic arrangement of active particles, the dipolar flows generated by each particle cancel each other on sufficiently large length scales so there is no net flow. However, if a fluctuation generates a local nematic alignment of active particles, then this results in the generation of locally shearing flow fields. In such shear flows particles align further in a manner that enhances the shear. This again results in the strengthening of the nematic order creating a bootstrap effect: stronger alignment leads to stronger flows, which in turn generate stronger alignment — but this process will be balanced by the instability that we described earlier, the destruction of long-range nematic ordering by flow, which will be occurring simultaneously.

12.2.3 Experimental systems

An early identification of active turbulence was in suspensions of *Bacillus subtilis* bacteria swimming near the contact line of a sessile drop [21]. Since then, it has become apparent that this is a generic behavior of many dense suspensions of microswimmers. Another nematic model system is *living liquid crystals*. These are bacteria dispersed in aqueous-based liquid crystals, a set-up which allows the swimming characteristics of the bacteria and the orientational order of the medium to be controlled independently [22]. The bacteria align along the director of the liquid crystal and motility is tuned by the amount of dissolved oxygen. After an oxygen supply is initiated, bacteria start swimming and trigger a stripe-like instability that, at sufficiently high activity, gives rise to the nucleation of half-integer defects proliferating into active turbulence. Flows resembling active turbulence are also seen in assemblies of bacterial cells crawling on a surface, although here friction is expected to screen the hydrodynamic interactions. An example is *Pseudomonas aeruginosa*, which uses pili, hair-like appendages, to pull itself along [23].

A particularly important experimental system that has been key in investigating the properties of active turbulence is a dense mixture of microtubules, biopolymers about $1.5\text{ }\mu\text{m}$ in length, and double-headed versions of the motor protein kinesin, driven by ATP [24] [Add reference to Sagues experimental chapter?]. The addition of the depleting agent PEG concentrates the microtubules into bundles. The microtubules are polar and the kinesin bridges between pairs of microtubules and walks towards the plus end. If the microtubules are parallel, there is no net displacement, but if they are antiparallel the microtubules move relative to each other and this results in polarity-sorted bundles of biopolymers. The bundles are not stable, but buckle and fracture leading to the dynamical steady state of active turbulence. In many experiments, the active microtubule and motor protein mixture assembles at an oil–water interface giving a two-dimensional active layer where properties such as the velocity–velocity correlation function and the trajectories of active defects can be measured. More recently, three-dimensional assays have been developed by enhancing the nematic ordering by adding a passive colloidal liquid crystal based on filamentous viruses [25].

Perhaps more surprisingly, active turbulence and motile topological defects have been observed in confluent

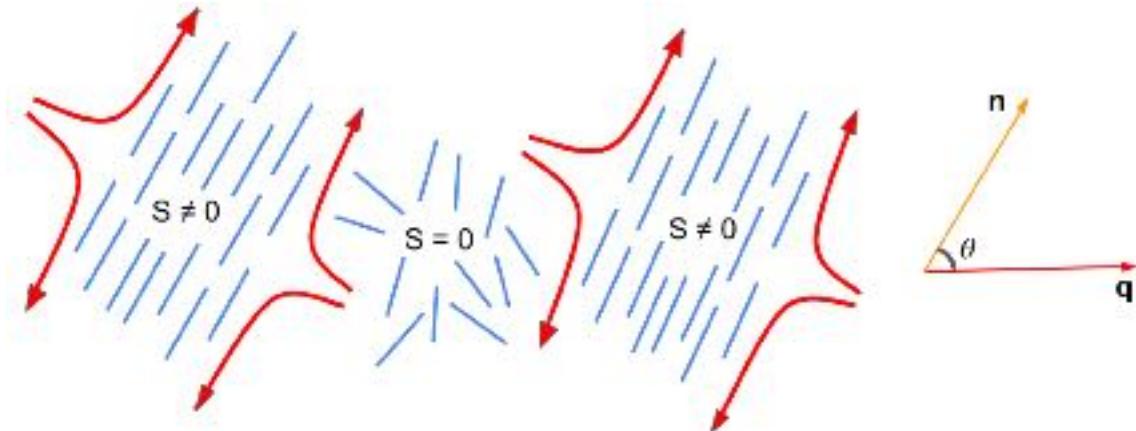


Figure 12.4: **Activity-induced nematic ordering in an isotropic system of active particles.** The blue solid lines indicate nematic directors and red arrows denote the activity-induced flows. The disordered region at the center is aligned by the shear flow set up by the neighboring ordered regions.

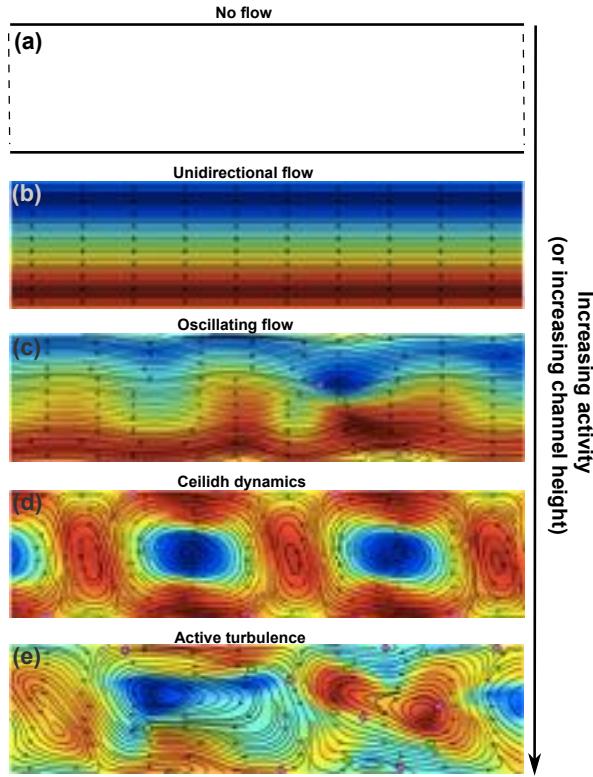


Figure 12.5: **Channel flow.** Flow states of an active nematic confined in a channel. The black lines are the streamlines of the velocity field and the colormap represents the vorticity field. Circles (green) and diamonds (magenta) mark $+1/2$ and $-1/2$ topological defects. Reproduced from Ref. [29]. [Agnese, can you rotate it so that it's horizontal and uses the whole page width? Also remove the letters](#)

cell layers [Add reference to Giavazzi theory chapter?]. Epithelial cells are tightly connected by means of cell-cell junctions and active inter-cellular forces constantly drive deformations of the shapes of the cells. By mapping out the (coarse-grained) direction of the long axis of the deformed cells, Saw et al. [26] identified nematic order and motile topological defects within a two-dimensional confluent layer of epithelial MDCK cells. The flow fields and stresses around the defects were in agreement with active nematic theories, and experiments showed that there is a correlation between the level of activity (controlled in the experiments by adding blebbistatin) and the number of defects in the cell layer. Kawaguchi et al. [27] worked with neural progenitor cells, showing that, at high densities and under confinement, they are capable of aligning over long length scales, forming migratory streams. The cells tended to deplete the neighborhood of $-1/2$ defects and instead to accumulate at $+1/2$ defects, forming mounds. Following these papers, active turbulence and topological defects are now being identified in an increasing number of cell monolayers.

12.3 Confinement

When active nematics are confined, the hydrodynamics is screened and active turbulence can be replaced by more regular flows. These depend sensitively not only on the fluid parameters and the confinement dimensions, but also on the boundary conditions and the strength of intrinsic fluctuations. Simulations of active flow in a one-dimensional channel show that, as the *activity number*, $A = \sqrt{(\zeta h^2)/K}$, where h is the channel height, is increased, the change in flow configurations is from no flow to a (shear or unidirectional) laminar flow to a one-dimensional line of flow vortices to active turbulence (Fig. 12.5) [28]. The system starts to flow when the active stresses can overcome the pinning effect of the boundaries, velocity vortices can form once the channel becomes wide enough to accommodate them, and then a further increase in channel width allows relative motion of the vortices, corresponding to active turbulence.

The motile, active defects add complexity to this sequence. Channel walls are preferential sites for defect formation. The stationary $-1/2$ defects remain close to the walls due to elastic interactions, whereas the self-propelled $+1/2$ defects move toward the center of the channel. The $+1/2$ defects can traverse the channel to be annihilated by the $-1/2$ defects at the opposite wall. In the vortex regime, however, they can alternatively be captured by the flow vortices and perform a *ceilidh dance*, with right- and left-moving defects moving past each other on sinusoidal trajectories in a way reminiscent of the great chain figure in country dancing. Similar interplays between active flow and the dynamics of motile topological defects govern the behavior of active materials confined to circles [30].

The transition from the flow vortex lattice to active turbulence in a channel has interesting properties. Turbulent patches continually appear, combine, and die out, giving a kymograph (space–time diagram) reminiscent of directed percolation [31]. Measuring the behavior of the transition where the turbulent cluster first spans the system indicates that it is indeed in the directed percolation universality class. The same behavior is associated with the transition to inertial turbulence in channel flow.

The confinement flows have been observed in experiments, although different systems are needed to probe different activity numbers. Confluent cell layers correspond to low activities and the transition from a quiescent state to flow has been observed in layers of spindle-shaped cells confined to a stripe on a micro-patterned glass substrate [32]. Experiments on microtubule-motor protein bundles correspond to higher activity numbers. By changing the channel width, Hardouin et al. [33] were able to demonstrate the transitions from laminar (shear) flow, to a flow-vortex state with dancing defects, and then to active turbulence. The experiments identified a state where well-defined shear flow alternates in a regular way with bursts of instability characterized by $+1/2$ topological defects moving across the channel. This occurs because the shear aligns the microtubules parallel to the walls so that they then undergo the usual active instability, bending to create defect pairs.

12.3.1 Friction

Friction with external fluids or boundaries screens hydrodynamic flows and leads to a crossover from a *wet* to a *dry* system. It has a complex effect on the properties of active nematics and is likely to be important in many experiments.

For active nematics with strong thermodynamic ordering, friction introduces memory into the system so that the motion of topological defects leaves trails, arch-like distortions in the director field, that persist for a time that increases with increasing friction. Defects in both extensile and contractile systems move in the same direction with respect to the polar axis of arches. They interact with the trails left by other defects, which leads to a polar order of the $+1/2$ defects. At very high friction, there is insufficient energy to create new topological defects, but, if there are defects already in the system, they create arches in the director field before eventually annihilating. The arches align parallel to each other and readjust to equal widths to form regular arch patterns that coexist with the nematic phase [34]. For active nematics with activity-induced ordering, however, it is much easier to create topological defects. The number of defects proliferate with increasing friction, and elastic and hydrodynamic interactions lead to nematic defect ordering on length scales many times larger than the active length scale [35].

If the friction is anisotropic so that the nematogens move more easily along than perpendicular to their length, yet other flow states and defect configurations are possible. If the active particles are flow-aligning, the chaotic flows can be streamlined into flow lanes with alternating directions, and widths large compared to the active length scale. This reproduces the laning state that has been observed in experiments by interfacing microtubule-motor protein mixtures with smectic liquid crystals. By contrast, for flow-tumbling particles, the synergistic effects of friction anisotropy and flow tumbling can lead to the emergence of bound pairs of active defects that align at an angle to the easy flow direction and move together [36].

12.4 Three dimensions

Three-dimensional active turbulence is again characterized by spatiotemporally chaotic flows. However, point defects are replaced by disclination lines, which constantly undergo transformation events such as breakup, recombination, nucleation, and annihilation. In bulk systems, disclination lines typically form closed, charge-neutral loops [25]. However, they can also terminate at surfaces and the dynamics of the resultant defects on the surface is coupled to the disclination line dynamics in the bulk by elastic interactions and flows [37].

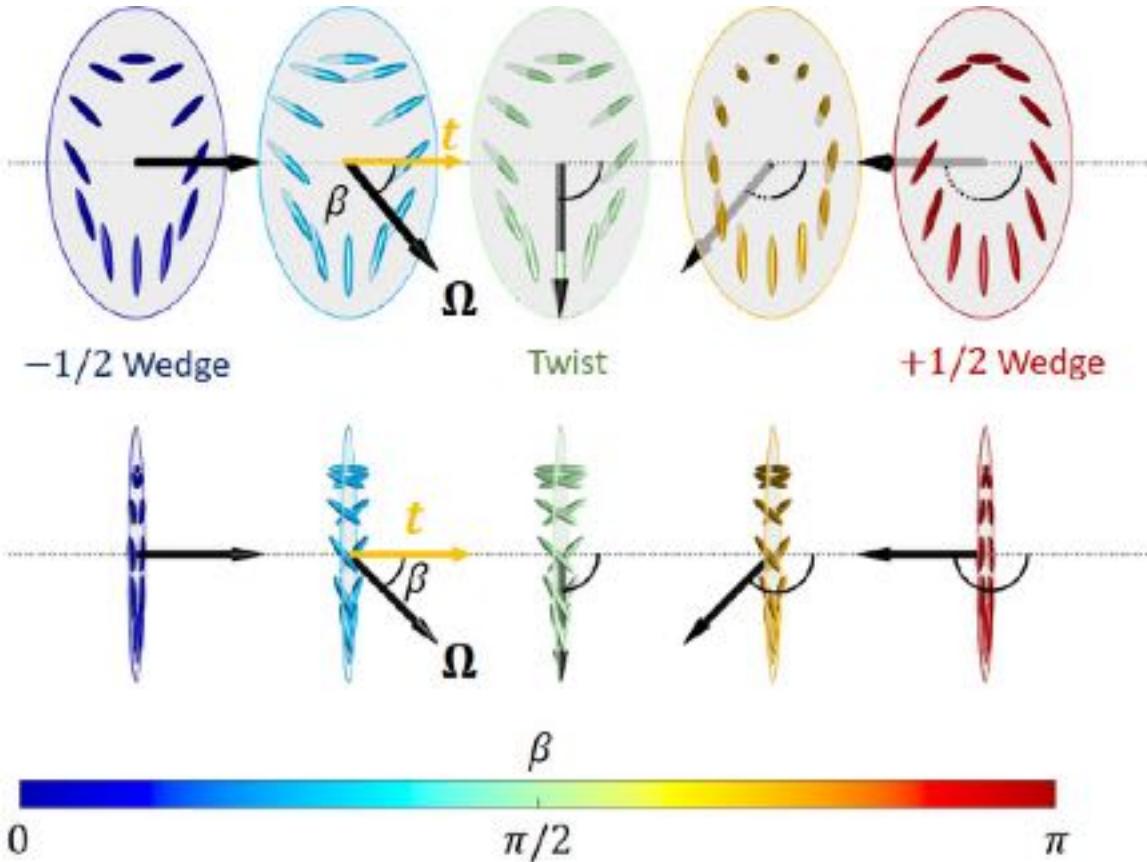


Figure 12.6: **Dislocation lines.** A local $-1/2$ wedge can continuously transform into a $+1/2$ wedge via an intermediate twist disclination. The director field winds about the rotation vector Ω (black arrows) by an angle π . The angle between the rotation vector Ω and the local disclination line tangent t (yellow arrow), called the *twist angle* β , varies continuously along a disclination line. Twist disclinations correspond to $\beta \approx \pi/2$ and $+1/2$ and $-1/2$ wedge disclinations to $\beta \approx \pi$ and 0 , respectively. Reproduced from Ref. [25]. Agnese: redraw figure not to trim it.

Disclination lines can continuously transform from a local $-1/2$ configuration in the plane perpendicular to the line into a $+1/2$ configuration through an intermediate twist winding, as indicated in Fig. 12.6. Moving around the core of a disclination in the plane perpendicular to the local disclination line segment, the director field winds around a specific axis, the rotation vector Ω (black arrows), by an angle π . The angle β between Ω and the local line tangent t (yellow arrow) is called the *twist angle* and can be used to locally characterize the disclination line. For $-1/2$ ($+1/2$) wedge-type defects, the twist angle corresponds to $\beta \approx 0(\pi)$, while line segments with local twist-type defects are indicated by $\beta \approx \pi/2$.

As a consequence of the activity, disclination lines act as self-propelled entities moving through the fluid. Based on a simplified model neglecting elastic interactions, each disclination line segment can be associated with a local self-propulsion velocity that has a component perpendicular to the local tangent of the line, which depends on the twist angle β as [38]

$$v_{\perp}^{\text{SP}} \propto (1 - \cos \beta)^2. \quad (12.24)$$

Thus, line segments with $\beta = 0$ are passive, while $\beta = \pi$ line segments are most active in pushing around the surrounding fluid and distorting the dislocation loops.

A combination of activity, defect, and disclination dynamics, and active anchoring can lead to an enormous range of behaviors in deformable active drops and shells. Fig. 12.7a shows simulations of a deformable active shell (with the director field strongly anchored to lie within the shell). The extending protrusion is caused by flows normal to the surface initiated near a $+1/2$ topological defect [39]. Analogous behavior has been

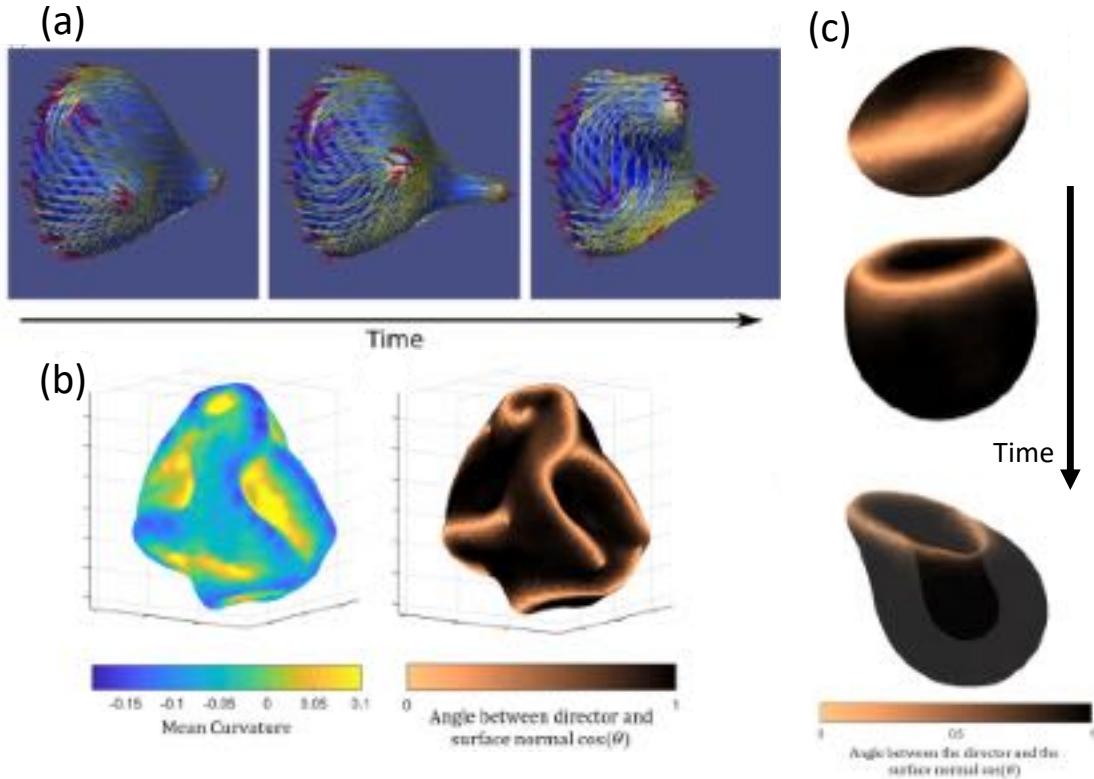


Figure 12.7: Deformable active droplets. (a) Active shell forming a protrusion at a $+1/2$ topological defect. Reproduced from Ref. [39]. (b) Wrinkle formation in a contractile active droplet. (c) Invagination of a smaller contractile active droplet. Reproduced from Ref. [41]. These results were obtained by solving the active nematohydrodynamic equations using a hybrid lattice Boltzmann algorithm. [Agnese: redraw figure not to trim it.](#)

implicated in driving tentacle formation in the marine polyp *Hydra* [40].

Fig. 12.7b shows simulations of a contractile deformable droplet. The wrinkles that form are related to points where dislocation lines reach the surface. Fig. 12.7c shows the time evolution of a contractile droplet, but now a much smaller one. In contractile systems, active flows favor normal surface alignment and, hence, there is at least one $+1$ disclination loop in the bulk due to topological constraints. These loops are associated with a large elastic energy cost. Therefore, a ring with in-plane surface alignment is formed encircling the droplet to maximize the area of perpendicular surface alignment favored by active anchoring while avoiding the formation of the $+1$ defect-loop in the bulk. The contractile activity of this director configuration produces flows that cause the droplet to invaginate, forming a cup shape [41]. A similar transition occurs in gastrulation, the stage of morphogenesis where a single cell sheet reorganizes into a multilayer structure, which then differentiates into the cell types that will initiate the formation of different organs.

12.5 Final remarks

Active nematics comprise a large number of particles that behave collectively but remain out of thermodynamic equilibrium. There is a lot of research still to be done to extend the theories of non-equilibrium statistical mechanics that are currently being developed for simpler active systems to active nematics. For example, although we understand a lot about active turbulence from simulations, predictive theories are still at an early stage. Increasingly the ideas underlying active nematic dynamics are being applied to biological systems. Active turbulence and motile defects have now been observed in many confluent cell layers. It is interesting to ask which additional features are important in biological active flows, and to start to apply the ideas arising from

active nematic physics to three-dimensional mechanobiology. Furthermore, active nematics have potential as devices, translating chemical energy to mechanical work. However, to achieve this potential it is necessary to design synthetic active systems that can be produced in bulk and are cheap and easy to handle.

Bibliography

- [1] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Löwen, and J. M. Yeomans. Meso-scale turbulence in living fluids. *PNAS*, 109:14308, 2012.
- [2] S. P. Thampi, A. Doostmohammadi, T. N. Shendruk, R. Golestanian, and J. M. Yeomans. Active micro-machines: Microfluidics powered by mesoscale turbulence. *Science Advances*, page e1501854, 2016.
- [3] J. Prost, F. Jülicher, and J. F. Joanny. Active gel physics. *Nature Physics*, 11:111, 2015.
- [4] D. Needleman and Z. Dogic. Active matter at the interface between materials science and cell biology. *Nature Reviews Materials*, 2:17048, 2017.
- [5] A. Doostmohammadi, J. Ignes-Mullol, J. M. Yeomans, and F. Sagues. Active nematics. *Nature Communications*, 9:3246, 2020.
- [6] A. N. Beris and B. J. Edwards. *Thermodynamics of Flowing Systems: With Internal Microstructure*. Oxford University Press, 1994.
- [7] P. G. de Gennes and J. Prost. *The Physics of Liquid Crystals*. Oxford University Press, 1995.
- [8] R. A. Simha and S. Ramaswamy. Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles. *Physical Review Letters*, 89:058101, 2002.
- [9] R. Voituriez, J. F. Joanny, and J. Prost. Spontaneous flow transition in active polar gels. *Europhysics Letters*, 70:404, 2005.
- [10] L. Giomi, M. J. Bowick, X. Ma, and M. C. Marchetti. Defect annihilation and proliferation in active nematics. *Physical Review Letters*, 110:228101, 2013.
- [11] S. P. Thampi, R. Golestanian, and J. M. Yeomans. Velocity correlations in an active nematic. *Physical Review Letters*, 111:118101, 2013.
- [12] S. P. Thampi, R. Golestanian, and J. M. Yeomans. Instabilities and topological defects in active nematics. *Europhysics Letters*, 105:18001, 2014.
- [13] L. Giomi. Geometry and topology of turbulence in active nematics. *Physical Review X*, 5:031003, 2015.
- [14] J. Urzay, A. Doostmohammadi, and J. M. Yeomans. Multi-scale statistics of turbulence motorized by active matter. *J. Fluid Mech.*, 822:762, 2017.
- [15] T. Gao, M. D. Betterton, A. S. Jhang, and M. J. Shelley. Analytical structure, dynamics, and coarse-graining of a kinetic model of an active fluid. *Physical Review Fluids*, 2:093302, 2017.
- [16] B. Martínez-Prat, J. Ignes-Mullol, J. Casademunt, and F. Sagues. Selection mechanism at the onset of active turbulence. *Nature Physics*, 15:362, 2019.
- [17] L. Giomi, M.J. Bowick, P. Mishra, R. Sknepnek, and M.C. Marchetti. Defect dynamics in active nematics. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 372:0365, 2014.
- [18] S. P. Thampi, A. Doostmohammadi, R. Golestanian, and J. M. Yeomans. Intrinsic free energy in active nematics. *Europhysics Letters*, 112:28004, 2015.
- [19] T. Gao and Z. Li. Self-driven droplet powered by active nematics. *Physical Review Letters*, 119:108002, 2017.

- [20] S. Santhosh, M. R Nejad, A. Doostmohammadi, J. M. Yeomans, and S. P. Thampi. Activity induced nematic order in isotropic liquid crystals. *Journal of Statistical Physics*, 180:699, 2020.
- [21] C. Dombrowski, L. Cisneros, S. Chatkaew, R. E. Goldstein, and J. O. Kessler. Self-concentration and large-scale coherence in bacterial dynamics. *Physical Review Letters*, 93:098103, 2004.
- [22] S. Zhou, A. Sokolov, O. D. Lavrentovich, and I. S. Aranson. Living liquid crystals. *PNAS*, 111:1265, 2014.
- [23] O. J. Meacock, A. Doostmohammadi, K. R. Foster, J. M. Yeomans, and W. M. Durham. Bacteria solve the problem of crowding by moving slowly. *Nature Physics*, 17:205, 2021.
- [24] T. Sanchez, D. T. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic. Spontaneous motion in hierarchically assembled active matter. *Nature*, 491:431, 2012.
- [25] G. Duclos, R. Adkins, D. Banerjee, M.S.E. Peterson, M. Varghese, I. Kolvin, A. Baskaran, R.A. Pelcovits, T.R. Powers, A. Baskaran, F. Toschi, M.F. Hagan, S.J. Streichen, V. Vitelli, D.A. Beller, and Z. Dogic. Topological structure and dynamics of three-dimensional active nematics. *Science*, 367:1120, 2020.
- [26] T. B. Saw, A. Doostmohammadi, V. Nier, L. Kocgozlu, S. Thampi, Y. Toyama, P. Marcq, C. T Lim, J. M. Yeomans, and B. Ladoux. Topological defects in epithelia govern cell death and extrusion. *Nature*, 544:212, 2017.
- [27] K. Kawaguchi, R. Kageyama, and M. Sano. Topological defects control collective dynamics in neural progenitor cell cultures. *Nature*, 545:327, 2017.
- [28] T. N. Shendruk, A. Doostmohammadi, K. Thijssen, and J. M. Yeomans. Dancing disclinations in confined active nematics. *Soft Matter*, 13:3853, 2017.
- [29] A. Doostmohammadi and J. M. Yeomans. Coherent motion of dense active matter. *European Physics Journal Special Topics*, 227:2401, 2019.
- [30] A. Opathalage, M. M. Norton, M. P. N. Juniper, B. Langeslay, S. A. Aghvami, S. Fraden, and Z. Dogic. Self-organized dynamics and the transition to turbulence of confined active nematics. *Proceedings of the National Academy of Sciences*, 116:4788, 2019.
- [31] A. Doostmohammadi, T. N. Shendruk, K. Thijssen, and J. M. Yeomans. Onset of meso-scale turbulence in active nematics. *Nature Communications*, 8:15326, 2017.
- [32] G. Duclos, C. Blanch-Mercader, V. Yashunsky, G. Salbreux, J.-F. Joanny, J. Prost, and P. Silberzan. Spontaneous shear flow in confined cellular nematics. *Nature Physics*, 14:728, 2018.
- [33] J. Hardouin, R. Hughes, A. Doostmohammadi, J. Laurent, T. Lopez-Leon, J. M. Yeomans, J. Ignes-Mullol, and F. Sagues. Reconfigurable flows and defect landscape of confined active nematics. *Communications Physics*, 2:1, 2019.
- [34] M. R. Nejad, A. Doostmohammadi, and J. M. Yeomans. Memory effects, arches and polar defect ordering at the cross-over from wet to dry active nematics. *Soft Matter*, 17:2500, 2021.
- [35] K. Thijssen, M. R. Nejad, and J. M. Yeomans. Role of friction in multidefect ordering. *Physical Review Letters*, 125:218004, 2020.
- [36] K. Thijssen, L. Metselaar, J. M. Yeomans, and A. Doostmohammadi. Active nematics with anisotropic friction: the decisive role of the flow aligning parameter. *Soft Matter*, 16:2065, 2020.
- [37] S. Čopar, J. Aplinc, Ž. Kos, S. Žumer, and M. Ravnik. Topology of three-dimensional active nematic turbulence confined to droplets. *Physical Review X*, 9:031051, 2019.
- [38] J. Binysh, Z. Kos, S. Čopar, M. Ravnik, and G. P. Alexander. Three-dimensional active defect loops. *Physical Review Letters*, 124:088001, 2020.

- [39] L. Metselaar, J. M. Yeomans, and A. Doostmohammadi. Topology and morphology of self-deforming active shells. *Physical Review Letters*, 123:208001, 2019.
- [40] Y. Maroudas-Sacks, L. Garion, L. Shani-Zerbib, A. Livshits, E. Braun, and K. Keren. Topological defects in the nematic order of actin fibers as organization centers of hydra morphogenesis. *Nature Physics*, 17:251, 2021.
- [41] L. J. Ruske and J. M. Yeomans. Morphology of active deformable 3d droplets. *Physical Review X*, 11:021001, 2021.

Chapter 13

Activity in the Cytoskeleton

FRANCOIS NEDELEC

The cytoskeleton, a crucial cellular infrastructure, is the focus of this Chapter, which emphasizes the active nature of intracellular filaments, focusing on actin and microtubules, and their pivotal role in controlling the dynamics and mechanics of cells. These filaments, as highlighted in Fig. 13.1, exhibit distinctive polar structures and bending elasticities, fundamental to their function in cellular processes like mitosis, endocytosis or cytokinesis. Special attention should be given to the polarity of these filaments, their assembly dynamics at barbed/plus and pointed/minus ends, and the implications of these properties for cellular function. Particularly, the polarity of cytoskeletal filaments allows them to support the movement of molecular motors, whereby chemical energy from ATP is transduced into mechanical work, driving various cellular activities. Advancements in microscopy and molecular biophysics have deepened our understanding of cytoskeletal dynamics. These techniques, ranging from *in vitro* reconstitution to *in vivo* imaging, have unraveled an unexpected complexity in the cytoskeletal machinery. This Chapter summarizes recent experimental insights which have been key

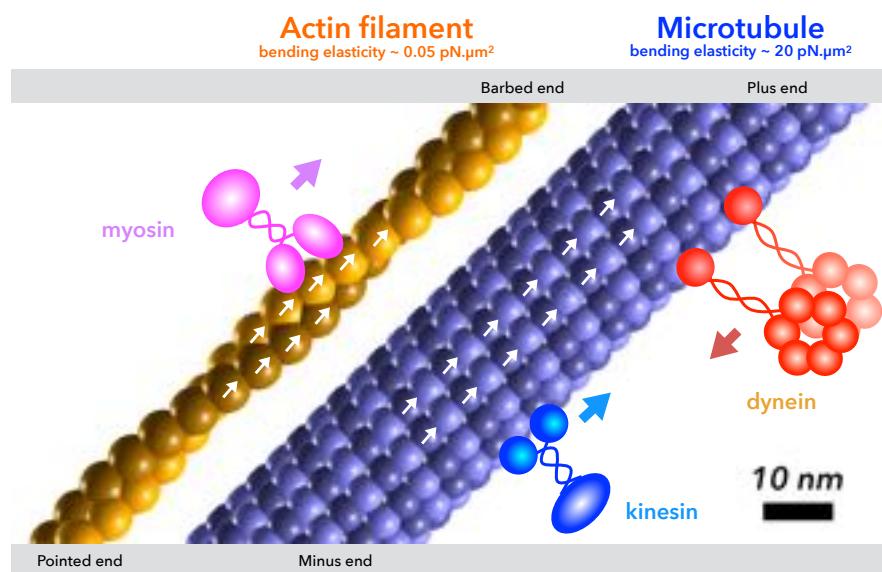


Figure 13.1: **Actin filaments and microtubules.** The cytoskeleton (i.e., the “skeleton of the cell”) is composed of different classes of filaments, of which only two are illustrated here: actin (left) and microtubule (right). Molecular motors (e.g., myosin, kinesin, dynein) are able to move along the filaments by stepping on the monomers. This requires a source of energy, which is provided by ATP. The bending elasticity of actin filaments and microtubules was measured in Ref. [1].

foundations of theoretical models used to study the cytoskeleton’s active role in cellular mechanics and movement. We provide examples of some common algorithms used in stochastic simulations of these processes. Additionally, we discuss how the cytoskeleton’s active matter characteristics contribute to cellular behavior, drawing parallels and contrasts with other systems of active matter studied in previous chapters. The goal is to give a brief sample of our current understanding of cytoskeletal dynamics, highlighting the challenges and specific methods associated with this fascinating field, which find application in cellular and developmental biology.

13.1 Introduction

Cells are the fundamental units of life, relying on a filamentous system known as the *cytoskeleton* to perform essential functions. Critical processes such as chromosome segregation (mitosis), cellular uptake (endocytosis), division (cytokinesis), polarization, and movement are all orchestrated by intricate cytoskeletal machinery composed of multiple filaments. These dynamic structures, powered by molecular motors and filament dynamics, consume energy to drive their functions, and are thus example of ‘Active Matter’. Nature provides a diverse array of cytoskeletal machines, ranging from simple assemblies with a few filaments to highly complex systems comprising millions. While these machines often involve numerous molecular components, significant progress has been made in understanding their mechanisms. Firstly, in some cases, we have identified nearly all the genes involved in these processes, assigning them names and categorizing them into functional families. The proteins they encode have been purified, and their chemical activities have been measured. Secondly, advancements in experimental techniques now enable precise, quantitative investigations both in vitro and in vivo. Researchers can purify individual components, and in many cases, combining these purified elements under controlled conditions is sufficient to reconstitute their collective behavior. Remarkably, even relatively crude mixtures can reproduce cellular phenomena, which is surprising given that the cellular biochemical context is often missing. Finally, digital microscopy allows for unprecedented imaging and measurement of live-cell behavior, while modern genetic tools enable precise manipulation of protein activity, abundance, and timing. These technological advancements have ushered in an exciting era for researchers, driving rapid progress in our understanding of cytoskeletal dynamics.

Since the fundamental laws governing matter at this scale — namely classical and statistical mechanics as well as biochemistry — are well understood, one might expect these processes to be amenable to theoretical investigation. Indeed, it is possible to build models, but the sheer complexity of biological systems remains a formidable challenge. Analytical approaches can sometimes distill the essence of a problem, but they rarely capture the full intricacy of real cells. Computer simulations provide a valuable complement. While simulating “an entire cell” remains far beyond our current capabilities, modern computing power allows for the simulation of sub-cellular processes such as endocytosis, cytokinesis, and the mitotic spindle in small cells, along with many other essential cytoskeletal systems. Although simulation methods continue to evolve, foundational principles have been established, integrating numerical physics with cytoskeletal phenomena that are absent in classical physics — such as molecular motors, filament self-assembly, and stochastic chemistry. Given the extraordinary diversity of life, much work remains to be done, but significant progress is already underway.

The role of theory and simulation is to complement experimental investigations, accelerating our understanding of these processes. By predicting emergent properties based on assumptions about molecular components, theoretical models and simulations can demonstrate whether these assumptions are sufficient to explain the observed collective behavior. Moreover, by requiring researchers to explicitly define these assumptions, theory and simulation help identify the critical parameters governing a system. Ultimately, the goal is to understand the function of cytoskeletal structures *in vivo* — for example, explaining how their architecture and properties are tailored to their specific biological tasks. In this chapter, we introduce the key elements needed to build quantitative models of two fundamental yet illustrative systems: (1) a molecular motor moving along a cytoskeletal filament and (2) a cytoskeletal filament propelled by a collection of molecular motors. These simple systems are perhaps not themselves of much current research interest, but they are sufficient to illustrate how cytoskeletal systems and intracellular mechanics differ from other Active Matter.

The cytoskeleton consists of various classes of filaments, but the most important are actin and microtubules (Fig. 13.1). While other filament types are essential to life, actin and microtubules stand out because they are polar — structurally oriented like arrows in space — an essential property for supporting the movement of molecular motors. Although actin and microtubules differ in many ways and perform distinct cellular functions,

they can also work together to accomplish complex tasks. *Actin filaments* are composed of actin monomers arranged into two protofilaments coiled around each other in a right-handed helix. *Microtubules*, in contrast, typically consist of 13 protofilaments connected laterally, forming a hollow tube with a central lumen. These protofilaments are slightly offset, creating three left-handed helices. Both actin filaments and microtubules are flexible, and they can bend under force without breaking. Their bending elasticities have been measured, revealing that microtubules are significantly stiffer than actin filaments. When observed *in vitro*, microtubules generally appear straight because their persistence length (≈ 7 mm) far exceeds their actual length (typically tens of micrometers). However, in many cells, microtubules appear bent due to forces stronger than Brownian motion, such as those exerted by molecular motors anchored to the cell cortex. Both actin filaments and microtubules are structurally polar because their constituent proteins assemble in a head-to-tail manner, resulting in a uniform orientation throughout the filament (Fig. 13.1, white arrows). While polarity exists along the entire filament, it is conventionally indicated by assigning distinct names to the filament ends: pointed and barbed for actin, and minus and plus for microtubules. This polarity makes the two ends functionally different. In general, filaments assemble more rapidly at their *barbed* or *plus* ends than at the opposite ends. This assembly is an active process with many intriguing aspects, which we will not cover here (see Ref. [2]).

13.2 Kinesins and other molecular motors

We now turn to molecular motors — remarkable proteins capable of "walking" along their associated filaments. The kinesin family consists of motors that travel along microtubules, while myosins move along actin filaments. Another important group, dyneins, are microtubule motors that move toward the minus end. Very similar motor proteins may be known by multiple names in the literature, particularly if they have been independently discovered in different organisms, with structural similarities only recognized later. Motors from various species have since been classified into distinct families based on their structural features [3, 4]. While many kinesins move toward the plus end of microtubules, some are specialized for minus-end-directed movement.

Conventional kinesin was the first motor discovered in the kinesin family [5] and has been extensively studied. In neurons, it transports vesicles from the cell body outward along extensions such as axons. Conventional kinesin is a plus-end-directed motor, meaning it moves toward the plus end of microtubules. This motor is a multi-protein complex composed of two heavy chains and two light chains. The light chains serve as adaptors for cargo binding [6], while each heavy chain contains a tubulin-binding domain. Remarkably, a single kinesin molecule is sufficient for motility [7, 8]. In the 1990s, conventional kinesin became a focus of remarkable single-molecule studies using the newly developed optical trap technique [9]. These studies revealed that kinesin moves in discrete steps of 8 nm [9, 10], a distance matching the size of the tubulin heterodimer, as determined by structural studies of microtubules. Each step requires the hydrolysis of one ATP molecule [11]. The speed of kinesin movement depends on ATP concentration, reaching approximately $1\text{ }\mu\text{m s}^{-1}$ at saturating ATP levels ($\approx 1\text{ mM}$). An opposing force gradually reduces the motor's speed, and kinesin is unable to advance against forces greater than $\approx 5\text{ pN}$ [12]. Under lower loads, the motor can travel about $1\text{ }\mu\text{m}$ before detaching, completing more than a hundred steps [13]. Motors capable of taking multiple consecutive steps without detaching are classified as processive. Detachment occurs stochastically but can also be triggered by microtubule defects or obstacles bound to the microtubule. Conventional kinesin predominantly moves parallel to the protofilaments of microtubules, following the simplest possible trajectory [14]. However, some motors instead follow a helical path around the microtubule [15, 16]. Additionally, the stiffness generated by a single kinesin between a microparticle and a microtubule has been estimated at approximately $400\text{ pN }\mu\text{m}^{-1}$ [17]. For further details, Ref. [18] provides an excellent overview.

The kinesin family is defined by homology within the motor domain — the portion of the heavy chain responsible for binding tubulin and ATP. However, other regions of the proteins vary significantly across the family, giving rise to diverse functions. While most kinesins are plus-end-directed motors, certain members, such as *NCD*, move toward the minus end of microtubules. This ability enables them to perform specialized tasks, such as retrograde transport — moving cargo toward the cell center. Some kinesins possess more than two tubulin-binding sites. For example, Kinesin-5 has a symmetric structure composed of four heavy chains, allowing it to bind and move along two microtubules simultaneously. This bifunctional capability makes Kinesin-5 an active crossbridge, playing a crucial role in organizing multi-microtubule structures, such as the mitotic spindle. Despite significant progress, motors with novel properties are still awaiting discovery and characterization, offering exciting prospects for future research.

13.2.1 Kinesin movement mechanism

The two motor domains of a kinesin molecule bind to adjacent tubulin heterodimers on a microtubule, functioning like feet (Fig. 13.2). Kinesin "walks" along the microtubule by alternating cycles of binding and unbinding. Similar to human walking, it moves by lifting its back foot, swinging it forward, and reattaching it ahead of the stationary foot. The key factor determining the motor's direction is that the foot nearest to the minus end moves forward, rather than the one already positioned closer to the plus end. Since tubulin heterodimers are 8 nm long, each step advances the trailing foot by 16 nm, resulting in an overall movement of 8 nm per step for the motor and its cargo, since it is positioned in the middle between the moving and the static feet. This motion occurs out of thermodynamic equilibrium because each step is powered by the hydrolysis of one ATP molecule to ADP. As a result, the system can break detailed balance, meaning the rate of forward steps typically exceeds that of backward steps.

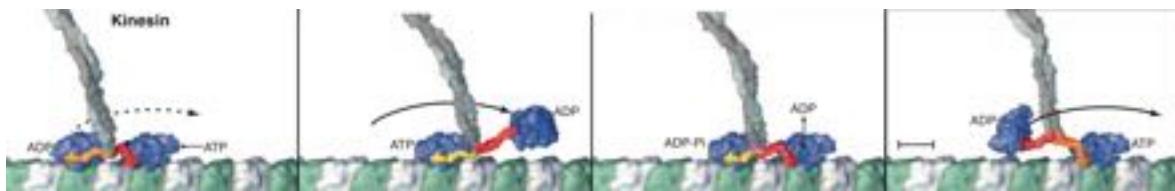


Figure 13.2: **Kinesins.** Molecular motors are remarkable molecules capable of binding to a filament using two distinct domains — often referred to as feet — and "walking" along it. These motors convert chemical energy from ATP into mechanical displacement, typically moving in discrete steps that correspond to the size of the filament's monomers or a multiple of that distance. For example, conventional kinesin takes 8 nm steps, matching the length of a single tubulin heterodimer. Structurally, these motors always bind to the filament's monomers in the same orientation and move forward by shifting the lagging foot ahead of the leading one. By continuously repeating this movement, they can travel long distances. Some molecular motors can take thousands of steps per second, reaching speeds of up to $10 \mu\text{m s}^{-1}$. Reproduced from [19].

13.2.2 Kinesin's forward and backward steps

The discrete nature of kinesin's movement can be observed using a specially-designed apparatus where all source of noise are minimized [20]. In this setup, the position of a bead — attached to a single kinesin — is measured optically with nanometer precision. The kinesin protein is anchored to this bead via its tail, away from the two motor "feet". Controlled forces can thus be applied to this kinesin by adjusting the focus point of the optical trap. Fig.13.3 shows various trajectories of a single motor moving along a microtubule under different conditions. Fig.13.4 illustrates how the stepping rate depends on the applied load. The data reveals that, when subjected to a force of approximately $\approx 5 \text{ pN}$, the motor does not stop moving entirely but instead ceases to make forward progress, oscillating back and forth around its position. This behavior indicates that the motor can take backward steps toward the minus end without detaching. At zero load, backward steps are rare, but their frequency increases as the force rises. Eventually, the rate of backward steps matches the rate of forward steps. At this point, the motor is unable to move forward, though it continues to oscillate occasionally. It is essential to consider the dependence of the stepping rates on force, to understand the role of molecular motors in most of their roles. In Example 13.5, you will use this measured data to build a simulation of this experiment. It is expected that motors such as dynein or myosins will respond differently to force, as they are adapted to perform their tasks under distinct conditions.

13.3 Bell's law and Kramers' theory of transitions under load

In this section, we review some classical results regarding how the detachment of a molecular link depends on the force exerted on it. While there are exceptions to this general trend, in most cases, applying force to a link promotes detachment, and this relationship is typically observed to be exponential:

$$k_{\text{off}} = k_0 \exp\left(\frac{fa}{k_B T}\right), \quad (13.1)$$

where k_0 is the load-free rate, f is the magnitude of the force, a is a molecular scale, k_B is the Boltzman constant, and T the absolute temperature. We will use $k_B T = 4.2 \text{ pN nm}$ corresponding to room temperature. The two parameters k_0 and a are characteristics associated with the attached state of each specific bond. The distance a is assumed to be on the order of the bond's dimension, typically in the nanometer range, and is often determined by fitting experimental data. This particular relationship between load and detachment rate is known as *Bell's law* [21], and it is widely observed in the non-covalent associations of biological molecules. However, not all types of bonds follow Bell's law. One notable exception is *catch bonds*, a unique type of bond that becomes stronger under tension. While a strong tensile force will eventually promote the dissociation of any molecular association, the rate of dissociation for a catch bond does not monotonically increase with force. Instead, the unbinding rate decreases with tension within a physiologically relevant range of force. For more information on dissociation under load, see Ref. [22]. Returning to Bell's law, it can be expressed equivalently as

$$k_{\text{off}} = k_0 \exp\left(\frac{f}{f_0}\right), \quad (13.2)$$

by defining the characteristic force $f_0 = k_B T/a$. This relationship can be derived as a particular limit of *Kramers' theory* [23]. Kramers considered a point-like particle moving in a one-dimensional potential landscape to calculate how the escape rate from a potential trap changes when an external force is applied to the particle.

For brevity, we will outline the essence of the argument, using Kramers' original illustration in Fig. 13.5. In Kramers' theory, the particle is initially trapped in the local minimum in A. To escape towards B, it must go over an energy barrier with a maximum in C. The height of this energy barrier is denoted as Q . The escape rate mainly depends on the probability of the particle climbing up to position C, as the descent from C to B occurs readily. Near equilibrium, the particle's positions follow Boltzmann's distribution: $\exp\left(-\frac{U}{k_B T}\right)$, where the probability depends solely on the potential value U at each position, omitting the normalization factor. When an external force is applied, the potential is altered everywhere. Specifically, at position x the potential is lowered by $-xf$, for a force f directed from A to B. The position x is measured from the equilibrium position A, but another reference could be used, as this would simply shift the potential uniformly without affecting the

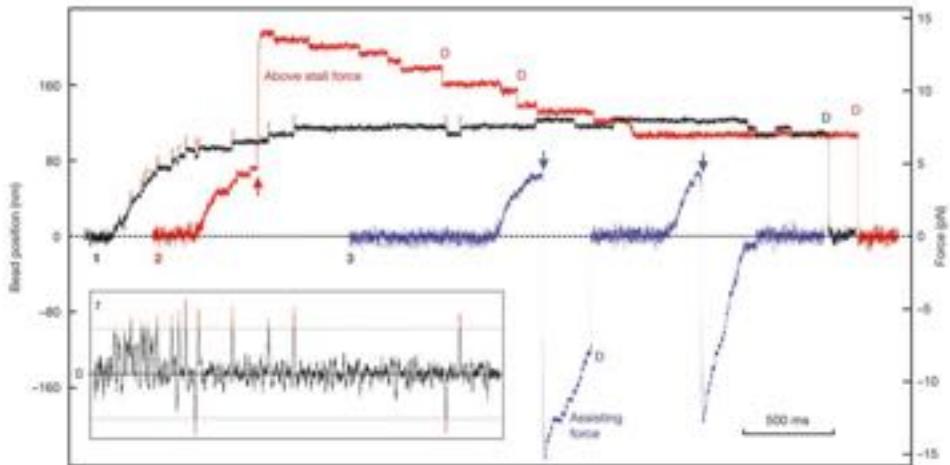


Figure 13.3: Trajectories of single kinesins trapped in an optical trap. 1) Black curve: The trap is stationary. This setup is similar to a kinesin pulling on a spring attached to a fixed position. The kinesin walks until it reaches its stall force. 2) Red curve: The optical trap is moved backward (indicated by the red arrow), creating excessive antagonistic forces on the kinesin. As a result, the kinesin moves slowly in the backward direction. 3) Blue curves: The optical trap is moved forward (indicated by the blue arrows), generating assisting forces. This allows the kinesin to move quickly in the forward direction. The two Y-axes indicate the position (left) and the corresponding force (right) exerted by the optical trap. Reproduced from [20].

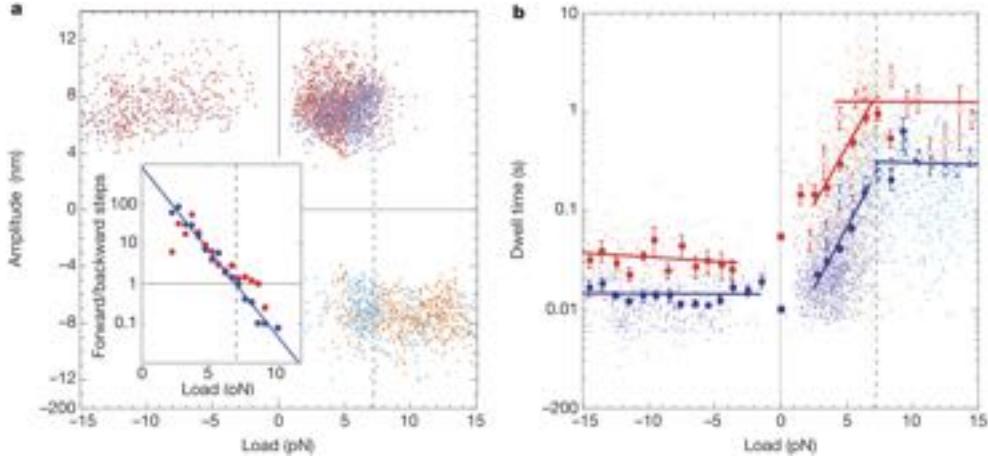


Figure 13.4: Characteristics of kinesin's forward and backward steps. The two panels were obtained by distributing all observed stepping events according to the load on the kinesin. Two conditions are represented in the graph: a high concentration of ATP (blue), corresponding to fast kinesin movements 1 mM ATP, and a low concentration (red), 10 μM ATP, where the stepping rates of kinesin are strongly reduced. On the left panel (a), the amplitude of the step is shown, and given the noise in the system, it can be seen that it is either -8 nm or $+8 \text{ nm}$ for respectively a forward or a backward step. The inset indicates the ratio of the rates of forward vs. backward steps as a function of the force applied: 1 mM ATP (blue) and 10 μM ATP (red). Note the logarithmic scale in the y-axis. The blue line is a fit: $802 \times \exp(-0.95 \text{ pN}^{-1} \times \text{Load})$. On the right panel (b), the dwell time is shown, which is the time during which the motor waited before taking the step. Given that the step itself is rapid, the dwell time ε is understood to be inversely proportional to the speed of movement of the motor, with $v = 8 \text{ nm}/\varepsilon$. Different symbols are used depending on whether a step was directed forward (closed symbols) or backward (open symbols): 1 mM ATP (blue) and 10 μM ATP (red). In (b), forward steps (closed symbols) are dependent on load: for loads between 0 and 7 pN, the fits for the dwell time are $\approx 0.0036 \text{ s} \exp(0.57 \text{ pN}^{-1} \times \text{Load})$ at 1 mM ATP (blue) and $\approx 0.0256 \text{ s} \exp(0.55 \text{ pN}^{-1} \times \text{Load})$ at 10 μM ATP (red). At high load, above 7 pN, backward steps (open symbols) are independent of force, but still depend on ATP. The nearly horizontal fits of the dwell time are $\approx 0.3 \text{ s}$ at 1 mM (blue) and $\approx 1.2 \text{ s}$ at 10 μM ATP (red). Reproduced from [20].

force. This adjustment is equivalent to adding the potential energy associated with a constant force. To a first approximation, the potential barrier is thus lowered by $-af$, where a is the horizontal distance from A to C. Consequently the rate of escape is changed:

$$\exp\left(-\frac{Q - af}{k_B T}\right) = \exp\left(-\frac{Q}{k_B T}\right) \exp\left(\frac{af}{k_B T}\right). \quad (13.3)$$

From this, we derive Eq. (13.1), where $k_0 = \exp\left(-\frac{Q}{k_B T}\right)$ is the escape rate in the absence of force. Hence Bell's law.

We now turn to some remarks about the three-dimensional world. In a one-dimensional framework, force is a scalar quantity. However, in the real three-dimensional world, force is a vector, and Kramers' theory might not apply directly. Specifically, pulling a molecular motor parallel to the microtubule on which it is attached, is expected to affect the unbinding rate differently than if one would pull in the perpendicular direction. In a more general context, sideways movements can also play a role in controlling detachment, as illustrated in many familiar devices. For example, think of how a hook or a key operates. At the molecular scale, these effects are difficult to measure experimentally, as the bond typically rotates in the direction of the applied force because of its flexibility, or the way in which it is anchored. Assuming such rotation occurs, we can consider f as the magnitude (the norm) of the force vector in the formulas above, and indeed Bell's law is often observed. This is the case for the detachment of a kinesin under load [24].

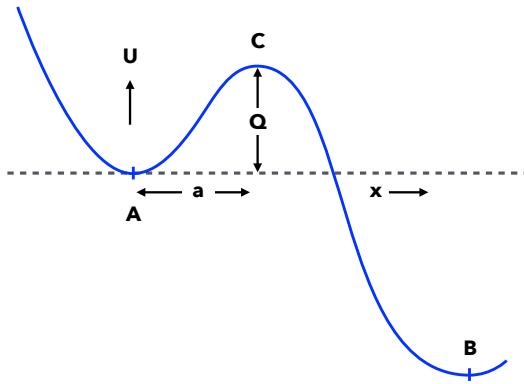


Figure 13.5: **Escaping from a one-dimensional potential landscape.** In Kramers' theory, the particle is initially trapped in the local minimum in A. To escape towards B, it must go over an energy barrier with a maximum in C. The rate of escape essentially depends on the probability of the particle to climb over the energy barrier (Ω). Adapted from [23].

13.4 Simulation of stochastic events with varying rates

We now turn to methods that are used to simulate a detaching bond, or a moving molecular motor, which are stochastic events that one can only predict statistically. It is common to simulate stochastic events in a computer using a source of randomness, a software facility known as a *random number generator*. The generation of good random numbers is a topic by itself. For simplicity, we assume that the function `random()` generates a random number uniformly distributed in the interval $]0,1]$. Each call to this function returns a new independent value, and for convenience we assume that zero is never returned. Let us first consider a *recurring* event characterized by a rate $\Gamma(t)$ in s^{-1} . By "recurring," we mean that the event can occur multiple times without memory of previous occurrences. The case of events that can only happen once is similar, but with the distinction that they do not recur. We will present three methods for simulating these events. While all of them produce correct results, they differ in the number of random numbers used and, consequently, their computational performance.

Example 13.1: Safe method. A common way to simulate stochastic events is to first divide time into small intervals of duration Δt , known as the *time step*. For each time step, one consumes one random number to compare with the probability of having an event within Δt . For our recurring event, this probability is $\Omega(t) = 1 - \exp(-\Delta t \Gamma(t))$. With this approach, the code is as follows:

```
t = 0;
while ( t < max_time )
{
    t = t + Delta_t;
    if ( random() < 1 - exp(-Delta_t * Gamma(t)) )
        do_event();
}
```

where `do_event()` invokes some code, and `Gamma(t)` returns the event's rate $\Gamma(t)$, which can be varying with time t .

The safe method is valid only if $\Delta t \Gamma(t) \ll 1$, and in this case $\Omega \approx \Delta t \Gamma(t)$. If the rate or the time step are so large as to reach $\Delta t \Gamma(t) \approx 1$, some events will be missed, since the code above will perform at most one event per time step, and the results will be wrong. Consequently,

- a. What time step can be used to simulate a system with two events, with rates $\Gamma_1(t)$ and $\Gamma_2(t)$?
- b. Generating fair random numbers can be an expensive task. Is it possible to reuse `random()` values over

multiple steps? Is this a safe practice?

The second method does not require the introduction of a time step, as it generates the time at which the event will fire directly:

Example 13.2: Direct method. By applying the *inverse transform sampling* method, one can generate exponentially-distributed random numbers using the logarithmic function:

```
delta = -log(random()) / Gamma(t);
```

Given that `random()` is a random number uniformly distributed in $]0, 1]$, the calculated stochastic time is appropriately positive. Note that `random()` must not return 0 for which the logarithm would not be defined. The pseudo code used to simulate the events until `max_time` with this method reads:

```
t = 0;
while ( t < max_time )
{
    delta = -log(random()) / Gamma(t)
    t = t + delta;
    do_event();
}
```

An important limitation of this approach is that it only applies when the rate $\Gamma(t)$ remains constant. The direct method serves as the foundation for a series of algorithms, the most well-known in the biology community being the *Gillespie algorithm*. These more advanced algorithms accommodate cases where the rate changes at discrete times, particularly when chemical reactions alter the number of reactive molecules, thereby modifying the propensity of reactions to occur.

Digression: Calculate the expected value of $-\log(\text{random}())$. Show that the associated distribution is exponential.

A third method was proposed in Ref. [25] to deal with time-varying rates:

Example 13.3: Hybrid method. A normalised time `esp` is first generated, as usual using a random number provided by `random()`. At each time step, `esp` is decreased as a function of the value of the rate during the interval. The associated event is performed if `esp` becomes negative. The timer `esp` is then updated by adding a random number. If this makes it positive, no further event is performed. The pseudo code is:

```
t = 0;
esp = -log(random());           // Exponentially distributed

while ( time < max_time )
{
    t = t + Delta_t;
    esp = esp - Delta_t * Gamma(t);
    while ( esp < 0 )
    {
        do_event();
        esp = esp - log(random()); // adding a positive number
    }
}
```

The method is simple to implement and will handle time-varying rates. It is based on a discrete integration

of the rate, within the precision dictated by the chosen time step. In this example, the event can be performed multiple times in the same time step, and this is necessary for the method to produce the correct number of events in general. This should naturally not be done for detachment and other events that can only occur once. In this case, the realization of the event effectively sets the rate to zero.

13.5 Final remarks and examples

In this chapter, we have introduced all the elements needed for simulating stochastic systems involving both mechanical processes and chemical reactions. First, we examined how certain events depend on force, such as the unbinding of a molecular bond or the stepping of a molecular motor along a filament. Second, we explored how to incorporate time-dependent reaction rates in stochastic simulations. The key part of this chapter lies in the following examples, where these concepts are combined.

In Example 13.4, one will compare different methods for simulating stochastic events with time-varying rates.

In Example 13.5, one will simulate the discrete movements of a kinesin motor along a microtubule.

In Example 13.6, one will consider a simple mechanical equilibrium, where the position of a filament is constrained by a molecular link.

These examples can be implemented in any programming language, with C++, Python, and Processing being commonly used. While this chapter focuses on key stochastic simulation methods, many other important aspects of cellular systems — such as Brownian motion, bending elasticity, and filament assembly dynamics — are not covered due to space constraints. However, these factors are essential for a more comprehensive model of the cytoskeleton. We thank Carlos Lugo-Velez, Eashan Saikia and Helen Saville for critically reading the document.

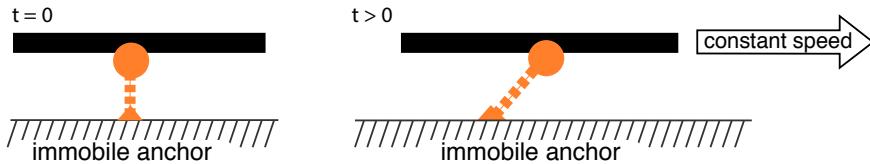


Figure 13.6: **Stretched link.** A molecule (orange) is anchored to a fixed position on an object that moves at a constant speed relative to a stationary substrate. As time progresses, the connecting link (dotted line) stretches increasingly. At what point will the bond break? See Example 13.4.

Example 13.4: Detachment of a stretched link under Bell's law. Consider a passive system in which a molecule is attached at a fixed position on a microtubule (Fig. 13.6). This molecule is anchored to a large immobile object (e.g., a coverslip), forming a Hookean link with stiffness k between the microtubule and the anchor ($k = 100 \text{ pN } \mu\text{m}^{-1}$). For simplicity, we assume the link has zero resting length and that the vertical distance between the coverslip and the microtubule is negligible. As the microtubule moves at a constant speed v , the link stretches progressively. Bond detachment follows Bell's law, with an unloaded dissociation rate of 0.01 s^{-1} and a characteristic force $f_0 = 3 \text{ pN}$.

a. Given that the bond is made unstretched at $t = 0$, derive the force on the bond at time t , assuming that it is still attached. Apply Bell's law to find the detachment rate. This rate ultimately defines how long the bond will stay attached.

b. Write the differential equations determining the evolution of the probabilities of being in the attached state, $p_{\text{att}}(t)$, and the detached state, $p_{\text{det}}(t)$, at time t . Write the conditions at $t = 0$.

c. Calculate the distribution of bound time, numerically or analytically. Start by estimating an order of magnitude for this time, as a function of the problem's parameters. What is the influence of each parameter on the bound time? Plot the histogram of bound time from simulations.

d. Implement two numerical methods to determine the probabilities of being in the attached and detached states, as described in this Chapter. Compare the distributions obtained with these methods. Evaluate the number of random numbers consumed in each case. Which advantages/disadvantages do

you find to each method?

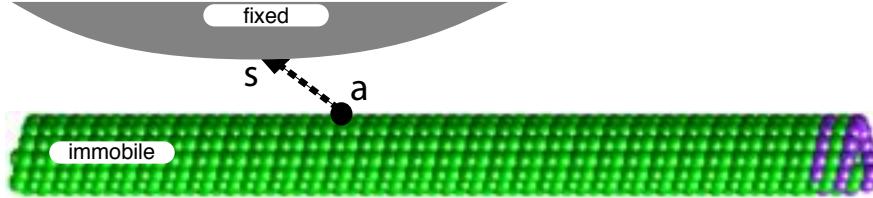


Figure 13.7: Kinesin walking along a microtubule. The kinesin (black arrow) moves along a microtubule, represented as a cylinder composed of green spheres, with the plus end (purple spheres) positioned to the right. The motor is anchored at a fixed position s on one side and attached to the microtubule at a on the other. See Example 13.5.

Example 13.5: Movement of a kinesin along an immobile microtubule. In this example, the objective is to develop a simulation of kinesin movement along a microtubule, based on experimentally measured forward and backward stepping rates (Fig. 13.4). Rather than explaining the detailed stepping mechanism, the focus is on reproducing the black trace in Fig. 13.3 by modeling how these rates are influenced by force. In this setup, the plus end of the microtubule is positioned to the right. For simplicity, we assume that kinesin moves along a single protofilament, reducing the problem to one dimension. Binding sites are spaced every 8 nm, and we consider a simplified model where the motor has a single binding domain (one foot). This allows us to represent its position using a single integer i , denoting the index of the binding site in the lattice. The task is to implement a stochastic model for the dynamics of i , capturing the random yet force-dependent movement of the kinesin motor along the microtubule.

One may assume initially that the motor is anchored at a fixed position s , and that it is attached to the microtubule at $a = i\delta$, with $\delta = 8 \text{ nm}$ the size of the tubulin heterodimer. The motor force is Hookean with stiffness k and zero resting length. The only possible transitions are to $i + 1$ (forward steps) and to $i - 1$ (backward steps), and both should depend on the force experienced by the link. Detachment can be ignored at first.

- a. Construct the backbone of the simulation, including the display. Draw the binding position of the motor on the microtubule. Optionally draw the lattice. Draw a thick line between s and a .
- b. Find the relationships for forward and backward steps measured experimentally (Fig. 13.4). From these relationships compute the expected stall force. Pay attention to units, and to the fact that rate and dwell time are each other's inverse.
- c. Implement stochastic transitions to the neighboring sites (i.e., evolution of the variable i). You may start with the simplest simulation method that can handle time-varying rates. Check that you are able to reproduce the black trace in Fig. 13.3.
- d. Instead of considering that the motor is anchored at a fixed position, modify the code to now anchor the motor on a position controlled by the mouse ($s = \text{mouse position}$). Verify that you can reproduce the behavior observed experimentally with optical traps (red and blue traces in Fig. 13.3).

Example 13.6: Gliding of a microtubule on immobile motors. Now that we have considered the movement of a motor on an immobile microtubule, we turn to the inverse problem: the movement of a microtubule propelled by immobilized motors (Fig. 13.8). This setup is widely used in laboratory experiments to assess motor activity, providing insights into their directionality and velocity. While this is the most challenging example, the following steps will help guide the development of the simulation. The objective is to simulate the movement of a gliding microtubule in one dimension, significantly simplifying the problem compared to a full two-dimensional simulation. As an initial step, consider a system with only a single motor attached. The microtubule's movement is governed by the external force acting on it, f ,

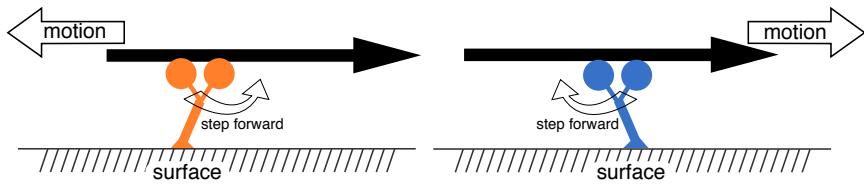


Figure 13.8: Gliding assays. Molecular motors immobilized on a surface via their tail domains can bind to microtubules floating in solution. If only one motor is attached, the microtubule remains free to pivot. However, the binding of additional motors can immobilize the microtubule on the surface. Upon encountering ATP, the motors' stepping motion propels the microtubules forward, allowing them to glide over the surface while engaging with other motors that continue their movement. When propelled by plus-end-directed (orange) motors, microtubules glide with their minus ends leading; conversely, when driven by minus-end-directed (blue) motors, they move with their plus ends leading. See Example 13.6.

divided by its mobility coefficient, which is determined by Stokes' drag:

$$\frac{dx}{dt} = \frac{f}{\gamma} \quad \text{with} \quad \gamma = 6\pi\eta R, \quad (13.4)$$

where η is the viscosity of the fluid ($\eta = 1 \text{ Pa}\cdot\text{s}$) and R the size of the microtubule, i.e., its length, a few micrometers. The links obey Hooke's law with a zero resting length, and stiffness $k = 400 \text{ pN}\cdot\mu\text{m}^{-1}$. Note that the viscosity used in this simulation will be 1000 times higher than that of water. However, this value accurately reflects the effective viscosity experienced by large objects, such as a microtubule, inside a living cell. The crowded cellular environment significantly increases resistance to movement compared to pure water.

a. Define the parameters of the problem. The microtubule position can be recorded by a scalar x , representing the position of the minus end. The orientation of the microtubule $d \in \{-1, 1\}$ will be constant during the simulation. The length of the microtubule L is also constant (e.g., $10 \mu\text{m}$). The motor is anchored at position s and it is attached on the microtubule at a distance a from the minus end. Verify the units of all quantities.

b. Derive the force in the link and the resulting force on the microtubule. Hooke's law specifies that $f = k(s - b)$ where s and b are the end positions of the link. Determine how b depends on the other parameters of the system. Show that the resulting force is an affine function of x , in the form $ax + b$.

c. Implicit integration. Different methods are used to integrate differential equations, but for this problem a first-order implicit scheme will work well. Given the equation of motion $\frac{dx}{dt} = v(x, t)$, implicit integration will expand to:

$$\frac{x_{t+\tau} - x_t}{\tau} = v(x_{t+\tau}, t). \quad (13.5)$$

Note that $x_{t+\tau}$ is used on the right-hand side, and not x_t as in the explicit scheme. Expand the force term and derive the formula giving $x_{t+\tau}$ as a function of x_t . Verify this formula by considering the results obtained with $\tau = 0$ and $\tau \rightarrow \infty$. Simulate with different values of the time step τ to test if/how it breaks down for higher values.

d. Finally implement the movement of the motor to simulate the system's dynamics. For simplicity, assume that the motor's movement is continuous: $\delta a = v\tau$, where the speed v is determined by a linear force-velocity relationship: $v = v_0(1 - f/f_0)$, where $v_0 = 1 \mu\text{m}\cdot\text{s}^{-1}$ and $f_0 = 5 \text{ pN}$. [Hint: pay attention to the sign of the force: an antagonistic force should hinder the motor movement.]

e. Extension: implement discrete movement of the motors, as seen in Example 13.5, to replace the continuous force-velocity relationship.

Bibliography

- [1] M. Kikumoto, M. Kurachi, V. Tosa, and H. Tashiro. Flexural rigidity of individual microtubules measured by a buckling force with optical traps. *Biophysical Journal*, 90:1687–96, 2006.
- [2] S. Banerjee, M. L. Gardel, and U. S. Schwarz. The actin cytoskeleton as an active adaptive material. *Annual Review of Condensed Matter Physics*, 11:421–439, 2020.
- [3] E. P. Sablin, F. J. Kull, R. Cooke, R. D. Vale, and R. J. Fletterick. Crystal structure of the motor domain of the kinesin-related motor ncd. *Nature*, 380:555–559, 1996.
- [4] F. J. Kull, E. P. Sablin, R. Lau, R. J. Fletterick, and R. D. Vale. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature*, 380:550–555, 1996.
- [5] R. D. Vale, T. S. Reese, and M. P. Sheetz. Identification of a novel class of proteins (kinesin) inducing organelle, bead, and microtubule movements: kinesin-induced movements. *Cell*, 42:39–50, 1985.
- [6] K. J. Verhey, D. L. Lizotte, T. Abramson, L. Barenboim, B. J. Schnapp, and T. A. Rapoport. Light chain-dependent regulation of kinesin’s interaction with microtubules. *The Journal of Cell Biology*, 143:1053–1066, 1998.
- [7] J. Howard, A. J. Hudspeth, and R. D. Vale. Movement of microtubules by single kinesin molecules. *Nature*, 342(6246):154–8, 1989.
- [8] R. D. Vale, T. Funatsu, D. W. Pierce, L. Romberg, Y. Harada, and T. Yanagida. Direct observation of single kinesin molecules moving along microtubules. *Nature*, 380:451, 1996.
- [9] K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365:721–7, 1993.
- [10] W. Hua, E. C. Young, M. L. Fleming, and J. Gelles. Coupling of kinesin steps to ATP hydrolysis. *Nature*, 388:390, 1997.
- [11] M. J. Schnitzer and S. M. Block. Kinesin hydrolyses one ATP per 8-nm step. *Nature*, 388:386, 1997.
- [12] E. Meyhöfer and J. Howard. The force generated by a single kinesin molecule against an elastic load. *Proceedings of the National Academy of Sciences*, 92:574–8, 1995.
- [13] E. Berliner, E. C. Young, K. Anderson, H. J. K. Mahtani, and J. Gelles. Failure of a single-headed kinesin to track parallel to microtubule protofilaments. *Nature*, 373:718, 1995.
- [14] S. Ray, E. Meyhofer, R. A. Milligan, and J. Howard. Kinesin follows the microtubule’s protofilament axis. *The Journal of Cell Biology*, 121:1083–1093, 1993.
- [15] S. Can, M. A. Dewitt, and A. Yildiz. Bidirectional helical motility of cytoplasmic dynein around microtubules. *eLife*, 2014:e03205, 2014.
- [16] A. Mitra, L. Meissner, R. Gandhimathi, R. Renger, F. Ruhnow, and S. Diez. Kinesin-14 motors drive a right-handed helical motion of antiparallel microtubules around each other. *Nature Communications*, 11:2565, 2020.
- [17] K. Kawaguchi and S. Ishiwata. Nucleotide-dependent single to double-headed binding of kinesin. *Science*, 291:667, 2001.
- [18] J. Howard. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Associates, 2001.
- [19] R. Vale and R. Milligan. The way things move: Looking under the hood of molecular motor proteins. *Science*, 288:88–95, 2000.
- [20] N. J. Carter and R. A. Cross. Mechanics of the kinesin step. *Nature*, 435:308–312, 2005.
- [21] G. I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200:618–627, 1978.

- [22] S. Walcott. The load dependence of rate constants. *Journal of Chemical Physics*, 128:215101, 2008.
- [23] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica VII*, 4:284–304, 1940.
- [24] Chris Coppin, Daniel Pierce, Long Hsu, and Ronald Vale. The load dependence of kinesin’s mechanical cycle. *Proc. Natl. Acad. Sci. USA*, 94:8539–8544, 1997.
- [25] A. Prados, J. J. Brey, and B. Sanchez-Rey. A dynamical Monte Carlo algorithm for master equations with time-dependent transition. *Journal of Statistical Physics*, 89:709–734, 1997.

Chapter 14

From Cells to Tissues

ROBERTO CERBINO, FABIO GIAVAZZI

Collective cellular movements are central to many physiologically and clinically relevant processes, such as embryogenesis, organogenesis, wound healing, as well as tumor progression and metastasis. This Chapter explores these phenomena through the lens of active network models, with a particular focus on *epithelial tissues*, shown on the left side of Fig. 14.3. Central to this analysis are two-dimensional *active network models*, especially those based on *Voronoi diagrams*. These models are adept at representing cell monolayers as polygonal tessellations, effectively capturing the dynamics of stationary and quasi-stationary tissues, as demonstrated on the right side of Fig. 14.1. We provide an overview of active network models by gradually introducing their key elements: We will first start with the description of the connection between the physical and geometric properties of isolated cells; we will then introduce various adaptations of the vertex model to describe epithelial tissues; finally, we will focus on self-propelled Voronoi models with alignment, in which one also takes into account cell–cell mutual orientational interactions occurring through the alignment of the cell polarity with its migration velocity.

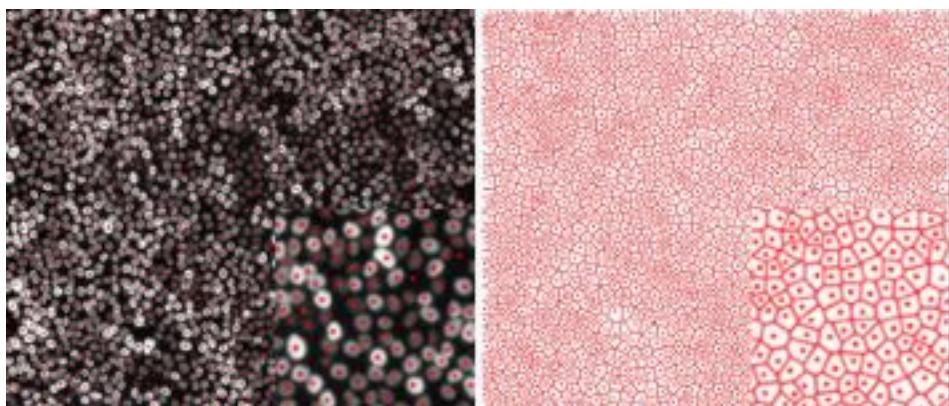


Figure 14.1: **Epithelial monolayer.** The nuclei of cells that form a simple epithelial tissue (monolayer) can be made visible in experiments by suitable fluorescent tagging and appear as bright spots over a dark background (left panel). Once the nuclei positions are determined (red dots), they can be used as generators of the corresponding Voronoi diagram (right panel), which is found to approximate well the edges of the cells to which the nuclei belong. In both panels, the inset is a $2 \times$ blow-up of a region in the main panel, and corresponds in real space to a square with side length of about $240 \mu\text{m}$.

14.1 Epithelial tissues as active networks

Many cellular processes, such as embryogenesis, organogenesis, wound healing, and tumor progression and metastasis, occur through collective cellular movements dictated by a complex interplay of biochemical and biophysical cues [1]. In this context, with biologists capable of producing remarkably controlled and flexible cellular model systems, physicists can ensure useful tools and unifying physical principles to characterize and model cellular collectives, as well as conceptual scaffolds to interpret the increasingly large amount of experimental data made available by the biologists.

A system of current interest for physicists is represented by epithelial tissues, which are cell layers that overlay skin and internal organs. According to their task, tissues are organized in very different structures. They can be *simple*, when formed by one layer of cells, or *stratified*, when they consist of more layers. Tissues made by cells whose diameter is larger (smaller) than their height are referred to as *squamous (columnar)*, whereas cells with diameter of the order of their height form the so-called *cuboidal* tissues. *Confluent tissues* are characterized by adjacent cells sharing the lateral surface with the nearest neighbors, so that no empty spaces are present inside the tissue.

A simple confluent tissue such as a monolayer is routinely obtained in the laboratory and, at the same time, can be modeled in a variety of ways ranging from considering cells as simple, isotropic, self-propelled particles to using phase field models that provide a more detailed description of the structural and dynamical properties of the cells. A good introduction to the topic, where the reader can get familiar with the majority of available models, is found in recent review articles in Refs. [2, 3].

With reference to the classification used in Ref. [3], we focus on two-dimensional (2D) *active network models* that, building on models originally developed for inert soft matter (e.g., foams), describe a cell monolayer as a network of prismatic cells that are self-propelling and interacting among each other with simple rules. The assumption that cells are prismatic is useful to describe the monolayer as a 2D polygonal pattern, typically located at the apical surface of the cell, i.e., the one far away from the basement membrane and facing instead the lumen or the external environment.

An elegant geometrical representation of a monolayer as a non-overlapping convex polygonal tessellation of the plane is represented by the *Voronoi model*. Each polygon of a Voronoi diagram is built around one of the generators (typically the cell nuclei for a monolayer) and encompasses all the points of the plane that are closer to that generator than to any other (Fig. 14.1). As shown in Refs. [4, 5], interpreting a monolayer as a Voronoi diagram works extremely well at the purely geometrical level for stationary or quasi-stationary tissues, whereas less accuracy in reproducing the cell boundaries is achieved in very dynamic settings such as the one encountered during development. Despite this, active network models [6, 7, 8, 9, 10, 11] have been shown to effectively describe the dynamic behavior observed in experiments with cells [6, 12, 13, 10, 14, 11], which are active entities capable of exerting a self-propulsion force. This force is characterized by a polarity that reflects the internal asymmetry within the cell, which in turn directs its motion.

Example 14.1: The Euler formula Consider a polygonal tessellation of the plane (not necessarily the Voronoi tessellation).

a. Prove the validity of the Euler formula

$$N - E + V = 1 \quad (14.1)$$

linking the number of polygons (N), edges (E), and vertices (V).

b. By assuming that each vertex has a connectivity of three (i.e., it is a three-branched object), show that the average number of edges of a cell $\langle e \rangle = 2E/N$ is given by

$$\langle e \rangle = 6 \left(1 - \frac{1}{N} \right). \quad (14.2)$$

14.2 Geometrical properties and mechanics of single cells

This section explores the geometrical properties and mechanics of single cells, crucial to understand complex cellular behavior. It begins with the *fluid drop model*, conceptualizing cells as incompressible fluid droplets, with their equilibrium shapes governed by the *Young–Laplace equation*. This model accounts for internal pressure and effective surface tension, influenced by the cell membrane's complex structure and the mechanics of the underlying cell cortex. The discussion then extends to *cell–cell adhesion*, examining its significant role in modulating effective surface tension at cell interfaces and in tissue formation. We also explore a *confluent monolayer* model of *prismatic cells*, to understand how variations in cell perimeter affect monolayer energy dynamics. Finally, the section addresses *traction forces* in cell motility, focusing on how cells exert forces on their substrate, leading to movement, and the collective dynamics within a cell monolayer.

14.2.1 Fluid drop model

An isolated cell in a homogeneous environment often takes a nearly spherical shape. This observed behavior resembles that of a liquid drop that, when immersed in a fluid with which it is not miscible, spontaneously takes the shape that minimizes its interfacial area. This similarity suggests that a minimal model for an isolated cell could be an incompressible fluid droplet, characterized by an internal pressure p_i and an effective surface tension γ_s .

Within this framework, the equilibrium shape of the cell is determined by the Young–Laplace equation

$$\Delta P = \gamma_s \left(\frac{1}{R_1} + \frac{1}{R_2} \right), \quad (14.3)$$

where $\Delta P = P_i - P_o$ is the pressure difference across a given portion of the cell surface (P_o is the external pressure), whose local curvature is specified by the principal radii of curvature R_1 and R_2 .

For soft cells, like granulocytes [15], the predictions of the liquid droplet model are in fair quantitative agreement with experiments probing the response of isolated cells to mechanical perturbations.¹ For example, results of pipette aspiration experiments can be consistently interpreted within this model, and the effective surface tension γ_s can be measured [16] (Example 14.2)

Example 14.2: Pipette aspiration experiment. One of the simplest experiments to probe the mechanical properties of an isolated cell is micropipette aspiration, where the tip of a pipette (a hollow cylinder with internal radius R_p) is put in contact with the cell surface. The pressure inside the tip is then progressively lowered, and the cell gradually enters the micropipette.

a. Modeling the cell as an incompressible liquid droplet with initial radius $R_0 \gg R_p$, internal pressure P_i and constant effective surface tension γ_s , calculate the critical value P_c of the suction pressure ΔP for which the portion of the cell aspirated into the pipette tip is a hemispherical cap of radius R_p . The suction pressure ΔP is defined as the difference between the atmospheric pressure P_o and the pressure P_p inside the micropipette tip.

b. Demonstrate that, if the suction pressure exceeds the critical value P_c determined in a, the whole cell tends to enter the micropipette.

14.2.2 Effective surface tension

Typical values of γ_s obtained from experiments are in the range 10^{-2} to $10^{-1} \text{ mN m}^{-1}$ [16]. For comparison, these values are about three orders of magnitude smaller than both the surface tension of water at room temperature and the tension of the phospholipid bilayer [17]. This is because γ_s is an effective quantity, which incorporates the effect of the complex structure of the cell membrane as well as the one of the underlying cell cortex, a layer of cytoplasmic proteins (mainly actin and myosin) whose thickness and contractile behavior modulate cell surface properties.

¹In some cases, modeling the cell interior as a viscous fluid is not consistent with the experimental results, which are better interpreted treating the cell as a piece of elastic material

At a first glance, the cell cortex can be modeled as an elastic shell of thickness δ much smaller than the cell radius enclosing the whole cytoplasm. When a square patch of a thin elastic sheet of thickness δ and linear size L is isotropically stretched by an in-plane force F acting on each one of its sides (Fig. 14.2a), the resulting isotropic strain is given by [17]

$$\varepsilon = \varepsilon_{xx} = \frac{1}{Y}(\sigma_{xx} - v\sigma_{yy}) = \frac{1-v}{Y}\sigma \quad (14.4)$$

where $\sigma_{xx} = \sigma_{yy} = \sigma = \frac{F}{L\delta}$ is the normal elastic stress, Y is the Young modulus of the elastic material, and v its Poisson's ratio. The internal stress within the stretched elastic sheet can be also expressed in terms of an effective surface tension

$$\gamma_s = \sigma\delta = \frac{\delta Y}{1-v}\varepsilon \quad (14.5)$$

whose value is determined by the elastic properties of the shell (via Y and v), by its current state of stretching (via ε), and by its thickness δ . An important ingredient that was not considered in writing Eq. (14.5) is the active contractile nature of the cell cortex. This can be accounted for by including, in the total internal stress σ_{tot} of the shell, beside the elastic term σ , also an active term σ_a , which we assume for simplicity to be isotropic, i.e., [19]

$$\sigma_{tot} = \sigma + \sigma_a. \quad (14.6)$$

By replacing σ with σ_{tot} and introducing the target deformation $\varepsilon_a = \frac{1-v}{Y}\sigma_a$ associated with the active stress, Eq. (14.5) becomes

$$\gamma_{s,tot} = \sigma_{tot}\delta = \frac{\delta Y}{1-v}(\varepsilon - \varepsilon_a). \quad (14.7)$$

It is worth noticing that, since active stresses in the cell cortex are contractile, σ_a (and so ε_a) is expected to be negative, its effect being to *increase* the effective surface tension of the cell.

14.2.3 Cell-cell adhesion

The liquid droplet model, which is based on including all the relevant contributions to the cell surface mechanics into an effective surface tension, can be extended to include also cell-cell interactions. At the molecular scale, cell-cell adhesion is mediated by transmembrane proteins (cadherins), which are anchored in the plasma membrane and are connected to the actin cytoskeleton. They can bind to other cadherins on the surface of neighboring cells, forming the so-called adherens junctions. When two cells with similar properties come in contact, cadherin bonds (each one with binding energy $u < 0$) are formed with a characteristic surface density ρ_b . Since the energetic contribution of this interaction is proportional to the contact area, it can be included in the description as an additive term in the surface tension.

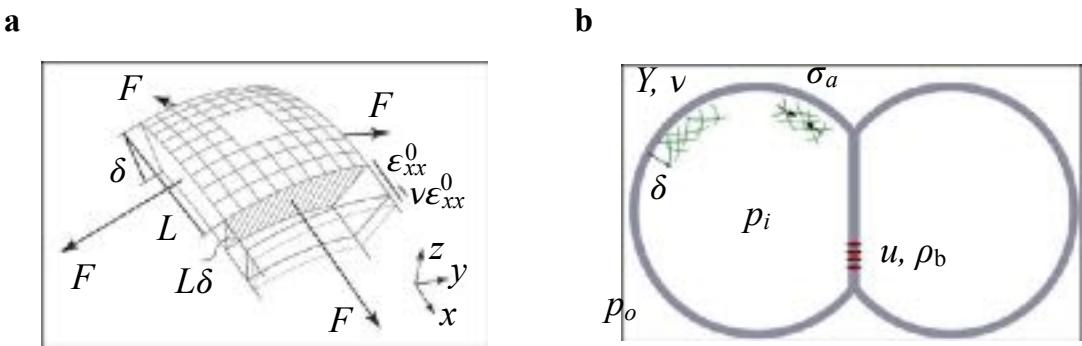


Figure 14.2: **Effective surface tension in cellular interfaces.** Schematic representation of the main contribution to the effective surface tension at the cell–medium interface and in the cell–cell contact regions. (a) Local deformation of an isotropically stretched thin elastic membrane. (b) Schematic representation of the main contributions to the effective surface tension at the cell–medium interface and in the cell–cell contact regions. Reproduced from Ref. [18].

The energy per cell and per unit contact area becomes

$$\gamma_{s,c} = \gamma_{s,tot} + \frac{1}{2}u\rho_b. \quad (14.8)$$

This expression shows how the effective surface tension $\gamma_{s,c}$ at the interface between two adherent cells is the result of the competition between two contributions: a positive term associated to cortical tension, and a negative one from cell-cell adhesion.

To get an idea of how cell-cell interactions impact on the interfacial properties of a tissue, let us consider the ideal case of an aggregate of N identical adhering cells, each one of area A . Let N_s be the number of cell at the boundary of the aggregate, while $N - N_s$ is the number of cells that are fully surrounded by other cells. Under the simplifying assumptions that the cortical tension $\gamma_{s,tot}$ is constant across the whole cell surface (see Ex. 14.3) we can estimate the total energy of the aggregate as

$$E_{tissue} = (N - N_s)\gamma_{s,c}A + N_s\gamma_{s,c}\frac{A}{2} + N_s\gamma_{s,tot}\frac{A}{2} = N\gamma_{s,c}A - N_s\frac{1}{2}u\rho_b\frac{A}{2}. \quad (14.9)$$

This expression is obtained under the additional simplifying assumption that exactly half of the area of a cell at the boundary is exposed to the external medium, while the other half is in contact with other cells. Using Eq. (14.8) and introducing the interfacial area of the tissue $A_{tissue} = N_s\frac{A}{2}$, Eq. (14.9) becomes $E_{tissue} = N\gamma_{s,c}A - \frac{1}{2}u\rho_bA_{tissue}$, from which we can readily identify $-\frac{1}{2}u\rho_b$ as the effective surface tension $\gamma_{s,tissue}$ of the tissue

$$\gamma_{s,tissue} = -\frac{1}{2}u\rho_b. \quad (14.10)$$

It is worth noticing that, while the adhesion term $\frac{1}{2}u\rho_b$ appears with the + sign in Eq. (14.8), its effect being of *decreasing* the effective surface tension at the cell-cell interface, it contributes with an opposite (−) sign to the effective surface tension of the tissue (Eq. (14.10)).

Example 14.3: Adhesion-induced changes in cortical tension. The formation of cell-cell contacts is known to induce significant local remodeling of the cell cortex (for example in terms of thickness and contractile activity), leading to a change in cortical tension and thus in the value of the total effective surface tension $\gamma_{s,tot}$ [20]. Let $\tilde{\gamma}_{s,tot}$ be the value of the total effective surface tension for a cell surface which is in contact with other cells. Determine how Eq. (14.10) should be modified to account for this effect.

A minimal model of planar cell assembly is that of a confluent monolayer of prismatic cells with constant height h , whose projected area is almost fixed, as shown in Fig. 14.3. In this geometry, if we neglect contributions from the boundaries, the total energy of the monolayer is determined by the shape of cell-cell contacts.

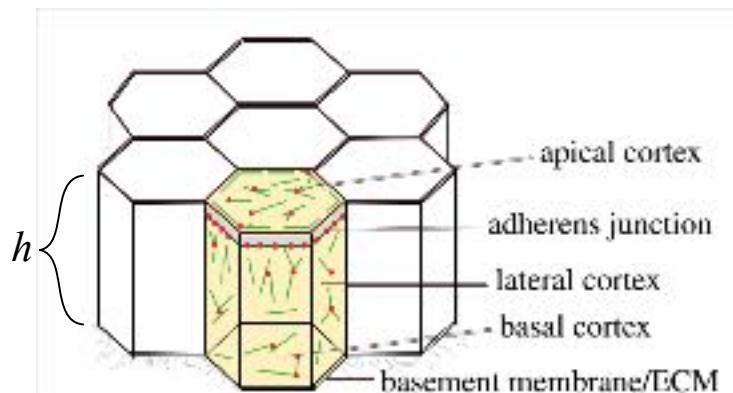


Figure 14.3: **Schematic representations of a cell monolayer.** Each cell is represented by a right prism with fixed height h . Reproduced from Ref. [21].

Indeed, the contribution to the total energy of a cell which is completely surrounded by other cells reads

$$E_c = \gamma_{s,c} h p, \quad (14.11)$$

where p is the cell perimeter and hp is the area of the lateral surface, which is in contact with neighbouring cells. It is worth noticing that, if one considers large perimeter fluctuations, the effective interfacial tension $\gamma_{s,c}$ in Eq. (14.11) cannot be assumed to be a constant. On one hand, a significant variation in the cell perimeter impacts on the deformation state of the cortex. In particular, increasing the cell perimeter has the effect of stretching the cell cortex, and this leads, according to Eqs. (14.5)-(14.11), to an increase in the cortical tension $\gamma_{s,tot}$ and, so, in $\gamma_{s,c}$ [22]. On the other hand, large variations in the perimeter (and thus in the cell-cell contact area) are likely to alter the density ρ_b of cadherins, which are present on the cell surface in a limited number [20]. In particular, an increase in the cell perimeter is expected to reduce ρ_b , and thus the value of the negative contribution to $\gamma_{s,c}$ due to cell-cell adhesion (Eq. (14.8)), leading, even in this case, to an increased $\gamma_{s,c}$.

Compared to the drastic assumption $\gamma_{s,c} = \text{constant}$, a more accurate description could be obtained by considering at least the first term of a Taylor expansion of $\gamma_{s,c}$ as a function of p at some given reference value \bar{p}

$$\gamma_{s,c} \simeq \gamma_{s,c} \Big|_{\bar{p}} + \frac{\partial \gamma_{s,c}}{\partial p} \Big|_{\bar{p}} (p - \bar{p}). \quad (14.12)$$

By plugging this expression into Eq. (14.11), we obtain

$$E_c \simeq \frac{\Lambda^{\text{el}}}{2} p + \frac{\Gamma^{\text{el}}}{2} p^2, \quad (14.13)$$

where $\Lambda^{\text{el}} \equiv 2h \left(\gamma_{s,c} \Big|_{\bar{p}} - \frac{\partial \gamma_{s,c}}{\partial p} \Big|_{\bar{p}} \bar{p} \right)$ and $\Gamma^{\text{el}} \equiv 2 \frac{\partial \gamma_{s,c}}{\partial p} \Big|_{\bar{p}}$. We note that Γ^{el} , according to the discussion above, is expected to be a positive constant, whereas the sign of Λ^{el} is not prescribed.

14.2.4 Traction forces and cell motility

Cells in contact with an external substrate exert traction forces on it. This typically occurs via the formation of protrusions that are attached to the substrate. Adhesion to the substrate takes place in localized contact patches called focal adhesion and is mediated by specialized transmembrane proteins (integrins). Traction forces are then generated by the actively contracting actin cytoskeleton that pulls on the adhesion domains. If the distribution of the adhesion domains and the contractile activity of the cell are fairly uniform, traction forces tend to balance and result in a contraction of the substrate, as the cell pulls on it from the boundary inwards

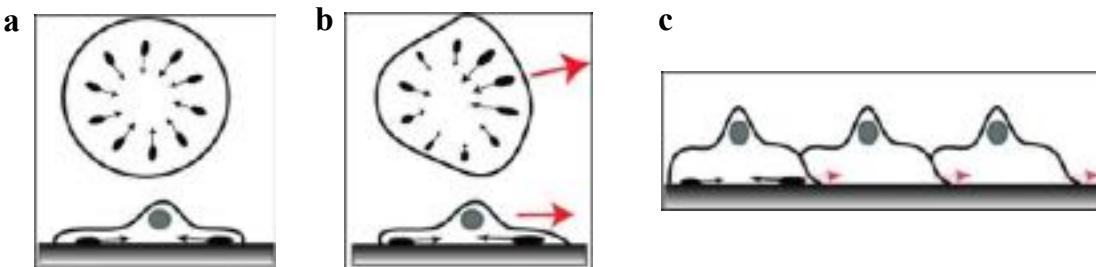


Figure 14.4: Traction forces in single and collective cell migration. Traction forces (black arrows) are exerted at the cell–substrate contacts (focal adhesion domains, represented by black ovals) due to active contraction of the actin–myosin cytoskeleton. (a) Within a nonmotile cell, these forces are uniformly distributed around the cell perimeter, pointing inwards, and result in a vanishing net traction. (b) In a motile cell, instead, the traction forces are unbalanced, being larger at the leading edge (lamellipodia), and the net traction force pulls the cell forward (red arrow). (c) Inside a confluent monolayer, cells generate strong traction forces in correspondence of polarized protrusion (cryptic lamellipodia, identified by red arrows), which tend to polarize neighboring cells in the same direction, promoting collective migration. Reproduced from Ref. [23].

(Fig. 14.4a). When instead traction-generating mechanisms are not balanced over the cell, a net force exists acting on the cell in one particular direction, promoting its directed motion (Figure 14.4b). The leading edge of the migrating cell usually develops large protrusions (called lamellipodia), where strong traction forces pull the cell forward. It is common to assume that the cell polarity is aligned in the same direction of the net traction force (Figure 14.4c).

14.3 Geometrical properties and mechanics of tissues

In this section, we explore the geometrical properties and mechanics of tissues through active network models, specifically *vertex models* and *self-propelled Voronoi models*. These models, based on space-filling polygonal networks, share similar energy functionals that incorporate deviations in cell area and perimeter. The primary difference lies in the degrees of freedom: Self-propelled Voronoi models focus on cell centers, making them more convenient for numerical implementation. The monolayer energy is expressed through elastic-like terms and edge tension, capturing the effects of actomyosin contractility and cell-cell adhesion. Mechanical equilibrium studies reveal a phase diagram with different tissue behaviors, indicating a rigidity transition between regular hexagonal networks and irregular soft networks, as well as a glass transition marked by a structural order parameter and cell self-diffusivity. Incorporating cell motility into self-propelled Voronoi models reveals a complex phase diagram with solid, liquid, and flocking states, influenced by cell motility and the target shape index. Furthermore, the introduction of alignment interactions, aligning cell propulsion with migration velocity, predicts flocking liquid and solid states, highlighting the role of alignment efficiency and structural lifetime. This comprehensive framework correlates well with experimental observations in epithelial monolayers, emphasizing the significance of unjamming-via-flocking in processes like metastasis, and bridging individual cell mechanics with collective tissue dynamics.

14.3.1 Cell tissues as active networks

Active network models encompass *vertex models* (Fig. 14.5) and *self-propelled Voronoi models* [3]. Both are confluent by construction, as they are based on space-filling polygonal networks, i.e., tessellations. As we

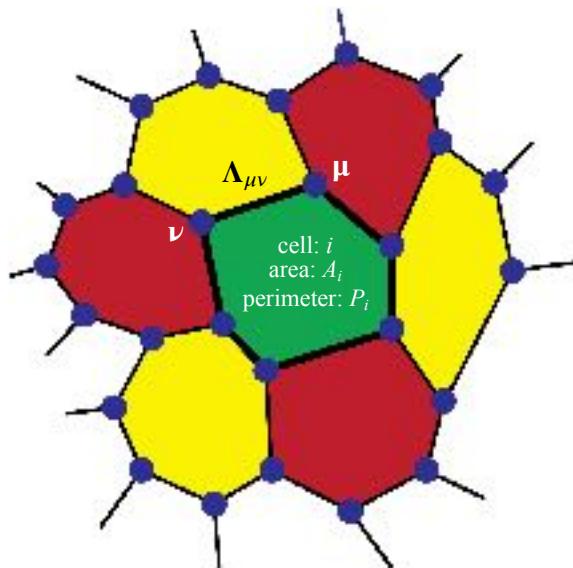


Figure 14.5: **Illustration of the parameters of the vertex model.** In the vertex models, three edges meet at a vertex (dark blue dots). Three parameters describe cell i : (1) reference area $A_{i,0}$, (2) area elastic modulus K_i^{el} , and (3) perimeter elastic modulus Γ_i^{el} . In addition, an edge connecting vertices μ and ν has tension $\Lambda_{\mu\nu}^{\text{el}}$. Reproduced from Ref. [8].

will shortly see, both models share also the same expression for the multi-body energy functional, which is essentially the sum of elastic-like terms encoding a quadratic penalty whenever the area and the perimeter of each cell deviate from their preferred values.

The main difference between self-propelled Voronoi models and vertex models is in the number of degrees of freedom that are kept to describe the tissue: In self-propelled Voronoi models, the degrees of freedom are the cell centers, whereas in vertex models they are the vertices of the polygons, which for a Voronoi tessellation are twice the number of cells. Another difference is that, when one thinks about forces exerted on/by a cell, it is more natural to think in terms of the cell center as a point of application of such forces. These considerations make self-propelled Voronoi models somehow more convenient, especially for numerical implementation. On the other hand, the vertices of the polygonal cells in the self-propelled Voronoi model are defined indirectly via the centers of neighboring cells, which complicates its use for cells with contracting edges (cell junctions). Finally, it must be taken into account that the two models can lead to subtle yet relevant differences when applied to the same polygonal network [24].

For our two-dimensional polygonal monolayer, the Euler formula holds (Eq. (14.1)), and one observes experimentally that almost all the vertices have a connectivity equal to three. According to Eq. (14.2), the average number of sides per cell becomes $\langle e \rangle = 6$ in the limit of large number of cells ($N \rightarrow \infty$). This relation holds on average and describes cases in which all the cells are hexagons (though they may be irregularly shaped), as well as cases in which a wider distribution of side numbers is observed. In the latter case, given that the minimum number of sides of a polygon is three with no upper limit, the distribution is often asymmetric.

It is interesting to notice that trivalent vertices are also observed in inert systems, even though the underlying physics can be different [25]. Nevertheless, this similarity in the topology was one of the main reasons driving Honda and coworkers (see Ref. [26, 27, 28] and the more recent review in Ref. [29]) to develop vertex models for the description of two-dimensional monolayers. Modern implementations of these vertex models are based on writing an energy function that encodes the predominant mechanisms acting to regulate the cell shapes within the monolayer. These mechanisms include cell-cell adhesion, which encodes the willingness of a cell to share an edge with another one; they also include the (active) elasticity of the cell cytoskeleton, which is mainly due to actin and myosin being localized at the cell cortex (acto-myosin contractility); finally, the quasi-incompressibility of the cytoplasm (mostly composed of water) and the resistance of the monolayer to thickness fluctuations lead to a sort of bulk elasticity.

14.3.2 The energy of a monolayer

Mathematically, the monolayer energy is typically written as [6, 8]:

$$E = \sum_{i=1}^N \frac{K_i^{\text{el}}}{2} (a_i - a_{i,0})^2 + \sum_{i=1}^N \frac{\Gamma_i^{\text{el}}}{2} p_i^2 + \sum_{\langle \mu, v \rangle} \Lambda_{\mu v}^{\text{el}} l_{\mu v}, \quad (14.14)$$

where the first two sums run over the $i = 1, \dots, N$ cells, the last one is a sum over all pairs of vertices that share an edge. K_i^{el} has units J m^{-4} and expresses the cell i rigidity to changes of its area a_i with respect to the reference value $a_{i,0}$. Similarly, Γ_i^{el} has units J m^{-2} and mirrors the cell rigidity to changes of its perimeter p_i from the zero value. Finally, $\Lambda_{\mu v}^{\text{el}}/2$ is the tension (units J m^{-1}) of the edge between the vertices μ and v , whose length is $l_{\mu v}$. It must be noted that Eq. (14.14) is valid for cells that belong to the bulk and neglects boundary effects, as already discussed in obtaining Eq. (14.11).

A useful simplification is obtained for a system of identical cells: All the cells have thus the same parameters ($a_{i,0} = a_0$, $\Gamma_i^{\text{el}} = \Gamma^{\text{el}}$, $K_i^{\text{el}} = K^{\text{el}}$); similarly, for all the edges, we have the same line tension ($\Lambda_{\mu v}^{\text{el}} = \Lambda^{\text{el}}$), which allows us to rewrite $\sum_{\langle \mu, v \rangle} \Lambda_{\mu v}^{\text{el}} l_{\mu v} = \Lambda^{\text{el}} p_i$. As a result, we have

$$E = \sum_{i=1}^N \left[\frac{K^{\text{el}}}{2} (a_i - a_0)^2 + \frac{\Gamma^{\text{el}}}{2} p_i^2 + \frac{\Lambda^{\text{el}}}{2} p_i \right]. \quad (14.15)$$

In light of the discussion made above for single cells, it is possible to sketch an interpretation of the different terms in Eq. (14.15). The first term is due to the fact that cells in three dimensions tend to maintain a constant volume. In our two-dimensional model, in which we assume that the thickness is almost constant, this implies that cells tend to maintain a constant area a_0 . In active network models, it is normally assumed that having a

cell with area $a_i \neq a_0$ in the tissue has an energetic cost that depends on the elastic constant K^{el} , and which is for each cell given by $\frac{K^{\text{el}}}{2} (a_i - a_0)^2$. Summing all these costs for the cells belonging to the monolayer provides the first term. As discussed in connection with Eq. (14.13), the second and the third terms can be interpreted as the result of a first-order expansion of the effective interfacial tension $\gamma_{\text{s,c}}$ of a cell around a reference perimeter, which gives two contributions. The third term in Eq. (14.15), proportional to the cell perimeter p_i , describes the energy cost of increasing the latter because of the line tension Λ^{el} exerted by the actin cable close to the apical surface of the cell (Fig. 14.2). The line tension Λ^{el} encodes the competition between the cortical tension along cell-cell contacts and the cell-cell adhesion energy. Finally, the second term in Eq. (14.15) can be interpreted as a linear change in tension due to a change in the cell perimeter, and is motivated by the fact that the actomyosin ring that underlies the adherens junctional network is expected to exert a tension that depends on cell perimeter.

By completing the square involving the cell perimeter, we obtain

$$E = \sum_{i=1}^N \left[\frac{K^{\text{el}}}{2} (a_i - a_0)^2 + \frac{\Gamma^{\text{el}}}{2} (p_i - p_0)^2 - \frac{\Gamma^{\text{el}}}{2} p_0^2 \right], \quad (14.16)$$

where we have introduced the reference perimeter $p_0 = -\Lambda^{\text{el}}/(2\Gamma^{\text{el}})$.

Choosing as a unit length the square root $\sqrt{a_0}$ of the target area allows us to define the dimensionless area $\alpha_i = a_i/a_0$, the dimensionless perimeter $\rho_i = p_i/\sqrt{a_0}$, the dimensionless reference perimeter (also known as *target shape index* in more recent literature) $\rho_0 = p_0/\sqrt{a_0}$, and to introduce the dimensionless energy

$$\epsilon = \frac{E}{K^{\text{el}} a_0^2} = N\epsilon_0 + \frac{1}{2} \sum_{i=1}^N \left[(\alpha_i - 1)^2 + \bar{\Gamma}^{\text{el}} (\rho_i - \rho_0)^2 \right], \quad (14.17)$$

where $\bar{\Gamma}^{\text{el}} = \Gamma^{\text{el}}/(Ka_0)$, $\bar{\Lambda}^{\text{el}} = \Lambda^{\text{el}}/(Ka_0^{3/2})$, and where $\epsilon_0 = -\bar{\Gamma}^{\text{el}}\rho_0^2/2 = -(\bar{\Lambda}^{\text{el}})^2/(8\bar{\Gamma}^{\text{el}})$ represents the minimum value of the energy, obtained when for all cells $\rho_i = \rho_0$ and $\alpha_i = 1$. Given that the relation $\rho_0 = -\bar{\Lambda}^{\text{el}}/(2\bar{\Gamma}^{\text{el}})$ holds, the monolayer energy depends on two dimensionless parameters, to be chosen among ρ_0 , ϵ_0 , $\bar{\Gamma}^{\text{el}}$, and $\bar{\Lambda}^{\text{el}}$. For instance, in Eq. (14.17) the two parameters are $\bar{\Gamma}^{\text{el}}$ and ρ_0 , but the dimensionless equivalent of Eq. (14.15) would depend explicitly on $\bar{\Gamma}^{\text{el}}$ and $\bar{\Lambda}^{\text{el}}$.

14.3.3 Ground states of a monolayer

The energy just found can be used to calculate the ground states (i.e., the most relaxed configurations) of the monolayer, by imposing mechanical equilibrium. If we perform this calculation (Problem 14.1), we obtain the phase diagram shown in Fig. 14.6, where the parameters chosen are $\bar{\Gamma}^{\text{el}}$, which as discussed in relation to Eq. (14.13) is expected to be a positive constant, and $\bar{\Lambda}^{\text{el}}$, which can be both positive and negative. When $\bar{\Lambda}^{\text{el}} < 0$ ($\bar{\Lambda}^{\text{el}} > 0$), cell boundaries tend to expand (shrink) and the cells tend to be elongated (roundish). Similarly, $\bar{\Gamma}^{\text{el}}$ is small when contractile forces are dominated by area elasticity forces, and vice versa.

In the striped region of the phase diagram, the ground state is a somehow uninteresting collapsed network of cells with $\rho_i = 0$ and $\alpha_i = 0$. In the gray region, we have a ground state in which a regular packing geometry is observed: For each value of ρ_i there is a well-defined value of α_i and the network is made of regular hexagons. This is a solid state with non-zero bulk and shear elastic moduli (i.e., we need to make work to compress/expand and shear the tissue), which is mainly caused by cortical tension dominating over cell-cell adhesion. The opposite is true in the cyan region, where the dominance of cell-cell adhesion causes a degenerate ground state, with different cell configurations having the same minimal energy: These are defined by all the cells having irregular shape but the same perimeter $\rho_i = \rho_0$, area $\alpha_i = 1$, and $E_{VM} = N\epsilon_0$. This state has a vanishing shear modulus and is thus more similar to a liquid, in which the system can change configuration among the possible ones without energy cost. The line separating the two regions can be thought of as a rigidity transition line, arising as the competition between cortical tension and cell-cell adhesion, with cell-cell adhesion having a rather counterintuitive effect of favoring cell motility within the tissue (known as the “adhesion paradox” [12]).

14.3.4 The effect of cell division

The just-described model does not place any bounds on the tissue density (e.g., in the striped region the tissue collapses), and in Ref. [6] simulations based on it are conducted, in which cell division (mitosis) is introduced as

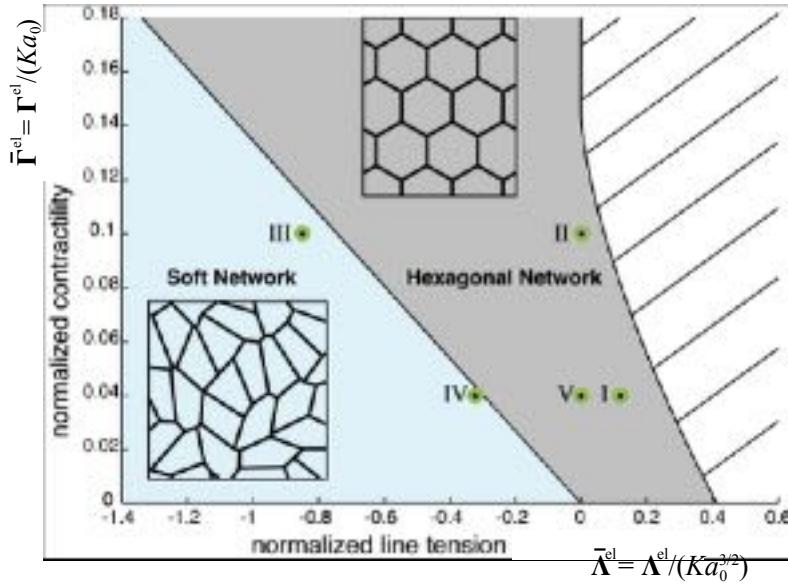


Figure 14.6: **Ground-state diagram of the vertex model.** Diagram of the ground states of the energy function E as a function of the normalized line tension $\bar{\Lambda}^{\text{el}}$ and contractility $\bar{\Gamma}^{\text{el}}$. In the grey region, the ground state is a hexagonal network. In the cyan region, the ground states are irregular soft networks, and many configurations coexist. In the striped region to the right, cell areas vanish and the model breaks down. Green dots indicate parameter values of five different cases (see Ref. [6] for details): Case I ($\bar{\Gamma}^{\text{el}} = 0.04$ and $\bar{\Lambda}^{\text{el}} = 0.12$) is the one whose properties agree more closely with experiments performed on *Drosophila* (see Fig. 14.8). Reproduced from Ref. [6].

a source of disorder in the tissue. Due to cell division, the system configurations deviate from the ground states described above, and the tissue finds local energy minima by relaxation that involves topological transitions similar to those occurring in foams: *T1 events* involve the exchange of neighbors, with one edge connecting two cells that shrinks to zero length and gives rise to a new line of contact between cells that were not sharing an edge before (see Fig. 14.7) *T2 events* involve the shrinking of a polygon to a point, and are typically related to apoptosis in cells. In this framework, mitosis is captured by a process (a sort of inverse of a T2 event), in which a cell is replaced by two thanks to the appearance of an edge. After a mitosis event, cells can then undergo T1 and T2 events to evolve toward a local equilibrium configuration. In Fig. 14.8, we show a typical equilibrium configuration that is obtained in simulations with cell division, conducted in the gray area of the phase diagram in Fig. 14.6. In this region, it is possible to find a steady-state pattern that compares well with experiments on *Drosophila* dividing tissues, pointing to a solid-like nature of real tissues at steady state [6].

14.3.5 Rigidity transition

In a confluent tissue where mitosis and apoptosis are rare events, T1 events are the only mechanism by which cells can rearrange. The rate at which these rearrangements occur, which is determined by the height of the mechanical energy barrier associated with a T1 event, determines the mechanical response of the tissue.

In Ref. [30], Bi *et al.* studied in more detail the vertex models by focusing on T1 transitions as the sole mechanism of cell rearrangements. To describe the monolayer, Bi *et al.* choose the target shape index p_0 and the inverse perimeter modulus $r = 1/\bar{\Gamma}^{\text{el}}$ and studied numerically how much the energy barriers for T1 transitions change with r and p_0 for homogeneous disordered polygonal monolayers. It is worth mentioning that in Ref. [30], all the results were obtained by assuming that the reference area a_0 is equal to the average cell area $\langle a \rangle = a_{\text{ml}}/N$ within the monolayer of area a_{ml} , so that cells are on average neither stretched nor compressed. Incidentally, this assumption makes the single cell minimum energy ϵ_0 a constant, which is the reason why it was discarded in Ref. [30], and will be discarded in the rest of this Chapter.

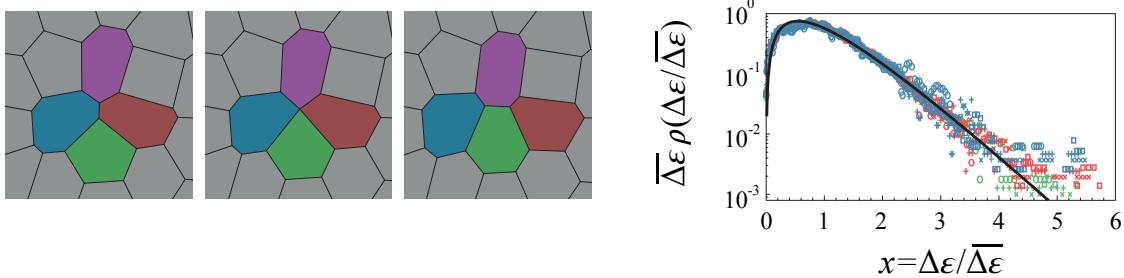


Figure 14.7: **Energy barriers for local cellular rearrangements in the vertex models (I).** Illustration of a T1 transition in a confluent tissue and the normalized distribution ρ of normalized energy barrier heights $\Delta\varepsilon/\bar{\Delta\varepsilon}$ for various parameters ($r = 1/\Gamma^{\text{el}} = 0.5, 1, 2$ and $p_0 = 3.2\text{--}3.7$). The solid line is a $k\Gamma^{\text{el}}$ distribution that completely specifies the mechanical response with the sole parameter $\bar{\Delta\varepsilon}$. Reproduced from Ref. [30].

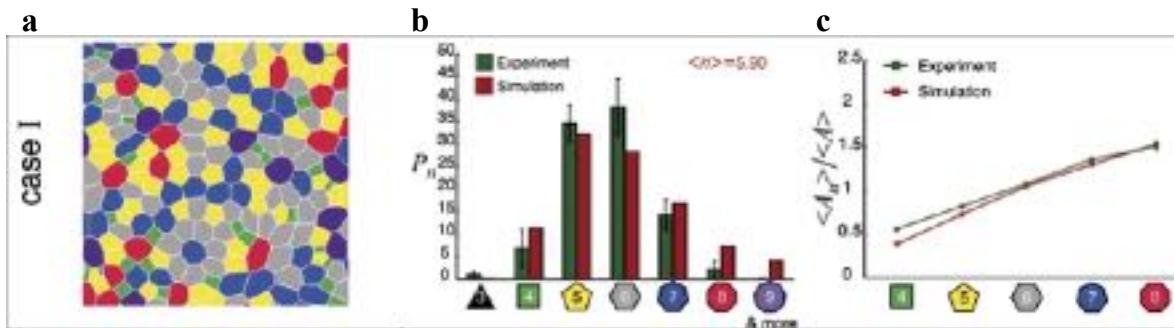


Figure 14.8: **Matching of the vertex model with experiments.** The stationary state of the vertex model including cell division matches experimental observations on *Drosophila*. When the ground states in Fig. 14.6 are perturbed by taking into account disorder induced by cell division, the steady state obtained with simulations performed in case I is the one that more closely matches the properties of the experimental pattern observed with *Drosophila*. (a) Stationary network pattern generated by repeated cell division for case I in Fig. 14.6. The color code for the polygon classes is the same as that used in b. (b) Stationary distribution of neighbor numbers P_n for the pattern in a. The average neighbor number is indicated. The green bars indicate the same experimentally determined distribution of neighbor number, for comparison. (c) Average areas of different polygon classes normalized to the average area of cells in the network, for the pattern in a (red line). The green line describes the experimentally determined values. Reproduced from Ref. [6].

As can be seen in Fig. 14.9, for $p_0 \lesssim 3.81$, energy barriers attain a finite value for every value of r , i.e., cells must do some work in order to undergo a deformation and execute a T1 transition. For these values of the target shape factor, the tissue behaves thus like a solid. As p_0 is increased, the barrier height diminishes until it becomes vanishingly small in proximity of $p_0 \simeq 3.81$, value for which the T1 transition occurs without energy cost. The vertex models presents thus a rigidity transition, whose sharpness is set by r (Fig. 14.10), for a value of p_0 that is close to the value $p_{0,\text{pent}} = 2\sqrt{5}(5 - 2\sqrt{5})^{1/4} \simeq 3.81$ of a regular pentagon.

Inspection of Fig. 14.10 also reveals that the analysis conducted in Ref. [30] to study disordered metastable states and transitions between them brings-in some new physics. Explicitly including nonlinear effects via the energy barriers uncovers a rigidity transition (confirmed also by numerical studies of the shear modulus performed in Ref. [30]) at a value $p_0 = p_0^* \simeq 3.81$ that is substantially identical to $p_{0,\text{pent}}$, and remarkably different from the value $p_{0,\text{hex}} = 3.72$ predicted by linear stability analysis of the single cell behavior [6, 22]. A more detailed study performed in Ref. [31] shows that in the region $3.72 \lesssim p_0 \lesssim 3.81$, the ordered ground states of the two-dimensional vertex model are linearly unstable but non-linearly stable with respect to localized cell rearrangements, which confirms the correctness of locating the rigidity transition at $p_0 = 3.81$. In addition, the predictions of Ref. [30] are also in agreement with experimental results obtained with bronchial epithelial cells

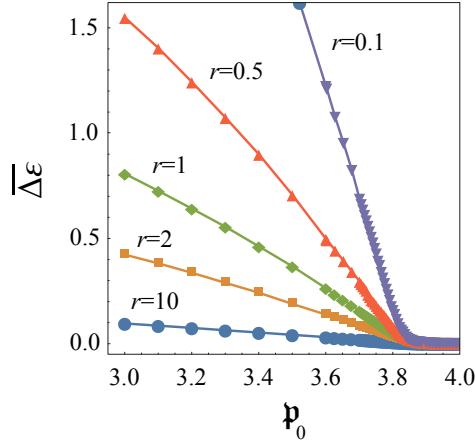


Figure 14.9: **Energy barriers for local cellular rearrangements in the vertex models (II).** $\overline{\Delta\epsilon}$ as function of the target shape index p_0 for various values of the inverse perimeter modulus r . Reproduced from Ref. [30].

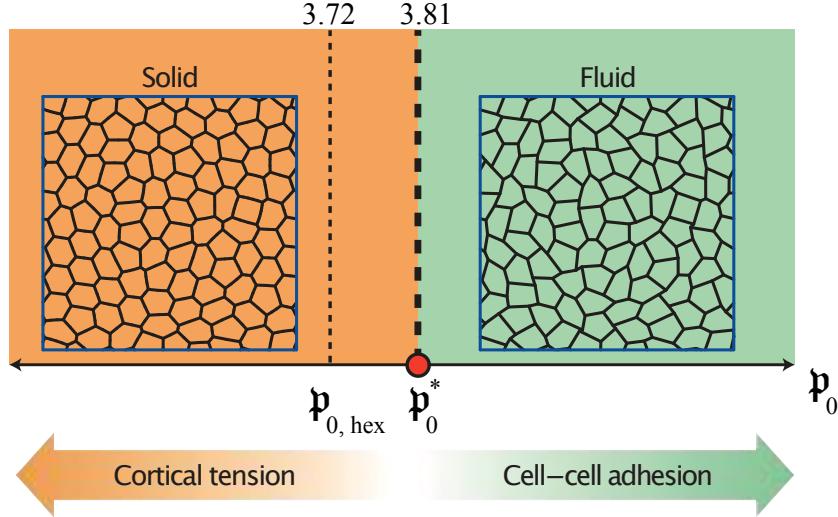


Figure 14.10: **Rigidity transition in confluent tissues.** The rigidity transition is demonstrated in a simple phase diagram as a function of p_0 , snapshots are taken from a typical rigid tissue ($p_0 = 3.7$) and fluid-like tissue ($p_0 = 3.96$). A rigidity transition occurs at $p_0 = p_0^* \approx 3.813$ for disordered metastable tissue configurations. The vertical dashed line is the prediction of Eq. (14.22) with $n = 6$, which leads to the value $p_{0,\text{hex}} = 3.722$: Below $p_{0,\text{hex}}$, the ground state is a hexagonal lattice, and above $p_{0,\text{hex}}$, the ground state is disordered. Reproduced from Ref. [30].

[12], when the *mean cellular shape* q :

$$q = \frac{1}{N} \sum_i \frac{p_i}{\sqrt{a_i}} \quad (14.18)$$

is measured. The parameter q is an experimental observable obtained by averaging over all cells, and thus describes a given sample; by contrast, p_0 is the target shape index that tunes the glass transition, and describes a preferred value that each single cell tries to reach. In the language of phase transitions, we would thus say that p_0 is the control parameter of the solid-to-liquid transition, where q is the order parameter, a fact that we will further discuss in the next paragraph (see discussion of Fig. 14.11).

14.3.6 Adding cell motility

The first attempt to incorporate cell motility in active network models was made in Ref. [7], where a self-propelled Voronoi model was preferred to a vertex model because of the more natural implementation of forces acting on cell centers rather than vertices: each cell is characterized by its position \mathbf{r}_i , which simplifies the consideration of individual cell motility. Assuming overdamped dynamics, the equation of motion for the center of the i -th cell is given by

$$\frac{\partial \mathbf{r}_i}{\partial t} = \mu(\mathbf{f}_i + \mathbf{F}_i), \quad (14.19)$$

where $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E$ is the force arising from the tissue energy in Eq. (14.16) (or equivalently by its dimensionless version Eq. (14.17)) and where $\mathbf{f}_i = f_0 \mathbf{n}_i = (v_0/\mu) \mathbf{n}_i$ of fixed magnitude f_0 , pointing along the direction $\mathbf{n}_i = (\cos \theta_i, \sin \theta_i)$ of cell polarity. In Ref. [7], the simplest dynamics for the cell polarity vector is considered, which corresponds to rotational diffusion according to the Langevin equations

$$\partial_t \theta_i = \eta_i(t), \quad \langle \eta_i(t) \eta_j(t') \rangle = 2D_r \delta(t - t') \delta_{ij}, \quad (14.20)$$

where θ_i is the polar angle associated to \mathbf{n}_i and η_i is a delta-correlated white-noise process with zero mean and variance equal to $2D_r$. The value of the angular noise D_r sets the persistence time $\tau_n = 1/D_r$ of the polarity vector: For small D_r , each cell tends to keep a constant polarity for long time, whereas, for large D_r , the polarity changes randomly at every step.

In addition to choosing $\sqrt{A_0}$ as the natural length scale, we can express all times in units of $1/(\mu K^{\text{el}} A_0)$. The free parameters of the problem are thus the target shape factor p_0 , the normalized contractility $\bar{\Gamma}^{\text{el}}$, the normalized self-propulsion speed $v_0/(\mu K^{\text{el}} A_0^{3/2})$, and the rotational noise strength $D_r/(\mu K^{\text{el}} A_0)$. The phase diagram for the case $\bar{\Gamma}^{\text{el}} = 1$ and $D_r = 1$ (obtained by neglecting proliferation as previously done in Ref. [30]) is shown in Fig. 14.11a.

As expected, in the absence of cell motility a critical target shape $p_0^* \simeq 3.81$ distinguishes a solid state ($p_0 < p_0^*$) from a liquid one ($p_0 > p_0^*$). However, as noted in Ref. [24], this is a glass transition rather than a rigidity transition (jamming) like it was for the vertex model, as it does not occur in the limit of zero temperature, in which the system is solid for all values of p_0 due to over-constraints (the degrees of freedom in a self-propelled Voronoi model are half of those in a vertex models, because the number of vertices is twice the number of cells).

Further inspection of the phase diagram shows that p_0 plays the role of the control parameter of the glass transition. For arbitrary values of the self-propulsion speed v_0 , a critical value of p_0 exists: For large enough values p_0 , the cells in the tissue are free to move as in a liquid (Fig. 14.11b, upper plot); by contrast, a tissue with small enough p_0 is stuck in a certain configuration that constitutes a local minimum of the energy (Fig. 14.11b, lower plot); in this solid-like state, cells lie in potential wells determined by their neighbors and in the absence of an external energy input they cannot escape from those cages; self-propulsion speed provides additional kinetic energy that allows cells to have larger probability to get over the energy barrier associated with T1 transition, exchange their neighbors, and move in a diffusive-like way (Fig. 14.11b, upper plot).

The existence of a glass transition in Fig. 14.11a is supported by an analysis of the mean-squared displacement $\langle \Delta r(t)^2 \rangle$ of the cell trajectories. For a fixed value of v_0 ($v_0 = 0.1$ in Fig. 14.11c), we observe that for small values of p_0 , $\langle \Delta r(t)^2 \rangle$ exhibits the typical intermediate plateau of glassy systems. Interestingly, the system eventually displays a long-time diffusive regime, but only for large values of p_0 , which suggests how the self-diffusivity $D_s = \lim_{t \rightarrow \infty} \langle \Delta r(t)^2 \rangle / (4t)$ may be a proxy of this glass transition. This hypothesis is supported by the results in Fig. 14.11d: The normalized self-diffusivity $D_{\text{eff}} = D_s/D_0$ of the cells (here, $D_0 = v_0^2/(2D_r)$ is the free diffusion constant of an isolated cell) is a dynamical order parameter for the glass transition becoming non-zero in its proximity. Remarkably, the glass transition can also be identified by a structural order parameter given by the mean cellular shape q , defined in Eq. (14.18), which is equal to 3.81 at the phase boundary in Fig. 14.11a.

14.3.7 Adding cell-cell alignment

Introducing alignment in the self-propelled Voronoi model can be done in a variety of ways [8]. Here, we follow Ref. [9], where an active feedback mechanism was introduced at the single cell level to align each cell's polarity vector with its migration velocity (Fig. 14.12), which is in turn controlled by interactions with other

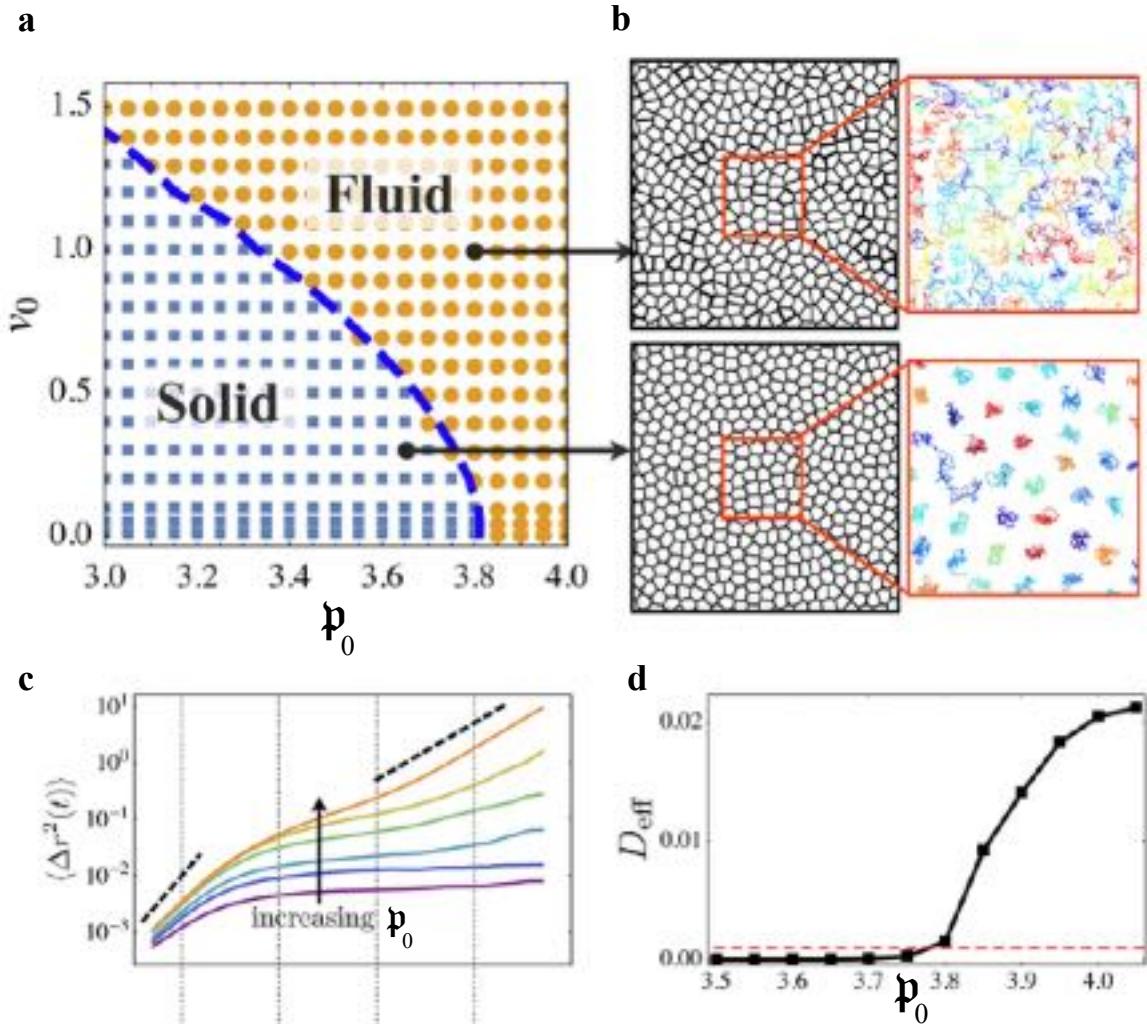


Figure 14.11: **Glass transition in confluent tissues with cellular motility.** (a) Phase diagram for confluent tissues as a function of cell motility v_0 and target shape index p_0 . Blue data points correspond to solid-like tissue; orange points correspond to a liquid-like tissue. The dashed phase line coincides with the locations in phase space where the mean cellular shape q (Eq. (14.18)) is equal to 3.81, which means that q is a structural order parameter for the glass transition — in the solid phase for $q \simeq 3.81$, whereas for $q > 3.81$ in the fluid phase. (b) Instantaneous tissue snapshots show the difference in cell shape across the transition. Cell tracks also show dynamical arrest due to caging in the solid phase and diffusion in the fluid phase. (c) The mean-squared displacement of cell centers for $D_r = 1$ and $v_0 = 0.1$ and various values of p_0 (bottom to top: $p_0 = 3.5, 3.65, 3.7, 3.75, 3.8, 3.85$) show the onset of dynamical arrest as p_0 is changed indicating a glass transition. The dashed lines indicate a slope of 2 (ballistic) and 1 (diffusive) on a log-log plot. (d) The effective self-diffusivity D_{eff} as a function of p_0 at $v_0 = 0.1$. At the glass transition D_{eff} becomes nonzero. Reproduced from Ref. [7].

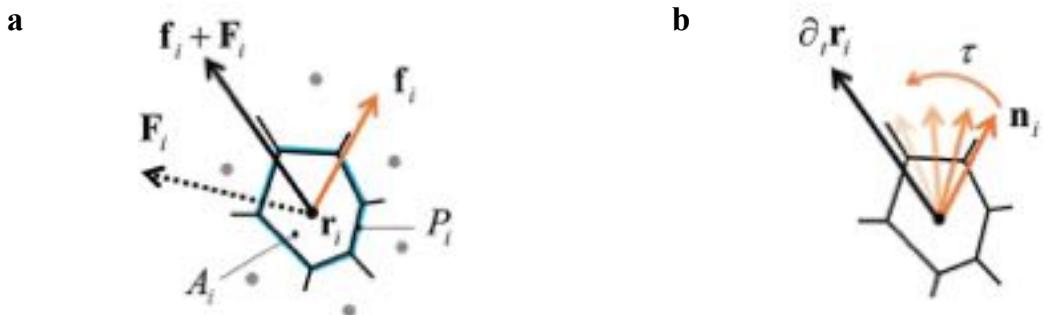


Figure 14.12: **Alignment interactions in a confluent tissue.** (a) Each cell is a polygon obtained by the Voronoi tessellation of initially random cell positions \mathbf{r}_i , characterized by the area A_i and the perimeter P_i of the polygon. The cell experiences a force $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} E$ due to its neighbors and an internal propulsive force \mathbf{f}_i along the direction \mathbf{n}_i of its polarity vector (Eq. (14.19)). (b) An active orientation mechanism reorients each cell's propulsive force towards its migration velocity over a characteristic response time $\tau = J^{-1}$ (Eq. (14.21)). Reproduced from Ref. [9].

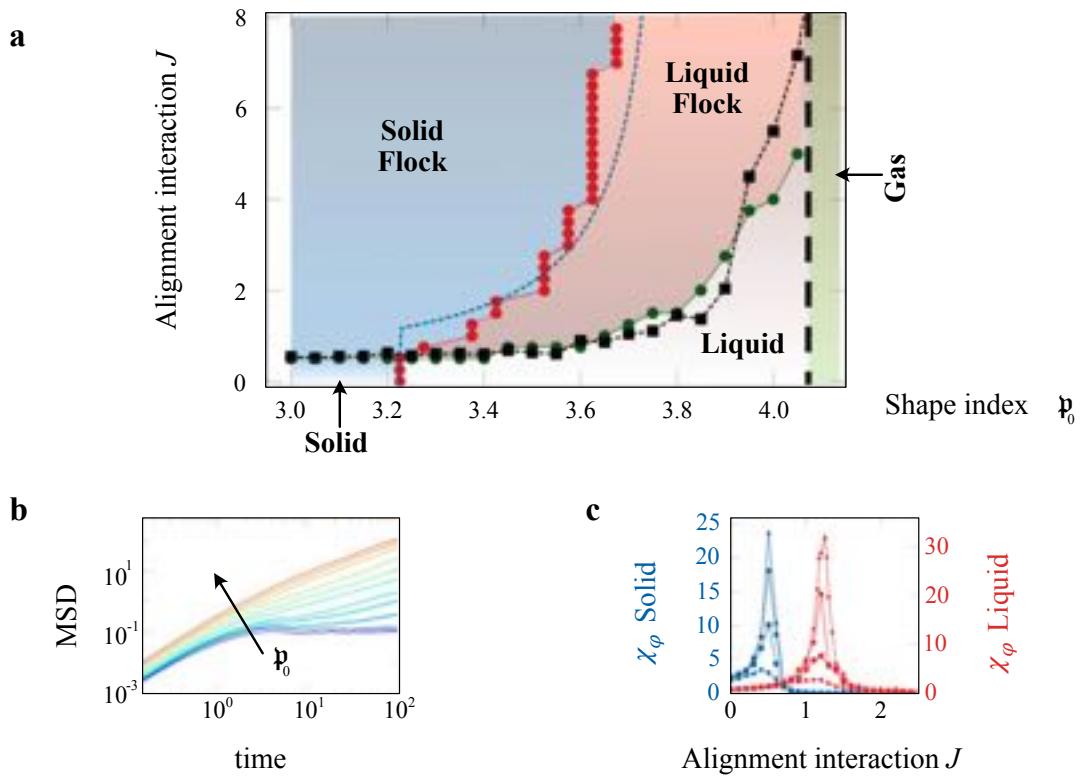


Figure 14.13: **Phase diagram for a confluent tissue in the presence of alignment.** (a) Different phases in the (p_0, J) plane. The solid–liquid transition line (red circles) is obtained from the vanishing of D_{eff} and the flocking transition line (green circles) corresponds to the peak in the susceptibility χ_φ . The dashed blue curve is an approximate analytical prediction for the boundary line given in Ref. [9]. The black squares (the dashed line is a guide to the eye) are the estimate for $J_{\text{flock}}(p_0)$ in terms of the numerically calculated cage lifetime τ_{cage} at $J = 0$. The vertical dashed black line marks the transition to a gas-like state, observed for $p_0 \gtrsim 4.2$, where flocking cannot occur. (b) The mean square displacement for $J = 2.0$ for a range of $p_0 \in [3.4, 4]$ across the liquid–solid transition (curves from red to violet). (c) The susceptibility χ_φ for $p_0 = 3.1$ (blue symbols, solid) and $p_0 = 3.7$ (red symbols, liquid) and sizes $N = 100, 400, 1600, 3200$, diamonds, squares, circles, and triangles, respectively. Reproduced from Ref. [9].

cells. In practice, this is done by replacing the time evolution equation for θ_i in Eq. (14.20) with the following equation

$$\partial_t \theta_i = -J \sin(\theta_i - \phi_i) + \eta_i, \quad (14.21)$$

where ϕ_i is the direction of the cell velocity $\partial_t \mathbf{r}_i = \mathbf{v}_i = v_i (\cos \phi_i, \sin \phi_i)$. Compared to Eq. (14.20), here we have the competition between the rotational diffusivity D_r and the alignment rate J . The latter is related to the response time $\tau = 1/J$ required by the cell to reorient its polarization in the direction of the resultant force exerted by its neighbors.

In Fig. 14.13, we show the phase diagram that was obtained numerically in the absence of proliferation for the case $\bar{\Gamma}^{\text{el}} = 1$, $D_r = 0.5$, and $v_0 = 1$ [9]. The phase diagram shows four distinct phases that can be obtained by varying the alignment rate J and the target shape index p_0 . In addition to the liquid and solid states that were found in the previous paragraph for the case $J = 0$, here we observe the emergence of flocking states, characterized by a large degree of alignment of neighboring cells.

To study the solid–liquid transition, the mean-square-displacement of the cells is evaluated in the reference frame of the center of mass, to get rid of possible collective directional motility. The associated normalized self-diffusivity D_{eff} serves as a dynamical order parameter for the onset of glassiness. The transition line $D_{\text{self}} \leq 10^{-3}$ (red circles in Fig. 14.13a) corresponds to $q = 3.813$. This model has also an additional transition at $p_0 \sim 4.2$, which is very close to the shape index ($p_0 = 4.186$) for a Voronoi pattern of uniformly distributed points, suggesting that above this value cells are effectively non-interacting and behave like a gas.

The emergence of flocking is quantified by using the Vicsek order parameter $\varphi \equiv N^{-1} \langle |(\sum_i \mathbf{v}_i / |\mathbf{v}_i|)| \rangle$, where the angular brackets indicate the average over trajectories. This quantity vanishes when cells are moving in random directions and attains a value of 1 when all cells coordinate their motion. The susceptibility $\chi_\varphi = \langle (\varphi(t) - \langle \varphi \rangle)^2 \rangle$ exhibits a maximum at the flocking transition, which was used in Ref. [9] to separate flocking from non-flocking states (green circles in Fig. 14.13a).

The four states of the phase diagram are thus defined as follows: (1) a *stationary solid* with vanishing D_{eff} , corresponding to the absence of cellular rearrangements, and $\varphi = 0$; (2) a *stationary liquid* with finite D_{eff} and vanishing mean motion ($\varphi = 0$); (3) a *flocking liquid* where cells flow collectively (D_{eff} and φ are both finite); and (4) a *flocking solid* where the tissue migrates as a unit (φ finite), while maintaining its rigidity.

Broadly speaking, alignment is possible when the alignment rate J is larger than the rate at which structural changes occur in the monolayer. Starting from the solid state, where cells are translationally caged, alignment occurs as soon as alignment overcomes the orientational noise $J > D_r$, which explains the horizontal asymptote at 0.5 of the flocking boundary for small p_0 . When p_0 is increased and the system liquefies, translational rearrangements become possible, which increase the critical value of J needed to observe flocking. The black squares in Fig. 14.13 are an estimate of the flocking transition obtained by comparing the lifetime τ_{cage} of the local cages with the characteristic alignment time $\tau = 1/J$. The agreement between this estimate of the flocking line and the one based on the susceptibility (green circles) confirms the validity of the interpretation of flocking in terms of competition between alignment efficiency and structural lifetime.

Interestingly, and somehow counterintuitively, alignment additionally promotes solidification. This can be understood as follows. If for $p_0 < p_0^*$ a tissue is a liquid also in the absence of motility, when $p_0 < p_0^*$ it can be melted only if cells have a self propulsion speed that allows to overcome T1 energy barriers. In the presence of alignment, one needs to distinguish between the velocity component parallel to the average velocity, and the component perpendicular to it; it is only the latter that increases the probability of a cell to overcome the energy barrier and undergo a rearrangement, which implies that for increasing J there is less kinetic energy available for melting and the tissue remains solid unless more energy is provided via increase of v_0 .

The conclusions of this model are in agreement with experimental results obtained in experiments on epithelial monolayers [13]: Overexpression of the endocytic protein RAB5A in a solid tissue was found to liquefy the tissue while at the same time causing directed collective motion; both features are compatible with the flocking liquid state predicted by the self-propelled Voronoi model with alignment in the regime where both the polar interaction strength and the target perimeter p_0 are large. This novel mode of collective motility, called unjamming-via-flocking was later shown to play a key role also in three dimensions and in more complex environments with physiological and clinical relevance [14], pointing to a key role of unjamming-via-flocking in cancer dissemination and metastasis.

To put these results in the more general context of active matter, it is worth mentioning that the flocking transition in the self-propelled Voronoi model of Ref. [9] appears to be smooth and continuous. This property matches the behavior of Vicsek models with topological interactions and contrasts with those having metric

interaction, which are known to exhibit a first-order transition [32]. This result is consistent with the fact that cells in the self-propelled Voronoi model interact with topologically neighboring cells.

14.4 Final Remarks

In this Chapter, we have summarized the main features and results of some two-dimensional active network models developed to capture the properties of cell monolayers. While we have restricted our description to the two-dimensional case, it is worth mentioning that three-dimensional extensions are also possible and successful [21, 33]. These network models, however simplified, provide a natural framework to describe the structure and dynamics of simple confluent tissues in terms of the geometrical properties of the constituent cells.

14.5 Problems

Problem 14.1: Phase lines of the vertex model. The energy of a polygonal tessellation representing a cell monolayer can be expressed by using Eq. (14.17). Consider a single isolated cell with n sides.

a. Use a scaling factor s to change the linear size of the cell and obtain the following ground states:

- i) $p = p_0$ and $\alpha = 1$, which is the absolute minimum ϵ_0 of the cell energy;
- ii) $p = 0$ and $\alpha = 0$, which corresponds to an unstable case with energy $\epsilon_0 + \frac{1}{2}(1 + p_0^2)$;
- iii) $p < p_0$ and $\alpha > 1$, where, for a fixed area, the cell energy can be decreased by increasing the perimeter;
- iv) $p > p_0$ and $\alpha < 1$, where, for a fixed area decrease, the cell energy can be decreased by decreasing the perimeter, which is satisfied by a regular polygon (local minimum).

b. For case (i) find the minimum value of the perimeter p_0 that makes it possible to satisfy simultaneously the conditions $p = p_0$ and $\alpha = 1$ and use it to show that this solution is meaningful only when

$$\bar{\Lambda}^{\text{el}} \leq -4\bar{\Gamma}^{\text{el}} \sqrt{n \tan(\pi/n)}. \quad (14.22)$$

c. Show that, for $n = 6$, Eq. (14.22) describes the line separating the cyan and gray regions (valid for a tissue) in Fig. 14.6.

Problem 14.2: Phase line of the motility induced solid-to-liquid transition in confluent tissues. Use the fact that the mean energy barriers $\Delta\epsilon$ for T1 transitions scales as $\Delta\epsilon \propto p_0^* - p_0$ when the target shape index p_0 approaches its critical value $p_0^* = 3.813$ from the solid side to show that the phase boundary line in Fig. 14.11a is described by the equation

$$p_c(v_0) = p_0^* - \frac{v_0^2}{\mu(\mu k + D_r)}. \quad (14.23)$$

[Hint: Calculate the mean square displacement of a motile cell tethered to a spring of force constant k in the presence of orientational noise and compare it to the corresponding result for a tethered Brownian particle. What can you learn from such comparison?]

Bibliography

- [1] X. Trepat and E. Sahai. Mesoscale physical principles of collective cell organization. *Nature Physics*, 14:671–682, 2018.
- [2] B. A. Camley and W.-J. Rappel. Physical models of collective cell motility: from cell to tissue. *Journal of Physics D: Applied Physics*, 50:113002, 2017.
- [3] R. Alert and X. Trepat. Physical models of collective cell migration. *Annual Review of Condensed Matter Physics*, 11:77–101, 2020.

- [4] D. Sánchez-Gutiérrez, M. Tozluoglu, J. D. Barry, A. Pascual, Y. Mao, and L. M. Escudero. Fundamental physical cellular constraints drive self-organization of tissues. *The EMBO journal*, 35:77–88, 2016.
- [5] S. Kaliman, C. Jayachandran, F. Rehfeldt, and A.-S. Smith. Limits of applicability of the voronoi tessellation determined by centers of cell nuclei to epithelium morphology. *Frontiers in Physiology*, 7:551, 2016.
- [6] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17:2095–2104, 2007.
- [7] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning. Motility-driven glass and jamming transitions in biological tissues. *Physical Review X*, 6:021011, 2016.
- [8] D. L. Barton, S. Henkes, C. J. Weijer, and R. Sknepnek. Active vertex model for cell-resolution description of epithelial tissue mechanics. *PLoS Computational Biology*, 13:e1005569, 2017.
- [9] F. Giavazzi, M. Paoluzzi, M. Macchi, D. Bi, G. Scita, M. L. Manning, R. Cerbino, and M. C. Marchetti. Flocking transitions in confluent tissues. *Soft Matter*, 14:3471–3477, 2018.
- [10] E. Lång, A. Połec, A. Lång, M. Valk, P. Blicher, A. D. Rowe, K. A. Tønseth, C. J. Jackson, T. P. Utheim, L. M. C. Janssen, J. Eriksson, and S. O. Bøe. Coordinated collective migration and asymmetric cell division in confluent human keratinocytes without wounding. *Nature Communications*, 9:1–15, 2018.
- [11] V. Petrolli, M. Le Goff, M. Tadrous, K. Martens, C. Allier, O. Mandula, L. Hervé, S. Henkes, R. Sknepnek, T. Boudou, G. Cappello, and M. Balland. Confinement-induced transition between wavelike collective cell migration modes. *Physical review letters*, 122:168101, 2019.
- [12] J.-A. Park, J. H. Kim, D. Bi, J. A. Mitchel, N. T. Qazvini, K. Tantisira, C. Y. Park, M. McGill, S.-H. Kim, B. Gweon, J. Notbohm, R. Jr Steward, S. Burger, S. H. Randell, A. T. Kho, C. Tambe, D. T. Hardin, S. A. Shore, E. Israel, D. A. Weitz, D. J. Tschumperlin, E. P. Henske, S. T. Weiss, M. L. Manning, J. P. Butler, J. M. Drazen, and J. J. Fredberg. Unjamming and cell shape in the asthmatic airway epithelium. *Nature materials*, 14:1040–1048, 2015.
- [13] C. Malinverno, S. Corallino, F. Giavazzi, M. Bergert, Q. Li, M. Leoni, A. Disanza, E. Frittoli, A. Oldani, E. Martini, T. Lendenmann, G. Deflorian, G. V. Beznoussenko, D. Poulikakos, K. H. Ong, M. Uroz, X. Trepat, D. Parazzoli, P. Maiuri, W. Yu, A. Ferrari, R. Cerbino, and G. Scita. Endocytic reawakening of motility in jammed epithelia. *Nature materials*, 16(5):587–596, 2017.
- [14] A. Palamidessi, C. Malinverno, E. Frittoli, S. Corallino, E. Barbieri, S. Sigismund, G. V. Beznoussenko, E. Martini, M. Garre, I. Ferrara, C. Tripodo, F. Ascione, E. A. Cavalcanti-Adam, Q. Li, P. P. Di Fiore, D. Parazzoli, F. Giavazzi, R. Cerbino, and G. Scita. Unjamming overcomes kinetic and proliferation arrest in terminally differentiated cells and promotes collective motility of carcinoma. *Nature materials*, 18:1252–1263, 2019.
- [15] E. Evans and A. Yeung. Apparent viscosity and cortical tension of blood granulocytes determined by micropipet aspiration. *Biophysical journal*, 56(1):151–160, 1989.
- [16] R. M. Hochmuth. Micropipette aspiration of living cells. *Journal of Biomechanics*, 33(1):15–22, 2000.
- [17] A. Šiber and P. Zicherl. *Cellular patterns*. CRC Press, 2017.
- [18] B. Smeets, M. Cuvelier, J. Pešek, and H. Ramon. The effect of cortical elasticity and active tension on cell adhesion mechanics. *Biophysical journal*, 116:930–937, 2019.
- [19] H. Berthoumieux, J.-L. Maître, C.-P. Heisenberg, E. K. Paluch, F. Jülicher, and G. Salbreux. Active elastic thin shell theory for cellular deformations. *New Journal of Physics*, 16:065005, 2014.
- [20] M. L. Manning, R. A. Foty, M. S. Steinberg, and E.-M. Schoetz. Coaction of intercellular adhesion and cortical tension specifies tissue surface tension. *Proceedings of the National Academy of Sciences*, 107:12517–12522, 2010.

- [21] S. Alt, P. Ganguly, and G. Salbreux. Vertex models: from cell mechanics to tissue morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372:20150520, 2017.
- [22] D. B. Staple, R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. Mechanics and remodelling of cell packings in epithelia. *The European Physical Journal E*, 33:117–127, 2010.
- [23] N. S. Gov. Traction forces during collective cell motion. *HFSP Journal*, 3:223–227, 2009.
- [24] D. M. Sussman and M. Merkel. No unjamming transition in a voronoi model of biological tissue. *Soft matter*, 14:3397–3403, 2018.
- [25] F. Graner and D. Riveline. ‘The forms of tissues, or cell-aggregates’: D’Arcy Thompson’s influence and its limits. *Development*, 144:4226–4237, 2017.
- [26] H. Honda. Description of cellular patterns by Dirichlet domains: the two-dimensional case. *Journal of Theoretical Biology*, 72:523–543, 1978.
- [27] H. Honda and G. Eguchi. How much does the cell boundary contract in a monolayered cell sheet? *Journal of Theoretical Biology*, 84:575–588, 1980.
- [28] T. Nagai and H. Honda. A dynamic cell model for the formation of epithelial tissues. *Philosophical Magazine B*, 81:699–719, 2001.
- [29] H. Honda and T. Nagai. Cell models lead to understanding of multi-cellular morphogenesis consisting of successive self-construction of cells. *The Journal of Biochemistry*, 157(3):129–136, 2015.
- [30] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning. A density-independent rigidity transition in biological tissues. *Nature Physics*, 11:1074–1079, 2015.
- [31] P. Sahu, J. Kang, G. Erdemci-Tandogan, and M. L. Manning. Linear and nonlinear mechanical responses can be quite different in models for biological tissues. *Soft Matter*, 16:1850–1856, 2020.
- [32] F. Ginelli. The physics of the Vicsek model. *The European Physical Journal Special Topics*, 225:2099–2117, 2016.
- [33] M. Merkel and M. L. Manning. A geometrically controlled rigidity transition in a model for confluent 3D tissues. *New Journal of Physics*, 20:022002, 2018.

Part II

Simulation

Chapter 15

Numerical Models for Active Matter

LUKAS HECHT, JOSÉ CARLOS UREÑA MARCOS, BENNO LIEBCHEN

Numerical models play a crucial role in our understanding of active matter. In this part of the Book, we will explore the main ones, going from *dry models* (involving only equations of motion for the active particles) to *wet models* (explicitly including a solvent in addition to the embedded active particles, which ensures momentum conservation) to *continuum descriptions* for systems consisting of many active particles (Fig. 15.1). Dry models are naturally used to describe active systems that do not involve a liquid solvent, such as granular particles on vibrating plates [1, 2, 3, 4, 5, 6], self-vibrating granular particles [7, 8, 9], bacteria gliding on a rigid surface [10],

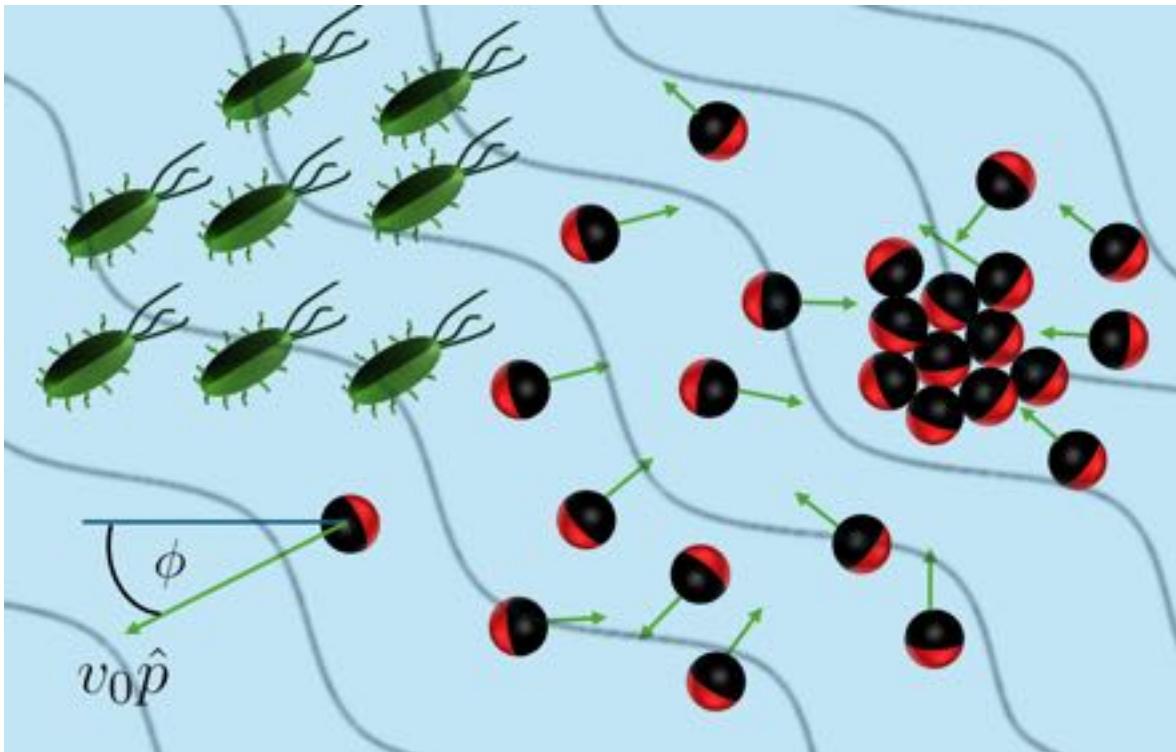


Figure 15.1: **Modeling the dynamics of active particles.** The collective dynamics of self-propelled particles, like active Janus colloids and E-coli bacteria, that are schematically shown in the figure, is often described by minimal models. In general, we distinguish between *dry models*, that involve equations of motion for the particles only and treat the solvent implicitly, and *wet models*, that explicitly couple the particle dynamics to the dynamics of the surrounding solvent.

flocks of birds, animal herds, swarms of locusts [11, 12, 13, 14, 15], human crowds [16, 17], autonomous robots [18, 19], and flying drones [20]. Furthermore, they are also frequently used as simplified descriptions of active matter systems involving a solvent, which is only represented in an effective way and commonly acts as a thermal bath leading to fluctuations in the equations of motion of the individual particles [21]. In contrast, wet models are used to describe microswimmers such as synthetic active colloids [22, 23, 24], droplet swimmers [25, 26, 27, 28], and biological microorganisms like bacteria [29, 10, 30], algae [31], or sperm cells [29], including their interaction with the surrounding solvent and the cross-interactions among different microswimmers. A particular example of a wet system at larger scales, i.e., beyond the soft-matter realm, which microswimmers belong to, can be found in schools of fish [13, 32]. Continuum descriptions significantly reduce the number of degrees of freedom necessary to describe the dynamics. This is specially relevant to study the emergence of patterns with characteristic length and time scales that are much larger than the ones for the individual particles [33, 34, 35]. In this Chapter, we will provide an overview of these models, which will be then discussed in detail in the following Chapters of this part of the Book.

Agnese: Finalize revision of bibliography.

15.1 The effect of the solvent

To understand the applicability regime of the various models for active matter, it is instructive to first discuss the impact of the solvent on active systems:

Fluctuations and dissipation. Active particles are typically orders of magnitude larger than the molecules of the surrounding solvent and are subject to collisions with the latter. This leads to fluctuations in their motion, analogously to *Brownian motion* (or *translational diffusion*) of passive colloids in equilibrium, such as the trajectories shown in Fig. 15.2a. The trajectory of an isolated active particle is then typically given by the combination of ballistic motion due to self-propulsion and fluctuations due to collisions with the solvent molecules. As shown in Fig. 15.2b, the motion of an active particle is not straight because the collisions of the solvent molecules with the active particles feature both a radial and a tangential component. The latter induces a stochastic turn of the particle orientation and, hence, reorientation of the self-propulsion direction (*rotational Brownian motion* or *rotational diffusion*). As established by the fluctuation–dissipation theorem, these fluctuations are necessarily linked to dissipation occurring, e.g., in the form of Stokes drag for spherical particles. For microswimmers, i.e., for (active) particles at the microscale, *dissipation* normally dominates over inertia; hence, the motion is overdamped. These effects of the solvent, namely translational diffusion, rotational diffusion, and dissipation, are the only effects of

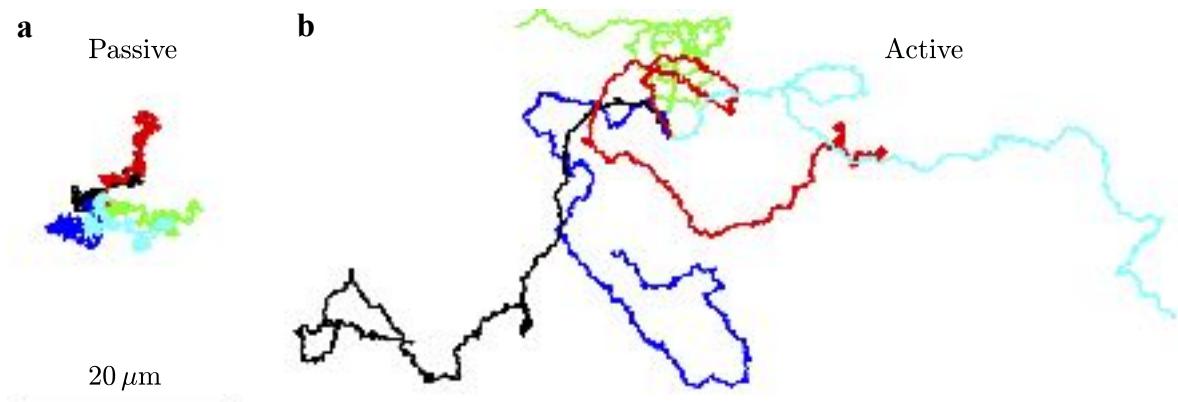


Figure 15.2: **Passive and active particles.** Experimental trajectories of a Janus colloid performing either (a) passive or (b) active Brownian motion. The Janus colloids are platinum-coated spherical particles with $1.62\text{ }\mu\text{m}$ diameter, which behave as passive particles when suspended in pure water. To make them active, a solution of water and hydrogen peroxide (H_2O_2) is used (concentration 20%). The duration of each acquisition is 25 s. Reproduced from Ref. [22].

the solvent typically taken into account in *dry models*, such as the active Brownian particle (ABP) model, which we will discuss further below.

Momentum conservation. Physically, in the absence of external fields or boundaries, the overall momentum of an active system has to be conserved. For example, when a microorganism or an active Janus colloid moves forward, there is necessarily a counter-propagating solvent flow such that the overall momentum of the active particle and the surrounding solvent is conserved. Thus, the solvent not only acts as a bath providing fluctuations and drag, but also ensures momentum conservation and allows self-propulsion — swimming in vacuum is impossible.

Hydrodynamic interactions. The solvent mediates hydrodynamic interactions among active particles. These arise because the flow pattern induced by each active particle as a consequence of its swimming acts onto all other particles in the system. These solvent-mediated interactions are often long-ranged. In particular, in the absence of external forces, they often decay as Δr^{-2} (where Δr is the distance from the swimmer) for force-dipole swimmers, such as various bacteria or algae (explicit measurements of the flow field exist, e.g., for *E. coli* bacteria [36]), and as Δr^{-3} for source-dipole swimmers, such as *Paramecium* [37] or (idealized) Janus colloids with a uniform surface mobility [38, 39]. However, they can be weakened or decay faster in the presence of a substrate or other boundaries [40].

Hydrodynamic boundary effects. If the active particles are in contact with boundaries, such as a glass substrate, which is frequently used in experiments with active colloids, or another liquid–solid or liquid–air interface, the solvent can lead to additional interesting effects. An example of these is constituted by *osmotic flows* at fluid–solid interfaces, such as those induced by auto-phoretic particles [41] or by some modular swimmers involving ion-resins [42, 43]. At fluid–air interfaces, active particles can cause *Marangoni flows* [44, 45, 46, 47], which act on all particles in the system and can elicit interesting collective behaviors [48, 49, 50, 51].

15.2 Dry active particles

In this section, we describe some of the most commonly employed dry models for active matter at the single-particle level. Some example trajectories are shown in Fig. 15.3. An overview of these models is provided in Tab. 15.2.

15.2.1 Active Brownian particle

One of the simplest and most popular models to describe active particles is the *active Brownian particle* (ABP) model [54, 55, 56, 52, 57, 58, 59, 21], originally introduced to describe the motion of colloidal particles that smoothly change their self-propulsion direction due to rotational diffusion. It treats the solvent as a bath providing only fluctuations and drag without ensuring momentum conservation and, at least in its most commonly used form, without accounting for hydrodynamic interactions among particles. The ABP model does not explicitly describe the mechanism leading to self-propulsion (which arises through the interactions of the active particles with the surrounding solvent or with a substrate), but simply replaces it with an effective force that drives the particle forward. Microscopically, this is not correct because, as discussed above, microswimmers are force free, but it leads to a simple generic model for the dynamics of active particles, agnostic to many details of the specific underlying realization. In its simplest form, the ABP model in two-dimensional space is defined by the overdamped Langevin equations

$$\frac{d\mathbf{r}_i(t)}{dt} = v_0 \hat{\mathbf{p}}_i(t) - \frac{1}{\gamma} \nabla_{\mathbf{r}_i} U + \sqrt{2D} \boldsymbol{\xi}_i(t), \quad (15.1)$$

$$\frac{d\theta_i(t)}{dt} = \sqrt{2D_R} \eta_i(t), \quad (15.2)$$

where $\mathbf{r}_i = (r_{x,i}, r_{y,i})$ and θ_i are the position and the orientation angle of the i -th spherical ABP, respectively, v_0 is the self-propulsion speed, γ is the Stokes drag coefficient, $U = \sum_{i < j} u(\Delta r_{ij})$ is the interaction energy with interaction potential $u(\Delta r_{ij})$, $\Delta r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the pairwise distance, D and D_R are the translational and

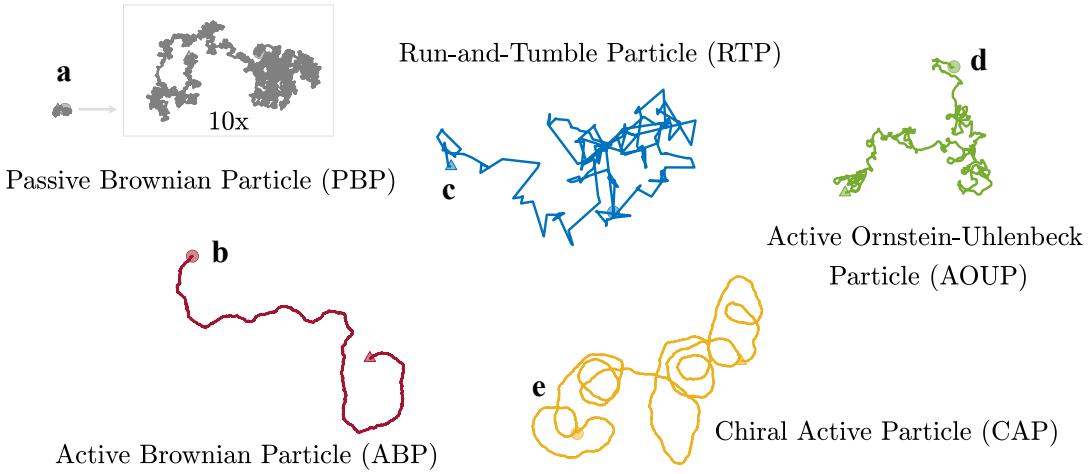


Figure 15.3: Dry models. Example trajectories obtained in the overdamped regime from simulations of (a) the passive Brownian particle (PBP) model, (b) the active Brownian particle (ABP) model, (c) the run-and-tumble particle (RTP) model, (d) the active Ornstein-Uhlenbeck particle (AOUP) model, and (e) the chiral active particle (CAP) model. The trajectory of the passive Brownian particle is also shown magnified 10 times in size in order to highlight its features. For each trajectory, the initial particle position is indicated with a triangle, the final position with a circle.

rotational diffusion coefficients, respectively, and $\xi_i(t) = (\xi_{x,i}(t), \xi_{y,i}(t))$ and $\eta_i(t)$ represent Gaussian white noise with unit variance and zero mean. The self-propulsion direction is given by $\hat{\mathbf{p}}_i = (\cos \theta_i, \sin \theta_i)$.

It is instructive to first consider a single ABP in the absence of external forces, i.e., $U = 0$. Different from the trajectory of a *passive Brownian particle* (PBP) shown in Fig. 15.3a, the trajectory of an ABP is characterized by an initial period of directed motion followed by a randomization of the self-propulsion direction due to rotational diffusion, as shown in Fig. 15.3b. The initial directed motion persists for a time $\tau_p = 1/D_R$ (*persistence time*) over a distance $l_p = v_0 \tau_p$ (*persistence length*). Accordingly, the mean trajectory of an ABP is given by [56, 52, 57]

$$\langle \mathbf{r}_i(t) - \mathbf{r}_i(0) \rangle = l_p \left(1 - e^{-t/\tau_p} \right) \hat{\mathbf{p}}_i(0). \quad (15.3)$$

Therefore, an ABP moves, on average, over a distance l_p along its initial orientation $\hat{\mathbf{p}}_i(0)$ before its orientation is randomized, which rationalizes the term “persistence length”. Additionally, the mean square displacement (MSD) of an ABP (in two-dimensional space) reads [22, 56, 52, 57]

$$\langle (\mathbf{r}_i(t) - \mathbf{r}_i(0))^2 \rangle = 2l_p^2 \left(\frac{t}{\tau_p} - 1 + e^{-t/\tau_p} \right) + 4Dt, \quad (15.4)$$

which provides valuable insights into the different dynamical regimes of the ABP model. Three regimes are observed when expanding Eq. (15.4) for short, intermediate, and long times. The motion of an ABP is initially diffusive with diffusion coefficient D for $t \ll D/v_0^2$. For $D/v_0^2 \ll t \ll \tau_p$, a ballistic regime which represents directed motion due to the activity of the particle comes about. Finally, for $t \gg \tau_p$, the motion is again diffusive with effective diffusion coefficient $D_A = D + l_p^2/(2\tau_p)$. These three regimes are shown by the blue line in Fig. 15.4.

The relative importance of activity in comparison with diffusion can be characterized by the *Péclet number* $\text{Pe} = v_0/\sqrt{2DD_R}$ [52]. Assuming spherical active particles of radius R , the Stokes-Einstein relation yields $D = \frac{4}{3}R^2D_R$, and thus the Péclet number can be rewritten in terms of the particle diameter $2R$ (as customarily done in the literature [62, 63, 64]) as $\text{Pe} = \sqrt{3/2}v_0/(2RD_R)$ or $\text{Pe} = v_0 \cdot (2R)/(\sqrt{6}D)$.

For sufficiently dense ensembles of active particles with a sufficiently large Péclet number and purely repulsive interactions arising because, e.g., the individual particles cannot overlap, the ABP model predicts a

Model	Equations of motion	Parameters	Natural units
ABP	$\dot{\mathbf{r}}(t) = \hat{\mathbf{p}}(t) + Pe^{-1} \xi(t)$ $\dot{\theta}(t) = \sqrt{2}\eta(t)$	$Pe = \frac{v_0}{\sqrt{2DD_R}}$	Time scale: $\tau_p = D_R^{-1}$ Length scale: $l = l_p = v_0 D_R^{-1}$
RTP	$\dot{\mathbf{r}}(t) = \hat{\mathbf{p}}(t)$ $\dot{\theta}(t) = \sum_n \Delta\theta_n \delta(t - \tilde{T}_n)$	None (*)	Time scale: $\tau_p = \lambda_t^{-1}$ Length scale: $l = l_p = v_0 \lambda_t^{-1}$
AOUP	$\dot{\mathbf{r}}(t) = \mathbf{v}_0(t)$ $\dot{\mathbf{v}}_0(t) = -\mathbf{v}_0(t) + \sqrt{2}\xi(t)$	None	Time scale: τ_p Length scale: $l = \sqrt{D\tau_p}$
CAP	$\dot{\mathbf{r}}(t) = \hat{\mathbf{p}}(t) + Pe^{-1} \xi(t)$ $\dot{\theta}(t) = \tilde{\omega} + \sqrt{2}\eta(t)$	$Pe = \frac{v_0}{\sqrt{2DD_R}}$ $\tilde{\omega} = \omega \tau_p$	Time scale: $\tau_p = D_R^{-1}$ Length scale: $l = l_p = v_0 D_R^{-1}$

Table 15.1: **Dry active matter models.** Dimensionless equations of motion and parameters of a single active particle in the active Brownian particle (ABP), run-and-tumble particle (RTP), active Ornstein-Uhlenbeck particle (AOUP), and chiral active particle (CAP) models. The dimensionless variables in the equations of motion are obtained by rescaling the original dimensional variables by their natural time and length scales, i.e., $\mathbf{r} = \mathbf{r}/l$, $\mathbf{v}_0 = \mathbf{v}_0 \tau_p/l$ and $t = t/\tau_p$. The dot over these variables denotes the derivative with respect to the dimensionless time t . Pe is the Péclet number [52]. (*) Note for the RTP particle: For equally spaced T_n or Poisson-distributed tumbling events as found in *E. coli* [53] and without translational diffusion; with the latter, the equation of motion for the position reads $\dot{\mathbf{r}}(t) = \hat{\mathbf{p}}(t) + Pe^{-1} \xi(t)$ with $Pe = v_0/\sqrt{2D\lambda_t}$.

spectacular phenomenon known as *motility-induced phase separation* (MIPS) [65, 24, 66, 67, 57, 68]. A sequence of snapshots of the state of an ensemble of ABPs that interact via the purely repulsive Weeks-Chandler-Anderson (WCA) potential [69] and for which MIPS occurs is shown in Fig. 15.5a-d. Initially, the ensemble is uniformly distributed. For suitable parameters (large Péclet number and high density), the uniform state loses stability and the particles aggregate in small clusters. These clusters grow following the coarsening law shown in Fig. 15.5e until a single macrocluster, which coexists with a low-density active gas, is eventually formed. Overall, while phase separation in equilibrium generally requires inter-particle attractions, active systems can phase separate even in their complete absence [70, 67]. The mechanism underlying MIPS is shown in Fig. 15.5f: When particles collide, they block each other until their orientations are randomized and they can separate from each other. Broadly, MIPS occurs if the active particles are fast and numerous enough for collisions with existing clusters to occur more often than particles in these clusters leave them due to rotational diffusion.

The ABP model can also be generalized to account for inertial effects [57, 71, 72, 73, 74], which is used, e.g., to model active granular particles on vibrating plates [1, 2].

Several alternative models have been designed that have a similar scope as the ABP model in the sense that they also treat the solvent as a bath that only provides fluctuations and drag rather than accounting for momentum conservation and hydrodynamic interactions. We will present them in the following sections.

Example 15.1: Nondimensionalization of the Langevin equations. Transform the Langevin equations for a single overdamped ABP (Eqs. (15.1) and (15.2) with $U = 0$) into dimensionless form, as in Tab. 15.2.

Solution: First, we define the dimensionless variables $\mathbf{r} = \mathbf{r}/l_p$ and $t = t/\tau_p$. Eqs. (15.1) and (15.2)

then read

$$\frac{l_p}{\tau_p} \frac{d\mathbf{r}(t)}{dt} = v_0 \hat{\mathbf{p}}(t) + \sqrt{2D} \boldsymbol{\xi}(\tau_p t), \quad (15.5)$$

$$\frac{1}{\tau_p} \frac{d\theta(t)}{dt} = \sqrt{2D_R} \eta(\tau_p t). \quad (15.6)$$

Since the translational and rotational noises are delta-correlated and $\delta(\tau_p t) = \frac{1}{|\tau_p|} \delta(t)$, both $\boldsymbol{\xi}$ and η fulfill $\boldsymbol{\xi}(\tau_p t) = \frac{1}{\sqrt{\tau_p}} \boldsymbol{\xi}(t)$. Now defining $\tau_p = D_R^{-1}$, $l_p = v_0 D_R^{-1}$ and $\text{Pe} = \frac{v_0}{\sqrt{2D D_R}}$, the dimensionless equations in Tab. 15.2 are obtained.

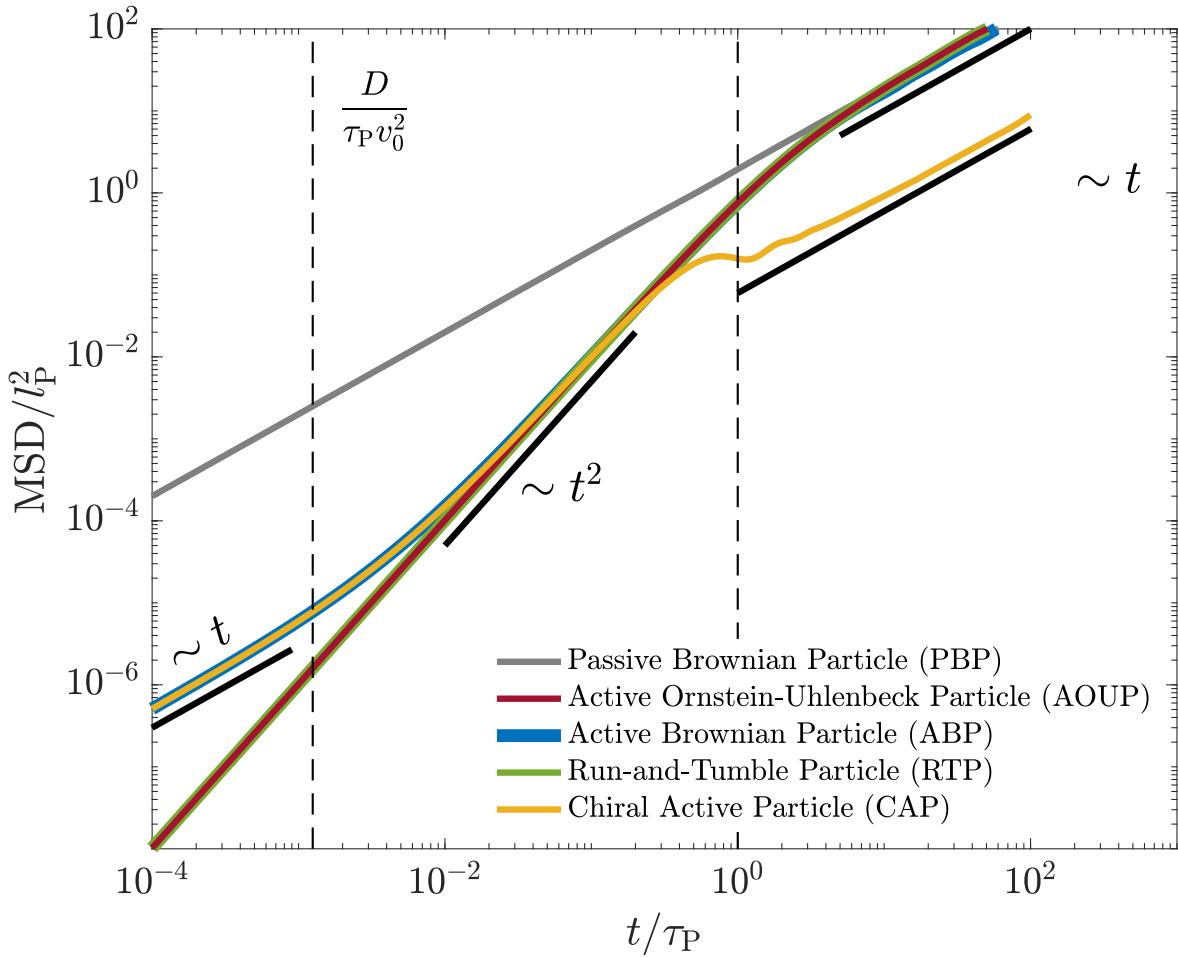


Figure 15.4: **Mean square displacement.** Mean square displacement (MSD) over time in the ABP, RTP, AOUP, and CAP models calculated based on a $10^3 \tau_p$ -long trajectory of the dimensionless equations in Tab. 15.2 (simulation parameters: time step $\Delta t/\tau_p = 10^{-4}$, $\text{Pe} = 20$ for the ABP and CAP models, $\tilde{\omega} = 5$ for the CAP model, and probability that a tumbling event occurs during a certain time step given by a Poisson distribution with rate $\lambda_p = \lambda_t \Delta t = 10^{-3}$ for the RTP model). The dashed vertical lines are placed at $t/\tau_p = D/(v_0^2 \tau_p)$ (left) and $t/\tau_p = 1$ (right) separating the different regimes exhibited by the particles in each model. The reader is referred to Refs. [60, 61] for more details on the MSD of the CAP model.

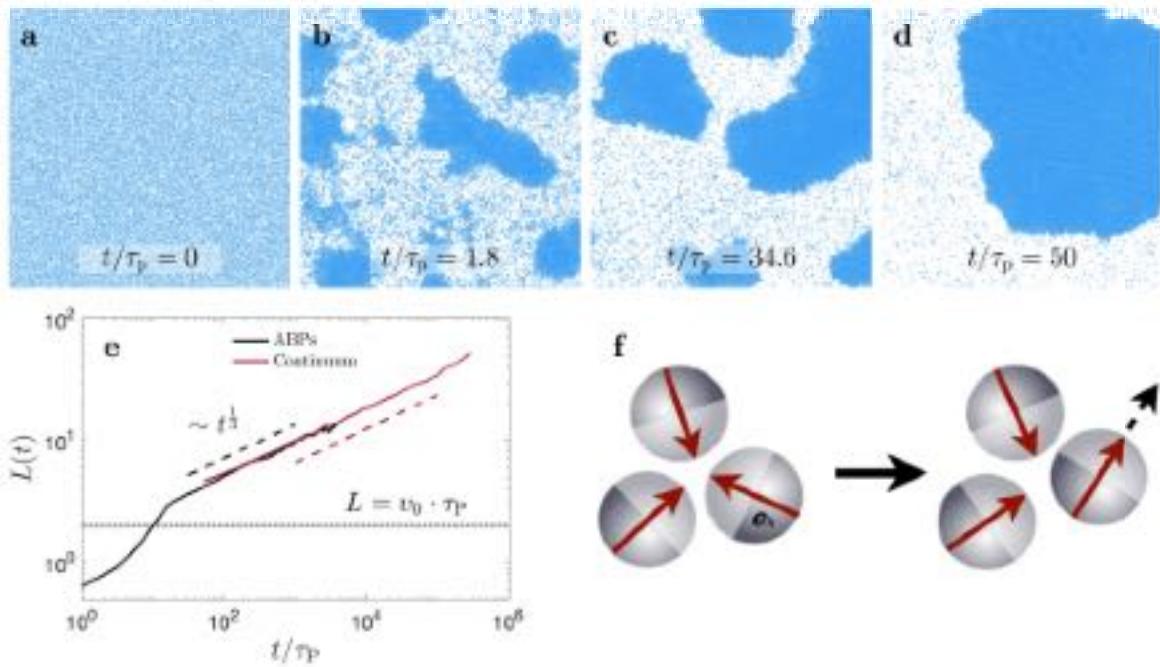


Figure 15.5: **Motility-induced phase separation.** (a-d) Evolution of motility-induced phase separation (MIPS) for an ensemble of $N = 39200$ ABPs of diameter $2R$ with number density $\rho(2R)^2 = 0.6$, Péclet number $\text{Pe} = 200/\sqrt{2}$, and time step $\Delta t/\tau_p = 5 \times 10^{-6}$. The ABPs interact via the purely repulsive Weeks-Chandler-Anderson (WCA) potential [69] with strength $\epsilon = 10k_B T$. (e) Time evolution of the mean cluster size $L(t)$ for the ABP model. Reproduced from Ref. [62]. (f) Schematic of the mechanism which leads to MIPS. Reproduced from Ref. [24].

15.2.2 Run-and-tumble model

The *run-and-tumble particle* (RTP) model [75, 76, 77, 78] was originally introduced to describe the characteristic motion patterns of certain bacteria such as *E. coli* [79, 80, 81], but it has now advanced to a standard model for the description of active particles.¹ In contrast to ABPs, RTPs alternate running periods, during which the self-propulsion direction remains unchanged, with idealized tumbling events, in which the orientation of the particles is randomized, as the trajectory shown in Fig. 15.3c. The equations of motion for the i -th RTP read

$$\frac{d\mathbf{r}_i(t)}{dt} = v_0 \hat{\mathbf{p}}_i(t) - \frac{1}{\gamma} \nabla_{\mathbf{r}_i} U, \quad (15.7)$$

$$\frac{d\theta_i(t)}{dt} = \sum_n \Delta\theta_n \delta(t - T_n), \quad (15.8)$$

where the parts and parameters of Eq. (15.7) are defined as in Eq. (15.1). The values of $\Delta\theta_n$ are uniformly distributed between 0 and 2π , with tumbling events taking place at discrete times T_n [78]. In practice, the times T_n are chosen either equally timed or randomly timed with $\langle T_{n+1} - T_n \rangle = \lambda_t^{-1}$ (i.e., tumbling events following a Poisson distribution, which leads to exponentially distributed times between tumbling events, as originally found for *E. coli* [53]). In any case, the (mean) tumbling rate λ_t is fixed, yielding a persistence time $\tau_p = 1/\lambda_t$, which plays the role of the (mean) time between tumbling events.

Remarkably, the many-particle dynamics following from the RTP and the ABP models turn out to be equivalent at coarse-grained scales if $(d-1)D_R = \lambda_t$, where $d > 1$ is the spatial dimension [76, 77]. At the single-particle level, this is also reflected by the MSD in Fig. 15.4 showing almost identical time evolution for RTPs and ABPs, with the exception that RTPs always feature a ballistic behavior for $t \ll \tau_p$, since tumbling events are statistically unlikely on this timescale and translational diffusion is not considered. The latter can also be taken into account, resulting in the emergence of a diffusive regime for $t \gg D/v_0^2$, as in the ABP model, where D is the translational diffusion coefficient.

15.2.3 Active Ornstein-Uhlenbeck model

Another alternative to the ABP model is the *active Ornstein-Uhlenbeck particle* (AOUP) model [82, 83, 84, 85], which has certain advantages compared with the ABP model regarding the theoretical description of the many-body dynamics of dry active particles. This is due to the fact that the AOUP model avoids the strongly nonlinear dependence of the center-of-mass motion on the particle orientation, which is present in the ABP model (see Eqs. (15.1) and (15.2)), by using colored noise to generate self-propulsion. The equation of motion for particle i in the AOUP model (in the overdamped regime) reads

$$\frac{d\mathbf{r}_i(t)}{dt} = \mathbf{v}_{0,i}(t) - \frac{1}{\gamma} \nabla_{\mathbf{r}_i} U, \quad (15.9)$$

where γ is the Stokes drag coefficient and U is the total interaction potential. Whereas the self-propulsion speed v_0 remains constant for a single particle in the ABP and RTP models, it evolves with time in the AOUP model according to

$$\tau_p \frac{d\mathbf{v}_{0,i}(t)}{dt} = -\mathbf{v}_{0,i}(t) + \sqrt{2D} \xi_i(t), \quad (15.10)$$

where τ_p is the persistence time and $\xi_i(t)$ is Gaussian white noise with unit variance and zero mean. As a result, the velocity components of an isolated AOUP are represented by colored Gaussian noise² with correlation function $\langle v_{0,i}^{(\alpha)}(0)v_{0,j}^{(\beta)}(t) \rangle = \delta_{ij}\delta_{\alpha\beta}(D/\tau_p)e^{-t/\tau_p}$ between components α and β of particles i and j . Interestingly, a single particle in the AOUP model without external forces (i.e., $U = 0$) is formally identical to an *underdamped* passive Brownian particle with $m/\gamma = \tau_p$, as shown by the trajectory in Fig. 15.3d. Hence, a single AOUP shows a ballistic regime for $t \ll \tau_p$ followed by a diffusive regime for $t \gg \tau_p$, as shown by the MSD in Fig. 15.4.

Since it involves colored noise, the AOUP model does not permit the formulation of an exact Fokker-Planck equation for the corresponding probability density function (PDF). However, it is still possible to derive

¹In fact, the first theory for MIPS was formulated for RTPs [75] and MIPS has been observed in simulations of this model as well [76].

²Whereas white noise is delta-correlated in time, the correlation function of colored noise takes finite values for finite time differences.

an approximate Fokker-Planck equation for the many-body dynamics, which does not depend on the particle orientations but only on the particle positions [84, 85, 82, 86] and can be solved to obtain the stationary PDF of the latter [87]. MIPS has also been reported for the AOP model [83, 85] suggesting that it provides a useful alternative for the description of the many-body dynamics of active particles although the single-particle properties significantly differ from those of the ABP and RTP model.

15.2.4 Chiral particle model

A further class of models describes *chiral active particles* (CAPs) [57, 88, 60, 89, 90, 61, 91, 92, 93, 94], which experience an additional effective torque arising from an anisotropy in their shape or propulsion mechanism. For an isolated CAP, this leads to circular trajectories in the limit of zero noise, whereas the orientation angle of the i -th CAP in the presence of noise evolves according to

$$\frac{d\theta_i(t)}{dt} = \omega + \sqrt{2D_R} \eta_i(t), \quad (15.11)$$

where ω is a constant angular velocity. As in the ABP model, the position \mathbf{r}_i of the i -th CAP generally evolves with time according to Eq. (15.1). An exemplary trajectory of a CAP and the time evolution of its MSD are shown in Figs. 15.3e and 15.4, respectively. Examples of circle swimmers include *E. coli* bacteria near surfaces and interfaces [95, 96], sperm cells [97, 98] and artificial microswimmers such as L-shaped particles [99], spherical-cap particles near a substrate [100], and asymmetric Quincke rollers [101].

15.2.5 Models with alignment interactions

So far, we have focused on particles without alignment interactions. We will now consider models with such interactions. The most popular model for describing (dry) active particles with (polar) alignment interactions is the *Vicsek model* [102, 103], which accounts for self-propelled particles that align their orientation with that of their neighbors. A generalized continuous-time variant of the Vicsek model comprising CAPs with alignment interactions can be defined by Eq. (15.1) with $U = 0$ for the translational dynamics and

$$\frac{d\theta_i(t)}{dt} = \omega + \frac{K}{\pi R_0^2} \sum_{j \in S_{R_0}^{(i)}} \sin(\theta_j - \theta_i) + \sqrt{2D_R} \eta_i(t), \quad (15.12)$$

where ω is an angular velocity, K is the strength of the alignment interactions (for $K = 0$ this model reduces to the CAP model) and the sum is calculated over all particles within a circle $S_{R_0}^{(i)}$ of radius R_0 centered at the position of particle i [104, 105, 106, 88]. The hallmark of this model is that particles tend to follow the orientation of their neighbors, which can induce polar order, e.g., in the form of the traveling bands shown in Fig. 15.6a for $\omega = 0$. When considering CAPs with polar interactions (Eq. (15.12) with $\omega > 0$), one finds two remarkable phenomena: The formation of rotating macrodroplets with late-time sizes comparable to the system size, which is indicative of phase separation, at low angular velocity (Fig. 15.6b), and a pattern of rotating microflocks exhibiting phase synchronization and a self-selected length scale at high angular velocity (Fig. 15.6c).

Another class of models with alignment interactions accounts for nematic alignment interactions [33, 107, 108, 109], which arise in systems of head-tail symmetric particles, such as (self-propelled) rods [110] featuring apolar interactions.

15.2.6 Advantages and limitations

The ABP model and its alternatives are commonly used to perform particle-based simulations of active particles and also as a starting point for the formulation of continuum theories, as we shall discuss hereunder. These models have proven useful when applied to, e.g., the following problems:

- When we are concerned with single active particles, the ABP model has been very successful, e.g., to predict correlation functions in close agreement with experiments of Janus colloids [111].

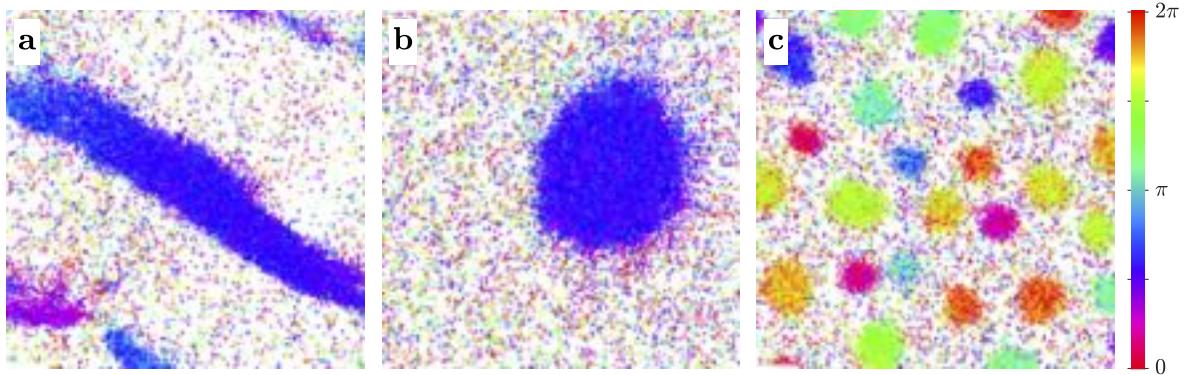


Figure 15.6: Brownian dynamics simulations of active particles with alignment interactions. Simulation snapshots of active particles with alignment interactions (Eqs. (15.1) and (15.12)) for $U = 0$, $D = 0$, and (a) $\omega = 0$ (smooth variant of the Vicsek model) and (b-c) $\omega > 0$ (chiral active particles) for (b) low and (c) high angular velocity. Colors represent particle orientation angles such that equally colored particles are aligned or phase-synchronized. Reproduced from Ref. [88].

- When hydrodynamic interactions play a minor role such as for certain active colloids, their many-body behavior is reasonably well described by ABPs [24]. Similarly, when hydrodynamic interactions are dominated by other interactions such as, e.g., phoretic interactions of autophoretic colloids with a near-uniform surface mobility, the ABP model serves as a useful starting point for the derivation of simple models with effective phoretic pair interactions [39].
- When a solvent is absent but fluctuations are still relevant as, e.g., for granular particles on vibrating plates where quasi-deterministic chaos arises and leads to effective randomness that can be described as Brownian noise, the ABP model can be used as a numerical model [4, 6].
- The ABP model is also useful for fundamental theoretical explorations, e.g., when we are more interested in the fundamental consequences of activity on the collective behavior of active particles rather than in the specific link to experimental realizations.

Compared with most wet models, a key advantage of the ABP model and its alternatives is their simplicity from both a conceptual and a computational viewpoint. In particular, these models permit simulations of very large ensembles of active particles (state-of-the-art simulations often use 10^5 to 10^7 particles [112, 71, 113, 62, 114, 70]). One key limitation of these models regarding the description of soft active matter systems is that they do not account for momentum conservation and often neither for hydrodynamic interactions. This can be particularly relevant for the description of collective behaviors or even for the description of single microswimmers near walls. The ABP model is popular when simulating the collective behavior of autophoretic active colloids as well. Here, beside hydrodynamic interactions, also phoretic interactions can play a crucial role and are also neglected by the standard ABP model [39], which can however be extended to take them into account [115, 116, 117, 118, 119, 120, 121].

15.3 Dry continuum models

To understand the collective behavior of active particles, one often uses continuum models, for either purely theoretical analyses or numerical analyses based on continuum simulations. In general, one can distinguish between phenomenological and microscopic theories.

Phenomenological theories are often based on the identification of the relevant slow variables (e.g., the density field $\rho(\mathbf{r}, t)$ in the case of isotropic active systems with particle number conservation, or the density field and polarization density for polar active systems with polar alignment interactions) and on writing down all terms allowed by symmetry and conservation laws up to a certain order. These theories are sometimes called *Landau theories*. A key advantage of phenomenological theories is that they predict the structure of the

field equations essentially based on symmetry, conservation laws, and dimensionality of the system without requiring any reference to the details of the underlying particle system (such as the precise form of the interactions). Thus, these field theories are sometimes called *generic theories* in this sense and can be formulated (and numerically solved) even if no underlying particle-based model is known. However, phenomenological field theories do not provide information about the values of the coefficients. Thus, one often treats all occurring coefficients as independent parameters and studies the phenomenology of the field equations as a function of these parameters. A related important drawback of this approach is that it then remains unclear whether there is an underlying particle-based model or realization which leads to the corresponding parameter values. A specific example of a phenomenological theory is discussed in Example 15.2.

Example 15.2: Phenomenological theory for isotropic active matter. Collective phenomena of isotropic active matter, such as phase separation, can be described, e.g., by the phenomenological active model B+, which is based on the common model B that describes phase separation in equilibrium systems [122]. Here, the density field $\rho(\mathbf{r}, t)$ is assumed to be the only slow variable of the system and the order parameter ϕ is related to it by the linear transformation $\phi = (2\rho - \rho_H - \rho_L)/(\rho_H - \rho_L)$, where ρ_H and ρ_L denote the density at the low-density and the high-density critical point, respectively [123, 124]. The active model B+ is given by the equations

$$\frac{\partial \phi}{\partial t} = -\nabla \cdot \left[-M\nabla \left(\frac{\delta \mathcal{F}}{\delta \phi} + \lambda |\nabla \phi|^2 \right) + \zeta M (\nabla^2 \phi) \nabla \phi + \sqrt{2D}\Lambda \right], \quad (15.13)$$

$$\mathcal{F}[\phi] = \int d^3 r \left[\frac{a}{2} \phi^2 + \frac{b}{4} \phi^4 + \frac{K}{2} |\nabla \phi|^2 \right]. \quad (15.14)$$

Here, the free-energy functional \mathcal{F} is approximated up to the order ϕ^4 and up to square-gradient terms [34]. Eq. (15.13) has the form of a continuity equation, ensuring particle number conservation, whereas reac-

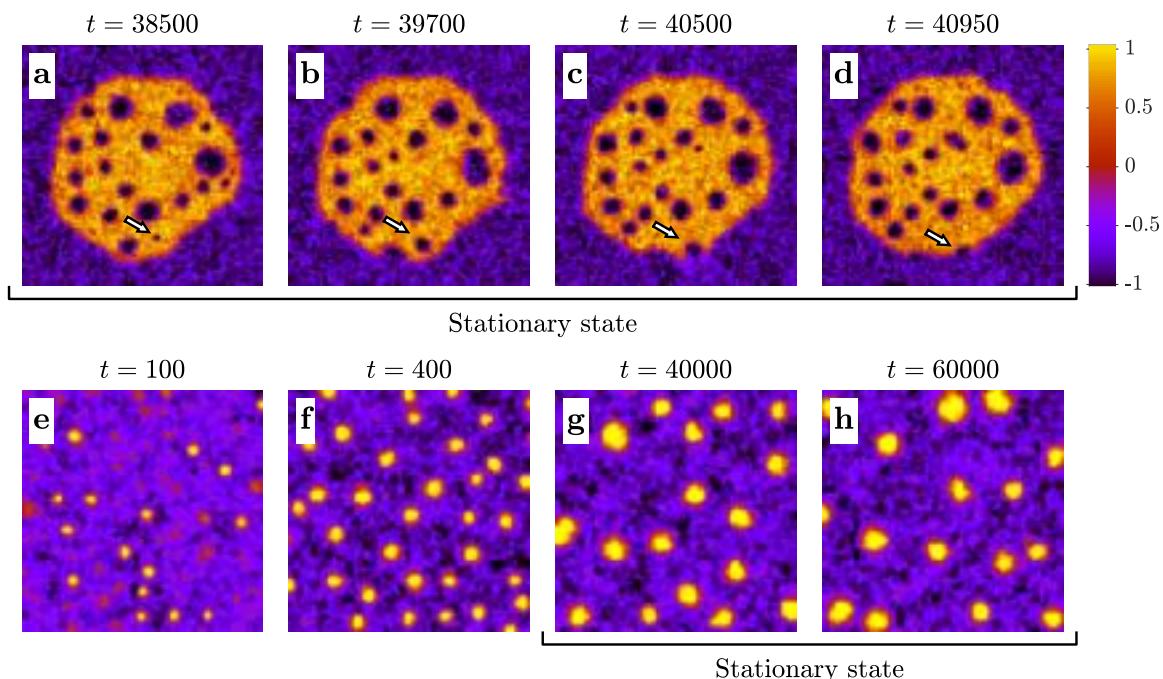


Figure 15.7: **Numerical results of the active model B+.** (a-d) Coexistence of a liquid phase (yellow to red colors) comprising continuously created vapor bubbles and a gas phase (purple to black colors). (e-h) Phase separation into a dense (yellow to red colors) and a dilute phase (black to purple colors). Dense clusters stabilize at a certain cluster size in the steady state. Reproduced from Ref. [34].

tion terms are not allowed. The order parameter ϕ is subject to a Gaussian white noise field $\Lambda(\mathbf{r}, t)$ with zero mean and unit variance. The diffusion coefficient is denoted by D and the mobility of the active particles by M . For active particles, the time-reversal symmetry is broken locally. This fact is included in the active model B+ by the additional terms proportional to λ and ζ . The active model B+ describes the phase separation behavior of isotropic active matter and predicts two types of patterns: The first one is characterized by phase separation into a dense and a dilute phase and the additional occurrence of vapor bubbles inside the dense phase, which are continuously created and move to the surface of the dense phase (Figs. 15.7a-d). The second pattern is characterized by the emergence of dense clusters that do not grow beyond a certain characteristic size (Fig. 15.7e-h). The coefficients a, b, K, λ, ζ are not known in this phenomenological approach and are treated as parameters of the model. Thus, there is no obvious connection to particle-based models such as the ABP model, whereas in microscopic theories all parameters are directly related to the underlying particle-based model, as will be discussed next.

Microscopic theories involve a systematic derivation of the field equations, typically from the underlying equations of motion for the individual active particles. This approach yields equations of motion for the relevant fields, which directly follow from the underlying particle-based model. Thus, in contrast to the former class of theories, one advantage of this second approach is that one obtains, in addition to the structure of equations, an explicit link between the coefficients of the particle-based model and the continuum theory. This typically leads to a (much) smaller number of independent parameters than one would obtain from phenomenological approaches. Another advantage is that, following the microscopic approach, terms allowed by symmetry cannot be missed, which has happened for various standard models of active matter in the past when following the phenomenological approach.

Example 15.3: Microscopic theories for isotropic active matter. There are several approaches to develop microscopic theories. To exemplify one of them, we consider a system of N active particles without alignment interactions. Then, we write down the Smoluchowski equation for the N -particle PDF and integrate out variables to obtain the one-particle density field. This approach has been used, for instance, to formulate a microscopic theory of MIPS in overdamped ABPs with positions \mathbf{r}_i and orientations θ_i [125]. Let $\mathbf{X} = \{\mathbf{r}_1, \dots, \mathbf{r}_N, \theta_1, \dots, \theta_N\}$ denote the state of the N -particle system. The corresponding Smoluchowski equation [126] for the joint PDF $\psi_N(\mathbf{X}, t)$ reads

$$\frac{\partial \psi_N}{\partial t} = \sum_{k=1}^N \nabla_{\mathbf{r}_k} \cdot \left[\frac{(\nabla_{\mathbf{r}_k} U)}{\gamma} - v_0 \hat{\mathbf{p}}_k + D \nabla_{\mathbf{r}_k} \right] \psi_N + D_R \sum_{k=1}^N \frac{\partial^2 \psi_N}{\partial \theta_k^2}, \quad (15.15)$$

with $U = \sum_{k < k'} u(|\mathbf{r}_k - \mathbf{r}_{k'}|)$, interaction potential $u(\Delta r_{kk'})$, and self propulsion along $\hat{\mathbf{p}}_k$ with speed v_0 . The Smoluchowski equation ensures probability conservation and its physical interpretation is illustrated in Fig. 15.8. Starting from Eq. (15.15), one usually derives an equation of motion for the one-particle PDF $\psi_1(\mathbf{r}_1, \theta_1, t)$ by integration. Due to the pair interactions, the resulting equation still contains terms which include the two-particle PDF. Similarly, one can derive an equation for the two-particle PDF, which then includes the three-particle PDF, and so on, leading to a hierarchy of coupled differential equations that have to be closed by a suitable closure scheme [127, 128]. Afterwards, an equation of motion for the particle density $\rho(\mathbf{r}, t)$ can be derived by integrating over the orientation ϕ , which typically couples again to higher moments and leads to a second hierarchy of equations, which has again to be closed using a suitable closure scheme. To study phase separation, one possible approximation to avoid the first type of hierarchy is to assume that the density varies slowly in space such that the local density is constant within the range of the interaction potential resulting in an effective self-propulsion speed $v(\rho) = v_0 - \zeta \rho$ with constant ζ . This density-dependent self-propulsion speed effectively accounts, to some extend, for the net effect of the repulsive interactions, namely the slowdown of active particles in regions of high density. The result of this microscopic approach fits well to computer simulations of ABPs and predicts MIPS in overdamped ABPs [125].

An alternative approach, sometimes called the *Dean approach* [129], is based on an explicit coarse-graining of the Langevin equations for the individual particles. This approach has been applied in several works, e.g., to describe MIPS in systems of RTPs [75], pattern formation in self-propelled particles with

Contribution of particle k to the probability current:

$$\begin{aligned}
 \mathbf{j}^{(k)}(\mathbf{X}, t) &= \mathbf{j}_t^{(k)}(\mathbf{X}, t) + \mathbf{j}_r^{(k)}(\mathbf{X}, t) \\
 &\quad \downarrow \\
 \mathbf{j}_D^{(k)}(\mathbf{X}, t) + \mathbf{j}_{\text{drift}}^{(k)}(\mathbf{X}, t) & \\
 \text{translational diffusion} & \quad \left(-\frac{\nabla_{\mathbf{r}_k} U}{\gamma} + v_0 \hat{\mathbf{e}}_k \right) \psi_N \\
 -D \nabla_{\mathbf{r}_k} \psi_N & \\
 \text{drift term} & \\
 & \quad \downarrow \\
 & \quad \mathbf{F} = -\nabla U \\
 & \quad \mathbf{v}_{\text{SP}} = v_0 \hat{\mathbf{e}} \\
 \text{rotational diffusion} & \quad -D_r \frac{\partial \psi_N}{\partial \phi_k}
 \end{aligned}$$

Probability conservation \rightarrow Smoluchowski equation:

$$\frac{\partial \psi_N}{\partial t} = \sum_{k=1}^N \left[\nabla_{\mathbf{r}_k} \mathbf{j}_t^{(k)} + \frac{\partial \mathbf{j}_r^{(k)}}{\partial \phi_k} \right]$$

Figure 15.8: Interpretation of the Smoluchowski equation. The Smoluchowski equation (Eq. (15.15)) for an ABP as probability conservation law. The contribution of each particle to the probability current can be decomposed into a translational and a rotational current. The former includes translational diffusion and a drift term due to the interaction potential as well as the self-propulsion velocity, whereas the latter considers rotational diffusion. The Smoluchowski equation can then be interpreted as a continuity equation ensuring probability conservation.

alignment interactions [105], collective phenomena in systems of CAPs [88], pattern formation in systems of phoretically interacting active colloids [115], or active systems showing nematic order [130]. Here, one uses Itô calculus [129, 131, 132] to deduce a stochastic differential equation, which involves multiplicative noise, for the (fluctuating) combined probability density $f(\mathbf{r}, \theta, t) = \sum_{i=1}^N \delta(\mathbf{r}_i(t) - \mathbf{r})\delta(\theta_i - \theta)$ for finding a particle with orientation θ at position \mathbf{r} at time t . To derive the one-particle density field $\rho(\mathbf{r}, t)$, one can then, for example, choose to neglect the multiplicative noise term (mean field) and derive a hierarchy of equations in a similar way to the Smoluchowski approach.

Example 15.4: Microscopic theories for polar active matter. The aforementioned continuum theories for dry active matter were focused on isotropic active matter that can be described by only considering the density field. However, if the particles feature alignment interactions such as in the Vicsek model, polar order can arise. Thus, describing these systems additionally requires the consideration of the mean local orientation of the particles by means of a polarization density $\mathbf{p}(\mathbf{r}, t)$. Corresponding theories for the density field and the polarization density can be derived based on the Smoluchowski approach or the Dean approach.

Another approach, which is aimed at describing the collective behavior of the Vicsek model (which is discrete in time in its original formulation) and is given by Ref. [133], is based on the Liouville equation for the N -particle PDF $\psi_N(\mathbf{X}, t)$ and applied to the well-known Vicsek model [102]. Within this model, the particles only interact during a collision event by aligning their orientation to that of their next neighbors and the orientation is subject to Gaussian white noise. Under the assumption that the particles are uncorrelated prior to a collision, the N -particle density is written as a product of one-particle densities, which is a good approximation if the noise strength is large and if the mean-free path between two collisions is larger than the interaction radius. Then, the one-particle PDF is obtained by integration. However, the solution contains complicated collision integrals that are approximated using the Chapman-Enskog expansion [134], which takes the stationary state as a reference and expands around it in powers of the gradients. Finally, this leads to a set of two coupled differential equations for ρ and \mathbf{p} . This set of equations is similar to that of the phenomenological Toner-Tu model [11] except for additional gradient terms, which occur only in the microscopic approach.

Independently of whether a theory is phenomenological or microscopic, the relevant field equations can then be studied based on various analytical and numerical techniques ranging from perturbation theories, linear stability analyses, or dynamical renormalization group calculations in the presence of additional noise terms to explicit numerical solutions based on, e.g., finite difference, finite volume, or finite element methods.

Example 15.5: Non-interacting overdamped active Brownian particles: the dynamics of the particle density follows the diffusion equation. Use the N -particle Smoluchowski equation for non-interacting overdamped ABPs in two dimensions to show that the dynamics of the particle density $\rho(\mathbf{r}, t)$ approximately follows the diffusion equation in the hydrodynamic regime (time scales larger than the persistence time and length scales much larger than the particle radius) with an enhanced diffusion coefficient.

Solution: As already shown in Eq. (15.15), the N -particle Smoluchowski equation for the N -particle PDF $\psi_N(\mathbf{X}, t)$, where $\mathbf{X} = \{\mathbf{r}_1, \dots, \mathbf{r}_N, \theta_1, \dots, \theta_N\}$, of non-interacting overdamped ABPs reads

$$\frac{\partial \psi_N}{\partial t} = \sum_{k=1}^N \nabla_{\mathbf{r}_k} \cdot \left[-v_0 \begin{pmatrix} \cos \theta_k \\ \sin \theta_k \end{pmatrix} + D \nabla_{\mathbf{r}_k} \right] \psi_N + D_R \sum_{k=1}^N \frac{\partial^2 \psi_N}{\partial \theta_k^2}. \quad (15.16)$$

One-particle PDF:

$$\psi_1(\mathbf{r}_1, \theta_1, t) = N \int d^2 r_{2\dots N} \int d\theta_{2\dots N} \psi_N(\mathbf{X}, t), \quad (15.17)$$

where $d^2 r_{2\dots N} = d^2 r_2 \dots d^2 r_N$ and $d\theta_{2\dots N} = d\theta_2 \dots d\theta_N$. Combining Eq. (15.17) and Eq. (15.16) leads to

(boundary terms vanish)

$$\psi_1 = D\nabla^2\psi_1 - v_0 \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \end{pmatrix} \cdot \nabla\psi_1 + D_R \frac{\partial^2\psi_1}{\partial\theta^2}. \quad (15.18)$$

Moment expansion (with $\mathbf{r} \equiv \mathbf{r}_1$ and $\theta \equiv \theta_1$):

- 0th moment: Density field

$$\rho(\mathbf{r}, t) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \psi_1(\mathbf{r}, \theta, t) \quad (15.19)$$

- 1st moment: Polarization density

$$\mathbf{p}(\mathbf{r}, t) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \psi_1(\mathbf{r}, \theta, t) \quad (15.20)$$

- 2nd moment: Nematic order parameter

$$Q(\mathbf{r}, t) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \left[\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} (\cos\theta, \sin\theta) - \frac{1}{2} \mathbb{I} \right] \psi_1(\mathbf{r}, \theta, t) \quad (15.21)$$

From Eqs. (15.19) and (15.20) together with Eq. (15.18), we get the following equations of motion for the density field and the polarization density:

$$\dot{\rho} = D\nabla^2\rho - v_0\nabla \cdot \mathbf{p} \quad (15.22)$$

$$\dot{\mathbf{p}} = D\nabla^2\mathbf{p} - \frac{v_0}{2}\nabla\rho - v_0\nabla^T Q^T - D_R \mathbf{p}. \quad (15.23)$$

To close the hierarchy of equations, we assume $Q \approx 0$. Furthermore, we make the following approximations:

- (i) For time scales larger than $1/D_R$, $\dot{\mathbf{p}}$ is much smaller than $D_R \mathbf{p}$. Therefore, $\dot{\mathbf{p}}$ is neglected.
- (ii) For length scales larger than the particle radius, the diffusion term $D\nabla^2\mathbf{p}$ can be neglected.

Finally, we get

$$\mathbf{p} = -\frac{v_0}{2D_R} \nabla\rho, \quad (15.24)$$

$$\boxed{\dot{\rho} = D_a \nabla^2 \rho}, \quad (15.25)$$

with active diffusion coefficient $D_a = D + v_0^2/(2D_R)$.

15.4 Wet active particles

The ABP model and its alternatives do not resolve the self-propulsion mechanism, but instead involve an effective force to phenomenologically model the resulting directed motion. To understand and describe the self-propulsion mechanism of a microswimmer, one has to explicitly model the flow field produced by the microswimmer and its interaction with the body of the swimmer.

Let us now briefly discuss the basic equations that are involved in the modeling of a single microswimmer. While swimming at the macroscale involves inertia and leads to flow fields described by the Navier–Stokes equation, microswimmers have to employ swimming mechanisms that work even in the absence of inertia since, at the microscale, viscous effects dominate over inertial effects [135]. This is quantified by the *Reynolds number*, which measures the relative importance of inertial versus viscous forces and is given by $\text{Re} = \rho Lv/\eta$, where the numerator represents the product of the fluid density ρ , a characteristic length scale L , and a typical flow speed v , whereas the denominator contains the solvent viscosity η . For microswimmers, $\text{Re} \ll 1$. For example, *E. coli* bacteria in water have $L \sim 3 \mu\text{m}$, $v \sim 30 \mu\text{m s}^{-1}$, $\eta = 0.001 \text{ Pa s}$, and $\rho = 1 \text{ g cm}^{-3}$ [52]; thus,

$\text{Re} \sim 10^{-5}$ to $10^{-4} \ll 1$ and inertial effects can safely be neglected. For comparison, phenomena occurring at high Reynolds numbers, such as turbulence, often occur at $\text{Re} \sim 10^3$ to 10^4 [136].

At low Reynolds number, the *Navier–Stokes equation* reduces to the *Stokes equation*, which describes creeping flow and reads

$$\eta \nabla^2 \mathbf{u} - \nabla p + \mathbf{f} = 0, \quad (15.26)$$

where $\mathbf{u}(\mathbf{r}, t)$ and $p(\mathbf{r}, t)$ are the solvent velocity field and the pressure field, respectively, and $\mathbf{f}(\mathbf{r}, t)$ is the force density representing the forces exerted by the microswimmers on the solvent. The Stokes equation is typically complemented by the incompressibility condition $\nabla \cdot \mathbf{u} = 0$ leading to a complete set of equations to determine $\mathbf{u}(\mathbf{r}, t)$ and $p(\mathbf{r}, t)$ for given $\mathbf{f}(\mathbf{r}, t)$ and boundary conditions. Notably, the Stokes equation does not contain any time derivatives, and therefore, the solvent responds instantaneously to the applied forces (no motion would take place once the forcing term is switched off), which reflects the absence of inertia. Accordingly, the swimming mechanism of scallops, which move by periodically opening and closing their shells, would not work at low Reynolds numbers (Fig. 15.9a). Likewise, any other mechanism based on reciprocal motions would not lead to directed motion. This is Purcell’s *scallop theorem* [135].

The general procedure to model microswimmers that move by body-shape deformations (or squirmers) at low Reynolds numbers consists in solving the Stokes equation with appropriate boundary conditions for the solvent velocity field \mathbf{u} on the surface of the microswimmers. This yields the solvent velocity field \mathbf{u} , from which the stress tensor $\sigma = \eta (\nabla \otimes \mathbf{u} + (\nabla \otimes \mathbf{u})^T)$ can be obtained. The latter then allows one to calculate the total force $\mathbf{F} = \int_S dS \sigma(\mathbf{r}, t) \hat{n}$ and the torque $\mathbf{T} = \int_S dS \mathbf{r} \times (\sigma(\mathbf{r}, t) \hat{n})$ that act on the microswimmer, where S and dS denote the surface of the microswimmer and a differential element of it, respectively. Then, for a solid particle, the rigidity condition

$$\mathbf{u}(\mathbf{r}) = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}, \quad \mathbf{r} \in S \quad (15.27)$$

is typically assumed to apply at the surface S of the particle and links the particle velocity \mathbf{v} and angular velocity $\boldsymbol{\omega}$ to \mathbf{F} and \mathbf{T} . Finally, the torque-free ($\mathbf{T} = 0$) and force-free ($\mathbf{F} = 0$) conditions allow one to solve for \mathbf{v} and $\boldsymbol{\omega}$ [52, 140]. Since microswimmers often deform in a cyclic way, the net displacement during one cycle of period T is given by $\int_0^T dt \mathbf{v}(t)$, which is zero for reciprocal movement in the regime of low Reynolds numbers [40]. Thus, non-reciprocal body-shape deformations are required to produce directed motion. Two examples of biological microswimmers that self propel by non-reciprocal motion are shown in Figs. 15.9b and 15.9c. A minimal microswimmer model can be constructed, e.g., based on three spheres connected by two

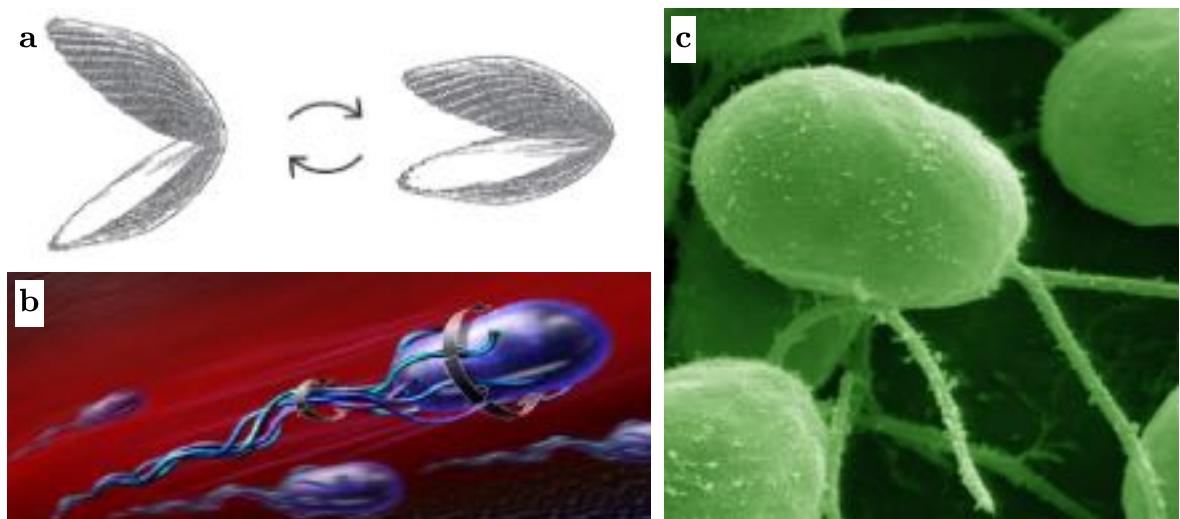


Figure 15.9: **The scallop theorem.** (a) Motion of a scallop. By quickly closing and slowly opening the two shells, the scallop produces a net flow and starts moving. At low Reynolds number, the net displacement is zero for this reciprocal motion. Reproduced from Ref. [137]. (b) Schematic of the non-reciprocal motion of an *E. coli* bacterium. Reproduced from Ref. [138]. (c) Electron microscope image of *Chlamydomonas reinhardtii* algae showing the flagella producing self propulsion by non-reciprocal motion. Reproduced from Ref. [139].

arms, which periodically change their length (*three-sphere swimmer*) [141, 142, 143, 144, 145], or based on two spheres which can contract or expand radially and are connected by an elastic arm [142, 146, 147].

15.5 Hydrodynamic interactions

In ensembles of microswimmers, each of them generates a specific flow pattern that typically decays slowly in space and leads to long-ranged hydrodynamic interactions among different microswimmers as well as to hydrodynamic (self-)interactions with walls and interfaces. These hydrodynamic interactions are typically not included in models of dry active matter such as the ABP model and its alternatives. One way of simulating several interacting microswimmers is to explicitly model the detailed self-propulsion mechanism of each microswimmer, i.e., to alternately solve the Stokes equation with the microswimmer-solvent boundary conditions for all swimmers simultaneously and to propagate the swimmers based on the force- and torque-free conditions. While such an approach is conceptually relatively simple and accurate in principle, it entails a huge numerical effort and typically becomes unfeasible even for moderately large microswimmer ensembles. In the following sections, we briefly discuss some alternative approaches, which provide more efficient numerical descriptions of microswimmer ensembles.

15.5.1 Minimal models

To model the dynamics of large microswimmer ensembles, an explicit modeling of the solvent flow including the detailed particle–solvent boundary conditions occurring in real microswimmers is often numerically so demanding that very large system sizes remain unreachable. Therefore, one often looks for a compromise between the ABP model, which neglects hydrodynamic interactions and momentum conservation altogether, and an explicit modeling of the self-propulsion mechanism of all interacting microswimmers in a given ensemble. One common approach involves formulating hydrodynamically consistent minimal models for the collective behavior of microswimmers where one does not explicitly describe the self-propulsion mechanism of each microswimmer but replaces each microswimmer with a simpler version that creates a similar (far-field) flow pattern. To this end, one uses a multipole expansion of the flow field (similar to that used, e.g., in electrodynamics) [148, 149, 150, 151] and only considers the leading-order terms. In the simplest case, these are the so-called *singularity solutions* of the Stokes equation (e.g., the flow field of a force dipole), which are then used to replace the flow field created by each microswimmer and are equivalent to the far-field flow pattern generated by the actual microswimmer to be modeled. For example, it is well known that *E. coli* bacteria produce essentially the same far-field flow pattern as a force dipole (pusher) [36] and *Chlamydomonas* algae produce a far-field flow pattern that can be represented by the flow field produced by an oscillatory force dipole [152]. Let us briefly discuss three common singularity solutions of the Stokes equation:

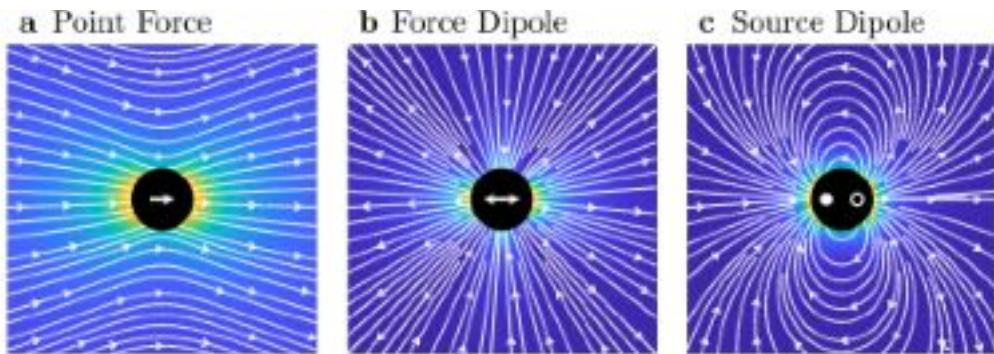


Figure 15.10: **Singularity solutions of the Stokes equation.** Illustration of the velocity field $\mathbf{u}(\mathbf{r})$ of (a) a point force, (b) a force dipole, and (c) a source dipole. The arrows indicate the direction of the flow field and the background color indicates its magnitude.

Point force (*stokeslet*). The flow generated by a point force $\mathbf{f}_p = f\hat{\mathbf{e}}\delta(\mathbf{r} - \mathbf{r}_0)$ placed at position \mathbf{r}_0 and pointing along the direction $\hat{\mathbf{e}}$ is similar to the far-field flow of a particle that is driven by an external force [140, 149]. By setting $\mathbf{f} = \mathbf{f}_p$ in the Stokes equation (Eq. (15.26)), the resulting velocity field reads

$$\mathbf{u}_{PF}(\mathbf{r}) = \frac{f}{8\pi\eta r} [\hat{\mathbf{e}} + (\hat{\mathbf{r}} \cdot \hat{\mathbf{e}}) \hat{\mathbf{r}}], \quad (15.28)$$

where $r = |\mathbf{r} - \mathbf{r}_0|$, $\hat{\mathbf{r}} = (\mathbf{r} - \mathbf{r}_0)/r$, and η denotes the viscosity of the solvent. The velocity field is shown in Fig. 15.10a. Since microswimmers are force free (momentum conservation), the Stokeslet solution alone is unsuitable to represent them.

Force dipole. The far-field solution of the Stokes equation in the presence of two point forces $\mathbf{f}_+ = f\hat{\mathbf{e}}\delta(\mathbf{r} - \mathbf{r}_0 - (l/2)\hat{\mathbf{e}})$ and $\mathbf{f}_- = -f\hat{\mathbf{e}}\delta(\mathbf{r} - \mathbf{r}_0 + (l/2)\hat{\mathbf{e}})$, which are separated by a distance l , reads

$$\mathbf{u}_{FD}(\mathbf{r}) = \frac{fl}{8\pi\eta r^2} [3(\hat{\mathbf{e}} \cdot \hat{\mathbf{r}})^2 - 1] \hat{\mathbf{r}} \quad (15.29)$$

in the limit $l \rightarrow 0$ or at distances $r \gg l$ and it is represented in Fig. 15.10b for $f > 0$ [140, 149]. These force dipoles push fluid molecules in the forward and backward directions. Hence, microswimmers that show this kind of far-field flow pattern are called *pushers*, whereas the case of $f < 0$, where all flow field lines are reverted, corresponds to a *puller*.

Source dipole. The point-force and force-dipole solutions are obtained by solving the Stokes equation together with the incompressibility condition $\nabla \cdot \mathbf{u} = 0$. In the presence of sources of solvent molecules, the Stokes equation is unchanged, but the incompressibility condition changes to $\nabla \cdot \mathbf{u}(\mathbf{r}) = s(\mathbf{r})$, where $s(\mathbf{r})$ denotes the source density [149]. While a point source is of limited relevance (it would lead to a net flow of solvent molecules entering or leaving the domain), the source dipole is an important singularity solution to the Stokes equation. Its source density consists of two point sources $s_+(\mathbf{r}) = Q\delta(\mathbf{r} - \mathbf{r}_0 - (l/2)\hat{\mathbf{e}})$ (source of solvent molecules) and $s_-(\mathbf{r}) = -Q\delta(\mathbf{r} - \mathbf{r}_0 + (l/2)\hat{\mathbf{e}})$ (sink of solvent molecules) that are separated by a distance l , where $Q > 0$ denotes the magnitude of the source densities. The corresponding solution to the Stokes equation in the limit $l \rightarrow 0$ reads [149]

$$\mathbf{u}_{SD}(\mathbf{r}) = \frac{Ql}{4\pi r^3} [3(\hat{\mathbf{e}} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{e}}] \quad (15.30)$$

and its velocity field is shown in Fig. 15.10c.

Since self-propelled particles are force free, the simplest representation of active particles by singularity solutions of the Stokes equation is given by force and source dipoles. Examples of simulations of microswimmer models based on these singularity solutions comprise, e.g., studies of motile suspensions of active rod-like particles [153], of the dynamics of a single molecule composed of microswimmers [154], of RTPs with hydrodynamic interactions [155], or of microswimmers near boundaries [156].

To simulate microswimmers based on singularity solutions of the Stokes equation, one often models the external fluid velocity field \mathbf{u} as a sum of all microswimmer singularity solutions and applies certain boundary conditions on the surface of each microswimmer. The velocity \mathbf{v} of each microswimmer is then calculated using the force-free and torque-free conditions based on the stress tensor, as previously discussed for a single microswimmer, via numerical integration. To obtain self propulsion, one shifts the force or source dipole away from the center of the particles [153, 154]. Moreover, one can also combine the singularity solutions with numerical solvers such as the Lattice-Boltzmann method [157] discussed below. Beside simulations, the force and source dipole models are used to develop continuum theories for active matter with hydrodynamic interactions, which we will discuss in the last section of this chapter.

Example 15.6: Flow field of a Stokeslet. Calculate the flow field of a Stokeslet (point force).

Solution: Let us define the point force as

$$\mathbf{f} = f\hat{\mathbf{e}}\delta(\mathbf{r} - \mathbf{r}_0). \quad (15.31)$$

Without loss of generality, we set $\mathbf{r}_0 = 0$ and consider the boundary conditions $\lim_{r \rightarrow \infty} p = 0$ and $\lim_{r \rightarrow \infty} \mathbf{u} = 0$

with $r = |\mathbf{r}|$. From the Stokes equation [cf. Eq. (15.26)] and the incompressibility condition $\nabla \cdot \mathbf{u} = 0$ we get

$$\begin{aligned} \nabla \cdot (\eta \nabla^2 \mathbf{u} - \nabla p + \mathbf{f}) &= 0 \\ \Rightarrow \eta \nabla^2 \underbrace{\nabla \cdot \mathbf{u}}_{=0} - \nabla^2 p + \nabla \cdot \mathbf{f} &= 0 \\ \Rightarrow -\nabla^2 p &= -\nabla \cdot \mathbf{f}. \end{aligned} \quad (15.32)$$

We now define $\nabla^2 \phi = -\delta(\mathbf{r})$, to which the Green's function $\phi(\mathbf{r}) = 1/(4\pi r)$ is a solution (as in electrostatics). Inserting Eq. (15.31) into Eq. (15.32) yields

$$-\nabla^2 p = \nabla \cdot \hat{\mathbf{e}} f \nabla^2 \phi$$

and we get

$$p = p_\infty - \nabla \cdot \hat{\mathbf{e}} f \phi \stackrel{p_\infty = 0}{=} -f \nabla \phi \cdot \hat{\mathbf{e}}.$$

Inserting the latter into the Stokes equation, Eq. (15.26), gives

$$\eta \nabla^2 \mathbf{u} = -f [\nabla \nabla \phi - \mathbb{I} \nabla^2 \phi] \cdot \hat{\mathbf{e}}.$$

Next, we define a function ζ such that $\nabla^2 \zeta = \phi$. Hence, we obtain the velocity field

$$\mathbf{u} = -\frac{f}{\eta} [\nabla \nabla \zeta - \mathbb{I} \phi] \cdot \hat{\mathbf{e}} + \mathbf{c}$$

with a function \mathbf{c} that fulfills $\nabla \cdot \mathbf{c} = 0$ and $\nabla^2 \mathbf{c} = 0$. Since $\mathbf{u} \rightarrow 0$ for $r \rightarrow \infty$, we set $\mathbf{c} = 0$. Now, from $\phi = 1/(4\pi r)$ we obtain $\zeta = r/(8\pi)$. Thus, we get

$$\mathbf{u} = -\frac{f}{\eta} \left[\nabla \nabla \frac{r}{8\pi} - \mathbb{I} \frac{1}{4\pi r} \right] \cdot \hat{\mathbf{e}}$$

Observing that $\nabla \nabla \cdot \hat{\mathbf{e}} = (\nabla \cdot \hat{\mathbf{e}}) \nabla$ we have:

$$\mathbf{u} = \frac{f}{8\pi\eta r} \left[\frac{2}{r} - (\nabla \cdot \hat{\mathbf{e}}) \nabla r \right]$$

Because $(\nabla \cdot \hat{\mathbf{e}}) \nabla r = \frac{1}{r} - \frac{(\hat{\mathbf{r}} \cdot \hat{\mathbf{e}}) \hat{\mathbf{r}}}{r}$, one can obtain:

$$\mathbf{u} = \frac{f}{8\pi\eta r} [\hat{\mathbf{e}} - (\hat{\mathbf{r}} \cdot \hat{\mathbf{e}}) \hat{\mathbf{r}}]$$

with $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$.

15.5.2 Squirmer models

An alternative approach to formulate hydrodynamically consistent models of microswimmers is to consider *squirmers*, i.e., spherical particles with a prescribed solvent flow along the surface (without explicitly modeling the origin of the latter) [158, 159, 160, 161, 162, 148, 163, 164, 165, 166, 167]. On the surface of the squirmer particle, the vertical fluid velocity is set to zero and the tangential surface velocity is prescribed by a series of first derivatives of Legendre polynomials, which can be used, e.g., to model the net effect of autophoresis, which leads to a slip velocity across the surface of Janus particles [168]. The squirmer model has been used in several works, e.g., in combination with the lattice-Boltzmann method [165, 169, 170] or multi-particle collision dynamics simulations [158, 159, 161, 162, 163, 164, 167, 166].

In contrast to the ABP model and its alternatives, microswimmer models based on combinations of singularity solutions of the Stokes equation or on squirmers are momentum conserving and can correctly describe

hydrodynamic interactions at large inter-particle distances for a given active system. However, they do not necessarily account for the correct hydrodynamic near-field interactions and are therefore mainly useful to model active systems at low density (squirmer models, when used to represent Janus particles, may serve as an exception, which is expected to correctly describe hydrodynamic interactions down to distances on the order of the slip length [168]). These effective models are often used also as a starting point for continuum theories as briefly discussed further below.

15.5.3 Explicit simulation of the solvent

In the following, we briefly introduce several numerical methods, which will be discussed more in detail in the following chapters. These methods are frequently used in active matter physics to explicitly determine the flow field and to simulate hydrodynamic interactions, often beyond the far-field approximation.

The *lattice-Boltzmann method* (LBM) is a popular method to solve fluid dynamics problems, where one solves the Boltzmann equation instead of the (Navier–)Stokes equation and exploits the fact that the latter equation can be derived from the former [171, 172, 173, 174, 175, 176, 165, 177]. Interestingly, the Boltzmann equation is numerically often more convenient when combined with suitable approximations. It describes the time evolution of the PDF $f(\mathbf{r}, \mathbf{v}, t)$ with velocity \mathbf{v} at position \mathbf{r} and time t . With the so-called collision operator $\Omega(f)$, the Boltzmann equation reads [178, 172]

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{v}} f = \Omega(f), \quad (15.33)$$

where m denotes the mass of the particles and \mathbf{F} is the external force field acting on them. The second term on the left-hand side describes advection of the particles with velocity \mathbf{v} , whereas the third term describes external forces acting on the solvent particles and affecting their velocity. The source term on the right-hand side of Eq. (15.33) describes the local redistribution of the solvent particles due to collisions. This collision operator is often approximated by $\Omega(f) = -(f - f_{\text{eq}})/\tau$, which describes the relaxation of f towards the equilibrium probability density f_{eq} on the time scale τ and is known as the Bhatnagar–Gross–Krook (BGK) collision operator [179]. In the LBM, the continuous Boltzmann equation (i.e., Eq. (15.33)) is discretized in position, velocity, and time and numerically solved on a lattice with spacing Δx at discrete times with time step Δt . The velocity \mathbf{v} can only take discrete values \mathbf{c}_i , which are given by a discrete set $\{\mathbf{c}_i, w_i\}$ with weights w_i . The discretized Boltzmann equation is then solved numerically as discussed, e.g., in Ref. [172] and in Chapter 24. To simulate microswimmers that, e.g., create directed motion through body-shape deformations, one often describes the microswimmer surface as a set of boundary links that define a closed surface and solves the discretized Boltzmann equation together with suitable boundary conditions [172].

Multi-particle collision dynamics (MPCD) is another popular approach to simulate the dynamics of microswimmers, where, in contrast to the LBM, the solvent is represented by point-like particles which have continuous positions and velocities [180, 181, 182, 183, 184, 185, 186]. As discussed in Chapter 23, to model active particles, one usually combines the MPCD method for the solvent molecules with molecular dynamics (MD) simulations of the active particles. These are coupled to the solvent and are represented either as a single particle or by a quasi-continuous distribution of particles which are connected with (time-dependent) springs and represent the surface of a (deformable) microswimmer [187]. The MPCD method has been used in several works to investigate, e.g., chemotactic Janus colloids [188], active particles with phoretic interactions [189], dynamics of active particles in chemically active media [190], the motion of squirmers [164, 166, 159], the influence of hydrodynamic interactions on phase separation in systems of microswimmers [162], collective behavior of sperm cells [191], and active particles in filament networks [192].

Dissipative-particle dynamics (DPD) is another coarse-grained approach to modeling the solvent. Here, each DPD particle represents a small solvent region and, similar to the MPCD simulations, the positions and velocities of the DPD particles take continuous values. The DPD particles interact via three types of effective forces: A weak conservative force models the soft repulsion of the solvent molecules, a dissipative force models the friction, and a random force accounts for thermal fluctuations. Knowing these forces, Newton's equation of motion is solved for the DPD particles to obtain the hydrodynamics of the solvent [193, 194, 195]. This model has been adapted, e.g., to active suspensions [196] and to model the self-propulsion of Janus colloids [197].

Microscopic solvent models. Finally, beside the previously discussed mesoscale-simulation methods, particle-based simulations of the solvent molecules based on direct MD simulations, which allow one to resolve very

small spatial and temporal scales, are possible. Nevertheless, these simulations are computationally very intense, which makes it impossible to study systems on the microscale over time scales of seconds, which are relevant to most active matter systems. Still, this explicit modeling of the solvent has been successfully used to model a self-propelled particle in a Lennard–Jones solvent [198].

Overall, the LBM, MPCD, and the DPD methods are mesoscale simulation methods, which can be applied to many hydrodynamic problems in soft and active matter physics and beyond. Since the DPD method is based on particles moving in continuous space, it avoids lattice artifacts and allows simulations capturing much larger length and time scales than typically possible in MD simulations. However, DPD simulations include a large number of parameters (in order to model the different forces), which have to be chosen carefully. The MPCD method, on its part, which models the net effect of individual collisions rather than accounting for every collision event, is computationally very efficient, and can be efficiently parallelized. This applies also to the LBM, which numerically solves the Boltzmann equation and is well suited, e.g., for implementing complex (moving) boundaries [172].

15.6 Wet continuum models

Based on the previously discussed effective microswimmer models, continuum theories for large ensembles of particles can be formulated that explicitly account for hydrodynamic interactions, at least at low density. These theories describe wet active matter and can be formulated, e.g., based on the puller and pusher solutions of the Stokes equation. One popular approach to account for hydrodynamic far-field interactions is to write down the (overdamped) equations of motion for the position and orientation of each microswimmer, which couple with the overall fluid velocity field. The contribution of each microswimmer to the overall velocity field is modeled by singularity solutions of the Stokes equation such as force or source dipoles (which can be superimposed thanks to the linearity of the Stokes equation). One then derives a continuity equation for the N -particle PDF, which typically takes the form of a Fokker-Planck equation [199, 200, 201, 202]. From here, one can proceed in a similar way to that of microscopic theories for dry active matter in order to derive an equation of motion for the one-particle density. Since the described approach to formulate continuum theories for wet active matter is based on the singularity solutions of the Stokes equation, which only describe the far-field flow pattern of active particles, near-field hydrodynamic effects are not included in this approach. However, although complicated in practice, one can go beyond the far-field regime in principle, e.g., by using superimposed singularity solutions to represent the flow field contribution due to each swimmer or by starting with squirmer models.

Let us finally mention that one can alternatively formulate phenomenological minimal models of wet active matter. Following a similar spirit to the case of dry active matter, these models are generic in the sense that they are largely based on considerations of symmetry, conservation laws, and dimensionality and do not refer to details such as the specific self-propulsion mechanism, which is employed by the microswimmers. One example of such a minimal model for wet active matter is given by the phenomenological active model H [203], which accounts for momentum conservation. It is based on the active model B [124] and is closely related to the model H for equilibrium systems [122]. The active model H addresses the phase separation behavior of wet active matter and couples the generalized density field $\phi(\mathbf{r}, t)$ to the velocity field $\mathbf{v}(\mathbf{r}, t)$ of the solvent. The general idea is that diffusive dynamics of the active particles take place in the moving frame of the solvent and the velocity field of the solvent is given by the corresponding Navier–Stokes equation. There are also phenomenological models for specific phenomena such as bacterial turbulence, which are based on phenomenological equations to describe the fluid velocity field [35].

More generally, there is a large number of alternative approaches to formulating continuum theories for microswimmers. Readers interested in further details are referred to Refs. [33, 148, 40].

Bibliography

- [1] Christian Scholz, Soudeh Jahanshahi, Anton Ldov, and H. Löwen. *Nature Communications*, 9:5156, 2018.
- [2] Christian Scholz, Michael Engel, and Thorsten Pöschel. *Nature Communications*, 9:931, 2018.
- [3] Christian Scholz, Sean D’Silva, and Thorsten Pöschel. *New Journal of Physics*, 18:123001, 2016.

- [4] Yann Lanoiselée, Guillaume Briand, O. Dauchot, and Denis S. Grebenkov. *Physical Review E*, 98:062112, 2018.
- [5] Arshad Kudrolli, Geoffroy Lumay, Dmitri Volfson, and Lev S. Tsimring. *Physical Review Letters*, 100:058001, 2008.
- [6] Lee Walsh, Caleb G. Wagner, Sarah Schlossberg, Christopher Olson, Aparna Baskaran, and Narayanan Menon. *Soft Matter*, 13:8964, 2017.
- [7] O. Dauchot and V. Démery. *Physical Review Letters*, 122:068002, 2019.
- [8] A. Deblais, T. Barois, T. Guerin, P. H. Delville, R. Vaudaine, J. S. Lintuvuori, J. F. Boudet, J. C. Baret, and H. Kellay. *Physical Review Letters*, 120:188002, 2018.
- [9] G. A. Patterson, P. I. Fierens, F. Sangiuliano Jimka, P. G. König, A. Garcimartín, I. Zuriguel, L. A. Pugnaloni, and D. R. Parisi. *Physical Review Letters*, 119:248301, 2017.
- [10] Charles Wolgemuth, Egbert Hoiczyk, Dale Kaiser, and George Oster. *Curr. Biol.*, 12:369, 2002.
- [11] John Toner and Yuhai Tu. *Physical Review E*, 58:4828, 1998.
- [12] J. Buhl, D. J.T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. *Science*, 312:1402, 2006.
- [13] Charlotte K. Hemelrijk and Hanno Hildenbrandt. *Interface Focus*, 2:726, 2012.
- [14] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. *Proc. Natl. Acad. Sci.*, 105:1232, 2008.
- [15] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak. *Proc. Natl. Acad. Sci.*, 109:4786, 2012.
- [16] Jesse L. Silverberg, Matthew Bierbaum, James P. Sethna, and Itai Cohen. *Physical Review Letters*, 110:228701, 2013.
- [17] Nicolas Bain and Denis Bartolo. *Science*, 363:46, 2019.
- [18] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe. Engineering sensorial delay to control phototaxis and emergent collective behaviors. *Physical Review X*, 6(1):011008, 2016.
- [19] M. Leyman, F. Ogemark, J. Wehr, and G. Volpe. Tuning phototactic robots with sensorial delays. *Physical Review E*, 98(5):052606, 2018.
- [20] Gábor Vásárhelyi, Csaba Virág, Gergő Somorjai, Tamás Nepusz, Agoston E. Eiben, and Tamás Vicsek. *Science Robotics*, 3:eaat3536, 2018.
- [21] G. Volpe, S. Gigan, and G. Volpe. Simulation of the active brownian motion of a microswimmer. *American Journal of Physics*, 82(7):659–664, 2014.
- [22] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. *Physical Review Letters*, 99:048102, 2007.
- [23] Marjolein N. Van Der Linden, L. C. Alexander, D. G. A. L. Aarts, and O. Dauchot. *Physical Review Letters*, 123:098001, 2019.
- [24] Ivo Buttinoni, Julian Bialké, Felix Kümmel, H. Löwen, Clemens Bechinger, and Thomas Speck. *Physical Review Letters*, 110:238301, 2013.
- [25] C. Jin, Babak V. Hokmabad, Kyle A. Baldwin, and C. C. Maass. *Journal of Physics: Condensed Matter*, 30:054003, 2018.
- [26] C. Jin, C. Krüger, and C. C. Maass. *Proc. Natl. Acad. Sci.*, 114:5089, 2017.

- [27] C. De Blois, M. Reyssat, S. Michelin, and O. Dauchot. *Physical Review Fluids*, 4:054001, 2019.
- [28] C. C. Maass, C. Krüger, S. Herminghaus, and C. Bahr. *Annu. Rev. Condens. Matter Phys.*, 7:171, 2016.
- [29] J. Elgeti, R. G. Winkler, and G. Gompper. *Rep. Prog. Phys.*, 78:056601, 2015.
- [30] Guannan Liu, Adam Patch, Fatmagül Bahar, David Yllanes, Roy D. Welch, M. Cristina Marchetti, S. Thutupalli, and Joshua W. Shaevitz. *Physical Review Letters*, 122:248102, 2019.
- [31] S. Rafaï, L. Jibuti, and P. Peyla. *Physical Review Letters*, 104:098102, 2010.
- [32] D. A. P. Reid, H. Hildenbrandt, J. T. Padding, and C. K. Hemelrijk. *Physical Review E*, 85:021901, 2012.
- [33] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. *Reviews of Modern Physics*, 85:1143, 2013.
- [34] Elsen Tjhung, Cesare Nardini, and M. E. Cates. *Physical Review X*, 8:031080, 2018.
- [35] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Lowen, and J. M. Yeomans. *Proc. Natl. Acad. Sci.*, 109:14308, 2012.
- [36] Knut Drescher, Jörn Dunkel, Luis H. Cisneros, Sujoy Ganguly, and Raymond E. Goldstein. *Proc. Natl. Acad. Sci.*, 108:10940, 2011.
- [37] P. Zhang, S. Jana, M. Giarra, P.P. Vlachos, and S. Jung. *European Physical Journal – Special Topics*, 224:3199, 2015.
- [38] F. A. Morrison. *J. Colloid Interface Sci.*, 34:210, 1970.
- [39] B. Liebchen and H. Löwen. *J. Chem. Phys.*, 150:061102, 2019.
- [40] E. Lauga and Thomas R Powers. *Rep. Prog. Phys.*, 72:096601, 2009.
- [41] Mojdeh Heidari, Andreas Bregulla, Santiago Muinos Landin, Frank Cichos, and Regine von Klitzing. *Langmuir*, 36:7775, 2020.
- [42] Ran Niu and Thomas Palberg. *Soft Matter*, 14:7554, 2018.
- [43] B. Liebchen, Ran Niu, Thomas Palberg, and H. Löwen. *Physical Review E*, 98, 2018.
- [44] Manoj Manjare, Fengchang Yang, Rui Qiao, and Yiping Zhao. *J. Phys. Chem. C*, 119:28361, 2015.
- [45] Wenjie Fei, Yang Gu, and Kyle J.M. Bishop. *Curr. Opin. Colloid Interface Sci.*, 32:57, 2017.
- [46] M. Schmitt and H. Stark. *Phys. Fluids*, 28, 2016.
- [47] A. Domínguez, P. Malgaretti, M. N. Popescu, and S. Dietrich. *Physical Review Letters*, 116, 2016.
- [48] Darren Crowdy. *Physical Review Fluids*, 5:124004, 2020.
- [49] M. Wittmann, M. N. Popescu, A. Domínguez, and J. Simmchen. *arXiv:2101.02499 [cond-mat.soft]*, 2021.
- [50] S. Thutupalli, R. Seemann, and S. Herminghaus. *New Journal of Physics*, 13:073021, 2011.
- [51] A. Domínguez and M. N. Popescu. *Soft Matter*, 14:8017, 2018.
- [52] Clemens Bechinger, Roberto Di Leonardo, H. Löwen, Charles Reichhardt, G. Volpe, and G. Volpe. *Reviews of Modern Physics*, 88:045006, 2016.
- [53] S. M. Block, J. E. Segall, and H. C. Berg. Adaptation kinetics in bacterial chemotaxis. *J. Bacteriol.*, 154:312, 1983.

- [54] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier. *European Physical Journal – Special Topics*, 202:1, 2012.
- [55] B. ten Hagen, S. van Teeffelen, and H. Löwen. Non-Gaussian behaviour of a self-propelled particle on a substrate. *Condens. Matter Phys.*, 12:725, 2009.
- [56] B ten Hagen, S van Teeffelen, and H Löwen. *Journal of Physics: Condensed Matter*, 23:194119, 2011.
- [57] H. Löwen. *J. Chem. Phys.*, 152:040901, 2020.
- [58] A. Callegari and G. Volpe. *Numerical Simulations of Active Brownian Particles*, page 211. Springer International Publishing, 2019.
- [59] Étienne Fodor and M. Cristina Marchetti. *Physica A*, 504:106, 2018.
- [60] S. van Teeffelen and H. Löwen. *Physical Review E*, 78:020101, 2008.
- [61] H. Löwen. *European Physical Journal – Special Topics*, 225:2319, 2016.
- [62] Joakim Stenhammar, D. Marenduzzo, Rosalind J. Allen, and M. E. Cates. *Soft Matter*, 10:1489, 2014.
- [63] G. Gonnella, D. Marenduzzo, A. Suma, and A. Tiribocchi. *Comptes Rendus Phys.*, 16:316, 2015.
- [64] Shibananda Das, Gerhard Gompper, and Roland G. Winkler. *New Journal of Physics*, 20, 2018.
- [65] Yaouen Fily and M. Cristina Marchetti. *Physical Review Letters*, 108:1, 2012.
- [66] Jeremie Palacci, Stefano Sacanna, Asher Preska Steinberg, David J. Pine, and Paul M. Chaikin. *Science*, 339:936, 2013.
- [67] M. E. Cates and J. Tailleur. *Annu. Rev. Condens. Matter Phys.*, 6:219, 2015.
- [68] L. Caprini, U. Marini Bettolo Marconi, and A. Puglisi. *Physical Review Letters*, 124:078001, 2020.
- [69] John D. Weeks, David Chandler, and Hans C. Andersen. *J. Chem. Phys.*, 54:5237, 1971.
- [70] Gabriel S. Redner, Michael F. Hagan, and Aparna Baskaran. *Physical Review Letters*, 110:055701, 2013.
- [71] Suvendu Mandal, B. Liebchen, and H. Löwen. *Physical Review Letters*, 123:228001, 2019.
- [72] Mario Sandoval. *Physical Review E*, 101:012606, 2020.
- [73] Luis L. Gutierrez-Martinez and Mario Sandoval. *J. Chem. Phys.*, 153:044906, 2020.
- [74] Isabella Petrelli, Leticia F. Cugliandolo, G. Gonnella, and A. Suma. *Physical Review E*, 102:012609, 2020.
- [75] J. Tailleur and M. E. Cates. *Physical Review Letters*, 100:218103, 2008.
- [76] M. E. Cates and J. Tailleur. *EPL*, 101:20010, 2013.
- [77] A. P. Solon, M. E. Cates, and J. Tailleur. *European Physical Journal – Special Topics*, 224:1231, 2015.
- [78] Maryam Khatami, Katrin Wolff, Oliver Pohl, Mohammad Reza Ejtehadi, and H. Stark. *Sci. Rep.*, 6:37670, 2016.
- [79] H. C. Berg and D. A. Brown. *Nature*, 239:500, 1972.
- [80] H. C. Berg. *Nature*, 254:389, 1975.
- [81] H. C. Berg. *E. coli in Motion*. Biological and Medical Physics, Biomedical Engineering. Springer, New York, NY, 2004.
- [82] T. F. F. Farage, P. Krininger, and J. M. Brader. *Physical Review E*, 91:042310, 2015.

- [83] Étienne Fodor, Cesare Nardini, M. E. Cates, J. Tailleur, Paolo Visco, and Frédéric van Wijland. *Physical Review Letters*, 117:038103, 2016.
- [84] L. L. Bonilla. *Physical Review E*, 100:022601, 2019.
- [85] David Martin, Jérémie O’Byrne, M. E. Cates, Étienne Fodor, Cesare Nardini, J. Tailleur, and Frédéric van Wijland. *arXiv:2008.12972 [cond-mat.stat-mech]*, 2020.
- [86] Lorenzo Caprini, Umberto Marini Bettolo Marconi, Andrea Puglisi, and Angelo Vulpiani. *J. Stat. Mech. Theory Exp.*, 2019:053203, 2019.
- [87] Claudio Maggi, Umberto Marini Bettolo Marconi, Nicoletta Gnan, and Roberto Di Leonardo. Multidimensional stationary probability distribution for interacting active particles. *Sci. Rep.*, 5:10742, 2015.
- [88] B. Liebchen and Demian Levis. *Physical Review Letters*, 119:058002, 2017.
- [89] M. Mijalkov and G. Volpe. *Soft Matter*, 9:6376, 2013.
- [90] G. Volpe, S. Gigan, and G. Volpe. *Am. J. Phys.*, 82:659, 2014.
- [91] Zhi-Feng Huang, Andreas M. Menzel, and H. Löwen. *Physical Review Letters*, 125:218002, 2020.
- [92] Raúl Josué Hernández, Francisco J. Sevilla, Alfredo Mazzulla, Pasquale Pagliusi, Nicola Pellizzi, and Gabriella Ciparrone. *Soft Matter*, 16:7704, 2020.
- [93] Guo-Jun Liao and Sabine H. L. Klapp. *Soft Matter*, 14:7873, 2018.
- [94] Jens Bickmann, Stephan Bröker, Julian Jeggle, and Raphael Wittkowski. *arXiv:2010.05262v1 [cond-mat.soft]*, 2020.
- [95] E. Lauga, W. R. DiLuzio, G. M. Whitesides, and H. A. Stone. *Biophys. J.*, 90:400, 2006.
- [96] R. Di Leonardo, D. Dell’Arciprete, L. Angelani, and V. Iebba. *Physical Review Letters*, 106:038101, 2011.
- [97] Ting-Wei Su, Inkyum Choi, Jiawen Feng, Kelvin Huang, Euan McLeod, and Aydogan Ozcan. *Sci. Rep.*, 3:1664, 2013.
- [98] Jens Elgeti, U. Benjamin Kaupp, and Gerhard Gompper. *Biophys. J.*, 99:1018, 2010.
- [99] Felix Kümmerl, Borge ten Hagen, Raphael Wittkowski, Ivo Buttinoni, Ralf Eichhorn, G. Volpe, H. Löwen, and Clemens Bechinger. *Physical Review Letters*, 110:198302, 2013.
- [100] Yogesh Shelke, N. R. Srinivasan, Sumesh P. Thampi, and Ethayaraja Mani. *Langmuir*, 35:4718, 2019.
- [101] Bo Zhang, Andrey Sokolov, and Alexey Snezhko. *Nature Communications*, 11:4401, 2020.
- [102] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. *Physical Review Letters*, 75:1226, 1995.
- [103] Francesco Ginelli. *European Physical Journal – Special Topics*, 225:2099, 2016.
- [104] F. Peruani, A. Deutsch, and M. Bär. *European Physical Journal – Special Topics*, 157:111, 2008.
- [105] F. D.C. Farrell, M. C. Marchetti, D. Marenduzzo, and J. Tailleur. *Physical Review Letters*, 108:248101, 2012.
- [106] Oleksandr Chepizhko and F. Peruani. *Physical Review Letters*, 111:160604, 2013.
- [107] A. Ahmadi, T. B. Liverpool, and M. C. Marchetti. *Physical Review E*, 72:060901, 2005.
- [108] Aphrodite Ahmadi, M. C. Marchetti, and T. B. Liverpool. *Physical Review E*, 74:061913, 2006.
- [109] Hugues Chaté, Francesco Ginelli, and Raúl Montagne. *Physical Review Letters*, 96:180602, 2006.

- [110] Markus Bär, Robert Großmann, Sebastian Heidenreich, and F. Peruani. *Annu. Rev. Condens. Matter Phys.*, 11:441, 2020.
- [111] Christina Kurzthaler, Clémence Devailly, Jochen Arlt, Thomas Franosch, W. C. K. Poon, Vincent A. Martinez, and A. T. Brown. *Physical Review Letters*, 121:078001, 2018.
- [112] Claudio B Caporusso, Pasquale Digregorio, Demian Levis, Leticia F. Cugliandolo, and G. Gonnella. *Physical Review Letters*, 125:178004, 2020.
- [113] Joakim Stenhammar, A. Tiribocchi, Rosalind J. Allen, D. Marenduzzo, and M. E. Cates. *Physical Review Letters*, 111:145702, 2013.
- [114] Pasquale Digregorio, Demian Levis, A. Suma, Leticia F. Cugliandolo, G. Gonnella, and Ignacio Pagonabarraga. *Physical Review Letters*, 121:098003, 2018.
- [115] B. Liebchen, D. Marenduzzo, and M. E. Cates. *Physical Review Letters*, 118:268001, 2017.
- [116] Suropriya Saha, R. Golestanian, and Sriram Ramaswamy. *Physical Review E*, 89:062316, 2014.
- [117] Oliver Pohl and H. Stark. *Physical Review Letters*, 112:238303, 2014.
- [118] H. Stark. *Acc. Chem. Res.*, 51:2681, 2018.
- [119] B. Liebchen, D. Marenduzzo, Ignacio Pagonabarraga, and M. E. Cates. *Physical Review Letters*, 115:258301, 2015.
- [120] B. Liebchen and H. Löwen. *Acc. Chem. Res.*, 51:2982, 2018.
- [121] Eva Kanso and Sébastien Michelin. *J. Chem. Phys.*, 150:044902, 2019.
- [122] P. C. Hohenberg and B. I. Halperin. *Reviews of Modern Physics*, 49:435, 1977.
- [123] M. Reza Shaebani, Adam Wysocki, Roland G. Winkler, Gerhard Gompper, and Heiko Rieger. *Nat. Rev. Phys.*, 2:181, 2020.
- [124] Raphael Wittkowski, A. Tiribocchi, Joakim Stenhammar, Rosalind J. Allen, D. Marenduzzo, and M. E. Cates. *Nature Communications*, 5:4351, 2014.
- [125] Julian Bialké, H. Löwen, and Thomas Speck. *EPL*, 103:30008, 2013.
- [126] Hannes Risken. *The Fokker-Planck Equation*. Springer Series in Synergetics. Springer, Berlin, Heidelberg, 1984.
- [127] Jean Pierre Hansen and Ian R. McDonald. *Theory of Simple Liquids*. Elsevier, Burlington, 3rd edition, 2006.
- [128] Eric Bertin, Michel Droz, and Guillaume Grégoire. *J. Phys. A Math. Theor.*, 42:445001, 2009.
- [129] David S. Dean. *J. Phys. A. Math. Gen.*, 29:L613, 1996.
- [130] Eric Bertin, Hugues Chaté, Francesco Ginelli, Shradha Mishra, Anton Peshkov, and Sriram Ramaswamy. *New Journal of Physics*, 15:085032, 2013.
- [131] CW Gardiner. *Stochastic methods*. Springer, Berlin Heidelberg, 4 edition, 2009.
- [132] Sergio Chibbaro and Jean Pierre Minier. *Stochastic Methods in Fluid Mechanics*, volume 548 of *CISM International Centre for Mechanical Sciences*. Springer, Vienna, 2014.
- [133] Thomas Ihle. *Physical Review E*, 83:030901, 2011.
- [134] Thomas Ihle. *Phys. Chem. Chem. Phys.*, 11:9667, 2009.
- [135] E. M. Purcell. *Am. J. Phys.*, 45:3, 1977.

- [136] Georg Wolschin. *Hydrodynamik*. Springer, Berlin, Heidelberg, 2016.
- [137] Tian Qiu, Tung-Chun Lee, Andrew G. Mark, Konstantin I. Morozov, Raphael Münster, Otto Mierka, Stefan Turek, Alexander M. Leshansky, and Peer Fischer. *Nature Communications*, 5:5119, 2014.
- [138] Nicolle Rager Fuller. How Escherichia coli Move. https://www.nsf.gov/news/mmg/media/images/ecoli_h1.jpg. (visited on 2020-12-08).
- [139] Louisa Howard. Green algae Chlamydomonas reinhardtii (Image 3). https://www.nsf.gov/news/mmg/media/images/PF3393_03_h.jpg. (visited on 2020-12-08).
- [140] Andreas Zöttl and H. Stark. *Journal of Physics: Condensed Matter*, 28:253001, 2016.
- [141] R. Golestanian and Armand Ajdari. *Physical Review E*, 77:036308, 2008.
- [142] F. Alouges, A. DeSimone, and A. Lefebvre. *Eur. Phys. J. E*, 28:279, 2009.
- [143] Babak Nasouri, Andrej Vilfan, and R. Golestanian. *Physical Review Fluids*, 4:073101, 2019.
- [144] Abdallah Daddi-Moussa-Ider, Maciej Lisicki, Christian Hoell, and H. Löwen. *J. Chem. Phys.*, 148, 2018.
- [145] Abdallah Daddi-Moussa-Ider, Maciej Lisicki, Arnold J T M Mathijssen, Christian Hoell, Segun Goh, Jerzy Bławzdziewicz, Andreas M. Menzel, and H. Löwen. *Journal of Physics: Condensed Matter*, 30:254004, 2018.
- [146] Qixuan Wang and Hans G. Othmer. *J. Math. Biol.*, 76:1699, 2018.
- [147] O. Silverberg, E. Demir, G. Mishler, B. Hosoume, N. Trivedi, C. Tisch, D. Plascencia, O. S. Pak, and I. E. Araci. *Bioinspir. Biomim.*, 15:064001, 2020.
- [148] J. M. Yeomans. *Riv. Nuovo Cim.*, 40:1, 2017.
- [149] Michael D. Graham. *Microhydrodynamics, Brownian Motion, and Complex Fluids*. Cambridge University Press, Cambridge, 2018.
- [150] Sangtae Kim and Seppo J. Karrila. *Microhydrodynamics*. Elsevier, 1991.
- [151] Roland G. Winkler and Gerhard Gompper. *Hydrodynamics in Motile Active Matter*, page 1. Springer, Cham, 2018.
- [152] Jeffrey S. Guasto, Karl A. Johnson, and J. P. Gollub. *Physical Review Letters*, 105:168102, 2010.
- [153] David Saintillan and Michael J. Shelley. *J. R. Soc. Interface*, 9:571, 2012.
- [154] S. Babel, H. Löwen, and A. M. Menzel. *EPL*, 113:58003, 2016.
- [155] R. W. Nash, R. Adhikari, J. Tailleur, and M. E. Cates. *Physical Review Letters*, 104:258101, 2010.
- [156] Saverio E. Spagnolie and E. Lauga. *J. Fluid Mech.*, 700:105, 2012.
- [157] R. W. Nash, R. Adhikari, and M. E. Cates. *Physical Review E*, 77:026709, 2008.
- [158] Matthew T. Downton and H. Stark. *Journal of Physics: Condensed Matter*, 21:204101, 2009.
- [159] Ingo O. Götze and Gerhard Gompper. *Physical Review E*, 82:041921, 2010.
- [160] Kenta Ishimoto and Eamonn A. Gaffney. *Physical Review E*, 88:062702, 2013.
- [161] Andreas Zöttl and H. Stark. *Physical Review Letters*, 112:118101, 2014.
- [162] Johannes Blaschke, Maurice Maurer, Karthik Menon, Andreas Zöttl, and H. Stark. *Soft Matter*, 12:9821, 2016.

- [163] Jan-Timm Kuhr, Johannes Blaschke, Felix Rühle, and H. Stark. *Soft Matter*, 13:7548, 2017.
- [164] Andreas Zöttl and H. Stark. *Eur. Phys. J. E*, 41:61, 2018.
- [165] Michael Kuron, Philipp Stärk, Christian Burkard, Joost de Graaf, and Christian Holm. *J. Chem. Phys.*, 150:144110, 2019.
- [166] Kai Qi, Elmar Westphal, Gerhard Gompper, and Roland G. Winkler. *Physical Review Letters*, 124:068001, 2020.
- [167] Arne W. Zantop and H. Stark. *Soft Matter*, 16:6400, 2020.
- [168] M. N. Popescu, W. E. Uspal, Z. Eskandari, M. Tasinkevych, and S. Dietrich. *Eur. Phys. J. E*, 41:145, 2018.
- [169] Michael Kuron, Philipp Stärk, Christian Holm, and Joost de Graaf. *Soft Matter*, 15:5908, 2019.
- [170] I. Llopis and I. Pagonabarraga. *J. Nonnewton. Fluid Mech.*, 165:946, 2010.
- [171] Ignacio Pagonabarraga. *Lattice Boltzmann Modeling of Complex Fluids: Colloidal Suspensions and Fluid Mixtures*, page 279. Springer, 2004.
- [172] Timm Krüger, Halim Kusumaatmaja, Alexandr Kuzmin, Orest Shardt, Goncalo Silva, and Erlend Magnus Viggen. *The Lattice Boltzmann Method*. Graduate Texts in Physics. Springer International Publishing, Cham, 2017.
- [173] Livio Nicola Carenza, G. Gonnella, Antonio Lamura, Giuseppe Negro, and A. Tiribocchi. *Eur. Phys. J. E*, 42:81, 2019.
- [174] Jean-Christophe Desplat, Ignacio Pagonabarraga, and Peter Bladon. *Comput. Phys. Commun.*, 134:273, 2001.
- [175] S. Ramachandran, P. B. Sunil Kumar, and I. Pagonabarraga. *Eur. Phys. J. E*, 20:151, 2006.
- [176] M. E. Cates, K Stratford, R Adhikari, P Stansell, J-C Desplat, I. Pagonabarraga, and A J Wagner. *Journal of Physics: Condensed Matter*, 16:S3903, 2004.
- [177] Joost de Graaf, Henri Menke, Arnold J. T. M. Mathijssen, Marc Fabritius, Christian Holm, and Tyler N. Shendruk. *J. Chem. Phys.*, 144:134106, 2016.
- [178] Gilberto Medeiros Kremer. *An Introduction to the Boltzmann Equation and Transport Processes in Gases*. Interaction of Mechanics and Mathematics. Springer, Berlin, Heidelberg, 2010.
- [179] P. L. Bhatnagar, E. P. Gross, and M. Krook. *Phys. Rev.*, 94:511, 1954.
- [180] G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler. *Multi-Particle Collision Dynamics: A Particle-Based Mesoscale Simulation Approach to the Hydrodynamics of Complex Fluids*, page 1. Springer, 2009.
- [181] Anatoly Malevanets and Raymond Kapral. *Mesoscopic Multi-particle Collision Model for Fluid Flow and Molecular Dynamics*, page 116. Springer, 2004.
- [182] José Ruiz-Franco, Diego Jaramillo-Cano, Manuel Camargo, Christos N. Likos, and Emanuela Zaccarelli. *J. Chem. Phys.*, 151:074902, 2019.
- [183] R.G. Winkler, M. Ripoll, K. Mussawisade, and G. Gompper. *Comput. Phys. Commun.*, 169:326, 2005.
- [184] R G Winkler, K Mussawisade, M Ripoll, and G Gompper. *Journal of Physics: Condensed Matter*, 16:S3941, 2004.
- [185] M Ripoll, K Mussawisade, R. G Winkler, and G Gompper. *EPL*, 68:106, 2004.
- [186] A Lamura, G Gompper, T. Ihle, and D. M Kroll. *EPL*, 56:319, 2001.

- [187] Andreas Zöttl. *Chinese Phys. B*, 29:074701, 2020.
- [188] Mu-Jie Huang, Jeremy Schofield, and Raymond Kapral. *New Journal of Physics*, 19:125003, 2017.
- [189] Pierre de Buyl and Raymond Kapral. *Nanoscale*, 5:1337, 2013.
- [190] Snigdha Thakur and Raymond Kapral. *J. Chem. Phys.*, 135:024509, 2011.
- [191] Yingzi Yang, Jens Elgeti, and Gerhard Gompper. *Physical Review E*, 78:061903, 2008.
- [192] Liyan Qiao, Mu-Jie Huang, and Raymond Kapral. *Phys. Rev. Res.*, 2:033245, 2020.
- [193] Dmitry A. Fedosov, Ankush Sengupta, and Gerhard Gompper. *Soft Matter*, 11:6703, 2015.
- [194] P. J. Hoogerbrugge and J. M. V. A Koelman. *EPL*, 19:155, 1992.
- [195] Robert D. Groot and Patrick B. Warren. *J. Chem. Phys.*, 107:4423, 1997.
- [196] Alexander Panchenko, Denis F. Hinz, and Eliot Fried. *Phys. Fluids*, 30:033301, 2018.
- [197] Shaltiel Eloul, W. C. K. Poon, O. Farago, and D. Frenkel. *Physical Review Letters*, 124:188001, 2020.
- [198] Ken Tokunaga and Ryo Akiyama. *Physical Review E*, 100:062608, 2019.
- [199] Sebastian Heidenreich, Jörn Dunkel, Sabine H. L. Klapp, and Markus Bär. *Physical Review E*, 94:020601, 2016.
- [200] David Saintillan and Michael J. Shelley. *Physical Review Letters*, 100:178103, 2008.
- [201] David Saintillan and Michael J. Shelley. *Comptes Rendus Phys.*, 14:497, 2013.
- [202] Joakim Stenhammar, Cesare Nardini, Rupert W. Nash, D. Marenduzzo, and Alexander Morozov. *Physical Review Letters*, 119:028005, 2017.
- [203] A. Tiribocchi, Raphael Wittkowski, D. Marenduzzo, and M. E. Cates. *Physical Review Letters*, 115:188302, 2015.

Chapter 16

Molecular Dynamics

AGNESE CALLEGARI, GIORGIO VOLPE, GIOVANNI VOLPE

Classical molecular dynamics is a set of numerical methods to study many-particle systems. These numerical methods are extensively employed in physics, chemistry, biophysics and materials science to study both microscopic (e.g., molecules, polymers, biomolecules, membranes) and macroscopic systems (e.g., gases, liquids and solids, phase diagrams, phase transitions), some examples of which are shown in Fig. 16.1. The basic idea of a molecular dynamics simulation is to resolve the Newton equations of motion for each individual particle composing the many-particle system [1, 2, 3, 4]. Particles are subject to both external fields and mutual interactions via first-principle (e.g., gravitational, electrostatic) and effective (e.g., Van der Waals, screened Coulomb interactions) potentials. This approach works even for systems that are not prototypical classical systems, such as atoms and molecules [5, 6, 7]. For instance, one can effectively use classical dynamics to predict the three-dimensional structure of complex biomolecules such as proteins [8]. Although understanding similar molecular features would typically require advanced concepts of quantum mechanics, in the molecular dynamics approach, the protein is divided in different units, their interaction is modeled with an appropriate effective potential [9], and finally these units are allowed to evolve according to the laws of classical dynamics. Thus, molecular dynamics bridge statistical physics concepts to basic physics notions from classical mechanics and are, therefore, a good starting point for the more advanced numerical methods employed in active matter that we will explore in the following chapters. In this Chapter, we will first present the basic algorithms of classical molecular dynamics, i.e., the Euler algorithm and the leapfrog algorithm. To illustrate the implementation of each method with its strengths and weaknesses, we will use these algorithm to simulate problems with well-known analytical solutions as the harmonic oscillator in one dimension. We will then extend these methods

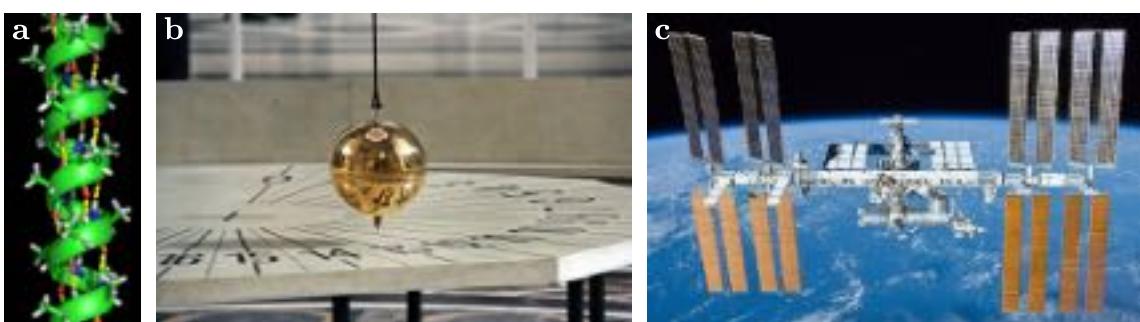


Figure 16.1: **Molecular dynamics across the scales.** Examples of systems studied with molecular dynamics at different length scales: (a) the α -helix of a protein (microscopic scale); (b) the Foucault pendulum at the Pantheon in Paris (macroscopic scale); and (c) the orbiting international space station, a prelude to the dynamics of celestial bodies and astronomical scales. Sources: From the original picture (a) [alpha-helix](#) under CC BY-SA 3.0 licence, (b) by [Misko Flickr](#) under CC BY 2.0 licence, and (c) [International Space Station](#) by NASA.

to two-dimensional problems for systems with many particles. We will also see how stochastic behavior (e.g., Brownian motion) can emerge from simulating a system of many particles regulated by deterministic Newton's laws. Finally, we will discuss how to implement the appropriate conservation laws or set the boundary conditions to control a numerical simulation and make sure it reproduces faithfully the behavior of the system of interest in the desired experimental conditions, e.g., constant temperature or constant pressure, or the interaction with a boundary with or without some heat exchange.

16.1 Basic algorithms

Let us consider a system of N classical particles in three dimensions (Fig. 16.2). For example, you may think of the gas molecules inside a party balloon. Each particle i traces out a trajectory in space, which in general depends on the collisions with the other particles and with the container's walls, and usually also depends on the action of external forces, such as gravity. One of the fundamental problems of classical mechanics is to determine the time evolution of such a system. At any instant in time, its state is described by three spatial coordinates $\mathbf{r}_i = [r_{x,i}, r_{y,i}, r_{z,i}]$ and by three momenta $\mathbf{p}_i = m_i \mathbf{v}_i = m_i [v_{x,i}, v_{y,i}, v_{z,i}]$, where m_i and \mathbf{v}_i are the mass and velocity of particle i , respectively. To describe the positions and momenta of the N particles, we need two $3N$ -dimensional vectors: $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_N]$ and $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_N]$. The $3N$ -dimensional space of positions \mathbf{R} is called *configuration space*, the $3N$ -dimensional space of momenta \mathbf{P} is called *momentum space* and the $6N$ -dimensional space $[\mathbf{R}, \mathbf{P}]$ is called *phase space*.

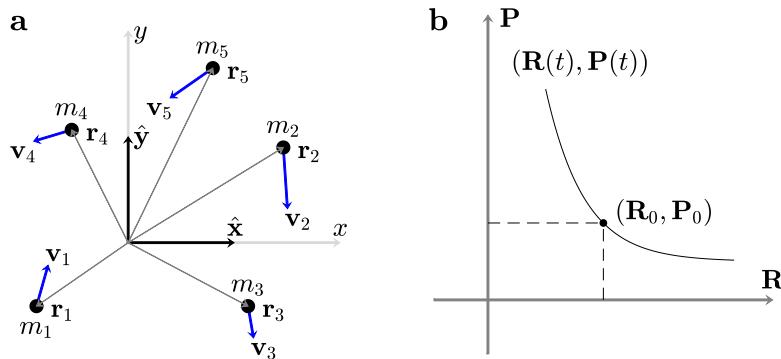


Figure 16.2: **Representation of a many-particle system.** (a) Schematic representation of a many-particle system with point-like particles with mass m_i , position \mathbf{r}_i , and velocities \mathbf{v}_i . (b) A two-dimensional representation of its time evolution in phase space.

Knowing the initial positions \mathbf{r}_{0i} and velocities \mathbf{v}_{0i} of the particles, it is possible to determine their time evolution (i.e., the trajectories $\mathbf{r}_i(t)$) at any instant t using Newton's laws:

$$\begin{cases} \frac{d^2\mathbf{r}_i}{dt^2}(t) &= \frac{1}{m_i} \left(\mathbf{F}_i^{\text{ext}} + \sum_{j \neq i} \mathbf{F}_{ij} \right), \\ \mathbf{r}_i(0) &= \mathbf{r}_{0i}, \\ \frac{d\mathbf{r}_i}{dt}(0) &= \mathbf{v}_{0i}, \end{cases} \quad (16.1)$$

where $\frac{d\mathbf{r}_i}{dt}(t)$ and $\frac{d^2\mathbf{r}_i}{dt^2}(t)$ are respectively the instantaneous velocity $\mathbf{v}_i(t)$ and acceleration $\mathbf{a}_i(t)$ of the i -th particle, \mathbf{F}_{ij} is a force representing the mutual interaction between particles i and j , and $\mathbf{F}_i^{\text{ext}}$ is an external force acting on particle i .

The set of equations in Eq. (16.1) includes the second-order derivative of the positions $\mathbf{r}_i(t)$. Without loss of generality, if we also consider the velocities $\mathbf{v}_i(t)$ as independent variables, we can transform Eq. (16.1) into

a set of equations constituted only by first-order derivatives:

$$\left\{ \begin{array}{l} \frac{d\mathbf{r}_i}{dt}(t) = \mathbf{v}_i(t), \\ \frac{d\mathbf{v}_i}{dt}(t) = \frac{1}{m_i} \left(\mathbf{F}_i^{\text{ext}} + \sum_{j \neq i} \mathbf{F}_{ij} \right), \\ \mathbf{r}_i(0) = \mathbf{r}_{0i}, \\ \mathbf{v}_i(0) = \mathbf{v}_{0i}. \end{array} \right. \quad (16.2)$$

In order to understand how we can solve this set of equations numerically, we can start considering the simpler case of a single particle in one dimension. Equations (16.2) in this case reduce to

$$\left\{ \begin{array}{l} \frac{dx}{dt}(t) = v(t), \\ \frac{dv}{dt}(t) = \frac{1}{m} F^{\text{ext}}, \\ x(0) = x_0, \\ v(0) = v_0, \end{array} \right. \quad (16.3)$$

where the mutual interaction has been dropped¹. Often (but not always), an exact solution for $x(t)$ can be determined analytically. In general, whether an analytical solution can be determined or not, we suppose that the problem in Eq. (16.3) is well defined, i.e., its solution $x(t)$ exists and is unique for all times t at which we are trying to retrieve it numerically. We are therefore trying to reconstruct the regular function $x(t)$ with a finite sequence of discrete values x_1, x_2, \dots, x_n at the times t_1, t_2, \dots, t_n . Ideally, we would like to use very small time increments $\Delta t_k = t_{k+1} - t_k$ in order to approximate our solution $x(t)$ as well as possible.

16.1.1 Euler algorithm

We now describe the simplest algorithm to calculate the time evolution of the positions and momenta in phase space. Having chosen an appropriate (i.e., small enough) time step Δt , we can reconstruct the function $x(t_n)$ evaluated at regular time steps ($t_n = n\Delta t$) using a *finite-difference* method known as the *Euler's algorithm*.

Assuming both the trajectory $x(t)$ and its first-order derivative to be continuous, we can write the Taylor series of Eq. (16.3) truncated to the first order as

$$\left\{ \begin{array}{l} x(t_0 + \Delta t) = x(t_0) + \frac{dx}{dt}(t_0)\Delta t + O(\Delta t^2) \\ \quad = x(t_0) + v(t_0)\Delta t + O(\Delta t^2), \\ v(t_0 + \Delta t) = v(t_0) + \frac{dv}{dt}(t_0)\Delta t + O(\Delta t^2) \\ \quad = v(t_0) + \frac{F(t_0)}{m}\Delta t + O(\Delta t^2), \end{array} \right. \quad (16.4)$$

where $O(\Delta t^2)$ is an infinitesimal of order two. From Eq. (16.4), we define the Euler's algorithm as

$$\left\{ \begin{array}{l} x_{n+1} = x_n + v_n \Delta t, \\ v_{n+1} = v_n + \frac{F_n}{m} \Delta t. \end{array} \right. \quad (16.5)$$

¹Even if the mutual interaction is crucial in many-particle systems, nothing is lost in the illustration of the basic numerical techniques by not considering it.

The force F_n is calculated from the position and the velocity at time t_n . In general, the force can depend both on the position (e.g., an elastic force and a gravitational force), on the velocity (e.g., a viscous force in a fluid), and also explicitly on time (e.g., a periodic force). For conservative forces, if an analytical expression for the potential is available, then the force is simply calculated as

$$F(x_n) = -\nabla U(x_n) = -\frac{\partial U(x_n)}{\partial x}. \quad (16.6)$$

Otherwise, it is directly estimated numerically as F_n :

$$F(x_n) = -\frac{U(x_n + \Delta x) - U(x_n - \Delta x)}{2\Delta x}, \quad (16.7)$$

in one dimension, where Δx is a very small increment of the position. This expression can be easily generalized to two and three dimensions.

Euler's method is a *first-order method*, i.e., the global error in estimating the true equation of motion $x(t)$ via this method starting from $x(0) = x_0$ with a time step Δt scales as Δt .

Example 16.1: Harmonic oscillator with Euler's algorithm. Consider a point-like mass m attached to a perfect spring with elastic constant k . The mass is set on a horizontal, frictionless surface. Initially, the mass is at position x_0 with velocity v_0 . The equation of the motion is:

$$m \frac{d^2x}{dt^2} = -kx \quad (16.8)$$

a. Solve the problem analytically. What is the trajectory of the particle as a function of time? Show that $x(t) = A \cos(\omega t + \theta)$, with $\omega = \sqrt{k/m}$, $A = \sqrt{x_0^2 + (v_0/\omega)^2}$, and θ defined by $\cos \theta = \frac{x_0}{A}$, $\sin \theta = -\frac{v_0}{A\omega}$. Also show that $v(t) = -\omega A \sin(\omega t + \theta)$.

b. Write a program that propagates the trajectory with the Euler method. What is a good value for Δt ? How does it compare with the period of the motion T , which is the inverse of the characteristic frequency of the motion, $v = \omega/(2\pi)$?

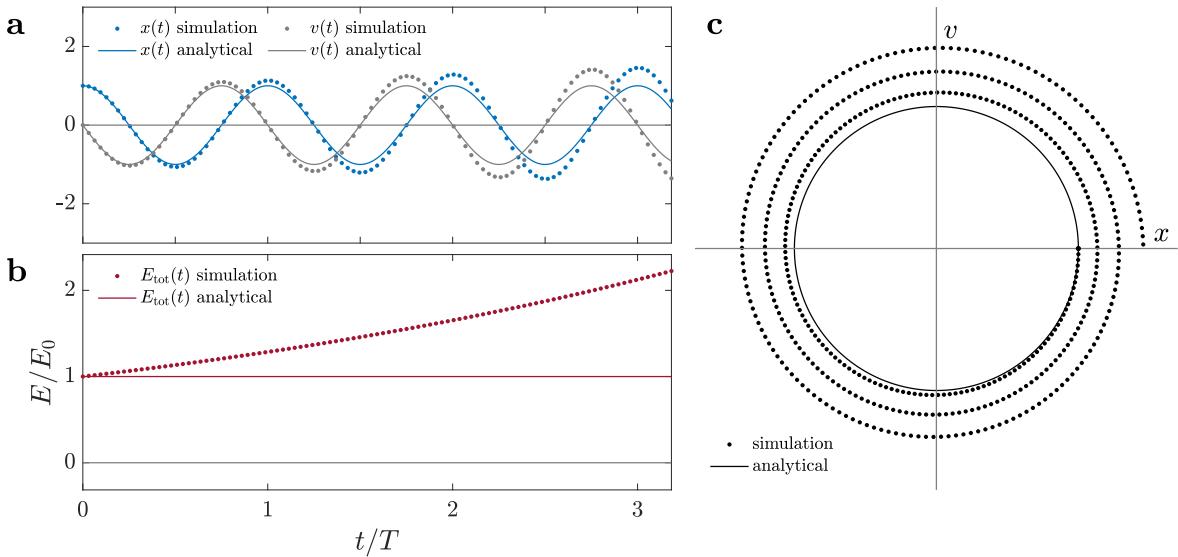


Figure 16.3: Harmonic oscillator with Euler's algorithm. The numerical solutions (dots) for (a) the position $x(t)$ and the velocity $v(t)$, and for (b) the total energy $E_{\text{tot}}(t)$ for the simple harmonic oscillator (Eq. (16.8), Example 16.1) compared to the respective analytical solutions (lines). (c) Representation of the solution in phase space. The Euler's algorithm introduces a drift in the energy and, no matter how small the time step of the simulation Δt , the energy eventually increases.

c. Calculate the total energy of the particle as a function of time, analytically and in simulation. Compare them.

d. Is the total mechanical energy conserved in simulation?

e. Show numerically and analytically that, no matter how small the Δt is taken, the total mechanical energy is never conserved and will increase in time without limit.

[You can compare your results with those shown in Fig. 16.3.]

Example 16.2: Multidimensional Euler's algorithm. Derive the Euler's algorithm for N particles in two and three dimensions.

16.1.2 Leapfrog algorithm

As shown in Fig. 16.3, the Euler's algorithm gives rise to an energy drift. Example 16.1 shows that the solution provided by the algorithm eventually deviates from the analytical solution, no matter how small we choose Δt . This is due to the fact that the force is calculated using the position at the beginning of each time interval $[t_n, t_{n+1}]$ leading to a systematic error in the estimation of the change in velocity. A much better approximation is obtained if one uses also the position in the middle of the time interval $[t_n, t_{n+1}]$. In the *leapfrog algorithm*, the position is first propagated by a half timestep to obtain $x_{n+\frac{1}{2}}$, then the velocity v_{n+1} is calculated using $F_{n+\frac{1}{2}}$, and finally the position is propagated by another half time step to x_{n+1} :

$$\left\{ \begin{array}{l} x_{n+\frac{1}{2}} = x_n + v_n \frac{\Delta t}{2}, \\ v_{n+1} = v_n + \frac{F(x_{n+\frac{1}{2}})}{m} \Delta t, \\ x_{n+1} = x_{n+\frac{1}{2}} + v_{n+1} \frac{\Delta t}{2}, \end{array} \right. \quad (16.9)$$

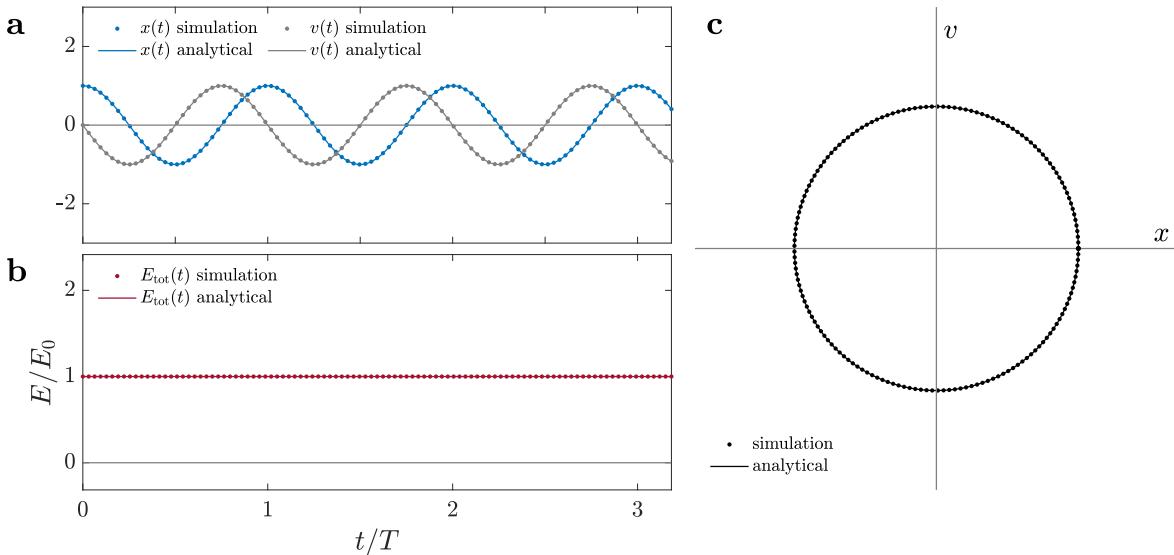


Figure 16.4: **Harmonic oscillator with the leapfrog algorithm.** The numerical solutions (dots) for (a) the position $x(t)$ and the velocity $v(t)$, and for (b) the total energy for the simple harmonic oscillator (Eq. (16.8), Example 16.4) compared to the respective analytical solutions (lines). (c) Solution plotted in phase space. The leapfrog algorithm does not introduce any drift in the energy and the discrepancy from the analytical solution is negligible.

Example 16.3: Precision of the leapfrog algorithm. Consider the leapfrog algorithm given in Eq. (16.9). Write x_{n+1} as a function of x_n , v_n , F_n , m , and Δt only. Show that

$$x_{n+1} = x_n + v_n \cdot \Delta t + \frac{1}{2} \cdot \frac{F_n}{m} \cdot \Delta t^2 + O(\Delta t^3) \quad (16.10)$$

where $O(\Delta t^3)$ is an infinitesimal of order 3.

Example 16.4: Harmonic oscillator with leapfrog algorithm. Consider a point-like mass m attached to a perfect spring with elastic constant k like in Example 16.1.

- a. Write a program that propagates the trajectory with the leapfrog method.
- b. Compare the simulated trajectory with the analytical trajectory, and the trajectory simulated with the Euler method. What do you observe?
- c. Compare the total energy of the analytical and simulation solutions. Which methods seems to have better performance?
- d. Try the leapfrog method for different values of $\Delta t < T = \frac{2\pi}{\omega}$. Show numerically that the error on the calculation of the energy remains always below a certain threshold and does not increase in time, if $\Delta t \ll T$.
- e. Show numerically and analytically that the global error (after n steps) on the position and on the velocity is of the order of $n(\Delta t)^2$ [Is this correct? Do you need the brackets?], and thus, the smaller the Δt , the more precise the numerical total energy $E_n = \frac{m}{2}v_n^2 + \frac{k}{2}x_n^2$.

[You can compare your results with those shown in Fig. 16.4.]

Example 16.5: Multidimensional leapfrog algorithm. Derive the leapfrog algorithm for N particles in two and three dimensions.

From the previous exercises on the harmonic oscillator, the leapfrog method should be preferred, because it does not introduce any drift in the energy, even for those choices of the time step that create a conspicuous drift using Euler's algorithm (Fig. 16.4). This behavior is related to the concept of time reversibility. Newton's equations, and in fact the whole of classical physics, are time-reversible. This means that by changing the sign of the velocities of a particle at a certain time, the solution exactly back-traces the trajectory in phase space. Energy conservation and time reversibility are different sides of the same coin. In molecular dynamics algorithms, it is very important to implement correctly the time reversibility of the trajectory [10] to avoid numerical problems in the solution. In a numerical sense, reversibility means the following: first, generate a trajectory by propagating a particle's position and velocity for a number of time steps N ; then, invert the sign of all velocities and propagate again for other N time steps; if the new trajectory exactly retraces the original trajectory, then the algorithm is *reversible*.²

Example 16.6: Time reversibility. Verify that the Euler algorithm is not reversible, while the leapfrog algorithm is reversible, as described in the text above.

Finally, it is also important to note that, if the force depends on the velocity as well as on the position (e.g., in presence of a viscous friction), then the total energy of the system is not conserved, as non-conservative forces are acting on the particles. We will see later in the Chapter how this feature can emerge in systems of many particles if we describe the system through an effective model, even in the presence of only mutual conservative interactions between particles. Moreover, it must be noted that, in the presence of non-conservative, dissipative forces, the Euler method performs better, especially with Δt small enough to balance the unphysical increase in total energy with the dissipation term.

²To be more precise, it does not need to *exactly* retrace the original trajectory: it is enough that it retraces it with a global error scaling like $(\Delta t)^2$. [check this!](#)

16.1.3 Velocity Verlet algorithm

Another widely used integration scheme is the *velocity Verlet algorithm* [7], because of its properties of being drift-free and introducing only errors in the order of Δt^4 in the calculation of the position. Let us see how this method is derived and why it is widely preferred.

We start from the finite difference equations expressing the Taylor series expansion of $x(t)$ to obtain x_{n+1} and x_{n-1} from x_n , v_n , a_n , b_n , where b_n is the third derivative of the position $x(t)$ in $t_n = n\Delta t$, i.e., the rate of change of the acceleration (also known as *jolt* or *jerk*):

$$\begin{cases} x_{n+1} &= x_n + v_n \Delta t + \frac{1}{2} a_n (\Delta t)^2 + \frac{1}{6} b_n (\Delta t)^3 + O(\Delta t^4), \\ x_{n-1} &= x_n - v_n \Delta t + \frac{1}{2} a_n (\Delta t)^2 - \frac{1}{6} b_n (\Delta t)^3 + O(\Delta t^4), \end{cases} \quad (16.11)$$

where $O(\Delta t^4)$ is an infinitesimal of order Δt^4 . Summing and subtracting the two equations in the system of Eq. (16.11) and rearranging the terms, we obtain:

$$\begin{cases} x_{n+1} &= 2x_n - x_{n-1} + a_n (\Delta t)^2, \\ v_n &= \frac{x_{n+1} - x_{n-1}}{2\Delta t}, \end{cases} \quad (16.12)$$

which is the *Verlet algorithm* that has a precision of $O(\Delta t^4)$ on the position and $O(\Delta t^2)$ on the velocity. An equivalent formulation of the Verlet algorithm is the *velocity Verlet algorithm*, represented by the set of equations:

$$\begin{cases} x_{n+1} &= x_n + v_n \Delta t + \frac{1}{2} a_n (\Delta t)^2, \\ v_{n+1} &= v_n + \frac{a_{n+1} + a_n}{2} \Delta t. \end{cases} \quad (16.13)$$

In the velocity Verlet algorithm, we keep track of both the velocity v_n and the position x_n at each time step t_n , while, in the Verlet algorithm, we keep track of the two last positions x_n and x_{n-1} . The acceleration $a_n = F_n/m$ is found as a function of the position through the force, either with an explicit formula $F_n = F(x_n)$ or obtained via the gradient of the potential (Eq. 16.7), for both cases.³

Example 16.7: Harmonic oscillator with velocity Verlet algorithm. Simulate a one-dimensional harmonic oscillator, like in Exercises 16.1 and 16.4. Show that the velocity Verlet algorithm performs similarly to the leapfrog algorithm.

Example 16.8: Multidimensional velocity Verlet algorithm. Derive the velocity Verlet algorithm for N particles in two and three dimensions.

³The Verlet algorithm and its equivalent velocity Verlet algorithm are called *symplectic* algorithms because their formulation is compatible with the *canonical transformations* of the *Hamiltonian formalism*, and thus preserves the *symplectic 2-form* $dp \wedge dq$. Although this is an important concept for a thorough mathematical definition of the Hamiltonian formalism, we will not discuss it here, as it is out of the scope of this Book. The interested reader can refer to classical texts of Hamiltonian mechanics, differential geometry and tensor calculus such as Refs. [11]. Differently from the Verlet, velocity Verlet and leapfrog algorithms, the Euler's algorithm and other integration schemes such as Runge-Kutta are not symplectic (for the Runge-Kutta family of integration scheme see, for instance, Ref. [12]). Symplectic integrators have the advantage that they preserve (or nearly preserve) always the physically conserved quantities of a system (e.g., energy, momentum, angular momentum): the error in the estimated conserved quantities remains limited, and self-compensates in time. The global error of the Verlet and the velocity Verlet algorithm, i.e., the error accumulated after $n\Delta t$ from the starting condition (position and velocity), is of the order of $O(\Delta t^2)$ on both position and velocity, despite the local error on the position (i.e., the error done between two consecutive time steps) being of the order of $O(\Delta t^3)$ or less. This is because the position is calculated in a *cumulative* way, while the velocity is calculated in a *non-cumulative* fashion, see Eqs. (16.12)-(16.13). In a numerical simulation, the global error is a good measure of the accuracy of the method and is arguably the most important parameter.

16.2 Interacting particles

Molecular dynamics are often applied to systems of atoms or molecules subject to a potential, which depends both on the interaction between the particles and on the presence of external forces. Now that we have seen how to simulate a single particle, we are ready to study a system with many particles. The potential functions describing these interactions are typically obtained from fits to experiments or approximations to solutions of the full electronic problem. Often only scalar pair-interaction potentials $U_2(\Delta r_{ij})$ are considered, where Δr_{ij} is the distance between the two particles i and j ; in such a case, the total potential energy for a system consisting of N particles is given by

$$U(\mathbf{R}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N U_2(\Delta r_{ij}) = \sum_{i < j}^N U_2(\Delta r_{ij}). \quad (16.14)$$

More complex situations may require the inclusion of three-body or multi-body potentials. For example, for biological systems, such as proteins, there are databases that store potential functions between basic constituents such as chemical functional groups.

16.2.1 Hard-core potential

Hard spheres are impenetrable objects of radius ρ and their pair-potential is (Fig. 16.5a)

$$U_{\text{hc}}(\Delta r) = \begin{cases} 0 & \Delta r > 2\rho, \\ +\infty & \Delta r < 2\rho. \end{cases} \quad (16.15)$$

16.2.2 Yukawa-like potential

Hard-core potentials are often implemented using repulsive Yukawa-like potentials (Fig. 16.5b), which have the numerical advantage of being smooth:

$$U_Y(\Delta r) = \begin{cases} \frac{2\rho\epsilon}{\Delta r} \exp\left[-\kappa\left(\frac{\Delta r}{2\rho} - 1\right)\right] & \Delta r > 2\rho, \\ +\infty & \Delta r < 2\rho, \end{cases} \quad (16.16)$$

where $\epsilon > 0$ is the energy scale and $\kappa > 0$ controls the interaction range.

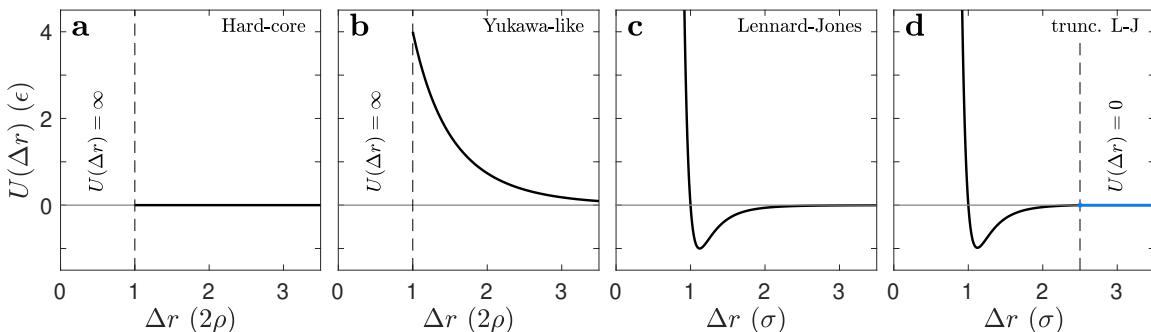


Figure 16.5: **Interaction potentials.** Most common interaction potentials: (a) hard-core potential (Eq. (16.15)). (b) Yukawa potential (Eq. (16.16)). (c) Lennard–Jones potential (Eq. (16.17)). (d) Truncated Lennard–Jones potential (Eq. (16.19)), where the truncation is represented with a dashed line (at this scale, the change of slope in the potential is not visible).

16.2.3 Lennard–Jones potential

The Lennard–Jones potential is a simple and convenient potential function, usually used for monatomic systems in the liquid and gas phases. The energy as a function of distance between the two particles is:

$$U_{\text{LJ}}(\Delta r) = 4\epsilon \left[\left(\frac{\sigma}{\Delta r} \right)^{12} - \left(\frac{\sigma}{\Delta r} \right)^6 \right], \quad (16.17)$$

where $\epsilon > 0$ corresponds to the well depth and σ sets the length scale.

The Lennard–Jones potential is an effective potential that describes the essential interaction between simple atoms and molecules [7]: two particles repel each other at very close distance, attract each other at intermediate distance, and do not interact at infinite distance, like shown in Fig. 16.5c.

It is usual to define the quantities in a Lennard–Jones system by means of the adimensional unit systems. The standard parameters defining a system of particles interacting via the Lennard–Jones potential are the depth of the potential energy ϵ , the particle size σ , the mass of a particle m . From these quantities, we can derive a force unit $F_0 = \epsilon/\sigma$, a time unit $t_0 = \sigma\sqrt{m/\epsilon}$, a velocity unit $v_0 = \sqrt{\epsilon/m}$, and a temperature unit $T_0 = \sqrt{\epsilon/k_B}$ where $k_B = 1.38064852 \cdot 10^{-23} \text{ J K}^{-1}$ is the Boltzmann constant. It is a standard procedure to define the simulation using these units. Therefore, distances and positions will be expressed in units of σ , energies in units of ϵ , velocities in terms of v_0 , time intervals in terms of t_0 , and so on. In the literature, the Lennard–Jones two-dimensional gas has been extensively investigated in terms of both molecular dynamics and Monte Carlo simulations. To refer to a particular system, one chooses the most appropriate values of ϵ , σ , and m . The simulation can be performed in terms of adimensional units and converted later to physical units in order to get the proper time, energy, and length scales.

Dimensionless units for a Lennard–Jones interacting system.

ϵ	energy scale
σ	length scale
m	particle mass
$F_0 = \epsilon/\sigma$	unit of force
$t_0 = \sigma\sqrt{m/\epsilon}$	unit of time
$v_0 = \sqrt{\epsilon/m}$	unit of velocity
$T_0 = \sqrt{\epsilon/k_B}$	unit of temperature

16.2.4 Truncated Lennard–Jones potential

The truncated Lennard–Jones is defined from the Lennard–Jones potential introducing the *cut-off* radius r_c (Fig. 16.5):

$$U_{\text{LJ}}^{\text{trunc}}(\Delta r) = \begin{cases} U_{\text{LJ}}(\Delta r) - U_{\text{LJ}}(r_c) & 0 < \Delta r \leq r_c, \\ 0 & \Delta r > r_c. \end{cases} \quad (16.19)$$

This definition sets to zero any interaction between particles that are more distant than the cut-off radius (i.e., $\Delta r > r_c$). Usually, $r_c = 2.5\sigma$. Therefore, for the calculation of the force on a given particle, only the particles that are closer than r_c are included, reducing the computation time of the force, which is the slow step of every molecular dynamics simulation. The truncated Lennard–Jones potential is continuous in $\Delta r = r_c$, but its derivative is not. Hence, the force $F_{\text{LJ}}(\Delta r) = -\frac{\partial U_{\text{LJ}}}{\partial r}(\Delta r)$ is not well-defined in $\Delta r = r_c$ and has a discontinuity of the first kind (a jump). Taking $r_c = 2.5\sigma$ makes the jump in the force very close to zero: $\Delta F_{\text{LJ}}(2.5\sigma) \approx \frac{\epsilon}{\sigma} \cdot 2.4 \cdot 10^{-4}$. To overcome the problem of the discontinuity in the force, a truncated and splined form of the Lennard–Jones potential can be used, where the potential is made derivable near r_c by using a spline function [13].

Example 16.9: Gas of identical particles interacting via the Lennard–Jones potential. Start by simulating a classical model, the gas of identical particles mutually interacting via the Lennard–Jones potential in Eq. (16.17) [14]:

- a. Show that the force corresponding to the potential in Eq. (16.17) can be written as:

$$F(r) = \frac{24\epsilon}{r} \left[2 \left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (16.20)$$

b. *Initializing the simulation.* Using either the leapfrog or the velocity Verlet algorithm (see Example 16.1f or 16.8) write a program to simulate N particles in two dimensions, enclosed in a squared arena $L \times L$, where L is an appropriate multiple of σ . Start from an initial configuration in which the N particles are set at random positions $\mathbf{r}_{0j} = (x_{0j}, y_{0j})$, with the only condition of being farther away than σ from each other and from the boundary of the arena. Set the initial velocity of the particles with a random orientation, and set its magnitude v in terms of a multiple (or a fraction) of $v_0 = \sqrt{\epsilon/m}$ depending on how much energy you want in the system.

c. *Force calculation.* As the interaction force rapidly diverges for $r \rightarrow 0$, special care should be used to choose the time step for the molecular dynamics simulation. The time step Δt should be small enough to account for the collisions of the particles accurately. In order to fix an upper bound to the value of Δt , consider the displacement occurring during a single time step Δt if the particle has a velocity v . We should make sure that $v\Delta t \ll \sigma$.

d. *Boundary.* We aim at simulating a real gas in two dimensions, in a closed rectangular container. In this case, we will have to implement the effect of the boundaries: when a particle hits the boundary, we suppose it bounces back without losing energy. To numerically implement this, we first define the location of our boundaries. If, for instance, the box containing the gas is a rectangle delimited by $x = 0$ and $x = L_x$ on the x-axis and $y = 0$ and $y = L_y$ on the y-axis, whenever a particle i crosses the boundary $x = L_x$ assuming a position $\mathbf{r}_i = (L_x + \Delta x_i, y_i)$ with a velocity $\mathbf{v}_i = (v_{xi}, v_{yi})$, then we set it back inside the container in the position $\mathbf{r}_{i,\text{refl}} = (L_x - \Delta x_i, y_i)$ and with velocity $\mathbf{v}_{i,\text{refl}} = (-v_{xi}, v_{yi})$. We implement the reflections on the other boundaries similarly.

e. *Time step and evolution in time.* Now that the initialization, the force calculation, and the effect of the boundary are implemented, try your simulation with $\Delta t = \alpha t_0$, where $t_0 = \sigma\sqrt{m/\epsilon}$ is the time dimensionless unit for a Lennard–Jones system (see Box 16.2.3) for different choices of α . Try, e.g., $\alpha = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$. Which α works better?

f. *Energy conservation, control of the numerical simulation.* Monitor the total energy $E = K + U_{\text{pot}}$ of the system

$$\begin{aligned} K &= \frac{1}{2}m \sum_{i=1}^N (\mathbf{v}_i)^2 && \text{kinetic energy} \\ U_{\text{pot}} &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N U(|\mathbf{r}_i - \mathbf{r}_j|) && \text{potential energy} \end{aligned} \quad (16.21)$$

as a function of time and compare simulations with different time steps and same initial condition over the same time interval.

[You can compare your results to Fig. 16.6. Furthermore, you should observe that in the reflection process, the total linear momentum of the system changes. The potential energy stored in the configuration may change slightly, as the position of some particles changes. The kinetic energy of the system before and after the interaction with the boundary, instead, does not change, because the speed is the same before and after the reflection. Depending on the density of particles in the systems and the time step of the simulation, the change in potential energy can be quite significant if, after the reflection, the reflected particle is set very close to another particle in proximity of the boundary. This can lead to a spurious sudden increase of the force acting on the involved particles and of the total energy of the system. We will see later in this Chapter how to control the amount of energy in a simulation (Section 16.3) and in the Chapter 17 on Brownian Motion how to deal with the numerical spurious superposition of particles (Section 17.5).]

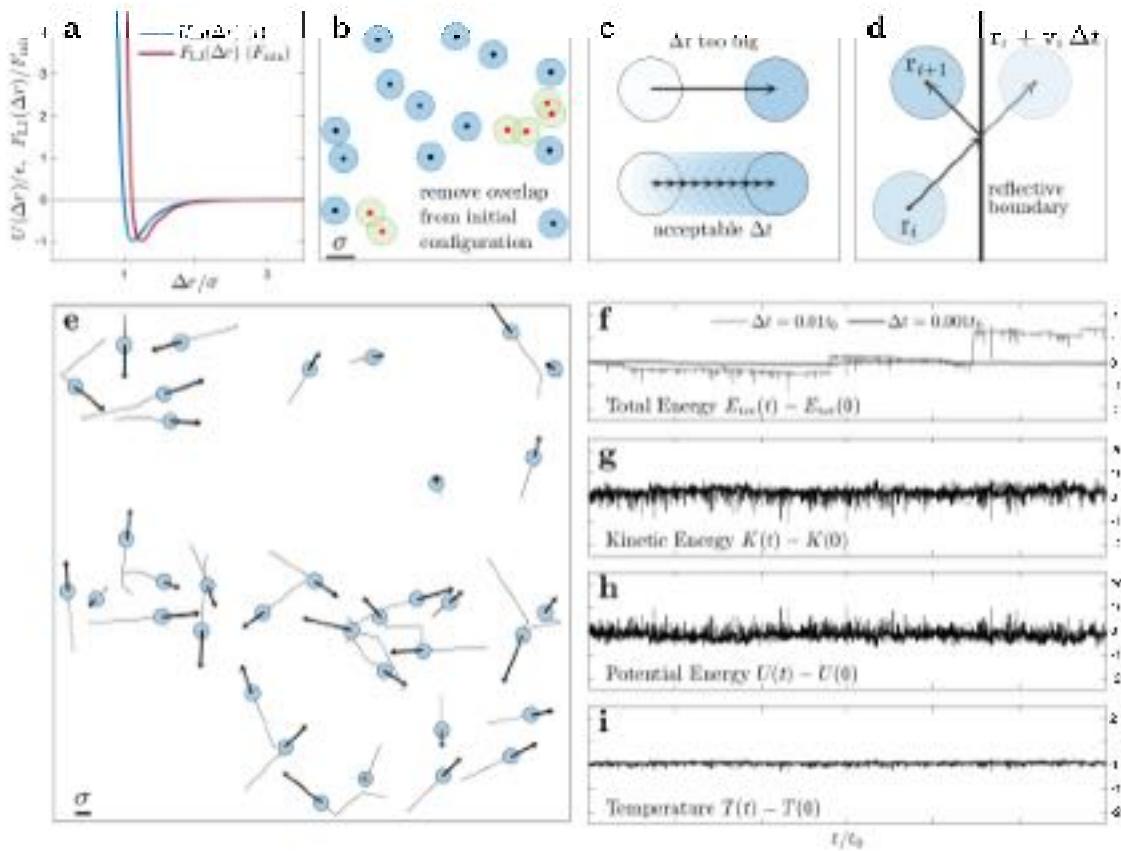


Figure 16.6: Two dimensional Lennard–Jones gas. (a) Lennard–Jones potential and force. (b) Initial configuration of the particles. The initial positions are generated at random within the simulation arena. It is good practice to remove the initial overlaps that might occur, i.e., when two particles are closer than σ . Here, the circles around the particle centers have diameter equal to σ , and the overlapping ones are colored in green. (c) The choice of the time step of the simulation Δt must be such that the particle moves each step less than the reference length scale σ . (d) Illustration of the effect of the boundary on the particle: in this case, the boundary is reflective. (e) Screenshot of the particles’ positions, velocities (arrows) and trajectories (gray lines) in a typical Lennard–Jones simulation. For simplicity’s sake, the simulation includes only $N = 36$ particles. The particles’ speeds, initially set equal for all particles, are different after a number of collisions occurred. (f) Total energy, (g) kinetic energy, (h) potential energy, and (i) temperature of the Lennard–Jones system of particles. The kinetic and potential energies fluctuate around the initial values (in this case, the fluctuation is within 1ϵ). The total energy is supposed to remain constant but, in this simple simulation, reveals jumps of even 1ϵ in energy, if the time step is not small enough. This effect is due to the non-physical amount of superposition that the finite time step simulation introduces when it is not small enough.

Example 16.10: Emergence of Brownian motion. Implement numerically the two-dimensional gas as in Example 16.9. For simplicity, neglect the interactions between the gas particles. The gas particles have mass m_0 . Add a substantially larger particle with mass $M \gtrsim 10m_0$ and radius R that interacts with the gas particles via a Lennard–Jones potential characterized by σ and ϵ . The particle is initially at rest at the center of the box. Generate the positions of the gas particles such that, initially, they are more distant than R from the different particle set at rest at the center. Set the velocities of the gas particles at random, following Example 16.9.

- Run the simulation and plot the time evolution, like in Fig. 16.7. Choose an appropriate time step.
- Plot the trajectory of the big particle as a function of the time. Compare with Fig. 16.7. We should

obtain that the trajectory looks like proceeding at random, following an unpredictable, erratic path.

c. Calculate the *mean square displacement* (MSD) of the particle trajectory. Show that the trend follows the one represented in Fig. 16.7, i.e., it is linear for large time intervals. [For the definition of mean square displacement and how to calculate it, see Chapter 17, Box 17.1.]

d. Monitor the kinetic energy K_0 of the particle of mass M as a function of time. Show that initially it is zero, then, through collision with the gas particles, it increases until reaching a steady state value, around which the kinetic energy fluctuates.

e. Monitor the total kinetic energy K_{gas} of the gas of particles with mass m_0 . How much does it vary with time? What fraction of the initial kinetic energy K_{gas} is transferred to the mass M ?

f. Repeat the numerical simulation increasing the number of particles and varying the simulation parameters to obtain a more unpredictable- and erratic-looking trajectory for the big particle.

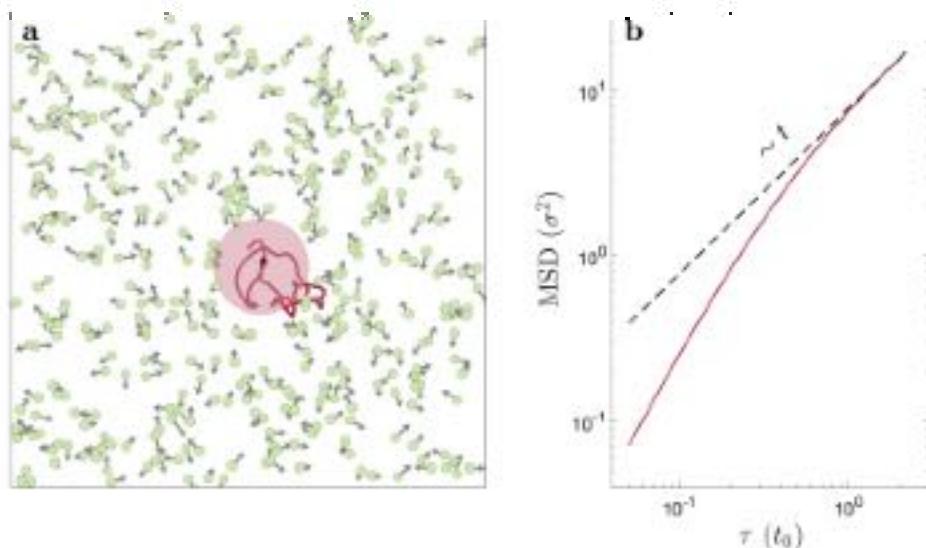


Figure 16.7: **Emergence of Brownian motion.** (a) Screenshot of the simulation after several thousands of time steps. A large disk (*Brownian disk*, red) interacts with several smaller gas particles via the Lennard–Jones potential. The small particles are, for simplicity’s sake, considered mutually non-interacting. The particles instantaneous velocity is represented with an arrow. The trajectory followed by the large particle is represented in red. The more the gas particles, the more collisions the Brownian disk experiences during the observation time interval t_{obs} , and the more erratic the motion of the particle results. (b) Mean square displacement (MSD) of the trajectory of the Brownian disk. The MSD scales linearly with t for large t .

16.3 Controlling the simulation

In a molecular dynamics simulation, the system is defined by a set of *state variables* such as, for example, the number N of particles in the system, the occupied volume V , the temperature T , the pressure P , and the total energy E . Such state variables are not all independent; for example, in the case of a gas, the state variables describing it are connected through its *equation of state*.

The worked Examples 16.9 and 16.10 depicted two cases of *isolated systems* where the total energy is conserved. This is what, in statistical mechanics, is called a *microcanonical ensemble*: a system that has a well defined volume V , number of particles N , and total energy E . There is no volume variation, and no particle or energy exchange with the environment occurs.

However, not all physical processes occur at constant volume, particle number, and energy. Often processes of interest occur at constant temperature (e.g., many biological reactions), or at constant pressure (e.g., physical

processes in the atmosphere). The system we want to simulate might be in thermal contact with a heat bath so the temperature is constant, or it might be free to expand or reduce in size so that its pressure is constant, or might even exchange particles with the environment at constant chemical potential.

Under the assumption of a constant number of particles, in a numerical dynamics simulation it is straightforward to deal with the case of constant volume and energy, because the Newton's law regulating the dynamics are equivalent to conserving the energy, when only conservative forces are at play. Instead, simulating a process at constant temperature or at a constant pressure requires specific algorithms to ensure that the temperature or the pressure of the system remains constant. Such algorithms are called respectively *thermostats* (to maintain a constant temperature) and *barostats* (to maintain a constant pressure). In the following paragraphs, we give a short overview of their most common implementations. For a deeper, more detailed approaches, we refer the reader to Refs. [15, 16, 17, 18, 19, 20].

16.3.1 Thermostats

A system that is characterized by constant N , V , and T is called, in statistical mechanics, a *canonical system*: the system of interest is in thermal contact with a heat reservoir with which it can exchange energy, while maintaining a constant temperature. The temperature of the system T is allowed to fluctuate around the average value T_0 imposed by the thermal bath. The features of these fluctuations (their amplitude and their characteristic decay time) are described in the framework of statistical mechanics [21]. For a system of N particles in d dimensions, the temperature can be defined as:

$$T = \frac{1}{dNk_B T} \sum_{i=1}^N m\mathbf{v}_i^2. \quad (16.22)$$

There are several algorithms to control the temperature in a numerical simulation. Let us consider a gas of identical particles. The equipartition theorem states that, at equilibrium, each degree of freedom is associated to an amount of energy $e_0 = \frac{1}{2} k_B T$. If we have N spherical particles in 3D with no internal degrees of freedom, the system in equilibrium at temperature T has a total average kinetic energy equal to $\frac{3}{2} N k_B T$:

$$\langle E_{\text{kin}} \rangle = \frac{1}{2} \sum_{i=1}^N m \langle \mathbf{v}_i^2 \rangle = \frac{3}{2} N k_B T \quad (16.23)$$

The simplest thermostat, called *velocity rescaling*, makes use of the equipartition theorem to keep the temperature constant: if T is the temperature we want for the system, then we rescale the system kinetic energy E_{kin} by the factor $\frac{3}{2} N k_B T / E_{\text{kin}}$. This thermostat, though of very simple implementation, is a rather crude method to correct the energy drift of a simulation, and must be implemented as infrequently as possible to minimize its effect on the trajectories. Moreover, it does not ensure that the new velocity distribution is the one corresponding to the equilibrium distribution for the system at temperature T .

Other more refined thermostats are the *Andersen* [15], *Berendsen* [17], and *Nosè-Hoover* thermostats [18]. Example 16.10, where the massive dust particle is coupled to the gas particles, can be seen as an analogue of the *Nosè-Hoover* thermostat, where the coupling with a virtual external degree of freedom stabilizes the temperature of the system. An in-depth discussion of thermostats algorithms and how to reproduce the different thermodynamics *ensembles* is out of the scope of this Chapter, but it is well documented in the literature [15, 16, 17, 18, 19, 20]. We refer the reader to the aforementioned references for an in-depth exploration of these concepts.

Example 16.11: A simple thermostat: velocity rescaling. Starting from the simulation of the 2D Lennard-Jones gas (Example 16.9), keep the initial temperature of the system T_0 constant by rescaling the kinetic energy of the velocities according to $Nk_B T_0 / E_{\text{kin}}$ every n_T iterations. Starting with the same initial condition, check the effect of the rescaling on the distribution of the particles' velocity by different simulation parameters (Δt , n_T). [The rescaling should happen as seldom as possible, for stability reasons. Can you make sense of why this happens? Can you see this in your simulation?]

Example 16.12: Berendsen thermostat. In the approach proposed by Berendsen *et al* [17], the system of interest is coupled to an external bath with a fixed reference temperature T_0 by inserting a stochastic term ($R(t)$) and a friction term ($-\gamma_i v_i$) in the equations of motion:

$$m_i \frac{dv_i}{dt} = F_i - \gamma_i v_i + R(t). \quad (16.24)$$

This is equivalent to the equation:

$$m_i \frac{dv_i}{dt} = F_i + \gamma_i \left(\frac{T_0}{T} - 1 \right) v_i, \quad (16.25)$$

where the local stochastic term is in the temperature T_0 . This allows a proportional scaling for the velocities, from v to $\lambda_s v$, with

$$\lambda_s = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T} - 1 \right) \right]^{\frac{1}{2}}, \quad (16.26)$$

where $\tau_T = \frac{m}{2\gamma}$ is the time constant of the coupling of the system with the heat bath, which determines the change in the temperature of the system per unit time. The scaling of the particles' velocities by λ_s determines a change in temperature during a single time step Δt of $\Delta T = \frac{T_0 - T}{\tau_T} \Delta t$, which is consistent with the assumption of the coupling with the external bath.

We now implement numerically the Berendsen thermostat. For simplicity, we suggest to implement the steps for a monoatomic gas, but we give the formulation as it appears in Ref. [17], i.e., for molecules, where in general a constraint on their three-dimensional shape and on the distances between the constituents atoms applies. Here, we give the algorithm for T constant only. The case in which also P is constant is given in Example 16.12.

Step 0. Configurations $\mathbf{r}(t)$ and velocities $\mathbf{v}(t - \frac{1}{2}\Delta t)$ are given for all atoms.

Step 1. Evaluate the (unconstrained) forces on all atoms:

$$\mathbf{a}_i(t) = \frac{\mathbf{F}_i(t)}{m_i}. \quad (16.27)$$

Step 2. Evaluate the temperature of the system:

$$T \left(t - \frac{1}{2}\Delta t \right) = \frac{2}{(d(N-1)-C)k_B} E_{kin} \left(t - \frac{1}{2}\Delta t \right), \quad (16.28)$$

where d is the dimensionality ($d = 2$ or 3), N is the number of atoms, and C is the number of constraints (in case the system is not with monoatomic components only) where

$$E_{kin} \left(t - \frac{1}{2}\Delta t \right) = \sum_i \frac{1}{2} m_i \mathbf{v}_i^2 \left(t - \frac{1}{2}\Delta t \right). \quad (16.29)$$

The temperature scaling factor becomes:

$$\lambda_s = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t - \frac{1}{2}\Delta t)} - 1 \right) \right]^{\frac{1}{2}}. \quad (16.30)$$

[Note: the kinetic energy is not available at time t , only at time $t - \frac{1}{2}\Delta t$, but this does not introduce any significant error because the coupling time constant τ_T is long compared to Δt .]

Step 3. Compute the new velocities as in the leapfrog methods:

$$\mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) = \mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) + \Delta t \mathbf{a}_i(t). \quad (16.31)$$

Step 4. Scale the velocities \mathbf{v}_i :

$$\mathbf{v}_i \left(t + \frac{1}{2} \Delta t \right) \leftarrow \lambda_s \mathbf{v}_i \left(t + \frac{1}{2} \Delta t \right). \quad (16.32)$$

[Note: the scaling factor λ_s is given at time $t - \frac{1}{2} \Delta t$, but it can be employed here because of its slow variation with Δt .]

Step 5. We compute the new (unconstrained) positions:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \left(t + \frac{1}{2} \Delta t \right) \Delta t. \quad (16.33)$$

For a constrained system (i.e., when we have molecules instead of monoatomic components only), we must apply the constraints and find the new constrained coordinates $\mathbf{r}_i(t + \Delta t)$, and on the base of this we calculate the new constrained velocities: $\mathbf{v}_i(t + \frac{1}{2} \Delta t) = [\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)] / \Delta t$.

Starting from your code in Example 16.9, implement the Berendsen thermostat following the steps given above. Monitor the temperature at each time step, and compare the results with those of Example 16.11.

16.3.2 Barostats

A system that is characterized by constant N , P , and T is called an *isothermal-isobaric ensemble*. Processes happening in the atmosphere are usually happening at constant pressure. A typical example of a process happening at constant pressure and temperature is a chemical reaction happening (for instance) at atmospheric pressure in contact with a thermal bath, where n_A molecules A combine with n_B molecules B to give n_{AB} molecules AB. In this case, the number of atoms in the system does not change, but the organization of the atoms in molecules changes. There are different algorithms to ensure constant pressure, called *barostats*. Examples of barostats are those defined by Parrinello-Rahman [16], Andersen [15], and Berendsen [17]. An important point for all these methods is to know how the pressure is defined. One way to define the pressure is as the average of the impulse per unit time and per unit area that the walls of the container exert on a gas. However, in this way only the particles that are close to the walls contribute to the pressure. A better definition can be given by using Clausius's virial function:

$$W(\mathbf{R}) = \frac{1}{2} \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i^{\text{tot}}, \quad (16.34)$$

where $\mathbf{F}_i^{\text{tot}} = \mathbf{F}_i + \mathbf{F}_i^{\text{ext}}$ is the total force, due to internal forces ($\mathbf{F}_i = \sum_{j \neq i}^N \mathbf{F}_{ij}$) and external ones ($\mathbf{F}_i^{\text{ext}}$), acting on the particle i , and by defining the pressure with the virial equation:

$$pV = Nk_B T + \frac{1}{d} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle, \quad (16.35)$$

where the sum is on the internal forces only, and d is the dimensionality of the system. Equivalently, in three dimensions one can define the pressure as:

$$p = \frac{1}{V} \left[\frac{1}{3} \sum_{i=1}^N \left(m\mathbf{v}_i^2 + \sum_{j>i}^N F_{ij} |\mathbf{r}_i - \mathbf{r}_j| \right) \right]. \quad (16.36)$$

Example 16.13: Berendsen thermostat and barostat. Building on the knowledge acquired in Example 16.12, we implement a Berendsen barostat [17]. To maintain the pressure constant, we need a rescaling of the coordinates $\mathbf{r}(t)$ and of the size $L(t)$ of the box containing our system, and, as a consequence, of its volume $V(t)$. Without repeating what already mentioned in Example 16.12, the scaling on the coordinates

is quantified by the parameter

$$\mu_s = \left[1 - \beta \frac{\Delta t}{\tau_T} (P_0 - P) \right]^{\frac{1}{3}}, \quad (16.37)$$

where β is the isothermal compressibility of the system. The scaling of the particles position ($\mathbf{r}(t) \leftarrow \mu_s \mathbf{r}(t)$) and the system dimensions ($L(t) \leftarrow \mu_s L(t)$) of μ_s determines a change in the pressure during a single time step Δt of $\Delta T = \frac{P_0 - P}{\tau_T} \Delta t$, which is consistent with the assumptions.

Now that have completed the picture including the pressure, we implement numerically the Berendsen thermostat and barostat. Again, we suggest to implement the step for a monoatomic gas, but we give the formulation as it appears in [17].

Step 0. Configurations $\mathbf{r}(t)$ and velocities $\mathbf{v}(t - \frac{1}{2}\Delta t)$ are given for all atoms. The box lengths $L(t)$ and the volume $V(t)$ are also given.

Step 1. (same as in Example 16.12) Evaluate the (unconstrained) forces on all atoms:

$$\mathbf{a}_i(t) = \frac{\mathbf{F}_i(t)}{m_i}. \quad (16.38)$$

Step 2. Evaluate the pressure from the virial and the kinetic energy in the center of mass. The virial is:

$$\Xi(t) = \frac{1}{2} \sum_{\alpha < \beta} \mathbf{R}_{\alpha\beta}(t) \mathbf{F}_{\alpha\beta}(t), \quad (16.39)$$

where $\mathbf{R}_{\alpha\beta} = \mathbf{R}_\alpha - \mathbf{R}_\beta$ is the relative coordinate of the center of mass of the molecule α with respect to β , and $\mathbf{F}_{\alpha\beta} = \sum_{i\alpha,j\beta} \mathbf{F}_{i\alpha,j\beta}^{\text{pair}}$ is the sum of all pair interactions between elements of molecule α and molecule β . The kinetic energy based on the center of mass of each molecule is:

$$E_{kin}^{CM} \left(t - \frac{1}{2}\Delta t \right) = \sum_{\alpha} \frac{1}{2} M_{\alpha} \mathbf{V}_{\alpha}^2 \left(t - \frac{1}{2}\Delta t \right), \quad (16.40)$$

where M_{α} is the total mass of molecule α ($M_{\alpha} = \sum_i m_{i,\alpha}$) and \mathbf{V}_{α} is the center of mass velocity of the molecule α : $\mathbf{V}_{\alpha} = (\sum_i m_{i,\alpha} \mathbf{v}_{i,\alpha}) / M_{\alpha}$. [Note: the kinetic energy is not available at time t , only at time $t - \frac{1}{2}\Delta t$, but this does not introduce any significant error because the coupling time constant τ_T is long compared to Δt .] The pressure becomes

$$P(T) = \frac{1}{3V(t)} \left[E_{kin}^{CM} \left(t - \frac{1}{2}\Delta t \right) - \Xi(t) \right] \quad (16.41)$$

and the pressure scaling

$$\mu_s = \left[1 + \beta \frac{\Delta t}{\tau_T} (P(t) - P_0) \right]^{\frac{1}{3}}. \quad (16.42)$$

Step 3. (same as Step 2 in Example 16.12) Evaluate the temperature od the system:

$$T \left(t - \frac{1}{2}\Delta t \right) = \frac{2}{(d(N-1)-C)k_B} E_{kin} \left(t - \frac{1}{2}\Delta t \right), \quad (16.43)$$

where d is the dimensionality (2 or 3), N is the number of atoms, and C is the number of constraints (in case the system is not with monoatomic components only) where

$$E_{kin} \left(t - \frac{1}{2}\Delta t \right) = \sum_i \frac{1}{2} m_i \mathbf{v}_i^2 \left(t - \frac{1}{2}\Delta t \right). \quad (16.44)$$

The temperature scaling factor becomes

$$\lambda_s = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t - \frac{1}{2}\Delta t)} - 1 \right) \right]^{\frac{1}{2}}. \quad (16.45)$$

[Note: the kinetic energy is not available at time t , only at time $t - \frac{1}{2}\Delta t$, but this does not introduce any significant error because the coupling time constant τ_T is long compared to Δt .]

Step 4. (same as Step 3 in Example 16.12) Compute the new velocities as in the leapfrog methods:

$$\mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) = \mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) + \Delta t \mathbf{a}_i(t). \quad (16.46)$$

Step 5. (same as Step 4 in Example 16.12) Scale the velocities \mathbf{v}_i :

$$\mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) \longleftarrow \lambda \mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right). \quad (16.47)$$

[Note: λ is given at time $t - \frac{1}{2}\Delta t$, but it can be employed here because of its slow variation with Δt .]

Step 6. (same as Step 5 in Example 16.12) Compute the new (unconstrained) positions:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) \Delta t. \quad (16.48)$$

In case we have a constrained system (i.e., when we have molecules instead of monoatomic components only) we must apply the constraints and find the new constrained coordinates $\mathbf{r}_i(t + \Delta t)$, and on the base of this we calculate the new constrained velocities: $\mathbf{v}_i(t + \frac{1}{2}\Delta t) = [\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)] / \Delta t$. In the case of a constant pressure, we have an additional last step to implement:

Step 7. Scale the coordinates \mathbf{r}_i and box size

$$\begin{aligned} \mathbf{x}_i(t + \Delta t) &\longleftarrow \mu_s \mathbf{x}_i(t + \Delta t) \\ L(t + \Delta t) &= \mu_s L(t) \\ V(t + \Delta t) &= \mu_s^3 V(t) \end{aligned} \quad (16.49)$$

The steps described above give the implementation of the Berendsen thermostat and barostat. Starting from your code in Example 16.9, implement the Berendsen thermostat. Monitor the temperature and pressure at each time step, and compare the results with those of Example 16.11.

16.4 Boundary conditions

Typically, in a molecular dynamics simulation one wants to keep the particles within a certain volume of phase space. This can be achieved either using an external potential or employing some boundary conditions. In this Section, we will see the most common boundary conditions.

16.4.1 Reflecting boundary condition

The simplest one is a *reflecting boundary* condition (Fig. 16.8). As the name suggests, when a particle reaches a reflecting boundary, it is reflected back into the bulk. This situation is quite natural and happens, for example, in the case of gas molecules in a party balloon. In the case of a plane boundary, the reflecting boundary condition can be implemented in two steps: one should check whether a particle has crossed the plane and, in this case, reflect its momentum orthogonal to the plane. Other shapes can be dealt with similarly as long as they can be locally approximated as planes.

Example 16.14: Circular box with reflective boundaries. The simulation in Example 16.9 was per-

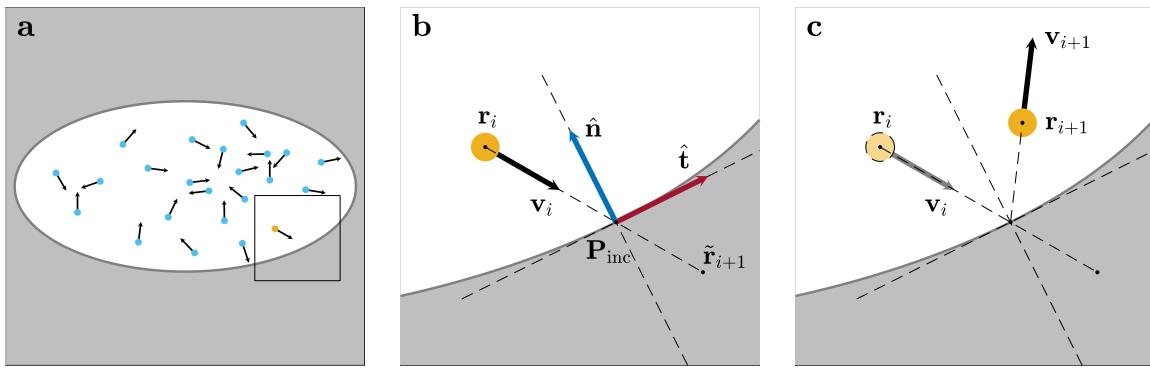


Figure 16.8: Reflective boundaries. (a) Particles inside an elliptical box. (b) The yellow particle in \mathbf{r}_i after one time step Δt , crosses the boundary of the box and would end up in \mathbf{r}_{i+1} . To determine the particle position after reflection one first determines where the trajectory of the particle intersects the boundary (\mathbf{P}_{inc}), the tangent direction (red arrow, $\hat{\mathbf{t}}$), and the normal direction (blue arrow, $\hat{\mathbf{n}}$) in that point. (c) The position \mathbf{r}_{i+1} of the particle is reflected with respect to the tangent plane to the boundary in the point of incidence and the final position $\tilde{\mathbf{r}}_{i+1}$ is determined. The final velocity of the particle (grey arrow) is reflected accordingly.

formed implementing reflective boundary conditions. Implement the same boundary conditions for the case of a circular arena of radius R_{arena} .

Example 16.15: Elliptical box with reflective boundaries. Implement the reflective boundary conditions for the case of an elliptical arena of semiaxes a and b . [Hint: from the Cartesian equation of the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ write the normal \mathbf{n} and tangential \mathbf{t} unit vectors in a generic point of the ellipse. Use the unit vectors for implementing correctly the reflective boundary conditions.]

16.4.2 Periodic boundary condition

A small volume with $N \approx 10^3$ is not representative of a condensed matter system, which will typically have on the order of 10^{23} particles. While such an approximation can work well for gases, where there is no long-range order, it clearly does not work for liquids and solids. In order to mimic a liquid or solid without increasing the number of particles to be simulated, it is possible to use *periodic boundary conditions*. This is like simulating a unit cell of a very large three-dimensional tiling: a *unit cell* or simulation box of a geometry suitable for perfect three-dimensional tiling is defined, and when an object passes through one of its faces, it reappears on the opposite face with the same velocity (Fig. 16.9). The *minimum image convention* is typically employed to decide with which image of the other particles a given particle should interact. Each individual particle in the simulation interacts with the closest image of the remaining particles in the system: given two atoms for which we wish to evaluate the potential and the corresponding force, in each of the three dimensions we check the distances between the two atoms in the same cell, as well as one atom being in one cell, the other periodically displaced into a nearby cell, and of the two distances we choose the smaller one. According to this scheme the maximum distance which enters the calculation of the potential is $L/2$. In order not to introduce any artificial anisotropy into the calculation, a cutoff of $L/2$ is introduced into the potential. To save processing time, the cutoff distance can often be decreased from $L/2$ to a smaller value r_c .

Example 16.16: Square arena with periodic boundary conditions. Repeat the simulation in Example 16.9 implementing periodic boundary conditions.

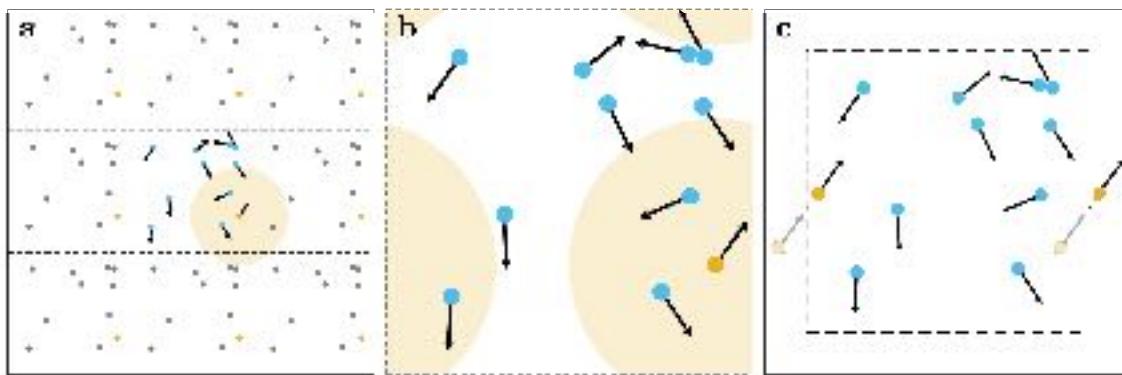


Figure 16.9: Periodic boundaries. (a) System of particles with periodic boundary conditions. The central square is the arena, and the eight neighboring squares represent the *unwrapped system* with the replicas of the arena and the particles (in grey). The dashed lines highlight the arena boundaries. One of the particles is highlighted in yellow to show its replicas. The interaction radius is represented with a circular light yellow area. The replica of one of the particles to the left is shown to belong to the interaction radius of the highlighted particle. (b) Here only the central arena is shown. The interaction area is *wrapped* onto the central arena and the apparently disconnected circular sections are shown. (c) The highlighted particle is moved to its next position, which falls out to the right of the arena boundary. Therefore, the particle is relocated by wrapping it onto the main arena. This is equivalent to move the particle replica on the left neighboring arena instead of moving the main particle.

16.4.3 Adsorbing boundary condition

Finally, there can be adsorbing boundary conditions. In this case, when a particle hits an absorbing boundary, it sticks to it and does not move anymore: it is taken away from the dynamics (Fig. 16.10). This might be of interest, for example, if the molecules in the simulation can interact with the boundaries such as in the case of some chemical reactions. Obviously, in this case the number of particles will not be conserved and, therefore, neither the momentum nor the energy.

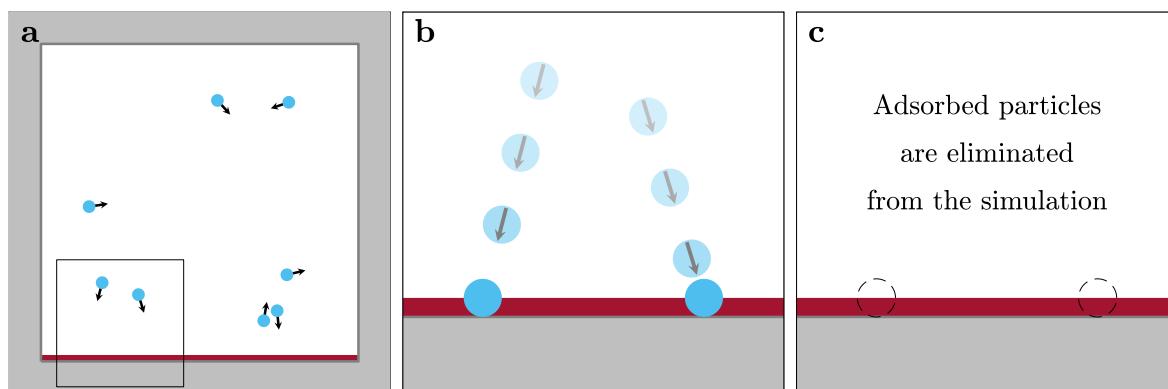


Figure 16.10: Adsorbing boundaries. (a) System of particles with adsorbing boundary conditions (highlighted with a thick red line) on the bottom boundary. (b) Two particles are moving in the direction of the bottom boundary. The particles, when crossing the bottom adsorbing boundary, get stuck on it. (c) Once adsorbed, the particles are removed from the simulation.

Example 16.17: Square arena with one boundary with adsorbing boundary conditions. Repeat the simulation in Example 16.9 where all the boundaries are reflective, except one which is adsorbing. Do not

remove the adsorbed particle from the simulation, but consider it when a second particle hits the adsorbing boundary: if the second particle hits the adsorbing boundary is in the range of σ from the adsorbed particle, it does not get adsorbed but interacts with the adsorbed particle in the usual way.

16.5 Problems

Problem 16.1: Damped harmonic oscillator. Consider a damped harmonic oscillator, characterized by the equation:

$$m \frac{d^2x}{dt^2}(t) = -b \frac{dx}{dt}(t) - kx(t) \quad (16.50)$$

where $b > 0$. Depending on b , we have three different regimes: oscillating ($b < 2\sqrt{mk}$), critically damped ($b = 2\sqrt{mk}$), and overdamped ($b > 2\sqrt{mk}$).

- a. Find the analytical solutions for the initial conditions $x(0) = x_0$ and $v(0) = v_0$ for all regimes.
- b. Solve numerically the problem with the Euler method, for all the three different regimes.
- c. Solve numerically the problem with the leapfrog method, for all the three different regimes. Use the same parameters used for the Euler method.
- d. Compare the trajectories obtain with the two methods for each regime. Observe where and how they differ.
- e. Plot the total energy of the system, and check whether the work–energy theorem is numerically fulfilled. Comment about the performances of the two methods in the overdamped regime.

Problem 16.2: Orbit of the Earth around the Sun. Consider a simulation of the Earth orbiting around the Sun. Let v_a be the Earth's velocity at the *aphelion*, the most distant point (distance: r_a) of Earth's trajectory around the Sun. The gravitational force acting on the Earth is $F(r) = \frac{GM_S m_E}{r^2}$. In the limit of considering the Sun as fixed at the origin of the reference frame (which is a reasonable approximation, given that $M_S \gg m_E$) the angular momentum of the Earth $\mathbf{L} = m_E \mathbf{r} \times \mathbf{v}$ is conserved, and so is the total energy of the Earth $E_{\text{tot}} = -\frac{GM_S m_E}{r} + \frac{1}{2} m_E \mathbf{v}^2$. The conservation of the Earth's angular momentum \mathbf{L} is equivalent to Kepler's second law, and Kepler's third law states that $T^2 = \frac{4\pi^2}{GM_S} R_0^3$, where T is the period and R_0 is the length of the major semiaxis of the orbit ($R_0 = r_a + r_p$, where r_p is the distance of the *perihelion*, i.e., the closest point of Earth's trajectory around the Sun.)

- a. Perform a simulation with $v_x(0) = 0$, $v_y(0) = v_a$, $y(0) = 0$, $x(0) = r_a$, and $\Delta t = 1$ day. Can you reproduce the trajectory of the Earth around the Sun reasonably? Is the trajectory reasonably periodical? Can you reproduce the values of the angular momentum, the perihelion, and the period? Simulate $N = 10$ revolutions, check if the total energy and the angular momentum is conserved.
- b. Perform simulations with $v_y(0) = 0.9 v_a$, $v_y(0) = 0.8 v_a$, and $\Delta t = 1$ day. Observe if the simulation behaves in the expected way. Are the trajectories still reasonably periodical? Do you observe that the perihelion is closer to the sun? Does the period of the orbit decrease?
- c. Perform simulations with $v_y(0) = 0.6 v_a$, $v_y(0) = 0.5 v_a$, $v_y(0) = 0.4 v_a$, and $\Delta t = 1$ day. Is the simulated trajectory still periodical?
- d. Perform simulations with $v_y(0) = 0.6 v_a$, $v_y(0) = 0.5 v_a$, $v_y(0) = 0.4 v_a$, and $\Delta t = 0.1$ day. Is there any improvement in the simulated trajectory? How important is the choice of the time step? [Hint: in this case, the force is not bounded, it diverges as r^{-2} .]

Problem 16.3: Solar system. Inspired by Problem 16.2, simulate the solar system. Use realistic data. Use an appropriate time step. Check the stability of your simulation and check the period you obtain in the simulation with the astronomical data.

Problem 16.4: Man to the Moon. Try to reproduce the trajectory of the Apollo 11 in the mission to the Moon.

Problem 16.5: Oscillations of an elastic band. Consider an elastic band moving in two dimensions, where N identical point-like masses of mass m are connected sequentially by identical perfect, massless springs with

elastic constant $k = 1$. Use $N = 120$ masses. Each mass $m = 1$ is connected to the two masses at its left and at its right with a spring of elastic constant k and rest length l_0 . The mass at the right end is kept fix in place. The mass at the left end is forced to move sinusoidally along the vertical direction, with law $y(t) = A \sin(\omega t)$. Initially, all masses are set in $x_1^{(j)} = (j - 1)d_0 \mathbf{x}$ and $y_1^{(j)} = 0 \mathbf{y}$ and are at rest.

a. Use the leapfrog method. Suppose $d_0 = 1$, $l_0 = 0.5d_0$, i.e., the small spring connecting the masses are all stretched. Take $A = 0.2d_0$, $\omega = 2\pi/T$ with $T = 25$. Simulate the time evolution of the elastic band. Choose an appropriate time step.

b. This time, let the mass at the right corner free to move vertically. Simulate the time evolution of the elastic band in this case (the other conditions are like before).

Problem 16.6: Guitar chords, elastic carpets, spider webs. Get inspired from the movies uploaded in the directory with the material for today and choose another simulation to perform [Check. It doesn't make much sense]: (i) simulate the oscillations in Problem 16.5 with different value of the masses [Which problem are you referring to here? Check I have added the correct one]; (ii) simulate a guitar chord pinched in the middle; (iii) simulate the effect of gravity on an elastic band with fix extrema, set horizontally; (iv) simulate bungee jumping, (v) simulate a ball bouncing on an elastic carpet; (vi) simulate the oscillations of a spider web when a fly gets captured.

Problem 16.7: Truncated Lennard–Jones potential. Repeat Example 16.9 using a truncated Lennard–Jones potential given in Eq. (16.19). Show that the behavior of the system does not change significantly with respect to Example 16.9.

Problem 16.8: Diverging potentials. Interaction potentials like the Lennard–Jones, that are diverging very rapidly for mutual distances $r \rightarrow 0$, create many problems in simulations, especially when the initial starting position of the particles is chosen randomly. [A note: one might prefer to start from a random position rather than from a predetermined one, i.e., for example with the particles set in a regular lattice, depending on the purpose of the simulation.] If two particles start too close, i.e., at a distance smaller than the parameter σ (the particles size) of the potential, then the force that the particles experience is too large and this might compromise the simulation with unrealistic values to the velocities after a few time steps. We must keep in mind that the simulation is performed at a fixed Δt and that, during each time step, the displacement of each particle should not exceed a small fraction of the particle size σ . In such situations, especially when the density of the particles is high, it is good practice to put a cap on the potential (and the force), so that the dynamics is not compromised from the initial random choice of the particles' positions, and progressively lift the cap so that the full Lennard–Jones dynamics is recovered.

a. Similarly to a the definition of a truncated Lennard–Jones potential seen in Eq. (16.19), define a *capped* Lennard–Jones force as:

$$F_{c\text{LJ}}(r) = \begin{cases} F_{\text{LJ}}(r_0) & 0 < r \leq r_0 \\ F_{\text{LJ}}(r) & r > r_0 \end{cases} \quad (16.51)$$

where $F_{\text{LJ}}(r)$ is the force in Eq. (16.20), and $0 < r_0 \leq \sigma$. As the force increases rapidly for $r < \sigma$, it could be a good starting point to take $r_0 = \sigma$ for the initial part of the simulation.

b. Implement the dynamics in such a way that initially $r_0 = \sigma$ and, every n_0 iterations, r_0 changes and becomes a fraction $q < 1$ of its previous value. After $k \cdot n_0$ iterations $r_0 = q^k \sigma$ and $\lim_{k \rightarrow \infty} q^k \sigma = 0$. With time, the capping radius r_0 tends to zero and the force represent the full Lennard–Jones force.

c. Compare the behavior of two simulations starting with the same initial conditions but one evolving with the full Lennard–Jones potential, the other with the capped Lennard–Jones potential, whose force is expressed in Eq. (16.51) and with r_0 decreasing gradually as described in **b**. In particular, try a case with a high density of particles and a case with a low particle density. Which method performs better in the two cases?

Problem 16.9: Solid, liquid, and gas. The Lennard–Jones interaction can describe solid, liquid and gaseous phases, depending on the relative values of the temperature T and of the density ρ of the system compared to the dimensionless units $T_0 = \epsilon/k_B$ (temperature) and $\rho_0 = \sigma^{-3}$ (particle density).

a. Show that, if $T/T_0 \approx 2$ and $\rho/\rho_0 \approx 0.01$ then the system is a gas.

b. Show that, if $T/T_0 \approx 1$ and $\rho/\rho_0 \approx 0.8$ then the system is a liquid.

- c. Show that, if $T/T_0 \approx 0.1$ and $\rho/\rho_0 \approx 0.9$ then the system is a solid.

Problem 16.10: Gas in a box with a movable wall. Consider a gas of N monoatomic particles (mass m) in a rectangular box. Suppose the particles interact with a Lennard–Jones potential and are in a regime where they behave as a gas (see Problem 16.9). Set one of the walls as movable, with a mass $M_w \gg m$. Suppose that, in case the movable wall moves with velocity $V_w > 0$, a constant friction force F_w acts on it and opposes to its motion.

- a. Simulate the evolution of the system in time, and in particular the value of the volume V and of the pressure p of the system.

Problem 16.11: Compressing a gas. Consider a gas of N monoatomic particles (mass m) in a rectangular box. Suppose the particles interact with a Lennard–Jones potential and are in a regime where they behave as a gas (see Problem 16.9). Suppose that one of the walls, with a mass $M_w \gg m$, moves slowly with constant velocity V_w and compresses the gas.

- a. Simulate the evolution of the system in time, and in particular the value of the temperature T and of the pressure p of the system.

Problem 16.12: Brownian rod. Instead of a Brownian disk, simulate a Brownian rod, taking inspiration from Example 16.10. Suppose the rod has a length L and a mass M . Its center of mass is located in the middle point of the segment. When considering the dynamics of the rod, you have to take into account also the torque acting on it, related to the point in which a particle hits the rod.

- a. Implement a routine that calculates the force and the torque that a particle exerts on the rod.
- b. Implement the dynamics taking into account also the rotational motion of the rod around its center of mass.
- c. Show that not only the position of the center of mass of the rod has an erratic motion, but also its orientation presents an erratic behavior.
- d. Is the time scale of the two different erratic behaviors the same time scale?

Bibliography

- [1] D. Frenkel and B. Smit. *Understanding Molecular Simulation: from algorithms to applications*. Academic Press, 1987.
- [2] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publications, 1987.
- [3] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 1996.
- [4] M. J. Field. *A Practical Introduction to the Simulation of Molecular Systems*. Cambridge University Press, 1999.
- [5] J. B. Gibson, A. N. Goland, M. Milgram, and G. H. Vineyard. Dynamics of radiation damage. *Physical Review*, 120:1229–1253, 1960.
- [6] A. Rahman. Correlations in the motion of atoms in liquid argon. *Physical Review*, 136:A405, 1964.
- [7] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159:98, 1967.
- [8] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.
- [9] B. R. Gelin and M. Karplus. Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences*, 72:2002–2006, 1975.
- [10] G. Söderlind. *Numerical Methods for Differential Equations, An Introduction to Scientific Computing*. Springer, 2017.

- [11] H. Goldstein, C. P. Poole, and J. Safko. *Classical Mechanics*. Pearson Education, 2011.
- [12] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2016.
- [13] B. Hafskjold, K. P. Travis, A. Bailey Hass, M. Hammer, A. Aasen, and Ø. Wilhelmsen. Thermodynamic properties of the 3D Lennard-Jones/spline model. *Molecular Physics*, 117:3754–3769, 2019.
- [14] J. E. Lennard-Jones. Cohesion. *Proceedings of the Physical Society*, 43:461–482, 1931.
- [15] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics*, 72:2384, 1980.
- [16] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Physical Review Letters*, 45:1196, 1980.
- [17] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81:3684, 1984.
- [18] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81:511, 1984.
- [19] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31:1695, 1985.
- [20] P. H. Hünenberger. *Thermostat Algorithms for Molecular Dynamics Simulations*, pages 105–149. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [21] R. K. Pathria and P. D. Beale. *Statistical Mechanics*. Elsevier, third edition edition, 2011.

Chapter 17

Brownian Dynamics

AYKUT ARGUN, AGNESE CALLEGARI, GIORGIO VOLPE, GIOVANNI VOLPE

Microscopic objects suspended in a liquid or gas move in an apparently random way. This phenomenon, known as *Brownian motion*, is a consequence of continuous collisions between these micro-objects and the molecules in their surrounding medium. The first experimental observation of such motion was done by Robert Brown (hence the name), a botanist who observed that microscopic particles within pollen grains were moving at random independently from each other [1] (Fig. 17.1a). In 1905, Albert Einstein developed a mathematical description for the diffusion of small particles in liquids and connected this phenomenon to the existence of atoms [4]. Later, in 1909, Jean Baptiste Perrin quantified such motion, putting to rest the century-old debate whether atoms had a physical reality or were just nice mathematical abstractions [3] (Fig. 17.1b). Understanding the dynamics of Brownian particles can provide important insights into microscopic thermodynamics and biophysical transport phenomena [5]. Furthermore, Brownian motion plays a role in many different phenomena at all scales, including foraging animals, financial markets, and human organizations [6]. In this Chapter, we start by simulating random walks with different step distributions, showing that many of them belong to the same universality class. Next, we simulate the Brownian motion of microscopic particles in a liquid environment in

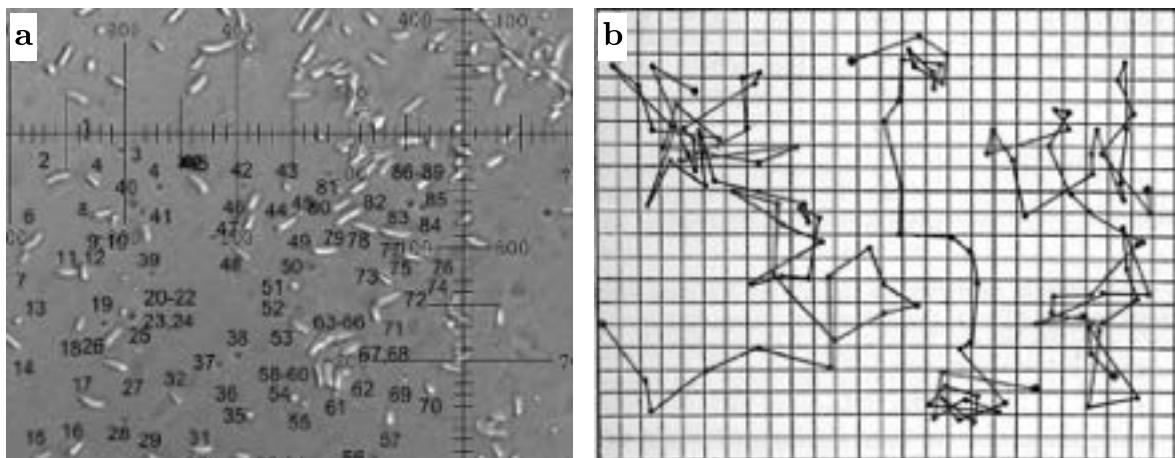


Figure 17.1: **Brownian motion and Perrin’s experiment.** (a) *Clarkia pulchella* pollen contents after dehiscence. The scale is $2\text{ }\mu\text{m}$ per division. This is what Brown likely observed in his experiments [1]. Reproduced from Ref. [2]. (b) Positions of microscopic spherical particles suspended in water recorded by Jean Baptiste Perrin in his famous 1909 experiment. The points are the measured positions of particles, taken at time intervals of 30 seconds. The straight lines connect in sequence the positions of the same particle. Here, the trajectories of three different particles are represented. The lines of the squared grid are separated by distance of $3.2\text{ }\mu\text{m}$ and the radius of the particles was $0.53\text{ }\mu\text{m}$. Reproduced from Ref. [3].

the inertial and viscous regimes. Then, we introduce how to simulate the motion of active Brownian particles, also in complex and crowded environments. We illustrate this with two examples of active particle systems where complex behaviors emerge: the formation of *living crystals* in a system with many active particles with phoretic interactions; and the creation of *metastable channels* by active particles moving through passive particles. Finally, we consider Brownian motion in the presence of *multiplicative noise* and its applications to the self-organization of active particles.

17.1 Random walks

Random walks are paths traveled by an agent taking successive steps in random directions. This mathematical construct can be used to model Brownian motion. In one dimension, a random walk is obtained by summing up the terms of a sequence of random numbers. Although the microscopic details of the motion depend on the exact step distribution, the macroscopic outcome features some universal properties. One of the simplest ways to obtain a one-dimensional random walk is by flipping a coin. At each coin flip, an agent will take a step to the right with a probability of 50% and a step to the left with a probability of 50%. The corresponding iterative equation for this walk can be written as:

$$x_{i+1} = x_i + \xi_{\text{flip}}, \quad (17.1)$$

where ξ_{flip} represents a random equal probability selection between -1 and 1 .

The precise form of a random walk is obviously not predictable, since it depends on a sequence of random events. However, if you realize several random walks (for example, by repeating the coin flipping sequence several times), you can identify some *average properties*, which are deterministic. For example, it is possible to predict the average value of x_n , i.e., $\langle x_n \rangle$. The $\langle \dots \rangle$ notation denotes an *ensemble average*, i.e., an average over all the possible realizations of the random walk; practically, you can take an ensemble average by generating a large number of random walks as described above and then averaging them. If you start with $x_0 = 0$ (and obviously also $\langle x_0 \rangle = 0$), due to the equal probability of taking steps to the right or to the left, it is possible to show that

$$\langle x_i \rangle = 0 \quad (17.2)$$

for every $i \geq 0$.

The mean displacement $\langle x_i \rangle$ is not a very informative measure for the motion of a random agent, as its value is always null. Nonetheless, we can measure the mean square displacement (MSD) of the agent, which can be shown to be $\langle x_i^2 \rangle = i$.

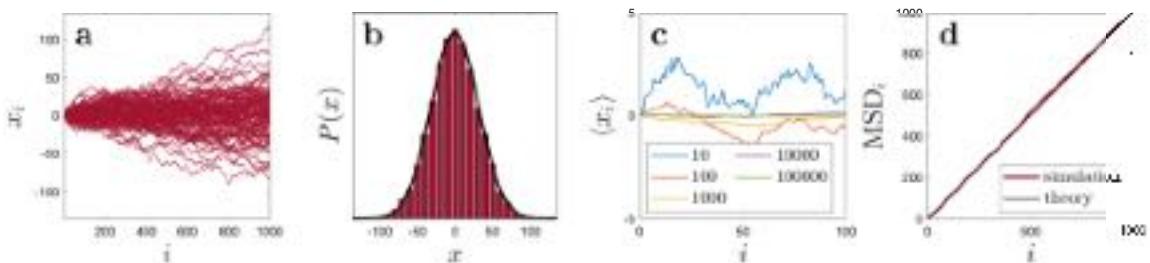


Figure 17.2: Flipping of a coin and random walk. A random walk can be realized by flipping a coin and taking one step to the left or to the right with equal probability. **(a)** Plot of $N_P = 100$ independent realizations $x(i)$ of the coin-flip random walk as a function of i . **(b)** Distribution of the random walk after $N = 1000$ steps, calculated over $N_P = 10^4$ independent realizations. The distribution is closely fitted by a Gaussian distribution with average $\mu = 0$ and standard deviation $\sigma = \sqrt{1000} \approx 31.6$. **(c)** Ensemble average of the displacement $\langle x(i) \rangle$ as a function of i . The different lines correspond to different numbers of independent trajectories used in the ensemble average: the larger this number, the closer to zero the average displacement. **(d)** Mean square displacement $MSD(i)$ of the coin-flip random walk, obtained by calculating the ensemble average over $N_P = 10^4$ independent realizations. The expected theoretical behavior ($MSD(i) = i$) is shown by the dark straight line.

Example 17.1: Random walk by flipping a coin. Write a program that simulates a random walk for N steps, where the agent has equal probability of stepping to the right (+1) or to the left (-1).

- Plot 10 sample trajectories (independent realizations of the random walk) of $N = 1000$ steps each.
- Run $N_P = 10^4$ random walks with $N = 1000$ steps. Record the end value x_N for each realization. Plot the distribution $P(x_N)$.
- Calculate (numerically) the ensemble average $\langle x_i \rangle$. Show (numerically) that $\langle x_N \rangle$ goes to zero as $N_P \rightarrow \infty$.
- Calculate the MSD (see Box 17.1) as an ensemble average. Show (numerically) that $\text{MSD}(i) = i$. [Compare your results to those shown in Fig. 17.2.]
- Random walk in n dimensions. Simulate a random walk in n dimensions with $n = 2$ or 3 .

Box 17.1: Mean square displacement (MSD). The MSD is defined as an *ensemble average* over N_P different trajectories $x^{(j)}(t)$, where $j = 1, 2, \dots, N_P$ indicates the particle:

$$\text{MSD}_{\text{ea}}(t) = \left\langle |x^{(j)}(t) - x^{(j)}(0)|^2 \right\rangle = \frac{1}{N_P} \sum_{j=1}^{N_P} |x^{(j)}(t) - x^{(j)}(0)|^2. \quad (17.3)$$

The MSD of a single trajectory $x(t)$ can also be calculated as a *time average*:

$$\text{MSD}_{\text{ta}}(\tau) = \overline{|x(t + \tau) - x(t)|^2} = \text{MSD}(\tau) = \frac{1}{N-n} \sum_{i=1}^{N-n} |x_{i+n} - x_i|^2, \quad (17.4)$$

where $x_0, x_1, \dots, x_i, \dots$ are the trajectory values at discrete times $t_0, t_1, \dots, t_i, \dots$ with $t_i = i\Delta t$ and $\tau = n\Delta t$.

For an ergodic system (such as Brownian particles), the ensemble average of an observable is equivalent to its time average value. For non-ergodic systems (such as systems characterized by anomalous diffusion; see Chapter 6) the time average and the ensemble average might differ.

Until now, we have generated random walks using binary random numbers at each step. Interestingly, even if we draw the random numbers from a different distribution (e.g., uniform random numbers in $[-1, 1]$ or Gaussian random numbers), all the random walks look the same on time scales long enough for the individual steps not to be distinguishable, as can be seen in Fig. 17.3. This independence of behavior on the microscopic details is called *universality*, which we will explore in some detail in the next example.

Example 17.2: Universality of the random walk. Write a program that simulates random walks for N steps for the following step distributions:

- Coin flip (50% +1, 50% -1; Figs. 17.3a-e).
 - Gaussian distribution with mean 0 and variance 1 (Figs. 17.3f-j).
 - An asymmetric distribution ($25\% -\sqrt{6}/2, 25\% -3\sqrt{6}/10, 50\% 2\sqrt{6}/5$; note that also in this case the mean is 0 and the variance is 1; Figs. 17.3k-o).
- Plot $N_P = 10$ sample trajectories for each case (Figs. 17.3b, 17.3g, and 17.3l). Increase the number of trajectories to $N_P = 10^4$ and the number of steps to $N = 1000$. Plot them for each case and compare with Figs. 17.3c, 17.3h, and 17.3m.
 - Plot the distributions of x_N for $N = 1000$ ($N_P = 10000$). Show that all three cases can be fitted with a Gaussian distribution with average $\mu = 0$ and standard deviation $\sigma = \sqrt{N}$ (Figs. 17.3d, 17.3i, and 17.3n).
 - Calculate average and MSD like in Example 17.1 and compare your results with Figs. 17.3e, 17.3j, and 17.3o.

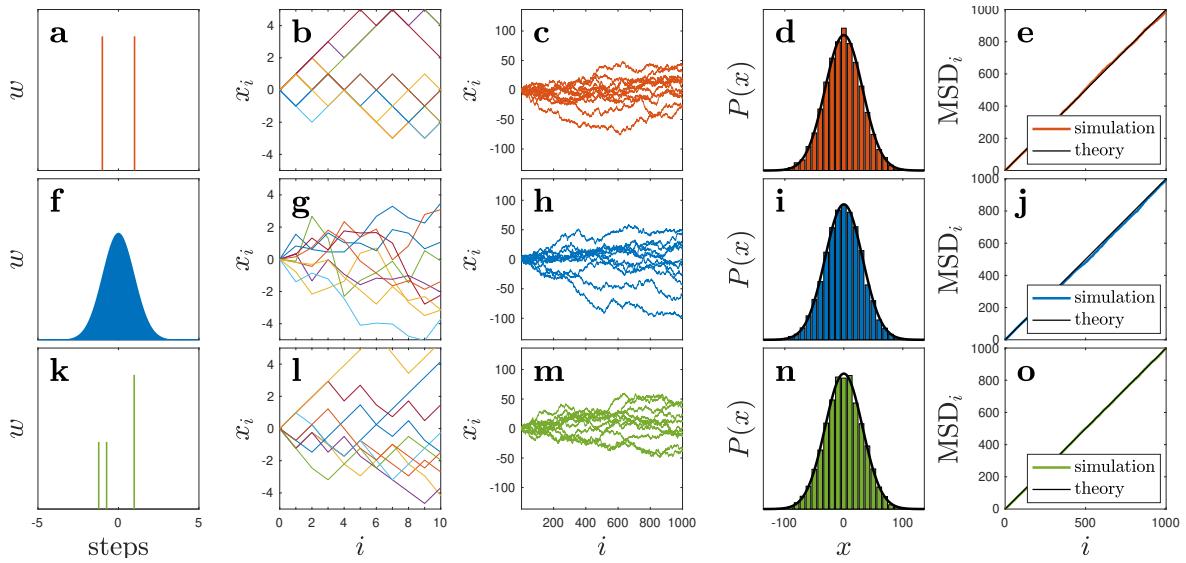


Figure 17.3: Universality of random walks. Step distributions for (a) coin-flip steps, (f) Gaussian steps, and (k) asymmetric steps. All distributions have the same mean ($\langle \Delta x \rangle = 0$) and variance ($\langle (\Delta x)^2 \rangle = 1$). (b), (g), (l) Sample trajectories for each case look different at short time scale (in this case, 10 steps). (c), (h), (m) At longer time scales (1000 steps), the trajectories look indistinguishable. (d), (i), (n) The distributions of the final position after 1000 steps tend to the same Gaussian limit distribution and (e), (j), (o) The relative mean square displacements (MSDs) also tend to the same theoretical $MSD(i) = i$.

17.2 White noise

White noise is a key concept in the study of stochastic phenomena. In order to gain an intuitive picture of white noise, we can think of it as the continuous limit of a sequence of random numbers as the time step between them goes to zero. We can start from a one-dimensional random walk expressed by

$$x_{i+1} = x_i + r_i, \quad (17.5)$$

where $\{r_i\}$ is a sequence of independent random numbers with mean 0 and variance σ_r^2 . If we imagine this random walk to happen in time, we can think that each step corresponds to a certain amount of time (say $\Delta t = 1$ s) and the random walk described by Eq. (17.5) is characterized by the fact that after a time $t = n\Delta t$ its average is 0 and its variance is $n\sigma_r^2$. If we now decrease the time step, we would still like to retain the global long-term properties of the resulting random walk (e.g., its variance) so that they do not depend on the

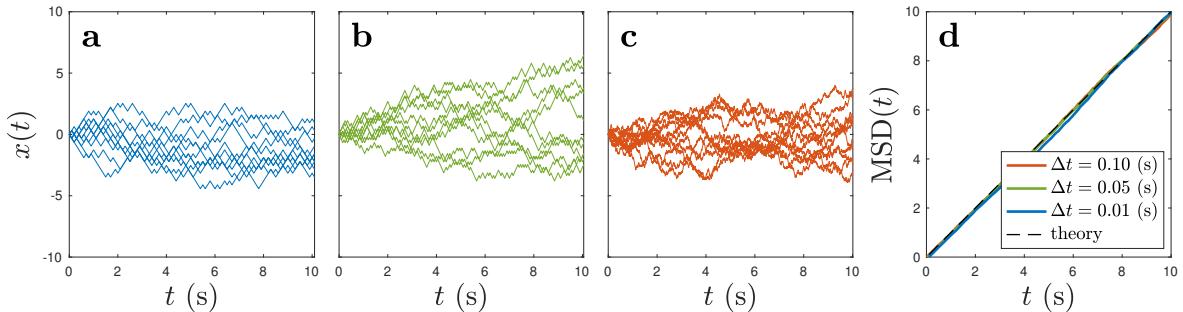


Figure 17.4: Scaling of diffusion with decreasing time step. (a-c) Sample trajectories of the simulated diffusion equation (Eq. (17.6)) with $\sigma_r = 1$ and a total simulation time $T = 10$ s for (a) $\Delta t = 0.1$ s, (b) 0.05 s, and (c) 0.01 s. As expected the qualitative behavior of the trajectories is the same, independent of Δt . (d) For example, they all have the same MSD (calculated over 10^4 realizations).

sampling time. Therefore, if we decrease the time step to half, i.e., to $\Delta t/2$, we need to increase the variance of the random numbers used in the construction of the random walk by 2, in order to retain the variance of the random walk at time t unchanged. If we decrease the time step further, we have to increase the variance of the single step further, as shown in Fig. 17.4. The corresponding process can be described as

$$x_{i+1} = x_i + \sqrt{\sigma_f^2 \Delta t} \xi_i, \quad (17.6)$$

where ξ_i is a sequence of random numbers with mean 0 and variance 1. This is the simplest version of a free diffusion equation.

A white noise $\xi(t)$, also known as a *Wiener process*, is obtained by taking the process described in the previous paragraph to the limit $\Delta t \rightarrow 0$. $\xi(t)$ is characterized by the following properties:

1. the mean $\langle \xi(t) \rangle = 0$ for all t ;
2. $\langle \xi(t)^2 \rangle = 1$ for each value t ;
3. $\xi(t_1)$ and $\xi(t_2)$ are independent of each other for $t_1 \neq t_2$: $\langle \xi(t_1) \xi(t_2) \rangle = \delta(t_1 - t_2)$

Because of these properties, white noise cannot be treated as a standard function. In particular, it is almost everywhere discontinuous and has infinite variation. Thus, it cannot be approximated by its instantaneous values at times t_i , because these values are not well-defined (due to the lack of continuity) and their magnitude varies wildly (due to the infinite variation).

Equation (17.6) enables simulations with consistent results independent of the time step. Normally, the time step Δt should be much smaller than the characteristic time scales of the stochastic process to be simulated. If Δt is comparable to or larger than the smallest time scale, the numerical solution will not converge and typically shows a nonphysical oscillatory behavior or divergence. The case of free diffusion treated in this section is special because there is no characteristic time scale, as can be seen from the fact that Eq. (17.6) is self-similar under a rescaling of the time, and therefore there is no optimal choice of Δt .

Example 17.3: Random walks happening in time with different time steps. Simulate the diffusion equation (17.6) for $\Delta t = 0.1$, $\Delta t = 0.05$ and $\Delta t = 0.01$.

- a. Obtain $N_P = 10$ different 10-s-long trajectories for each Δt and plot them as in Figs. 17.4a-c. Observe that they qualitatively look very similar.
- b. Simulate $N_P = 10^4$ trajectories for each Δt and calculate the $MSD(t)$ for each ensemble (Fig. 17.4d). Show that the MSD does not depend on our choice of Δt .

17.3 Brownian motion

One of the simplest examples of a stochastic system is a Brownian particle, which is a microscopic particle suspended in a fluid. Such particles perform random walks in their fluid environment with Gaussian-distributed steps. A free Brownian particle undergoes a diffusion process that is very similar to what was described in the previous section. Brownian particles are often used to study random phenomena, because their motion due to collisions with thermal-agitated surrounding fluid molecules provides a well-defined random background dependent on the temperature and the fluid viscosity (Chapter 16).

The motion of a Brownian particle in one dimension can be modeled by the following Langevin equation:

$$m\ddot{x} = -\gamma\dot{x} + \sqrt{2k_B T \gamma} \xi(t), \quad (17.7)$$

where x is the particle position, m is its mass, γ is the friction coefficient, $\sqrt{2k_B T \gamma} \xi(t)$ is the fluctuating force due to random impulses from the many neighboring fluid molecules, k_B is Boltzmann's constant, and T is the absolute temperature. Equation (17.7) can be solved numerically by considering the corresponding finite-difference equation,

$$m \frac{x_i - 2x_{i-1} + x_{i-2}}{\Delta t^2} = -\gamma \frac{x_i - x_{i-1}}{\Delta t} + \sqrt{2k_B T \gamma} \frac{1}{\sqrt{\Delta t}} \xi_i. \quad (17.8)$$

whose solution for x_i is

$$x_i = \frac{2 + \Delta t \frac{\gamma}{m} x_{i-1}}{1 + \Delta t \frac{\gamma}{m}} - \frac{1}{1 + \Delta t \frac{\gamma}{m}} x_{i-2} + \frac{\sqrt{2k_B T \gamma}}{m [1 + \Delta t \frac{\gamma}{m}]} \Delta t^{\frac{3}{2}} \xi_i. \quad (17.9)$$

The ratio $\tau = m/\gamma$ is the momentum relaxation time, i.e., the time scale of the transition from smooth ballistic behavior to diffusive behavior. The time τ is very small, typically of the order of a few nanoseconds [5, 7] for a micron-sized colloidal particle suspended in water. Note that this is very often orders of magnitude smaller than the time scales of typical experiments. Only since 2010, it has been possible to experimentally measure the particle position sufficiently fast to probe its instantaneous velocity with such a resolution to observe the transition from the ballistic to the diffusive regime [8]¹. Thus, for the typical acquisition frame rate in experiments, the effect of the inertia for microscopic particles suspended in a liquid is negligible. We can therefore ignore the inertial term ($m = 0$) in Eq. (17.7) and obtain the *overdamped Langevin equation*

$$\gamma \dot{x} = \sqrt{2k_B T \gamma} \xi(t), \quad (17.10)$$

corresponding to the finite-difference equation:

$$\gamma \frac{x_i - x_{i-1}}{\Delta t} = \sqrt{2k_B T \gamma} \frac{1}{\sqrt{\Delta t}} \xi_i. \quad (17.11)$$

The solution of Eq. (17.11) for x_i is:

$$x_i = x_{i-1} + \sqrt{2D \Delta t} \xi_i. \quad (17.12)$$

where $D = \frac{k_B T}{\gamma}$ is the particle's *diffusion coefficient*. If we compare this equation to the original difference equation with mass (Eq. (17.9)), Eq. (17.12) is a very good approximation to Brownian motion for long time steps, i.e., for $\Delta t \gg \tau$, but it shows clear deviations at short time scales, i.e., $\Delta t \ll \tau$. We will observe these differences in the next example.

¹This was possible in air environment where the viscosity is approximately 200 times smaller than in water environment. Since then many experiments have managed to measure inertial time scales also in water [9].

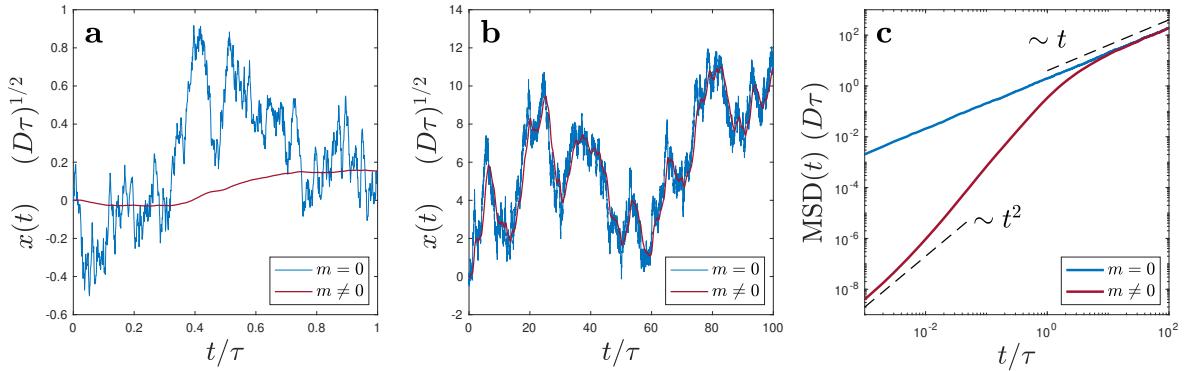


Figure 17.5: **Brownian motion in inertial and viscous regimes.** (a) Trajectories of a Brownian particle using the full Langevin equation with inertia (Eq. (17.7), red line) and the overdamped Langevin equation without the inertial term (Eq. (17.10), blue line). Here, $\Delta t = 0.01\tau$, with $\tau = m/\gamma$ the momentum relaxation time. Since the timescales are smaller than or comparable to τ , there is a clear difference in the solutions. (b) The same trajectories but simulated for a much longer time ($T_{\text{tot}} = 100\tau$) largely overlap to the point that it becomes difficult to observe the microscopic details. (c) Mean square displacement (MSD) measurements of the solutions with inertia (blue line) and without inertia (red line). The statistical properties of the two trajectories become identical after $t > 10\tau$. The black dashed lines highlight the behavior of the MSD at small and at large t/τ ($\text{MSD}(t) \propto t^2$ and $(\text{MSD}(t) \propto t)$, respectively).

Example 17.4: Brownian particle in one dimension. Consider a spherical Brownian particle with radius $R = 1 \mu\text{m}$ and mass $m = 11.1 \cdot 10^{-15} \text{ kg}$, suspended in a liquid ($\eta = 0.001 \text{ N s m}^{-1}$ for water at room temperature) and freely diffusing at room temperature ($T = 300 \text{ K}$). Simulate the trajectory of the particle according to the Langevin equation. First, consider the inertia (Eq. (17.9)) and then the case without the inertial term (Eq. (17.12)). Take $\Delta t < 0.1\tau$ and simulate trajectories at least $T_{\text{tot}} \approx 100\tau$ long. [Use the same sequence of random numbers in the two cases so that the two trajectories experience the same white noise and can be easily compared.]

Let us call $x_m(t_i)$ and $x_0(t_i)$ the trajectory with and without mass.

a. Plot the $x_m(t_i)$ and $x_0(t_i)$ for $0 \leq t_i \leq \tau$ and observe that the outcome is quite different, as shown in Fig. 17.5a. This means that ignoring the mass would not give accurate results for timescales that are comparable to or smaller than τ ($m/\gamma = 0.588 \mu\text{s}$ in this case).

b. Plot $x_m(t_i)$ and $x_0(t_i)$ for $0 \leq t_i \leq T_{\text{tot}}$ and observe that the outcome is very similar and the two trajectories overlap almost perfectly, as shown in Fig. 17.5b.

c. Compute the MSDs for both cases as an ensemble average. Show that the MSDs converge for long time scales, as shown in Fig. 17.5c.

d. *Brownian particle in n dimensions.* Simulate Brownian motion in n dimensions, with $n = 2$ and 3.

17.4 Active Brownian motion

In order to model the behavior of biological and artificial microscopic self-propelling objects (or microswimmers), we consider Brownian particles that are able to propel themselves in a certain direction. Such particles exist widely both in nature and in laboratories [10]. Differently from Brownian particles, they display a directed motion that creates interesting non-equilibrium behaviors. In this section, we will have a closer look at their motion dynamics.

17.4.1 Active Brownian particles

Let us consider a single active Brownian particle characterized by a constant swimming speed v . This particle is subject to collisions with the molecules of the surrounding fluid, which perturb its trajectory and cause the particle to move randomly (described by a *translational diffusion* term). The orientation of this active particle will also fluctuate due to these collisions (described by a *rotational diffusion* term). Therefore, the motion of the particle is described by the following set of equations:

$$\begin{cases} \frac{dx(t)}{dt} = v \cos \theta(t) + \sqrt{2D} \xi_x, \\ \frac{dy(t)}{dt} = v \sin \theta(t) + \sqrt{2D} \xi_y, \\ \frac{d\theta(t)}{dt} = \sqrt{2D_R} \xi_\theta, \end{cases} \quad (17.13)$$

where x and y are the position coordinates, θ is its instantaneous orientation, and D and D_R are the translational and rotational diffusion coefficients, respectively. ξ_x , ξ_y , and ξ_θ are independent random processes with zero mean and unitary variance. Numerically, they consist of a sequence of random numbers chosen from a Gaussian distribution with zero mean and variance equal to one.

Example 17.5: Trajectories of a single active Brownian particle. Write a program generating and plotting active particle trajectories, according to Eq. (17.13), using a finite-difference method with a constant time step Δt . Compare with Fig. 17.6.

a. As a starting point, you can consider the following values for the various parameters: $\Delta t = 0.001 \text{ s}$, $D = 5 \cdot 10^{-14} \text{ m}^2 \text{s}^{-1}$, $D_R = 10^{-1} \text{ s}^{-1}$, $v = 3 \cdot 10^{-6} \text{ m}^2 \text{s}^{-1}$, which are values in the range of what is normally observed in experiments with artificial microscopic active particles or with biological microswimmers. Generate a trajectory and compare it with the typical trajectory for ABP illustrated in Fig. 15.3b.

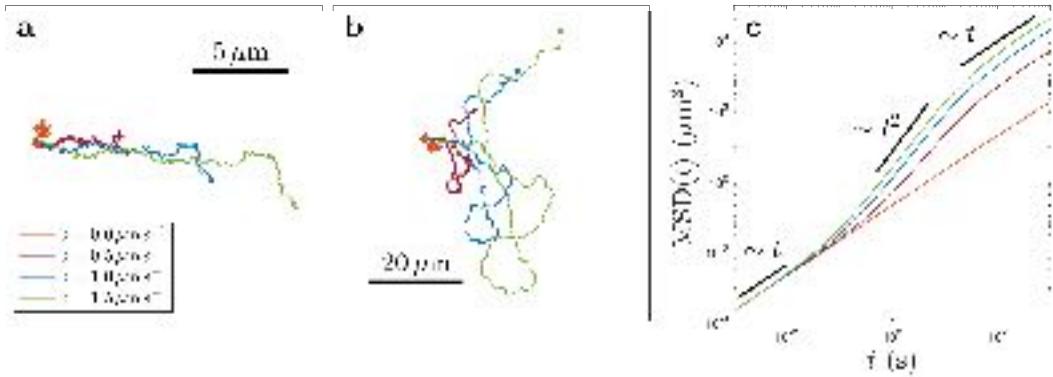


Figure 17.6: Active Brownian motion. Sample trajectories of active Brownian particles with different swimming velocities. The trajectories have been generated with the same thermal noise. The other parameters of the simulation are $D = 5 \cdot 10^{-2} \mu\text{m s}^{-1}$, $D_R = 1 \cdot 10^{-1} \text{s}^{-1}$ (which gives a characteristic time $\tau_R = 10\text{s}$), and $\Delta t = 10^{-4} \tau_R$. **(a)** The trajectories show a total time evolution $T_{\text{tot}} = \tau_R$. On this time scale, the motion of the active Brownian particles is approximately linear and the displacement is proportional to the self-propulsion speed v with a good approximation. The comparison with a passive Brownian particle ($v = 0$) is shown. **(b)** The trajectories show a total time evolution $T_{\text{tot}} = 10\tau_R$. At this time scale, the average speed starts deviating significantly from the self-propulsion velocity of the particle. The motion starts its transition to the superdiffusive regime at long time intervals. **(c)** Corresponding mean square displacements of the particles: the motion is diffusive ($\text{MSD} \sim t$) for $t \ll \tau_R$, ballistic ($\text{MSD} \sim t^2$) for $t \approx \tau_R$, and superdiffusive ($\text{MSD} \sim t$ with an enhanced diffusion constant D_{eff}) for $t \gg \tau_R$.

b. Play with the parameters: What happens if you change the swimming speed v ? Compare the plots for different values of v .

c. What happens if you change the value of the translational diffusion D ?

d. What happens if you change the value of the rotational diffusion D_R ?

e. What happens if you use a different value of the time step Δt for the simulation?

f. Calculate the MSD and compare it with the theoretical one for an active Brownian particle in two dimensions:

$$\text{MSD}(t) = (4D + v^2 \tau_R) t + \frac{v^2 \tau_R^2}{2} \left(e^{-\frac{2t}{\tau_R}} - 1 \right) \quad (17.14)$$

where $\tau_R = D_R^{-1}$ is the relaxation time for the orientation of the particle and t is the time interval after which we measure the MSD from the initial position of the particle. Verify that the motion of an active particle is ergodic. Compare the MSD dependence obtained with the ABP trend illustrated in Fig. 15.4.

g. *Active Brownian particle in 3D.* Simulate active Brownian motion in three dimensions.

17.4.2 Chiral active Brownian particles

Some microorganisms (like swimming bacteria, or plankton) or artificial swimmers can move in quasi-circular motion in a clockwise or counterclockwise fashion [11, 12, 13, 14, 15, 16, 17]. This motion is commonly referred to as *chiral active Brownian motion*. To describe this kind of dynamics, we have to take into account that now the orientation varies with a defined *angular velocity* ω :

$$\begin{cases} \dot{x} = v \cos \theta + \sqrt{2D} \xi_x, \\ \dot{y} = v \sin \theta + \sqrt{2D} \xi_y, \\ \dot{\theta} = \omega + \sqrt{2D_R} \xi_\theta. \end{cases} \quad (17.15)$$

According to the standard convention on the direction of angles, $\omega > 0$ will be associated to a *counterclockwise* motion, while $\omega < 0$ will characterise a *clockwise* motion. When translated to a finite-difference equation

formalism, we have:

$$\begin{cases} x_{n+1} = x_n + v \cos \theta \Delta t + \sqrt{2D\Delta t} \xi_{x,n}, \\ y_{n+1} = y_n + v \sin \theta \Delta t + \sqrt{2D\Delta t} \xi_{y,n}, \\ \theta_{n+1} = \theta_n + \omega \Delta t + \sqrt{2D_R\Delta t} \xi_{\theta,n}. \end{cases} \quad (17.16)$$

Example 17.6: Chiral active Brownian motion in 2D. Implement the finite-difference equations (17.16) and generate trajectories for different values of ω , D_R and D . Calculate the mean square displacement of the chiral trajectories for different ω . Compare the trajectory obtained with the typical trajectory for CAP illustrated in Fig. 15.3e. Compare the MSD dependence obtained with the CAP trend illustrated in Fig. 15.4.

Example 17.7: Chiral active Brownian motion in 3D. Generalize Eq. (17.15) for the case of a chiral active Brownian particle in three dimensions. It can be of use to associate a unit vector $\hat{\mathbf{u}}$ and a polar axis $\hat{\mathbf{p}}$ to the active particle, in such a way that the propulsion is always along $\hat{\mathbf{u}}$ and the angular velocity ω acts around the polar axis $\hat{\mathbf{p}}$. Write the corresponding finite-difference equation and generate trajectories for different values of ω , D_R , and D . Calculate the mean square displacement of the chiral trajectories for different ω .

17.4.3 Run-and-tumble particles

A very common pattern for the motion of microorganisms like *E. coli* is what is known as the *run-and-tumble motion*. The observed motion can be described as a sequence of rectilinear forward steps and, occasionally, a sudden stop in the motion followed by a reorientation and by rectilinear motion along the new direction. This kind of motion is typical of chemotactic microorganisms and its mechanism has been thoroughly studied in *E. coli* [18]. Various models have been developed to simulate it. In the absence of the chemical substance to which the chemotactic organism is responding, the reorientation events happen with a timing well-described by a Poisson process, characterized by the probability distribution:

$$P_\lambda(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad (17.17)$$

where N is the number of events observed in the time interval Δt , λ is the average number of events expected in Δt , and n is a natural number. The probability that tumbling happens is

$$P_{\text{tumble}} = 1 - P(N = 0) = 1 - e^{-\lambda}. \quad (17.18)$$

Therefore, to describe a run-and-tumble active particle, we add then a variable ρ to keep track of the status of the active particle (either run or tumble), and the characteristic angle Θ_0 characterizing the distribution of the tumble angle, that we think as a flat distribution in the interval from $[-\Theta_0, \Theta_0]$. The Langevin equation for this particle is

$$\begin{cases} \dot{x} = v \cos \theta + \sqrt{2D} \xi_x, \\ \dot{y} = v \sin \theta + \sqrt{2D} \xi_y, \\ \dot{\theta} = \sum_k \Theta_{\text{tumble},k} \delta(t - T_k), \end{cases} \quad (17.19)$$

where the sum over k is over the tumbling events happening at T_k . Eq. (17.19) corresponds to the finite difference set of equations:

$$\begin{cases} x_{n+1} = x_n + v \cos \theta_n \Delta t + \sqrt{2D\Delta t} \xi_{x,n}, \\ y_{n+1} = y_n + v \sin \theta_n \Delta t + \sqrt{2D\Delta t} \xi_{y,n}, \\ \theta_{n+1} = \theta_n + (1 - \rho_n) \Theta_{\text{tumble},n}, \\ \Theta_{\text{tumble},n} = \Theta_0 \xi_{\theta,n}, \\ \rho_{n+1} = 0 \text{ or } 1 \text{ with probability } 1 - \lambda_t \text{ and } \lambda_t \text{ each step,} \end{cases} \quad (17.20)$$

where $\xi_{u,n}$ represents a random number extracted from a uniform distribution on the interval $[-1, 1]$.

Example 17.8: Run-and-tumble motion in 2D. Implement the finite-difference equations (17.20) corresponding to the run-and-tumble particle and generate trajectories for different values of λ_t , Θ_0 , and D . Calculate the mean square displacement of the run-and-tumble trajectories for different λ_t . Compare the trajectory obtained with the typical trajectory for RTP illustrated in Fig. 15.3c. Compare the MSD dependence obtained with the RTP trend illustrated in Fig. 15.4.

Example 17.9: Run-and-tumble motion in 3D. Implement some active particles performing a run-and-tumble motion in three dimensions.

17.4.4 Active Ornstein-Uhlenbeck particles

Another possible description for active particles is the *active Ornstein-Uhlenbeck particle*, which is a particle with mass m and obeying the equations of motion (in 2D):

$$\begin{cases} \dot{x} = v_x, \\ \dot{y} = v_y, \\ \tau_p \dot{v}_x = -v_x + \sqrt{2D}\xi_x, \\ \tau_p \dot{v}_y = -v_y + \sqrt{2D}\xi_y, \end{cases} \quad (17.21)$$

where $\tau_p = m/\gamma$ is the momentum relaxation time. It is worth noting that this motion does not have any stochastic component acting directly on the coordinates, but there is a noise term affecting the velocity. This means that, contrarily to what happens in all the other models, the speed of the active motion is not constant in time, but changes slightly. The system of stochastic equations (17.21) corresponds to the finite-difference system ($\Delta t \ll \tau_p$):

$$\begin{cases} x_{n+1} = x_n + v_{x,n} \Delta t, \\ y_{n+1} = y_n + v_{y,n} \Delta t, \\ v_{x,n+1} = v_{x,n} - \frac{v_{x,n}}{\tau_p} \Delta t + \frac{\sqrt{2D\Delta t}}{\tau_p} \xi_{x,n}, \\ v_{y,n+1} = v_{y,n} - \frac{v_{y,n}}{\tau_p} \Delta t + \frac{\sqrt{2D\Delta t}}{\tau_p} \xi_{y,n}. \end{cases} \quad (17.22)$$

Example 17.10: Active Ornstein-Uhlenbeck particle motion in 2D. Implement the finite-difference equations (17.22) corresponding to the active Ornstein-Uhlenbeck particle and generate trajectories for different values of τ_p and D . Compare with the typical trajectory for AOUP illustrated in Fig. 15.3d. Calculate the mean square displacement of the generated trajectories. Compare with the trend illustrated in Fig. 15.4.

Example 17.11: Active Ornstein-Uhlenbeck particle motion in 3D. Implement an active Ornstein-Uhlenbeck particle in three dimensions.

An interesting exercise is to compare the MSDs for the different models of active particle described above. As mentioned in Chapter 15, the behavior of the MSD is *universal* at long times.

Example 17.12: MSD of the different models of active particle. From the trajectories generated in Examples 17.5-17.10, calculate the MSD and compare with the trend illustrated in Fig. 15.4. [Hint: in order to obtain Fig. 15.4 you have to express t in units of the characteristic time scale and the MSD in units of the square of the characteristic length scale. Check the definition given in Table 15.2.]

17.4.5 Non-spherical active particles

Motile microorganisms and particles are often of non-spherical shape, and, especially in the case of living microorganisms, their shape changes in time while self-propelling. While how to describe a soft deformable particles is beyond the scope of this Chapter, we will give a brief introduction on how to deal with rigid, non-spherical particles.

Non-spherical particles feel a different interaction with the surrounding fluid because of their shape, i.e., their diffusion matrix is, in the most simple case, separable into two contributions, the *translational* diffusion matrix \mathbb{D}_T , and the *rotational* diffusion matrix \mathbb{D}_R . Both are 3×3 matrices that can be put in a diagonal form, but they are not necessarily proportional to the 3×3 identity matrix \mathbb{I}_3 . In general, we can write them as:

$$\mathbb{D}_T = \begin{bmatrix} D_x & 0 & 0 \\ 0 & D_y & 0 \\ 0 & 0 & D_z \end{bmatrix}, \quad \mathbb{D}_R = \begin{bmatrix} D_{x,R} & 0 & 0 \\ 0 & D_{y,R} & 0 \\ 0 & 0 & D_{z,R} \end{bmatrix}, \quad (17.23)$$

where the three translational diffusion constants ($D_{x,T}$, $D_{y,T}$, and $D_{z,T}$) and the three rotational diffusion constants ($D_{x,R}$, $D_{y,R}$, $D_{z,R}$) are not necessarily equal. Considering, for example, the case of a prolate ellipsoid where the long semi-axis is indicated with z , we have that $D_{x,T} = D_{y,T} \neq D_{z,T}$ and $D_{x,R} = D_{y,R} \neq D_{z,R}$. We emphasize that the denomination x , y , and z is relative to the *particle* reference frame, and not to the *lab* reference frame [19, 20, 21]. In the case of more complex shapes, there might also be a roto-translational coupling. The generic diffusion matrix for a generic 3D rigid shape is then

$$\mathbb{D} = \begin{bmatrix} \mathbb{D}_T & \mathbb{D}_{RT} \\ \mathbb{D}_{TR} & \mathbb{D}_R \end{bmatrix}, \quad (17.24)$$

where \mathbb{D}_{TR} and \mathbb{D}_{RT} are the roto-translational terms (3×3 matrices) of the full diffusion matrix \mathbb{D} , which is a 6×6 matrix.

The overdamped Langevin equation to describe the particle motion is:

$$\begin{bmatrix} \mathbf{v} \\ \boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} v_0 \hat{\mathbf{u}} \\ \omega_0 \hat{\mathbf{p}} \end{bmatrix} + \boldsymbol{\Xi}, \quad (17.25)$$

where v_0 is the active speed of the particle, ω_0 its active angular speed (in the case it is a chiral active Brownian particle), \mathbf{v} is the instantaneous 1×3 particle velocity vector, $\boldsymbol{\omega}$ is the instantaneous 1×3 particle angular velocity vector, $\boldsymbol{\Xi}$ is a 1×6 noise term whose correlation matrix has variance defined via the diffusion matrix \mathbb{D} [20]. It is worth noting that, because the noise term is written in the particle reference frame, to calculate the actual contribution in the lab reference frame one must transform from lab to particle reference frame and back [20, 21]. If we restrict ourselves to considering a 2D motion of an ellipsoidal particle, the set of equation simplify considerably. In fact, the torque acting on the particle is described by a single diffusion parameter D_R acting on the orientation θ , and the translational diffusions in the particle's x and y direction are characterized by two different parameters D_x and D_y .

Example 17.13: Elliptical active Brownian particle. Inspired by Eq. (17.13) and by the former paragraph, write the finite-difference equations for an elliptical active Brownian particle in 2D. Implement a program to generate trajectories and calculate the mean square displacement. Compare with the case of a *spherical* active Brownian particle (i.e., described by Eq. (17.13)).

Example 17.14: Ellipsoidal active Brownian particles. Derive the finite-difference equations for the system Eq. (17.25) in the case of a prolate ellipsoid. For the diffusion matrix, refer to the Perrin's coefficients [22].

17.5 Complex and crowded environments

In realistic systems and applications, active Brownian particles explore a complex environment, with external force fields, obstacles, and other particles. In this section, we discuss how to simulate Brownian systems in complex and crowded environments. For simplicity's sake, the examples will be given for the active Brownian particle model, but the same formalism can be used to describe the effects of complex and crowded environments on other active particle models.

17.5.1 External force fields

Active particles can be affected by an external force field: for example, they can feel gravity, or the effect of an electric field, or of an optical force. A typical case could be an active Brownian particle in a harmonic potential: for instance, an optically trapped motile microorganism, or a Janus particle in an optical potential [23, 24, 25, 26, 27].

When an active Brownian particle is subject to a force field, its equation of motion includes a term involving the deterministic force felt by the particle. The resulting Langevin equation is then:

$$\begin{cases} \dot{x} = v \cos \theta + F_x(x, y)/\gamma + \xi_x, \\ \dot{y} = v \sin \theta + F_y(x, y)/\gamma + \xi_y, \\ \dot{\theta} = \xi_\theta, \end{cases} \quad (17.26)$$

where $F_x(x, y)$ and $F_y(x, y)$ are the force fields acting in the x- and y-directions, respectively. When translated to a finite-difference formalism, we have:

$$\begin{cases} x_{n+1} = x_n + (v \cos \theta + F_x(x_n, y_n)/\gamma) \Delta t + \sqrt{2D\Delta t} \xi_{x,n}, \\ y_{n+1} = y_n + (v \sin \theta + F_y(x_n, y_n)/\gamma) \Delta t + \sqrt{2D\Delta t} \xi_{y,n}, \\ \theta_{n+1} = \theta_n + \sqrt{2D_R\Delta t} \xi_{\theta,n}. \end{cases} \quad (17.27)$$

If there is also a torque, the equation for the angular variable is $\dot{\theta} = T_z(x, y)/\gamma_R + \xi_\theta$ and the corresponding finite-difference equation becomes $\theta_{n+1} = \theta_n + (T_z(x_n, y_n)/\gamma_R) \Delta t + \sqrt{2D_R\Delta t} \xi_{\theta,n}$, where γ_R is the rotational friction coefficient.

Example 17.15: Active Particle in an optical potential. An active Brownian particle moving in 2D with self-propulsion speed v is subject to an optical trap, which is described by a harmonic potential $U(r) = \frac{1}{2}kr^2$, with $r = \sqrt{x^2 + y^2}$ the radial distance from the center of the trap. Simulate its trajectory. Calculate the MSD and compare it with the case of a passive Brownian particle in the same optical trap.

Example 17.16: Active Brownian particle subject to external force and torque in 3D. Generalize the Langevin equations for a particle subject to an external force field in 3D. Consider also the case of an external torque.

Example 17.17: Critical motor. Ref. [25] shows a microscopic motor powered by critical demixing. A silica particle with iron oxide nano-inclusions suspended in a critical solution of water-lutidine becomes an active particle when the amount of optical power absorbed by the iron oxide nano-cores is enough to determine a partial demixing of the critical solution in its surroundings. When optically trapped, a circular motion is observed: part of the optical power absorbed by the particle is transformed into rotational motion, *de facto* generating a microscopic engine.

A simplified model to obtain a microscopic engine is the following: an active Brownian particle set in the position (x, y) has propulsion velocity $v(r) = v_0 e^{-\frac{r^2}{2\sigma^2}}$, where $r = \sqrt{x^2 + y^2}$. The particle is subject to an effective potential $U(r) = \frac{1}{2}k(r - r_0)^2$, and to a torque T_z whose magnitude depends on its radial distance r : $T_z(r) = T_0 e^{-\frac{r^2}{2\sigma^2}}$.

Implement a code and generate trajectories for the particle trying different ratios T_0/v_0 . Said $x(t)$ and $y(t)$ the two components of the trajectory with respect to the center of the potential, calculate the differential cross correlation function $\mathcal{D}_{x,y}$. Show that the ensuing behavior resembles the corresponding $\mathcal{D}_{x,y}$ of figure 2 of Ref. [25]. Calculate the frequency and the power of the engine.

Example 17.18: Active particle in a speckle potential. Generate an optical potential

$$U(x, y) = \sum_{i=1}^N -U_0 e^{-\frac{(x-x_{0,i})^2+(y-y_{0,i})^2}{2\sigma_i^2}},$$

which could correspond to the one of a speckle light field. Generate the trajectories for an active Brownian particle with characteristic propulsion speed v_0 . Compare the trajectories for different v_0/U_0 , and show that below a certain v_1 the motion of the particle remains localized in a small set of potential wells, while above a certain v_2 the motion is not localized. [See [28] for further details.]

17.5.2 Obstacles

A complex environment can be created by placing obstacles that occupy part of the space and prevent the particle to transit across them. The interaction between the particles and the obstacles can be relatively simple (e.g., reflective boundary conditions, as described in Section 16.4) or more complex (e.g., inducing also a torque causing a reorientation of the particle). While passive Brownian particles diffuse uniformly through environments with obstacles, the interaction between active particles and obstacles can lead to interesting emerging phenomena, e.g., the concentration of bacteria in certain regions [29, 30] and the emergence of a preferential direction when crossing a periodic lattice [31].

Example 17.19: Active particle moving in a periodic lattice of obstacles. Inspired by Ref. [31], we simulate the trajectory of an active particle in a periodic lattice of obstacles.

a. Build a simplified version of the experimental lattice of Ref. [31] by taking a rectangular lattice. Place an obstacle on each lattice point, and an additional obstacle in the center of each elementary cells. Simulate the interaction of the active particle with the obstacle with a constant torque T_0 applied to the particle when approaching the obstacle surface, as exemplified in Fig. 17.7a. The torque starts acting on the particle when at a distance h_{int} from the wall and acts in such a way to divert the particle from the obstacle. In the case the particle hits the obstacle it cannot trespass it, but the torque continues acting and the particle is able to slide along the obstacle in the direction of the component of the velocity parallel to the obstacle surface.

b. Generate the trajectory of a particle for different values of its propulsion speed v , in the presence of a constant force drift F_d directed perpendicular to the obstacles. Show that the observed behavior is consistent with Fig. 17.7b-d.

c. Show that the microscopic self-propelled particles, in the presence of periodic obstacles, can steer

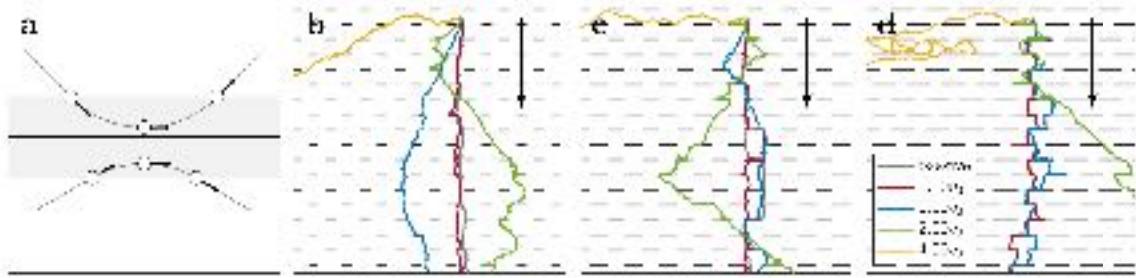


Figure 17.7: Active particle moving in a periodic lattice of obstacles. (a) Interaction of the active particle with the obstacle. When the particle enters the interaction range of the obstacle (gray shade) it feels a constant torque that drives the particle away from the obstacle. The black arrows highlight the propulsion velocity of the particle in the entrance point of the interaction zone, in the nearest point to the obstacle, and in the exit point of the interaction zone. Two example trajectories (particle approaching from above and from below, for two different angles) are shown. (b-d) Sample trajectories of an active particle for different values of its propulsion speed v (color coded according to v) in differently-spaced periodical lattice of obstacles under the effect of a constant drift $F_d = 2.65 F_0$ (arrow), where F_0 is the reference value for the force $F_0 = k_B T \cdot v_0 / D$. For generating this picture we took the reference values of translational and rotational diffusion of a spherical particle with radius $R = 1 \mu\text{m}$ in water (viscosity $\eta = 0.001 \text{ N s m}^{-1}$): $D = 0.22 \mu\text{m}^2 \text{s}^{-1}$ and $D_r = 0.16 \text{ s}^{-1}$. The reference propulsion speed is $v_0 = 1 \mu\text{m s}^{-1}$, the force drift is $F_d = 0.1 \text{ pN}$, and the interaction torque is $T_{\text{obs}} = 0.05 \text{ pN } \mu\text{m}$. See also [31].

even perpendicularly to an applied drift force [31].

[A choice for the reference values for the simulation: $D = 0.22 \mu\text{m}^2 \text{s}^{-1}$ and $D_r = 0.16 \text{ s}^{-1}$ (which are the values for a spherical particle of radius $R = 1 \mu\text{m}$ in water), reference propulsion speed $v_0 = 1 \mu\text{m s}^{-1}$, force drift $F_d = 0.1 \text{ pN}$, interaction torque $T_{\text{obs}} = 0.05 \text{ pN } \mu\text{m}$. The lattice has a unit cell of $10 \mu\text{m} \times 8.66 \mu\text{m}$. The length of the obstacles are taken between 1 and 9 μm .]

Example 17.20: Active particle in a wall of funnels. Inspired by Ref. [29], we simulate the trajectory of

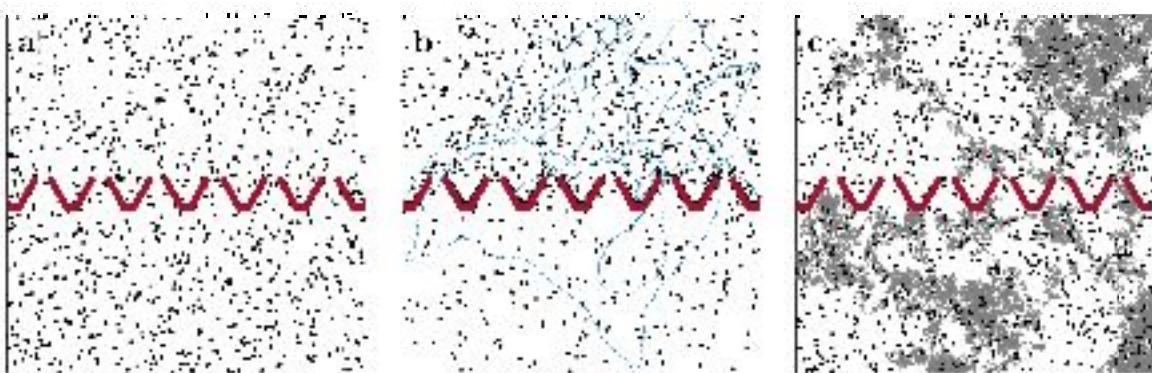


Figure 17.8: Active particle with a wall of funnels. (a) Initial condition for the simulation with $N = 1000$ active particles ($v = 10 \mu\text{m s}^{-1}$). (b) Final position of the active particles after 100 s. See also Ref. [29, 32]. From an initial homogeneous density configuration, the active particles get concentrated in the top of the arena, because of the presence of the wall of funnels. On the contrary, (c) the distribution of passive Brownian particles remains homogeneous even after a much longer time (10^4 s). Representative trajectories are shown in blue (active Brownian particle) and in grey (passive Brownian particle).

an active particle in a box that is split in a half by a wall of funnels.

a. Build a simplified version of the experimental setup of Ref. [29] and implement the interaction of the particles with the obstacles with the standard reflective boundary conditions (see Section 16.4), though without altering their orientation after the interaction with the obstacle. It is useful to consider two dimensional obstacles, i.e., by assigning a thickness to each of them. This will prevent spurious crossing of the obstacles due to the finite time step of the simulation.

b. Generate the trajectory of an active Brownian particle for different values of its propulsion speed v , and for different size w of the funnel aperture. Note that it is more likely for an active particle to move in the *natural* direction of the funnel than opposite to it.

c. Generate several trajectories and show that the density of the active particle increases asymmetrically in time, i.e., the particles end up concentrating in one of the two compartments. Also show that, for passive Brownian particles, the density of the system remains homogeneous in time.

[Compare the results of your simulation with Fig. 17.8.]

17.5.3 Interacting particles

Both passive and active systems are often made up of several identical particles. Unlike what we have seen in Chapter 16, where we considered exclusively examples of ideal point-like particles, we consider here the case of particles with a finite, though microscopic, size. For example, in a system of colloidal spheres, the particles have a well defined size and are rigid (their constituent material is commonly silica or polystyrene) so that they behave like hard spheres. Moreover, additional interactions between the particles could be relevant. It is therefore crucial to describe correctly such interactions.

Let us start by considering the case of a system composed of many particles. In order to have a realistic simulation, one should keep into account that the particles will never overlap, i.e., they should not occupy the same physical space. When two particles are approaching, collisions might happen. In the case of hard spherical colloids this is well reproduced, in simulation, using the so-called *volume exclusion method* or *hard-sphere correction*: when an overlap between two particles is found, both particles are moved away from each other along the direction connecting the centers to a new configuration in which they touch each other, but are not superposed [33, 21]. Note that the common center of mass of the two particles should remain in the same place before and after the particle rearrangement.

Example 17.21: Hard-sphere interaction with the excluded volume algorithm. Write a program gen-

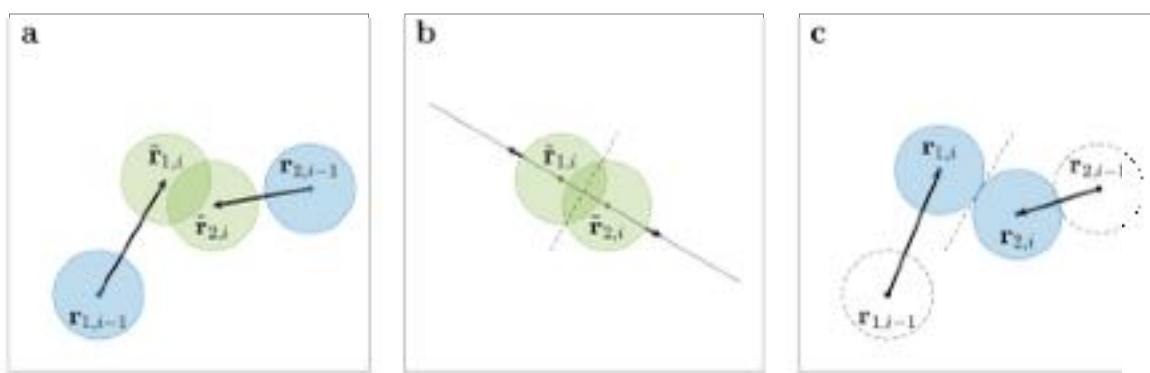


Figure 17.9: **Volume exclusion method.** (a) Two particles with finite radius, initially in the non-overlapping configuration $\mathbf{r}_{1,i-1}$ and $\mathbf{r}_{2,i-1}$ at time step $t_{i-1} = (i-1)\Delta t$ (blue circles), overlap at t_i . The temporary positions are $\hat{\mathbf{r}}_{1,i}$ and $\hat{\mathbf{r}}_{2,i}$ (green circles). (b) The particles are displaced equally (arrows) along the line connecting their centers (solid straight line). The displacement amount is such that the particle are tangent to each other, but do not overlap any longer. (c) The new displaced positions $\mathbf{r}_{1,i}$ and $\mathbf{r}_{2,i}$ become the new particle positions at time t_i . It is important to note that, in the process, the center of mass of the two particles has not changed.

erating and plotting active particle trajectories, according to Eq. (17.13), using a finite-difference method with a constant time step Δt , for a system of N particles, implementing the *volume exclusion method* (Fig. 17.9).

a. Write a function implementing the hard-sphere correction for a single pair of particles with radius R . Test your function: set the particles with their centres at a distance $d < 2R$, and check the new positions given by your function. The midpoint of the segment connecting the centres of the two particles should not have changed, and the two particles should touch each other, but not overlap, after the function has been applied.

b. After your function for the hard-sphere correction works as required, write a code for the dynamics of $N = 100$ active particles. To begin with, use the following set of parameters, which are reasonable for a set of identical colloidal particles with $R = 1 \mu\text{m}$ in water: $D = 0.1 \mu\text{m}^2\text{s}^{-1}$, $D_R = 1 \text{s}^{-1}$, $v = 3 \mu\text{m s}^{-1}$, $\Delta t = 0.01 \text{s}$. Check that the behavior of the simulation is appropriate. Use periodic boundary conditions and choose an appropriate box dimension to contain all the particles with no superposition.

c. How important is the choice of the time step Δt for a correct, physical behavior of the simulation? What if, for instance, you take $\Delta t = 10 \text{s}$? Is the choice of an appropriate Δt affected by the size of the particles? And by their active speed?

d. How is the time step Δt affected by the concentration of the particles?

Living crystals

We now study the behavior of a set of active hard spherical particles interacting only at short range via a phoretic attractive force. This model has been proposed to explain the experimental observation of Ref. [23], where a set of light-activated colloids, propelled by a phoretic force, formed metastable clusters that were called living crystals. The active building blocks of the clusters were polymer colloidal spheres with a protruding hematite cube, immersed in a basic solution containing hydrogen peroxide, and were activated via a blue light, which triggered the decomposition of the hydrogen peroxide at the exposed hematite surface. The interaction between the particles was phoretic and induced an attractive force, giving the particles a phoretic velocity $v_P(\Delta r)$ that scales according to the inverse of the square distance between the particles as

$$v_P(\Delta r) = v_{P0} \frac{r_0^2}{(\Delta r)^2}. \quad (17.28)$$

This force is significant only between particles lying within a given radius R_{ref} from one another.

This phoretic interaction is due to the advective flow generated by the decomposition of hydrogen peroxide on the exposed hematite surface of the light-activated colloids. The inverse of the square distance dependence is the expected behavior for a phoretic attraction to a reaction source, and fits the behavior observed experimentally, as explained in Ref. [23].

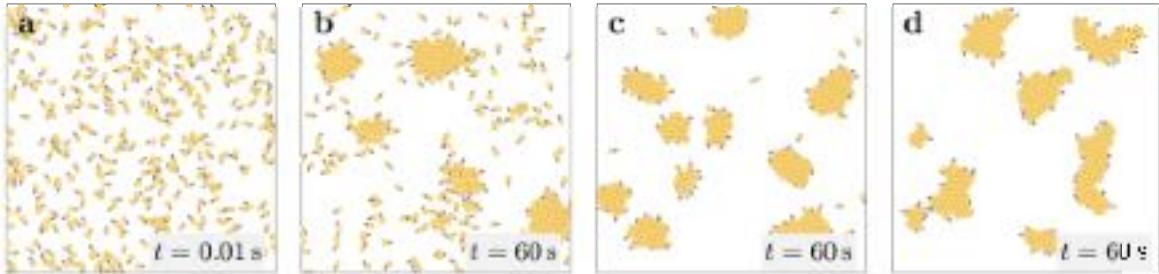


Figure 17.10: **Living crystals.** Formation of metastable living crystals for different values of the phoretic velocity parameter. **(a)** Initial configurations with $N = 400$ particles. Final configuration after 60 s for a phoretic velocity value of **(b)** $v_{P0} = 1 \mu\text{m}$, **(c)** $v_{P0} = 1.5 \mu\text{m}$, **(d)** $v_{P0} = 6 \mu\text{m}$ at $r_0 = 1 \mu\text{m}$. The phoretic interaction radius is set to $R_{\text{max}} = 10 \mu\text{m}$ in this simulation. The bigger the reference value of v_{P0} , the more stable the resulting clusters. In the case of **(b)**, the crystals are dynamical and many particles can detach from a cluster. Instead, in **(c)** and **(d)**, the clusters are more stable.

Example 17.22: Living crystals. Using the code in Example 17.21, write a program generating and plotting active particles interacting according to Eq. (17.28).

a. Write a function implementing the phoretic interaction for a single pair of particles. Your function should set properly the phoretic velocity of both particles, in function of the parameters v_{p0} and r_0 characterizing the interaction. For the set of parameters given above, start with $v_{p0} = 2 \mu\text{m}$ and $r_0 = 2 \mu\text{m}$.

b. After your function for the phoretic works as required, write a code for the dynamics of $N = 100$ active particles interacting according to Eq. (17.28), using the hard-sphere correction to avoid overlap. Choose a convenient time step Δt . Use periodic boundary conditions: consider a square box with a side of $L = 100 \mu\text{m}$ (concentration of $\theta_V = 0.03$). Describe the behavior of the system after a long time.

[Compare your results with Fig. 17.10.]

The strength of the interaction is set by the magnitude of the reference speed v_{p0} at a given reference distance r_0 . Increasing v_{p0} , the interaction becomes stronger. In the following example, we will see that a stronger interaction determines the formation of larger, more stable clusters.

Example 17.23: Living crystals revisited. Using the code in Example 17.22, explore different combinations of v_{p0} and r_0 .

a. (Stronger interaction) Take $v_{p0} = 2 \mu\text{m}$ and $r_0 = 2.5 \mu\text{m}$, which leads to a stronger interaction between the particles. What is the observed behavior of the system after a long time, if the concentration of the particles and the other parameters are left constant?

b. Vary the concentration of the particles (you can either vary the number of particles, or vary the dimension of the arena). Describe the resulting behavior. Does a higher concentration of particles determine a bigger size for the clusters?

[Compare your results with Fig. 17.10.]

Aligning interactions

We now consider a system of active Brownian particles subject to an aligning interaction in a homogeneous environment. First, we focus on how these particles interact with each other, and then how they form metastable clusters [23] thanks to the interplay of the aligning interaction and the activity [34]. The elementary particle-particle interaction is chosen to cause a torque on the interacting particles, and therefore a change in their

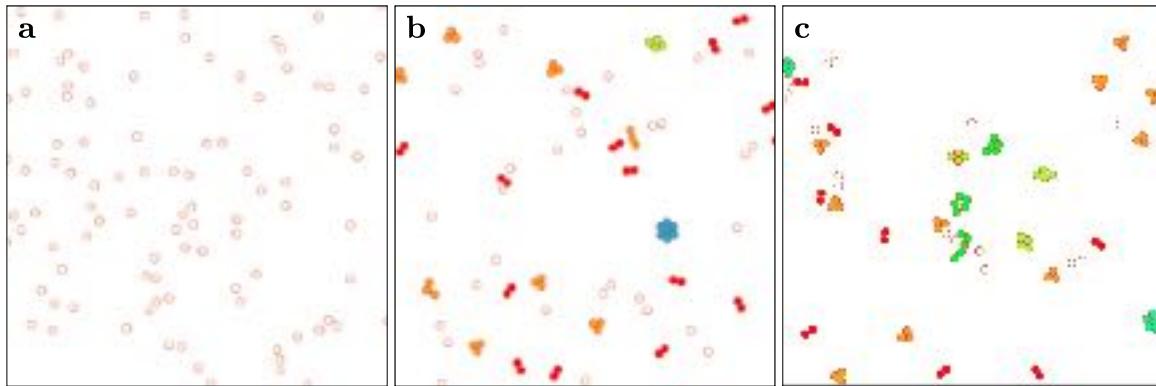


Figure 17.11: **Clustering behavior of active Brownian particles with aligning interactions.** Steady state of the system with $N = 100$ active particles with aligning interaction given in Eq. (17.29). The noise level is chosen equal to (a) $\eta = 2\pi$, (b) $\eta = 0.2\pi$, (c) $\eta = 0.002\pi$. When the amount of rotational noise is high, the particles do not cluster and their trajectories are highly erratic. Reducing appreciably the noise leads to the formation of small clusters of two or three particles, in configurations that are either non-motile or motile. Further reducing the noise favours the formation of clusters of larger size. Clusters are color-coded by size. See also [34]. Figure adapted from Ref. [34].

orientation. This interaction is considered additive, i.e., the total torque acting on a particle that is interacting with N different particles is the sum of the elementary torques given by the pair interactions. The overall torque on the particle n is then given by the following expression:

$$T_n = T_0 \sum_{i \neq n} \frac{\hat{\mathbf{v}}_n \cdot \hat{\mathbf{r}}_{ni}}{r_{ni}^2} \hat{\mathbf{v}}_n \times \hat{\mathbf{r}}_{ni} \cdot \hat{\mathbf{e}}_z \quad \text{for } r_{ni} < r_c, \quad (17.29)$$

where T_0 represents the strength of the interaction, $\hat{\mathbf{v}}_n$ is the direction of the n^{th} particle instantaneous velocity, and $\hat{\mathbf{r}}_{ni}$ is the position of particle i with respect to particle n .

Example 17.24: Aligning interaction, two particles. Write a code showing the dynamics of two active particles interacting via the aligning interaction defined in Eq. (17.29). It might be useful to write a function implementing the aligning interaction defined in Eq. (17.29) between two particles only. Run the dynamics for different values of the interaction strength T_0 . Observe that the particles tend to align and move against each other when T_0 is strong enough.

If instead of considering a system made by only two particles, we consider a system composed of a large number of particles, we find that the aligning interaction (Eq. (17.29)) is the cause of a rich and complex collective behavior, and allows the formation of metastable clusters whose size depend on the strength of the interaction with respect to the noise (Fig. 17.11).

Example 17.25: Aligning interaction, many particles. Generalize the code of Example 17.24 to run with $N = 100$ particles (or more). Describe the motion of each particle with a scheme similar to the active Brownian particle of Eq. (17.13) but with a finite radius R and without any translational noise term. Consider noise only on the orientation of the particle, according to the equation below:

$$\begin{cases} x_n(t+1) &= x_n(t+1) + v_0 \cos \theta_n(t) \\ y_n(t+1) &= y_n(t+1) + v_0 \sin \theta_n(t) \\ \theta_n(t+1) &= \theta_n(t) + T_n + \xi_n \end{cases}$$

where n is the particle index, v_0 its propulsion speed, T_n the total torque on particle n due to the interaction with the surrounding particles given by Eq. (17.29), and ξ_n is a white noise term, uniformly distributed on the interval $[-\frac{\eta}{2}, \frac{\eta}{2}]$ of amplitude η . Consider the values $v_0 = 0.05$, $R = 1$, $T_0 = 1$. The time step is 1. Set the particle positions and orientations initially at random, resolve the possible overlapping according to the excluded volume procedure.

a. Start with a noise characterised by $\eta = 2\pi$, i.e., a very high value for the noise. Show that, in this case, the time evolution of the system resembles Fig 17.11a, and no stable cluster is visible at any time.

b. Reduce the noise level to $\eta = 0.2\pi$ and show that the steady state (i.e., after an enough long time) of the system resembles Fig 17.11b: clusters are formed and many of them are static or stable, like, for example, most configurations with three particles arranged according to an equilateral triangle. There are also configurations where particles are arranged in a line and keep moving altogether in line. Clusters of sizes larger than three are present, but statistically clusters of smaller dimensions are more represented in number.

c. Reduce the noise level to $\eta = 0.002\pi$, i.e., a very low amount of noise. Show that the steady state (i.e., after an enough long time) of the system resembles Fig 17.11c: now, more exemplars of clusters of sizes larger than three are formed.

For a reference, see [34].

Metastable channels

A natural further step is the addition of passive particles in the system. It is possible to define an active–passive aligning interaction such that, in the case of active particles swimming in an environment of densely packed passive particles, channels are formed by the active agents [34].

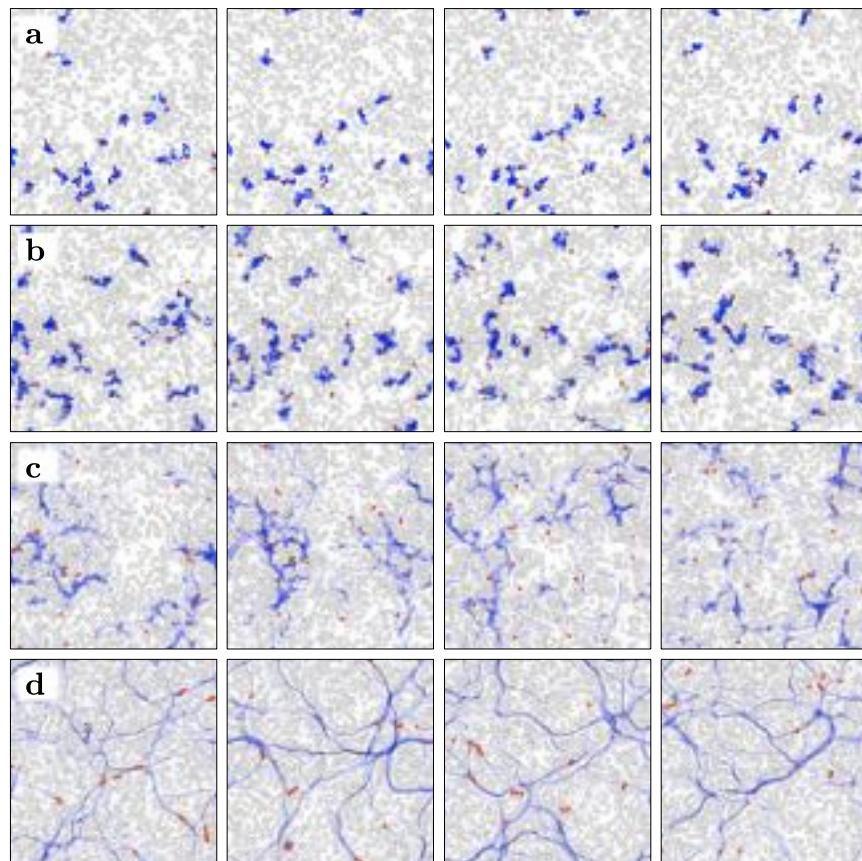


Figure 17.12: Metastable channel formation. The figure is obtained by simulating a system with 20 active particles and 900 passive particles. The noise level is set equal to: (a) $\eta = 2\pi$, (b) $\eta = \pi$, (c) $\eta = 0.5\pi$, and (d) $\eta = 0.03\pi$. From left to right, the plots correspond to time steps $t = 25000, 50000, 75000, 100000$. The blue shades represent the trail left by the active particles in the last 25000 steps before the represented configuration. A low enough noise level allows the active particles to open and stabilize metastable channels in the landscape of passive particles. See also [34]. Figure adapted from Ref. [34].

Let us consider a system with N_a active Brownian particles interacting with the aligning torque in Eq. (17.29), and with N_p passive Brownian particles that mutually interact only via the hard-sphere interaction. The active particles interact with the passive particles by mean of an aligning torque similar to the one of Eq. (17.29), but with opposite sign. This expresses the fact that, when an active particle interacts with a passive one, it tends to turn away from it. The passive particle instead does not feel any torque. Let us suppose that $N_p \gg N_a$, i.e., there are many more passive particles than active ones. We want to show that there are choices of the parameters describing the system (density of active and passive particles, strength of the aligning interaction, speed of the active particles, level of the thermal noise affecting the active particle orientation) such that the active particles can dig metastable channels in the configuration of the passive particles. [34].

Example 17.26: Metastable channels. Generalize the code of Example 17.25 to run with N_a active Brownian particles and N_p passive Brownian particles. Choose a small number of active particles and a number of passive particle that almost fills the available space. Consider periodic boundary conditions.

- a. Write a function that implements the aligning interaction between active and passive particles. In the following, consider that the magnitude of the reference torque in the interaction between active and passive particles is exactly equal to the torque acting between active particles. [*They are opposite in sign though.*]
- b. Start with a large value for the noise: $\eta = 2\pi$. Show that the behavior of the system resembles the one given in the sequence of images in Fig. 17.12a. The active particles are confined by the passive particles in a region of size comparable to few particle diameters.
- c. Lower the level of the noise to $\eta = \pi$. The active particles are still confined but on a more extended and elongated area. Compare with Fig. 17.12b.
- d. Lower the level of the noise to $\eta = 0.5\pi$. The active particles start *drilling* their path in the landscape of passive particles, and channels are formed, though not long-lived. Compare with the behavior represented in Fig. 17.12c.
- e. Lower further the noise level to $\eta = 0.03\pi$. In this case, channels are more long-lived and stabilized by other active particles following the trail of the active particle opening the channel. Compare with Fig. 17.12d.

For a reference, see Ref. [34].

17.6 Multiplicative noise

In the previous sections, we considered the case of particles moving in a homogeneous and isotropic space, e.g., like in the bulk of a fluid. However, in many cases the Brownian particles are close to a boundary, e.g., they are close to a solid–liquid or to a liquid–air interface, or near a boundary or an obstacle limiting their free diffusion. In such cases, the thermal noise acting on the Brownian particle depends explicitly on their position, and we are in the case of *multiplicative noise* acting on the particle.

17.6.1 Spurious drift

When multiplicative noise is at play, special care should be taken when writing the Langevin equation for the particle and simulating the particle trajectory in the finite-difference formalism. For example, let us consider a colloid diffusing above (and very close to) a microscope slide. The colloid can move in three dimensions: two directions are parallel to the near planar surface, the third is perpendicular to it. If the colloid were suspended in the bulk of the solvent, very far away from the slide, the three directions would be equivalent, and the diffusion coefficient entering the equation of Brownian motion would be the same for each of the three directions. When close to a planar solid–liquid interface, instead, the particle diffuses differently in the direction perpendicular to the interface than in the directions along the interface. Focusing specifically on the perpendicular direction, that we call here x -direction, we have that the diffusion coefficient $D_{\perp}(x)$ depends explicitly on the distance of the particle from the boundary x [35, 36].

In general, a multiplicative noise is described as:

$$\xi = \sqrt{2 D(x)} \xi_{0,1} , \quad (17.30)$$

where the diffusion constant depends on the value of the variable describing the status of the system. In such a case, the corresponding equation is [37]:²

$$\dot{x} = \frac{D(x)}{k_B T} F_{\text{ext}} + \underbrace{\frac{dD(x)}{dx}}_{\text{spurious drift}} + \xi(x), \quad (17.31)$$

where the additional term $\frac{dD}{dx}$ is the *spurious drift* and is necessary for the correct convergence of the solution to the original Langevin's equation with multiplicative noise [35, 38]. Finally, the corresponding finite-difference equation describing a particle in a gradient of diffusion is:

$$x_{n+1} = x_n + \left(\frac{D(x_n)}{k_B T} F_{\text{ext}} + \frac{dD(x_n)}{dx} \right) \Delta t + \sqrt{2D(x_n) \Delta t} \xi_n. \quad (17.32)$$

[Shouldn't the derivative in the previous expression be discretized too?] It is important to include the spurious drift term in both the Langevin equation and in the finite-difference equation in order to describe the system correctly.

17.6.2 At thermodynamic equilibrium

We clarify here the meaning of *describing the system correctly* and the role of the spurious drift starting with a system at thermodynamic equilibrium.

The force $F(x)$ acting on a microscopic particle suspended in a liquid can be assessed either by their underlying potential $U(x)$ or by the effect on the particle trajectory. In the former approach, which can be used at thermodynamic equilibrium only, one samples the equilibrium distribution $\rho(x) \propto e^{\frac{U(x)}{k_B T}}$ and extracts the force by taking the gradient of the potential:

$$F(x) = -\frac{dU(x)}{dx} = \frac{k_B T}{\rho(x)} \frac{d\rho(x)}{dx} \quad (17.33)$$

In the latter approach, which can be used both in the case of equilibrium and non-equilibrium, one calculates the drift $C(x)$ and the diffusion $D(x)$ from the trajectories (see Box 17.2), and obtains the force through:

$$F(x) = \gamma(x) C(x) - \gamma(x) \frac{dD(x)}{dx} \quad (17.34)$$

where $\gamma(x)$ is the friction coefficient and $\gamma(x) \frac{dD(x)}{dx}$ is the spurious force.

Box 17.2: Given a trajectory x_j with positions measured at equispaced time intervals $t_j = j\Delta t$, a partition of the space into bins centered at L_k with width ΔL , and a reference observation number of time steps n (corresponding to the observation time $\tau = n\Delta t$), we define the following quantities:

$$\text{drift:} \quad C(L_k, \Delta L, n) = \frac{1}{n\Delta t} \langle x_{j+n} - x_j \rangle, \quad (17.35)$$

$$\text{biased diffusion:} \quad D_b(L_k, \Delta L, n) = \frac{1}{2n\Delta t} \langle (x_{j+n} - x_j)^2 \rangle, \quad (17.36)$$

$$\text{unbiased diffusion:} \quad D_u(L_k, \Delta L, n) = \frac{1}{2n\Delta t} \langle (x_{j+n} - x_j - C(L_k) \cdot n\Delta t)^2 \rangle, \quad (17.37)$$

where the average is taken over all x_j belonging to the interval $[L_k - \frac{\Delta L}{2}, L_k + \frac{\Delta L}{2}]$. It is possible to use different values of n for Eqs. (17.35), (17.36), and (17.37). In particular, it might be convenient to use a very small n for the measurement of the diffusion, and a larger n for the calculation of the drift. In all

²The equation here is written in the Ito form. Alternative forms are possible, see details in [37].

cases, for an accurate estimate of the drift and the diffusion from a sequence of experimental data, a high sampling rate is to be preferred.

In [35], there is a systematic study of the spurious drift correction, necessary when comparing the forces acting on various Brownian particles derived from equilibrium-distribution and drift measurements. Here, we propose a simplified exercise, related to Ref. [35], to bring a concrete example of the necessity of considering the spurious drift when simulating a system in a gradient of diffusion and when extracting physical quantities using the trajectories of particles that are in a diffusion gradient (i.e., by calculating the *drift*, see Box 17.2), and not their equilibrium distribution.

Example 17.27: A minimal model for the spurious drift. Consider a particle moving in one dimension in a harmonic potential $U(x) = \frac{1}{2}kx^2$. The particle is moving in a gradient of diffusion characterised by $\sigma(x) = \sigma_0 (\arctan x + \frac{\pi}{2})$, where $\sigma = \sqrt{2D}$.

a. Simulate the dynamics of a Brownian particle according to Eq. (17.31) in the finite-difference formalism. Simulate trajectories first by omitting the spurious drift, then by accounting for the spurious drift, and generate the equilibrium distribution for both cases by taking the ensemble average of the trajectories. Show that the correct equilibrium distribution $P(x) \propto e^{-\frac{1}{2}kx^2}$ is reproduced by the trajectories where the spurious drift is taken into account.

b. Using the trajectories generated with the spurious drift, calculate the drift $C(x)$ and the diffusion $D(x)$. Verify that Eq. (17.34) holds.

[See also Problem 17.6.]

17.6.3 Out of equilibrium

At equilibrium the friction coefficient γ and the diffusion coefficient D satisfy the Einstein's relation $\gamma D = k_B T$. This is valid also in the case that the diffusion coefficient D depends on the position: $\gamma(x)D(x) = k_B T$.

In non-equilibrium systems this relation does not hold any longer, and $\gamma(x)$ and $D(x)$ vary independently. In general, calling $\sigma(x) = \sqrt{2D(x)}$, the Langevin equation for a particle with mass $m \neq 0$ is:

$$m\ddot{x} = F(x) - \gamma(x)\dot{x} + \sigma(x)\xi. \quad (17.38)$$

In the limit of zero mass, the effective equation becomes:

$$dx = \left[\frac{F(x)}{\gamma(x)} - \frac{\sigma(x)^2}{2\gamma(x)^3} \frac{d\gamma(x)}{dx} \right] dt + \frac{\sigma(x)}{\gamma(x)} d\xi. \quad (17.39)$$

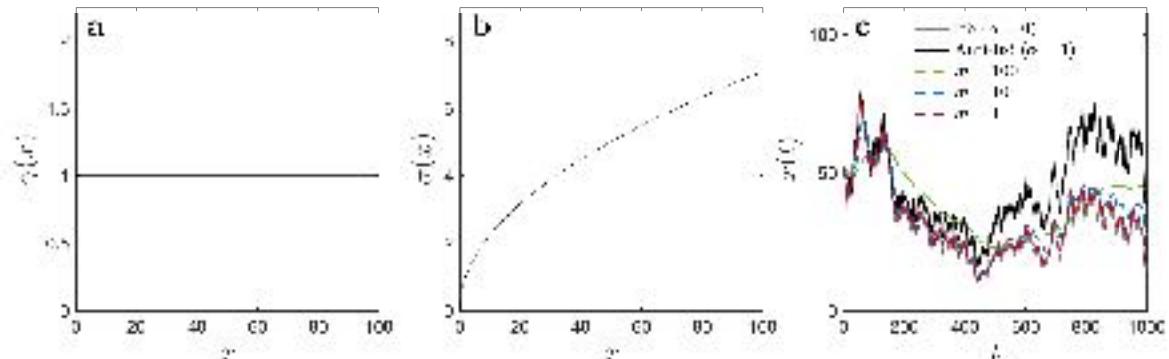


Figure 17.13: **Brownian particle in a non-equilibrium environment.** (a) Dissipation $\gamma(x)$ and (b) fluctuation $\sigma(x)$ for a Brownian particle. (c) Trajectories obtained from Eq. (17.38) for masses $m = m_0$, $m = 0.1 m_0$, $m = 0.01 m_0$ and from Eq. (17.39). The last is the limit for $m \rightarrow 0$ of the trajectories obtained implementing Eq. (17.38). Figure adapted from Ref. [37].

Example 17.28: Brownian particle in a non-equilibrium environment. Consider a Brownian particle in one dimension subject to a constant dissipation $\gamma(x) = 1$ and to a position-dependent fluctuation $\sigma(x) = x^2$.

a. Simulate its dynamics according to Eq. (17.38) for $m = 10^{-n}m_0$ with $n = 0, 1, 2$. Use the same realization of white noise for the different cases.

b. Simulate the dynamics of a particle with zero mass, according to Eq. (17.39). Use the same white noise realization used in point a.

c. Compare the trajectories obtained with mass and without mass. Verify graphically that the trajectory generated with Eq. (17.39) corresponds to the limit $m \rightarrow 0$

See Fig. 17.13.

Problems

Problem 17.1: Rigid colloidal molecule. Consider a colloidal molecule that consists of two identical Brownian particles ($R = 0.2 \mu\text{m}$) that are connected with a nanorod of length $L = 1 \mu\text{m}$ that is negligible in thickness. Assume that the connecting rod imposes the condition on the particles that the distance between them has to stay constant.

a. Simulate this two-particle system in 2 dimensions under free diffusion with this constraint. This can be achieved simply by letting both particles diffuse freely in each step and push or pull the particles back to the same distance from each other at the end of the step. Make an animation of the motion of the two particles and observe their motion.

b. Calculate the orientation of the rod φ in each step and plot this as a function of time **with removed jumps of 2π [What does it mean? Every 2π ?]**.

c. Compute the mean square angular displacement ($\langle \varphi^2(t) \rangle$) as a function of time for different values of L and observe that the rotational diffusion of this colloidal molecule is inversely proportional to L^3 .

Problem 17.2: Bistable potential. Consider a colloidal particle that is held in a double well potential: $U(x) = -ax^3 + bx$. Such a particle will have two stable equilibrium locations that are placed symmetrically around the origin L that are separated by an energy barrier E_b .

a. Calculate a and b analytically in terms of L and E_b .

b. Simulate the trajectory of the particle for fixed parameters. Plot the trajectory and observe that the particle is following a bistable motion. Verify that the **particle's positions obey?** the Boltzmann distribution.

c. Show that the frequency of the jumps on the potential barrier decays exponentially with the increasing barrier height, i.e., $f \propto \exp\left(-\frac{E_b}{k_B T}\right)$ where f stands for the average frequency of the jumps.

Problem 17.3: Brownian particle in an oscillating trap. Consider a Brownian particle that is subject to an optical trapping potential where the center of the optical trapping potential oscillates with a frequency f and an amplitude x_c :

$$F(x(t), t) = -k[x(t) - x_c \sin(2\pi ft)]$$

a. Simulate the motion of this particle under this condition and explore what happens with different values of x_c and f .

b. Show that, if the values of f and k/γ are comparable, the variance of the particle position does not decrease monotonically as k increases, meaning that there is an optimum value of k for the minimum variance of the particle position [39].

Problem 17.4: Stochastic resonance. Consider a Brownian particle that is in a double well potential where the potential is tilted periodically on either direction:

$$F(x(t), t) = -ax^3 + bx + c \sin(2\pi ft) \quad (17.40)$$

Under such conditions stochastic resonance occurs [40]. If f is comparable to the particle's average jump frequency between the wells, there can be a partial synchronization of the jumps with the oscillating force. This synchronization strongly depends on the temperature of the system.

a. Simulate a Brownian particle that is placed in the potential given above. Observe what happens with different parameters and plot trajectories.

b. Compute the correlation of the particle location with the potential oscillations $C = \langle x(t) \sin(2\pi ft) \rangle$ and plot it as a function of temperature.

Problem 17.5: Brownian particle trapped in air. Consider a Brownian particle in an optical trap that is immersed in air where the friction coefficient γ is much lower than the one of water. In this case, we need to simulate the motion of the Brownian particle in the optical trap without ignoring its mass.

a. Derive the difference equation of the solution for x_i for a Brownian particle that is optically trapped without ignoring the mass.

b. Show that the autocorrelation function of the particle position C_x shows oscillatory behavior for low values of γ and transitions into an exponential decay as in Fig. 3.4 (c) when the γ values are high.

Problem 17.6: Brownian colloid suspended above a flat substrate. A Brownian spherical particle is suspended above a substrate: the only deterministic forces acting on the particle are its weight F_{weight} and the electrostatic colloidal repulsion with the substrate F_{ES} , where $F_{\text{ES}}(z) = -\frac{\partial U}{\partial z}$ with

$$U(z) = k_B T e^{-\frac{z-h_0}{\lambda_D}}$$

where h_0 is a parameter depending on the surface charges and λ_D is the Debye length in the fluid. The system is at thermodynamic equilibrium. We are monitoring the motion of the particle in the direction perpendicular to the substrate (z). The equilibrium distribution of the particle in z at equilibrium is given by $P(z) \propto e^{-\frac{U(z)}{k_B T}}$.

a. Simulate the vertical trajectory of a Brownian particle (radius R) subject to weight F_{weight} and electrostatic repulsion $F_{\text{ES}}(z)$ accounting for the spurious drift, according to Eq. (17.32). For the diffusion coefficient along z , use the expression $D_{\perp}(z) = \frac{D_{\text{SE}}}{\xi_{\perp}(z)}$ [36, 35], where

$$\xi_{\perp}(z) = \frac{4}{3} \sinh[a(z)] \sum_{n=1}^{\infty} \frac{n(n+1)}{(2n-1)(2n+3)} \left(\frac{2 \sinh[(2n+1)a(z)] + (2n+1) \sinh[2a(z)]}{4 \sinh^2[(n+0.5)a(z)] - (2n+1)^2 \sinh^2[a(z)]} - 1 \right) \quad (17.41)$$

$a(z) = \cosh^{-1}\left(1 + \frac{z}{R}\right)$, $D_{\text{SE}} = \frac{k_B T}{6\pi\eta R}$ is the Stokes-Einstein bulk diffusion coefficient, and η is the viscosity of the liquid. By simulating long trajectories, obtain numerically the equilibrium distribution, and compare it with the expected theoretical distribution.

b. Repeat the previous simulation without accounting for the spurious drift correction (use the same values for all the other parameters, including the random sequence for the thermal noise). For particles of small radius you will notice a discrepancy in the equilibrium distribution obtained numerically.

c. On a short time scale, plot the two trajectories (with and without the spurious drift, starting from the same point, with the same noise contribution and all the other parameters) for particles of radius $R = 1 \mu\text{m}$ and $R = 0.1 \mu\text{m}$. Show that a bigger discrepancy between the trajectory with and without spurious term is observed for the smaller particle.

d. Use the trajectories $z(t)$ obtained in the previous steps to calculate numerically the *drift* and the *diffusion* experienced by the particle for a given point z_0 (see definition in Box 17.2).

In the case of constant diffusion coefficient, the external force $F_{\text{ex}} = -\nabla U_{\text{ext}}$ in a point z_0 are revealed by calculating the drift from a particle trajectory $z(t)$. If the case of multiplicative noise (i.e., diffusion coefficient varying with the position), taking the drift of the *experimental* particle trajectories in z_0 will not give the information about the force, but also about the noise induced drift. [This sentence is not clear. Please, rewrite it]

As a reference, see [35].

Problem 17.7: Free-falling, overdamped active Brownian particle and sedimentation. Consider a single free-falling overdamped ABP subjected to a gravitational field $\mathbf{F}_g = -\gamma v_s \hat{e}_z$ in the z direction, where γ is the Stokes drag coefficient and v_s is the sedimentation speed.

a. Write the Langevin equations for such an ABP.

b. Write the Fokker-Planck equation for the PDF of the ABP and use separation of variables $(P(z, \theta) = Z(z)\theta(\theta))$ to show that, in the steady state, the sedimentation profile $Z(z)$ decays exponentially with z .

Problem 17.8: Active Brownian particle. Consider the Langevin equation for the orientation of a single ABP [cf. Eq. (15.2)] with initial orientation θ_0 .

- a. Show that $\langle \theta(t) \rangle = \theta_0$ and $\langle (\theta(t) - \langle \theta(t) \rangle) \rangle = 2D_R t$.
- b. Show that the PDF of the orientation at time t is given by $P(\theta, t) = \frac{1}{\sqrt{4\pi D_R t}} \exp -\frac{(\theta-\theta_0)^2}{4D_R t}$. Do not assume periodicity of the orientation.
- c. Show that $\langle \cos \theta(t) \rangle = e^{-D_R t} \cos \theta_0$ and $\langle \sin \theta(t) \rangle = e^{-D_R t} \sin \theta_0$, as was used in Example 1.
- d. Show that the MSD of a single ABP is given by Eq. (15.4).

Problem 17.9: Run-and-tumble particle. Consider the dimensionless equations of motion for a single RTP (see Tab. 15.2).

- a. Perform simulations of these equations assuming that tumbling events are equally spaced, i.e., $T_{n+1} - T_n = 1$ for all n . Plot the resulting trajectories and compare them with the exemplary trajectory of an RTP shown in Fig. 1 (e).
- b. Consider an ensemble of non-interacting RTPs and include a gravitational field acting on the center of mass of the particles. Is the resulting sedimentation profile also exponential, as shown for ABPs in Problem 17.7?
- c. Create a histogram with the steady-state angular distribution of the RTPs of part b). Does polar order arise? Compare your results with those of Ref. [41].

Problem 17.10: Chiral active particle. Consider the equations of motion for a single CAP (Eqs. (15.1) and (15.11)) with $U=0$.

- a. By defining the time and length scales of the system as in Tab. 15.2, rewrite these equations in dimensionless form.
- b. Carry out Brownian dynamics simulations of the obtained equations of motion and plot the resulting trajectories, i.e., $r_y(t)$ vs. $r_x(t)$, for increasing values of the Péclet number Pe and the dimensionless angular velocity ω .
- c. Compute the MSD numerically and compare it with Refs. [42, 43].

Problem 17.11: Smooth variant of the Vicsek model. Consider an ensemble of N active particles obeying the smooth variant of the Vicsek model as given by Eq. (15.1) with $U = 0$ and Eq. (15.12). Carry out simulations of the ensemble in a box of side L with periodic boundary conditions and produce a video showing the behavior of the system. Explore the effect of varying v_0 while keeping K constant and vice versa. Use the following parameter values: $D = 0$, $D_R = 0.5$, $\omega = 0$, $R_0 = 1$, $N/L^2 = 20$, time step $\Delta t = 10^{-2}$, $v_0 \in [0.1, 1.5]$ and $K \in [0, \pi/5]$.

Problem 17.12: Ensemble of active Brownian particles and MIPS. Perform particle-based simulations of an ensemble of $N = 4000$ overdamped ABPs [cf. Eqs. (15.1) and (15.2)] with short-ranged repulsive interactions (e.g., WCA potential [44]). Find a parameter regime for which MIPS occurs and analyze the time dependence of the mean cluster size. Would you expect this to change when considering inertial ABPs?

Problem 17.13: Active model B+. Consider the active model B+ given by Eqs. (15.13) and (15.14) for $-a = b = 0.25$, $M = 1$, $K = 1$, $\xi = 4$, $\lambda = 1$, and $D = 0.2$. Solve the active model B+ numerically in two dimensions in a square box with periodic boundary conditions with time step $\Delta t = 0.001$ and a grid spacing of $\Delta x = \Delta y = 1$ at global density $\phi_0 = -0.4$.

Problem 17.14: Flow field by a force dipole. [isn't this problem more appropriate for a different chapter?] Calculate the flow field created by a force dipole. You can find useful information in Refs. [45, 46].

Problem 17.15: Inertial active Brownian particle. Consider an inertial ABP of mass m and moment of inertia I .

- a. Write the corresponding Langevin equations and calculate the mean square displacement. Which regimes can be identified?
- b. Derive the Langevin equations in dimensionless form. How many parameters determine the dynamics of the inertial ABP?

c. Perform simulations of the dimensionless Langevin equations and determine the mean square displacement of the particle. Compare your simulation results with the analytical expression found in part **a**.

Problem 17.16: Fokker-Planck equation for non-interacting inertial active Brownian particles. Write the Fokker-Planck equation for non-interacting inertial ABPs with mass m and moment of inertia I in two dimensions.

a. Derive the steady-state PDF $P(\theta, \mathbf{v})$ and determine $P(\mathbf{v})$ by numerical integration, where θ denotes the orientation angle of a particle and \mathbf{v} denotes its velocity. Compare your result with Ref. [47].

b. From $P(\mathbf{v})$, calculate the mean square speed $\langle \mathbf{v}^2 \rangle$ and compare your results with particle-based simulations.

Bibliography

- [1] R. Brown and J. J. Bennett. *Microscopical Observations*, pages 463–486. Published for the Ray society by R. Hardwicke, London, 1866.
- [2] P. Pearle, B. Collett, K. Bart, D. Bilderback, D. Newman, and S. Samuels. What Brown saw and you can too. *American Journal of Physics*, 78:1278–1289, 2010.
- [3] J. Perrin. Mouvement brownien et molécules. *Journal de Physique Théorique et appliquée*, 9:5, 1910.
- [4] A. Einstein. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der Physik*, 17:208, 1905.
- [5] E. Nelson. *Dynamical theories of Brownian motion*, volume 106. Princeton University Press, 2020.
- [6] J. M. T. Thompson and H. B. Stewart. *Nonlinear dynamics and chaos*. John Wiley & Sons, 2002.
- [7] E. M. Purcell. Life at low Reynolds number. *American Journal of Physics*, 45:3, 1977.
- [8] T. Li, S. Kheifets, D. Medellin, and M. G. Raizen. Measurement of the instantaneous velocity of a Brownian particle. *Science*, 328:1673, 2010.
- [9] R. Huang, I. Chavez, K. M. Taute, B. Lukić, S. Jeney, M. G. Raizen, and E.-L. Florin. Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid. *Nature Physics*, 7:576, 2011.
- [10] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [11] J. Yuan, K. A. Fahrner, L. Turner, and H. C. Berg. Asymmetry in the clockwise and counterclockwise rotation of the bacterial flagellar motor. *Proceedings of the National Academy of Sciences*, 107:12846, 2010.
- [12] L. Lemelle, J.-F. Palierne, E. Chatre, and C. Place. Counterclockwise circular motion of bacteria swimming at the air-liquid interface. *Journal of Bacteriology*, 192:6307, 2010.
- [13] S. van Teeffelen and H. Löwen. Dynamics of a Brownian circle swimmer. *Physical Review E*, 78:020101, 2008.
- [14] B. ten Hagen, S. van Teeffelen, and H. Löwen. Brownian motion of a self-propelled particle. *Journal of Physics: Condensed Matter*, 23:194119, 2011.
- [15] R. Di Leonardo, D. Dell’Arciprete, L. Angelani, and V. Iebba. Swimming with an image. *Physical Review Letters*, 106:038101, 2011.
- [16] F. Kümmel, B.T. Hagen, R. Wittkowski, I. Buttinoni, R. Eichhorn, G. Volpe, H. Löwen, and C. Bechinger. Circular motion of asymmetric self-propelling particles. *Physical Review Letters*, 110:198302, 2013.
- [17] M. Mijalkov and G. Volpe. Sorting of chiral microswimmers. *Soft Matter*, 9:6376, 2013.

- [18] H. C. Berg. *E. coli in Motion*. Biological and Medical Physics, Biomedical Engineering. Springer, New York, NY, 2004.
- [19] H. Brenner. Coupling between the translational and rotational Brownian motion of rigid particles of arbitrary shape. *Journal of Colloid and Interface Science*, 23:407, 1967.
- [20] M. X. Fernandes and J. G. de la Torre. Brownian dynamics simulation of rigid particles of arbitrary shape in external fields. *Biophysical Journal*, 83:3039, 2002.
- [21] A. Callegari and G. Volpe. Numerical simulations of active Brownian particles. In F. Toschi and M. Sega, editors, *Flowing Matter, Soft and Biological Matter*, page 211. Springer International Publishing, Cham, 2019.
- [22] F. Perrin. Mouvement brownien d'un ellipsoïde (II). Rotation libre et dépolarisation des fluorescences. translation et diffusion de molécules ellipsoidales. *Journal de Physique et le Radium*, 7:1, 1936.
- [23] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339:936–940, 2013.
- [24] A. Argun, A.-R. Moradi, E. Pinçé, G. B. Bagci, A. Imparato, and G. Volpe. Non-Boltzmann stationary distributions and nonequilibrium relations in active baths. *Physical Review E*, 94:062150, 2016.
- [25] F. Schmidt, A. Magazzù, A. Callegari, L. Biancofiore, F. Cichos, and G. Volpe. Microscopic engine powered by critical demixing. *Physical Review Letters*, 120:068004, 2018.
- [26] F. A. Lavergne, H. Wendehenne, T. Bäuerle, and C. Bechinger. Group formation and cohesion of active particles with visual perception-dependent motility. *Science*, 364:70–74, 2019.
- [27] F. Schmidt, B. Liebchen, H. Löwen, and G. Volpe. Light-controlled assembly of active colloidal molecules. *Journal of Chemical Physics*, 150:094905, 2019.
- [28] G. Volpe, G. Volpe, and S. Gigan. Brownian motion in a speckle light field: Tunable anomalous diffusion and selective optical manipulation. *Scientific Reports*, 4:3936, 2014.
- [29] P. Galajda, J. Keymer, P. Chaikin, and R. Austin. A wall of funnels concentrates swimming bacteria. *Journal of Bacteriology*, 189:8704–8707, 2007.
- [30] G. Vizsnyiczai, G. Frangipane, C. Maggi, F. Saglimbeni, S. Bianchi, and R. Di Leonardo. Light controlled 3D micromotors powered by bacteria. *Nature Communications*, 150:15974, 2017.
- [31] G. Volpe, I. Buttinoni, D. Vogt, H.-J. Kümmerer, and C. Bechinger. Microswimmers in patterned environments. *Soft Matter*, 7:8810–8815, 2011.
- [32] G. Volpe, S. Gigan, and G. Volpe. Simulation of the active Brownian motion of a microswimmer. *American Journal of Physics*, 82:659, 2014.
- [33] W. Schaertl and H. Sillescu. Brownian dynamics of polydisperse colloidal hard spheres: Equilibrium structures and random close packings. *Journal of Statistical Physics*, 77:1007–1025, 1994.
- [34] S. Nilsson and G. Volpe. Metastable clusters and channels formed by active particles with aligning interactions. *New Journal of Physics*, 19:115008, 2017.
- [35] T. Brettschneider, G. Volpe, L. Helden, J. Wehr, and C. Bechinger. Force measurement in the presence of Brownian noise: Equilibrium-distribution method versus drift method. *Physical Review E*, 83:041113, 2011.
- [36] H. Brenner. The slow motion of a sphere through a viscous fluid towards a plane surface. *Chemical Engineering Science*, 16:242, 1961.
- [37] G. Volpe and J. Wehr. Effective drifts in dynamical systems with multiplicative noise: a review of recent progress. *Reports on Progress in Physics*, 79:053901, 2016.

- [38] S. Hottovy, A. McDaniel, G. Volpe, and J. Wehr. The Smoluchowski-Kramers limit of stochastic differential equations with arbitrary state-dependent friction. *Communications in Mathematical Physics*, 336:1259, 2015.
- [39] G. Volpe, S. Perrone, J. M. Rubi, and D. Petrov. Stochastic resonant damping in a noisy monostable system: Theory and experiment. *Physical Review E*, 77:051107, 2008.
- [40] L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Reviews of Modern Physics*, 70:223, 1998.
- [41] A. P. Solon, M. E. Cates, and J. Tailleur. Active Brownian particles and run-and-tumble particles: A comparative study. *European Physical Journal Special Topics*, 224:1231–1262, 2015.
- [42] S. van Teeffelen and H. Löwen. Dynamics of a Brownian circle swimmer. *Physical Review E*, 78:020101, 2008.
- [43] H. Löwen. Chirality in microswimmer motion: From circle swimmers to active turbulence. *European Physical Journal Special Topics*, 225:2319–2331, 2016.
- [44] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *Journal of Chemical Physics*, 54:5237–5247, 1971.
- [45] M. D. Graham. *Microhydrodynamics, Brownian Motion, and Complex Fluids*. Cambridge University Press, Cambridge, 2018.
- [46] J. M. Yeomans, D. O. Pushkin, and H. Shum. An introduction to the hydrodynamics of swimming microorganisms. *European Physical Journal Special Topics*, 223:1771, 2014.
- [47] P. Herrera and M. Sandoval. Maxwell-Boltzmann velocity distribution for noninteracting active matter. *Physical Review E*, 103:012601, 2021.

Chapter 18

Accelerating Molecular Dynamics

JOOST DE GRAAF

In this chapter, we cover intermediate to advanced methods for *molecular dynamics* (MD) simulations aimed at working toward the hydrodynamic limit, see the cover picture (Fig. 18.1) for an illustration. At their core, MD simulations are a means to solve the dynamics in a many-body system described by the Hamilton formalism, as seen in Chapter 16.

The conservative nature of the particle collisions leads to hydrodynamic response at sufficiently large scale — the lecture notes *Kinetic Theory* by David Tong (University of Cambridge) are a very accessible introduction to this topic [1]. In brief, it should be understood that when there are only two particles in a system, it is not sensible to talk about a flow or even a fluid. Hydrodynamic flows and interactions emerge when very many particles interact, i.e., sufficiently many that it is possible to speak of statistical averages on a local scale. For a physicist, it is then natural to think of a *probability density description* of the underlying dynamics.

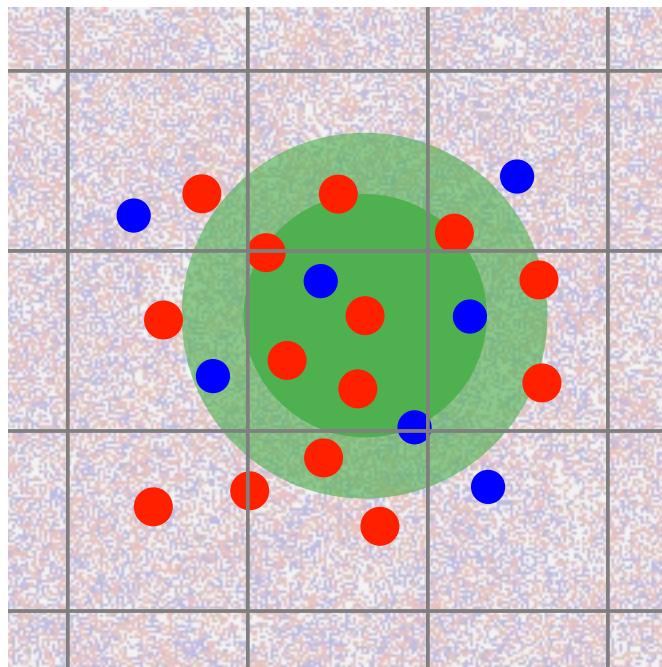


Figure 18.1: **Accelerating molecular dynamics using cell and Verlet lists.** The figure sketches the concepts of a Verlet list and cell list on top of a snapshot of a simulation containing 16384 particles. Such large systems would not be computationally tractable without the use of advanced simulation techniques.

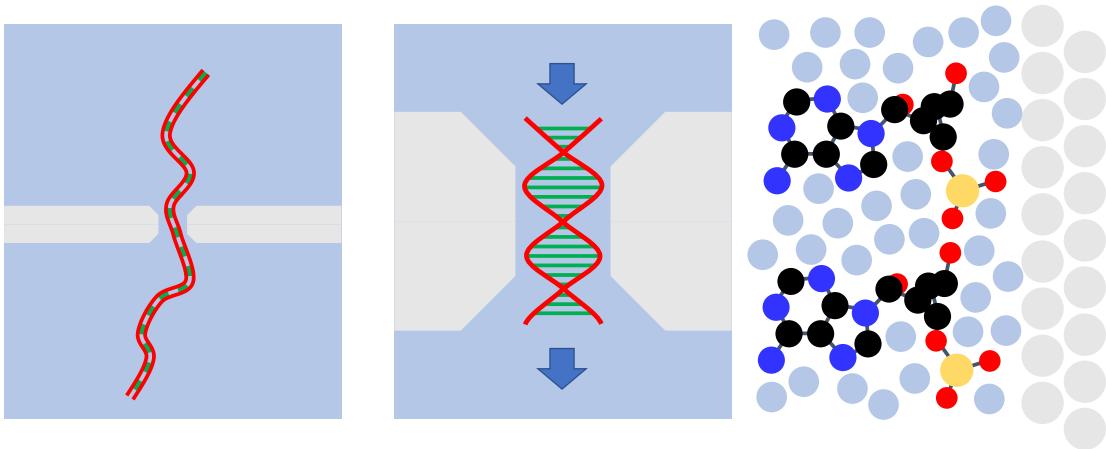


Figure 18.2: **Three scales involved in DNA translocation through a microfabricated nanopore.** The resolution increases from left to right. At the largest scale, provided in the left panel, the continuum strand is immersed in a continuum fluid. At the intermediate scale, provided in the central panel, some features of the DNA are resolved but not those of the molecular fluid. At the finest scale, provided in the right panel, the water molecules the DNA molecules, and also the atoms comprising the pore, are resolved — this illustration shows two halves of base pairs, as well as fluid molecules and atoms comprising the wall.

Boltzmann was the first to show that rewriting the dynamical equations in such a manner results in a transport equation for phase-space probability density, known as the *Boltzmann transport equation* (BTE), which governs the dynamics at intermediate scales. At even larger scales, the well-known *Navier-Stokes equations* for fluid dynamics emerge. However, there are situations where it is interesting to consider the emergence of flow from a molecular level, as we will discuss in Section 18.1. This then naturally leads back to simulating flow with particles. Of course, even if one is not interested in flow, it would still be desirable to perform MD simulations that involve tens to hundreds of thousands of molecules efficiently.

In the following sections, we will cover the concept of *neighbor list*, *Verlet list*, *cell list*, and *linked list*, which are necessary to efficiently simulate systems with many particles. We will cover the basics and give arguments as to why these methods can be useful to speed up a plain MD simulation, referring to standard textbooks for additional details. We will also discuss simulations involving non-differentiable interaction potentials and the associated technique of *event-driven MD* (EDMD). While this is a method that is not usually encountered in describing atomistic and molecular systems, EDMD can be a powerful technique to study simple model systems. EDMD can also be sped up using a clever form of book keeping, a *collision tree*, which we briefly comment upon and which can be combined with a cell list.

18.1 Relevant scales

Imagine a system of water molecules. It should be intuitively clear that when a sufficiently large amount of water is simulated, there can be flow on macroscopic length and time scales — provided this water is subjected to some suitable external force. Obviously, it is preposterous to simulate a bucket of water emptying out using molecular dynamics simulations that resolve the individual molecules. Some of the methods mentioned in Section 18.1.1 are more appropriate. Plain MD simulations, as introduced in Chapter 16, are typically suitable to resolve systems comprising several hundred molecules, i.e., systems with length and times scales of several nanometers and nanoseconds, respectively. Is this sufficient to see hydrodynamic effects? This section aims to address this question.

Let us first gain a feeling for scale and for convenience considering the example of transport through a nanopore. This could be the transport of biomolecules through pores in the cell membrane [2], or at a larger scale, the motion of a DNA strand through a microfabricated pore [3]; Fig. 18.2 shows an illustration of this process. Referencing the work of Garaj *et al.* [4], their smallest nanofabricated graphene pore has a diameter of

5 nm. A water molecule is about 3 Å in diameter, which implies that the pore itself is 15 water diameters wide. Thus, resolving a small cylinder of diameter 5 nm and length 5 nm around the pore mouth already involves the simulation of about 4,000 water molecules. For a biological example, we consider an α -hemolysin heptameric protein pore, which has been used by Oxford Nanopore in their sequencing technology [5]. This pore has a minimum inner diameter of 14 Å and a length of 100 Å, implying that it can contain about 600 water molecules.

The above numbers give some indication that only the smallest of structures can be resolved fully using MD simulations involving explicit simulation of water or other fluid molecules. A method such as *lattice Boltzmann* (Chapter 24) or *multi-particle collision dynamics* (Chapter 23) will be much more suited to resolve flow on the scale shown in the left panel of Fig. 18.2. However, we have yet to answer our original question. For this, we turn to another pore system, namely water flowing through carbon nanotubes. It was experimentally found that these carbon nanotubes have a curiously low resistance to the transport of water [6, 7]. This was simulated, amongst others, by Gravelle *et al.* [2], who reported excellent agreement between continuum hydrodynamics and even for pores with diameters around 3 nm. The implication being that for a number of water molecules on the order of 1,000 aspects of continuum flow may be recovered¹.

If continuum methods work at such small scales, what is the point of simulating flow with explicit particles? One answer is that because MD resolves individual molecules, it is possible to study aspects of flow that are only handled in an effective manner within meso- and macroscale descriptions. In hydrodynamic equations, one has to specify the boundary condition for the flow at a surface, with no-slip (zero fluid velocity) being a common choice. However, it is known that microscopic surface roughness and/or chemical functionalization can lead to an effective surface slip (think hydrophobic surfaces). Molecular dynamics offer opportunities to figure out how this comes about [8].

18.1.1 Other approaches for solving hydrodynamic equations

For completeness, we should point out a few other methods by which hydrodynamics may be solved here, before moving on. The *Boltzmann transport equation* (BTE) is a mesoscale description, because it combines a continuum description with aspects of pair interactions through a collision kernel. In simulations involving macromolecules, nanoparticles, and colloids, it is often too expensive to resolve all the particles in the fluid. When it proves necessary to account for hydrodynamic interactions to describe the physics of the system in this size range, mesoscale methods such as *lattice Boltzmann* (based on the BTE, e.g., see Ref. [9]) and *multi-particle collision dynamics* (MPCD, e.g., see Ref. [10]) simulations, which will be covered in Chapters 23 and 24, respectively, may be appropriate. For even larger scales, the collisional aspects of the molecules comprising the fluid can be further reduced to a few emergent quantities, such as viscosity and density, that fully describe the behavior of the fluid. This is the regime governed by *continuum Navier-Stokes equations*. Here, both LB and MPCD may also be used, as they recover the large-scale flow behavior described by these equations (but also resolve finer detail). However, other techniques that solve the Navier-Stokes equations directly, e.g., finite-volume, finite-element, boundary-element methods, or spectral methods, e.g., see Ref. [11], can be employed. Alternatively, the linear form of the Navier-Stokes equations, the Stokes equation — applicable to the colloidal regime — is amenable an approach based on *Green's functions*, which lead to numerical methods such as Stokesian dynamics, see the introduction of Ref. [12] for a detailed overview. All of these methods have seen success in describing wet active matter systems, for which there are hydrodynamic (and possibly phoretic) couplings.

18.2 Neighbor lists

When discussing MD, it is important to observe that a large part of the simulation revolves around computing the forces between particles. These are necessary in order to advance the particle positions and determine the evolution of the system. In plain MD simulations, the basic forces between particles act in a pairwise manner², which is what we will assume throughout. This suggests that a sensible way of organizing the particles in the algorithm is to loop over a list of pairs, a primitive form of *neighbor list*, rather than to have a nested loop

¹Basic statistical mechanics teaches us that, in equilibrium, the relative fluctuations about a (local) mean of n particles scale as $1/\sqrt{n}$. This gives an impression of the numbers of particles required to pin down averages to say 10% and speak of local quantities.

²Angular potentials are relatively common in MD simulations and these would involve three particles, or even four, if torques resulting from a dihedral or torsion-angle based potential are required. However, for the sake of understanding, we will ignore these aspects.

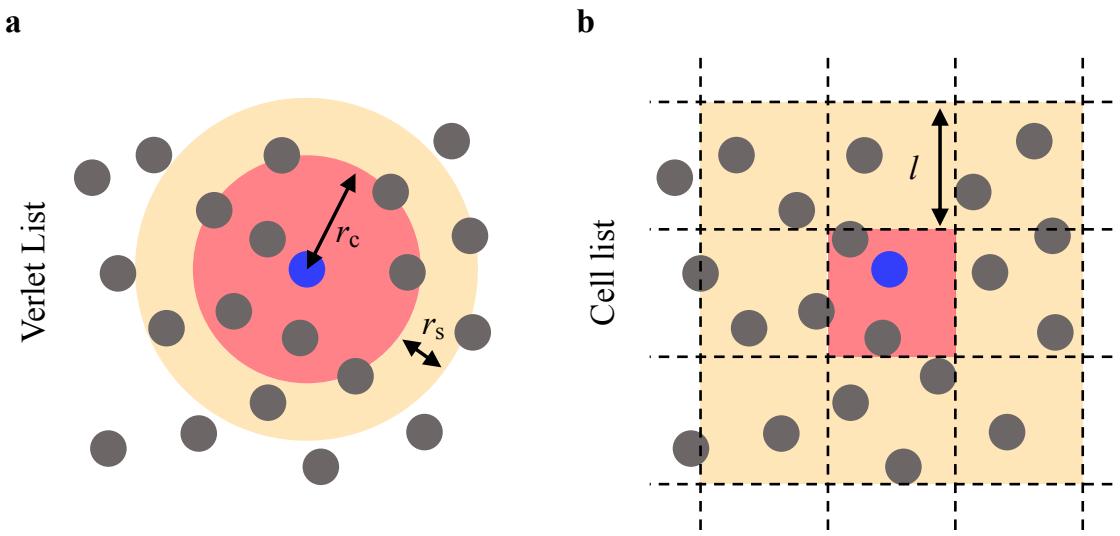


Figure 18.3: Two-dimensional illustration of the concept of a Verlet list and a cell list. The particles are indicated by gray disks. (a) In the *Verlet list*, the cut-off for the interaction range, r_c , is indicated by the red region, while the range of the shell, r_s , is given by the yellow region. The blue particle of interest has 11 neighbors. (b) In the *cell list*, the system is subdivided into squares with edge length l . Instead of checking all pairs in the system, only those in the central cell (red) and the neighboring cells (yellow) have to be checked. In this case, the particle of interest has 18 neighbors.

over particle coordinates. However, there does not seem to have an obvious advantage, as both algorithms must give identical results and involve the same number of calculations. In this section, we discuss how to go about making a list of neighbors such that there is substantial improvement in computational efficiency.

A naive implementation of the force calculation involves looping over all particle pairs and has an *algorithmic complexity* of $\mathcal{O}(N^2)$ with N the number of particles in the system. That is, the speed of the simulation will scale quadratically with N , because there are $N(N - 1)/2$ particle pairs to compute the force between. This is immediately problematic, if one is, for example, interested in removing finite-size effects from a system. That is, scaling out the artifacts that come from using periodic boundary conditions³. After all, doubling the system size – that is, the edge length of the simulation box – leads to a 64-fold increase in force calculations for a three-dimensional (3D) system at constant density when the algorithmic complexity is $\mathcal{O}(N^2)$.

A desirable complexity is $\mathcal{O}(N)$, where the computational cost scales linearly with the number of particles. However, even in this favorable case, doubling the system size as above will lead to an 8-fold increase in force calculations in 3D. This strongly limits the sizes of the systems that can be studied⁴. The multiplicative factor for two- and one-dimensional systems is more favorable, obviously. However, unfortunately (especially in two dimensions, 2D) the intrinsic range of the physical correlations can be much greater, which necessitates the simulation of systems with much larger values of N than the 3D equivalent to eliminate finite-size effects [13, 14]. A good rule of thumb in project design is to spend additional time thinking about what aspects of the system you want to simulate and how to best go about this physically, especially once the run time of a single simulation starts to exceed a single day, as you will typically need many simulations to gather statistics and there is a finite span to a project.

Returning to the matter of the algorithmic complexity of the force calculation, the question now becomes: How can we lower the exponent 2 of the scaling to 1 and speed up the algorithm as much as possible? We will see that these are actually two distinct questions. Clearly, it is desirable to reduce the number of pairs

³In performing finite size scaling, it is often convenient to graph your data as $1/L$ of the length scale L that you are varying. This allows you to more accurately ascertain the asymptotic behavior and extrapolate toward the limit of infinite system size $L \rightarrow \infty$.

⁴A quick back-of-the-envelope calculation reveals that a single-precision (floating-point number) MD algorithm that only accounts for positions and velocities will require at a minimum 6 floats to be stored in memory (24 bytes), meaning that memory is generally not a bottleneck for scaling up the system.

over which we loop as much as possible. Most molecular interactions are short ranged⁵. That is, for typical molecular interactions there is a finite scale r_c — usually referred to as the cut-off range — beyond which the interaction can be neglected, as shown in Fig. 18.3. It makes sense that the algorithm can be sped up considerably by only taking into account those particles which can interact with each other, rather than particles which are separated by a distance $r \gg r_c$. Provided that an appropriately reduced list of local neighbors is in place, there will be $m \geq 0$ neighbors around every particle, for which the distance $r < r_c$. For convenience, we can say that there are on average \bar{m} neighbors. Accounting for the dynamics then requires looping over particles and their local neighborhoods, rather than looping over all pairs in the system. This algorithm should be faster than the original, since $\bar{m}^2 \ll N$.

Unfortunately, there are two flaws in this reasoning. (i) Creating the list of neighbors takes time. If this is done too frequently, any gain from reducing the number of pair calculations in the dynamics is lost. The algorithm might even be slower than a naive implementation, certainly for small numbers of particles. (ii) Constructing a list of neighbors is still an $\mathcal{O}(N^2)$ operation, which means that the overall scaling of the simulation with a reduced neighbor list has not improved. To overcome these issues there are two additional steps that should be taken.

18.2.1 Verlet list

The first was proposed by Loup Verlet [16] and involves adding a *skin* to the interaction range, r_s , such that particles loose their property of being neighbors only when $r > r_c + r_s \equiv r_v$, as shown in Fig. 18.3a. This prevents the neighbor list from being constantly updated when particles come in and out of the interaction range r_c , but do not experience substantial local rearrangement. The skin size would have to be optimized for the specific MD parameters and interaction. Too large a skin and \bar{m}^2 becomes significant, leading to a slower simulation. Too small a skin and the neighbor list has to be updated too frequently, also slowing down the simulation. If you wish the resulting *Verlet list* to be updated every N_m simulation steps, then the skin size should roughly be $r_s \approx N_m \bar{v} \Delta t$, with \bar{v} the average particle speed and Δt the time step. Beyond this advice, there is no universally applicable approach to speeding up your simulation using a Verlet list. Depending on where in parameter space you are, larger or smaller skins may need to be used. Performing a few trial simulations at various points in parameter space before gathering statistics is advised to improve the time it will take you to complete your project.

The second requirement is to construct the Verlet list (based on the $r < r_v$ criterion) in an efficient manner. That is, the list of pairs should be constructed (infrequently) in such a way that it has low algorithmic complexity. This brings us to the concept of *cell lists*.

18.2.2 Cell list

The idea is to cut space into small cubes that have an edge length l greater (but not too much greater) than r_c , where l is a divisor of the simulation box length L , as shown in Fig. 18.3b. We first assign all particles to their respective cubes to determine whether a particle is neighboring another, which is an $\mathcal{O}(N)$ operation. Subsequently, we check for pairs, which only involves checking the cube the particle is in and its 26 neighbors (in 3D). This operation does not scale with N at constant volume. Reassignment to a new cube and recomputing the local neighbor list when updating the positions occurs on a per-particle basis throughout the remainder of the simulation. Therefore, the cell-list algorithm should have an overall $\mathcal{O}(N)$ scaling, see Fig. 18.4 and Example 18.1.

This discussion begs the question: What is the point of using a Verlet list in addition to a cell list, if the latter has the favorable scaling? The answer is that scaling only kicks in for large N , while aspects such as a prefactor can make a significant difference for intermediate values of N . Consider the example of a 3D system that consists of Lennard-Jones particles with diameter σ and cut-off radius $r_c = 2.5\sigma$. Further assume that we have a particle density ρ and we have the favorable situation of a cell-list edge length $l = r_c$. For the Verlet list, we use a skin size of $r_s = 0.2\sigma$ leading to $r_v = 2.7\sigma$. Then the average number of neighbor particles for which we need to compute pair interactions (per particle) in the cell list algorithm is $n_{\text{cell}} = 27\rho l^3$. The same

⁵This excludes electro- and magnetostatic interactions (or their dynamic equivalent), which are truly long ranged and which require their own set of techniques [15] that go beyond the scope of this Chapter. Though, it should be noted that in an electrolyte most electrostatic interactions are screened and can typically be replaced with effective, short-ranged Yukawa potentials, depending on the scale being resolved.

number for the algorithm based on a Verlet list is $n_{\text{verlet}} = (4/3)\pi\rho r_v^3$. This leads to the ratio $n_{\text{cell}}/n_{\text{verlet}} \approx 0.2$ for our example, i.e., a non-negligible factor of 5. The factor is less significant in 2D, but the concept can be understood referencing Fig. 18.3. This is why cell and Verlet lists are often combined to achieve optimal computational performance.

Example 18.1: Accelerating MD simulations using a cell list. In this example, we will construct a simple MD algorithm in Python for the simulation of a system of particles interacting via the Weeks-Chandler-Anderson potential [17] and accelerate it using a cell list. We have chosen the dimension of the system to be 2D for convenience such that the result is relatively easy to visualize. Please refer to the accompanying Python script `accelerating_md.py` to complete this example. The script does not function in its present state and you will need to complete parts of this example to get a working and verified MD simulation. The parts of the code that need modification have been indicated using `# TODO` comments.

- a. Look up the definition of the *Weeks–Chandler–Anderson potential* (WCA potential) [17], a variant of the well-known Lennard–Jones potential that eliminates the attractive tail. Derive the force between two WCA particles and implement it in the marked lines of the script.
- b. Modify the lines in the code that should account for the periodic boundary conditions in the distance calculation. It is advised to do so using the `d_sys` parameter, rather than using a 2D specific implementation. For the next example, it will prove convenient to assume a cubic simulation volume with edge length L centered on the origin and with its edges parallel to the coordinate frame.
- c. Implement the initialization of the particles and velocities. Recall which ensemble you are working in and the implications this has for determining the temperature and pressure. For the purpose of this example, it will not be necessary to thermostat the simulation.

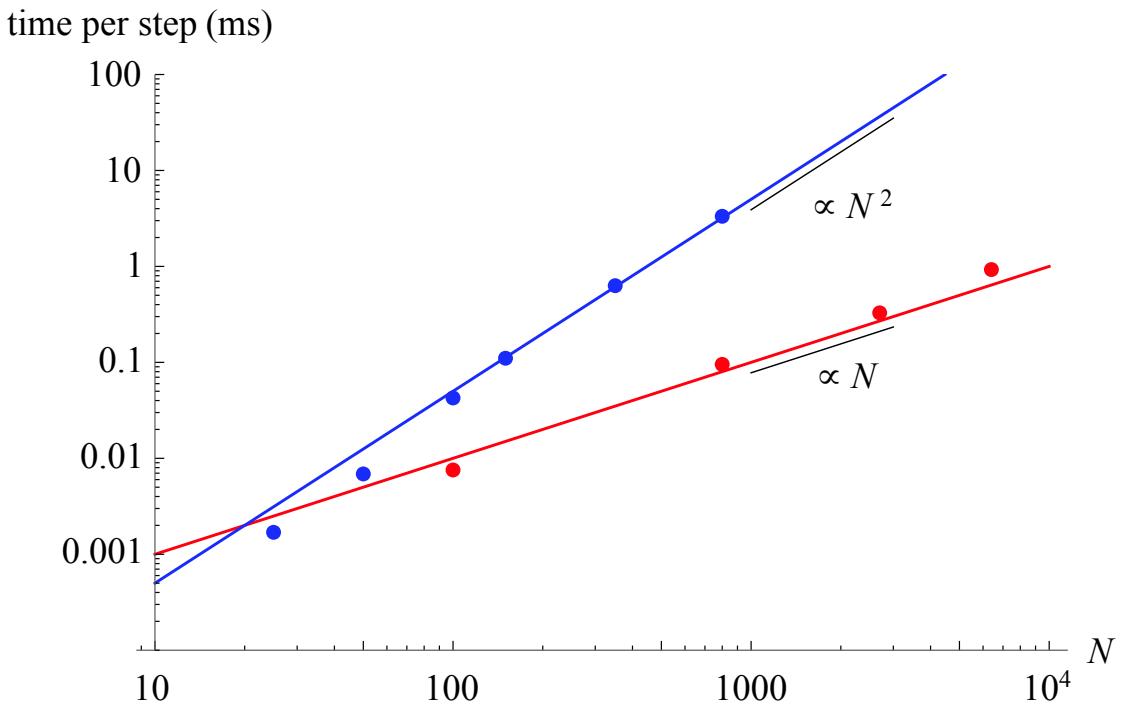


Figure 18.4: **Accelerating MD simulations using a cell list.** Comparison between a plain MD simulation (blue) showing quadratic scaling of the simulation time (ms) as a function of the number of particles N , and a cell-list accelerated MD simulation (red) showing linear scaling. The simulations were performed at a fixed density of $\rho = 0.1$ and the timings were obtained using a C++ variant of the solution to the first example.

d. Modify the code such that you can output the positions of the disks periodically, i.e., not too often otherwise it slows down the simulation, and visualize the dynamics. Check if your algorithm is visually behaving correctly. Note that at this point of coding you have two places where you can introduce bugs, in the visualization routine and the simulation itself. Be mindful of this aspect as you work to get the dynamics up and running.

e. Add code to compute the kinetic and potential energy and verify that your algorithm preserves the total energy. Use the former to establish the average temperature T ; do not forget to equilibrate your simulation.

f. The instantaneous virial contribution to the pressure is given by

$$p_{\text{vir}} = \frac{1}{Vd} \sum_{i=1}^N \sum_{j>i}^N \mathbf{r}_{ij} \cdot \mathbf{f}_{ij}, \quad (18.1)$$

where V is the system's volume (area in 2D), d is the dimension ($d = 2$ here), \mathbf{r}_{ij} is the distance vector between particles i and j , and \mathbf{f}_{ij} is the corresponding force. The average pressure p is now given by

$$p = \frac{k_B TN}{V} + \bar{p}_{\text{vir}}, \quad (18.2)$$

with k_B Boltzmann's constant and the bar indicating the (properly equilibrated) time average. In principle, you can now verify if your simulation gives the correct result by establishing a few (T, p) points.

g. Now, time your simulation for a few values of N to examine the scaling. Comment out all parts that are irrelevant for advancing the simulation, as they only reduce performance.

h. Implement a rudimentary cell list in the MD code following the directions below:

1. Start by defining a cell size which divides the box length and exceeds the interaction length.
2. Create a d -dimensional list, of which the elements are the indices of the particles belonging to the corresponding cell.
3. Loop over the particles and store the indices in the relevant cells.
4. Rewrite your force calculation to loop over the cells and relevant neighbors (do not forget the periodic boundary conditions) before computing the distances.

Show that this works by verifying that the trajectories obtained using the original and new algorithm are identical provided the initial configuration is the same.

i. Now, time your simulation for a few values of N to examine the (hopefully) improved scaling.

18.2.3 Linked list

Lastly, the concept of a *linked list* should be mentioned and, in connection to this, some aspects of computer architecture. A standard implementation of a list of neighbors might take the form of a long list of indices or even structs or classes that catalogue the positions and velocities of the involved pairs. However, it is also possible to take a different approach, wherein each element in a list points to the location of the next element in random access memory (RAM). This is referred to as a linked list, where the link is the pointer to the memory address of the next element. This bypasses the need for an indexing list and facilitates the restructuring of the neighbor list as the particles move in and out of the range r_v . The reduction comes from the computational overhead involved with restructuring a large block of memory by inserting/deleting elements from an indexing list. A linked list can be straightforwardly implemented using programming languages such as C, C++, or Fortran. We refer to Ref. [18] for an in-depth discussion of this topic and general pointers on MD simulations.

Beyond these observations, there are additional considerations to be made involving parallel programming, communication between nodes of a cluster, hyperthreading, having particles of different size, etc. Some speed gain may also be achieved by a clever use of memory organization and utilizing the cache sizes on modern CPUs to store elements of computations that occur frequently. However, this goes into very high-end compu-

tational physics and a full description of this lies beyond the scope of this book. As an example, we refer any interested reader to Ref. [19] to gain a flavor of the types of considerations that come into play. To close this section and because we will touch upon the topic of hydrodynamics, we also note that thermalizing the MD simulation requires a Galilean-invariant thermostat to recover the results of fluctuating hydrodynamic continuum equations [20, 21].

18.3 Event-driven molecular dynamics

In discussing MD algorithms, we have thus far considered analytic potentials, meaning that they are infinitely differentiable and that the associated forces are equally well behaved. The *standard* MD introduced by Rahman [22] resolves trajectories in a discrete manner, i.e., this algorithm uses a finite time step to integrate the dynamics. Reducing the time step leads to a better approximation of the true trajectories, which are smooth. However, there is a class of interesting pair interactions that cannot be handled in this manner, namely those for which the potential has a jump or even an infinite value. In systems with these kind of interactions, the dynamics of the particles is not smooth. Widely used models that possess such features are the hard-sphere and square-well potential:

$$\Phi_{\text{HS}}(r) = \begin{cases} \infty & r \leq \sigma \\ 0 & r > \sigma \end{cases} \quad \text{and} \quad \Phi_{\text{SW}}(r) = \begin{cases} \infty & r \leq \sigma \\ -\varepsilon & \sigma < r \leq \sigma(1 + \lambda) \\ 0 & r > \sigma(1 + \lambda) \end{cases}, \quad (18.3)$$

respectively, see Fig. 18.5. Here, σ is the diameter, $\lambda > 0$ the (reduced) extent of the well, and ε the attraction strength. These two models are immensely popular in minimal modeling of complex systems, e.g., colloidal gels [23], and in fundamental studies, e.g., involving nucleation [24] and the glass transition [25]. In the following, we shall focus on the hard-sphere case, which is the simplest, but these considerations may be readily generalized.

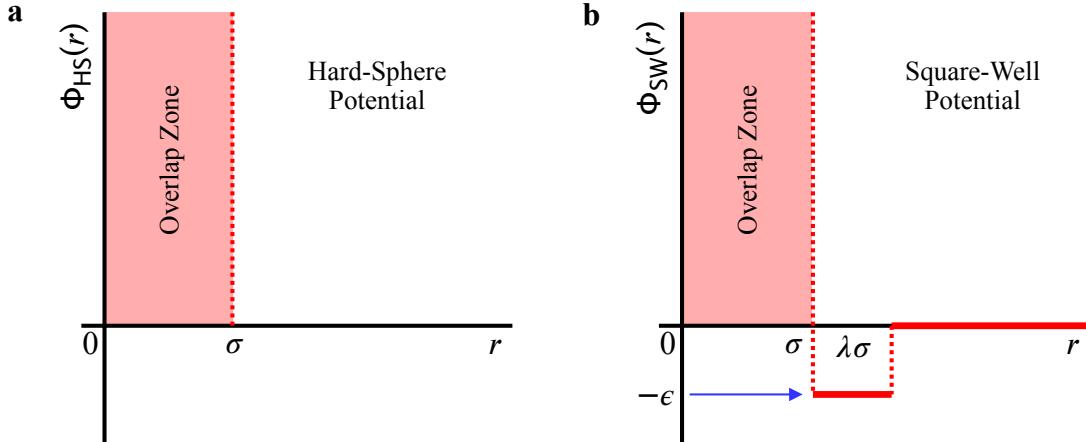


Figure 18.5: **MD models with a non-analytical pair potential.** (a) The hard-sphere and (b) square-well potential, $\Phi_{\text{HS}}(r)$ and $\Phi_{\text{SW}}(r)$, respectively, with r the radial distance between particles. The dashed lines indicate the discontinuities in the potentials and the red zone that part where the potential is ∞ . The square-well potential has a range $\lambda\sigma$ for which it has an attractive well ($\varepsilon > 0$). When $\varepsilon < 0$ the interaction is referred to as a square-shoulder potential.

Unfortunately, differentiation of the hard-sphere potential leads to a Dirac- δ distributed force. That is, the step-like nature of the potentials leads to a force $\mathbf{F}_{\text{HS}} = -\nabla\Phi_{\text{HS}}(r) = \delta(r - \sigma)\hat{\mathbf{r}}$, where ∇ denotes the gradient and $\hat{\mathbf{r}}$ is the radial unit vector. The shell-localized nature (and divergence) of the force is problematic, since no amount of time-step reduction allows for such an interaction to be accurately captured in Rahman MD. Effectively, the particles will move through each other, as particles coming into contact and experiencing a

(divergent) force has vanishing likelihood of occurring. This means that these particles are not suited to be studied using Rahman MD.

Fortunately, systems featuring these potentials lend themselves well to study by Monte Carlo (MC) simulations, for which only energy differences are taken into account. However, MC exploits the equivalence between the time-resolved picture of statistical physics and the phase-space sampling that applies in equilibrium [26]. Hydrodynamic effects fall in the realm of non-equilibrium physics and should not be targeted using MC methods.

The dynamics of this important class of models can, however, be recovered using *event-driven MD* or EDMD, which was introduced by Alder and Wainwright [27]. EDMD may also be applied to study very dilute systems, where integration using a fixed time step is highly inefficient, as collisions are infrequent. The idea behind the algorithm becomes clear by considering a single collision between two otherwise freely moving spherical particles. Before the collision the sphere move along analytically known rectilinear trajectories; this is also the case afterward. At the moment of collision, the speed and direction of motion of the involved spheres changes in a well-known way. The change follows from mass, (linear⁶) momentum, and energy conservation.

Thus, to know the evolution of a system of multiple spheres, we need to only focus on the collisions, which we shall refer to as *events*. This makes sense from the kinetic picture of matter, wherein the physical properties are defined through the collisions of the constituents. Consider a cuboidal simulation volume with periodic boundary conditions and two spheres (out of N) labelled i and j , which at time $t = 0$ are located at \mathbf{r}_i and \mathbf{r}_j and have velocities \mathbf{v}_i and \mathbf{v}_j , respectively. Introduce the difference vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$. Then, the spheres will either never collide or collide at some time τ_{ij} in the foreseeable future. It is straightforward to show that

$$\tau_{ij} = -\frac{1}{v_{ij}^2} \left(b_{ij} + \sqrt{b_{ij}^2 - v_{ij}^2 (r_{ij}^2 - \sigma^2)} \right), \quad (18.4)$$

where we have introduced the vector lengths $r_{ij} = |\mathbf{r}_{ij}|$ and $v_{ij} = |\mathbf{v}_{ij}|$, and the dot product $b_{ij} = \mathbf{r}_{ij} \cdot \mathbf{v}_{ij}$. The update algorithm of EDMD now simply consists of computing the collision time for all particles (including periodic images), selecting the minimum collision time (say τ^*), advancing all particles to their positions at τ^* , and updating the velocity vectors of the pair that has interacted according to the rules of the elastic scattering.

The above algorithm is, when naively implemented, rather inefficient, since it requires the calculation and sorting of N^2 interaction events. It may be considerably sped up by the use of cell lists, where an additional event is implemented, namely the crossing from one cell to another. This reduces the number of event calculations substantially. However, more clever methods of bookkeeping may be employed, namely by using a linked collision tree [28]. This has similar advantages to the linked list in standard MD.

Example 18.2: A simple event-driven MD simulation. In this example, we will establish a basic EDMD simulation using Python, which gives a flavor of the more mature algorithms that are used in production-level simulations. That is, we will not concern ourselves with features such as a linked collision tree, which are best implemented in another programming language. To get started, take your completed `accelerating_md.py` script and change it to 3D. Most of the initialization part can be maintained as is. We will assume a cubic simulation volume with edge length L centered on the origin and with its edges parallel to the coordinate frame. The spheres will bounce off of these elastically.

a. Derive Eq. (18.4). Establish an expression for the way the velocity of spheres changes when two collide. Note that the collision need not be central and that you should ignore rotational degrees of freedom throughout. Verify your answer by referring to Ref. [28]. Why can three-sphere collisions be safely ignored?

b. Derive an expression for the collision time of a single sphere with a wall — the wall is defined through a normal vector $\hat{\mathbf{n}}$ and a suitable base point \mathbf{w} — and an expression for how wall impact changes the sphere's velocity.

c. Implement two collision detection algorithms that determine the time until the next contact. One for interaction with the walls of the simulation box, for which you can use the expression derived in b. The other for interaction with another particle using Eq. (18.4). Note that for the latter you may assume that there are no walls, since collisions that occur *outside* the simulation volume, would lead to a wall collision

⁶Rotations are also neglected for simplicity's sake.

first.

- d.** Construct a loop over particles and pairs and establish all the collision times and the particle(s) involved.
- e.** Sort this list and establish the minimum collision time τ .
- f.** Advance all particles to their new positions for time $t + \tau$, and additionally update the velocity vector(s) for the particle(s) taking part in the collision. You have now completed a basic EDMD algorithm.
- g.** Challenge: Modify the algorithm to achieve the same Couette flow as you studied in the previous example. Hint: you will need to deal with partial periodic boundary conditions.
- h.** Perform the correct time averaging of the velocity profile in the channel; be careful with the length of the time steps. Do your results differ substantially? Why is this expected?

18.4 Problems

Problem 18.1: Couette flow. MD simulations conserve quantities such as mass, momentum, and energy. The associated Hamilton dynamics leads to the emergence of the transport of local quantities when the particles are subjected to external influences. Between these two limits — the particle and hydrodynamic — there is the probabilistic domain of the Boltzmann transport equation, see also Chapter 24. To gain some feeling for the emergence of hydrodynamic quantities, we will consider simple channel flow in this example, which is based on the principles laid out in Ref. [8]. To get started, take your completed `accelerating_md.py` script and change it to 3D. Most of the initialization part can be maintained as is.

- a.** Introduce two walls enclosing the simulation volume in the z -direction. Assume that the interaction potential between the walls and the particles is also the WCA potential. The relevant distance is now the separation in the normal direction. Keep in mind where you place the wall in terms of how wide the channel will be to the centers of mass of the particles.
- b.** Add a new *species* of particle to the simulation, which are just touching the top and bottom wall, and have a prescribed velocity $\pm v_w \hat{x}$ by which they are moving, respectively. What you require for the number and distribution of particles on the top and bottom wall in order to conserve momentum? As you can read in the paper, there are various ways of organizing the wall particles, which give rise to different hydrodynamic boundary conditions.
- c.** Update your force algorithm to take into account the presence of the new species which leads to different amounts of slip. Note that this species is not updated according to the Velocity verlet algorithm! Clearly not having any wall particles constitutes a full slip condition, which leads to no net flow.
- d.** Take a time average of the particle velocities in the channel and see if you can recover the Couette profile.

Bibliography

- [1] D. Tong. Lectures on kinetic theory. <http://www.damtp.cam.ac.uk/user/tong/kinetic.html>, 2017.
- [2] S. Gravelle, L. Joly, C. Ybert, and L. Bocquet. Large permeabilities of hourglass nanopores: From hydrodynamics to single file transport. *Journal of Chemical Physics*, 141:18C526, 2014.
- [3] C. Dekker. Solid-state nanopores. *Nature Nanotechnology*, 2:209, 2007.
- [4] S. Garaj, W. Hubbard, A. Reina, J. Kong, D. Branton, and J. A. Golovchenko. Graphene as a subnanometre trans-electrode membrane. *Nature*, 467:190–193, 2010.
- [5] J. Nivala, D. B. Marks, and M. Akeson. Unfoldase-mediated protein translocation through an α -hemolysin nanopore. *Nature Biotechnology*, 31:247–250, 2013.
- [6] G. Hummer, J. C. Rasaiah, and J. P. Noworyta. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature*, 414:188, 2001.
- [7] M. Majumder, N. Chopra, R. Andrews, and B. J. Hinds. Enhanced flow in carbon nanotubes. *Nature*, 438:44, 2005.

- [8] D. M. Huang, C. Sendner, D. Horinek, R. R. Netz, and L. Bocquet. Water slippage versus contact angle: A quasiuniversal relationship. *Physical Review Letters*, 101:226101, 2008.
- [9] T. Krüger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva, and E. M. Viggen. *The Lattice Boltzmann Method*. Springer Cham, 2017.
- [10] R. Kapral. *Multiparticle Collision Dynamics: Simulation of Complex Systems on Mesoscales*, pages 89–146. John Wiley & Sons, Ltd, 2008.
- [11] L. Quartapelle. *Numerical Solution of the Incompressible Navier-Stokes Equations*. Birkhäuser Basel, 2013.
- [12] A. M. Fiore and J. W. Swan. Fast stokesian dynamics. *Journal of Fluid Mechanics*, 878:544–597, 2019.
- [13] C. H. Mak. Large-scale simulations of the two-dimensional melting of hard disks. *Physical Review E*, 73:065104, 2006.
- [14] M. Engel, J. A. Anderson, S. C. Glotzer, M. Isobe, E. P. Bernard, and W. Krauth. Hard-disk equation of state: First-order liquid-hexatic transition in two dimensions with three simulation methods. *Physical Review E*, 87:042134, 2013.
- [15] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369:253, 1921.
- [16] L. Verlet. Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159:98–103, 1967.
- [17] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54:5237–5247, 1971.
- [18] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publications, 1987.
- [19] M. P. Howard, J. A. Anderson, A. Nikoubashman, S. C. Glotzer, and A. Z. Panagiotopoulos. Efficient neighbor list calculation for molecular simulation of colloidal systems using graphics processing units. *Computer Physics Communications*, 203:45, 2016.
- [20] C. P. Lowe. An alternative approach to dissipative particle dynamics. *Euro Physics Letters*, 47(2), 1999.
- [21] S. D. Stoyanov and R. D. Groot. From molecular dynamics to hydrodynamics: a novel Galilean invariant thermostat. *Journal of Chemical Physics*, 122:114112, 2005.
- [22] A. Rahman. Correlations in the motion of atoms in liquid argon. *Physical Review*, 136:A405, 1964.
- [23] G. Foffi, C. De Michele, F. Sciortino, and P. Tartaglia. Scaling of dynamics with the range of interaction in short-range attractive colloids. *Physical Review Letters*, 94:078301, 2005.
- [24] S. Auer and D. Frenkel. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature*, 409:1020, 2001.
- [25] R. J. Speedy. The hard sphere glass transition. *Molecular Physics*, 95:169, 1998.
- [26] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 1996.
- [27] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. General method. *Journal of Chemical Physics*, 31:459, 1959.
- [28] D. C. Rapaport. The event scheduling problem in molecular dynamic simulation. *Journal of Computational Physics*, 34:184, 1980.

Chapter 19

Implementation with LAMMPS

CRISTOVÃO S. DIAS

LAMMPS is an acronym for Large-scale Atomic/Molecular Massively Parallel Simulator. It is a classical Molecular Dynamics package optimized for parallel computing and mainly directed for materials modeling. Its development began in the mid 1990s, with the first versions in FORTRAN 77 and 90. It was built under a cooperative research and development agreement (CRADA) between two DOE labs (Sandia National Lab and LLNL) and three companies (Cray, Bristol Myers Squibb, and Dupont). The coding effort was led by Steve Plimpton from Sandia National Lab [1, 2, 3, 4, 5, 6]. The current versions are written in C++ with open source code starting from 2004, however, the interaction with LAMMPS is through a scripting language (see Fig. 19.1). The package is freely available for download under GPL, it is designed to be easy to modify or extend, and very well documented (see <https://lammps.sandia.gov>).

Packages like LAMMPS are very modular and highly optimized for parallel computing. However, due to the complexity of the code, it is always hard to start using them and to get to know all their possibilities. This chapter aims at providing a brief introduction to LAMMPS and how to use it to simulate soft active matter, starting by implementing some molecular dynamics methods. We present a hands-on introduction to first-time users and we continue with an advanced section focused on implementing active Brownian particles.

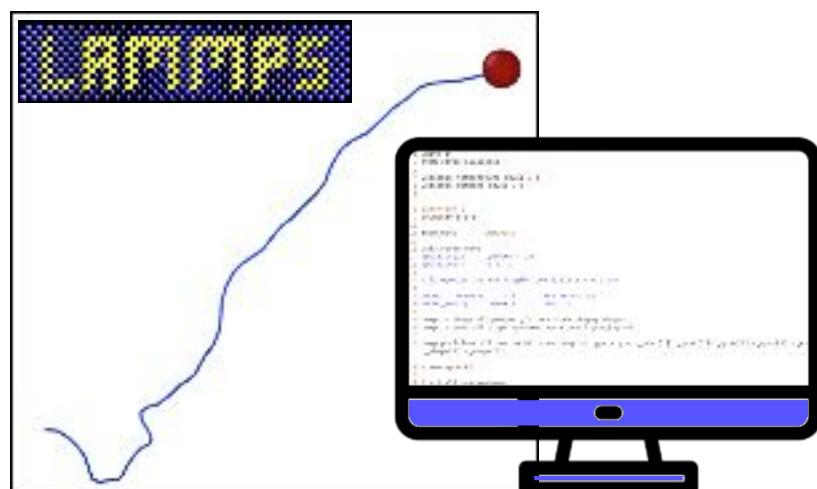


Figure 19.1: **LAMMPS scripting to generate Active Brownian Particle trajectory.** Schematic representation of the scripting process to perform Molecular Dynamics simulations in LAMMPS. The proper trajectory of an Active Brownian Particle was generated.

19.1 Basic algorithms

In this section, we discuss the basic molecular dynamics methods implemented in LAMMPS. Since molecular dynamics simulations consist in integrating the Newton's equations of motion for many particles [7], we start by discussing the integration method. This method needs a force computation, which is discussed in the following section. Some concepts of parallelization are discussed next. We then show how Langevin dynamics simulations are performed in LAMMPS. Finally, we provide a brief discussion on reduced units.

19.1.1 Velocity Verlet algorithm

Many different methods can be used to integrate Newton equations of motion. LAMMPS uses the velocity Verlet method discussed in Section 16.1.

LAMMPS implementation One could implement directly the two steps given by Eqs. (16.13), however these steps need to save, at all time, two acceleration (or force) vectors, which take up memory and communication time between parallel jobs. The most common implementation, and the one used in LAMMPS, computes the velocity twice, using an intermediate half timestep jump forward. The steps in one dimension are the following:

1. Compute $v(t + \Delta t/2) = v(t) + a(t) \frac{\Delta t}{2}$.
2. Compute $x(t + \Delta t) = x(t) + v(t + \Delta t/2) \Delta t$.
3. Compute the forces at $x(t + \Delta t)$.
4. Compute $v(t + \Delta t) = v(t + \Delta t/2) + \frac{1}{2}a(t + \Delta t)\Delta t$.

With this implementation, only one vector of forces is used since the forces are computed between the two steps.

19.1.2 Force computation

The force calculation is usually the bottleneck for an efficient molecular dynamics simulation as discussed in Chapter 18.

LAMMPS implementation In LAMMPS, both methods presented in Figs. 18.3 are implemented. The cut-off radius, r_{cut} , depends on the potential and the skin is defined (command `neighbor` in the LAMMPS script). LAMMPS considers the length of the cells in the cell list, $r_{\text{cut}}/2$, as default but its value can be controlled (command `neigh_modify` in the LAMMPS script).

19.1.3 Parallel computing

One of the main advantages of molecular dynamics simulations is their scalability when split into multiple processes. Basically, most tasks can be done in parallel, needing only information from the positions of other particles during the force calculation. One of the most common method for parallel computing is the spatial decomposition method [6]. As you can see in Fig. 19.2, the space is divided into domains that are assigned to different processes. Particles inside the red region that belong to a certain process are also ghost particles for processes in adjacent domains. When the force is computed, all particles in the domain plus the ghost particles are used.

LAMMPS implementation In LAMMPS, the length of the red region is given by $r_{\text{cut}} + \text{skin}$ (the command `comm_modify` in the LAMMPS script can be used to change the length of the red region). While the mapping of the spacial domains can be controlled in LAMMPS (command `processors` in the LAMMPS script), by default, LAMMPS optimizes the domain distribution in such a way to minimize the red region volume. In the molecular dynamics pipeline, one needs to add a step where the communication between processes is done and the particles that belong to each process are updated. In LAMMPS, this communication step occurs after the first integration of the velocity Verlet method (before step 3 of the force computation in Section 19.1.1).

19.1.4 Langevin dynamics

Soft matter dynamics usually occur inside viscous environments (fluids) so, as discussed in Section 17.3, to perform molecular dynamics simulations one needs to add the interaction with the fluid, which can be given by the Langevin equation for the translational degrees of freedom in Eq. (17.7) (in one dimension). For a rigid body, one also needs the Langevin equation for the rotational degrees of freedom. The Langevin equation for the one-dimensional diffusion of a single-axis rotation is given by

$$I\dot{\omega}(t) = -\nabla_{\theta}V(\theta) - \gamma_R\omega(t) + \xi_R(t), \quad (19.1)$$

where I is the moment of inertia and γ_R the rotational damping coefficient along that single-axis rotation, while $\xi_R(t)$ is the stochastic torque from the thermal noise, given from a distribution of zero mean and satisfying the relation

$$\langle \xi_R(t)\xi_R(t') \rangle = 2k_B T \gamma_R \delta(t-t'). \quad (19.2)$$

For both translational and rotational motion, above a damping time (time for the inertia to become negligible), the Brownian diffusive regime is recovered. We obtain the translational diffusion coefficient given by the Stokes-Einstein relation as [8]

$$D = \frac{k_B T}{\gamma} \quad (19.3)$$

and the rotational diffusion coefficient as

$$D_R = \frac{k_B T}{\gamma_R}. \quad (19.4)$$

The damping times for the translational and rotational motions are $\tau = m/\gamma$ and $\tau_R = I/\gamma_R$, respectively, where m and I are mass and inertia of the colloidal particle.

The translational and rotational diffusion coefficients presented here are generic for passive or active particles. In the case of active particles, the rotational diffusion coefficient sets the time scale for the self-propulsion directional motion to forget its initial orientation (for active particles the effective diffusion coefficient depends also on the propulsion, which is discussed in Section 19.3). For passive particles, one can find a relation between translational and rotational diffusion. For spherical passive particles, these coefficients take the form [8]

$$D = \frac{k_B T}{6\pi\eta R} \quad (19.5)$$

and

$$D_R = \frac{k_B T}{8\pi\eta R^3}, \quad (19.6)$$

where η is the viscosity of the fluid medium and R the radius of the particles. It is possible to obtain a relation between the rotational and translational diffusion coefficient given by

$$\frac{D_R}{D} = \frac{3}{4R^2}. \quad (19.7)$$

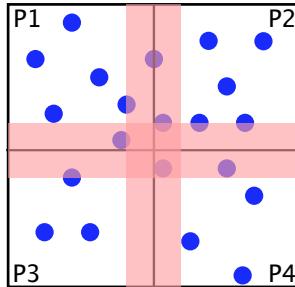


Figure 19.2: **Spatial domain decomposition by processes P1 to P4.** Each particle is allocated to a process. Particles in the red region that belong to a process are ghost particles in the processes of the adjacent domains.

LAMMPS implementation The LAMMPS variables that parameterize the Langevin dynamics are the translational damping time, τ and the mass m (command `fix langevin` in the LAMMPS script). Since only the translational damping time is defined, an isotropic scale factor $\alpha = \tau/\tau_R$ between translational and rotational damping, is introduced. For spherical particles, using Eq. (19.7), one can write the scale factor as $\alpha = 10/3$. For rigid aspherical particles, α is an external parameter (`angmom` in the Langevin fix) and, for rigid, more complex bodies, if they are approximately spherical, one needs to add the factor 10/3 directly into the code [9], while for more complex shapes the scale factor is 1 [10]. In LAMMPS, the Langevin equation for the translational motion, in one dimension, is given by

$$m\dot{v}(t) = -\nabla_r V(r) - \frac{m}{\tau} v(t) + \sqrt{\frac{2mk_B T}{\tau}} \xi(t) \quad (19.8)$$

and for the rotational motion, around one single axis of rotation, it is

$$I\dot{\omega}(t) = -\nabla_\theta V(\theta) - \frac{\alpha I}{\tau} \omega(t) + \sqrt{\frac{2\alpha I k_B T}{\tau}} \xi(t). \quad (19.9)$$

In Eqs. (19.8) and (19.9), v and ω are the translational and angular velocity, m and I are again the mass and inertia of the colloidal particle, and V is the external potential. In LAMMPS, for efficiency [11], $\xi(t)$ is a uniform distribution from -0.5 to 0.5 , which forces a correction in the stochastic term of the Langevin equation since the second moment of the uniform distribution is given by

$$\sigma^2 = \int_{-0.5}^{0.5} x^2 dx = \frac{2}{3} \left(\frac{1}{2}\right)^3 = \frac{2}{24} \quad (19.10)$$

and imposes a correction of $\sqrt{2/24}$, that can be seen in the source code in the file `fix_langevin.cpp`. In the same source file, the integration timestep is included in the denominator inside the square root because LAMMPS multiplies all forces (including the stochastic term from the thermal noise) by the timestep in the velocity Verlet integration (see Section 19.1.1). The timestep in the stochastic term comes from the variance of the random number distribution, which is $\sigma^2 = 2D\Delta t$ (the multiplication factor in the stochastic term of the Langevin equation). In this way, the diffusion coefficient does not depend on the timestep.

19.1.5 Reduced Units

When modeling soft matter, we are usually working with length scales on the order of nanometers to micrometers and energy scales on the order of 10^{-1} to 10^{-2} electronvolts. If we decide to use standard SI units, we will have to work with very small values. It is, of course, more convenient to work with values near unity, which can be done by using the size of the simulation elementary particles as reference. But reduced units are not only used for convenience. Performing numerical calculations with very small values can lead to rounding errors due to the machine precision. Using *reduced* (or *dimensionless* or *unitless*) units also facilitates scaling, where a single model can help to understand a large class of problems at different scales [12].

LAMMPS implementation A common scale for reduced units is the one of the Lennard-Jones potential [12], given by Eq. (16.17). LAMMPS uses the units of distance/energy from this equation and, without loss of generality, sets the fundamental quantities m , σ , ϵ , and k_B to unity. The masses, distances, and energies are then multiples of these fundamental values. The formulas relating the reduced quantity are show in Table 19.1.

The conversion to real units can be sometimes confusing. Nonetheless, it is always better to use reduced units instead of SI units (which LAMMPS allows). The conversion between real units and reduced units can be done by using the relations in Table 19.1. For instance, soft matter experiments are usually performed at room temperature ($T = 293$ K). So, taking Boltzmann constant, $k_B = 1.38 \times 10^{-23}$ m² kg s⁻² K⁻¹, we obtain the energy unit of $\epsilon = 4.04 \times 10^{-21}$ J. The same can be done with the length units, for instance, if we set the mass of a colloid as $M = 10m$ and the diameter as $d = 10\sigma$ in the simulations, and consider the real values of colloid mass 10^{-12} g and radius 0.5 μm, we recover the fundamental quantities of $\sigma = 10^{-7}$ m and $m = 10^{-16}$ kg respectively. The same can be done with the time units. If we set a timestep in the simulations of $\Delta t = 0.01$, this tells us that each simulation step takes 1.57×10^{-7} s. For Langevin dynamics simulations, it is sometimes clearer to use the *Brownian time* [13] (the time for a particle to diffuse the square of its diameter) to set the timescale.

Table 19.1: LAMMPS reduced units.

Variable	Symbol	Value
Mass	m	1
Distance	σ	1
Energy	ϵ	1
Boltzmann constant	k_B	1
Temperature	T	$\frac{\epsilon}{k_B}$
Time	τ	$\sqrt{\frac{m\sigma^2}{\epsilon}}$

19.2 Hands-on practice

In this section, we will have a more practical approach to LAMMPS. We will start by downloading and compile LAMMPS. We will then run a test example. Finally, we will perform some measurements on a Langevin dynamics simulation of passive particles.

19.2.1 Compiling LAMMPS

LAMMPS was primarily developed for Linux systems, mainly because most of the high-performance computers are Linux-based. With this in mind, if you have a Linux system, it is straightforward to download the most recent tarball (<https://lammps.sandia.gov/download.html>), uncompress and compile by performing `make mpi` or `make serial` in the source folder (src/). If you have a Windows 10 or newer I recommend to use the Windows Subsystem for Linux (WSL). Clear instructions are available at https://lammps.sandia.gov/doc/Howto_wsl.html. For older versions of Windows, I recommend you to install some Linux virtual machine. MacOS users can install as in the Linux case.

In LAMMPS, one can find extra packages to include some specific features. These packages were added by the LAMMPS developers or by other users, and should be consistent with the style of the rest of LAMMPS. To check what packages are installed just run the command `make ps` in the source folder. In this Chapter, we will use, for instance, the ASPHERE package and the COLLOID package. You can add them by running the commands `make yes-asphere` and `make yes-colloid`. You need to recompile after adding any new package.

19.2.2 Running LAMMPS

When compiling the code with the `make mpi` or `make serial`, an executable file is created in the source folder with name `lmp_mpi` or `lmp_serial`. To run, one just needs to use the command:

```
./lmp_mpi -in lj.in
```

to run in one core, or

```
mpirun -np 4 ./lmp_mpi -in lj.in
```

to run in four cores (the `lmp_serial` executable does not allow parallelization). The `lj.in` file is a LAMMPS script where the simulation parameters and methods are included. An example of the script for Lennard-Jones interacting particles in a box with periodic boundary conditions is provided in Box 19.1.

Box 19.1: Lennard-Jones interacting particles. LAMMPS script for Lennard-Jones interacting particles in a box with periodic boundary conditions.

```

1 units lj
2 atom_style atomic
3
4 dimension 3
5 boundary p p p
6 region box block 0 30 0 30 0 30 units box
7 create_box 1 box
8
9 lattice sc 0.1
10 create_atoms 1 box
11 mass 1 1.0
12 velocity all create 1 176217 dist gaussian
13
14 pair_style      lj/cut 2.5
15 pair_coeff      1 1  1 1
16
17 dump      DUMPXYZ      all      xyz 5 out.xyz
18
19 timestep 0.005
20 fix 1 all nve
21
22 thermo_style custom step temp pe ke etotal
23 thermo 100
24
25 run 5000

```

Lines 1 and 2 define main properties of the simulations like the reduced (LJ) units and the atom style atomic defines particles as point particles. The lines 4 to 7 define and create the simulation box in three dimensions with periodic boundary conditions in all directions (boundary p p p). A region named box is defined in line 6 and the simulation box with the properties of that region is created (line 7). Lines 9 to 12 create the particles in the region box. They are distributed on a square cube (sc) lattice with a reduced density of particles of $\rho^* = 0.1$, where $\rho^* = N/V$ with N the number of particles and $V = L_x \times L_y \times L_z$ the volume in reduced units. The command mass sets the mass of a particle of type 1 as 1 and the command velocity sets their velocities with a Gaussian distribution around 0 and a variance equal to the temperature 1. The value 176217 is the random number generator seed.

In line 14 and 15, the Lennard-Jones potential is defined with a cutoff radius of 2.5 and coefficients of $\sigma = 1$ and $\epsilon = 1$ (last two values). The first two values of the coefficients indicate which pair of particle types will have those coefficients, as different particles could have different interactions. In line 17, a xyz file is created with the positions of all the particles every 5 time steps. This type of files can be read in different external software; for example, Ovito (<https://www.ovito.org>) is simple to download and use, as well as very powerful. Fig. 19.3 shows the initial and final configurations of a simulation using the proposed script.

In line 19, we define the simulation timestep. One can test larger timesteps to check LAMMPS error outputs when the simulation is unstable. In line 20, the fix nve is applied to all particles and perform integration of the equations of motion in constant number of atoms (N), volume (V), and energy (E). The fix commands in LAMMPS are used to change something in the integration pipeline. In lines 22 and 23, the output information is defined (if no file is defined, the outputs are saved in the LAMMPS log file). In this case, the chosen output variables are: timestep (step), temperature (temp), potential energy (pe), kinetic energy (ke), and total energy (etotal). One simple test is to check the constant total energy of the NVE integration. Line 25 defines the total number of timesteps in the simulation.

Example 19.1: NVE integration of Lennard-Jones particles in a box. Consider the LAMMPS script in Box 19.1.

- Run the script and test if the total energy remains constant over time.
- Reproduce the images of Fig. 19.3.

19.2.3 Brownian motion

To test the concepts introduced in Sec. 19.1.4, we perform Langevin dynamics simulations in LAMMPS using the script in Box 19.2.

Box 19.2: Langevin dynamics of non-interacting particles. LAMMPS script for Langevin dynamics of non-interacting particles in a box with periodic boundary conditions.

```

1 units lj
2 atom_style sphere
3
4 variable temperature equal 1
5 variable damping equal 1
6
7 dimension 3
8 boundary p p p
9
10 read_data      inp.data
11
12 dump      DUMPXYZ      all      xyz 50 out.xyz
13
14 timestep 0.005
15 fix 1 all nve
16 fix 2 all langevin ${temperature} &
   ${temperature} ${damping} 82763871
17
18 compute msd_col all msd

```

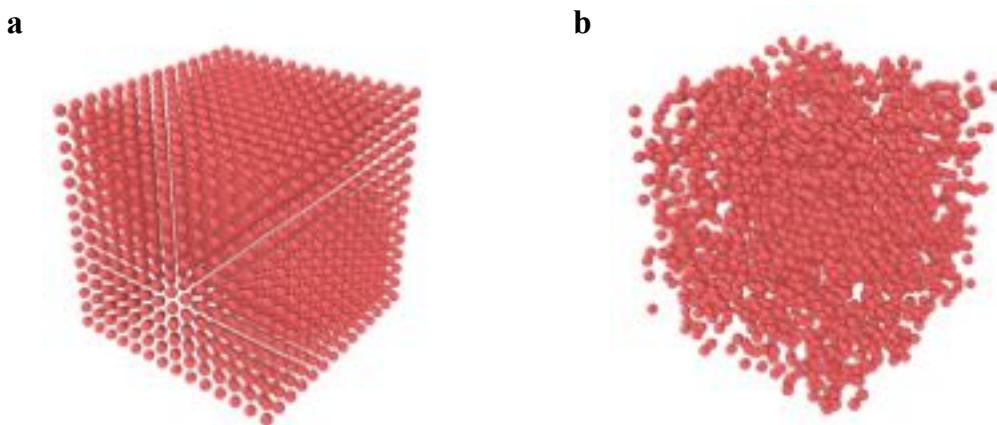


Figure 19.3: **NVE simulation with LAMMPS.** (a) Initial and (b) final configuration of the NVE simulation using the test script provided in Box 19.1. The visualization was performed using Ovito.

```

19
20 thermo_style custom step c_msd_col[1] &
   c_msd_col[2] c_msd_col[3] c_msd_col[4] ke
21
22 thermo 100
23 run 50000

```

Many commands are the same as in the previous script. The one relevant for the Langevin dynamics is the command in line 16 where the Langevin fix is added to the NVE integration. In this fix, extra terms are added to the forces and torque following Eqs. (19.8) and (19.9). We add the starting and stopping temperature (which, in this case, we defined constant over the full simulation run, see line 4) and the damping time introduced in Sec. 19.1.4. The last value is the random number generator seed.

In this script, we introduced a new atom style (`atom_style`) in line 2, which is the `sphere` type, where particles are considered as spheres, with a specific diameter and density. We also introduced the concept of variables in line 4 and 5, which we use in the Langevin fix. One major difference in this script is in line 10 where we read the particles positions from an external file (`inp.data`) with the following structure shown in Box 19.5.

Box 19.3: Structure of input data file. LAMMPS input data with atoms properties, positions and system size.

```

1 # lammps molecular data
2
3 1 atoms
4
5 1 atom types
6
7 0 200 xlo xhi
8 0 200 ylo yhi
9 0 200 zlo zhi
10
11 Atoms
12
13 1 1 1 1.9099 100 100 100

```

We define the number of particles and particle types in lines 3 and 5, and the size of the simulation box in lines 7, 8, and 9. The particles' positions are defined in the `Atoms` sections, where we can see the first (and only) particle (column 1 gives then particle number) of type one (column 2), with a diameter of one (column 3), a particle density of 1.9099 (column 4) and positions in x, y, and z (columns 5, 6, and 7). The order of the particles properties is specific for the sphere atom style (the manual of the `read_data` command has examples for other atom styles). The density was chosen such that a particle of unit diameter has unit mass.

The final new command introduced in the run script of Box 19.2 is in line 18. The command `compute` is used by LAMMPS to perform calculations during the simulation. Here, we compute the mean square displacement (`msd`) of all particles (only one in this data file). The name of the `compute` command is `msd_col` that is used in the output `thermo_style` (line 20), which is a vector with four coordinates (x, y, z, and total), summed and averaged over all particles.

Example 19.2: Langevin dynamics of non-interacting particles. Consider the LAMMPS script in Box 19.2.

- Use the script to perform Langevin dynamics simulations for 1 particle and plot the trajectory as in

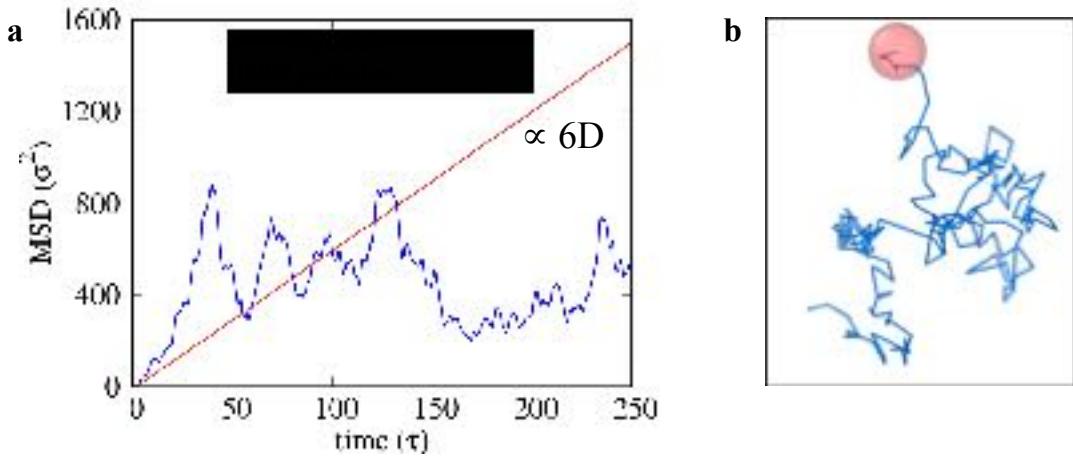


Figure 19.4: **MSD for a single particle and an ensemble of 10,000 non-interacting particles.** (a) Mean square displacement (MSD) as a function of time for a single particle and for 10,000 particles. (b) Typical trajectory of a single particle.

the inset of Fig. 19.4.

b. Reproduce the mean square displacement (MSD) plot of Fig. 19.4.

c. Compute the diffusion coefficient and compare the value with the one obtained from Eq. (19.3).

In Fig. 19.4, we plot the output of the proposed script for a single particle and for 10,000 particles. To implement the many particles case, just add 10,000 consecutive lines in the `inp.data` file. Since no pair potential has been defined, one can just define the same position for all particles. The proper MSD slope, in three dimensions, of $6D$ is recovered, where D is the diffusion coefficient of spherical passive particles given by the Stokes-Einstein relation of Eq. (19.5).

19.3 Customize LAMMPS

As stated before, LAMMPS is an open source code under GPL. This means you are free to download, study, use, and even alter its code. At this point, more than 95% of LAMMPS is made of add-on files to the core ones. Despite being a very complete and powerful tool, some specific cases are not implemented and a user could need to add new features.

One feature, relevant for our topic of active matter, is the inclusion of active Brownian particles simulations, i.e., particles that have a persistent motion in a specific direction. This feature has been recently implemented in LAMMPS (see command `fix propel/self`), while other groups have used custom implementations [14]. However, in this section, we discuss how to build a new class for active Brownian particles, which can be used as an example on how to implement custom methods into LAMMPS. The implementation of active Brownian particles consists of adding a force $\mathbf{F}_A = F\mathbf{e}(t)$ [15, 16], where F is the magnitude of the activity (a constant) and $\mathbf{e}(t)$ is the unit orientation vector. The orientation vector evolution over time is given by Eq. (19.9).

Most of the features to be added to LAMMPS require writing a new C++ class. The simplest way is to find some class that does something similar to the one we want to add. For instance, in the case of an active particle, one just needs to add a force to a specific preferred direction of the particle. Since this is not a pair interaction force, we should not build a new pair class. In this case, we just add a force to the ones already computed in the pair classes much like how it is done in the Langevin dynamics case. The simplest classes one could use as examples are the `fix_addforce` or the `fix_setforce`, which basically add a value or set a value to the force array in a specific function of the class that LAMMPS uses for those cases, called `post_force`. The only thing we need to worry about is how to define the orientation of the active particles. One option is to use ellipsoid particles (where we can define all axis of equal size if we want spheres), which then integrates the rotational motion through the use of quaternions [12]. In the case of ellipsoids, LAMMPS defines the x -axis as

the reference axis and the rotation is defined by the quaternions values using that same axis as a reference. You just need to have a fail check on your new build class that your `atom_style` is ellipsoid.

For a fast implementation of active particles in LAMMPS, one could follow a simpler approach, which includes the following steps:

Step 1: Copy the `fix_langevin` class to a `fix_langevin_active` one. Note that each class has two files, in this case `fix_langevin.cpp` and `fix_langevin.h`. You just need to change the class name inside the files, and in `fix_langevin_active.h` change the call of the fix from `langevin` to `langevin/active` for instance. This way, we use a fix, which is a bit more complex to fully understand than the two examples used before, but is far easier to implement. For starter, the use of ellipsoids is already contemplated.

Step 2: In the constructor function, in the beginning of the class, we can access the arguments of the fix. An option for ellipsoids is the `angmom` (which is the scale factor α of the rotational diffusion in Eq. (19.9)), which has a numerical value after the option. We only need to add a second value with the force (just copy the line with the variable `ascale` and add your new variable from `arg[iarg+2]`) that represents the activity, which will be saved into a variable defined in the header file.

Step 3: The final step to have a fully functioning class with activity is to add, in the `angmom_thermostat()` function, a few lines inside the cycle that runs over the local particles (particles in the same process) which will add a force to the forces vector, in the direction given by the ellipsoid quaternions. One can convert quaternions to the three main axis of the ellipsoids using a function in MathExtra called `q_to_exyz` (use only the principal axis of the ellipsoid).

After implementing this new fix, we just need to recompile LAMMPS with the new files in the source folder. We then run a very similar script as the one for the Langevin dynamics, as shown in Box 19.4.

Box 19.4: Langevin dynamics of active particles. LAMMPS script for the new active particles fix.

```

1 units lj
2 atom_style ellipsoid
3
4 variable t equal 0.1
5 variable d equal 0.1
6
7 dimension 3
8 boundary p p p
9
10 read_data      inp.data
11
12 compute shape all property/atom shapex shapey &
      shapez
13 compute quat all property/atom quatw quatj &
      quatj quatk
14
15 dump positions all custom 50 out.dump &
      id type x y z &
      c_quat[1] c_quat[2] c_quat[3] c_quat[4] &
      c_shape[1] c_shape[2] c_shape[3]
16
17 timestep 0.05
18
19 fix 1 all nve/asphere
20
21 fix 2 all langevin/active ${t} ${t} ${d}  &
      82763871 angmom 3.33 100
22

```

```

23 compute msd_col all msd
24
25 thermo_style custom step c_msd_col[1] &
   c_msd_col[2] c_msd_col[3] &
   c_msd_col[4] ke
26
27 thermo 100
28 run 50000

```

Note that the only change is in the atom style and the `langevin` fix (and a different dump with shape and orientation to test the ellipsoids proper dynamics). In this new fix `langevin/active`, after the `angmom` option we add the `angmom` value and the activity (100 in this script). You can use `angmom` to control the rotational diffusion effect, which in this case is set to $10/3$, the value for spherical passive particles. Since we are dealing with ellipsoids, the `inp.data` file changes as shown in Box 19.5.

Box 19.5: LAMMPS input data with atoms properties, positions and system size.

```

1 # lammps molecular data
2
3 1 atoms
4
5 1 atom types
6
7 1 ellipsoids
8
9 0 200 xlo xhi
10 0 200 ylo yhi
11 0 200 zlo zhi
12
14 Atoms
15
16 1 1 1 1.9099 100 100 100
17
18 Ellipsoids
19
20 1 1 1 1 -0.654 -0.089 -0.285 0.695

```

Here, the third column in the Atoms section is not the diameter, since the shape is defined in columns 2, 3 and 4 of the Ellipsoids section, but a flag indicating that this particle is an ellipsoid. In the Ellipsoids section, the shape is defined by the diameters in x, y, and z (columns 2, 3 and 4), which is a sphere in this example, and then the four last values are the quaternions quatw, quatx, quaty, and quatz.

Example 19.3: Langevin dynamics of active particles. Consider the LAMMPS script in Box 19.4.

- a. Implement the proposed new fix and recompile LAMMPS.
- b. Use the script to perform Langevin dynamics simulations of 1 particle and plot the velocity as a function of the propulsion force. Compare the slope of the function with the one obtained from Eq. (19.8) at zero temperature in the overdamped limit.
- c. Reproduce the MSD plot of Fig. 19.6.
- d. Compute the effective diffusion coefficient and compare the value with the one obtained from Eq. (19.11).

One simple test to this new active particles class is to measure the velocity at zero temperature for different values of the force. The particles rapidly reach the terminal velocity and, using Eq. (19.8) at zero temperature, we obtain the relation $v = \frac{\tau}{m}F$. This relation is shown in Fig. 19.5.

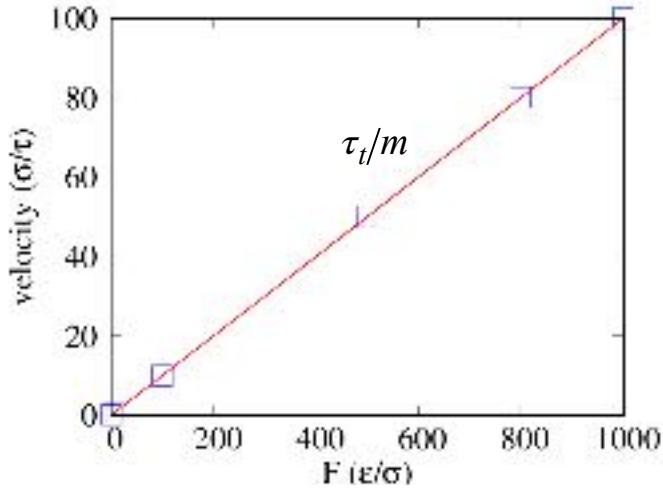


Figure 19.5: **Terminal velocity vs active force.** Terminal velocity as a function of the force, for active particles at zero temperature.

Another test is to measure the MSD of an active particle. In Fig. 19.6, we can see the two typical regimes of an active particle, with a ballistic motion at short times and a diffusive motion for longer times. The diffusive motion has an effective diffusion coefficient given by [17]

$$D_{\text{eff}} = D + \frac{1}{6} \frac{v^2}{D_R}, \quad (19.11)$$

where D is the translational diffusion coefficient, v the velocity, which we show in Fig. 19.5 to be τ/m , and D_R the rotational diffusion coefficient. D_R can be related to D using Eqs. (19.3) and (19.4), and the scaling factor $\alpha = \tau/\tau_R$, which gives the relation

$$\frac{D}{D_R} = \alpha \frac{I}{m} \quad (19.12)$$

and, for spherical passive particles, recovers Eq. (19.7).

19.4 Final Discussion

The LAMMPS package is a very powerful tool. However, similar when implementing molecular dynamics from scratch, we need to be very careful with the parameterization of our models. We always need to start with fewer particles and small systems, without parallelization. Then, as done in these notes, perform simple but effective tests to check if we recover the expected physics.

19.5 Problems

Problem 19.1: Parallelization scaling. Use the script of Box 19.1 to perform NVE simulations of Lennard-Jones particles.

- a. For 10^2 , 10^3 , 10^4 , 10^5 and 10^6 particles, plot the simulation time as a function of the number of parallel cores (`./lmp_mpi -np #cores`) for 1, 2, 4 and 8 (if possible) cores.

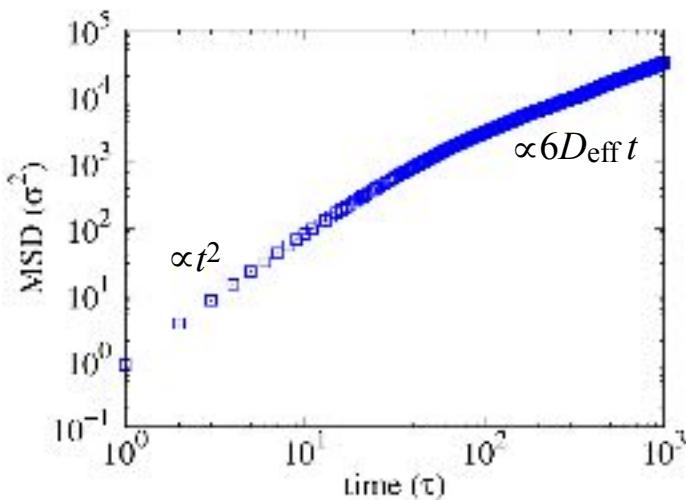


Figure 19.6: MSD of an active particle as a function of time. The calculation of the MSD is performed for 10,000 non-interacting active particles. The two regimes of ballistic and diffusive motion are seen, where the diffusive regime is linear and proportional to $6D_{\text{eff}}$.

- b.** For 10^6 particles compute the simulation time as a function of the skin.

Problem 19.2: NVT simulations. Adapt the script of Box 19.1 to perform NVT simulations of Lennard-Jones particles (replace the `fix nve` by `fix nvt`.)]

- a.** Solve Problem 16.9 using this adapted LAMMPS script.
- b.** Solve Problem 16.11 using the command `fix wall/reflect`.

Problem 19.3: Brownian motion. Use the script of Box 19.2 to perform Langevin dynamics simulations.

- a.** For 10^4 non-interacting particles, plot the diffusion coefficient as a function of the temperature, the damping constant, and the mass of the particles.
- b.** For particles interacting with a Yukawa potential (see Eq. (16.16)), plot the effective diffusion coefficient as a function of the density of particles.

Problem 19.4: Active Brownian particles. Use the script of Box 19.4 to perform Langevin dynamics simulations of active Brownian particles.

- a.** For 10^4 non-interacting particles, plot the effective diffusion coefficient as a function of the active force, the radius of the particles, the mass of the particles, and the scaling factor α of Eq. (19.12).
- b.** For active Brownian particles interacting with a Yukawa potential, plot the effective diffusion coefficient as a function of the density of active particles.
- c.** For a mixture of active and passive Brownian particles, plot the effective diffusion coefficient as a function of the density and fraction of passive particles.

Bibliography

- [1] P. J. in 't Veld, S. J. Plimpton, and G. S. Grest. Accurate and efficient methods for modeling colloidal mixtures in an explicit solvent using molecular dynamics. *Computer Physics Communications*, 179:320, 2008.
- [2] M. L. Parks, R. B. Lehoucq, S. J. Plimpton, and S. A. Silling. Implementing peridynamics within a molecular dynamics code. *Computer Physics Communications*, 179:777, 2008.

- [3] W. M. Brown, P. Wang, S. J. Plimpton, and A. N. Tharrington. Implementing molecular dynamics on hybrid high performance computers - Short range forces. *Computer physics communications*, 182:898, 2011.
- [4] W. M. Brown, T. D. Nguyen, M. Fuentes-Cabrera, J. D. Fowlkes, P. D. Rack, M. Berger, and A. S. Bland. An evaluation of molecular dynamics performance on the hybrid cray XK6 supercomputer. *Procedia Computer Science*, 9:186, 2012.
- [5] M. K. Petersen, J. B. Lechman, S. J. Plimpton, G. S. Grest, P. J. In 'T Veld, and P. R. Schunk. Mesoscale hydrodynamics via stochastic rotation dynamics: Comparison with Lennard-Jones fluid. *Journal of Chemical Physics*, 132:174106, 2010.
- [6] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117:1, 1995.
- [7] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Oxford University Press, Oxford, 1987.
- [8] M. G. Mazza, N. Giovambattista, H. E. Stanley, and F. W. Starr. Connection of translational and rotational dynamical heterogeneities with the breakdown of the Stokes-Einstein and Stokes-Einstein-Debye relations in water. *Physical Review E*, 76:031203, 2007.
- [9] C. S. Dias, C. Braga, N. A. M. Araújo, and M. M. Telo da Gama. Relaxation dynamics of functionalized colloids on attractive substrates. *Soft Matter*, 12:1550, 2016.
- [10] H. P.M. Melo, C. S. Dias, and N. A.M. Araújo. Optimal number of faces for fast self-folding kirigami. *Communications Physics*, 3:154, 2020.
- [11] B. Dunweg and P. Wolfgang. Brownian dynamics without Gaussian random numbers. *International Journal of Modern Physics C*, 2:817, 1991.
- [12] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, Cambridge, 1996.
- [13] C. S. Dias, N. A. M. Araujo, and M. M. Telo da Gama. Annealing cycles and the self-organization of functionalized colloids. *Journal of Physics: Condensed Matter*, 30:014001, 2018.
- [14] J. Stenhammar, D. Marenduzzo, R. J. Allen, and M. E. Cates. Phase behaviour of active Brownian particles: The role of dimensionality. *Soft Matter*, 10:1489, 2014.
- [15] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger. Computational models for active matter. *Nature Reviews Physics*, 2:181, 2020.
- [16] G. Volpe, S. Gigan, and G. Volpe. Simulation of the active Brownian motion of a microswimmer. *American Journal of Physics*, 82:659, 2014.
- [17] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.

Chapter 20

Interaction Rules in Models of Active Matter

DANIEL M. SUSSMAN

It is common in the study of a dizzying array of soft matter systems to perform agent-based simulations of particles interacting via conservative and often short-ranged forces. In this context, well-established algorithms for efficiently computing the set of pairs of interacting particles have established excellent open-source packages to efficiently simulate large systems over long time scales — a crucial consideration given the separation in time- and length-scales often observed in soft matter. What happens, though, when we think more broadly about what it means to construct a neighbor list? What if interactions are non-reciprocal, or if the *range* of an interaction is determined not by a distance scale but according to some other consideration? As the field of soft and active matter increasingly considers the properties of living matter — from the cellular to the super-organismal scale — these questions become increasingly relevant, and encourage us to think about new physical and computational paradigms in the modeling of active matter. In this Chapter, we examine case studies in the use of non-metric interactions.

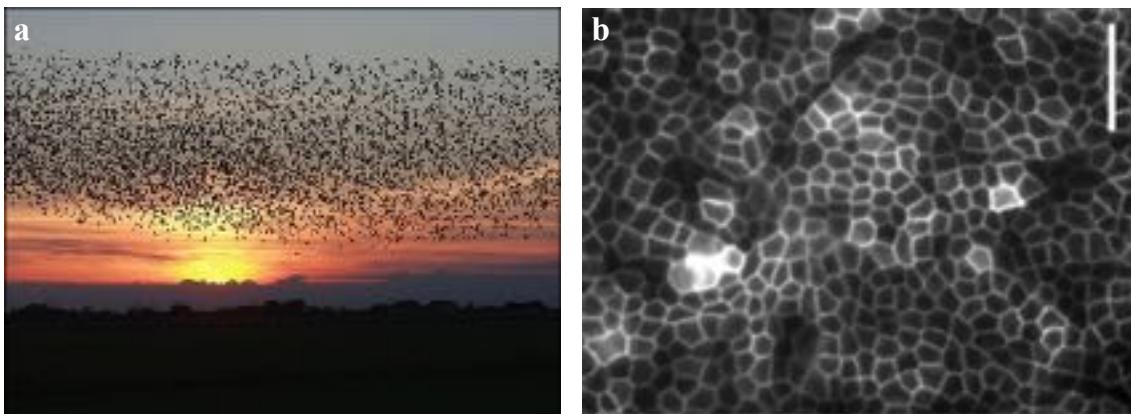


Figure 20.1: **Active matter systems with non-metric interaction.** Whereas classical particles typically interact via conservative potentials that depend on inter-particle separation, simple models in which biological entities are taken to be the degrees of freedom have few restrictions on the effective interactions that govern their displacements. (a) A collective group of starlings flocking, in which individual birds have their trajectories influenced by other birds in their cone of vision. Image from [Wikipedia](#) in the Public Domain. (b) A monolayer of Madin-Darby Canine Kidney cells grown on a collagen substrate (scale bar: 50 micrometers), in which individual cells exert forces on cells that they share an edge with. Image reproduced from Ref. [1].

20.1 Agent-based numerical modeling

In this Chapter, we are going to think about new phenomena that emerge as soft matter moves from the study of squishy equilibrium materials to thinking about active and sometimes living *material* systems. We will begin by thinking first *not* about this physics, but about generalizations of some of the algorithms we use to efficiently perform simulations. We will see that these algorithmic generalizations have natural analogs in physical systems.

Agent-based numerical modeling of soft matter systems have been enormously influential in the development of the field of soft matter. A number of powerful, flexible, open-source packages for performing large-scale molecular dynamics (MD) simulations are readily available and continue to be developed [2, 3, 4, 5, 6], helping radically lower the barrier to entry for researchers in the field.

At its most basic level, the essential computational pattern in a molecular simulation is sketched in this algorithm:

Algorithm 1 Fundamental MD loop

```

Initialize system,  $\mathbf{X}$ 
for Desired number of timesteps do
    Evaluate all pairwise forces,  $\mathbf{F}$ 
    Update  $\mathbf{X}$  based on  $\mathbf{F}$ 
end for
```

A system of N *particles* (e.g., colloids, soft spheres, interacting Lennard-Jones units) are initialized in a simulation domain equipped with a set of boundary conditions, and a sequence of time steps is executed. At each time step, given the current state of the system, all interparticle forces are computed and the information about these forces is used to forward integrate an equation of motion — Newton’s equations of motion, say, or the Langevin equation, and perhaps coupling the system degrees of freedom with a thermostat or barostat to perform simulations that appropriately sample different statistical ensembles [7]. Already, we are assuming that the forces we are interested in are simple, isotropic, pairwise, conservative ones — this situation arises quite naturally in many physical systems, but there are certainly soft matter systems that require going beyond such assumptions.

One reason that these simulation techniques have been so powerful in the study of soft matter is that they are readily parallelizable, allowing large simulations of many orders of magnitude in time to be performed on modern hardware. This is crucial, since many soft matter systems are characterized by widely separated time and length scales. For instance, a micron-scale colloid is orders of magnitude larger than the atoms in the solvent it interacts with, and the time scales of solvent motion are radically faster than the time scales of colloidal motion. Even in single-component systems this issue arises: in a melt of high-molecular-weight polymers — often represented by chains of repulsive spheres connected by stiff springs — the time scale of overall polymer motion can be much longer than the time scale associated with the vibration of adjacent units composing the polymer [8]. Thus, to faithfully integrate the equations of motion, soft matter simulations often involve very many, very short individual time steps.

In the previous Chapters, we learned that in a naive implementation of the fundamental pattern the algorithm above, the most computationally expensive piece is the calculation of the forces. In particular, even the evaluation of all pairwise distances between particles requires $\mathcal{O}(N^2)$ operations, and for many of the simulations we want to do N^2 is already too many. If the interparticle interactions are truly long-ranged, a variety of methods have been proposed that improve on the naive scaling, such as *Ewald summation*, particle-mesh methods, fast multipole methods [7].

In many soft matter systems, however, it is common to care about potentials that have a finite range. In this setting, we can radically reduce the computational cost of building lists of neighbors. For instance, before computing distances between particles, we can spatially partition the domain into a grid of *cells* and, during the force-computation step, only compute pairwise distances between points in nearby cells. If the simulation is of particles that all have roughly the same interaction range and are roughly evenly distributed throughout the domain, computing pairwise distances among particles that exert forces on each other becomes $\mathcal{O}(N)$. Not only is this version of the calculation computationally less expensive, it is also *embarrassingly parallel* (a nice feature given the current state of computational hardware). Note, by the way, that for simulating systems with

large disparities in particle sizes or densities other algorithms exist that similarly accelerate the simulation of short-ranged forces, such as using stenciled cell lists [2] or using a hierarchical tree-based structure to help partition the simulation domain rather than a grid of cells [9].

20.2 Won't you be my neighbor?

We spend so much of our time thinking about physical systems evolving according to conservative forces that this connection between *neighbor lists* — lists of interacting pairs of degrees of freedom — and the interactions according to which the degrees of freedom exert forces is often assumed without much further thought. As active matter increasingly looks to build highly coarse-grained models of living systems — at the level of cells or entire organisms, as shown in Fig. 20.1 — we should pause and question even these fundamental assumptions.

For example, in the case studies below we will discuss classic models of flocking birds, in which each *bird* tries to align with its neighbors. Looking at a flock, or thinking about the neural processes involved in birds deciding which way to turn, it is far from clear that the *neighbor list* specifying with which other birds a member of the flock is aligning should be tied to strict distance cutoffs. Similarly, looking at a dense packing of cells, it seems clear that cell *neighbors* should be essentially determined by whether two cells share an edge — an indirect function of distance, perhaps, but one which has the character of a graph or network of interactions more than a distance-based cutoff.

These are just two of a broader class of systems that we might say interact *topologically*, that is, in which we define neighbors via something other than a strict metric criterion. Generalizing our notion of neighbor relations takes us to the forefront of some very exciting current research in the soft active matter community, ranging from the new phenomena and scaling that can be found in such non-metric models [10, 11] to the role and importance of non-reciprocal forces [12, 13, 14, 15].

20.3 Metric and topological flocking models

The Vicsek model [16] is one of the canonical models in the study of active matter. Its original form studied out-of-equilibrium point particles evolving according to a simple dynamical rule, and showed that beautiful collective organizational states emerged. Many variations of the Vicsek model have been proposed and studied, but for our purposes we will focus on the simple *vectorial* version of the model [17].

The essence of the model is to posit the overdamped dynamics of N self-propelled particles. Each particle has a position, \mathbf{r}_i , and an orientation, $\hat{\mathbf{n}}_i$, and at each time step the positions of the particles are updated by moving at a constant speed, v_0 , in the direction of the particle's orientation. Computationally, in a simple Eulerian integration of the equations of motion, at each time step we update the position of particle i according to

$$\Delta\mathbf{r}_i = (v_0\Delta t) \hat{\mathbf{n}}_i(t). \quad (20.1)$$

So far, this is just an uninteresting gas of non-interacting particles, each moving at some velocity. The dynamical rule added to make this model both deeply fascinating and technically non-trivial is to say that particles tend to align their direction of motion with nearby particles. Thus, at each time step, we also update the orientation of particle i by

$$\hat{\mathbf{n}}_i(t + \Delta t) = \mathcal{N} \left(\eta \xi_i^t + \frac{1}{m_i^t} \sum_{\langle ij \rangle} \hat{\mathbf{n}}_j(t) \right). \quad (20.2)$$

In the above equation, η is a parameter that characterizes the strength of the noise in the model, ξ_i^t is a random unit vector delta-correlated in both time and particle index, the sum is over the m_i^t neighbors, j , that particle i has at time t , and \mathcal{N} is an operator that simply normalizes its argument. More simply: to determine a particles new direction of motion, we average the current direction of the particle's neighbors, add a bit of vectorial noise to this direction, and normalize the result so that $\hat{\mathbf{n}}_i$ stays a unit vector.

How, though, should we choose the set of neighbors with which each particle attempts to align at each time step? Early versions of the model proposed a straightforward and quite natural metric criterion: there is a distance scale associated with particle alignment, and so a parameter was added to the model. In a given

timestep, all pairs of particles within some distance, r_0 , would count as neighbors, and particles farther apart would not. This solution feels quite natural, and it also has the virtue of mapping perfectly onto some of the efficient computational techniques we have learned for dealing with finite-ranged interactions. With this choice, there is a high-noise regime in which the particles are in a disordered state, a low-noise regime in which there is a phase transition to a polar flocking state, and an intermediate regime in which a *banded* phase is observed, characterized by both overall polar order and periodic traveling density bands [17].

While this kind of distance threshold for alignment feels natural, *is it how birds actually align with each other?* A beautiful body of literature has attempted to infer simple, phenomenological interaction rules for flocking birds, shoaling fish, and other collectively moving groups of organisms [18, 19, 20], and a consistent theme seems to be that these natural metric thresholds are likely poor descriptions of the *social forces* that represent organismal-scale interactions in moving-group dynamics. Some brief self-reflection of your own behavior in, say, moving crowds of different densities might supply the intuition that this should probably not be a surprise.

For flocks of birds, the [StarFlag project](#) reconstructed full three-dimensional positions and trajectories of flocks of thousands of birds flying above a railway station in Rome. Statistical analyses of these trajectories revealed what your intuition perhaps suggested: a typical bird interacts not with all birds up to some distance, but rather with its six or seven *closest neighbors* — regardless of how far away those closest neighbors actually are [18, 21, 22].

We should be quite open, then, to interpreting the *sum over neighbors* in Eq. (20.2) without being restricted to the kinds of distance-based cutoffs we are so used to. What if, instead, we replaced the fixed-distance criteria with a rule that each particle interacted with its k nearest neighbors (with k a parameter)? Or the k nearest neighbors within some angular *field of view*? What about a rule that constructed some triangulation of the set of positions — say, a Delaunay triangulation — with neighbor relations chosen according to the set of edges in the triangulation? In all of these cases, the interactions are still *local*, but they are local in a topological rather than metric sense.

In pioneering work, Ginelli and Chaté showed that these changes *matter*, with implications (for instance) about the nature of the transition between the different dynamical phases of the model [10]. It is equally clear that the different topological criteria are *not equivalent to each other*. For instance, in a Delaunay triangulation the neighbor relation is still reciprocal: particles i and j share an edge, say, so each is a neighbor of the other. In contrast, k -nearest-neighbors need not be reciprocal: in the presence of density inhomogeneities particle i can be among the k nearest neighbors of particle j without j being among the k nearest neighbors of i . This opens up a universe of possibilities, with the effects of making these different choices for neighbor relations on the structural or dynamical phases of a model both interesting and poorly understood.

20.4 Voronoi neighbor lists

In our exploration of topological neighbor relations, we focus on Delaunay-based neighbor lists in this Chapter. They break away from strictly metric considerations, but they still maintain reciprocity of particle interactions; in this sense, they are a natural first generalization to consider. So, what is a *Delaunay triangulation* of a given set of points? One definition is that, for points in a two-dimensional plane, it is a triangulation that maximizes the minimum angle of all triangles in the triangulation (that is, the minimum angle in the Delaunay triangulation is at least as large as the minimum angle found in all possible other triangulations of the point set). Equivalently, the Delaunay triangulation can be defined as the triangulation with the *empty circumcircle* property: it is the triangulation such that no point in the point set is inside the circumcircle of any triangle in the triangulation (the circumcircle is simply the circle that passes through all of the vertices of a given polygon). Higher-dimensional versions of the Delaunay triangulation can be defined by the appropriate generalization of this to the empty circumsphere (or circum-hypersphere) property.

An example Delaunay triangulation, along with the (empty) circumcircles of each triangle, is shown in Fig. 20.2. The Delaunay triangulation is intimately related to the *Voronoi diagram* of a point set, which is a partition of a space into polygonal (or polyhedral, or appropriate higher-dimensional generalizations) regions that are closer to a given point than to any other point in the set. In fact, the Voronoi diagram and the Delaunay triangulation are dual to one another: the vertices of the Voronoi diagram correspond to the circumcenters of the Delaunay triangulation, and, if two triangles share an edge in the Delaunay triangulation, then the vertices

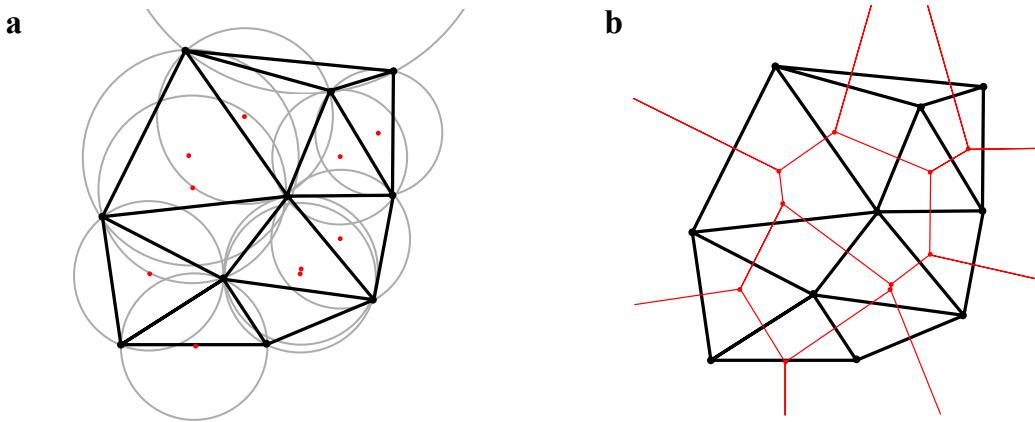


Figure 20.2: Delaunay triangulations and Voronoi diagrams. (a) Delaunay triangulation of a set of points (in black). Circumcircles corresponding to each triangle are shown, with circumcircle centers in red. (b) The Voronoi tessellation is dual to the Delaunay triangulation. Connecting the circumcircle centers of triangles that share an edge produces the Voronoi diagram of the original set of points.

corresponding to those circumcenters will be connected by an edge in the Voronoi diagram. This relationship is similarly shown in Fig. 20.2.

The dual relation between Delaunay triangulations and Voronoi diagrams can often be exploited in computational settings. A wide variety of algorithms have been created to construct both types of graph. If one has a Delaunay triangulation, it is a trivial ($\mathcal{O}(N)$) operation to extract the Voronoi diagram (and vice versa). For instance, if one is simulating a Vicsek-style model where at each timestep every particle aligns with any particle in the first shell of its Voronoi neighbors [10], it might be computationally beneficial to avoid recomputing the entire Voronoi diagram if no change in the diagram occurred as a result of the most recent update of positions. It might be computationally convenient to formulate this condition — *do I need to recompute the Voronoi diagram?* — as a test of the empty circumcircle property of the Delaunay triangulation.

These computational and algorithmic considerations can be quite important. Most MD algorithms exploit the easily parallelizable character of particles evolving according to short-ranged interactions to accelerate simulations (across multiple CPU cores, or using GPUs or TPUs), but most efficient algorithms for computing the edges and faces of Delaunay triangulations and Voronoi diagrams are fundamentally serial ones. In some ways, this is natural: even though, e.g., the Voronoi diagram can be thought of as a basically local characterization of the area around each point in a set, it must still obey global consistency conditions (e.g., in 2D the graph of the Voronoi diagram must obey Euler's formula for plane graphs).

This makes it more difficult to build efficient MD simulations in systems where neighbor relations are determined by such a criteria. One option is to have a subroutine in an otherwise highly parallelized simulation that makes a call to an efficient algorithm to construct the Voronoi diagram (or Delaunay triangulation) of the entire set of particle positions in the simulation. Although scripting languages are not typically well-suited to large-scale simulations, existing packages can provide this functionality in Python, Matlab, Mathematica, and similar environments. On the high-performance side, a wrapper to quality open-source libraries that have been optimized for geometric calculations — like the Computational Geometry Algorithms Library (CGAL) [23, 24] and Voro++ — can be incorporated into a MD simulation software and called when needed. Recently, some work has gone into developing MD software packages that incorporate these sorts of tessellations at a fundamental level, with a high level of parallelism exploited even in the triangulation protocol [25]. To date, most of these efforts have centered around efficiently simulating specific models of dense tissue, which we discuss next.

20.5 Geometric models of dense tissue

A natural setting in which this sort of tessellation-based neighbor list arises is in modeling dense, confluent tissue — tissues in which there are no gaps between cells. There is a rich history of looking at images of the sort seen in Fig. 20.1 and taking the idea of *cells as shapes* seriously, representing the confluent tissue by polygonal or polyhedral tilings of space [26, 27, 28, 29, 30, 31, 32, 33]. These fundamentally geometrical models of cells have fascinating properties — in some cases, they display the sort of behavior one might expect given roughly *any* reasonably-biologically-informed agent-based model in which each cell corresponds to just one or a handful of degrees of freedom, but in other cases they can be shown to support exotic mechanical states [32, 34, 35] and display unusual structural or dynamical scaling [36, 37]. To investigate these unusual properties, in recent years there have been substantial advances in the efficient numerical simulation of these and related agent-based models of dense tissue [38, 25, 39, 40, 41].

To take a specific example, we consider using a computational package to generate the Voronoi diagram for a set of N points in the plane. In the context of the sort of Voronoi-based neighbor list discussed above, it is natural to assign each *Voronoi cell* to its generating point: the location of the points will be the degrees of freedom of the model, and each point will have a particular geometric region associated with it at every point in time. Here, we are thinking of a *Voronoi model* in the context of dense tissue, but related models have been used to also study foams, geological formation processes, and the partitioning of animal territorial domains [42]. We proceed, as promised, to take the tessellation seriously and consider an energy functional that depends explicitly on the geometry of the tessellation:

$$E = \sum_i^N (K_A(A_i - A_0)^2 + K_P(P_i - P_0)^2). \quad (20.3)$$

Here, A_i and P_i are the area and perimeter of cell i , A_0 and P_0 are *preferred* values of the area and perimeter for cell i , and K_A and K_P are area and perimeter moduli. In the context of dense tissue, the quadratic dependence on cell area can be interpreted as a cell monolayer's resistance to height fluctuations due to adhesions between cells and cell incompressibility, and the quadratic dependence on cell perimeter as a competition between active contractility of the actomyosin sub-cellular cortex and tension due to both cell-cell adhesions and cortical tension [36]. More generally, one might view this as a subset of the possible low-order Taylor expansions that can be written in the various geometric properties of the tessellation. Indeed, in the absence of any area terms precisely this interpretation has been given, where the terms non-linear in perimeter are what separate out foam-like energy functionals from tissue-like energy functionals [43].

Since we will be simulating this model on a computer, it is convenient to non-dimensionalize the energy by choosing the unit of length to be $\sqrt{\langle A \rangle}$, where $\langle A \rangle$ is the average area of all of the cells in our simulation, letting $k_r = K_A \langle A \rangle / K_P$, and writing

$$\frac{E}{K_P \langle A \rangle} \equiv e = \sum_i k_r (a_i - 1)^2 + (p_i - p_0)^2, \quad (20.4)$$

where a and p refer to dimensionless areas and perimeters. This non-dimensionalization differs from Eq. (20.3) only by an overall constant (assuming a finite domain that is completely occupied by cells), and it makes clear that the density-dependence of the model enters via $p_0 = P_0 / \sqrt{\langle A \rangle}$ — an important point given the early interest of this model as one in which the jamming transition was one controlled by preferred shape rather than strict density considerations [44, 37, 45].

Now that we have a set of degrees of freedom (again, here, the location of points in 2D space), a rule for finding neighbors, and an energy functional (which depends on that neighbor-finding rule), we have all of the ingredients necessary to perform agent-based simulations. One could perform energy minimizations to find tissue ground states, use classic molecular dynamics algorithms to simulate them in NVE or NVT ensemble, or choose from any of a number of equations of motion corresponding to non-equilibrium, active-matter scenarios (e.g., self-propelled particle dynamics, self-propulsion with noise, chemotaxis) — the world is our oyster. All that is needed is an efficient way of taking the appropriate gradients of this energy functional in order to compute sets of forces acting on our degrees of freedom. For completeness, we show two variations of this calculation.

20.5.1 Vertex model forces

First, suppose we had a simpler version of our geometric model. Rather than taking a tessellation of a point set, suppose our degrees of freedom were actually the *vertices* of a set of polygons tiling the plane. This sort of model — often, unsurprisingly, referred to as a *vertex model* — has a similar flavor to Voronoi-based models, in that once again the neighbor list of our degrees of freedom comes *not* from strictly distance-based considerations, but rather stems from a graph structure (i.e., from an explicit enumeration of vertex-vertex connections). Given such an enumeration, computing the forces on the vertices given the energy functional in Eq. (20.3) is straightforward.

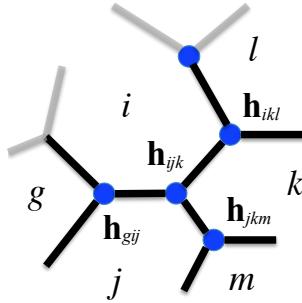


Figure 20.3: **Two-dimensional tissue model.** Schematic diagram of cell i and some of its neighbors, along with the associated Voronoi vertices (the circumcenters of three adjacent cells) labeled for convenient reference.

For reference, Fig. 20.3 provides a schematic picture of a relevant patch of a two-dimensional tissue model, with each vertex labeled by the three cells it is adjacent to. From here, computing the forces is simply an exercise in patience and application of the appropriate propagation of the chain rule. From the figure, since the motion of vertex \mathbf{h}_{ijk} only changes the shape of cells i , j , and k , the force on vertex \mathbf{h}_{ijk} is

$$-\frac{\partial E}{\partial \mathbf{h}_{ijk}} = -\left(\frac{\partial E_i}{\partial \mathbf{h}_{ijk}} + \frac{\partial E_j}{\partial \mathbf{h}_{ijk}} + \frac{\partial E_k}{\partial \mathbf{h}_{ijk}} \right). \quad (20.5)$$

Each of the cell-specific energy derivatives are straightforward, e.g.,

$$\frac{\partial E_i}{\partial \mathbf{h}_{ijk}} = 2K_A(A_i - A_0) \frac{\partial A_i}{\partial \mathbf{h}_{ijk}} + 2K_P(P_i - P_0) \frac{\partial P_i}{\partial \mathbf{h}_{ijk}}, \quad (20.6)$$

where the area- and perimeter-derivatives with respect to vertex positions can themselves be written as follows. Let $\mathbf{t}_{ij} = \mathbf{h}_{ijk} - \mathbf{h}_{gij}$, $\mathbf{t}_{ik} = \mathbf{h}_{ikl} - \mathbf{h}_{ijk}$, with $\hat{\mathbf{t}}_{ij}$ and $\hat{\mathbf{t}}_{ik}$ being the unit vectors in those directions. Similarly, let l_{ij} be the length of the edge between cell i and j , and $\hat{\mathbf{n}}_{ij}$ be the unit vector pointing outwardly normal to that cell edge. Then, one can write

$$\frac{\partial A_i}{\partial \mathbf{h}_{ijk}} = \frac{1}{2} (l_{ij} \hat{\mathbf{n}}_{ij} + l_{ik} \hat{\mathbf{n}}_{ik}), \quad (20.7)$$

$$\frac{\partial P_i}{\partial \mathbf{h}_{ijk}} = -(\hat{\mathbf{t}}_{ij} + \hat{\mathbf{t}}_{ik}). \quad (20.8)$$

20.5.2 Voronoi model forces

Returning to the class of models where the degrees of freedom are the positions of a point set we have tessellated, the same approach to computing forces can be used. Of course, the positions of the vertices of the Voronoi tessellation can be explicitly computed from the original point set, and so ultimately we are simply forced to carry out an extra layer of chain-rule calculations. Let us see this explicitly.

Following the notation in Bi *et al.* [36], the force on cell i in Cartesian direction λ (indicated with $r_{i\lambda}$) can be computed as

$$F_{i\lambda} = -\frac{\partial E}{\partial r_{i\lambda}} = -\frac{\partial E_i}{\partial r_{i\lambda}} - \sum_{\langle ij \rangle} \frac{\partial E_j}{\partial r_{i\lambda}}. \quad (20.9)$$

The motion of cell i changes the geometry, and hence the energy, of cell i itself and all of its $\langle ij \rangle$ neighbors, j . Again, using the configuration in Fig. 20.3, all of these terms can be expanded via the chain rule, for instance:

$$\frac{\partial E_k}{\partial r_{i\lambda}} = \sum_v \left(\frac{\partial E_k}{\partial h_{ijk,v}} \frac{\partial h_{ijk,v}}{\partial r_{i\lambda}} + \frac{\partial E_k}{\partial h_{ikl,v}} \frac{\partial h_{ikl,v}}{\partial r_{i\lambda}} \right). \quad (20.10)$$

Here, these are the only terms needed, since the other Voronoi vertices associated with cell k (the middle of the three neighboring cells in clockwise order) do not depend on the position of cell i . The partial derivatives depend on the positions of \mathbf{h}_{jkm} and \mathbf{h}_{klm} , where m is the cell other than i that has both j and k as neighbors, and n is the cell other than i that has both k and l as neighbors (for this energy functional there is a dependence on the nearest and some of the next-nearest neighbors of cell i).

The derivative of the energy with respect to the vertices was calculated in the previous subsection, so the only additional piece that is needed are the derivatives of the Voronoi tessellation's vertices with respect to the position of the cell, e.g., $(\partial \mathbf{h}_{ijk}) / (\partial \mathbf{r}_i)$. These derivatives can be calculated efficiently as follows. We let \mathbf{r}_{ij} denote the vector from i to j , and define the auxiliary quantities

$$c = \mathbf{r}_{ij,x} \mathbf{r}_{kj,y} - \mathbf{r}_{ij,y} \mathbf{r}_{kj,x}, \quad (20.11)$$

$$d = 2c^2, \quad (20.12)$$

$$\mathbf{z} = \beta d \mathbf{r}_{ij} + \gamma d \mathbf{r}_{ik}, \quad (20.13)$$

$$\beta d = -|\mathbf{r}_{ik}|^2 \cdot (\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}), \quad (20.14)$$

$$\gamma d = |\mathbf{r}_{ij}|^2 \cdot (\mathbf{r}_{ik} \cdot \mathbf{r}_{jk}), \quad (20.15)$$

with I_2 representing the 2×2 identity matrix and \otimes the dyadic product, the desired change in Voronoi vertex position with respect to cell position is

$$\frac{\partial \mathbf{h}_{ijk}}{\partial \mathbf{r}_i} = I_2 + \frac{1}{d} \left[\mathbf{r}_{ij} \otimes \left(\frac{\partial(\beta d)}{\partial \mathbf{r}_i} \right) + \mathbf{r}_{ik} \otimes \left(\frac{\partial(\gamma d)}{\partial \mathbf{r}_i} \right) - (\beta d + \gamma d) I_2 - \mathbf{z} \otimes \left(\frac{1}{d} \frac{\partial d}{\partial \mathbf{r}_i} \right) \right], \quad (20.16)$$

where

$$\frac{\partial(\beta d)}{\partial \mathbf{r}_i} = 2(\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}) \mathbf{r}_{ik} + |\mathbf{r}_{ik}|^2 \mathbf{r}_{jk}, \quad (20.17)$$

$$\frac{\partial(\gamma d)}{\partial \mathbf{r}_i} = -2(\mathbf{r}_{ik} \cdot \mathbf{r}_{jk}) \mathbf{r}_{ij} + |\mathbf{r}_{ij}|^2 \mathbf{r}_{jk}, \quad (20.18)$$

and

$$\frac{1}{d} \frac{\partial d}{\partial \mathbf{r}_i} = \frac{2}{c} (-\mathbf{r}_{jk,y} \hat{x} + \mathbf{r}_{jk,x} \hat{y}). \quad (20.19)$$

20.6 Discussion and outlook

The point of the explicit calculation is not simply to recapitulate expressions that have appeared previously in the literature for the efficient calculation of vertex- and Voronoi-model forces [36, 25], but to again emphasize the common character of these models with other agent-based simulations in the study of equilibrium and active matter. In this instance, the energy of a tissue configuration is dependent on the shape of the tessellated regions, and so in our computational work we need additional subroutines and data structures to calculate and keep track of these unusual sets of neighbor relations. Everything else, though, fits into the usual paradigm of a molecular dynamics simulation.

In the context of dense models of tissues, this perspective of starting with an unusual neighbor list and taking its implication seriously suggests a class of models with unusual behavior — in the way they can support stresses [32], in their *low-temperature* properties [46, 47], in the way they can organize into collectively moving states [48], in the way they support *curvotaxis* [49], and so on. In the context of active models with less complex energy functionals, it is known that this *topological* way of choosing your neighbors changes the nature of the flocking transition [10]. Beyond this, though, lies an expanse of open questions.

20.7 Problems

Problem 20.1: Simulating cells. Using existing packages to perform the Voronoi tesselation (Voro++ for C++, or `scipy.spatial.Voronoi` in python), implement the gradients described in the *Voronoi model forces* section above, and use them to perform dynamical simulations in the NVE ensemble.

- a. How does the mean square displacement of cells vary as p_0 is varied?

Problem 20.2: Energy minima of cellular configurations. Again implementing a model using the Voronoi model forces from above, use gradient descent (or any other minimization algorithm) to sample the energy minima when initializing the system with randomly placed points. There is a (potential) rigidity transition in the model when $p_0 \approx 3.81\dots$, a value which corresponds to the perimeter of a unit-area regular pentagon.

- a. Is there a corresponding transition that you observe in the energy minima? I.e., do you find a transition from ground states of zero to ground states of finite energy?
- b. What if you use not the Voronoi but the vertex model of cells?

Bibliography

- [1] J. Devany, D. M. Sussman, T. Yamamoto, M. L. Manning, and M. L. Gardel. Cell cycle-dependent active stress drives epithelia remodeling. *Proceedings of the National Academy of Sciences*, 118:e1917853118, 2021.
- [2] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117:1–19, 1995.
- [3] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.
- [4] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26:1701–1718, 2005.
- [5] J. A. Anderson, C. D. Lorenz, and A. Traverset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics*, 227:5342–5359, 2008.
- [6] S. S. Schoenholz and E. D. Cubuk. Jax, m.d.: End-to-end differentiable, hardware accelerated, molecular dynamics in pure python. *arXiv:1912.04232*, 2019.
- [7] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.
- [8] M. Rubinstein and R. H. Colby. *Polymer physics*, volume 23. Oxford University Press New York, 2003.
- [9] M. P. Howard, A. Statt, F. Madutsa, T. M. Truskett, and A. Z. Panagiotopoulos. Quantized bounding volume hierarchies for neighbor search in molecular simulations on graphics processing units. *Computational Materials Science*, 164:139–146, 2019.
- [10] F. Ginelli and H. Chaté. Relevance of metric-free interactions in flocking phenomena. *Physical Review Letters*, 105:168103, 2010.
- [11] D. M. Sussman, J. M. Schwarz, M. C. Marchetti, and M. L. Manning. Soft yet sharp interfaces in a vertex model of confluent tissue. *Physical Review Letters*, 120:058001, 2018.
- [12] Z. You, A. Baskaran, and M. C. Marchetti. Nonreciprocity as a generic route to traveling states. *Proceedings of the National Academy of Sciences*, 117:19767–19772, 2020.
- [13] L. P. Dadhichi, J. Kethapelli, R. Chajwa, S. Ramaswamy, and A. Maitra. Nonmutual torques and the unimportance of motility for long-range order in two-dimensional flocks. *Physical Review E*, 101:052601, 2020.

- [14] A. V. Ivlev, J. Barwick, M. Heinen, C.-R. Du, V. Nosenko, and H. Löwen. Statistical mechanics where Newton's third law is broken. *Physical Review X*, 5:011035, 2015.
- [15] E. A. Lisin, O. F. Petrov, E. A. Sametov, O. S. Vaulina, K. B. Statsenko, M. M. Vasiliev, J. Carmona-Reyes, and T. W. Hyde. Experimental study of the nonreciprocal effective interactions between microparticles in an anisotropic plasma. *Scientific Reports*, 10:1–12, 2020.
- [16] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75:1226, 1995.
- [17] F. Ginelli. The physics of the Vicsek model. *The European Physical Journal Special Topics*, 225:2099–2117, 2016.
- [18] Andrea Cavagna, Alessio Cimarelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences*, 107(26):11865–11870, 2010.
- [19] J. E. Herbert-Read, A. Perna, R. P. Mann, T. M. Schaerf, D. J. T. Sumpter, and A. J. W. Ward. Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*, 108:18726–18731, 2011.
- [20] J. Gautrais, F. Ginelli, R. Fournier, S. Blanco, M. Soria, H. Chaté, and G. Theraulaz. Deciphering interactions in moving animal groups. *PLoS Computational Biology*, 2012.
- [21] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak. Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences*, 109:4786–4791, 2012.
- [22] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, T. S. Grigera, A. Jelić, S. Melillo, L. Parisi, O. Pohl, E. Shen, and M. Viale. Information transfer and behavioural inertia in starling flocks. *Nature Physics*, 10:691–696, 2014.
- [23] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 5.0 edition, 2019.
- [24] S. Hert and S. Schirra. 3D convex hulls. In *CGAL User and Reference Manual*. CGAL Editorial Board, 5.0 edition, 2019.
- [25] D. M. Sussman. cellGPU: massively parallel simulations of dynamic vertex models. *Computational Physics Communications*, 219:400, 2017.
- [26] H. Honda. Description of cellular patterns by Dirichlet domains: The two-dimensional case. *Journal of Theoretical Biology*, 72:523–530, 1978.
- [27] Tatsuzo N. and Hisao H. A dynamic cell model for the formation of epithelial tissues. *Philosophical Magazine Part B*, 81:699–719, 2001.
- [28] L. Hufnagel, A. A. Teleman, H. Rouault, S. M. Cohen, and B. I. Shraiman. On the mechanism of wing size determination in fly development. *The Proceedings of the National Academy of Sciences*, 104:3835–3840, 2007.
- [29] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17:2095–2104, 2007.
- [30] L. Yan and D. Bi. Multicellular rosettes drive fluid-solid transition in epithelial tissues. *Physical Review X*, 9:011029, 2019.
- [31] S. Alt, P. Ganguly, and G. Salbreux. Vertex models: from cell mechanics to tissue morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372:20150520, 2017.
- [32] N. Noll, M. Mani, I. Heemskerk, S. J. Streichan, and B. I. Shraiman. Active tension network model suggests an exotic mechanical state realized in epithelial tissues. *Nature Physics*, 13:1221, 2017.

- [33] M. Popovic, V. Druelle, N. Dye, F. Jülicher, and M. Wyart. Inferring the flow properties of epithelial tissues from their geometry. *New Journal of Physics*, 2020.
- [34] M. Moshe, M. J. Bowick, and M. C. Marchetti. Geometric frustration and solid-solid transitions in model 2d tissue. *Physical Review Letters*, 120:268105, 2018.
- [35] R. Kupferman, B. Maman, and M. Moshe. Continuum mechanics of a cellular tissue model. *Journal of the Mechanics and Physics of Solids*, 143:104085, 2020.
- [36] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning. Motility-driven glass and jamming transitions in biological tissues. *Physical Review X*, 6:021011, 2016.
- [37] E. Teomy, D. A. Kessler, and H. Levine. Confluent and nonconfluent phases in a model of cell tissue. *Physical Review E*, 98:042418, 2018.
- [38] P. Van Liedekerke, M. M. Palm, N. Jagiella, and D. Drasdo. Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Computational Particle Mechanics*, 2:401–444, 2015.
- [39] D. L. Barton, S. Henkes, C. J. Weijer, and R. Sknepnek. Active Vertex Model for cell-resolution description of epithelial tissue mechanics. *PLoS Computational Biology*, 13:e1005569, 2017.
- [40] A. G. Fletcher, M. Osterfield, R. E. Baker, and S. Y. Shvartsman. Vertex models of epithelial morphogenesis. *Biophysics Journal*, 106:2291–2304, 2014.
- [41] A. Boromand, A. Signoriello, F. Ye, C. S. O’Hern, and M. D. Shattuck. Jamming of deformable polygons. *Physical Review Letters*, 121:248003, 2018.
- [42] D. Weaire and N. Rivier. Soap, cells and statistics — random patterns in two dimensions. *Contemporary Physics*, 25:59–99, 1984.
- [43] S. Kim, Y. Wang, and S. Hilgenfeldt. Universal features of metastable state energies in cellular matter. *Physical review letters*, 120:248001, 2018.
- [44] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning. A density-independent rigidity transition in biological tissues. *Nature Physics*, 11:1074–1079, 2015.
- [45] M. Merkel and M. L. Manning. A geometrically controlled rigidity transition in a model for confluent 3D tissues. *New Journal of Physics*, 20:022002, 2018.
- [46] D. M. Sussman, M. Paoluzzi, M. C. Marchetti, and M. L. Manning. Anomalous glassy dynamics in simple models of dense biological tissue. *EPL (Europhysics Letters)*, 121(3):36001, 2018.
- [47] D. M. Sussman and M. Merkel. No unjamming transition in a Voronoi model of biological tissue. *Soft matter*, 14:3397–3403, 2018.
- [48] F. Giavazzi, M. Paoluzzi, M. Macchi, D. Bi, G. Scita, M. L. Manning, R. Cerbino, and M. C. Marchetti. Flocking transitions in confluent tissues. *Soft Matter*, 14:3471–3477, 2018.
- [49] D. M. Sussman. Interplay of curvature and rigidity in shape-based models of confluent tissue. *Physical Review of Research*, 2:023417, 2020.

Chapter 21

Mesoscopic Modelling of Epithelial Tissues

KEVIN HÖLLRING, ANA-SUNČANA SMITH

Over the last two decades, the scientific literature has been blooming with various means of simulating epithelial cell colonies. Each of these simulations can be separated by their respective efficiency (expressed in terms of consumed computational resources), the amount of cells/the size of tissues that can be simulated, the time scale of the simulated dynamics, and the coarse grained level of precision. Choosing the right algorithm for the simulation of epithelial cells and tissues is a compromise between these key elements. Irrespective of the method, each algorithm includes part, or all, of the following features: short-range membrane-mediated attraction between cells, soft-core repulsion between cells, cell proliferation, cell death, cell motility, and fluctuations. In this Chapter, we will first give a non-exhaustive overview of commonly used modeling approaches for tissues at a mesoscopic level, giving a rough idea of the coarse-graining decisions made for every one of them. Then, we will dive into greater detail on how to implement a relaxation procedure according to the *vertex model*, refreshing aspects of the theoretical groundwork, describing required data structures and simulation steps, and pointing out details of the simulation that can present pitfalls.

21.1 Modeling tissues at various levels of granularity

When it comes to the representation of cells and tissues, there have been various, substantially different approaches trying to capture dynamics at different levels of detail [1]. Of course, there is the *molecular level*, where the objects of interest are proteins, their internal dynamics, and their interactions with other parts of the system. This level of detail can be tackled with the already established methods of molecular dynamics simulations and other atomistic approaches (see Chapters 16, 18, and ??).

At the other end of the scale, there are the *macroscopic models*, where a tissue is described via its material properties like its Young's modulus and its viscosity. The latter can be exploited because large colonies of cells behave like viscous fluids, opening the door to a more classical approach, where known techniques from material sciences and fluid dynamics can be applied to abstract away the processes that are relevant on an atomistic nano-scale.

In this Chapter, we will focus on the description of tissues on the *mesoscale*, where the overall scale of the system is still comparable to that of an individual cell but the dynamics concern more than just individual cells. In Figure. 21.2, we provide a visual overview for a certain set of cell models on this scale located between the sub-cellular and super-cellular scale in terms of their measures for capturing the dynamics of the individual actors of their systems.

Figure 21.1: **Introductory Figure Title.** Introductory Figure Caption.

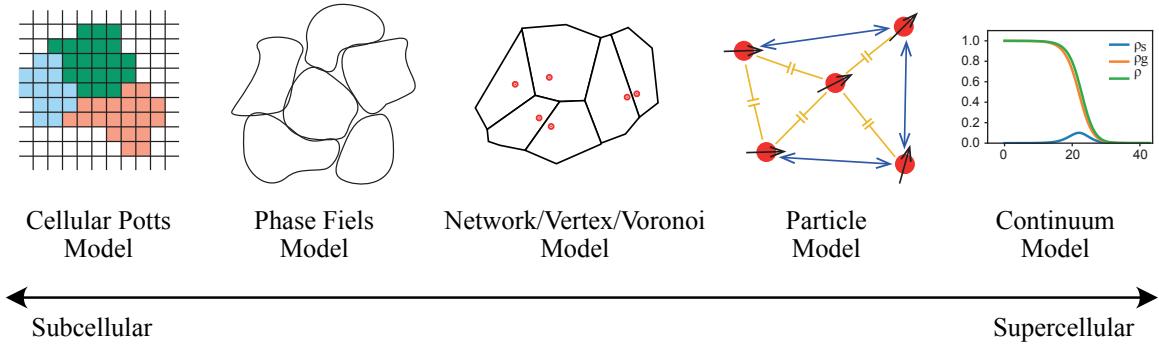


Figure 21.2: Overview of cell models at different levels of detail and granularity. Starting from the *cellular Potts model* and the *phase field model* at high level of detail, the *active network models* take cell descriptions towards a more abstract representation of shapes with fewer degrees of freedom. The *particle models* do not consider cell shape at all, only accounting for pairwise interaction forces of point-representations with the *continuum models* finally abstracting away the individual cell altogether.

21.1.1 Cellular Potts models

On the sub-cellular end of the scale, there are the *cellular Potts models*, which build on the techniques employed in Ising models to describe cell dynamics on a discrete grid [2]. A grid-cell can be assigned to a cell and a cell is represented by the entirety of grid-cells assigned to it. This approach allows for the investigation of how cells change their shape depending on parameters of interfacial cell–cell interaction in competition with internal cell processes. Further degrees of detail concerning cell polarity [3] as well as interaction with the medium can be modeled using an appropriate Hamilton operator with the full time-evolution dynamics usually being simulated employing well-known Metropolitan-algorithm techniques.

21.1.2 Phase field models

Contrary to the grid-based approach of cellular Potts models, the *phase field models* use a description of individual cells via a phase field ϕ_i , which takes on a value of 1 inside and 0 outside of cell [4]. Between these two domains lies the gradual transition corresponding to the interface or membrane of each individual cell, which can also be represented as a contour function [5]. Combining the different phase-field functions for all cells in the system, one can again combine the overall interaction and internal energies into a Hamiltonian and simulate the dynamics of the combined tissue via classical numerical techniques. This approach introduces more degrees of freedom to each individual cell than the cellular Potts model with the phase fields potentially overlapping and the grid used for the numerical simulations being more fine-grained than that employed in the cellular Potts model, but effectively both of these approaches work on a similar level of simulation.

21.1.3 Active network/vertex/Voronoi models

The class of *network models* lies in the middle of our scale of cell detail with increasingly many simplifications and abstractions being introduced. In these models, one tries to describe cell dynamics in terms of network interactions either by employing a simplified representation of a cell as a polygon representing the cell membrane and using the vertices of the polygons as the degrees of freedom (*vertex models*) or by representing the cell by one or two points (to model cell anisotropy/polarity [6]) from which the cell shape and tissue topology can then be reconstructed (*Voronoi models*) [7]. The interactions between cells in these models are described in terms of the aforementioned degrees of freedom like cell–cell adhesion and membrane contractility, with certain shape aspects still being retained, while the degrees of freedom per cell are reduced drastically compared to the prior two families of models. As our own group has been able to show, the Voronoi tesselation used to reconstruct the cell shape and tissue topology from the simplified description of cells via particles is able to capture key

topological and morphological aspects of real tissues, making this approach to simplification justified [8]. Additionally, Voronoi models have been extended to include cell polarity and directed active motion to elevate the investigation above that of simply passive tissues to being able to describe collective cell phenomena [9, 10].

21.1.4 Particle models

The trend towards a higher level of abstraction continues with the *particle models*, where cells are represented by point particles with different kinds of interactions (i.e., adhesion forces at short distances and repulsion in the medium distance regime) [11]. This approach is known from the material science of granular matter and allows for the simulation of much larger systems in terms of size with respect to the number of cells being simulated as we need to account for fewer degrees of freedom per cell. This class of models is the first where we see a loss of information concerning cell shape, because they are often resorting to cell representations as finite-size disks or point particles. Not only have we simplified the shapes that we allow for cells to take (as in the active network class of models) but instead we have completely abstracted that aspect of the cells away, with no clear representation of the interface being formed by the deformable membrane and only general pairwise interaction forces still being considered to describe cell–cell interactions. The strong simplification of individual particles allows for larger system sizes and also larger collective phenomena like flocking, velocity alignment and vortex formation to be investigated [6, 12].

21.1.5 Continuum models

As the ultimate consequence of the trend towards abstracting away the degrees of freedom of the individual cells, one eventually arrives at *continuum models*, where the individual cell is no longer modeled directly [13]. Instead, it is replaced by a density distribution describing the distribution and/or a flow field representing movement of cells in the system, which can then be supplemented by various energies and forces being experienced by the cell colony, which allow for predictions of overall colony or tissue evolution. The description of whole cells via a density distribution is similar to the approach in the phase field models, where the individual parts of a cell were abstracted away into a density distribution/phase field describing the positioning of the cell in the system. The continuum models then reach the next level of abstraction with the cells themselves being the parts no longer being considered individually but as an effective agglomerate whose dynamics are to be investigated.

21.1.6 Bridging scales with adaptive numerical simulations

Summing up the different approaches to tissue simulation at a mesoscopic level, one faces a trade-off forcing us to choose between the degrees of freedom for each cell or coarse-grained descriptions, the complexity and level of detail of interactions or abstracted effective interaction forces all while considering the need for computational resources to simulate the desired dynamics **Can you rephrase the previous sentence? at the moment it's too complex and unclear**. This makes it hard to relate low-level cell properties to the features of large scale tissues. Since tissue topology and cell morphology play an important role in the mechanical properties and collective behavior of living tissues [7], we want to put an emphasis here on a model that still retains this level of information especially since the effect of membrane dynamics on tissue properties is often overlooked at higher levels of abstraction, e.g., in the vertex model. This model allows us to bridge sub-cellular and super-cellular scales, from cell level dynamics to tissue morphogenesis and growth.

21.2 Vertex model of tissues

Just like many other models for describing multi-cell dynamics, the *vertex model* can in principle be used to describe expanding tissues. Instead, due to its more common application to structural optimization in tissues, we will focus on that approach here and lay out its application in the scenario of tissue relaxation processes.

Compared to other models, the vertex model is a rather young introduction to the field of biophysics. Starting in 2007, the vertex model in its most common form has been introduced as a good model to describe the structural properties of a tissue caused by its membrane dynamics through apical junctions [14]. This is a departure from other models trying to model the cell-level tissue dynamics, like the *dissipative particle dynamics* (DPD) approach [15], which mainly focuses on the idea of the tissue dynamics being governed by

the cells' nuclei. There, the membranes were simply an intermediate structure reconstructed from the nuclei positions to determine which cells neighbor each other, and its dynamics can be simplified to the representation via two points for each cell. In years following the introduction of the vertex model, extensive theoretical and in-silico work on the vertex model by the group of Lisa Manning has shown that it can predict a certain type of rigidity phase transition in tissues in terms of only one effective parameter, the *isoperimetric ratio* $I = P_0/\sqrt{A_0}$ defined via the cells' target area A_0 and target perimeter P_0 [16]. A plethora of different analyses and implementations has emerged since then, both for expanding tissues as well as for the optimization and relaxation of tissues with a fixed number of cells [17, 1].

We will focus on a very basic description of an optimization routine for the basic vertex model in two dimensions, which can easily be extended to account for even more constraints and interactions as well as higher spatial dimensions, or could be adapted to allow for the simulation of the time-dependent growth of tissues.

21.2.1 The physical foundation of the vertex model as an energy functional

In the vertex model, one considers a tissue to be a collection of cells. Each cell is here represented by a shape with a well-defined area A and perimeter P with the latter being a model description for the cell membrane confining the cytoplasm spread across the cell area. One does not usually keep track of a representation of the cell nucleus within this shape, especially not in the most simple form of the model. Depending on how complex the description of each cell is supposed to be, one has many options to describe its shape, but the simplest one is certainly the representation via a polygon — an oriented collection of points (or vertices) with edges in between defining the position of the cell's membrane. Then, from the position of the vertices, both the perimeter and the area of the cell can be calculated.

Based on the area A and the perimeter P of a cell, the vertex model assigns a certain energy to its configuration that is intended to describe a set of counteracting processes favoring either an extended or reduced length of the cell membrane. This energy functional can be written (in a simplified form, neglecting a constant coefficient) as

$$E_{\text{cell}}(A, P) = (A - A_0)^2 + \frac{(P - P_0)^2}{r}. \quad (\text{see Chapter 14 for detailed explanation}) \quad (21.1)$$

In this, A_0 describes a certain target size the cell is trying to attain and P_0 is an effective target perimeter arising from counteracting forces of cell-cell-adhesion and contractile forces within the cell membrane of each individual cell [18]. Additionally, r is a constant controlling the relative magnitude of the energies associated with cell size and cell perimeter.

The vertex model does not need to reconstruct cell shapes and neighbor relations from simplified representations but instead it keeps track of the tissue topology via shared vertices between neighboring cells. Overall, each vertex in the simplest model of a confluent tissue (i.e., fully covering the surface) has three adjacent cells and is affected by their combined dynamics. Let us by \mathbf{r}_i denote the position of vertex i and by ∇_i the spatial gradient according to that position. The detailed forces $F_{i,k}$ acting on a vertex i within a cell k can easily be calculated from 21.1 by calculating the energy-gradient of that cell:

$$\mathbf{F}_{i,k} = \nabla_i E_{\text{cell}}(A_k, P_k) = 2(A_k - A_0)\nabla_i A_k + \frac{2(P_k - P_0)\nabla_i P_k}{r}. \quad (21.2)$$

As both the area A and the perimeter P can be expressed analytically in terms of the positions of the vertices in the cell k , we are able to derive a precise analytic formula to link the positions \mathbf{r}_i of vertices to the forces experienced by each vertex (see Example 21.1). The total force acting on a vertex is then obtained by summing up the contribution from all its neighboring cells.

Once the energy gradients (or forces) at each vertex are known, we can then proceed to try and find local and global minima in the energy landscape of a tissue via gradient descent.

Example 21.1: Analytics of the energy functional. We want to develop the analytical expression for the gradient of the area $\nabla_i A$ and the gradient of the perimeter $\nabla_i P$ of a cell required for the analytical expression in Eq. (21.2).

Let \mathbf{r}_i , \mathbf{r}_{i+1} , and \mathbf{r}_{i+2} be the positions of the vertices of a triangle. Then, $\mathbf{a} = (\mathbf{r}_{i+1} - \mathbf{r}_i)$ and $\mathbf{b} =$

$(\mathbf{r}_{i+2} - \mathbf{r}_i)$ are the vectors representing the sides of the triangle starting from the vertex \mathbf{r}_i . Then, the oriented area of the triangle can be calculated via the following two-dimensional cross-product:

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = a_1 \cdot b_2 - a_2 \cdot b_1. \quad (21.3)$$

The area is given by $A = (\mathbf{r}_{i+1} - \mathbf{r}_i) \times (\mathbf{r}_{i+2} - \mathbf{r}_i)$. Here, the sign of the area depends on the orientation of the three vertices (i.e., clockwise or counterclockwise).

- a. Use the oriented formula for the area above to find an expression for the full area A (up to a correction of the sign) of a polygon with vertices at $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$. [Hint: Starting from one point, cut the polygon in triangles and apply the formula for the triangle.]
- b. Use the formula for the area of a full polygon to calculate $\nabla_i A$. [Hint: Consider how best to slice the polygon into triangles such that only very few triangles involve the point \mathbf{r}_i .]
- c. The length of an edge $l_{i,i+1}$ can be calculated via $l_{i,i+1} = |\mathbf{r}_{i+1} - \mathbf{r}_i| = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$. Use this formula to derive a sum expression for the full polygon perimeter P .
- d. Use the formula obtained in the previous point to calculate the gradients $\nabla_i l_{i,i+1}$ and $\nabla_{i+1} l_{i,i+1}$.
- e. Derive the formula for $\nabla_i P$.

21.2.2 Tissue boundary conditions and system size

To fully describe a tissue, one needs to consider how to deal with the boundary conditions at the tissue edge. In a growing tissue simulation, one would opt for a finite size tissue growing on an infinite plane with an optional special treatment being applied to the cells being detected at the edge of an expanding tissue. For the vertex model in its relaxation implementation, we will instead consider a *confluent* or *covering* tissue, i.e., we define a box with periodic boundary conditions such that cells cover the entire box and periodic copies of the box tile the plain. Hence the “outer cells” to the left and to the top of the box will share edges with the outer right and lower bottom ones, respectively.

In this periodic boundary conditions scenario, it is crucial to have a reasonable minimum number of cells in the system to avoid the observation of finite size effects due to cells interacting with their own copies. Such effects have been observed in various other applications like in molecular dynamics and can either be corrected for analytically or by scaling the system up [19]. Here we face two competing effects: On the one hand, we want the system to be reasonably large to avoid finite-size effects. On the other hand, the nature of the equilibration protocol will lead to quickly increasing relaxation simulation times with increasing numbers of cells. Depending on how many equilibrations need to be run up until convergence, one may either decide to prioritize one or the other. We often limit ourselves to tissues of sizes below 100 cells, as other groups have been forced to do, too [16].

21.3 Implementing the vertex model relaxation procedure

add a global paragraph to introduce this section.

21.3.1 Data structures for simulation

For the vertex model simulation, we require a set of data structures to represent the tissue and its cells as well as boundary conditions:

- We will call a data structure with two floating point coordinates a “vector”. This represents the two-dimensional position of a point in the plain.
- A cell is then described as a list of vectors representing the vertices of its polygonal shape. Additionally, an integer index is associated with each of the vertices. This index is used to identify which vertices of different cells are meant to represent the same point in space. As each vertex is shared between three cells, the same index should occur exactly three times across the whole tissue.

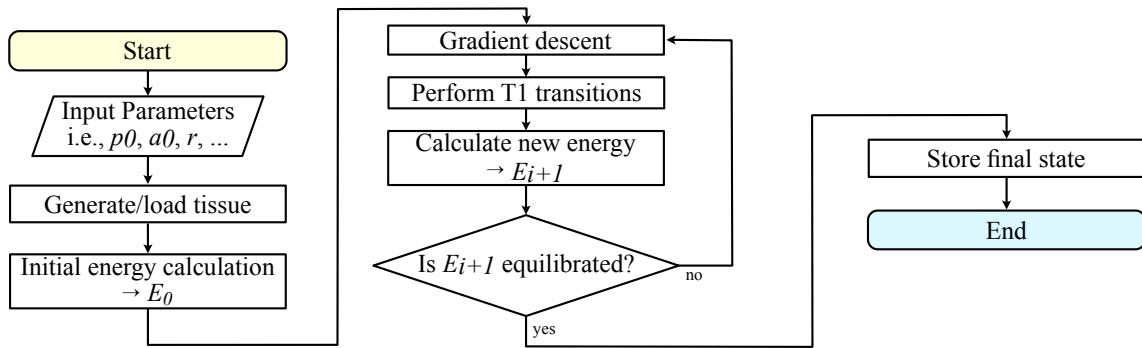


Figure 21.3: **Process flow chart of the main simulation process of the vertex model.** The main simulation process of the vertex model consists of some initial setup steps to obtain a base tissue configuration, an equilibration loop being run until convergence is reached, and a final step to store the resulting tissue configuration.

- A tissue is then represented by a list of cells as detailed before, combined with two vectors representing the upper left and bottom right edges of the periodic boundary box. Alternatively, the boundary condition box can also be represented by one vector with the period length in x and y direction as the respective coordinates.

21.3.2 Outline of the relaxation procedure

The overall vertex model simulation can be abstracted to a small set of main steps as presented in the flow chart shown in Fig. 21.3. We want to outline the overall procedure to give an idea of the relation between the steps as well as their function before going into more detail on their individual implementations.

First, the optimization and relaxation procedure requires a certain set of input parameters for the process — most notably the model parameters A_0 , P_0 , and r — as well as an initial tissue state to work on. For this initial state, it will either load a tissue configuration or generate it based on some settings, e.g., the number of cells to be considered and their average area. Then, some initial setup computations are carried out including the calculation of the initial total energy E_0 of the tissue to ensure that a state change only ever reduces total energy with each optimization step.

Then, the actual optimization and relaxation start. The optimization is run in multiple iterations with the i -th iteration starting with the energy E_i and ending with the energy E_{i+1} . This optimization continues until we consider the total energy to have converged to a minimum. Each iteration involves only two major modification steps required to describe the tissue dynamics being considered:

1. First, we have the energy minimization using gradient descent in the energy landscape described by the typical per-cell energy functional detailed in Eq. (21.1) for a fixed tissue topology — i.e., fixing which cell shares edges with which other cell. In this, the individual vertices are moved around to reduce the total energy of the tissue without changing the topology — i.e., the number of vertices per cell or which vertices are connected by edges. This step can also be referred to as the equilibration step, driving the system to the closest (local) energy minimum and thus a stable equilibrium given the fixed topology.
2. Then, there is a second step in which the topology of the tissue may be updated in an operation referred to as a *T1 transition*. A T1 transition is the process of an initial shared edge between two neighboring cells being replaced by a shared edge between the two common neighbors of those cells. This allows for cells to lose or gain neighbors, which is the only way in which the tissue can change its structure. In the process of overall energy optimization, this change in topology can be quite a significant contribution on top of the equilibration step. Again, as in the equilibration step, a T1 transition will only be performed in case the total tissue energy after the transition is reduced relative to the energy at its start.

After the performance of equilibration and T1 steps, the resulting energy E_{i+1} of the tissue is calculated and compared to the previous energy E_i at the beginning of the iteration. If the change $\Delta E = |E_{i+1} - E_i|$ is above

either an absolute or relative threshold, then the optimization step is repeated, starting from the equilibration step. Otherwise, the relaxation is considered to have converged and the final optimized state is stored.

21.3.3 The individual steps of the vertex model relaxation procedure

Add a global paragraph to provide an overview of this section

Generating and loading the initial state

In order to relax a tissue, we need an initial guess for a state from which the relaxation can start.

How to store a tissue configuration On the one hand, the state can be loaded from a file, if this has been generated previously. If you have no prior experience with reading and writing binary files, we recommend a textual representation of the tissue's state. The tissue can for example be represented as follows:

- First, the periodic boundary conditions information is provided, usually amounting to the vectors of the bottom left and the upper right corner of the bounding box if it is a rectangle.
- Then, the actual tissue information is provided:
 - The integer number n_v of distinct vertices in the tissue, followed by n_v vertices represented by two coordinates each.
 - The number of cells n_c of the tissue followed by n_c representations of a single cell:
 - * The number m_v of vertices the cell's polygon shape has followed by m_v entries for the vertices.
 - * Each of the vertices is represented by two integer indices i_{pos} and i_{pbc} . The index $0 \leq i_{\text{pos}} < n_v$ is the index of the vector in the initial list of n_v vectors representing the position of the vertex. The index $0 \leq i_{\text{pbc}} < n_c$ on the other hand identifies the vertex in the topology. If two vertices have the same i_{pbc} then they represent different copies of the same vertex in its 3 adjacent cells.

All programming languages provide tools to read such a textual format, which, thanks to its human-readable nature, is also way easier to check for errors. Here, we have gone for a split between the positional information (the initial list of vector coordinates) and the description of the tissue topology (the list of cells and their vertices). As an alternative, one may merge the two sets of data and instead of storing i_{pos} for a vertex in the cell, directly store the vertex coordinates. One would still need to retain i_{pbc} , due to shared nature of vertices in a confluent tissue with periodic boundary conditions.

How to generate a tissue If one does not have a previously generated tissue configuration, an initial state needs to be generated from scratch. The process for doing so in our simulator is visualized in Fig. 21.4.

First, we use a random number generator to obtain the coordinates of random points within the periodic boundary conditions box. For each of these points, we want to generate a corresponding cell in our tissue. To account for the periodic boundary conditions, we first copy the generated points to its first 8 neighboring positions (see Fig. 21.4a) so that we can account for all neighbor relations between cells in our system.

Then, we construct the *Delaunay triangulation*¹ of all the points in these 9 periodic boundary conditions cells to identify which cells will eventually be sharing a common edge (see Fig. 21.4b).

Finally, from the intermediate Delaunay triangulation, we construct the actual shape of the cells by constructing the *Voronoi cells*² of each point and eventually only retaining the Voronoi cells of the points within the main periodic boundary conditions cell (see Fig. 21.4c). This leaves us with a collection of cells that we can use as a basis for our relaxation procedure.

The way the initial points are distributed within the periodic boundary conditions box can lead to different initial configurations. Still, assuming that the relaxation procedure is efficient and stable enough, the vertex

¹The Delaunay triangulation [20] is a complex algorithm generating a set of edges between points fulfilling certain conditions concerning the position of points being outside of the circumcircles of the constructed triangles as well as conditions on the angles between edges. It is mainly used as an intermediate step for the generation of the cell shapes obtained by constructing the Voronoi cells from the triangulation. An edge in the triangulation indicates that two cells will be neighbors at the end of the generation process.

²A Voronoi cell [21] of a point is the set of points in the plane closer to this point than to any other point. It can be constructed using the perpendicular bisectors of the edges of the Delaunay triangulation.

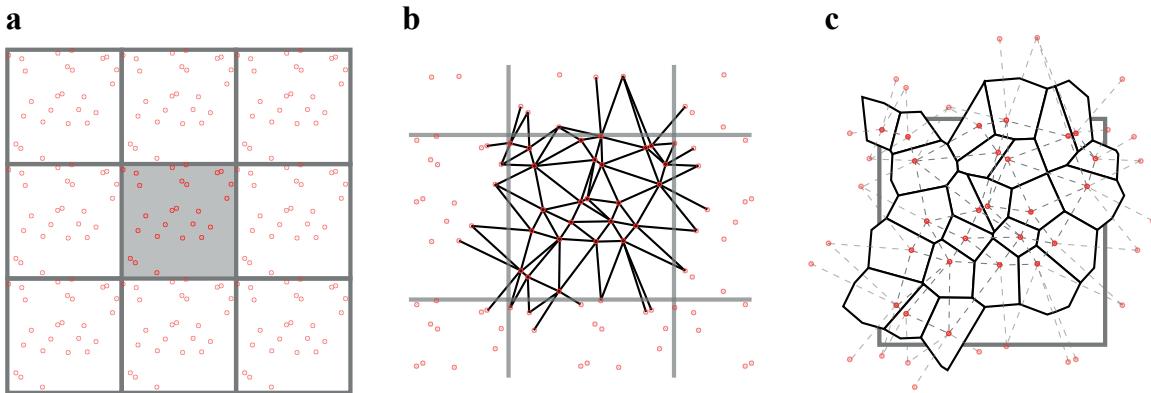


Figure 21.4: Scheme of tissue generation. (a) Generation of randomly distributed points within the main periodic boundary conditions cell (grey) and their copies in the 8 immediately neighboring periodic boundary conditions cells. At the end of the generation, one polygon/cell will be associated with each point. (b) Triangulation of the set of points using the Delaunay-algorithm [20]. An edge is drawn between two points if the respectively generated cells will touch in the end. (c) Construction of the Voronoi cells [21] from the intermediate Delaunay-triangulation obtained in step (b). The Voronoi cells constructed for the initially generated points are used as an initial set of cells for the relaxation process.

model optimization procedure should still eventually lead to statistically comparable results no matter how the cells were seeded initially. The transient evolution during the optimization will probably be different, though.

As usual, there are other options for the generation of the initial configurations depending on which properties one is trying to investigate. Tiling the periodic boundary conditions with a repeating pattern of cells with identical shapes is an option, but the nature of the vertex model makes it very unlikely for the degree of symmetry of the system to drop during tissue relaxation. The statistical results when starting from such a configuration may thus not be generalizable. We therefore recommend starting from a more random initial condition.

21.3.4 Energy calculation

The energy of the system is calculated by looping over all cells, calculating their perimeter and their cell area and, based on that, the cell energy via Eq. (21.1). By adding up those energies for all cells, we obtain the full-tissue energy.

Be aware that you may need to account for the periodic boundary conditions to obtain absolute positions for vertices when calculating P and A for a cell depending on how you decide to deal with the periodic boundary conditions. Accidentally using a periodic copy of a vertex for the calculation can lead to very large perimeters and negative or very large areas. Pay attention to the individual cells' values during the calculation to identify such issues of periodicity. A useful tool here is finding the closest periodic copy of a vector given the periodic boundary conditions of the box, which is a known problem in vector algebra and will not be detailed here.

21.3.5 Equilibration and gradient descent step

In the equilibration step, the tissue is supposed to be optimized with respect to the energy functional provided in Eq. (21.1). For this purpose, we can sum up the forces acting on the individual vertices resulting from the adjacent cells' energy functionals as derived in Eq. (21.2). This assigns the energy gradient $\nabla_i E$ to each vertex i in the tissue, pointing towards the direction of steepest increase of energy (see Fig. 21.5a). We can then follow the opposite direction of this gradient for individual vertices to converge towards the point of lowest energy (see Fig. 21.5b). As we do not strive for the simulation of an accurate time evolution but instead want to explore the configuration space for ideal tissue states, we can either perform gradient descent steps on each vertex individually or on all vertices in parallel. Also, we have the option to repeatedly perform the gradient descent steps until the energy cannot be optimized any further, or just once. If the step is repeated until convergence,

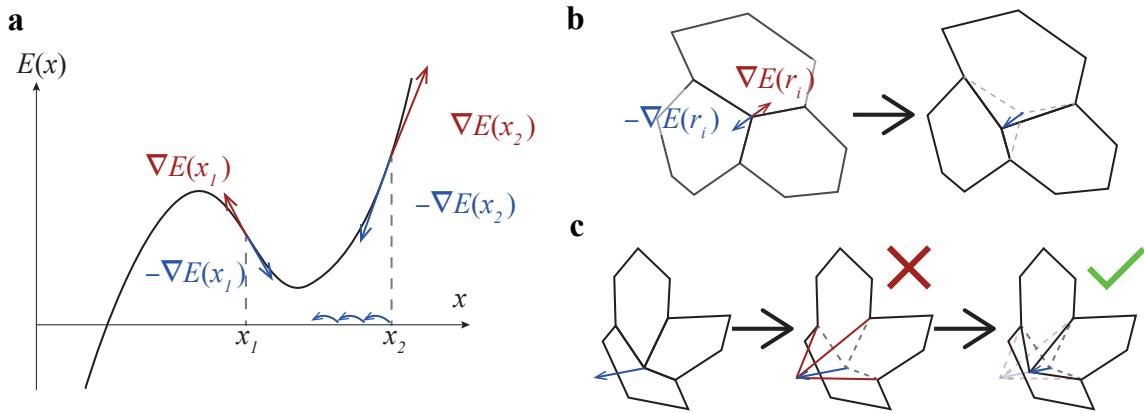


Figure 21.5: Gradient descent in tissues. (a) General visualization of the concept of the gradient descent. Starting at either position x_1 or x_2 , the gradient ∇E (red) will always point towards the highest increase of E . One can therefore follow the direction of $-\nabla E$ (blue) towards a (local) minimum. This may need to be repeated several times to actually reach the desired minimum. (b) Process of gradient descent for a vertex in the tissue. After calculating $-\nabla E$ for a vertex from the contribution of all three neighboring cells, the vertex is moved along $-\nabla E$, which changes the length of the cell edges (or rather the length of cell membranes) P as well as cell areas A . (c) Possible issues with geometric constraints. When following the negative gradient $-\nabla E$ for a vertex, it may happen that the vertex crosses another edge of one of its three adjacent cells, thus creating intersecting pairs of edges amounting to intersecting cell membranes. This is an undesired effect and the gradient descent step needs to be repeated with a shorter step length to try and avoid this from happening.

we have reached a local minimum in tissue energy and a change in topology would be required to make the tissue converge any further.

While performing the gradient descent, it is vital to account for constraints imposed on the cells, especially at high values of P_0 . In scenarios with long target perimeters for a given target area, the energy of the tissue when making the full step might actually increase or cells may be driven towards non-convex shapes or even towards the edges of their polygonal shapes intersecting (see Fig. 21.5c). While it may not be an issue to have a cell shift towards a non-convex shape, intersecting edges — amounting to intersecting cell membranes — do not represent a (physically) reasonable configuration of the tissue. Hence, such aspects need to be checked whenever a vertex has been moved and, if the move has forced the tissue into an unreasonable state or lead to an increase in energy, the step needs to be reverted and a smaller step size should be attempted (see Fig. 21.5c). It can happen that no stride length along the negative gradient allows for the constraints to be kept, this can for example be tested by repeatedly halving the step length and retrying again and again. If several reductions in step length still do not allow for a decrease in energy while adhering to constraints, the step should be considered as failed and either another vertex should be attempted or the next operation should be executed instead.

Instead of immediately moving a vertex along the entire length of its gradient vector, we often use a relative initial step length for our simulations. Once the gradient $\nabla_i E$ has been calculated, one should try a first step of length $\alpha \cdot \nabla_i E$ with, e.g., $\alpha = 10^{-1}$ or $\alpha = 10^{-2}$, depending on how fine-grained you want your relaxation to be.

21.3.6 T1-transition step

The gradient descent step in which the topology of the tissue is always maintained is then followed by a step where the topology is modified in a physically and biologically reasonable manner while maintaining the confluence of all cells, i.e., without allowing holes to form in the tissue. In real tissues, neighboring cells can detach from each other, which needs to be represented in this step via so-called *T1-transitions*. A T1-transition is the process of two neighboring cells losing their shared edge and a new edge forming between their two common neighbors instead (visual representation in Fig. 21.6).

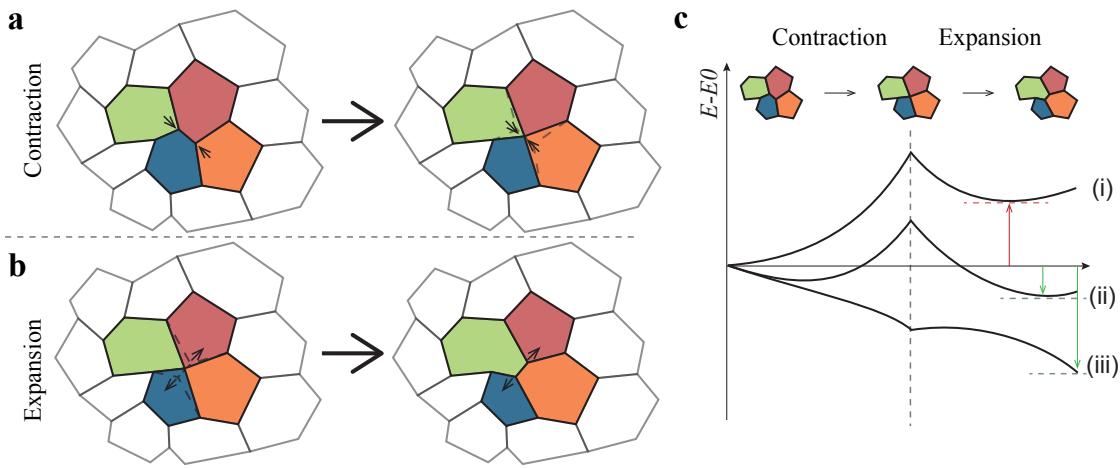


Figure 21.6: Tissue topology changes via T1 transitions. (a) Visualization of the *contraction step* of the T1 transition, where an edge between two cells is collapsed into a point. (b) Visualization of the *expansion step* of the T1 transition following the contraction in step a. A new edge is generated perpendicular to the original edge that has been collapsed and expanded. It is now a shared edge between the two common neighbors of the two cells whose initially shared edge has been collapsed. (c) Examples of qualitative energy profiles observed during T1 transitions. The vertical axis is labeled $E - E_0$. The horizontal axis represents the state of the tissue. Three curves are shown: (i) a double-well potential with a local maximum above the initial energy; (ii) a single-well potential with a minimum below the initial energy; (iii) a single-well potential with a minimum at the initial energy level. The latter is possible, if the collapsed state is actually energetically favorable, but, due to vertices only ever having three neighbors, the edge cannot fully collapse by gradient descent alone. The decision to accept or reject a T1 transition is then made based on the minimum energy observed during the expansion step. Here, (i) has a minimum expansion energy above the initial energy and is therefore rejected, whereas (ii) and (iii) have an energy minimum during expansion that is below the initial energy. Hence, the T1 transition is accepted and the expansion state with minimum energy is picked as the post-transition state.

We identify possible candidates for a T1 transition via the length of the shared edge between two cells, i.e., by checking whether the edge length is below a certain length threshold length l_T . If the gradient descent in the energy functional has driven an edge towards collapse therefore resulting in a short length $l < l_T$, then it might be energetically beneficial to replace it with an edge between two other cells instead, while maintaining the constraint that each vertex should have exactly 3 adjacent cells. The two cells between which the new edge is formed are the two cells that do not contain the relevant edge but do contain the vertices delimiting the edge. As each vertex is a part of exactly three cells, these two are uniquely identifiable. Once all possible candidate edges with length $l < l_T$ have been identified, we successively try for every one of them to perform a T1 transition and then check if it has reduced the total energy of the tissue.

Performing a T1 transition involves first actually collapsing the candidate edge to a point from its original state associated with the energy $E_{T1,0}$, as visualized in Fig. 21.6a. This can either be done instantaneously or with several intermediate steps. In a real tissue, this is a continuous process, thus being a gradual contraction of the edge with the surrounding tissue adapting to the change. It is therefore reasonable to perform the collapse in several intermediate steps with tissue relaxation and gradient descent steps being performed for each of the intermediate steps to mimic local tissue response to the topology change.

Once the edge has been collapsed into a vertex now neighboring four cells, a new edge is created perpendicular to the original edge and expanded up to a maximum length of l_E , as visualized in Fig. 21.6b. This is again performed in a gradual manner with relaxation of the tissue performed on the intermediate steps of the expansion. One vertex in the collapsed state having 4 neighboring cells is fine because it is only an intermediate configuration and is immediately transformed back into a valid state as soon as the expansion of the new edge starts. During the multiple intermediate steps of edge expansion, for each length of the expanded edge $0 < l \leq l_E$ the resulting relaxed total tissue energy $E_{T1}(l)$ is calculated and retained.

Once the full expansion has been performed, an energy profile depending on the expansion length l similar to that in Fig. 21.6c should have been observed. (The set of energy profiles in Fig. 21.6c is not exhaustive and just meant to qualitatively illustrate how energy can evolve during the transition process). If at no point during the expansion an energy $E_{T1}(l) < E_{T1,0}$ has been observed, then the tested transition is not favorable to overall tissue energy and it should be rejected while reverting to the original pre-T1-state with energy $E_{T1,0}$. If instead during expansion at length $l_m > 0$ a minimal energy $E_{T1}(l_m) = \min_l E_{T1}(l)$ below the initial energy before the transition (i.e., $E_{T1}(l_m) < E_{T1,0}$) has been observed, then the topology and tissue state associated with this length l_m should be chosen for any further operations. This is repeated until all possible candidate edges have either been rejected or collapsed.

Some notes on the implementations of T1 transitions:

- The data manipulations required for performing the T1 transitions include the calculation of the position of the center point of the edge to which the edge collapses, the subsequent calculation of the end points of the new edge from the direction of the original edge with the new edge being perpendicular as well as the removal and addition of vertices to the respective cells involved in the T1 transition.
- During the relaxation of the tissue performed during the intermediate steps of contraction and expansion, the positions of the vertices delimiting the old and the new edges respectively should not be relaxed along with other vertex positions.
- The number of intermediate steps as well as the number of relaxation steps within each intermediate step influence the outcome of the overall vertex model optimization process. Fewer T1 transitions are accepted if no relaxation is performed, as the relaxation reduces overall tissue energy and makes it more likely for a state during expansion to be below the initial energy $E_{T1,0}$.
- When the T1 transition moves from contraction to expansion, the topology of the tissue only needs to be updated once and only the four involved cells need to be updated but special care needs to be taken that the right vertices are part of the right cell after the transition. The first check should be that the correct cells have lost one vertex and the other two have gained one and then a visual inspection of the resulting tissue should be performed during testing to make sure the vertices are at the right position. This is made easier by keeping the vertices within each cell consistently in a (counter-)clockwise orientation because then the two vertices of the new expanding edge can be obtained by rotation of the original edge (counter-)clockwise and rescaling. Keeping the same orientation of vertex orientation in cells and the

rotation of the original cell then helps with assigning the correct new point to the correct cells and only requires visual inspection once or twice to make sure it is done correctly.

21.3.7 Check for convergence

Convergence of a series of energies $E_i > 0$ is usually identified via the combination of an absolute and a relative threshold of change. Let $\Delta E_i = |E_{i+1} - E_i|$ denote the absolute change and $\varepsilon_i = \frac{\Delta E_i}{E_i}$ the relative change of energy from iteration i to iteration $i + 1$. We consider the energy to be converged and therefore the tissue to have relaxed once $0 \leq \Delta E < E_T$ or $\varepsilon < \varepsilon_T$ is below a relative threshold ε_T .

The absolute threshold may at first glance seem sufficient to identify very minor changes in the occurring energy values and thus convergence of the process, but there are situations where this is not a good measure for how far the optimization has come. Indeed, if the optimum configuration is at a very high energy scale, then the absolute steps ΔE_i between iterations may stay at a large scale for a very long time without significant progress in terms of changes to the tissue configuration. The relative threshold is therefore a good measure for whenever we do not know the absolute scale of energies and want to estimate the chance for significant overall progress through the relaxation routine.

If the relative change is below, e.g., $\varepsilon_T = 10^{-3}$, we can be sure that independent of the absolute scale of the energy the achieved change does not significantly affect its scale anymore. Furthermore the effect of any further relaxation iterations is also expected to be small as big changes tend to happen early on and successive iterations generally exhibit diminishing returns with only occasional spikes in ΔE_i . Hence, we can consider the system to be converged.

If the absolute scale of the optimum energies is known precisely, the absolute threshold E_T alone can provide us with a good measure of convergence. In general, one should always consider combining the two for the most reliable identification of convergence.

On top of that, it is always a good idea to limit the maximum number of relaxation iterations being performed to a reasonable number when implementing a simulator oneself. This can help if there is an issue with the relaxation that has not originally been considered but we do not want to lose the progress that we have already made, for example when a large system has been processed for quite a long time.

21.3.8 Storage of tissue state

For the storage of the final tissue state, we again recommend using a textual description as presented for the generation and loading step or — for more advanced readers — going for a binary (i.e., machine-readable) custom format containing the same information. The latter has the advantage of being more storage-efficient but the downside of being non-human-readable so if there is an issue, it will be harder to debug.

For debugging purposes, you may also consider outputting a visual representation in the form of an SVG-file. SVG is a vector image format ideally suited to output polygonal shapes like the ones used here. Additionally, SVG is written in a human-readable text description of where to put vertices of polygonal shapes. There are various tutorials for outputting SVG format figures,³ and we would recommend following any one of them as, in our experience, the visual representation of the tissue state along the relaxation process has helped tremendously with debugging.

21.3.9 Technical and analytical limitations

The implementation of the vertex model as explained before is quite restricted in the possible system sizes that it can accommodate. Especially the complex relaxation process of the T1 transitions has a heavy toll in terms of computations as not only the number of candidates for T1 transitions increases but also the number of required relaxation steps during each T1 transition. Hence, an increase in the number of cells will very quickly increase the computation time required to fully optimize the tissue. This has been observed by other groups as well [16], which have therefore restricted their analysis to double-digit numbers of cells. Reasonable statistics of observables can then be obtained mainly by re-running the relaxation for different random initial configurations with the same relaxation parameters and analyzing the different configurations independently.

³See for example <https://developer.mozilla.org/en-US/docs/Web/SVG/Tutorial>

Another less obvious issue of the described relaxation process of the vertex model is the constraint imposed on the average cell area, which effectively reduces the number of degrees of freedom per cell [22]. Due to fixed periodic boundary conditions causing the overall area of the confluent tissue to be constant and no changes in the number of cells via apoptosis or proliferation, the average cell area does not change during the simulation. A relaxation operation of the periodic boundary conditions could help mitigate the issue, but with the described methods, the parameters should be chosen carefully.

While introducing cell apoptosis and proliferation may seem like the obvious solution to the above problem, we would like to point out some considerations for these processes that make it hard to bring them into the system without experimental data to answer these questions:

- How is cell division triggered? Is it a purely random process or is there a condition on cell parameters? If the decision to split a cell is made based on its perimeter or area, this will alter certain statistics of the tissue.
- Is there a prelude to cell division? When cells in real tissues divide, they will grow to a larger size before actually forming a separating membrane. It may therefore be a good idea to account for this growth by altering cell parameters A_0 and P_0 once the decision to divide has been made.
- In which direction is a cell split in half during proliferation? This is relevant due to the possible effect on average cell elongation and other statistics relating to cell orientation/polarization. As a rule of thumb it might be wise to split a cell in half along its main axis that can be determined via principal component analysis (PCA).
- How is cell apoptosis triggered? Is it a purely random process or is it coupled to energy and/or size considerations?
- How do we distribute the area of a dying cell among its neighbors? We cannot just remove it from the tissue without losing confluence. Instead we need to account for its former neighbors coming into contact with each other once the cell is removed. As we generally want vertices with only 3 neighboring cells, we cannot just collapse the cell to a point and be done with it, but instead we need to divide up its space and introduce a new topology for its neighborhood. How does one determine, which of the former neighbors will be sharing a new edge once the dying cell has been removed?

21.3.10 Possible tweaks and optimizations

In the implementation of the gradient descent operation both in its own step and as a substep of the T1 transition's relaxation, there is the option to go for a stochastic approach to optimize the convergence of the tissue. In our experience, the various constraints of the individual vertices makes it inefficient to perform the gradient descent on all vertices simultaneously, because any constrained vertex will make it hard to relax all other vertices. Picking one individual vertex and then relaxing it within the energy landscape defined by the remaining ones is therefore preferable. Furthermore, to avoid the effect of always performing the gradient descent in the same order on all vertices, the order of vertices within the gradient steps can be chosen randomly prevent cyclic updates from happening that could just amount to a shift of the overall tissue but not generate any real new tissue relaxation effect. This arbitrary choice of order of optimization is possible due to the nature of our simulation not trying to adhere to a real time evolution but instead performing an exploration of phase space.

Whenever there is an operation run in parallel across the entire tissue like the generation of gradients, the determination of T1 edge candidates and the likes, parallelization can be considered. Yet, a main issue is the sequential nature of certain steps like the intermediate relaxation of T1 transitions, which limits the ability to parallelize T1 transitions occurring on multiple edges at the same time. This is a main reason for the constraint on the number of cells that can be relaxed in a reasonable time span.

21.4 Final remarks

On the scale of the mesoscopic tissue description, we have opted to detail an implementation of the vertex model — a description technique, where aspects of cell shape are retained but simplified to a polygonal shape.

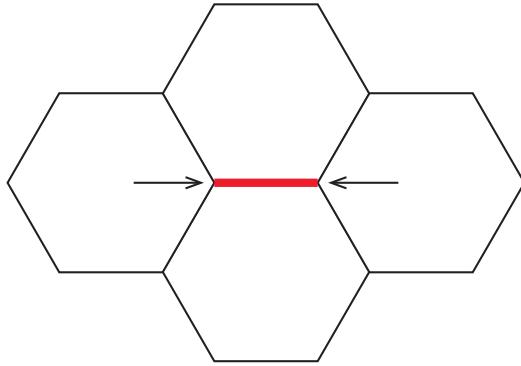


Figure 21.7: **Simplified system for analyzing a T1 transition.** A set of four hexagonal cells with its center edge, represented with a thick red segment and highlighted by two arrows.

This places it firmly between more complex descriptions of cell shape like the cell Potts and the phase field approach and more abstract techniques like the particle and continuum models.

We have laid out the parts of the vertex model theoretical foundations necessary for the implementation, going into detail on some analytical aspects. Building on that, we have laid out the necessary data structures for simulating a tissue in the vertex model formalism required to keep track of tissue topology. The detailed guide for the implementation of the individual steps for the use of the vertex model for tissue relaxation includes tips regarding pitfalls of implementation as well as possible extensions and optimizations of the simulation that the limited space of this section does not allow to go into detail on. The depicted algorithm is able to reproduce jamming behavior and tissue stratification previously described for in-silico investigations by other groups [16] and can easily be extended to account for more recent developments of the field like nematic orientation behavior. On request, the authors can provide sample code for how certain aspects of the model have been implemented in C++ to help with issues during implementation. *@authors: can't we make it publicly available?*

The interested reader may want to build upon the described algorithm to set up the simulation of a growing non-confluent tissue with cell proliferation and intermediate structure optimization/relaxation steps. We recommend this as an excellent challenge to reconsider the choices made for the implementation of the vertex model presented here, e.g. the assumption of a confluent tissue which reduces the amount of edge cases one needs to consider in the implementation of T1 transitions and the analytical form of the energy functional. The treatment of the edge has been a topic of ongoing investigation by other groups [23, 24] and we recommend taking a look at their work first for some inspiration for practical implementation of non-confluent tissues.

21.5 Problems

Problem 21.1: Tissue topology and T1 transitions. The energy profile observed during a T1 transition as depicted in Fig. 21.6c can be quite complex. It depends heavily on the choice of parameters P_0 and A_0 in 21.1, which influences both the optimum shape a cell can attain within the confluent tissue and how the energy functional during a transition evolves.

- a. Convince yourself that, in a confluent tissue with periodic boundary conditions where all vertices have three adjacent cells, on average each cell must have 6 vertices. [*Hint: Consider the angles at the vertices*]
- b. Let us now fix $A_0 = 1$ and only consider the influence of the choice of P_0 . Find the optimum shape of a cell depending on P_0 by distributing the points used for generating the tissue in a grid pattern and then running the relaxation step until it converges. How does the choice of P_0 influence the shape? Does the shape or orientation of cells depend on how you chose the initial grid pattern? There should be a change in behavior

of the cell shape depending on P_0 . Where is that critical value of P_0 ? [Hint: A circle is the shape with the shortest possible perimeter relative to the square root of its area]

c. Setup a simplified system of four regular hexagonal cells as depicted in Fig. 21.7 and perform a T1 transition on its center edge. Perform this T1 transition with intermediate steps and keep track of the overall energy profile depending on P_0 . How does the energy profile observed during the T1 transition change if you perform gradient descent/relaxation steps at each intermediate step of contraction and expansion? [Note: You can also go for a system with periodic boundary conditions but then you will need to include more than these four cells. Also the results may depend on your chosen system size.]

@authors: It would be good to also have some problems about numerical implementation or use of the code.

Bibliography

- [1] R. Alert and X. Trepaut. Physical models of collective cell migration. *Annual Review of Condensed Matter Physics*, 11:77–101, 2020.
- [2] F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Physical Review Letters*, 69(13):2013, 1992.
- [3] W.-J. Rappel, A. Nicol, A. Sarkissian, H. Levine, and W. F. Loomis. Self-organized vortex state in two-dimensional Dictyostelium dynamics. *Physical Review Letters*, 83(6):1247, 1999.
- [4] R. González-Cinca, R. Folch, R. Benítez, L. Ramírez-Piscina, J. Casademunt, and A. Hernández-Machado. Phase-field models in interfacial pattern formation out of equilibrium. *Advances in Condensed Matter and Statistical Mechanics*, pages 203–236, 2004.
- [5] L. Coburn, L. Cerone, C. Torney, I. D. Couzin, and Z. Neufeld. Tactile interactions lead to coherent motion and enhanced chemotaxis of migrating cells. *Physical Biology*, 10(4):046002, 2013.
- [6] M. Basan, J. Elgeti, E. Hannezo, W.-J. Rappel, and H. Levine. Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing. *Proceedings of the National Academy of Sciences*, 110(7):2452–2459, 2013.
- [7] S. Kaliman, M. Hubert, C. Wollnik, L. Nuić, D. Vurnek, S. Gehrer, J. Lovrić, D. Dudziak, F. Rehfeldt, and A.-S. Smith. Mechanical regulation of epithelial tissue homeostasis. *Physical Review X*, 11(3):031029, 2021.
- [8] S. Kaliman, C. Jayachandran, F. Rehfeldt, and A.-S. Smith. Limits of applicability of the voronoi tessellation determined by centers of cell nuclei to epithelium morphology. *Frontiers in Physiology*, 7:551, 2016.
- [9] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning. Motility-driven glass and jamming transitions in biological tissues. *Physical Review X*, 6(2):021011, 2016.
- [10] X. Yang, D. Bi, M. Czajkowski, M. Merkel, M. L. Manning, and M. C. Marchetti. Correlating cell shape and cellular stress in motile confluent tissues. *Proceedings of the National Academy of Sciences*, 114(48):12663–12668, 2017.
- [11] B. Smeets, R. Alert, J. Pešek, I. Pagonabarraga, H. Ramon, and R. Vincent. Emergent structures and dynamics of cell colonies by contact inhibition of locomotion. *Proceedings of the National Academy of Sciences*, 113(51):14621–14626, 2016.
- [12] B. Szabo, G. J. Szöllösi, B. Gönci, Zs. Jurányi, D. Selmecki, and T. Vicsek. Phase transition in the collective migration of tissue cells: experiment and model. *Physical Review E*, 74(6):061908, 2006.
- [13] S. Banerjee and M. C. Marchetti. Continuum models of collective cell migration. *Cell Migrations: Causes and Functions*, pages 45–66, 2019.

- [14] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17(24):2095–2104, 2007.
- [15] M. B. Liu, G. R. Liu, L. W. Zhou, and J. Z. Chang. Dissipative particle dynamics (DPD): an overview and recent developments. *Archives of Computational Methods in Engineering*, 22(4):529–556, 2015.
- [16] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning. A density-independent rigidity transition in biological tissues. *Nature Physics*, 11(12):1074–1079, 2015.
- [17] R. Alert and X. Trepat. Living cells on the move. *Physics Today*, 74(6):30–36, 2021.
- [18] R. Winklbauer. Cell adhesion strength from cortical tension—an integration of concepts. *Journal of Cell Science*, 128(20):3687–3693, 2015.
- [19] I.-C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *The Journal of Physical Chemistry B*, 108(40):15873–15879, 2004.
- [20] B. Delaunay. Sur la sphère vide. *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7(793-800):1–2, 1934.
- [21] G. Voronoi. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908(133):97–102, 1908.
- [22] D. M. Sussman and M. Merkel. No unjamming transition in a Voronoi model of biological tissue. *Soft Matter*, 14(17):3397–3403, 2018.
- [23] M. L. Manning, R. A. Foty, M. S. Steinberg, and E.-M. Schoetz. Coaction of intercellular adhesion and cortical tension specifies tissue surface tension. *Proceedings of the National Academy of Sciences*, 107(28):12517–12522, 2010.
- [24] E. Teomy, D. A. Kessler, and H. Levine. Confluent and nonconfluent phases in a model of cell tissue. *Physical Review E*, 98(4):042418, 2018.

Chapter 22

Active Fluids

LUCA BIANCOFIORE

An *active fluid* is a collection of particles, cells, or macromolecules, typically suspended in a viscous fluid, that are able to convert chemical energy into mechanical work. There are several kinds of active fluids, e.g., suspensions of different kinds of microswimmers [1] such as synthetic active colloids [2], droplet swimmers schools of fish, and biological microorganisms like bacteria, algae, or sperm cells, and active nematic fluids [3]. In an active fluid, we may need to model the interaction between the active particles and the solvent and the cross-interactions among different microswimmers. Since the definition of active fluids contains a broad variety of fluid systems, it is not surprising that their equations in the continuum description change significantly depending on what we need to model. In some cases, a *Lagrangian approach* is the best option to model the active particles moving in the fluid (that will described by the Navier–Stokes equations), while in other cases, in which computational efficiency is a must (such as for active fluids in complex geometries or systems where non-linear coupling is present), an *Eulerian model*, which describes the ensemble of the particles plus the fluid acting as a continuum, is the most suitable choice. The goal of this Chapter is to introduce the basic techniques of numerical modeling of fluid flows to the reader. These techniques will be very useful to tackle the diverse challenges given by the numerical treatment of active fluids.

In this Chapter, we first introduce some examples of governing equations for active fluid systems (Section 22.1). Afterwards, the Navier-Stokes equations, i.e., the most used continuum model for describing fluid motion are presented (Section 22.2). The finite volume method (FVM) that will be used in this Chapter to discretize the Navier-Stokes equations is introduced in Section 22.3 and the discretization itself will be conducted in Section 22.4. Some properties of the numerical scheme are presented in Section 22.5.

22.1 Equations for an active fluid

In the following subsections, we will describe the equations for three systems as examples: (i) a Janus particle in a fluid, (ii) a suspension of active Brownian swimmers in an incompressible and Newtonian fluid, and (iii) a suspension of active Brownian swimmers in a non-Newtonian fluid, in the presence of a shear flow. These systems are chosen due to their practical applications. The Janus particle system is an example of a synthetic colloidal particle, while the other two are examples of microswimmers suspensions in different fluids. It should be noted that non-Newtonian fluids are ubiquitous in biological, and particularly in soft matter, applications. These examples show how diverse active systems can be. This diversity poses problems in having a universal approach (both mathematically and numerically) for active fluids. For this reason, special ad-hoc numerical treatments could be often needed.

Figure 22.1: **Cover figure for the chapter.** Caption.

22.1.1 Motion of a Janus particle in a fluid

Let us consider first the motion of an *active Janus particle* in an *incompressible Newtonian fluid* [2]. We can describe the motion of the particle from the Newton's equation

$$m \frac{d\mathbf{v}}{dt} = \mathbf{F}_{\text{ext}} - \int_{\Sigma(t)} d\Sigma \mathbf{P}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}, \quad (22.1)$$

where m is the mass of the Janus particle, $\Sigma(t)$ is its surface, \mathbf{v} is its velocity, $\mathbf{P}(\mathbf{r}, t)$ is the fluid pressure tensor, \mathbf{F}_{ext} is the sum of all the external forces (if any), t is the time, and \mathbf{r} is the position in the domain. Moreover, the fluid exerts also a torque on the Janus particle and its angular velocity $\boldsymbol{\Omega}$ is given by

$$I \cdot \frac{d\boldsymbol{\Omega}}{dt} = \mathbf{T}_{\text{ext}} - \int_{\Sigma(t)} d\Sigma \nabla \mathbf{r} \times [\mathbf{P}(\mathbf{r}, t) \cdot \hat{\mathbf{n}}], \quad (22.2)$$

where I is the inertia tensor of the Janus particle and \mathbf{T}_{ext} is the sum of the external torques.

To find the fluid pressure acting on the particle surface, we need to solve the fluctuating Navier–Stokes equations for an incompressible Newtonian fluid:

$$\begin{cases} \nabla \cdot \mathbf{u} = 0 \\ \rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = -\nabla \cdot \mathbf{P} \end{cases} \quad (22.3)$$

where \mathbf{u} represents the fluid velocity vector and ρ is the density.

Note that the pressure tensor can be expressed as $\mathbf{P} = p\mathbf{I} + \boldsymbol{\Pi}$, where p is the hydrostatic pressure and the viscous pressure tensor components are given by

$$\Pi_{ij} = -\mu (\partial_i v_j + \partial_j v_i) + \pi_{ij}. \quad (22.4)$$

In Eq. (22.4), we introduced the dynamic viscosity μ and a Gaussian white noise field π_{ij} characterized by

$$\begin{cases} \langle \pi_{ij}(\mathbf{r}, t) \rangle = 0, \\ \langle \pi_{ij}(\mathbf{r}, t) \pi_{kl}(\mathbf{r}', t') \rangle = 2k_B T \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \delta(\mathbf{r} - \mathbf{r}') \delta(t - t'), \end{cases} \quad (22.5)$$

where k_B is the Boltzmann constant and T is the temperature. We need to couple this equation with the energy equation to solve for velocity, pressure, and temperature, as, e.g., in Ref. [2]. Obviously, all the equations need to be coupled with suitable boundary conditions.

By using a suitable model for the particle collision (see for instance Ref. [4]), we can extend this model also to a group of particles. Using a Lagrangian description for the motion of the particles offers some gains in accuracy, but it becomes computationally unbearable if the number of particles start to be too large.

22.1.2 Continuum model for active suspensions

A continuum model for active suspensions was developed first by Saintilllian and Shelly [5, 6]. Here, we consider a version of this model for a suspension of active Brownian swimmers as described in Ref. [1]. In particular, the fluid motion is assumed to be incompressible and Newtonian. Then, the velocity and the pressure must satisfy

$$\begin{cases} \rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = -\nabla p + \mu \nabla^2 \mathbf{u} + n \sigma_0 \mathbf{D}, \\ \nabla \cdot \mathbf{u} = 0. \end{cases} \quad (22.6)$$

Equations (22.6) are the Navier–Stokes equations corrected by the addition of the nematic alignment tensor \mathbf{D} , with n the average number density of particles in the suspension and σ_0 the force dipole strength that arises from the first moment of the force distribution on the particle surface [7]. This term captures the effect of the force dipoles exerted by the swimmers on the fluid while self-propelling. Again, we need to couple Eqs. (22.6) with suitable boundary conditions. For instance, in a confined domain, it could be a no-slip condition, i.e., $\mathbf{u} = 0$ at the boundaries. These corrected Navier–Stokes equations must be coupled with other transport equations for (i) the particle concentration c , (ii) the polarization \mathbf{m} , and (iii) the already introduced tensor \mathbf{D} .

22.1.3 Active suspension flow using generalized Navier–Stokes equations

A model based on generalized Navier–Stokes equations for active suspensions particularly adapted to shear flows was developed by Slomka and Dunkel [8, 9, 10]:

$$\begin{cases} \rho(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = -\nabla p + \Gamma_0 \nabla^2 \mathbf{u} - \Gamma_2 \nabla^4 \mathbf{u} + \Gamma_4 \nabla^6 \mathbf{u}, \\ \nabla \cdot \mathbf{u} = 0, \end{cases} \quad (22.7)$$

where $\nabla^{2n} \equiv (\nabla^2)^n$ takes into account non-Newtonian effects if $n > 2$. The constant Γ_0 and Γ_4 must be positive to ensure asymptotic stability, while the constant Γ_2 can have both signs. If $\Gamma_2 = \Gamma_4 = 0$, we find the standard Navier–Stokes equations where the dynamics viscosity $\mu = \Gamma_0$.

22.2 Navier–Stokes equations

The models presented in the previous sections (Sections 22.1.1, 22.1.2, and 22.1.3) are three examples of descriptions suitable for different kinds of active systems (or different flow configurations). They have something in common though: they are all based on the incompressible Newtonian Navier–Stokes equations:

$$\begin{cases} \rho(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = -\nabla p + \mu \nabla^2 \mathbf{u} \\ \nabla \cdot \mathbf{u} = 0. \end{cases} \quad (22.8)$$

The *momentum equation* (Eq. (22.8)) is a *transport equation*. The term $\partial_t \mathbf{u}$ is the *transient term*, $\mathbf{u} \cdot \nabla \mathbf{u}$ is the *advective term* (or *convective term*), ∇p is the *pressure term* (which is a term representing momentum sources and sinks), and $\mu \nabla^2 \mathbf{u}$ is the *viscous diffusion term*. The derivation of Eqs. (22.8) is out of the scope of this Chapter but the interested reader is referred to Ref. [11].

The list of models presented here is not exhaustive. For instance, we did not present all the models based on the Toner–Tu equations [12, 13], which are possible also to supplement with a Swift–Hohenberg-like term [3]. However, also in these approaches based on the Toner–Tu model common for active nematic fluids, terms typical of the Navier–Stokes equations (e.g., advection or viscous diffusion) are present. Briefly, the discretization of a Navier–Stokes-type equation will be the minimum common factor of all the available models of active fluids from a continuum mechanics point of view.

22.2.1 Non-dimensionalization

We can introduce the following non-dimensional quantities:

$$\mathbf{u}^* = \mathbf{u}/U, \quad \mathbf{x}^* = \mathbf{x}/L, \quad t^* = tU/L, \quad \text{and} \quad p^* = p/(\mu U/L), \quad (22.9)$$

where U is the reference speed and L is the reference length. There are other possible references scales for the pressure but the viscous pressure $\mu U/L$ is the most suitable for most of active matter applications. By substituting Eqs. (22.9) into Eqs. (22.8), we obtain the non-dimensional Navier–Stokes equations:

$$\text{Re}(\partial_t^* \mathbf{u}^* + \mathbf{u}^* \cdot \nabla^* \mathbf{u}^*) = -\nabla^* p^* + \nabla^{*2} \mathbf{u}^*, \quad (22.10)$$

$$\nabla^* \cdot \mathbf{u}^* = 0, \quad (22.11)$$

where we introduced the *Reynolds number* $\text{Re} = \frac{\rho UL}{\mu}$, which is the ratio of the inertial to the viscous forces. If the viscous forces are much stronger than the inertial ones, we can assume that $\text{Re} = 0$ to obtain the *Stokes flow equations* (or *creeping flow equations*):

$$\begin{aligned} -\nabla^* p^* + \nabla^{*2} \mathbf{u}^* &= 0, \\ \nabla^* \cdot \mathbf{u}^* &= 0. \end{aligned} \quad (22.12)$$

Stokes flow equations are widely used in microfluidics where the pressure gradient is counterbalanced by viscous diffusion. From now on, we will work on the non-dimensional Navier–Stokes equations, Eqs. (22.10)–(22.11), but neglecting the asterisks.

22.3 Discretization of transport equation via the finite volume method

Our goal is to arrive at the discretization of the incompressible Navier–Stokes equations (Eqs. (22.8)). To do so, we prefer to start working on the differential form of the transport equation for the generic transported quantity ϕ :

$$\frac{\partial(\rho\phi)}{\partial t} + \nabla \cdot (\rho\phi\mathbf{u}) = \nabla \cdot (\Gamma\nabla\phi) + S_\phi, \quad (22.13)$$

where Γ is the diffusion coefficient relative to the quantity ϕ . On the left hand side, we have the rate of increase of ϕ in the fluid element $\frac{\partial(\rho\phi)}{\partial t}$ and the net rate of flow of ϕ out of the fluid element $\nabla \cdot (\rho\phi\mathbf{u})$. On the right hand side, we have the rate of increase of ϕ due to diffusion $\nabla \cdot (\Gamma\nabla\phi)$ and the rate of increase due to sources (that can be due to the surface or body forces).

22.3.1 Finite volume method

We will now use the *finite volume method* (FVM) to discretize Eq. (22.13). We chose FVM because it is widely used in computational fluid dynamics in commercial (e.g., Ansys Fluent) and open-source software (e.g., OpenFoam). The interested reader is referred to Refs. [14, 15, 16] for a more comprehensive dissertation on this numerical method. Other possible discretizations could make use of *finite difference* or *finite elements*.

The key FVM step is to integrate Eq. (22.13) over a three-dimensional control volume V :

$$\int_V \frac{\partial(\rho\phi)}{\partial t} dV + \int_V \nabla \cdot (\rho\phi\mathbf{u}) dV = \int_V \nabla \cdot (\Gamma\nabla\phi) dV + \int_V S_\phi dV. \quad (22.14)$$

We can use the Gauss divergence theorem to rewrite the integral representing the convective and diffusive terms over the bounding surface A of the control volume:

$$\int_V \frac{\partial\rho\phi}{\partial t} dV + \int_A \hat{\mathbf{n}} \cdot (\rho\phi\mathbf{u}) dA = \int_A \hat{\mathbf{n}} \cdot (\Gamma\nabla\phi) dA + \int_V S_\phi dV. \quad (22.15)$$

Let us describe shortly the four terms of Eq. (22.15):

- $\int_V \frac{\partial(\rho\phi)}{\partial t} dV$ represents the rate of change of the total amount of ϕ in the control volume.
- In the term $\int_A \hat{\mathbf{n}} \cdot (\rho\phi\mathbf{u}) dA$, we have $\hat{\mathbf{n}} \cdot (\rho\phi\mathbf{u})$ that is the advective flux of ϕ due to the fluid flowing along the outward normal component $\hat{\mathbf{n}}$. Thus, the second term on the left hand side is the net rate of decrease of ϕ in the control volume due to advection.
- To understand the first term on the right hand side ($\int_A \hat{\mathbf{n}} \cdot (\Gamma\nabla\phi) dA$), we need first to remember that a diffusive flux is positive in the direction of a negative gradient (i.e., heat conducted from positive to negative temperature). For this reason, the $\hat{\mathbf{n}} \cdot (-\Gamma\nabla\phi)$ is the diffusive flux along the outward normal vector, i.e., out of the control volume. This means that $-\hat{\mathbf{n}} \cdot (-\Gamma\nabla\phi)$ can be seen as a positive diffusive flux along the inward normal vector (i.e., inside of the control volume). In conclusion, the first term of the right-hand side represents the net rate of increase of fluid property ϕ of the fluid element due to diffusion.
- $\int_{CV} S_\phi dV$ represents the rate of increase of ϕ as a result of sources/sinks inside the control volume.

Note that in this section we assume that the velocity field \mathbf{u} is somehow known. Otherwise, it would have been necessary to couple the transport equation with the Navier–Stokes equations (Eqs. (22.8)).

[Shouldn't the various P be consistently upright or italic in the following text of the Chapter?]

Advection–diffusion

Here, we focus on the steady state of Eq. (22.15):

$$\int_A \hat{\mathbf{n}} \cdot (\rho\phi\mathbf{u}) dA = \int_A \hat{\mathbf{n}} \cdot (\Gamma\nabla\phi) dA + \int_{CV} S_\phi dV. \quad (22.16)$$

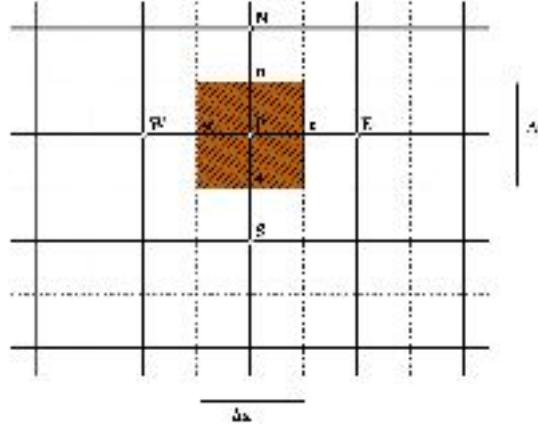


Figure 22.2: **Control volume in the finite volume method.** The 2D control volume used to discretize Eq. (22.16). The control volume P under examination in the figure has four neighbor control volumes defined with capital letters: northern N , southern S , eastern E , and western W neighbors. The faces of the control volume are defined with lowercase letters: northern n , southern s , eastern e , and western w faces. The cross-sectional areas of the control volume are defined as $A_e = A_w = \Delta y$ and $A_n = A_s = \Delta x$.

This equation needs to be coupled with the continuity equation

$$\nabla \cdot \mathbf{u} = 0. \quad (22.17)$$

First of all, we need to define the control volume. We work using the two-dimensional control volume centered in P shown in Fig. 22.2. The analyzed control volume has four neighbor control volumes defined with capital letters: northern N , southern S , eastern E , and western W neighbors. The faces of the control volume are defined with lowercase letters: northern n , southern s , eastern e , and western w faces. We define the cross-sectional areas of the control volume as $A_e = A_w = \Delta y$ and $A_n = A_s = \Delta x$.

Let us integrate Eqs. (22.16)-(22.17) over the control volume to obtain

$$(\rho u A \phi)_e - (\rho u A \phi)_w + (\rho v A \phi)_n - (\rho v A \phi)_s = \left(\Gamma A \frac{\partial \phi}{\partial x} \right)_e - \left(\Gamma A \frac{\partial \phi}{\partial x} \right)_w + \left(\Gamma A \frac{\partial \phi}{\partial y} \right)_s - \left(\Gamma A \frac{\partial \phi}{\partial y} \right)_n + \bar{S} \Delta V,$$

and

$$(\rho u A)_e - (\rho u A)_w + (\rho v A)_n - (\rho v A)_s = 0, \quad (22.18)$$

where ΔV is the volume of the control volume in question, and \bar{S} is the average value of the source S over the control volume. We can assume that (i) Γ and ρ are constant along the domain and (ii) the grid is uniform (i.e., all the control volumes have the same size). In this manner, we can write

$$\begin{aligned} & \rho \Delta y u_e \phi_e - \rho \Delta y u_w \phi_w + \rho \Delta x u_n \phi_n - \rho \Delta x u_s \phi_s = \\ &= \Gamma \Delta y \left(\frac{\partial \phi}{\partial x} \right)_e - \Gamma \Delta y \left(\frac{\partial \phi}{\partial x} \right)_w + \Gamma \Delta x \left(\frac{\partial \phi}{\partial y} \right)_n - \Gamma \Delta x \left(\frac{\partial \phi}{\partial y} \right)_s + \bar{S} \Delta V. \end{aligned} \quad (22.19)$$

Now, we have to decide how to approximate the values of ϕ and its derivatives at the faces. In this Chapter, we use the central differencing scheme, which corresponds to a linear interpolation. While for the diffusion term this scheme is perfectly suitable, for the convective terms it depends on the system and the parameters. We will briefly discuss this later. Therefore, by using central differencing, we can approximate the values of ϕ at the faces as:

$$\phi_e = \frac{\phi_E + \phi_P}{2}, \quad \phi_w = \frac{\phi_P + \phi_W}{2}, \quad \phi_n = \frac{\phi_N + \phi_P}{2}, \quad \text{and} \quad \phi_s = \frac{\phi_P + \phi_S}{2}, \quad (22.20)$$

and the fluxes as

$$\left(\frac{\partial \phi}{\partial x}\right)_e = \frac{\phi_E - \phi_P}{\Delta x}, \quad \left(\frac{\partial \phi}{\partial x}\right)_w = \frac{\phi_P - \phi_W}{\Delta x}, \quad \left(\frac{\partial \phi}{\partial y}\right)_n = \frac{\phi_N - \phi_P}{\Delta y}, \quad \text{and} \quad \left(\frac{\partial \phi}{\partial y}\right)_s = \frac{\phi_P - \phi_S}{\Delta y}. \quad (22.21)$$

Furthermore, to simplify the equation, we can introduce the convective fluxes F and the diffusive fluxes D as

$$F_e = \rho \Delta y u_e, \quad F_w = \rho \Delta y u_w, \quad F_n = \rho \Delta x v_n, \quad \text{and} \quad F_s = \rho \Delta x v_s, \quad (22.22)$$

and

$$D_e = D_w = \frac{\Gamma \Delta y}{\Delta x} \quad \text{and} \quad D_n = D_s = \frac{\Gamma \Delta x}{\Delta y}. \quad (22.23)$$

Finally, we need to treat the source term. In practical situations, S could be a function of the dependent variable ϕ . In these cases, FVM approximates the source term in a linear form:

$$\bar{S} \Delta V = S_u + S_p \phi_P. \quad (22.24)$$

Replacing Eqs. (22.20)-(22.24) in Eqs. (22.18) and (22.19), we obtain

$$F_e - F_w + F_n - F_s = 0, \quad (22.25)$$

$$\begin{aligned} \frac{F_e}{2}(\phi_P + \phi_E) - \frac{F_w}{2}(\phi_W + \phi_P) + \frac{F_n}{2}(\phi_P + \phi_N) - \frac{F_s}{2}(\phi_S + \phi_P) = \\ = D_e(\phi_E - \phi_P) - D_w(\phi_P - \phi_W) + D_n(\phi_N - \phi_P) - D_s(\phi_P - \phi_S) + S_u + S_p \phi_P. \end{aligned} \quad (22.26)$$

We can rearrange Eq. (22.26) as

$$\begin{aligned} \left[\left(D_w - \frac{F_w}{2} \right) + \left(D_e + \frac{F_e}{2} \right) + \left(D_s - \frac{F_s}{2} \right) + \left(D_n + \frac{F_n}{2} \right) - S_p \right] \phi_P = \\ = \left(D_w + \frac{F_w}{2} \right) \phi_W + \left(D_e - \frac{F_e}{2} \right) \phi_E + \left(D_s + \frac{F_s}{2} \right) \phi_S + \left(D_n - \frac{F_n}{2} \right) \phi_N + S_u. \end{aligned} \quad (22.27)$$

We can also let appear the continuity equation (Eq. (22.25)) in the coefficient of ϕ_P :

$$\begin{aligned} \left[\left(D_w + \frac{F_w}{2} \right) + \left(D_e - \frac{F_e}{2} \right) + \left(D_s + \frac{F_s}{2} \right) + \left(D_n - \frac{F_n}{2} \right) + (F_e - F_w + F_n - F_s) - S_p \right] \phi_P = \\ = \left(D_w + \frac{F_w}{2} \right) \phi_W + \left(D_e - \frac{F_e}{2} \right) \phi_E + \left(D_s + \frac{F_s}{2} \right) \phi_S + \left(D_n - \frac{F_n}{2} \right) \phi_N + S_u. \end{aligned} \quad (22.28)$$

The notation of Eq. (22.28) can be simplified by

$$a_P \phi_P = a_W \phi_W + a_E \phi_E + a_N \phi_N + a_S \phi_S + S_u, \quad (22.29)$$

where

$$a_W = D_w + \frac{F_w}{2}, \quad a_E = D_e - \frac{F_e}{2}, \quad a_S = D_s + \frac{F_s}{2}, \quad a_N = D_n - \frac{F_n}{2},$$

$$a_P = a_W + a_E + a_S + a_N + (F_e - F_w + F_n - F_s) - S_p.$$

The unknowns of Eq. (22.29) are ϕ_P , ϕ_E , ϕ_W , ϕ_N , and ϕ_S . Since we need to write a similar equation for every grid point of the domain, we have transformed a partial differential equation (PDE) like the steady general transport equation (i.e., the steady state of Eq. (22.13)) into a system of M -algebraic equations, where M is the number of grid points. If M is large, the solution of this system can be computationally expensive. There are several methods to solve it, such as the *Tri-Diagonal Matrix Algorithm* (TDMA or Thomas algorithm), the *Gauss-Seidel technique*, or the *Multigrid technique*. A detailed treatment of these (and other) algorithms are out of scope of this Chapter. The interested reader is referred to Ref. [14] to have more information. Libraries (or also simple commands) to use this kind of algorithms are available in the most common computational languages.

In most practical applications, however, an iterative algorithm to solve the system of M -algebraic equations must be implemented. The convergence of the numerical solution is not ensured for all the numerical methods. A numerical scheme must be *bounded* to avoid divergence and converge to an accurate solution. More details can be found in Section 22.5. One way to check whether a numerical scheme is bounded is to check whether the coefficients of ϕ_P , ϕ_E , ϕ_W , ϕ_N , and ϕ_S in the equivalent of Eq. (22.29) have the same sign. For central differencing, this is clearly not guaranteed. For instance, assuming u and v are positive and, as a consequence, a_W and a_S are positive as well, a_E and a_N are positive only if $D_e > \frac{F_e}{2}$ and $D_n > \frac{F_n}{2}$, respectively. We define as Péclet number Pe the ratio between convective and diffusive fluxes $\text{Pe} = \frac{F}{D}$. Then, the central differencing scheme is bounded only if the Pe at the eastern and the northern face is $\text{Pe}_e < 2$ and $\text{Pe}_n < 2$. Since the diffusive flux D does depend on the grid size, a grid refinement could help in ensuring boundedness. However a too fine grid means too expensive simulations. Furthermore, central differencing does not embed transportiveness at all (see Section 22.5). *Transportiveness* is an important feature for a numerical scheme if the advective term is important. In conclusion, if advection is strong in our flow system, it is preferable to discretize the convective terms with other schemes, such as *upwind* (see Problem 22.9), *linear upwind*, or *QUICK* [14], depending on the degree of accuracy that we need in our simulations. However, many applications in active matter do not present a large Pe, making central differencing a suitable numerical scheme for both convection and diffusion terms.

Boundary conditions

In the previous section, we have shown how it is possible to transform the steady general transport equation (i.e., the steady state of Eq. (22.13)) into a system of algebraic equations. However, the generic equation for the node P , Eq. (22.29), is valid just for internal nodes. If the node is close to a boundary (like in Fig. 22.3), we need to take into account the presence of the boundary in our final algebraic equation for the boundary node P . Two main kinds of boundary conditions are common in fluid mechanics:

- Prescribed distribution of the boundary value of ϕ , i.e., $\phi = \phi_B$. It corresponds to a *Dirichlet boundary condition*.
- Prescribed distribution of the total (convective plus diffusive) flux at the boundary. It corresponds to a *Robin boundary condition*.

Let us work on the boundary control volume depicted in Fig. 22.3, where the east face of the cell coincides with the boundary B . The neighboring control volume lies in the boundary. In reality, this node does not exist.

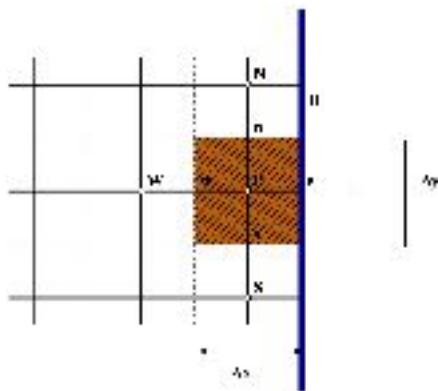


Figure 22.3: **A boundary control volume.** Instead of the nodal neighbor E , we have a boundary denoted with B . In this case, we cannot use the approximation used for the internal node of Fig. 22.2 since we do not have a neighboring node E to approximate the flux. An expression can still be deduced from the relevant boundary condition.

The fluxes at the boundary e can be written as

$$(\rho u \phi)_e - \Gamma \left(\frac{\partial \phi}{\partial x} \right)_e , \quad (22.30)$$

where evidently the cell face e coincides with the boundary. We cannot use the approximation used for internal nodes since we do not have a neighboring node E to approximate the flux. However, an expression can easily be deduced from the relevant boundary condition.

- **Prescribed ϕ .** In this case, we need to modify the usual expression considering that the cell face and the boundary coincide, thus:

$$(\rho u \phi)_e - \Gamma \left(\frac{\partial \phi}{\partial x} \right)_e = \rho u_B \phi_B - \Gamma \left(\frac{\phi_B - \phi_P}{\Delta x / 2} \right) . \quad (22.31)$$

Note that in this expression we introduced also the velocity boundary condition u_B . In this scenario we introduce a source term $S_u + S_p \phi_P$, in which $S_u = (\frac{2\Gamma}{\Delta x} - \rho u_B) \Delta y = 2D_e - F_e$ and $S_p = \frac{2\Gamma}{\Delta x} = 2D_e$.

- **Prescribed total flux.** Since we know the total flux at the cell face e , the complete expression for the flux, Eq. (22.30), is simply replaced by the specified value. Note that the boundary value of ϕ_B does not enter in the solution directly. However, we can evaluate it after the solution, using the expression for the flux and working out the value of ϕ_B .

Transient

In the previous sections, we analyzed the steady state of a generic transport equation (Eq. (22.13)). Now, it is time to work on the transient term $\frac{\partial(\rho\phi)}{\partial t}$. We can approximate the transient term by using the FVM method:

$$\int_{CV} \frac{\partial \rho \phi}{\partial t} dV \approx \rho \frac{d\phi_P}{dt} \Delta V . \quad (22.32)$$

However, now we need to discretize the time derivative. There is a large variety of schemes to use, often based on finite-difference schemes. Let us first write Eq. (22.13) as

$$\rho \frac{d\phi_P}{dt} = f ,$$

where f comes out from the steady discretized transport equation, Eq. (22.29):

$$a_P \phi_P - a_W \phi_W - a_E \phi_E - a_N \phi_N - a_S \phi_S - S_u = 0 . \quad (22.33)$$

Since f depends on $\phi_P, \phi_W, \phi_E, \phi_S, \phi_N$, and t , we can write $f = f(\phi_I, t)$. In general, we can discretize Eq. (22.32) as

$$\rho \frac{\phi_P^{n+1} - \phi_P^n}{\Delta t} = (1 - \theta) f(\phi_I, t_n) + \theta f(\phi_I, t_{n+1}) , \quad (22.34)$$

where $0 \leq \theta \leq 1$. However, we have three main options for choosing the value of θ :

1. $\theta = 0$. In this case, we evaluate all the values of ϕ at the old time t_n . This is called *explicit Euler method*. This is the simplest way to discretize the temporal derivative. The main advantage is that the evaluation of ϕ_P^{n+1} does not depend on the neighboring values at time t_{n+1} . This means that we can obtain the value of ϕ_P^{n+1} by straightforward substitution of the values ϕ_{nb}^n on the right hand side. We can develop a simple algorithm for the explicit Euler method, starting from the known variable field at t_n and advancing with time Δt to the next time level t_{n+1} :
 - (a) Assemble all the coefficients a_I , and the source terms S_u and S_p .
 - (b) Scan the grid and compute ϕ_P^{n+1} at each point. The order in which each point is targeted is irrelevant, as the equation for ϕ_P^{n+1} does not depend on the neighboring values at t_{n+1} .
 - (c) Advance the time from n to $n + 1$ by adding Δt .

The main setback of the explicit Euler method is that stable solutions can only be obtained under strict conditions (see Problem 22.4).

2. $\theta = 1$. In this case, we evaluate all the values of ϕ at the old time t_{n+1} . We call this scheme the *implicit Euler method*. Each ϕ_p^{n+1} depends on its neighbors at the new time level ϕ_{nb}^{n+1} . This results in the impossibility of writing an explicit expression for ϕ_p^{n+1} in terms of known quantities. Rather, each ϕ_p^{n+1} is related implicitly to its neighbor values ϕ_{nb}^{n+1} . This means that a set of algebraic equations must be solved to determine the values of ϕ_p^{n+1} .
3. $\theta = \frac{1}{2}$. We use both the old and the new time, equally weighted, to evaluate the values of ϕ . This is called the *Crank-Nicolson scheme*. This is the most accurate between the three presented scheme (it is second order accurate, while implicit and explicit Euler are only first order, see Subsection 22.5.1). The Crank-Nicolson algorithm is unconditionally stable [17] but boundedness is not ensured if the time step is too large (see Problem 22.4).

In computational fluid dynamics, we need to respect the following condition:

$$\frac{\mathcal{U} \Delta t}{\Delta x_{max}} \leq 1 , \quad (22.35)$$

where \mathcal{U} is the maximum propagation speed of the information, Δt is the time step, and Δx_{max} is the maximum distance in the domain between grid points. Eq. (22.35) is called the *Courant, Friedrichs and Lewy (CFL) condition* and derives from the principle of domain of dependence, i.e., the numerical solution in each time step must not advance further in space than the physical solution over the same time increment. If the CFL condition, which couples the numerical and the physical solutions, is not satisfied an explicit discretization of the governing equation is not stable. Even if the numerical schemes are unconditionally stable (so they are allowed to exceed the CFL condition), violation of Eq. (22.35) may lead to unphysical results.

22.4 Discretization of Navier–Stokes

The incompressible Navier–Stokes momentum equation (Eq. (22.10)) is a transport equation of momentum. However, there are two main complications in Eq. (22.10) with respect to the discretization shown in Subsection 22.3.1: (i) the non-linearity in the advection term ($\mathbf{u} \cdot \nabla \mathbf{u}$) and (ii) the role of the pressure.

The non-linearity in the advection term results from the convective fluxes F , which depend on the unknowns $\mathbf{u}(u, v)$. There is then the need of initially “guessing” the values of the velocity in the convective fluxes and, through an iterative procedure, reach the final solution by solving both the momentum and the continuity equations (Eqs. (22.10) and (22.11)). There are several schemes available in the literature but before briefly presenting one of the most used (i.e., the *SIMPLE algorithm* [18]), we must talk about the second challenge inherent in discretizing the Navier–Stokes equations: the pressure gradient (∇p) and its role.

22.4.1 Pressure gradient

The main difficulty arising if the pressure field is not a priori known lies in the fact that we have no explicit equation for the pressure (for incompressible flows). The pressure merely appears as a gradient in the momentum equation. This means that the pressure field has to be chosen in such a way that the solution of the momentum equation gives a velocity field satisfying both the momentum equation and the continuity equation.

The first task is to decide where to store the pressure. In general, the most obvious choice would be to store all the variables (velocities, pressure, etc.) at the nodal points, in the center of the control volume. We call *collocated grid* this kind of arrangement (see Fig. 22.4a). Unfortunately, a collocated grid presents some setbacks. In particular, if we use the central differencing scheme to approximate the pressure gradient, the dependency of the pressure node at the center falls out (see Problem 22.5). This could bring to oscillations and instabilities in the numerical scheme. There are ways to overcome this issue [16] but are out of the scope of this Chapter.

However, the easiest (and most common) solution is to apply a *staggered grid* in which the pressure values (and other scalar variables) are stored in the center of the control volume, while the u -velocities and v -velocities

are stored in the vertical and horizontal faces of the control volume, respectively (see Fig. 22.4b). In a staggered grid implementation, for instance, the approximation of the pressure gradient in the horizontal momentum equation can directly use the pressure values at the western and eastern faces since it is where they are defined. Furthermore the u - and v -control volumes are built directly around the velocity components. Thus, they are staggered by half a cell with respect to the control volume of the pressure (and the other scalar variables) as it is possible to observe in Fig. 22.5.

22.4.2 Discretization of the Navier–Stokes momentum equation

First of all, we need to discretize the momentum (Eqs. (22.10)). For the u -momentum equation, we can derive

$$\left(\rho \frac{du}{dt} \Delta x \Delta y + a_p^u \right) u_P = a_N^u u_N + a_S^u u_S + a_E^u u_E + a_W^u u_W + b_P^u - (p_P - p_W) \Delta y , \quad (22.36)$$

where the coefficients are

$$\begin{aligned} a_N^u &= \left(D_n - \frac{F_n}{2} \right) = \left(\frac{1}{\Delta y} - \text{Re} \frac{v_n}{2} \right) \Delta x , & a_E^u &= \left(D_e - \frac{F_e}{2} \right) = \left(\frac{1}{\Delta x} - \text{Re} \frac{u_e}{2} \right) \Delta y , \\ a_S^u &= \left(D_s + \frac{F_s}{2} \right) = \left(\frac{1}{\Delta y} - \text{Re} \frac{v_s}{2} \right) \Delta x , & a_W^u &= \left(D_w + \frac{F_w}{2} \right) = \left(\frac{1}{\Delta y} - \text{Re} \frac{u_w}{2} \right) \Delta y , \\ a_p^u &= a_N^u + a_S^u + a_E^u + a_W^u , & b_P^u &= S_p^u \Delta x \Delta y . \end{aligned}$$

If we use the central differencing scheme also for the convective mass fluxes, we obtain, for instance,

$$\begin{cases} F_n = \frac{\rho}{2} (v_N + v_P) \Delta x , \\ F_e = \frac{\rho}{2} (u_E + u_P) \Delta y . \end{cases}$$

Similarly, we can obtain the discretized version of the y -momentum equation:

$$\left(\rho \frac{dv}{dt} \Delta x \Delta y + a_p^v \right) u_P = a_N^v v_N + a_S^v v_S + a_E^v v_E + a_W^v v_W + b_P^v - (p_P - p_W) \Delta y , \quad (22.37)$$

where the expressions of the coefficients are left to be found by the interested reader (see Problem 22.6).

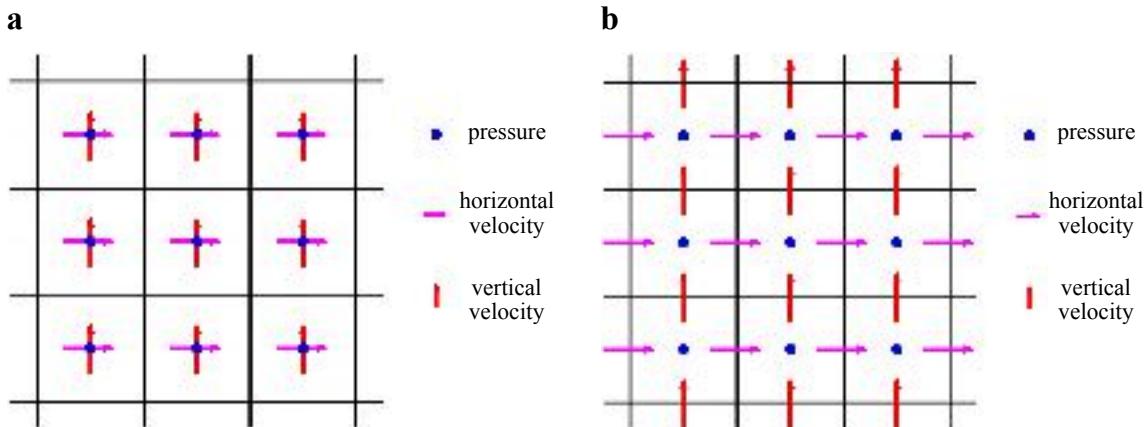


Figure 22.4: **Discretization of the Navier–Stokes equations: collocated grid and staggered grid.** Schematic of (a) a collocated grid and (b) a staggered grid. We show the storage of the two velocity components u and v and of the pressure p . In the collocated grid, all the variables are stored at the nodal points. In a staggered grid, the different variables are stored at different points of the control volume. The pressure values (and other scalar variables) are stored in the center of the control volume, while the u and v velocities are stored in the vertical and horizontal faces of the control volume, respectively.

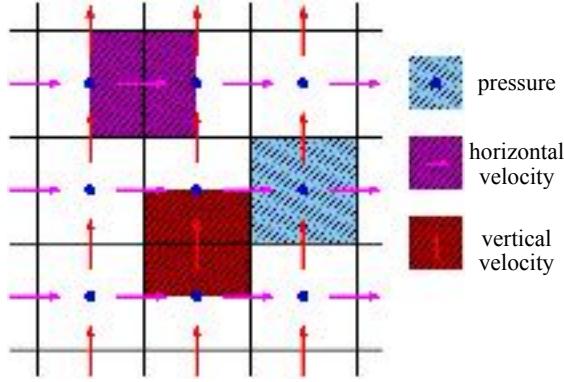


Figure 22.5: **Control volumes in a staggered grid.** The three control volumes (pressure p , vertical velocity u , and horizontal velocity v) used in a staggered grid to solve the Navier–Stokes equations.

22.4.3 Discretization of the continuity equation

The approximation of the continuity equation, Eq. (22.11), through the FVM is straightforward. We use the control volume around the scalar variables (see Fig. 22.5). After some calculations, we obtain

$$(u_E - u_P)\Delta x + (v_N - v_P)\Delta y = 0. \quad (22.38)$$

22.4.4 SIMPLE algorithm

The *Semi-Implicit Method of Pressure Linked Equations* (SIMPLE) algorithm [18] has two different steps: (i) the predictor step and (ii) the corrector step. We leave the detailed discretization to the reader (see Problem 22.6), but here we will give the general details.

In the predictor step, we solve the momentum equations, Eqs. (22.36)-(22.37), to obtain u_P^* and v_P^* by using the pressure field and coefficients evaluated at the previous time step or previous iteration, whichever is the most recent. This predicted velocity field closely satisfies the momentum equations but it will not necessarily satisfy also the continuity equation Eq. (22.38).

To solve this issue we need a corrector step. Let us introduce a pressure correction

$$p'_I = p_I^{n+1} - p_I^n. \quad (22.39)$$

After some calculations, we can write expressions for the velocity corrections u' and v'

$$\begin{aligned} u'_{P'} &= u_P^{n+1} - u_P^* &= -d_P^u (p'_{P'} - p'_{W'}) , \\ v'_{P'} &= v_P^{n+1} - v_P^* &= -d_P^v (p'_{P'} - p'_{S'}) , \end{aligned} \quad (22.40)$$

and similarly for the neighbor nodes, as for example:

$$\begin{aligned} u'_{E'} &= u_E^{n+1} - u_E^* &= -d_E^u (p'_{E'} - p'_{P'}) , \\ v'_{N'} &= v_N^{n+1} - v_N^* &= -d_N^v (p'_{N'} - p'_{P'}) . \end{aligned} \quad (22.41)$$

The d coefficients are given by

$$\begin{aligned} d_P^u &= \frac{r\Delta y}{\rho \frac{du}{dt} \Delta x \Delta y + a_P^u} , & d_E^u &= \frac{r\Delta y}{\rho \frac{du}{dt} \Delta x \Delta y + a_E^u} , \\ d_P^v &= \frac{r\Delta x}{\rho \frac{dv}{dt} \Delta x \Delta y + a_P^v} , & d_N^v &= \frac{r\Delta x}{\rho \frac{dv}{dt} \Delta x \Delta y + a_N^v} . \end{aligned} \quad (22.42)$$

Note than in the SIMPLE algorithm the dependence on the neighboring nodes is neglected. This is the major assumptions of SIMPLE and it could slow down the convergence speed but it does not infer with the accuracy of the solution.

In Eqs. (22.40), the only unknowns are the new velocity u^{n+1} and v^{n+1} , and the pressure correction p' . The new velocities can be eliminated by inserting Eqs. (22.40) into the continuity equation (Eq. (22.38)) to obtain a Poisson-type equation for the pressure correction:

$$a_P^p p'_P = a_N^p p'_N + a_S^p p'_S + a_E^p p'_E + a_W^p p'_W + b_P , \quad (22.43)$$

where

$$\begin{aligned} a_N^p &= d_N^v, & a_S^p &= d_N^v, & a_E^p &= d_E^u, & a_W^p &= d_W^u, \\ a_P^p &= a_N^p + a_S^p + a_W^p + a_E^p, & b_P &= -(u_E^* - u_P^*) - (v_N^* - v_P^*). \end{aligned}$$

From Eq. (22.43), we can find the pressure correction since all the coefficients are known. We can find then the velocity correction from Eqs. (22.40) that can be used to correct the velocity field (that will automatically satisfies the continuity equation). However, it is not given that the corrected velocity field satisfies the momentum equations as well. For this reason, we have to re-evaluate the coefficients of both momentum and continuity equations. In practice, we need to repeat the entire procedure until reaching a velocity field that satisfies both continuity and momentum equations (at least to a reasonable level).

Finally, we summarize the main parts of the SIMPLE algorithm:

1. Initialize the velocity and the pressure fields.
2. Calculate the momentum equations coefficients (in particular d^u and d^v) and solve for the predictor velocity field u^* and v^* using the latest available pressure p^n .
3. Calculate the coefficients for the pressure correction (p') equation (Eq. (22.43)) and solve for p' . With the pressure correction, we can obtain p^{n+1} since $p^{n+1} = p^n + p'$.
4. Correct the velocity field by means of the pressure correction (Eqs. (22.40)) and then the continuity equation is satisfied.
5. Calculate the coefficients of the other scalar equations connected to the Navier–Stokes equations (energy, concentration, etc.) and solve for the new values.

The sequence 2 to 5 must be repeated until reaching convergence. SIMPLE is used by many CFD practitioners (and commercial and open source softwares as well) and it is observed to be quite robust despite the lack of a formal proof of its convergence. In general, a necessary (but not sufficient) condition to achieve convergence is to have a decrease in the pressure correction while iterating.

There are several variants of SIMPLE available in the literature (see Ref. [14] for further details). For instance, the *SIMPLEC algorithm* does not neglect the dependence on the neighboring nodes. This will increase the convergence speed. The *SIMPLER algorithm* adds a preliminary step, involving the so-called pseudo-velocities, to have a better initial value for the pressure to find the predicted velocities u^* and v^* . More in general, the SIMPLE algorithm has originally been put forward for steady state problems, while PISO could be a better choice for unsteady flows.

22.5 Properties of numerical schemes

In this section, we present the most important numerical properties of numerical schemes, i.e., accuracy, stability, boundedness, conservativeness, and transportiveness.

22.5.1 Accuracy

Numerical solutions of fluid flows are only approximate solutions. To measure the degree of accuracy of a numerical method, we should compare how much the approximated solution is close to an analytical one. However, in many cases it is not possible to have an analytical solution. For this reason, we have to find another method to systematically define the accuracy of a numerical scheme.

Let us use as an example the discretization of the term $\frac{d\phi}{dx}$. We need to approximate its value in P (i.e. the center of the CV shown in Fig. 22.2) given by

$$\left(\frac{d\phi}{dx} \right)_P = F(\phi_I, \Delta x) + O(\Delta x^k), \quad (22.44)$$

where F is a function depending on the values ϕ in some nodes (which nodes it depends on the numerical scheme) and the grid size along the x -direction (Δx), while $O(\Delta x^k)$ represents the truncated term. In practice, the power k in the truncated term governs the rate at which the error tends to zero as the grid is refined, and is called the *order* of the difference approximation. We define this power k as the degree of accuracy of a numerical scheme.

For the central differencing scheme, we can write (see Problem 22.1)

$$\left(\frac{d\phi}{dx} \right)_P = \frac{\phi_E - \phi_W}{2\Delta x} + O(\Delta x^2). \quad (22.45)$$

For this reason, the accuracy of the central differencing scheme is second order.

22.5.2 Stability

A numerical solution is said to be *stable* if it does not magnify the errors appearing during the course of the numerical process. In the case of transient flows, stability guarantees to reach a bounded solution (see Section 22.5.3) if, of course, the exact solution is bounded. If iterations are involved, a simple definition of a stable numerical method is a scheme which does not diverge. If complex boundary conditions or (worse) nonlinearities are present, it could be complicated to analyze the stability of a numerical method. However, for these complex cases, we can often investigate the stability of the method for linear problems without boundary conditions. In general, it was observed that the results obtained in this manner can be often used also for the more complex problems (but be careful since there are exceptions).

22.5.3 Boundedness

Numerical solutions should lie in proper bounds. Quantities that are always positive (such as density, viscosity, kinetic energy of turbulence) have to be always positive. Concentrations, moreover, must lie between 0 and 1. If sources/sinks are not applied, some equations (like the heat equation for instance) want that the minimum and the maximum values of the variables are located at the boundaries of the domain. Therefore, we would like ideally to have these constraints inherited by the numerical scheme. Schemes known to produce unbounded solutions could have problems in stability easily. If the discretization scheme does not satisfy boundedness, it is highly possible that the solution does not converge at all, or, if it does, that it contains oscillations.

There are two methods to control the boundedness of a numerical method:

- Check if the *Scarborough's criterion* is satisfied. This criterion states a sufficient condition for a convergent iterative method:

$$\frac{\sum |a_{nb}|}{|a_P|} \begin{cases} \leq 1 & \text{at all nodes} \\ < 1 & \text{at one node at least} \end{cases} \quad (22.46)$$

This corresponds to have a diagonal dominant matrix of the coefficients a (see Eq. (22.29) for instance).

- Control whether all the coefficients of the discretized equations have the same sign (usually positive). Physically, this implies that an increase in the variable ϕ at one node should result in an increase in ϕ at its neighbors.

22.5.4 Conservativeness

In fluid systems, the equations to be solved are conservation laws. Ideally, the numerical schemes should respect these laws. In absence of sources/sinks at steady state, the amount of a quantity ϕ leaving a closed volume is equal to the amount of the same quantity going into the same volume. This is a fundamental property since it stabilizes a constraint on the solution error. The error can only wrongly distribute conserved quantities (such

as mass, momentum, or energy) over the domain. Non-conservative schemes can produce artificial sources and sinks. However, if these schemes are consistent and stable, they can still lead to correct solutions if the grid is very fine (sometimes not easy to guarantee though). A good feature of the FVM is that, if coupled with the strong conservation form of the equations, it guarantees conservativeness for each individual CV and for the solutions in the entire domain.

22.5.5 Transportiveness

The *transportiveness* of a numerical method measures how much the directionality of the flow is taken into account. Let us assume that our CV is represented in Fig. 22.2 and the velocity is positive and completely horizontal (i.e., $v = 0$). [Check this sentence] In this case, the value of the quantity ϕ will be influenced more by the western neighboring node (W) than by the eastern neighbor (E). If the velocity is negative instead, we will have the opposite case (i.e., E more influential than W). If a numerical scheme embeds this quality, we say that it has a good transportiveness. Flows with strong advection (i.e., with high Péclet number, see section 22.3.1) need to be treated with schemes with good transportiveness, while this property is less important if diffusion is dominant. The central differencing scheme does not take into account at all the flow direction. For this reason, it is not a suitable scheme for the convective terms if the Péclet number is large.

22.6 Problems

Problem 22.1: Accuracy of the central differencing scheme. Show that the accuracy of central differencing scheme is second order. [Hint: How will you express the value of a function ϕ in E and W if you know the value of ϕ in P by using Taylor series approximations? Look at the control volume in Fig. 22.2.]

Problem 22.2: Upwind scheme. If the Péclet number $\text{Pe} > 2$, the central differencing scheme is not a suitable choice anymore since it is not bounded and a good transportiveness is required in this case. The easiest solution is to apply the so-called *upwind scheme* (or *donor cell*). In this scheme, the values of the transported quantity ϕ , needed at the faces in the convective terms in Eq. (22.19), are approximated simply by the value of ϕ of the control volume upwards (downwards) if the velocity is positive (negative). For instance, if u is positive, ϕ_e is approximated by ϕ_P .

Write the equivalent of Eq. (22.29) using upwind instead of central differencing to discretize the convective terms.

Problem 22.3: Non-constant density and diffusivity coefficient, non-uniform grid. [Check this sentence] Obtain all the coefficients a , the convective and the diffusive fluxes (F and D) as we have done for Eq. (22.29) if (i) the density ρ and the diffusivity coefficient Γ are not constant along the domain and (ii) the grid is non-uniform.

Problem 22.4: Maximal time step for boundedness in different schemes for the general transport equation. Find what is the maximal allowed time step to ensure boundedness for the general transport equation (Eq. (22.13)) without sources/sinks for the following temporal schemes:

1. The explicit Euler scheme if only the diffusion term is present.
2. The Crank-Nicolson scheme if only the diffusion term is present.
3. The explicit Euler scheme if both advection and diffusion terms are present.

Problem 22.5: Checkerboard pressure. Assume that we have a checkerboard pressure field (i.e., the pressure is point-by-point oscillating between p_{\max} and p_{\min}). Calculate the pressure gradients $\frac{\partial p}{\partial x}$ and $\frac{\partial p}{\partial y}$ for two grids similar to those shown in Fig. 22.2 but (i) the first one is collocated and (ii) the second one is staggered (see Fig. 22.4).

Problem 22.6: SIMPLE algorithm step-by-step. In this problem, we want to obtain all the relevant equations for the SIMPLE algorithm step-by-step. Note that a staggered grid is used (Fig. 22.4b).

- a.** Starting from the non-dimensional Navier–Stokes equations, Eqs. (22.10)–(22.11), obtain Eqs. (22.36), (22.37), and (22.38), and all the relevant coefficients.
- b.** Find the predictor velocities u^* and v^* .
- c.** Write the velocities for the control volume centred in the point P for the iteration level $n + 1$ ($u_{P,n+1}$ and $v_{P,n+1}$) by using Eqs. (22.36) and (22.37).
- d.** Find the velocity correction u'_P and v'_P by introducing the pressure correction (Eq. (22.39)) and the d coefficients (Eqs. (22.42)). [Hint: To find the velocity correction, you should subtract the expressions for u^* and v^* found in the question **a** from the expressions for $u_{P,n+1}$ and $v_{P,n+1}$ found in the question **b**.]
- e.** Which term can we neglect in the expressions for u'_P and v'_P ? Neglect this term to obtain Eqs. (22.40).
- f.** Replace Eqs. (22.40) in the discretized continuity equation, Eq. (22.38), to find the pressure correction equation, Eq. (22.43).
- g.** In general, we use *under-relaxation* in SIMPLE to improve convergence. This means that we do not use directly the velocity field and the pressure obtained in the last time-step or iteration to solve the momentum equations to find \mathbf{u}^* and $\mathbf{u}_{P,n+1}$ as done in question **a** and **b**. The solution for u , v and p after each iteration are regarded as provisional and so we can write

$$\begin{cases} p_{P,n+1} = \alpha_p \hat{p}_{P,n+1} + (1 - \alpha_p) p_{P,n} \\ u_{P,n+1} = \alpha_u \hat{u}_{P,n+1} + (1 - \alpha_u) u_{P,n} \\ v_{P,n+1} = \alpha_v \hat{v}_{P,n+1} + (1 - \alpha_v) v_{P,n} \end{cases} \quad (22.47)$$

where \hat{p} , \hat{u} , and \hat{v} are the provisional solution, and α_p , α_u , and α_v are the under-relaxation factors for the pressure, horizontal and vertical velocities, respectively. The under-relaxation factors α_i can have values $0 \leq \alpha_i \leq 1$. Note that if they are larger than 1 they become over-relaxation factors, but this is warmly discouraged for SIMPLE. Assume that $\alpha_p = \alpha_u = \alpha_v$ and write again all the SIMPLE algorithm introducing under-relaxation. In particular, the expressions for the predicted velocities, the velocities for the iteration level $n + 1$ and the d coefficients must be changed.

Problem 22.7: Transport equation for local particle concentration: non-dimensionalization. The transport equation for the local particle concentration c from Ref. [6] is

$$\frac{\partial c}{\partial t^*} + \mathbf{u} \cdot \nabla^* c = d \nabla^{*2} c - U_0 \nabla \cdot \mathbf{m}, \quad (22.48)$$

where \mathbf{u} is the velocity, d is the diffusion coefficient, U_0 is the swimming velocity, and \mathbf{m} is the polarization capturing the mean direction of swimming.

- a.** Non-dimensionalize equation (22.48) by using L as length scale and T as time scale. [Hint: You have to introduce the swimming Péclet number $\text{Pe}_s = \frac{U_0 T}{L}$ and the parameter $\Lambda = \frac{d}{U_0^2 T}$.

Problem 22.8: Local particle concentration for zero polarization field. Start from the non-dimensional form of Eq. (22.48) found in Problem 22.7.

- a.** Assuming that the polarization field is zero, and \mathbf{u} is given by the Taylor-Green vortex

$$\mathbf{u}(u, v) = \left(\frac{1}{2} \cos x \sin z, \frac{1}{2} \sin x \cos z \right)$$

discretize the non-dimensional version of Eq. (22.48) in a squared domain $(0, 2\pi) \times (0, 2\pi)$. Use central differencing for both advection and diffusion terms and Euler implicit/explicit as temporal scheme.

- b.** Find the spatiotemporal evolution of the concentration c by choosing your favorite programming language (comparing the two different temporal schemes you used to discretize). Set $\Lambda = 0.1$ and $\text{Pe}_s = \sqrt{10}$. Use a uniform grid and choose the optimal values for the number of grid points $N_x = N_y = N$ and the time step Δt . The boundary condition are periodic in both directions and the initial condition is $c(x, z, 0) = \cos^2 x$.

Problem 22.9: Local particle concentration for non-zero polarization field. Start from the non-dimensional form of Eq. (22.48) found in Problem 22.7.

a. Let us assume that the polarization vector is

$$\mathbf{m}(m_x, m_y) = (\cos z \sin x, \sin z \cos x)$$

for (i) $\Lambda = 0.1$ and $\text{Pe}_s = \sqrt{10}$ and (ii) $\Lambda = 10$ and $\text{Pe}_s = \sqrt{\frac{1}{10}}$. Discretize the non-dimensional version of Eq. (22.48) using the same spatial discretization used in Problem 22.8 and choosing your preferred temporal scheme.

b. Show the spatiotemporal evolution of the concentration c by choosing your favorite programming language. Use a uniform grid and choose the optimal values for the number of grid points $N_x = N_y = N$ and the time step Δt . The boundary conditions are periodic in both directions and the initial condition is $c(x, z, 0) = \cos^2 x$.

Bibliography

- [1] M. Theillard and D. Saintillan. Computational mean-field modeling of confined active fluids. *Journal of Computational Physics*, 397:108841, 2019.
- [2] P. Gaspard and R. Kapral. The stochastic motion of self-thermophoretic Janus particles. *Journal of Statistical Mechanics: Theory and Experiment*, 2019:074001, 2019.
- [3] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E Goldstein, H. Löwen, and J. M. Yeomans. Meso-scale turbulence in living fluids. *Proceedings of the National Academy of Sciences*, 109:14308–14313, 2012.
- [4] A. Callegari and G. Volpe. Numerical simulations of active brownian particles. *Flowing Matter*, page 211, 2019.
- [5] D. Saintillan and M. J. Shelley. Instabilities and pattern formation in active particle suspensions: kinetic theory and continuum simulations. *Physical Review Letters*, 100:178103, 2008.
- [6] D. Saintillan and M. J. Shelley. Instabilities, pattern formation, and mixing in active suspensions. *Physics of Fluids*, 20:123304, 2008.
- [7] D. Saintillan and M. J. Shelley. Orientational order and instabilities in suspensions of self-locomoting rods. *Physical Review Letters*, 99:058102, 2007.
- [8] J. Słomka and J. Dunkel. Generalized Navier-Stokes equations for active suspensions. *The European Physical Journal Special Topics*, 224:1349–1358, 2015.
- [9] J. Słomka and J. Dunkel. Geometry-dependent viscosity reduction in sheared active fluids. *Physical Review Fluids*, 2:043102, 2017.
- [10] J. Słomka and J. Dunkel. Spontaneous mirror-symmetry breaking induces inverse energy cascade in 3d active fluids. *Proceedings of the National Academy of Sciences*, 114:2119–2124, 2017.
- [11] P. Kundu, I. Cohen, and D. Dowling. Fluid mechanics, 2015.
- [12] J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical Review E*, 58:4828, 1998.
- [13] J. Toner, Y. Tu, and S. Ramaswamy. Hydrodynamics and phases of flocks. *Annals of Physics*, 318:170–244, 2005.
- [14] H. K. Versteeg and W. Malalasekera. *An introduction to computational fluid dynamics: the finite volume method*. Pearson Education, 2007.
- [15] B. Andersson, R. Andersson, L. Håkansson, M. Mortensen, R. Sudijo, and B. Van Wachem. *Computational fluid dynamics for engineers*. Cambridge University Press, 2011.

- [16] F. Moukalled, L. Mangani, and M. Darwish. *The finite volume method in computational fluid dynamics*, volume 6. Springer, 2016.
- [17] C. A. J. Fletcher. *Computational techniques for fluid dynamics 2: Specific techniques for different flow categories*. Springer Science & Business Media, 2012.
- [18] S. V. Patankar and D. B. Spalding. A calculation procedure for heat, mass and momentum transfer in three-dimensional parabolic flows. In *Numerical prediction of flow, heat transfer, turbulence and combustion*, pages 54–73. Elsevier, 1983.

Chapter 23

Multiparticle Collision Dynamics

MARISOL RIPOLL

Active matter refers to systems where at least one of their components is able to transform energy into motion [2, 3]. These systems are ubiquitous in biological systems and have components ranging from the meter scale of sheep or birds to the micrometer scale of bacteria or algae, or even to the nanometer scale of sub-cellular structures such as microtubules. Synthetic active systems are currently a subject of great technological interest due, for example, to the development of new bio-mimetic materials [4, 5]. Biological and synthetic systems with their essential components in the micrometer scale are typically suspended in a solvent such as water, which plays an important role in determining the system properties. The solvent-induced interactions among different suspended particles are commonly known as *hydrodynamic interactions*.

In the absence of any interparticle solvent interactions, a force \mathbf{F}_i applied to a particle will result in a linear increase of the particle velocity $\mathbf{v}_i = \frac{D_0}{k_B T} \mathbf{F}_i$, where k_B is the Boltzmann constant, T the system temperature, and D_0 the diffusion coefficient [6]. D_0 is a constant that depends solely on the solute and solvent properties.

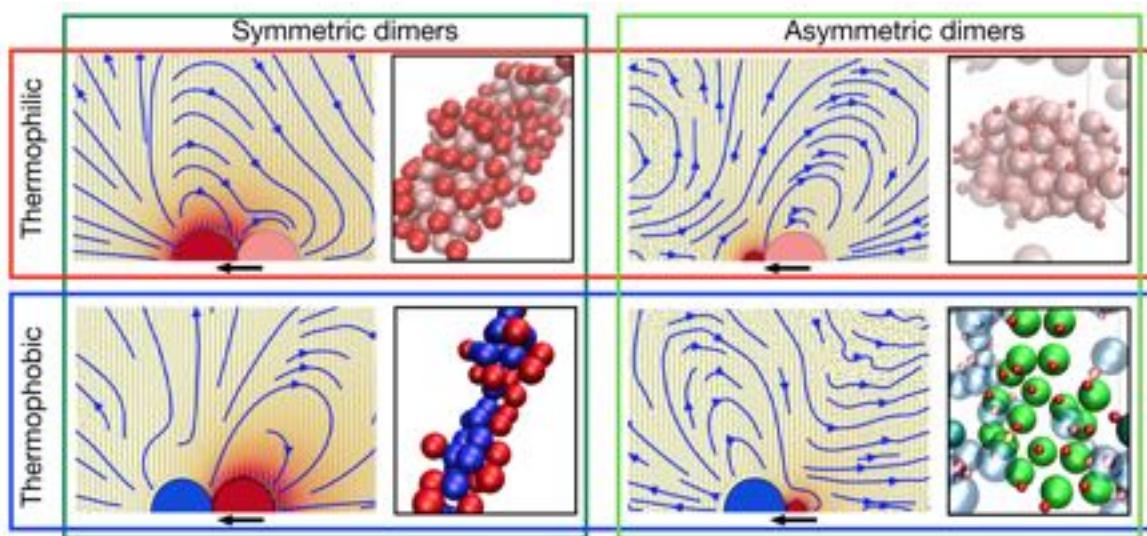


Figure 23.1: **Thermophoretic colloidal systems simulated with the MPC method.** Solvent-induced flow fields of single self-propelled dimeric colloids and representative snapshots of corresponding dense systems. The solvent is not shown here, bright red denotes the hot beads, translucent colors non-clustered dimers. These flow fields show the dependence of hydrodynamics with shape and phoretic character, from pusher- to puller-type, which here result in large colloidal aggregates for thermophilic dimers, and oriented swarms for thermophobic dimer systems. Figure adapted from Ref. [1].

In the presence of hydrodynamic interactions, the velocity increase of a particle due to an applied force is not linear anymore but depends on the position and velocities of all neighboring particles, $\mathbf{v}_i = \frac{1}{k_B T} \sum_j \mathbb{D}_{ij}(\mathbf{r}^N) \mathbf{F}_j$, where $\mathbb{D}_{ij}(\mathbf{r}^N)$ is the hydrodynamic mobility tensor. In cases where only two point particles are considered and the velocities are not too large, an approximation for the fluid motion is well characterized by the *Oseen tensor*, which in three dimensions decays like the inverse of the particles separation $\sim r^{-1}$. This means that hydrodynamic interactions are long range. This, together with the fact that they are dynamical many-body forces, makes that their consideration is not always easily taken into account.

The precise consideration of solvent effects in the simulation model of an active matter system is challenging due to the large difference in the relevant length and time scales. The suspended swimmers are of the order of micro to millimeters, while the solvent molecules have sizes of a few tenths of nanometer. The need to bridge the length and time scales gap and the increasing availability of computing power have stimulated the development of several mesoscale simulation techniques [7, 8]. This has been done fundamentally either with *top-down approaches* which consist of discretizations of the continuum equations or with *bottom-up approaches* which consist in coarse-grained descriptions of the fluid where the microscopic scale is strongly simplified, but relevant effects are still taken into account. Mass and momentum local conservation are necessary requirements for hydrodynamic interactions to be correctly modeled.

Two of the most well-known and extended hydrodynamic mesoscopic simulation techniques are *lattice Boltzmann* (LB) [9, 10] and *dissipative particle dynamics* (DPD) [11, 12, 13]. In the LB method, particle densities move in the nodes of a lattice with discretized velocities, such that mass and momentum fulfill local conservation laws [10] (see also Chapter 24). In standard DPD, particles positions and velocities are updated according to Newton's equations of motion by considering conservative, dissipative, and stochastic interactions. Mass and momentum are then conserved quantities and thermal fluctuations are taken into account. A third and alternative approach has been proposed by Malevanets and Kapral [14, 15] with the method *multiparticle collision dynamics* (MPC), also known as *stochastic rotation dynamics* (SRD). This method conserves in its basic implementation mass, momentum, and energy, including also thermal fluctuations, and the motion is not restricted by the presence of an underlying lattice. This simplicity presents MPC as a prominent alternative for the investigation of numerous systems.

In this Chapter, the implementation details of the mesoscopic method MPC are introduced. A couple of example applications are also reviewed as illustrated in Fig. 23.1. For further explanations of the method, I refer the interested reader to more extended reviews [16, 17, 18].

23.1 Multiparticle collision dynamics method

This mesoscale simulation technique is developed from a previous method known as *direct simulation Monte Carlo* (DSMC). DSMC was developed for simulating flow of relatively dilute gases [19], and consists in a set of particles that alternate streaming and collision steps. In the streaming step all particles move ballistically, and then collisions are performed among randomly selected pairs of particles. The space is divided into collision boxes inside which the mentioned pairs are chosen. The number of collisions per time step is fixed by the known collision frequency at the specified density. The precise rules for the collisions depend on the molecular model that one wants to reproduce. This is still a widely employed method for the simulation of gases, important for example in aeronautic problems, but mostly restricted to diluted gases, although an extension of the DSMC for dense gases is also available [20].

Multiparticle collision dynamics is a variant of the DSMC method where binary collisions are replaced by multi-particle collisions in a prescribed collision volume. Mass, momentum, and energy are locally conserved quantities by construction and it has been demonstrated that the hydrodynamic equations are satisfied. MPC was first introduced by Malevanets and Kapral [14, 15] in 1999 and since then there has been a very large number of works. The MPC method has already been extensively tested, showing for example to reproduce the Navier-Stokes equation [14], to include hydrodynamic interactions for example in polymer solutions [21, 22], and to fulfill the fluctuation theorem [23].

In these more than two decades, MPC has successfully been applied to a large number of systems like colloidal solutions [24, 25], rod-like colloids [26, 27], star polymers [28, 29, 30], polyelectrolyte solutions [31], and red-blood cells [32, 33]. These systems have been simulated in various conditions like equilibrium, confinement, shear, and capillary flows, and under temperature and density gradients. More recently, MPC is

being very successful in the simulations of various systems of both biological [34, 35, 36, 37, 38] and synthetic [39, 40, 41, 42, 43, 44] microswimmers. MPC has furthermore served to propose novel microscale machines [45, 46].

23.1.1 MPC algorithm

The fluid is modeled by N point particles. Each of these particles is characterized by its position \mathbf{r}_i and velocity \mathbf{v}_i , and labeled with $i = 1, \dots, N$. Positions and velocities are continuous variables, which evolve in discrete increments of time. The mass m_i associated with the particles is usually taken to be the same, but more generally, different masses can be assigned.

The MPC algorithm consists of two steps, *streaming* and *collision*, which are illustrated in Fig. 23.2. In the streaming step, the particles do not interact with each other (Fig. 23.2a), they move ballistically according to their velocities during a fixed amount of time h . This is not just an integration time, since it partly encodes the system properties, such that we refer to it as the *collision time*. Thereby, the evolution rule is

$$\mathbf{r}_i(t+h) = \mathbf{r}_i(t) + h\mathbf{v}_i(t). \quad (23.1)$$

In the collision step, the particles are sorted into collision boxes (Fig. 23.2b), and interact with all other particles in the same collision box. This multi-body interaction takes place through $\mathbf{v}_{cm,i}(t)$, the velocity of the center of mass of all particles j , which are located in the collision box of particle i at the considered time t . This is

$$\mathbf{v}_{cm,i}(t) = \frac{1}{\sum_j m_j} \sum_j^{(i,t)} m_j \mathbf{v}_j. \quad (23.2)$$

The collision boxes are typically the unit cells of a d -dimensional cubic lattice with lattice constant a , although other geometries would be in principle possible [47]. The collision is then defined as a rotation of the velocities of all particles in a box in a co-moving frame with its center of mass. Thus, the velocity of the i -th particle after the collision is

$$\mathbf{v}_i(t+h) = \mathbf{v}_{cm,i}(t) + \mathcal{R}(\alpha) [\mathbf{v}_i(t) - \mathbf{v}_{cm,i}(t)], \quad (23.3)$$

where $\mathcal{R}(\alpha)$ is a stochastic rotation matrix. This implies that, during the collision, each particle changes the magnitude and the direction of its velocity (Fig. 23.2c), in such a way that the total momentum and kinetic energy are still conserved within every collision box.

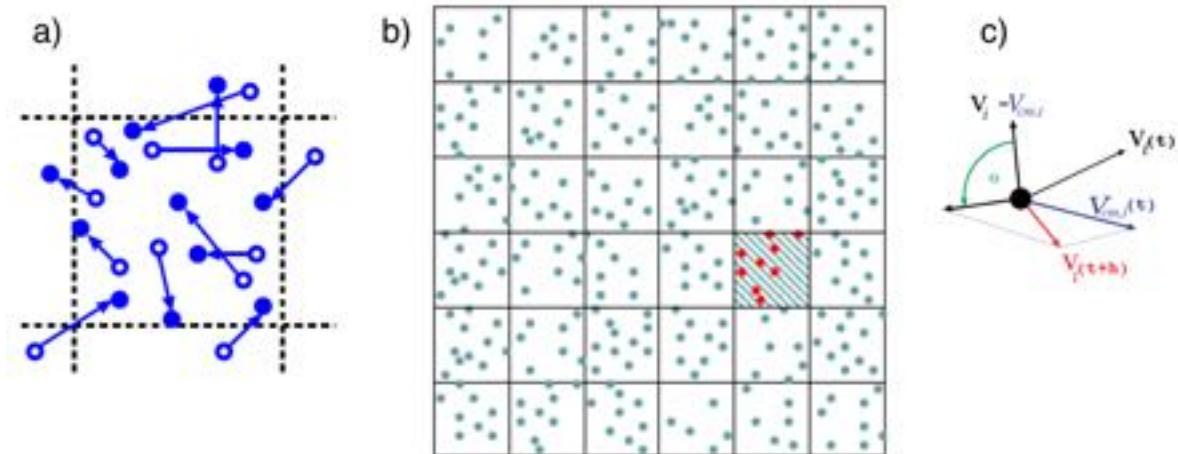


Figure 23.2: **Diagram of the MPC dynamics in two dimensions.** (a) Streaming step, (b) particles sorted into collision boxes, and (c) rotation of the particle velocity relative to the center of mass.

Example 23.1: Conserved quantities: Considering the MPC collision rule in Eq. (23.3), prove that both linear momentum and kinetic energy are in fact conserved quantities.

Solution: This can be checked by comparing the total momemtum and kinetic energy of the box center of mass before and after the collision

$$\begin{aligned} \sum_j^{(i,t)} \mathbf{v}_j(t+h) &= \sum_j^{(i,t)} (\mathbf{v}_{\text{cm},i}(t) + \mathcal{R}(\alpha)[\mathbf{v}_j(t) - \mathbf{v}_{\text{cm},i}(t)]) = \sum_j^{(i,t)} \mathbf{v}_j(t), \\ \sum_j^{(i,t)} v_j^2(t+h) &= \sum_j^{(i,t)} (\mathbf{v}_{\text{cm},i}(t) + \mathcal{R}(\alpha)[\mathbf{v}_j(t) - \mathbf{v}_{\text{cm},i}(t)])^2 = \sum_j^{(i,t)} v_j^2(t). \end{aligned}$$

Therefore, with the collision rule in Eq. (23.3), the conservation of mass, local momentum, and kinetic energy are guaranteed by construction. Note that, for length scales smaller than the box size a , the above calculation does not apply. This means, for example, that perturbations with wavelengths smaller than a are in principle not guaranteed to propagate hydrodynamically. Most systems and phenomena of interest are therefore significantly larger than a .

MPC units. In the simulations, N particles are initially placed at random in a cubic system of linear extension L . The average number of particles in a collision box is $\rho = N(a/L)^d$, the *scaled number density*. Starting from an arbitrary distribution of velocities, only a few steps are required to reach the *Maxwell–Boltzmann velocity distribution*. The equilibrium temperature T is then given by the average kinetic energy $m \langle \mathbf{v}_i^2 \rangle = 3k_B T$, where k_B is the Boltzmann constant. It is necessary to choose some reference units. These can for instance be

Mass unit:	m	≡ particle mass
Length unit:	a	≡ collision box size
Energy units:	$k_B T$	≡ system temperature

which correspond to measuring length and time according to $\hat{x} = x/a$ and $\hat{t} = t \sqrt{k_B T / m a^2}$. A first estimation for the scaled mean free path is then given by $\lambda = \hat{h}$, although a more accurate value should account for the number density and the collision time as well.

Random rotation. Together with the collision step the stochastic rotation matrix $\mathcal{R}(\alpha)$ has been introduced, such that α is a parameter of the model. The direction around which the rotation is performed is different in neighboring boxes, but inside a **particulate [particular?]** box all particles suffer the same rotation. In two dimensions, the rotation of the relative velocity to the box center of mass is simply given by an angle $\pm\alpha$, where the sign is independently and randomly chosen for each cell (Fig. 23.3a). In three dimensions, two main schemes have been described for the random collisions. The first one [48] chooses the rotation direction among the three main axes and the rotation is performed by an angle $\pm\alpha$. The second scheme [49] consists in choosing a random direction in space for each box around which the relative velocities are rotated by the angle α . This second scheme is slightly more difficult to implement but in return it reduces the anisotropy eventually introduced by the underling lattice.

Random shift. In order to perform the multiparticle collision, as already discussed, particles are sorted into cells where the collision takes place. The choice of these collision boxes defines a preferential grid, and the first naive choice would be a fixed grid whose outside borders coincide with the system boundaries, similar to the sketch in Fig. 23.2b. Nevertheless, such a fixed grid does not fulfill Galilean invariance if the mean free path is not large enough. This can be visualized in the case of two systems, one at rest and the second moving with a constant velocity. If the displacement of particle i is smaller than the size of the collision box a , the particles in the collision box will not be the same in the moving system and in the system at rest, which will lead to different dynamics in the two cases and therefore to a breakdown of the Galilean invariance. Therefore, a *random shift* of the collision grid has to be performed in the execution of the collision step in order to ensure Galilean invariance in the full range of possible parameters [50, 51].

The random shift is performed by displacing the collision grid a random number uniformly distributed in the interval $(0, 1)$ which is chosen independently in each collision. In Fig. 23.3b, the solid black grid represents

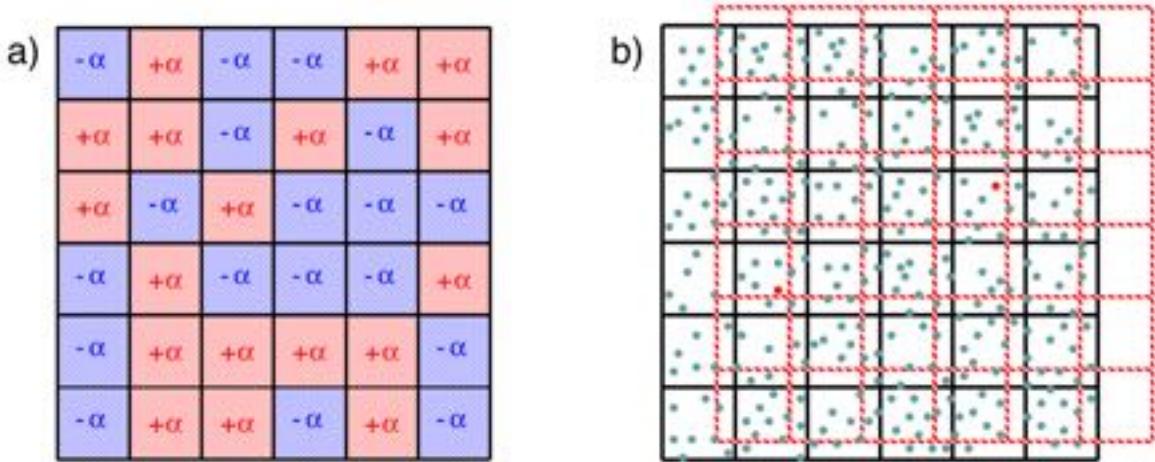


Figure 23.3: **Random rotation and shift in two dimensions.** (a) Realization of the random direction rotation in two dimensions (red $+ \alpha$, blue $- \alpha$). (b) Diagram of the random shift of the collision grid.

the fixed grid, while the discontinuous red grid would be one of the possible displaced grids. Note that periodic boundary conditions are applied, such that for example, the last-right column of the shifted grid would include some particles of the first-left column of the fixed grid. As a consequence of such a shift, no special frame exists and Galilean invariance is restored. The random shift implementation produces also that the probability of two particles to interact will be inversely proportional to their relative distance, similar to any soft range potential. Random shift also facilitates the transfer of momentum between neighboring particles, since the different positions of the grid in two consecutive collision steps make possible transfer of momentum over larger distances per time unit.

23.1.2 Capillary flow

For many interesting situations, it is necessary to implement walls as boundary conditions. Two main types of boundary conditions are possible when simulating a solid wall: *stick boundary conditions* and *slip boundary conditions*. This classification is performed attending the difference of velocity between the fluid and the wall. *Slip boundary conditions* allow the fluid in contact with the wall and the wall itself to have different velocities. *Stick boundary conditions* (also called *no-slip boundary conditions*) impose a continuous change of velocity between the wall and the fluid. The standard procedure to obtain stick is known as *bounce-back*: when a particle hits the wall its trajectory is reverted and its velocity is inverted (Fig. 23.4a). These considerations apply to many simulation techniques, such as Monte Carlo or Molecular Dynamics.

For simulating fixed walls with MPC, standard bounce-back is applied during the streaming step. This would be enough when the walls exactly coincide with the boundaries of the collision cells. However, the walls will generally not coincide with the cell boundaries, due to special geometries or to random shift. In these cases (Fig. 23.4b), the cells in the boundary will be generally partially filled, which will not lead to the desired stick boundary conditions. An efficient solution to overcome this problem has been proposed [52]. The idea is that for all the cells of the channel that are cut by walls and have as a consequence a number of particles n smaller than the average number of particles in the bulk ρ , extra *virtual particles* will be added. The function of these virtual particles is to obtain the behavior of a fluid with constant density and temperature in the cells at the boundaries. The velocities of the virtual-wall particles are drawn from a Maxwell-Boltzmann distribution of zero average velocity and the same temperature T as the fluid. The collision step in Eq. (23.3) is then carried out with the average velocity of all particles in the cell. Since the sum of random vectors drawn from a Gaussian distribution is again Gaussian-distributed, it is not necessary to calculate the velocities of the individual particles. Instead, the center of mass velocity in Eq. (23.3) can be written as

$$\mathbf{v}_{\text{cm},i} = \frac{\sum_j^i m \mathbf{v}_j + \mathbf{a}}{\rho}, \quad (23.4)$$

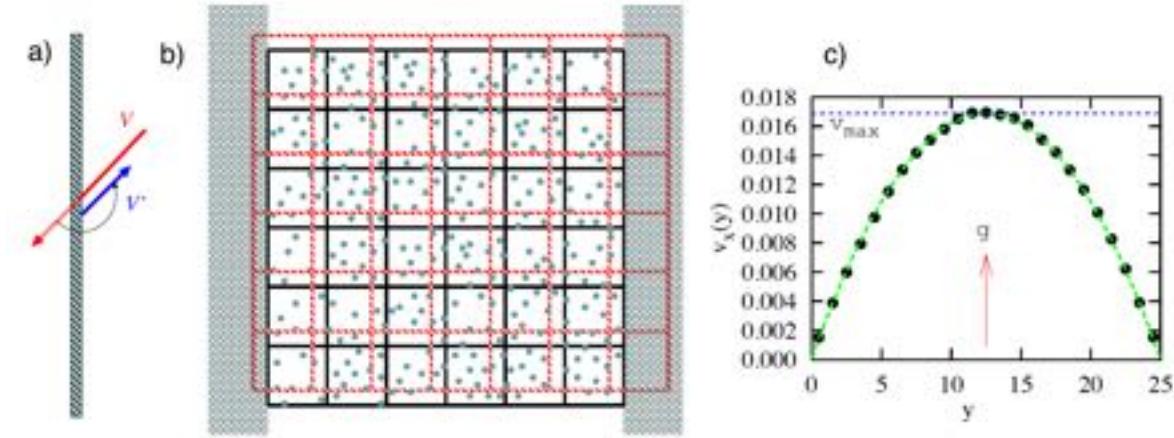


Figure 23.4: **Effect of walls and boundary conditions.** (a) Diagram of the bounce-back rule. (b) Random-shift in the presence of walls. (c) Parabolic velocity profile in the presence of a gravitational field reproduced by MPC.

where \mathbf{a} is a vector whose components are numbers from a Maxwell-Boltzmann distribution with zero average and variance $(\rho - n)k_B T$.

Capillary flow. The *capillary flow* or *Poiseuille flow* is the typical flow arising when considering a fluid confined between two planar walls, and a gravitational field g applied in one direction parallel to the walls. After a relaxation time, the system reaches a stationary state with a parabolic velocity profile between the walls and in the direction of the force. See an example in Fig. 23.4c, where planar fixed walls are implemented at $y = 0$ and $y = L_y = 25$. In the limit of low Reynolds number, the Stokes equation can be solved [53], which explains that the measured maximum velocity of the parabola is inversely proportional to the kinematic viscosity ν of the fluid like $v_{\max} = gL_y^2/(8\nu)$. This behavior is reproduced by MPC simulations, and it can be used as a measurement for the viscosity displayed by the MPC fluid [52, 21, 54]. Alternative methods to determine the viscosity from simulations have been employed in Refs. [55, 56], where a system under shear flow and vorticity correlations have been respectively used.

23.2 Solvent properties

The transport coefficients of the MPC solvent have been intensively studied. Analytical expressions have been derived from kinetic theory by generalizing point-like collisions to finite collision volumes [51, 56, 48, 55]. The theoretical expressions describe numerical results very well. This is quite convenient since different properties of the fluid can be tuned by choosing the adequate set of model parameters.

23.2.1 Viscosity

The *total kinematic viscosity*, $\nu = \nu_{\text{kin}} + \nu_{\text{coll}}$, is the sum of two contributions, the *kinetic viscosity* ν_{kin} and the *collisional viscosity* ν_{coll} , which have been calculated in two and three dimensions. In three dimensions [48, 55], these expressions are

$$\nu_{\text{coll}} = \frac{1}{\lambda} \frac{(1 - \cos \alpha)}{18} \left(1 - \frac{1}{\rho} \right), \quad (23.5)$$

$$\nu_{\text{kin}} = \lambda \left[\frac{1}{(4 - 2 \cos \alpha - 2 \cos 2\alpha)} \frac{5\rho}{\rho - 1} - \frac{1}{2} \right]. \quad (23.6)$$

Check the references to these equations in the text to be sure they are correct.

The viscosity data obtained from capillary flow simulations are presented in Fig. 23.5 together with the theoretical predictions of Eqs. (23.5) and (23.6). The obtained agreement is quite remarkable, in contrast to the case of other mesoscopic simulation techniques such as dissipative particle dynamics [57]. Density fluctuations can also be included in the theory [55], which noticeably improves the agreement with the simulations results for small number densities; for $\rho \geq 5$ these contributions are negligible.

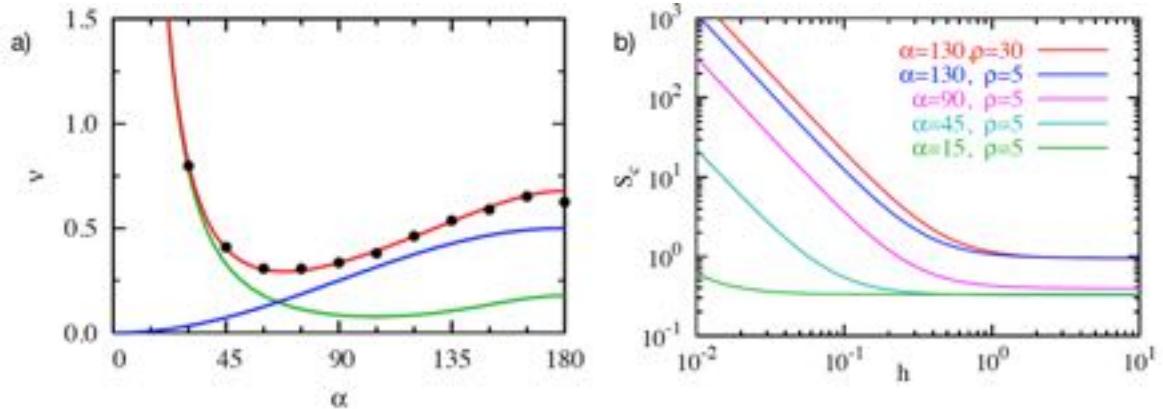


Figure 23.5: **Kinematic viscosity.** (a) Kinematic viscosity as a function of the collision time. The red line is the total theoretical prediction in Eq. (23.7) [Is the equation reference correct?], with $\alpha = 130$ and $\rho = 5$, the blue line is the collisional contribution and green line the kinetic contribution. The symbols are measurements from capillary flow simulations, with system size $L/a = 20$. (b) Schmidt number versus collision time from the theoretical expressions in Eqs. (23.5), (23.6), and (23.7). The parameters α and ρ are modified as specified in the legend. Figures adapted from Ref. [54]. Check notation used in the figure to be sure it's consistent with the text.

The ratio between the kinetic and the collisional contributions to the kinematic viscosity varies considerably with the model parameters, as can be seen easily from the theoretical expressions in Eqs. (23.5) and (23.6). In Fig. 23.5, the total kinematic viscosity and its two contributions are plotted as a function of the collision time step. The collisional contribution is dominant for large collision small collision times, while the kinetic viscosity dominates in the opposite case of small collision angles and large collision times.

23.2.2 Diffusion coefficient

The self-diffusion coefficient can be calculated by the *Green–Kubo formalism* from the velocity autocorrelation function (VACF) as $D = \frac{1}{3} \int_0^\infty dt \langle \mathbf{v}(t) \mathbf{v}(0) \rangle$ [56]. In order to obtain an analytical prediction, the molecular chaos approximation is employed for the VACF, yielding

$$D = \lambda \left(\frac{1}{\gamma} - \frac{1}{2} \right), \quad \gamma = \frac{2}{3} (1 - \cos \alpha) \left(1 - \frac{1}{\rho} \right), \quad (23.7)$$

where the decorrelation factor γ has been specified for the three-dimensional case. A simulation measurement can be obtained through the VACF or by calculating the averaged mean square displacement. The agreement with simulation results is quite reasonable for large values of the mean free path λ , but noticeably deviates for small ones. This is due to the fact that the molecular chaos approximation neglects the effect of correlations between particles. This deviation is therefore an indication of the importance of the enhancement of the collective behavior in the regime where the mean free path is smaller than the box size [21, 54].

23.2.3 Dimensionless numbers

Dimensionless numbers provide an important way of quantitatively comparing simulation and experimental parameters. Numerous numbers are available and can be described depending on the problem under study.

Schmidt number. The ratio of the kinematic viscosity, ν , and the diffusion coefficient, D , is known as the Schmidt number $Sc \equiv \nu/D$, this is the ratio between momentum transport and mass transport. For gases, this number is on the order of unity, while in fluids like water it is on the order of 10^2 to 10^3 . A prediction for the Schmidt number of an MPC fluid can be obtained from the theoretical expressions for the kinematic viscosity in Eq. (23.6) and the diffusion coefficient in Eq. (23.7). In Fig. 23.5, the theoretical prediction is plotted for Sc as a function of the collision time for different values of the rotation angle and number density. This shows that Sc becomes considerably larger than unity for the same range of parameters where the collisional viscosity is considerably larger than the kinetic viscosity.

In order to perform simulations where the fluid-like hydrodynamic effects are properly taken into account, one needs to employ large values of the rotation angle α and small values of the collision time. These restrictions, together with the election of the number density ρ , still give a large margin to chose particular values for the transport coefficients like the viscosity $\eta = \nu\rho$. Simulations will become computationally more expensive for smaller collision times and for larger number densities, which call for an optimization of the chosen parameters depending on the problem requirements. As an example, an extensively used set of parameters is $h = 0.1$, $\alpha = 130^\circ$, and $\rho = 10$, for which $Sc = 13$. This value is clearly smaller than that of water, but shows that the propagation of momentum is faster than that of mass, which has extensively been shown to provide a very efficient approach to include hydrodynamic interactions [54, 58].

Prandtl number. The Prandtl number relates the transport of momentum and the transport of energy, and is defined as $Pr \equiv \nu/k_T$, with k_T being the thermal diffusivity, which has also been extensively characterized for the MPC fluid [48, 59, 60]. With the same above example parameters $Pr = 5.3$, which is very close to that of various fluids such as water, and most importantly, enables local temperature gradients to remain stable under adequate boundary conditions [44].

23.2.4 Equation of state

In spite of the liquid-like behavior displayed by the MPC solvent, it should be noted that, due to the lack of potential interactions, the MPC solvent is characterized by the ideal gas equation of state. This does not prevent the technique to be applicable to many relevant hydrodynamic systems, but it should be taken carefully. The investigation of binary fluids, or systems with phase transitions is therefore not possible with the MPC solvent in the form that it has been introduced so far. On the other hand, it should be noted that stable temperature gradients will always come together with density gradients and vice versa [60]. Although this could in principle result in some cases in the emergence of compressibility effects, simulations of phoretic colloids in the presence of temperature gradients have clearly shown that these effects are not significant [58]. This has been tested in the case of colloids in the presence of externally imposed constant temperature gradients, and in the case of Janus spherical colloids or dimers where the temperature decays with the inverse of the temperature due to the presence of a hot colloid surface [58, 41]. The velocity profiles can be compared with the solution of the Stokes equation for a problem with similar boundary conditions and constant density, showing only small and non-significant deviations.

23.3 MPC alternative implementations

Add a global paragraph for this section.

23.3.1 MPC with Andersen thermostat

The MPC algorithm has been constructed such that mass and translational momentum are locally conserved, which is essential for obtaining a correct hydrodynamic behavior. As already discussed, the original version of MPC defined with the collision rule in Eq. (23.3) also conserves energy. This means that the method is suited to be applied to investigate transport of energy or systems with temperature gradients [60], and in principle also that temperature is constant without the necessity of any additional thermostat. This is though not the case only in many cases the system has external influences **check this sentence — not sure about its meaning** like in the presence of a gravitational force, shear flow, or an electric field. In these cases, the interaction with a

thermostats might be necessary similar to most simulation techniques [61, 62]. A collision with an intrinsic thermostat can be defined [63] alternatively to Eq. (23.3) as

$$\begin{aligned}\mathbf{v}_i(t+h) &= \mathbf{v}_{\text{cm},i}(t) + \delta\mathbf{v}^{\text{ran}}, \\ &= \mathbf{v}_{\text{cm},i}(t) + \mathbf{v}_i^{\text{ran}} - \frac{1}{N_c} \sum \mathbf{v}_j^{\text{ran}},\end{aligned}\quad (23.8)$$

where N_c is the number of particles in the collision cell and the sum runs over all particles in the cell. Here, the new relative velocities are Gaussian random numbers with variance $\sqrt{k_B T/m}$, being T the system average temperature. This method is known as MPC-AT in contrast to the standard one which is now frequently called MPC-SRD.

23.3.2 MPC with angular momentum conservation

Besides the multiple advantages of the two described MPC methods, local angular momentum is not conserved in any of them. Although the violation of angular momentum conservation does not affect the generic form of the hydrodynamic equations [64, 61, 65], the fluid stress tensor is no longer symmetric [64, 66, 67, 68]. This affects the propagation of sound, but is unimportant for the evolution of vorticity of a bulk fluid. The absence of angular momentum conservation leads to a non-physical contribution to the torque in the case of circular Couette flow, or to incorrect velocity fields [67, 68], but only in some cases.

Angular momentum conservation can be restored by imposing constraints on the new relative velocities. The collision rule in this case is

$$\begin{aligned}\mathbf{v}_i(t+h) &= \mathbf{v}_{\text{cm},i}(t) + \mathbf{v}_i^{\text{ran}} - \frac{1}{N_c} \sum \mathbf{v}_j^{\text{ran}} \\ &\quad + \left\{ \frac{m}{\Pi} \sum [\mathbf{r}_{j,c} \times (\mathbf{v}_j - \mathbf{v}_j^{\text{ran}})] \times \mathbf{r}_{j,c} \right\},\end{aligned}\quad (23.9)$$

where Π is the moment of inertia tensor of the particles in the cell, and $\mathbf{r}_{j,c} = \mathbf{r}_j - \mathbf{R}_c$ is the relative position of particle j in the cell and \mathbf{R}_c is the center of mass of all particles in the cell. This is then an angular-momentum conserving modification of MPC-AT [63, 67], denoted MPC-AT+a.

23.3.3 Non-ideal MPC

Based on the MPC solvent, a model with generalized excluded volume interactions between fluid particles such that fluids with non-ideal-gas equation of state are produced is proposed in Ref. [69]. The additional consideration of interactions between different types of particles leads to the possibility of simulating multicomponent fluids [70]. More recently, the MPC technique has been generalized to the application of binary-fluid mixtures and multiphase flows, by coupling MPC to a Ginzburg-Landau free-energy functional for phase-separating binary fluids [71]. An additional density-dependent term is also introduced in order to describe fluids with a non-ideal equation of state.

23.4 Coupling with particles and obstacles

Until now, we have only discussed the implementation of a simple fluid simulated with MPC. The next important question is how to model complex structures with an MPC solvent. The general strategy is to define a hybrid algorithm where the solvent is simulated with the MPC technique and both, the solute description and the solute–solvent interactions, are accounted with specific models. For soft matter applications, different structures like colloids, polymers, or membranes have therefore developed different coupling schemes. For active matter applications, these structures are also different depending on the type of swimmers, and they are very similar to the ones employed in soft matter.

MPC coupling. In this case, only solute–solute interactions are described with MD, while both solvent–solvent and solute–solvent interactions are described with MPC interactions. This means that the MPC streaming step in Eq. (23.1) is used to update the positions just of the solvent particles, while in the collision step

in Eq. (23.3) both solvent and solute particles are taken into account. The collision is performed through the collision box center of mass velocity in Eq. (23.2), where the solute particles are generally considered to have larger mass than the surrounding solvent particles.

The solute–solute interactions are integrated with standard MD strategies, for instance, with the velocity–Verlet algorithm with a time step Δt generally 20 to 100 times smaller than the collision time. This discrete time evolution is then interrupted every $h/\Delta t$ steps such that solvent particles interchange momentum among themselves through an MPC collision (Fig. 23.6a). Typical potential parameters can then be related to the MPC units.

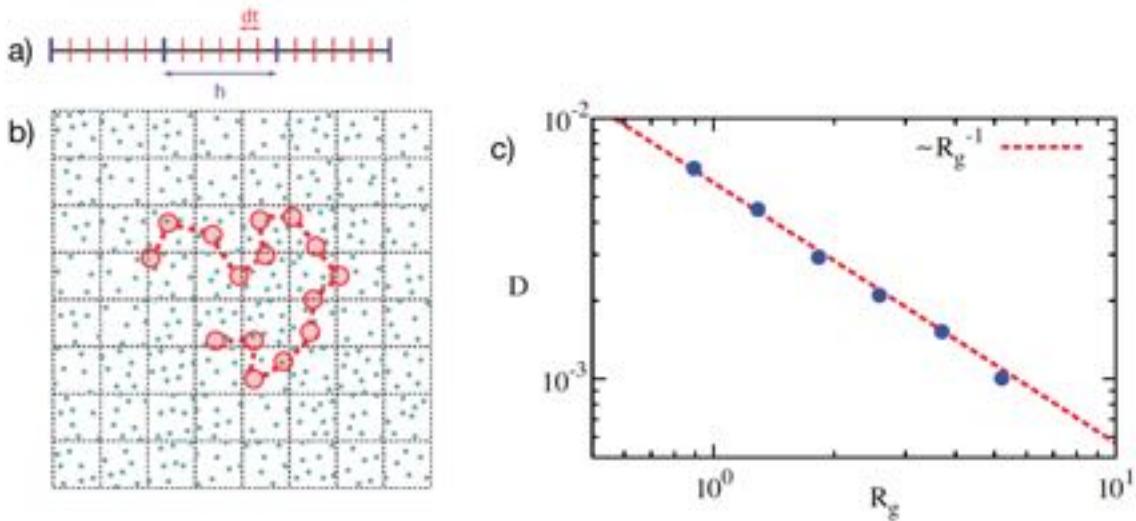


Figure 23.6: Inclusion of a polymer in an MPC solvent. (a) Diagram of the time evolution in MD and MPC couplings. (b) Diagram of the inclusion of a polymer in an MPC solvent. (c) Polymer center of mass diffusion coefficient with the MPC coupling together with the Zimm theoretical prediction. Reproduced from Ref. [22].

This coupling has been extensively used for simulating different types of polymers [22, 72]. Following MD polymer models, N_m particles of mass M are considered to be the polymer monomers. These monomers are connected by harmonic springs and may also interact via an excluded-volume potential. The typical distance between consecutive monomers (bond length) is of the order of the collision box size a , so that there are not many monomers in each collision box, and that consecutive monomers are not separated more than one collision box. The potential strength would be taken in units of the thermal energy $k_B T$. This MPC coupling of solvent and monomers does not have any potential interaction, which implies that the solvent particles can freely move inside the monomers' cores, but the monomers cannot inter-penetrate each other (Fig. 23.6b).

In order to test the validity of this approach, the MPC coupling has been used to study the dynamics of short polymer chains in solution [21, 22]. The predictions of the *Zimm theory*, in which the effect of hydrodynamic interactions is taken into account, are in excellent agreement with the simulation results for instance, for the center of mass diffusion coefficient (Fig. 23.6c) and for the relaxation times of the Rouse modes. This agreement constitutes a nice proof of the correct hydrodynamic behavior implemented by the MPC coupling approach. This is found in the parameter regime where the collective behavior has been described to dominate the dynamics, which also is when the Schmidt number displayed larger values. Polymer collapse has also been studied with this coupling showing the relevance of hydrodynamics interactions [73, 74].

MD coupling. Although the MPC coupling is very effective, there are cases in which stronger, or simply tunable, interactions between solvent and solute are crucial to determine the proper system dynamics. The idea of this coupling is that both the solute–solute and solute–solvent interactions are taken into account through explicit potentials with molecular dynamics (MD), and only the solvent–solvent interactions are mesoscopically described through MPC. Although computationally more costly than the MPC coupling, this method still leads to a very important speed up of the simulations respect to simulations that consider with MD all the solvent–solvent interactions. **With this coupling, it should be considered that *solvent-induced depletion interactions***

might appear if the solvent–solute interactions have larger interaction range than solute–solute interactions can result in unphysical effects [75] [Check this sentence, something is missing].

This was the first coupling introduced for MPC by Malevanets and Kapral [15] and it was applied to a colloidal dispersion. In Ref. [76], Lee and Kapral employ this coupling to study dynamics of a polymer with solvents of different qualities. This is achieved by changing the excluded volume interactions from purely repulsive to a Lennard-Jones interaction that accounts for repulsive interactions at short distances and attractive interactions at intermediate distances. The poor solvent is reproduced by considering attractive interactions between monomers and repulsive ones between solvent and solute particles. Oppositely, the good solvent considers repulsive interactions between monomers and repulsive [attractive?] between monomers and solvent particles.

Boundary coupling. Specific solvent–solute interactions can be defined at the boundaries between the solute surfaces and the surrounding solvent. In order to simulate the interaction of a flow with a spherical obstacle, the MPC solvent particles interact with the sphere surface as with a planar wall with a bounce back collision [52, 77, 49]. As can be seen in Fig. 23.7, the fluid instability expected for high Reynolds numbers is properly reproduced.

More elaborated rules have been introduced when simulating moving colloids [78]. These interactions take into account not only the transfer of linear momentum and energy between solvent and solute particles, but also the transmission of angular momentum from the solvent collisions to the solute particle.

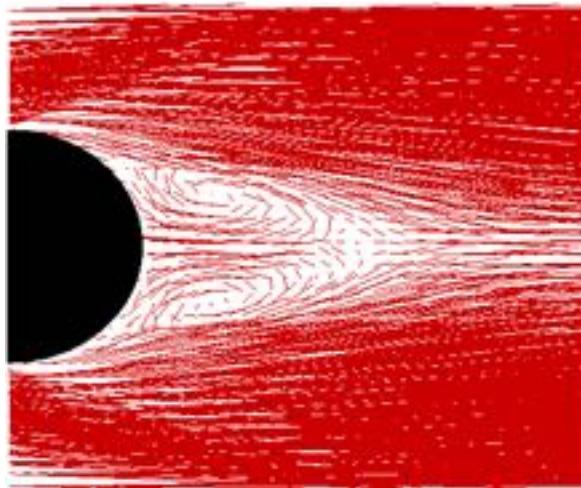


Figure 23.7: **Boundary coupling.** Velocity field of a left-to-right flow past a sphere for high Reynolds number conditions. Reproduced from Ref. [49].

Further couplings. In order to describe different specific situations other couplings can be defined. An example is the simulation of vesicles in flow [79, 32]. In this case, solvent-free models for the simulation of membranes like discretized curvature models or triangulated surfaces are employed in combination of the MPC solvent. Interestingly, in this case, the fluids inside and outside the vesicle do not mix and can have different fluid properties like density or viscosity.

23.5 Random MPC solvent

A standard procedure to determine the relevance of hydrodynamic interactions is to compare simulations with and without hydrodynamic interactions. In principle, existing methods without hydrodynamic interactions, like Brownian Dynamics (BD), provide a really valuable comparison. Nevertheless, the ideal approach to perform a good comparison is to ensure that the considered methods differ as little as possible apart from the inclusion of hydrodynamic interactions. With this spirit, we proposed the *random MPC* solvent [80], which is particularly

useful in the case of MPC coupling. The idea is to replace the solvent–solute interaction in the collision step by the interaction with a heat bath. Since the positions of the solvent particles within a cell are not required in the collision step, no explicit particles have to be considered. Instead, each monomer is coupled with an effective solvent momentum \mathbf{P} which is directly chosen from a Maxwell-Boltzmann distribution of variance $m\rho k_B T$ and a mean given by the average momentum of the fluid field – which is zero at rest, or $(m\rho \dot{\gamma} r_y^i, 0, 0)$ in the case of an imposed shear flow. This strategy is related to that proposed in Section 23.3.1 and Ref. [52] in order to model the no-slip boundary conditions of solvent particles at a planar wall. The total center-of-mass velocity is then given by

$$\mathbf{v}_{\text{cm},i} = \frac{M\mathbf{v}_i + \mathbf{P}}{m\rho + M}, \quad (23.10)$$

with which the collision step in Eq. (23.3) is performed. Note that, similarly to the MPC solvent, the complex structure trajectory is sampled by MD and the interaction with the solvent is performed every collision time h .

Solute particles, such as the monomers of a polymer, interacting with the random MPC solvent have similar properties as the MPC solvent, except those associated with hydrodynamic interactions. The relevant solvent parameters in both methods are the density ρ , the rotation angle α , and the collision time h which can be chosen to be the same. A very important advantage provided by this method is that *no* explicit solvent particles need to be considered. This translates into a computational speed-up by several orders of magnitude, depending on the considered system size. Moreover, to implement non-equilibrium conditions is straightforward. Examples of application of this technique for simulations with the MPC coupling can be found in Refs. [80, 81, 82].

23.6 Active matter examples

The described MPC method simulates an explicit solvent that includes hydrodynamic interactions. These interactions have shown to be the consequence of the momentum conservation in the MPC dynamics, where no specific fluid behavior is imposed. The flow is hydrodynamic and results therefore depend on the specific boundary conditions of the problem. In the following, a few examples of these flows in active matter are discussed. For comparison, in a few cases, analytical expressions for hydrodynamic flows can be obtained as solution of the Navier-Stokes equations with adequate boundary conditions. Systems of interest are frequently on the steady state, with small Reynolds and Mach numbers, such that time derivatives and non-linear terms can be neglected and the Stokes equation applies.

Active particles in the micrometer range are known to display different hydrodynamic flows, and these flows strongly influence the system dynamics when more than one active particle are swimming together. The first relevant example is the *source dipole* type of flow, which typically appears for swimmers that move under an externally applied force, known as *neutral swimmers*. This flow has a fast symmetric decay with $1/r^3$, which does not have a strong influence on other swimmers' motion. Swimmers with no external force applied are very common, in those cases the force that the swimmer exerts on the fluid is exactly opposite than the force the fluid exerts on the swimmer, such that the total simply vanishes. In the simplest case, the far-field hydrodynamics (at distances from the swimmer much larger than its size) can well be described by a force dipole [83, 84], and two classes of such dipole swimmers can be distinguished, known as *pullers* and *pushers*. Following some representative active matter systems simulated with MPC will be presented [3].

23.6.1 Flagellated swimmers

Detailed models of flagellated swimmers such as sperm and *E. coli* cells have been proposed [85, 37, 86, 87]. Bead–spring models with additional bending and torsion potentials conform the special geometries of the flagella and cell bodies of these biological swimmers which are implemented with molecular dynamics codes, and coupled to the solvent most typically with the MPC coupling. Independent rotation of flagella is considered via an applied torque, which comes together with opposite forces and torques to the cell body in order to ensure force-free and torque-free swimming. Propulsion is then due to frictional anisotropy of the thin flagellum [5, 84]. Simulations illustrate the complexity of the hydrodynamic flow field adjacent to a cell [86], or emphasize the importance of hydrodynamic interactions for phenomena such as the synchronization of bacterial flagella in the bundling process [85, 37].

23.6.2 Phoretic swimmers

Phoresis refers to the drift motion that particles display in the presence of an inhomogeneous solvent. These inhomogeneities can be due to a gradient of density, temperature, or ionic charge, being then referred to as *diffusiophoresis*, *thermophoresis*, or *electrophoresis* [88]. Janus or dimeric colloids generating local gradients are being studied as examples of synthetic active particles. Given the importance of the colloid–solvent interactions and to make them tunable, MPC simulations of phoresis are performed with the MD coupling. Diffusiophoretic swimmers have been investigated by considering two types of MPC fluid, attending solely to their interactions with the phoretic colloidal surface [39, 89, 90]. These are typically a catalytic reaction that induces and maintains the concentration gradient, and different potential interactions of the two species such that uneven concentrations of the two components result in the colloids’ self-propulsion. Thermophoretic swimmers can be simulated just by including local temperature inhomogeneities, this is considering a part of the colloid hot while keeping constant the system average temperature [91, 41]. Propulsion arises due to the dependence of the potentials with the temperature, and changes from being *thermophobic* (motion with the hot side to the back), to *thermophilic* (motion with the hot side to the front) when using attractive (solvophilic), or repulsive (solvophobic) colloid–solvent interactions potentials [92], all these resulting in different hydrodynamic interaction patterns. Simulations in the collective regime of these phoretic swimmers have already shown multiple possible cooperative behaviors, such as clustering, living cluster formations, or swarming [89, 93, 44].

23.6.3 Squirmers

Based on the motion of deformable swimmers at low Reynolds numbers, Lighthill proposed the squirmer model [94]. An individual squirmer is modeled as a rigid sphere with the prescribed tangential slip velocity u_{sq} on its surface,

$$u_{\text{sq}} = B_1 \sin \theta (1 + \beta \cos \theta) e_\theta , \quad (23.11)$$

where θ is the angle between the body-fixed propulsion direction and the considered point on the colloid surface with tangent vector e_θ . The generated propulsion velocity is $v_0 = 2B_1/3$ and the active stress is given by β , which characterizes the nature of the swimmer, which is a pusher for $\beta < 0$, neutral for $\beta = 0$, and a puller for $\beta > 0$. Numerous works have implemented squirmers in an MPC solvent with the velocity slip coupling at the boundary [35, 43, 95, 96] probing the corresponding hydrodynamic induced flow fields, and also extending the model to spheroidal microswimmers. The swimmer’s displacement in the MPC solvent needs to be implemented then by considering the rigid body dynamics. Simulations of two to many squirmers show the emergence of hydrodynamic collective effects and structure formation.

23.7 Concluding remarks

In the last twenty years, multiparticle collision dynamics has shown to be a very convenient and robust simulation method to investigate the dynamical properties of soft matter, active matter, and biological systems. The simplicity of its implementation, and its versatility is pushing the method forward for a large class of systems where thermal fluctuations are of importance. One additional advantage is that not only the hydrodynamic effects on the solute properties can be studied, but also the distortion that the fluid undergoes due to the presence of the solute.

The implementation details of MPC are summarized in this Chapter. A few relevant examples of applications of MPC to systems where the induced flow fields are of relevance are here also reviewed. In the cases where comparison with the solution of the Stokes equation is possible, MPC simulation results show to perfectly agree. Consequently, in cases where analytical solution cannot be obtained, the results obtained with MPC simulations are very reliable. Considering the increasing number of applications in active matter for example with the squirmer model, with flagellated biological swimmers, or with phoretic synthetic swimmers make clear that the method will play a decisive role in the investigation of novel characteristics and application possibilities of active matter systems.

23.8 Problems

Problem 23.1: Two dimensional fluid in bulk and equilibrium. Program an MPC fluid in two dimensions with periodic boundary conditions. Consider as a starting configuration N distributed particles all at velocity $|v| = 1$ but with stochastic orientations.

- a. Compute the velocity distribution function at various time steps and check how the distribution tends to the thermal equilibrium one given by the Maxwell-Boltzmann distribution.
- b. Compute total momentum and energy and check that they are perfectly conserved at all times.
- c. Compute the mean square displacement and check the value of the diffusion coefficient. Compare your results with the theoretical expression in Eq. (23.7) which in two dimensions should consider the decorrelation factor $\gamma = (1 - \cos \alpha)(1 - 1/\rho)$.
- d. Implement random shift and check again all the previous points

Problem 23.2: Two-dimensional fluid in confinement. Implement the interactions of the solvent particles with two parallel confining walls, as indicated in Fig. 23.4b, with the bounce-back interaction at the walls, as indicated in Fig. 23.4a, and considering the effect of the random shift.

- a. Compute total momentum and energy. It is still constant, but slightly varying around the conserved one. This is due to the interactions with the wall ghost particles which do not exactly conserve energy, but only on average.
- b. In the streaming step, include now a force with variable strength g in the component parallel to the walls. Measure the velocity profile in the direction perpendicular to the walls. Parabolic profile as in Fig. 23.4c should be recovered.
- c. Perform simulations with fixed input parameters, for example $\rho = 5$, $\alpha = \pi/2$, $h = 0.2$, and $L_x = L_y = 25a$, starting with $g = 10^{-4}$. The simulation time should be large enough for the velocity profile not to be time-dependent, and to allow for the fit to the parabola $v_x(y) = 4v_{\max}(L_y - y)y/L_y^2$. Compute the maximum velocity of the parabolic profile v_{\max} and check that it increases linearly with g . For small g values this should be the case; for larger values in the system, there will be a small deviation, since values might not correspond anymore to the laminar regime, and where the relation is linear.
- d. Fix g to a value in the laminar regime and then calculate the value of the kinematic viscosity $\nu = gL_y^2/(8v_{\max})$. Compare the measurement with the prediction from kinetic theory [59] which in two dimensions read $\nu = \nu_{\text{kin}} + \nu_{\text{col}}$ with

$$\begin{aligned} \nu_{\text{kin}} &= h \left\{ \frac{\rho / (1 - \cos 2\alpha)}{\rho - 1 + e^{-\rho}} - \frac{1}{2} \right\}, \\ \eta_{\text{col}} &= \frac{1}{12h} \frac{(1 - \cos \alpha)(\rho - 1 + e^{-\rho})}{\rho}. \end{aligned} \quad (23.12)$$

- e. An interesting exercise is to comment the line in the code where random shift is implemented, such that this is cancelled, and run some values of the analysis in the previous point. The parabolic profile should still be recovered, but the maximum velocity significantly varies. This is a sign that random shift does not only restore Galilean invariance in the method, but also varies the transport properties such as the viscosity.

Bibliography

- [1] M. Wagner, S. Roca-Bonet, and M. Ripoll. Collective behavior of thermophoretic dimeric active colloids in three-dimensional bulk. *European Physics Journal E*, 44:43, 2021.
- [2] G. Gompper, R. G. Winkler, T. Speck, A. Solon, C. Nardini, F. Peruani, H. Löwen, R. Golestanian, U. B. Kaupp, L. Alvarez, T. Kiørboe, E. Lauga, W. C. K. Poon, A. DeSimone, S. Muñoz-Landin, A. Fischer, N. A. Söker, F. Cichos, R. Kapral, P. Gaspard, M. Ripoll, F. Sagues, A. Doostmohammadi, J. M. Yeomans, I. S. Aranson, C. Bechinger, H. Stark, C. K. Hemelrijk, F. J. Nedelev, T. Sarkar, T. Aryaksama, M. Lacroix, G. Duclos, V. Yashunsky, P. Silberzan, M. Arroyo, and S. Kale. The 2020 motile active matter roadmap. *Journal of Physics: Condensed Matter*, 32:193001, 2020.

- [3] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger. Computational models for active matter. *Nature Reviews Physics*, pages 181–199, 2020.
- [4] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [5] J. Elgeti, R. G. Winkler, and G. Gompper. Physics of microswimmers – single particle motion and collective behavior: a review. *Report on Progress in Physics*, 78:056601, 2015.
- [6] J. K. G. Dhont. *An Introduction to Dynamics of Colloids*. Elsevier, 1996.
- [7] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2002.
- [8] U. D. Schiller, T. Krüger, and O. Henrich. Mesoscopic modelling and simulation of soft matter. *Soft Matter*, 14:9–26, 2018.
- [9] G. R. McNamara and G. Zanetti. Use of the Boltzmann equation to simulate lattice-gas automata. *Physical Review Letters*, 61:2332–2335, 1988.
- [10] S. Succi. *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*. Oxford University Press, 2001.
- [11] P. J. Hoogerbrugge and J. M. V. A. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhysics Letters*, 19:155, 1992.
- [12] P. Espa  ol and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhysics Letters*, 30:191, 1995.
- [13] R. D. Groot and P. B. Warren. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *The Journal of Chemical Physics*, 107:4423–4435, 1997.
- [14] A. Malevanets and R. Kapral. Mesoscopic model for solvent dynamics. *The Journal of Chemical Physics*, 110:8605–8613, 1999.
- [15] A. Malevanets and R. Kapral. Solute molecular dynamics in a mesoscale solvent. *The Journal of Chemical Physics*, 112:7260–7269, 2000.
- [16] J. M. Yeomans. Mesoscale simulations: Lattice Boltzmann and particle algorithms. *Physica A: Statistical Mechanics and its Applications*, 369:159–184, 2006.
- [17] R. Kapral. *Multiparticle Collision Dynamics: Simulation of Complex Systems on Mesoscales*, pages 89–146. John Wiley & Sons, Ltd, 2008.
- [18] G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler. Multi-particle collision dynamics: A particle-based mesoscale simulation approach to the hydrodynamics of complex fluids. *Advances in Polymer Science*, 221:1–87, 2009.
- [19] G. A. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Oxford University Press, 1976.
- [20] F. J. Alexander, A. L. Garcia, and B. J. Alder. A consistent Boltzmann algorithm. *Physical Review Letters*, 74:5212–5215, 1995.
- [21] M. Ripoll, K. Mussawisade, R. G. Winkler, and G. Gompper. Low-Reynolds-number hydrodynamics of complex fluids by multi-particle-collision dynamics. *Europhysics Letters*, 68:106, 2004.
- [22] K. Mussawisade, M. Ripoll, R. G. Winkler, and G. Gompper. Dynamics of polymers in a particle-based mesoscopic solvent. *The Journal of Chemical Physics*, 123:144905, 2005.
- [23] M. Belushkin, R. Livi, and G. Foffi. Hydrodynamics and the fluctuation theorem. *Physical Review Letters*, 106:210601, 2011.

- [24] J. T. Padding and A. A. Louis. Hydrodynamic and Brownian fluctuations in sedimenting suspensions. *Physical Review Letters*, 93:220601, 2004.
- [25] A. Wysocki, C. P. Royall, R. G. Winkler, G. Gompper, H. Tanaka, A. van Blaaderen, and H. Löwen. Direct observation of hydrodynamic instabilities in a driven non-uniform colloidal dispersion. *Soft Matter*, 5:1340–1344, 2009.
- [26] M. Ripoll, P. Holmqvist, R. G. Winkler, G. Gompper, J. K. G. Dhont, and M. P. Lettinga. Attractive colloidal rods in shear flow. *Physical Review Letters*, 101:168302, 2008.
- [27] Marisol Ripoll. Helicopter rotation and smectic-isotropic coexistence of strongly attractive rods. *Physical Review E*, 83:040701, 2011.
- [28] M. Ripoll, R. G. Winkler, and G. Gompper. Star polymers in shear flow. *Physical Review Letters*, 96:188302, 2006.
- [29] Sunil P. Singh, Roland G. Winkler, and Gerhard Gompper. Nonequilibrium forces between dragged ultrasoft colloids. *Physical Review Letters*, 107:158301, 2011.
- [30] D. A. Fedosov, S. P. Singh, A. Chatterji, R. G. Winkler, and G. Gompper. Semidilute solutions of ultra-soft colloids under shear flow. *Soft Matter*, 8:4109–4120, 2012.
- [31] S. Frank and R. G. Winkler. Polyelectrolyte electrophoresis: Field effects and hydrodynamic interactions. *Europhysics Letters*, 83:38004, 2008.
- [32] H. Noguchi and G. Gompper. Dynamics of fluid vesicles in shear flow: Effect of membrane viscosity and thermal fluctuations. *Physical Review E*, 72:011901, 2005.
- [33] J. L. McWhirter, H. Noguchi, and G. Gompper. Flow-induced clustering and alignment of vesicles and red blood cells in microcapillaries. *Proceedings of the National Academy of Sciences*, 106(15):6039–6043, 2009.
- [34] J. Elgeti, U. B. Kaupp, and G. Gompper. Hydrodynamics of sperm cells near surfaces. *Biophysical Journal*, 99:1018–1026, 2010.
- [35] I. O. Götze and G. Gompper. Mesoscale simulations of hydrodynamic squirmers interactions. *Physical Review E*, 82:041921, 2010.
- [36] Y. Yang, V. Marceau, and G. Gompper. Swarm behavior of self-propelled rods and swimming flagella. *Physical Review E*, 82:031904, 2010.
- [37] S. Y. Reigh, R. G. Winkler, and G. Gompper. Synchronization and bundling of anchored bacterial flagella. *Soft Matter*, 8:4363–4372, 2012.
- [38] J. Elgeti and G. Gompper. Emergence of metachronal waves in cilia arrays. *Proceedings of the National Academy of Sciences*, 110:4470–4475, 2013.
- [39] G. Rückner and R. Kapral. Chemically powered nanodimers. *Physical Review Letters*, 98:150603, 2007.
- [40] P. de Buyl and R. Kapral. Phoretic self-propulsion: a mesoscopic description of reaction dynamics that powers motion. *Soft Matter*, 5:1337–1344, 2013.
- [41] M. Yang, A. Wysocki, and M. Ripoll. Hydrodynamic simulations of self-phoretic microswimmers. *Soft Matter*, 10:6208–6218, 2014.
- [42] O. Pohl and H. Stark. Dynamic clustering and chemotactic collapse of self-phoretic active particles. *Physical Review Letters*, 112:238303, 2014.
- [43] A. Zöttl and H. Stark. Hydrodynamics determines collective motion and phase behavior of active colloids in quasi-two-dimensional confinement. *Physical Review Letters*, 112:118101, 2014.

- [44] M. Wagner and M. Ripoll. Hydrodynamic front-like swarming of phoretically active dimeric colloids. *Europhysics Letters*, 119:66007, 2017.
- [45] M. Yang and M. Ripoll. A self-propelled thermophoretic microgear. *Soft Matter*, 10:1006–1011, 2014.
- [46] M. Yang, R. Liu, M. Ripoll, and K. Chen. A microscale thermophoretic turbine driven by external diffusive heat flux. *Nanoscale*, 6:13550–13554, 2014.
- [47] Sebastian Mühlbauer, Severin Strobl, and Thorsten Pöschel. Isotropic stochastic rotation dynamics. *Physical Review Fluids*, 2:124204, 2017.
- [48] E. Tüzel, M. Strauss, T. Ihle, and D. M. Kroll. Transport coefficients for stochastic rotation dynamics in three dimensions. *Physical Review E*, 68:036701, 2003.
- [49] E. Allahyarov and G. Gompper. Mesoscopic solvent simulations: Multiparticle-collision dynamics of three-dimensional flows. *Physical Review E*, 66:036702, 2002.
- [50] T. Ihle and D. M. Kroll. Stochastic rotation dynamics: A Galilean-invariant mesoscopic model for fluid flow. *Physical Review E*, 63:020201, 2001.
- [51] T. Ihle and D. M. Kroll. Stochastic rotation dynamics. I. Formalism, Galilean invariance, and Green-Kubo relations. *Physical Review E*, 67:066705, 2003.
- [52] A. Lamura, G. Gompper, T. Ihle, and D. M. Kroll. Multi-particle collision dynamics: Flow around a circular and a square cylinder. *Europhysics Letters*, 56:319, 2001.
- [53] D. J. Tritton. *Physical Fluid Dynamics*. Oxford University Press, 1988.
- [54] M. Ripoll, K. Mussawisade, R. G. Winkler, and G. Gompper. Dynamic regimes of fluids simulated by multiparticle-collision dynamics. *Physical Review E*, 72:016701, 2005.
- [55] N. Kikuchi, C. M. Pooley, J. F. Ryder, and J. M. Yeomans. Transport coefficients of a mesoscopic fluid dynamics model. *The Journal of Chemical Physics*, 119:6388–6395, 2003.
- [56] T. Ihle and D. M. Kroll. Stochastic rotation dynamics. II. Transport coefficients, numerics, and long-time tails. *Physical Review E*, 67:066706, 2003.
- [57] I. Pagonabarraga, M. H. J. Hagen, and D. Frenkel. Self-consistent dissipative particle dynamics algorithm. *Europhysics Letters*, 42:377, 1998.
- [58] M. Yang and M. Ripoll. Thermophoretically induced flow field around a colloidal particle. *Soft Matter*, 9:4661–4671, 2013.
- [59] E. Tüzel, T. Ihle, and D. M. Kroll. Dynamic correlations in stochastic rotation dynamics. *Physical Review E*, 74:056702, 2006.
- [60] D. Lüsebrink and M. Ripoll. Temperature inhomogeneities simulated with multiparticle-collision dynamics. *The Journal of Chemical Physics*, 136:084106, 2012.
- [61] C.-C. Huang, G. Sutmann, G. Gompper, and R. G. Winkler. Tumbling of polymers in semidilute solution under shear flow. *Europhysics Letters*, 93:54004, 2011.
- [62] C.-C. Huang, A. Varghese, G. Gompper, and R. G. Winkler. Thermostat for nonequilibrium multiparticle-collision-dynamics simulations. *Physical Review E*, 91:013310, 2015.
- [63] H. Noguchi, N. Kikuchi, and G. Gompper. Particle-based mesoscale hydrodynamic techniques. *Europhysics Letters*, 78:10005, 2007.
- [64] T. Ihle, E. Tüzel, and D. M. Kroll. Equilibrium calculation of transport coefficients for a fluid-particle model. *Physical Review E*, 72:046707, 2005.

- [65] J. T. Padding and A. A. Louis. Hydrodynamic interactions and brownian forces in colloidal suspensions: Coarse-graining over time and length scales. *Physical Review E*, 74:031402, 2006.
- [66] C. M. Pooley and J. M. Yeomans. Kinetic theory derivation of the transport coefficients of stochastic rotation dynamics. *The Journal of Physical Chemistry B*, 109(14):6505–6513, 2005.
- [67] Ingo O. Götze, Hiroshi Noguchi, and Gerhard Gompper. Relevance of angular momentum conservation in mesoscale hydrodynamics simulations. *Physical Review E*, 76:046705, 2007.
- [68] M. Yang, M. Theers, J. Hu, G. Gompper, R. G. Winkler, and M. Ripoll. Effect of angular momentum conservation on hydrodynamic simulations of colloids. *Physical Review E*, 92:013301, 2015.
- [69] T. Ihle, E. Tüzel, and D. M. Kroll. Consistent particle-based algorithm with a non-ideal equation of state. *Europhysics Letters*, 73:664, 2006.
- [70] E. Tüzel, G. Pan, T. Ihle, and D. M. Kroll. Mesoscopic model for the fluctuating hydrodynamics of binary and ternary mixtures. *Europhysics Letters*, 80:40010, 2007.
- [71] T. Eisenstecken, R. Hornung, R. G. Winkler, and G. Gompper. Hydrodynamics of binary-fluid mixtures – An augmented multiparticle collision dynamics approach. *Europhysics Letters*, 121:24003, 2018.
- [72] R. G. Winkler, K. Mussawisade, M. Ripoll, and G. Gompper. Rod-like colloids and polymers in shear flow: a multi-particle-collision dynamics study. *Journal of Physics: Condensed Matter*, 16:S3941, 2004.
- [73] N. Kikuchi, A. Gent, and J. M. Yeomans. Polymer collapse in the presence of hydrodynamic interactions. *The European Physical Journal E*, 9:63–66, 2002.
- [74] N. Kikuchi, J. F. Ryder, C. M. Pooley, and J. M. Yeomans. Kinetics of the polymer collapse transition: The role of hydrodynamics. *Physical Review E*, 71:061804, 2005.
- [75] M. Wagner and M. Ripoll. Solvent-induced depletion interactions in multiparticle collision dynamic simulations. *International Journal of Modern Physics C*, 30:1941008, 2019.
- [76] S. H. Lee and R. Kapral. Mesoscopic description of solvent effects on polymer dynamics. *The Journal of Chemical Physics*, 124:214901, 2006.
- [77] A. Lamura and G. Gompper. Numerical study of the flow around a cylinder using multi-particle collision dynamics. *The European Physical Journal E*, 9:477–485, 2002.
- [78] J. T. Padding, A. Wysocki, H. Löwen, and A. A. Louis. Stick boundary conditions and rotational velocity auto-correlation functions for colloidal particles in a coarse-grained representation of the solvent. *Journal of Physics: Condensed Matter*, 17:S3393, 2005.
- [79] H. Noguchi and G. Gompper. Fluid vesicles with viscous membranes in shear flow. *Physical Review Letters*, 93:258102, 2004.
- [80] M. Ripoll, R. G. Winkler, and G. Gompper. Hydrodynamic screening of star polymers in shear flow. *The European Physical Journal E*, 23:349–354, 2007.
- [81] J. Elgeti and G. Gompper. Self-propelled rods near surfaces. *Europhysics Letters*, 85:38002, 2009.
- [82] M. P. Lettinga, J. K. G. Dhont, Z. Zhang, S. Messlinger, and G. Gompper. Hydrodynamic interactions in rod suspensions with orientational ordering. *Soft Matter*, 6:4556–4562, 2010.
- [83] T. Ishikawa, M. P. Simmonds, and T. J. Pedley. Hydrodynamic interaction of two swimming model micro-organisms. *Journal of Fluid Mechanics*, 568:119–160, 2006.
- [84] E. Lauga and T. R. Powers. The hydrodynamics of swimming microorganisms. *Reports on Progress in Physics*, 72:096601, 2009.
- [85] Y. Yang, J. Elgeti, and G. Gompper. Cooperation of sperm in two dimensions: Synchronization, attraction, and aggregation through hydrodynamic interactions. *Physical Review E*, 78:061903, 2008.

- [86] J. Hu, M. Yang, G. Gompper, and R. G. Winkler. Modelling the mechanics and hydrodynamics of swimming *E. coli*. *Soft Matter*, 11:7867–7876, 2015.
- [87] J. Hu, A. Wysocki, R. G. Winkler, and G. Gompper. Physical sensing of surface properties by microswimmers – Directing bacterial motion via wall slip. *Scientific Reports*, 5:9586, 2015.
- [88] J. L. Anderson. Colloid transport by interfacial forces. *Annual Review of Fluid Mechanics*, 21:61–99, 1989.
- [89] S. Thakur and R. Kapral. Collective dynamics of self-propelled sphere-dimer motors. *Physical Review E*, 85:026121, 2012.
- [90] R. Kapral. Perspective: Nanomotors without moving parts that propel themselves in solution. *The Journal of Chemical Physics*, 138(2):020901, 2013.
- [91] M. Yang and M. Ripoll. Simulations of thermophoretic nanoswimmers. *Physical Review E*, 84:061401, 2011.
- [92] D. Lüsebrink, M. Yang, and M. Ripoll. Thermophoresis of colloids by mesoscale simulations. *Journal of Physics: Condensed Matter*, 24:284132, 2012.
- [93] P. H. Colberg and R. Kapral. Many-body dynamics of chemically propelled nanomotors. *The Journal of Chemical Physics*, 147:064910, 2017.
- [94] M. J. Lighthill. On the squirming motion of nearly spherical deformable bodies through liquids at very small reynolds numbers. *Communications on Pure and Applied Mathematics*, 5:109–118, 1952.
- [95] M. Theers, E. Westphal, G. Gompper, and R. G. Winkler. Modeling a spheroidal microswimmer and cooperative swimming in a narrow slit. *Soft Matter*, 12:7372, 2016.
- [96] J. Clopés, G. Gompper, and R. G. Winkler. Hydrodynamic interactions in squirmer dumbbells: active stress-induced alignment and locomotion. *Soft Matter*, 16:10676–10687, 2020.

Chapter 24

Lattice Boltzmann Method

RODRIGO COELHO, NUNO ARAÚJO, IGNACIO PAGONABARRAGA

The peculiar specificities of wet active matter systems pose serious challenges to their numerical simulation. Basic methods, such as *molecular dynamics* (MD), for both the active and solvent particles are ruled out because they cannot reach the relevant time and length scales in which colloids evolve. MD can be used to explore the behavior at short times (usually at the expense of not resolving appropriately the separation in time scales between solvent and active particle relaxation) [1]. Nonetheless, this approach is useful because it takes the solvent molecular details into account and can address basic issues, as for example the relation between friction and diffusion, or the validity of effective boundary conditions used on effective models and theories [2].

In order to reach longer time scales and to enforce the appropriate scale separation between active particles and the embedding molecular solvent, it is necessary to reduce the degrees of freedom of the solvent and introduce coarse grained descriptions that focus on a set of prescribed solvent variables. Such a choice requires physical insight, and different approaches have been followed during recent years. The methods discussed in Chapters 16 and 17, based on the Langevin picture of particle/solvent interaction, considers the solvent as a passive medium that exerts a friction on the colloids. In this method, *Brownian dynamics* (BD), a friction force proportional to the particles velocity is added to the conservative colloid–colloid interactions, and a random

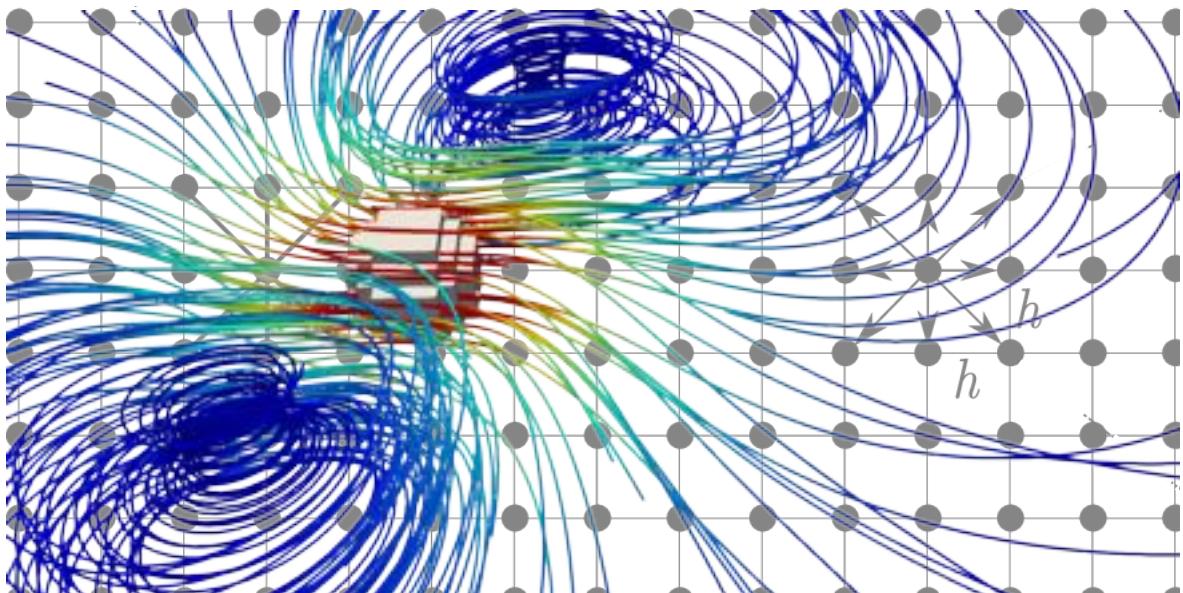


Figure 24.1: **Illustration of an active particle moving in a fluid simulated with the lattice Boltzmann method.** The colors in the streamlines represent the velocity magnitude.

force is also included to ensure that the fluctuation–dissipation theorem is fulfilled and that proper thermal equilibrium can be reached [3]. Even if this method has proved very useful in the exploration of the structures that active particles may form, and in the dynamics on time scales in which configuration relaxes, it misses the fact that momentum conservation gives rise to induced solvent flows in response to active particle motion.

In an attempt to ensure a more realistic description of the solvent, in recent years there has been a growing interest in the development and use of *mesoscopic simulation methods* that capture, at a coarse grained level, the collective modes of the solvent. The goal is to eliminate degrees of freedom of the solvent, so that the computational effort can be focused on the study of active particles, while ensuring that momentum is conserved, a basic feature missing in BD. One approach consists in the use of particle-based methods, in the same spirit as molecular methods such as MD, but introducing effective forces. It is well known that the reduction of degrees of freedom introduces friction forces. One can accordingly extend MD schemes by including friction and random forces such that equilibrium is recovered. The use of momentum conserving thermostats has opened the possibility to enforce the development of collective hydrodynamic modes on shorter scales than in standard MD. For passive particles, *dissipative particle dynamics* (DPD) proposed a thermostat where there is a friction force that depends on the relative velocity between particles, combined with pairwise additive random forces [4]. The amplitudes of these two forces should be carefully tuned to ensure that the fluctuation–dissipation theorem is satisfied in equilibrium. An alternative momentum-conserving thermostat has been proposed by Lowe [5], which is a local version of the traditional *Anderson thermostat*, and which captures proper hydrodynamics. For the *DPD thermostat*, theoretical analysis has proved how the Navier–Stokes equation follows from these non-Hamiltonian forces [6]. However, in order to reach meaningful time scales, these thermostats should be combined with effective conservative forces. At the simplest level, these are regarded as coming from soft potentials, although there is no derivation from first principle that allows us to link the parameters appearing in such potentials with their atomic counterparts [7]. Such forces allow the use of longer time steps, hence favoring reaching meaningful time scales. Thus, MD becomes impractical to simulate macroscopic fluids, although it can be used to simulate fluids at the microscale [8].

An alternative approach is based on the use of *kinetic theories*. The idea is to develop simplified models for the solvent that capture the collective solvent modes while disregarding any information related to its molecular character. Such approaches were introduced with the idea of producing minimalist dynamic models, analogous to the use of Ising models in critical phenomena, where the microscopic details are irrelevant. The *lattice gas cellular automata* (LGCA) [9, 10, 11] and the *lattice Boltzmann method* (LBM) [12, 13] have opened the door for kinetically-based models for fluids that can be coupled to colloids and other soft materials.

In order to overcome the computational limitations in the study of fluids at high Reynolds numbers, in the early nineties the LGCA was introduced to study fluid flow [14]. It was defined on a lattice, moves were restricted to a subset of neighbouring nodes, and single node occupation was enforced. Even if they were unconditionally stable and easily parallelizable, it was not possible to reach Reynolds numbers as large as expected. Moreover, the need to average fluctuations made them computationally intensive and they were mostly limited to two-dimensional applications since three-dimensional models were too complicated and expensive [12, 15].

In the context of LGCA, the LBM was introduced as a pre-averaged version of LGCA. It keeps both space and time discretization, but considers as the fundamental dynamical variable the particle distribution function. The set of allowed velocities joins a given node with a prescribed set of neighbors. Hence, space and time still remain discretized. The symmetry of the lattice and the minimum allowed set of velocities should conform to a minimum set of symmetry properties that ensure that the underlying anisotropy of the lattice does not affect the response of the system at the Navier–Stokes level [16]. The collision matrix determining the relaxation of the distribution function toward equilibrium was introduced at first as a result of averaging the collision rules of the underlying LGCA. However, in a second stage, this collision mechanism was substituted by a linearized collision matrix [17].

The simplicity of the dynamics and its suitability to parallel computing have made LBM a powerful tool. LBM does not discretize directly the hydrodynamic equations as other *computational fluid dynamics* (CFD) methods (e.g., finite-differences [18] and finite-volumes [19]), but it can be shown that it recovers them in the macroscopic limit. LBM can treat flows in complex geometries and moving boundaries more easily than other CFD methods. Due to its physical basis on the kinetic theory, it is also possible to implement many physical models [12] including, multiphase/multicomponent flows, advection–diffusion, thermal flows and, beyond classical hydrodynamics, semi-classical [20, 21], relativistic [22, 23, 24], and quantum systems [25, 12, 26]. It also has disadvantages depending on the application as every numerical method. For instance, it

is naturally time-dependent and thus not appropriate to find steady state solutions in general. It is also very memory-intensive since it stores not only the velocity and density fields but also a set of distribution functions for each node.

24.1 Fundamental concepts

The foundation of any advanced methodology lies in its fundamental concepts, which provide the structural framework necessary for understanding its principles and applications. In this section, we explore distribution functions, the Boltzmann equation, collective dynamics, and the Gauss–Hermite quadrature.

24.1.1 Distribution function

The relevant quantity in kinetic theory is the distribution function $f(\mathbf{x}, \boldsymbol{\xi}, t)$, which refers to the probability of finding any of the particles at position \mathbf{x} with velocity $\boldsymbol{\xi}$ at time t . Thus, the average mass of fluid particles positioned between \mathbf{x} and $\mathbf{x} + d^D \mathbf{x}$ with a velocity between $\boldsymbol{\xi}$ and $\boldsymbol{\xi} + d^D \boldsymbol{\xi}$ is $f(\mathbf{x}, \boldsymbol{\xi}, t) d^D \mathbf{x} d^D \boldsymbol{\xi}$, where D is the spatial dimension. The total mass of particles inside a volume V is:

$$M = \int_V \int f(\mathbf{x}, \boldsymbol{\xi}, t) d^D \boldsymbol{\xi} d^D \mathbf{x}, \quad (24.1)$$

where the integral in $\boldsymbol{\xi}$ runs from $-\infty$ to ∞ in each direction. Notice that the inner integral is the density field. The average momentum, $M\mathbf{U}$, is obtained by multiplying the distribution function by the microscopic velocity of each particle in the expression above: $M\mathbf{U} = \int_V \int f(\mathbf{x}, \boldsymbol{\xi}, t) \boldsymbol{\xi} d^D \boldsymbol{\xi} d^D \mathbf{x}$. The density and macroscopic velocity $\mathbf{u}(\mathbf{x}, t)$ of the fluid are then:

$$\rho(\mathbf{x}, t) = \int f(\mathbf{x}, \boldsymbol{\xi}, t) d^D \boldsymbol{\xi} \quad \text{and} \quad \mathbf{u}(\mathbf{x}, t) = \frac{1}{\rho(\mathbf{x}, t)} \int f(\mathbf{x}, \boldsymbol{\xi}, t) \boldsymbol{\xi} d^D \boldsymbol{\xi}. \quad (24.2)$$

Typically, the microscopic velocity $\boldsymbol{\xi}$ is much higher than the macroscopic fluid velocity \mathbf{u} , which is usually compared with the sound speed c_s , through the Mach number: $Ma = u/c_s$, where u is the absolute value of \mathbf{u} .

In principle, for each macroscopic set of parameters (density and macroscopic velocity field), one can derive the equilibrium distribution function f^{eq} . For an idealized gas composed of weakly interacting particles, this distribution is the Maxwell–Boltzmann distribution, which, in reduced units, is

$$f^{eq}(\rho, \mathbf{u}, \theta, \boldsymbol{\xi}) = \frac{\rho}{(2\pi\theta)^{\frac{D}{2}}} \exp\left[-\frac{(\boldsymbol{\xi} - \mathbf{u})^2}{2\theta}\right], \quad (24.3)$$

where $\theta = T/T_r$ is the non-dimensional temperature and T_r is a reference temperature. For simplicity, we consider a constant temperature ($\theta = 1$) since, for many practical problems, the effect of thermal fluctuations in the fluid flow is negligible. Nevertheless, there are LBM schemes that consider thermal fluctuations [27, 28].

As discussed in Sec. 24.1.4, the integration method used in LBM, called *Gauss–Hermite quadrature*, is exact for polynomial functions up to a certain order, which depends on the quadrature. For the most common quadratures, the integration is exact for polynomials up to second order. For this reason, it is convenient to expand f^{eq} in Hermite polynomials up to second order [29]:

$$f^{eq}(\rho, \mathbf{u}, \boldsymbol{\xi}) = \rho \omega(\boldsymbol{\xi}) \left[1 + \boldsymbol{\xi} \cdot \mathbf{u} - \frac{u^2}{2} + \frac{(\boldsymbol{\xi} \cdot \mathbf{u})^2}{2} \right] + \mathcal{O}(u^3), \quad (24.4)$$

where we assumed $\theta = 1$ and $Ma \ll 1$, which is equivalent to $u/\xi \ll 1$ since $\boldsymbol{\xi}$ is of the same order of magnitude as c_s . The weight function for the Hermite polynomials is given by

$$\omega(\boldsymbol{\xi}) = \frac{1}{(2\pi)^{\frac{D}{2}}} \exp\left(-\frac{\boldsymbol{\xi}^2}{2}\right), \quad (24.5)$$

which is essentially the Maxwell–Boltzmann distribution for $u = 0$.

24.1.2 Boltzmann equation

The Boltzmann equation describes the time evolution of the distribution function. Consider a particle in position \mathbf{x} with velocity $\boldsymbol{\xi}$ at time t . If it suffers no collision and is subject to an external force, and thus an acceleration \mathbf{a} , it will move to the position $(\mathbf{x} + \boldsymbol{\xi}\Delta t, \boldsymbol{\xi} + \mathbf{a}\Delta t)$ after a small time interval Δt . Thus, for a fluid of collisionless particles subject to a body force f , conservation of mass implies that

$$f(\mathbf{x}, \boldsymbol{\xi}, t) d^D \mathbf{x} d^D \boldsymbol{\xi} = f(\mathbf{x} + \boldsymbol{\xi}\Delta t, \boldsymbol{\xi} + \mathbf{a}\Delta t, t + \Delta t) d^D \mathbf{x} d^D \boldsymbol{\xi}. \quad (24.6)$$

If particles collide among each other, one also needs to account for the net change due to collisions:

$$f(\mathbf{x} + \boldsymbol{\xi}\Delta t, \boldsymbol{\xi} + \mathbf{a}\Delta t, t + \Delta t) - f(\mathbf{x}, \boldsymbol{\xi}, t) = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}} \Delta t, \quad (24.7)$$

where the derivative on the right-hand side is the collision operator. This expression is the discrete version of the Boltzmann equation. The first term on the left-hand side can be expanded up to first order in Δt , giving the Boltzmann equation in its continuous form:

$$\frac{\partial f}{\partial t} + \boldsymbol{\xi}_i \frac{\partial f}{\partial x_i} + a_i \frac{\partial f}{\partial \xi_i} = \left(\frac{\partial f}{\partial t} \right)_{\text{coll}}, \quad (24.8)$$

where the repeated indices represent a sum over the dimensions. Notice that the left-hand side of the above equation is simply the total time derivative of $f(\mathbf{x}, \boldsymbol{\xi}, t)$, since $\boldsymbol{\xi} = \partial \mathbf{x} / \partial t$ and $\mathbf{a} = \partial \boldsymbol{\xi} / \partial t$. It can be interpreted as an advection equation, where the first two terms are the material derivative of f , the third term accounts for the changes in the velocity due to an external force, and the collision operator is the source term.

The original collision operator used by Boltzmann considered all possible collisions between two particles for a generic interaction between them [30]. It would be very demanding to compute it numerically and also unnecessary if the focus is on the macroscopic properties. A much simpler approach has been proposed by Bhatnagar, Gross and Krook (BGK) [31], which is the common choice in LBM simulations. It consists in assuming that, if collisions tend to converge, within a time scale τ , the distribution function f to the equilibrium distribution f^{eq} compatible with the macroscopic density and fluid velocity [Check this sentence. Something is missing], i.e.,

$$\left(\frac{\partial f}{\partial t} \right)_{\text{coll}} = -\frac{f - f^{\text{eq}}}{\tau}. \quad (24.9)$$

All the information about the collision between the particles (e.g., viscosity) is in the relaxation time τ . To understand its meaning, consider a spatially homogeneous distribution $f(\boldsymbol{\xi}, t)$ in the absence of external forces. The Boltzmann equation with the *BGK operator* (which we will call *Boltzmann–BGK equation* from now on) gives an exponential convergence towards the equilibrium distribution with a characteristic time τ :

$$f(\boldsymbol{\xi}, t) = f^{\text{eq}}(\boldsymbol{\xi}) + [f(\boldsymbol{\xi}, 0) - f^{\text{eq}}(\boldsymbol{\xi})] e^{-t/\tau}. \quad (24.10)$$

The collision operator in Eq. (24.9) is the simplest used in LBM and sufficient for most applications. There are other operators which are more accurate for specific applications, but are essentially generalizations of the BGK one [32, 13].

24.1.3 Collective dynamics

It can be shown that the Boltzmann equation recovers Navier–Stokes, in the macroscopic limit. For this reason, the LBM, which is based on the kinetic theory, is used to simulate macroscopic fluids.

Let us start by calculating the first moment of the Boltzmann–BGK equation:

$$\frac{\partial}{\partial t} \int f d^D \boldsymbol{\xi} + \frac{\partial}{\partial x_i} \int \boldsymbol{\xi}_i f d^D \boldsymbol{\xi} + a_i \int \frac{\partial f}{\partial \xi_i} d^D \boldsymbol{\xi} = -\frac{\Delta t}{\tau} \int (f - f^{\text{eq}}) d^D \boldsymbol{\xi}. \quad (24.11)$$

The integrals in the first and second terms are the density and momentum, as defined in Eq. (24.2). The integral in the third term is the first moment of the forcing term, which can be interpreted as a source of mass. Since the forcing term cannot “create” mass, this integral must vanish. The integral in the right-hand side is also zero,

since $\int f d^D \xi = \int f^{\text{eq}} d^D \xi = \rho$, as the equilibrium distribution is compatible with the macroscopic quantities, as explained in Sec. 24.1.1. Thus,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (24.12)$$

which is the continuity equation.

The second moment of the Boltzmann–BGK equation is:

$$\frac{\partial}{\partial t} \int \xi_i f d^D \xi + \frac{\partial}{\partial x_j} \int \xi_i \xi_j f d^D \xi + a_j \int \xi_i \frac{\partial f}{\partial \xi_j} d^D \xi = -\frac{\Delta t}{\tau} \int \xi_i (f - f^{\text{eq}}) d^D \xi. \quad (24.13)$$

Notice that ξ is not an explicit function of position and time. The first integral is the momentum and the third one yields $\int \xi_i \frac{\partial f}{\partial \xi_j} d^D \xi = -\int \frac{\partial \xi_i}{\partial \xi_j} f d^D \xi = -\rho \delta_{ij}$. The right-hand side is zero since $\int \xi_i f d^D \xi = \int \xi_i f^{\text{eq}} d^D \xi = \rho u_i$. The second integral is the momentum flux tensor and can be decomposed as

$$\int \xi_i \xi_j f d^D \xi = \rho u_i u_j + \int \chi_i \chi_j f d^D \xi, \quad (24.14)$$

where $\boldsymbol{\chi} = \boldsymbol{\xi} - \mathbf{u}$ is the relative velocity of the particle from the macroscopic velocity. Thus, Eq. (24.13) becomes

$$\frac{\partial}{\partial t} (\rho u_i) + \frac{\partial}{\partial x_j} (\rho u_i u_j) = \frac{\partial \sigma_{ij}}{\partial x_j} + \rho a_i, \quad (24.15)$$

where

$$\sigma_{ij} = - \int \chi_i \chi_j f d^D \xi \quad (24.16)$$

is the stress tensor. Equation (24.15) is the *Cauchy momentum equation*. To compute the stress tensor, one relies on an approximation for the non-equilibrium distribution. The *Chapman–Enskog method* provides a way to expand the distribution function in terms of perturbations to the equilibrium distribution: $f = f^{\text{eq}} + \varepsilon f^{(1)} + \varepsilon^2 f^{(2)} + \dots$, where ε is the *Knudsen number*, defined as the ratio between the particle mean free path to a representative physical length scale. Interestingly, if $f \approx f^{\text{eq}}$ in Eq. (24.15), the Euler equation for ideal fluids is obtained.

For the full calculation of the stress tensor, see e.g. Chapter 4 of Ref. [13], which provides, up to second order in the velocity field, the compressible Navier–Stokes equation:

$$\frac{\partial}{\partial t} (\rho u_i) + \frac{\partial}{\partial x_j} (\rho u_i u_j) = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \left[\eta \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right] + \rho a_i, \quad (24.17)$$

where the pressure is $p = \rho c_s^2$ and the shear viscosity is $\eta = \rho c_s^2 (\tau - \frac{\Delta t}{2}) = \rho v$, with v being the kinematic viscosity. Although it is not visible in Eq. (24.17), this analysis also provides the bulk viscosity as $\eta_B = 2\eta/3$. If the flow can be considered incompressible ($\nabla \cdot \mathbf{u} = 0$), Eq. (24.17) can be simplified to the incompressible Navier–Stokes equation:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla p + v \nabla^2 \mathbf{u} + \mathbf{a}. \quad (24.18)$$

It is possible to go beyond the hydrodynamic regime, analyzing the hydrodynamic modes of LBM at finite wave vectors. In the case of an ideal gas equation of state, it has been shown that such dependence is analytic in the wave vector [33]. This fact shows one of the useful aspects of working with a kinetic equation, rather than with the Navier–Stokes equations themselves. Even if LBM is tailored to reproduce hydrodynamics, the underlying kinetic model ensures a physically meaningful behavior in regimes that go beyond the hydrodynamic limit, avoiding an uncontrolled response in those regimes. Another important aspect related to the kinetic-equation structure of LBM modeling is the fact that the collision matrix can be also used to model the decay of the fast (non-conserved) modes. Although in general models this feature is not used (and such eigenvalues are set to their maximum value to ensure a proper separation of time scale that ensure proper hydrodynamic behavior), tuning them appropriately can be used to model, e.g., viscoelasticity [34].

24.1.4 Gauss–Hermite quadrature

In LBM, the moments in Eq. (24.2) are calculated for every position and time step, which is done in a discretized space-time as will be discussed in Sec. 24.2.1. An efficient way to calculate the integrals is through the Gauss–Hermite quadrature, which approximates integrals by a sum:

$$\int \omega(\xi) g(\xi) d\xi \approx \sum_{\alpha} w_{\alpha} g_{\alpha}, \quad (24.19)$$

where the continuous weight function $\omega(\xi)$ is replaced by discrete weights w_{α} and the generic function $g(\xi)$ is calculated only on a few points: $g_{\alpha} \equiv g(\xi_{\alpha})$. This integration method becomes exact if the function $g(\xi)$ is a polynomial up to a certain order that depends on the quadrature.

In the Gauss–Hermite quadrature, the weight function is the weight used in the Hermite polynomials, Eq. (24.5). Thus, the equalities below must hold for a given order N of the monomials:

$$\int \omega(\xi) \xi_{i_1} \xi_{i_2} \dots \xi_{i_N} d^D \xi = \sum_{\alpha=0}^{q-1} w_{\alpha} \xi_{i_1, \alpha} \xi_{i_2, \alpha} \dots \xi_{i_N, \alpha}, \quad (24.20)$$

where i_1, \dots, i_N denote the spatial dimensions and α the points of the quadrature. This gives, for the first orders:

$$\sum_{\alpha} w_{\alpha} = 1 \quad (24.21)$$

$$\sum_{\alpha} w_{\alpha} \xi_{i_1, \alpha} = 0 \quad (24.22)$$

$$\sum_{\alpha} w_{\alpha} \xi_{i_1, \alpha} \xi_{i_2, \alpha} = \delta_{i_1 i_2} \quad (24.23)$$

$$\sum_{\alpha} w_{\alpha} \xi_{i_1, \alpha} \xi_{i_2, \alpha} \xi_{i_3, \alpha} = 0 \quad (24.24)$$

$$\sum_{\alpha} w_{\alpha} \xi_{i_1, \alpha} \xi_{i_2, \alpha} \xi_{i_3, \alpha} \xi_{i_4, \alpha} = \delta_{i_1 i_2} \delta_{i_3 i_4} + \delta_{i_1 i_3} \delta_{i_2 i_4} + \delta_{i_1 i_4} \delta_{i_2 i_3} \quad (24.25)$$

$$\sum_{\alpha} w_{\alpha} \xi_{i_1, \alpha} \xi_{i_2, \alpha} \xi_{i_3, \alpha} \xi_{i_4, \alpha} \xi_{i_5, \alpha} = 0 \quad (24.26)$$

...

The usual nomenclature for quadratures in LBM, also known as *lattices*, is $DdQq$, where d is the spatial dimension and q is the number of velocity vectors, or points, where the integrand is calculated. Let us consider as an example the D2Q9, which is standard in 2D simulations. The velocity vectors for this lattice, shown in Fig. 24.2, have components:

$$\begin{aligned} c_x &= (0, 1, 0, -1, 0, 1, -1, -1, 1), \\ c_y &= (0, 0, 1, 0, -1, 1, 1, -1, -1), \end{aligned} \quad (24.27)$$

and discrete weights

$$w_{\alpha} = \begin{cases} w_0, & \text{if } \alpha = 0, \\ w_s, & \text{if } \alpha = 1, 2, 3, 4, \\ w_l, & \text{if } \alpha = 5, 6, 7, 8. \end{cases} \quad (24.28)$$

Besides the weights, another parameter of the lattice to be found is the speed of sound c_s , which gives $\xi_{\alpha} = \mathbf{c}_{\alpha}/c_s$. The lattice velocities \mathbf{c} have integer components and c_s can be interpreted as the inverse of the lattice parameter. Using Eq. (24.27) in Eqs. (24.21) to (24.26), one obtains a set of four equations, with four unknowns (the weights and the speed of sound). Notice that Eqs. (24.22), (24.24) and (24.26) lead to trivial equations, while Eq. (24.25) leads to two non-trivial equations (one for all indices being the same and other for indices equal in pairs). From them, we obtain

$$w_0 = \frac{4}{9}, \quad w_s = \frac{1}{9}, \quad w_l = \frac{1}{36}, \quad c_s = \frac{1}{\sqrt{3}}. \quad (24.29)$$

For integrals of monomials beyond the fifth order, Eq. (24.26), the D2Q9 lattice is not valid anymore. This is sufficient to calculate the two moments of the distribution function, Eq. (24.2), using the second order expansion of f^{eq} . For polynomials with other weight functions and higher order expansions, see Ref. [20, 35].

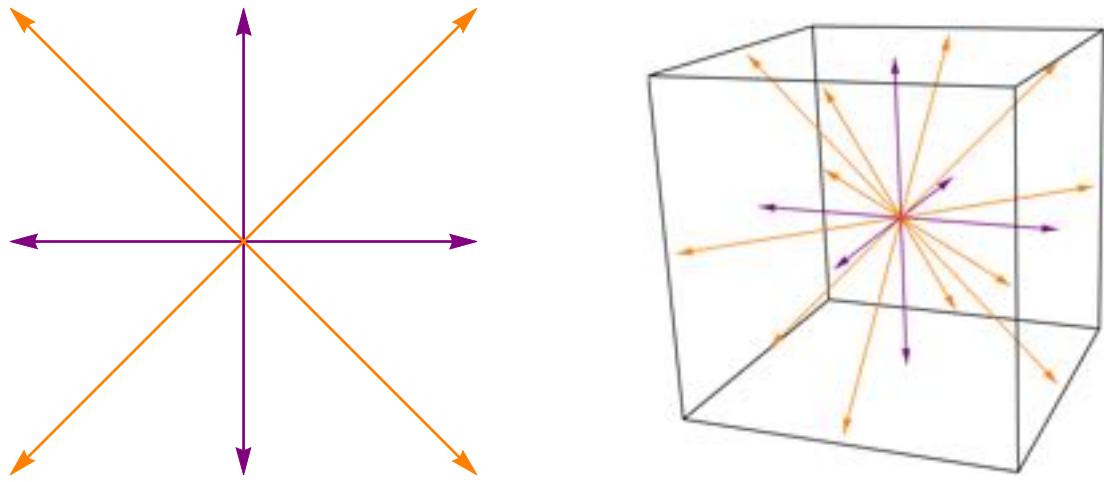


Figure 24.2: **Velocity vectors.** Velocity vectors of the D2Q9 (on the left) and D3Q19 (on the right) lattices. Vectors with the same color have the same length.

24.2 Numerical implementation

In this section, we will describe more practical aspects of the LBM implementation.

24.2.1 Discretization of the phase space

In LBM simulations, the position space is discretized in squared or simple cubic lattices. Hexagonal and other variants are also possible, but less usual and more complicated to apply boundary conditions to. The velocity space is also discretized following the chosen lattice. For instance, for the D2Q9 discussed in the previous section, the particles can be at rest ($\alpha = 0$), go to the first neighbours ($\alpha = 1, \dots, 4$), or go to the second neighbors ($\alpha = 5, \dots, 8$). Going further is not allowed. Due to these interactions between the closest neighbors and due to the space discretization in a regular lattice, the LBM simulations are efficiently implemented in parallel architectures in CPUs and GPUs [36, 37].

In the discrete phase space, Eq. (24.4) becomes

$$f_\alpha^{\text{eq}} = \rho w_\alpha \left[1 + \frac{\mathbf{c}_\alpha \cdot \mathbf{u}}{c_s^2} - \frac{u^2}{2c_s^2} + \frac{(\mathbf{c}_\alpha \cdot \mathbf{u})^2}{2c_s^4} \right], \quad (24.30)$$

where the macroscopic velocity \mathbf{u} is in the same units as \mathbf{c}_α , and thus divided by c_s . The moments in Eq. (24.2) are calculated using sums:

$$\rho = \sum_\alpha f_\alpha, \quad \mathbf{u} = \frac{1}{\rho} \sum_\alpha \mathbf{c}_\alpha f_\alpha. \quad (24.31)$$

The Boltzmann–BGK equation is split in two steps for implementation purposes: a *collision step*,

$$f_\alpha^{\text{aux}}(\mathbf{x}, t + \Delta t) = f_\alpha(\mathbf{x}, t) - \frac{f_\alpha - f_\alpha^{\text{eq}}}{\tau} \Delta t, \quad (24.32)$$

which is essentially the Boltzmann–BGK equation, but keeping the distributions in the same positions; and a *streaming step*,

$$f_\alpha(\mathbf{x} + \mathbf{c}_\alpha \Delta t, t + \Delta t) = f_\alpha^{\text{aux}}(\mathbf{x}, t + \Delta t), \quad (24.33)$$

which propagates the distributions to the neighbors according to the lattice. Notice that we need two sets of populations in the streaming step: the *current population* and the *buffer population*. This is because the propagation would erase the distributions of the neighbors if we had only one set of distributions. Equation (24.33) corresponds to the the streaming *scatter*, in which each node propagates its distributions to the neighbors. It is also possible to use a streaming *gather*, in which each node gather the distributions from the neighbors.

24.2.2 Algorithm and code hints

A basic algorithm for the LBM is (see the fluxogram in Fig. 24.3):

- 1) Initialization: Initialize the macroscopic fields ρ and \mathbf{u} .
- 2) Equilibrium distribution: Compute the equilibrium distributions using Eq. (24.30).
- 3) Collision: Compute the time evolution of the distributions using Eq. (24.32).
- 4) Streaming: Propagate the distributions to the neighbors, Eq. (24.33).
- 5) Moments update: Calculate the new ρ and \mathbf{u} using Eq. (24.31).
- 6) Back to step 2.

Each cycle corresponds to one time step Δt . In the next section, we will see how to implement boundary conditions and external forces to the fluid.

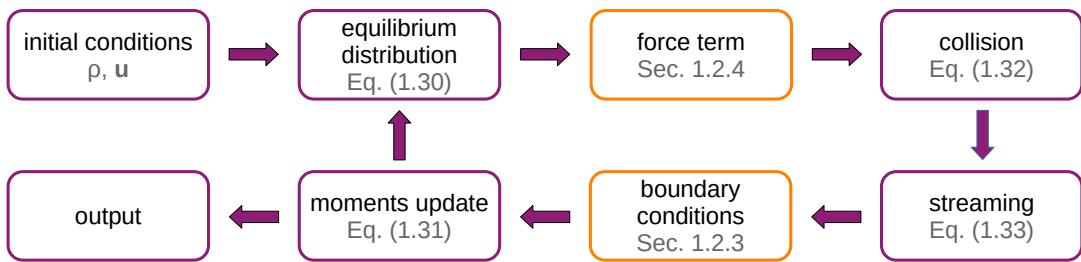


Figure 24.3: **LBM algorithm.** The orange boxes are optional steps, which will be discussed in the next sections.

We give here some implementation hints to optimize the code. First, N-dimensional (ND) arrays should be converted to 1D arrays to avoid inefficiencies as memory requirements increase. The conversion rules between ND and 1D should be consistent with their reading/writing, i.e., the 1D array should be read from the beginning to the end sequentially, without going back and forth. For example, the 3D coordinates of a scalar array can be mapped into a 1D index as

$$I_{3D}(z, y, x) = x + yL_X + zL_X L_Y, \quad (24.34)$$

where L_X , L_Y and L_Z are the domain dimensions. With this mapping, the loops to scan the positions should start with the x dimension, then go to the y dimension, and lastly the z dimension. Eq. (24.34) can be used to map the density and the components of the velocity field. The inverse mapping is

$$x = i \% L_X, \quad y = (i / L_X) \% L_Y \quad z = i / (L_X L_Y), \quad (24.35)$$

where $\%$ is the modulo operator, which gives the remainder of the division between two numbers. The distribution function has an extra dimension $\alpha = 0, \dots, q$, which indicates the velocity vectors in the lattice. Besides, we store the current and buffer distributions (see Eq. (24.33)) in the same vector, which will be explained next. Then, we add a fifth binary dimension b , which is $b = 1$ for the current and $b = 0$ for the buffer and the converter becomes

$$I_f(z, y, x, \alpha, b) = (x + yL_X + zL_X L_Y)q + \alpha + b(L_Z L_Y L_X q). \quad (24.36)$$

In this case, the reading/writing should start with α . In C/C++ language, these indices converters can be defined in the header using `#define` or as inline functions.

The next hint is to avoid the copying process to f^{aux} in the collision step. Initially, we define a variable `current` equal to 0 and another one `buffer` equal to 1. Then, up to the collision step, only one half of the distribution 1D vector, corresponding to the current one, will be used (see Eq. (24.36)). In the streaming step,

the distributions are propagated to the buffer (second half of the distribution 1D vector). After streaming, we call a function swap which converts buffer in current and vice-versa:

```
buffer = current
current = 1 - buffer.
```

Nowadays, the compilers already perform optimizations. For instance, with GNU compilers, one can compile with the flag `-O3`, which is a third level of optimization¹.

As mentioned before, LBM is naturally parallelizable due to its local operations in a lattice. A simple way to parallelize with little effort and already with fairly good results is to use openMP². The simplest is to add `omp_set_num_threads(omp_get_max_threads())` at the beginning of the main function in order to set the number of treads to be used (one can replace `omp_get_max_threads()` by an integer number to use a different number of threads). And add `#pragma omp parallel for collapse(3)` before each loop for, where `collapse(3)` is used to collapse the three loops in x , y , and z in only one, which will be parallelized with openMP. In the compilation, the flag `-fopenmp` must be used. Notice that some functions, e.g., the output, cannot be parallelized. For LBM implementations in MPI, see Palabos [36]³ and OpenLB⁴, and for parallelization in GPU, see Sailfish⁵.

Lastly, for visualizing the fields (density, velocity, pressure, etc) we recommend Paraview⁶. We make available a basic implementation of LBM in C++ with OpenMP at Ref. [38].

24.2.3 Boundary conditions

Here, we describe some useful *boundary conditions*. One needs to define the density and velocity at the boundaries, which can be, e.g., solid obstacles or the borders of the simulation domain.

The *periodic boundary conditions* are used to simulate periodic systems and they conserve density, momentum, and energy. They can be implemented as default in the code as they are commonly used in simulations and in order to avoid numerical instabilities. They can be simply implemented in the streaming step as

$$\begin{aligned}x_n &= (x + c_{x,\alpha} + L_X) \% L_X, \\y_n &= (y + c_{y,\alpha} + L_Y) \% L_Y, \\z_n &= (z + c_{z,\alpha} + L_Z) \% L_Z,\end{aligned}$$

where $\mathbf{x}_n = \mathbf{x} + \mathbf{c}_\alpha \Delta t$. This condition implies that, for example, the distributions that would go outside the system on the right are propagated to the first positions on the left. This condition can be used to solve 2D flow problems using a code able to simulate 3D flows. In this case, the periodic conditions is applied along one direction, which has only one node in length.

The *bounce-back boundary conditions* [39] are used to impose no-slip at solid surfaces. Initially, we label the nodes as being fluid or solid. This can be done with a boolean scalar field which is 1 inside the obstacles and 0 outside. For walls at rest, the distribution propagated to the solid node is reflected back in the opposite direction $\mathbf{c}_{\bar{\alpha}} = -\mathbf{c}_\alpha$ to the original fluid node (see Fig. 24.4):

$$f_{\bar{\alpha}}(\mathbf{x}_s, t + \Delta t) = f_\alpha^*(\mathbf{x}_s, t). \quad (24.37)$$

The vectors $\mathbf{c}_{\bar{\alpha}}$ are defined in the code together with the original ones and are used only in the streaming step, when the propagation finds a solid node. These boundary conditions can be easily used to simulate flows though complex geometries, such as porous media. The solid boundary (where the no-slip condition is imposed) is located approximately midway between the solid and fluid node. Note that curved boundaries with low resolutions become stair-shaped. To impose no-slip at interpolated curved surfaces, see Ref. [40, 41]. For a moving solid wall with velocity \mathbf{u}_s , the bounce-back method has an additional term:

$$f_{\bar{\alpha}}(\mathbf{x}_s, t + \Delta t) = f_\alpha^*(\mathbf{x}_s, t) - 2w_\alpha \rho_s \frac{\mathbf{c}_\alpha \cdot \mathbf{u}_s}{c_s^2}, \quad (24.38)$$

¹<https://gcc.gnu.org/onlinedocs/gcc-4.6.2/gcc/Optimize-Options.html>

²www.openmp.org/

³[https://palabos.unige.ch](http://palabos.unige.ch)

⁴[www.openlb.net](http://openlb.net)

⁵<http://sailfish.us.edu.pl>

⁶www.paraview.org

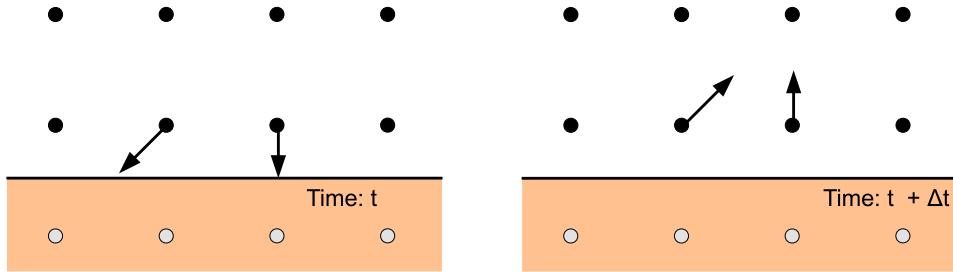


Figure 24.4: **Scheme of the bounce-back boundary condition.** The arrows represent the populations being reflected at the solid surface back to the fluid.

where the density of the solid ρ_s can be estimated as the local density of the fluid nodes in contact with the solid. This condition is also implemented in the streaming step.

For some problems, one wants to fix the density or velocity at the boundaries. This is the case, for instance, when flow is imposed in a channel with constant flow rate or constant pressure drop. A relatively simple to implement boundary conditions to impose such conditions is the *non-equilibrium extrapolation method* [42]. The desired solid density ρ_s and velocity \mathbf{u}_s are imposed in the equilibrium distribution of the solid while their non-equilibrium distribution $f^{neq} = f - f^{eq}$ are taken from the closest fluid neighbor:

$$f_\alpha(\mathbf{x}_b, t) = f_\alpha^{eq}(\rho_s, \mathbf{u}_s) + [f_\alpha(\mathbf{x}_f, t) - f_\alpha^{eq}(\rho_f, \mathbf{u}_f)], \quad (24.39)$$

where ρ_f and \mathbf{u}_f are the density and velocity of the fluid at \mathbf{x}_f . To implement this condition, the LBM steps discussed in Sec. 24.2.2 can be performed normally and, after the moments update, these boundary conditions are imposed to the distributions and to the moments at the solid.

24.2.4 Applied forces

Let us consider now the case where there is a *body force* applied to the fluid. It can be an external force (e.g., gravity) which drives the flow, an internal force (e.g., in multicomponent models [43, 44]), or a force to implement *force-based boundary conditions* (e.g., immersed boundary condition [45, 46]). The expression for the force term in the Boltzmann equation is calculated using the term $a_i \frac{\partial f}{\partial \xi_i}$ in Eq. (24.8). In the discrete space-time, the Boltzmann–BGK equation has a source term S_α :

$$f_\alpha(\mathbf{x} + \mathbf{c}_\alpha, t + \Delta t) = f_\alpha(\mathbf{x}, t) - \frac{f_\alpha - f_\alpha^{eq}}{\tau} \Delta t + S_\alpha. \quad (24.40)$$

This source term expanded up to second order, to be consistent with f^{eq} , reads

$$S_\alpha = \rho w_\alpha \left(1 - \frac{1}{2\tau}\right) \left[\frac{\mathbf{c}_\alpha}{c_s^2} + \frac{(\mathbf{c}_\alpha \cdot \mathbf{u}) \mathbf{c}_\alpha}{c_s^4} - \frac{\mathbf{u}}{c_s^2} \right] \cdot \mathbf{a}. \quad (24.41)$$

The fluid velocity in the presence of a force is redefined as

$$\mathbf{u} = \frac{1}{\rho} \sum_\alpha \mathbf{c}_\alpha f_\alpha + \frac{\mathbf{a} \Delta t}{2}. \quad (24.42)$$

This velocity must be used both in f_α^{eq} and in S_α as well as in the output. Thus, before the collision step (see Sec. 24.2.2), the velocity is redefined using Eq. (24.42), the source term is calculated using Eq. (24.41), and the collision with the source term is applied using Eq. (24.40). The streaming and moments update remain the same. This method for implementing forces is known as *Guo forcing scheme* [47].

24.3 Moving particles in a fluid

To simulate a solid particle moving in a fluid, we need to calculate the forces that the fluid exerts on the solid particle. The general approach is to integrate the fluid stresses acting on the solid surface:

$$F_i = \int \sigma_{ij} n_j dA, \quad (24.43)$$

where \mathbf{n} is the normal to the surface and dA its area element. This calculation in LBM simulations is laborious and involves calculations that are an additional source of numerical errors [40]. A much simpler approach is based on the momentum exchange $\Delta\mathbf{P}$ between solid and liquid [39, 48]. This is readily done by using the reflected distribution in the bounce-back boundary conditions (see Eq. (24.38)):

$$\mathbf{F} = \frac{\Delta\mathbf{P}}{\Delta t} = \frac{\Delta x^3}{\Delta t} \sum_{\mathbf{x}_s, \alpha} (f_\alpha^{\text{in}} + f_\alpha^{\text{out}}) \mathbf{c}_\alpha = \frac{\Delta x^3}{\Delta t} \sum_{\mathbf{x}_s, \alpha} \left[2f_\alpha^*(\mathbf{x}_s, t) - 2w_\alpha \rho_s \frac{\mathbf{c}_\alpha \cdot \mathbf{u}_s}{c_s^2} \right] \mathbf{c}_\alpha, \quad (24.44)$$

where the double sum runs over the solid nodes \mathbf{x}_s and the direction of lattice vectors α . Analogously, the torque is calculated as

$$\mathbf{T} = \frac{\Delta x^3}{\Delta t} \sum_{\mathbf{x}_s, \alpha} \left[2f_\alpha^*(\mathbf{x}_s, t) - 2w_\alpha \rho_s \frac{\mathbf{c}_\alpha \cdot \mathbf{u}_s}{c_s^2} \right] (\mathbf{x}_s - \mathbf{C}) \times \mathbf{c}_\alpha, \quad (24.45)$$

where \mathbf{C} is the point of reference. Thus, the solid velocity at each point is

$$\mathbf{u}_s = \mathbf{u}_{\text{CM}} + \boldsymbol{\Omega} \times (\mathbf{x}_s - \mathbf{C}), \quad (24.46)$$

where $\boldsymbol{\Omega}$ is the angular velocity. The particle dynamics is described by standard MD (see Chapter 16), where the fluid forces are considered.

As the solid moves through the fluid, some fluid nodes become solid nodes and vice versa. For the new solid nodes, the fluid information is erased and the fluid momentum is transferred to the solid. This is done by applying a force $\rho \mathbf{u} \Delta x^3 / \Delta t$ to the solid using the density and velocity of the destroyed fluid node. The creation of fluid nodes is more complicated since the fluid information is unknown at first, including the distributions. A simple approach [49] is to initialize the created node with the density equal to the average density of the neighbouring fluid nodes $\bar{\rho}$ (for instance, considering all nodes reached by the lattice vectors) and the velocity field equal to the solid velocity \mathbf{u}_s . The distributions are the equilibrium ones with these fields: $f = f^{\text{eq}}(\bar{\rho}, \mathbf{u}_s)$. This approach of destruction/creation of fluid nodes can also be employed in interpolated versions of bounce-back [50, 51, 40, 52]. An alternative method, introduced by Ladd [39, 53], consists in simulating the fluid inside the solid as well. The internal fluid also bounces back following Eq. (24.38). This avoids the creation and destruction of fluid nodes and conserves mass and momentum, but the dynamics of the internal fluid might affect the dynamics of the particle at high Reynolds numbers and it is limited to particles' densities larger than the fluid density.

As an example, we consider here an active particle in a fluid using LBM and the methods of MD discussed in Chapter 16. In a squared domain with dimension $L = 256$, we simulate a small particle with radius $R = 4$ in a fluid with kinematic viscosity $\nu = 0.1$ ($\tau = 0.8$). The domain is bounded by solid walls at the borders to avoid spurious advection currents [54]. The particle has density 50% higher than the fluid $\rho_s = 1.5\rho$. We model the activity as an external force $\mathbf{a} = |\mathbf{a}|(\cos \theta, \sin \theta)$ acting on the particle, where $|\mathbf{a}| = 5 \times 10^5$. The angle changes each $\Delta t' = 500$ LBM iterations accordingly to $\theta(t + \Delta t') = \theta(t) + \sqrt{2\Delta t' D_r} W_\theta(t)$. $W_\theta(t)$ is a random number following a Gaussian distribution with average 0 and standard deviation $\sigma = 1$. The particle reaches the overdamped regime rapidly. Thus, changing the direction of the external force is equivalent to changing the velocity of the particle as in particle based simulations in Chapter 2. Then, the position of the particle is updated based on this external force together with the fluid forces. Fig. 24.5 shows three samples of trajectories for three different values of D_r and the inset shows the velocity field close to the particle at a particular time step.

24.4 Problems

Problem 24.1: Expansion of the Maxwell–Boltzmann distribution in Hermite polynomials. Expand the

Maxwell–Boltzmann distribution, Eq. (24.3), up to second order in Hermite polynomials. Consider $\theta = 1$. See Ref. [21] for expansions up to fourth order.

Problem 24.2: Conservation of mass and momentum of the BGK operator. Show that the BGK collision operator, Eq. (24.9), conserves mass and momentum.

Problem 24.3: D3Q19 lattice. Calculate the weights of the D3Q19 lattice, Fig. 24.2, which has the following velocity vectors:

$$\begin{aligned} c_x &= (0, 1, -1, 0, 0, 0, 0, 1, -1, 1, -1, 1, -1, 1, -1, 0, 0, 0, 0, 0), \\ c_y &= (0, 0, 0, 1, -1, 0, 0, 1, 1, -1, -1, 0, 0, 0, 0, 1, -1, 1, -1), \\ c_z &= (0, 0, 0, 0, 0, 1, -1, 0, 0, 0, 0, 1, 1, -1, -1, 1, 1, -1). \end{aligned}$$

Answer:

$$w_0 = \frac{1}{3}, \quad w_s = \frac{1}{18}, \quad w_l = \frac{1}{36}, \quad c_s = \frac{1}{\sqrt{3}}.$$

Problem 24.4: Velocity field of a viscous incompressible fluid driven by a body force. Consider the flow of a viscous incompressible fluid driven by a body force \mathbf{a} (e.g., gravity) between two parallel static plates. Use Eq. (24.18) to show that the velocity field is given by

$$u_x(y) = \frac{a}{2\nu}y(L_Y - y). \quad (24.47)$$

Use your LBM code to solve this problem numerically and compare with the analytical result. Use bounce-back at the solid plates and wait until the flow converges. Note that the no-slip boundary conditions is applied in the midway between the solid and fluid nodes.

Problem 24.5: Vortex shedding behind an obstacle. Implement an inlet and outlet boundary condition with constant velocity V in a rectangular domain with a circular solid obstacle with diameter D centered at $(L_X/4, L_Y/2)$.

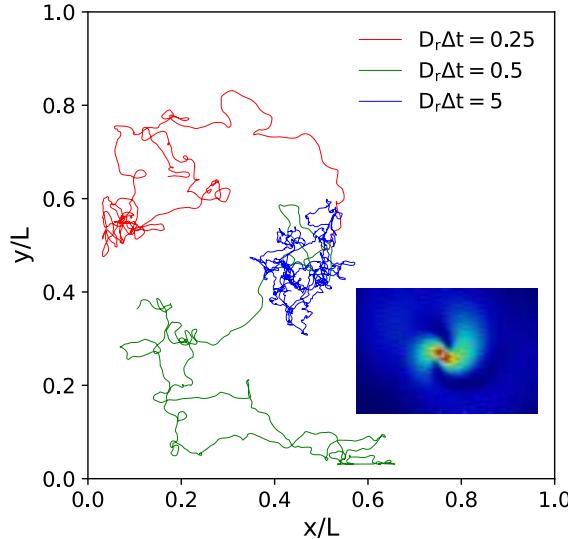


Figure 24.5: **Active particle in a fluid.** Trajectory of an active particle in a fluid for three different rotational diffusion coefficients D_r . The inset shows an example of the velocity field around the active particle, where the maximum velocity (red) is $|\mathbf{u}| = 0.0044$ l.u. Define l.u. Is it lattice units?

a. Observe that for high enough Reynolds number $\text{Re} = DV/v > 50$ the flow develops vortex shedding behind the cylinder.

b. Obtain that for 'Re = 100' the Strouhal number is $\text{St} = fD/V \approx 0.16$, where f is the shedding frequency [55].

Problem 24.6: Drag force. Consider a cylinder of diameter D centered in a squared domain with constant velocity V in the left and right borders and periodic boundary conditions in the vertical direction. Compute the drag force F_x in the static cylinder using Eq. (24.44). For $\text{Re} = DV/v = 10$, obtain that the drag coefficient is $C_D = \frac{2|F_x|}{\rho V^2 D} \approx 3.36$ [40].

Problem 24.7: Drag coefficient of a cylinder. Compute the drag coefficient of a cylinder (see Problem 24.6) in the two following situations:

a. First, a static cylinder with solid plates at the top and bottom moving with velocity V and periodic conditions in the horizontal direction. Bounce-back with moving walls or non-equilibrium extrapolation can be used to implement the moving plates.

b. Second, static plates and moving cylinder with velocity V .

c. Verify that the drag coefficient is approximately the same for small Reynolds numbers in the laminar regime.

Problem 24.8: Brownian particle with LBM. Simulate a passive Brownian particle (see Chapter 17) using the LBM model with a moving particle described in Sec. 24.3.

a. Measure the diffusion coefficient through the mean squared displacement and compare with the theoretical one. See Ref. [54].

Bibliography

- [1] M. J. Nuevo, J. J. Morales, and D. M. Heyes. Self-diffusion coefficients and shear viscosity of model nanocolloidal dispersions by molecular dynamics simulation. *Physical Review E*, 58:5845–5854, 1998.
- [2] F. Ould-Kaddour and D. Levesque. Determination of the friction coefficient of a Brownian particle by molecular-dynamics simulation. *The Journal of Chemical Physics*, 118:7888–7891, 2003.
- [3] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017.
- [4] P. Espa  ol and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhysics Letters (EPL)*, 30:191–196, 1995.
- [5] Lowe, C. P. An alternative approach to dissipative particle dynamics. *Europhysics Letters*, 47:145–151, 1999.
- [6] C. A. Marsh, G. Backx, and M. H. Ernst. Static and dynamic properties of dissipative particle dynamics. *Physical Review E*, 56:1676–1691, 1997.
- [7] E. G. Flekk  y and P. V. Coveney. From molecular dynamics to dissipative particle dynamics. *Physical Review Letters*, 83:1775–1778, 1999.
- [8] R. Giro, P. W. Bryant, M. Engel, R. F. Neumann, and M. B. Steiner. Adsorption energy as a metric for wettability at the nanoscale. *Scientific Reports*, 7, 2017.
- [9] P. Lallemand. Lattice-gas cellular automata, simple models of complex hydrodynamics. *European Journal of Mechanics B/Fluids*, 17:783–784, 1998.
- [10] J. Hardy, Y. Pomeau, and O. de Pazzis. Time evolution of a two-dimensional model system. I. Invariant states and time correlation functions. *Journal of Mathematical Physics*, 14:1746–1759, 1973.
- [11] U. Frisch, B. Hasslacher, and Y. Pomeau. Lattice-gas automata for the navier-stokes equation. *Physical Review Letters*, 56:1505–1508, 1986.

- [12] S. Succi. *The Lattice Boltzmann Equation: For Complex States of Flowing Matter*. Numerical Mathematics and Scientific Computation Series. Oxford University Press, 2018.
- [13] T. Krüger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva, and E. M. Viggen. *The Lattice Boltzmann Method - Principles and Practice*. Springer International Publishing, 10 2016.
- [14] U. Frisch, B. Hasslacher, and Y. Pomeau. Lattice-gas automata for the Navier-Stokes equation. *Physical Review Letters*, 56:1505–1508, 1986.
- [15] D. A. Wolf-Gladrow. *Lattice Gas Cellular Automata and Lattice Boltzmann Models*. Springer Berlin Heidelberg, 2000.
- [16] Stephen Wolfram. Cellular automaton fluids 1: Basic theory. *Journal of Statistical Physics*, 45(3-4):471–526, 1986.
- [17] F. J Higuera, S Succi, and R Benzi. Lattice gas dynamics with enhanced collisions. *Europhysics Letters (EPL)*, 9:345–349, 1989.
- [18] S. V. Patankar. *Numerical Heat Transfer and Fluid Flow*. CRC Press, 2018.
- [19] H. K. Versteeg and W. Malalasekera. *An introduction to computational fluid dynamics: the finite volume method*. Pearson education, 2007.
- [20] Rodrigo C. V. Coelho and Mauro M. Doria. Lattice Boltzmann method for semiclassical fluids. *Computers & Fluids*, 165:144 – 159, 2018.
- [21] R. C. V. Coelho, A. Ilha, M. M. Doria, R. M. Pereira, and V. Y. Aibe. Lattice Boltzmann method for bosons and fermions and the fourth-order hermite polynomial expansion. *Physical Review E*, 89:043302, 2014.
- [22] A. Gabbana, M. Mendoza, S. Succi, and R. Tripiccione. Towards a unified lattice kinetic scheme for relativistic hydrodynamics. *Physical Review E*, 95:053304, 2017.
- [23] A. Gabbana, M. Mendoza, S. Succi, and R. Tripiccione. Kinetic approach to relativistic dissipation. *Physical Review E*, 96:023305, 2017.
- [24] R. C. V. Coelho, M. Mendoza, M. M. Doria, and H. J. Herrmann. Fully dissipative relativistic lattice Boltzmann method in two dimensions. *Computers & Fluids*, 172:318–331, 2018.
- [25] S. Succi, F. Fillion-Gourdeau, and S. Palpacelli. Quantum lattice Boltzmann is a quantum walk. *EPJ Quantum Technology*, 2, 2015.
- [26] D. Lapitski and P. J. Dellar. Convergence of a three-dimensional quantum lattice Boltzmann scheme towards solutions of the Dirac equation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369:2155–2163, 2011.
- [27] I. V. Karlin, D. Sichau, and S. S. Chikatamarla. Consistent two-population lattice Boltzmann model for thermal flows. *Physical Review E*, 88:063310, 2013.
- [28] K. K. Mattila, P. C. Philippi, and L. A. Hegele. High-order regularization in lattice-Boltzmann equations. *Physics of Fluids*, 29:046103, 2017.
- [29] X. He and L.-S. Luo. Theory of the lattice Boltzmann method: From the Boltzmann equation to the lattice Boltzmann equation. *Physical Review E*, 56:6811–6817, 1997.
- [30] G. M. Kremer. *An Introduction to the Boltzmann Equation and Transport Processes in Gases*. Springer Berlin Heidelberg, 2010.
- [31] P. L. Bhatnagar, E. P. Gross, and M. Krook. A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. *Physical Review*, 94:511–525, 1954.

- [32] C. Coreixas, G. Wissocq, B. Chopard, and J. Latt. Impact of collision models on the physical properties and the stability of lattice Boltzmann methods. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378:20190397, 2020.
- [33] O. Behrend, R. Harris, and P. B. Warren. Hydrodynamic behavior of lattice Boltzmann and lattice Bhatnagar-Gross-Krook models. *Physical Review E*, 50:4586–4595, 1994.
- [34] R. Adhikari, K. Stratford, M. E. Cates, and A. J. Wagner. Fluctuating lattice Boltzmann. *Europhysics Letters (EPL)*, 71:473–479, 2005.
- [35] M. M. Doria and R. C. V. Coelho. Chebyshev, Legendre, Hermite and other orthonormal polynomials in D dimensions. *Reports on Mathematical Physics*, 81:243 – 271, 2018.
- [36] J. Latt, O. Malaspinas, D. Kontaxakis, A. Parmigiani, D. Lagrava, F. Brogi, M. Ben Belgacem, Y. Thorimbert, S. Leclaire, S. Li, F. Marson, J. Lemus, C. Kotsalos, R. Conradi, C. Coreixas, R. Petkantchin, F. Raynaud, J. Beny, and B. Chopard. Palabos: Parallel lattice Boltzmann solver. *Computers & Mathematics with Applications*, 81:334–350, 2021.
- [37] E. Calore, A. Gabbana, J. Kraus, E. Pellegrini, S. F. Schifano, and R. Tripiccione. Massively parallel lattice-Boltzmann codes on large GPU clusters. *Parallel Computing*, 58:1–24, 2016.
- [38] A. M. R. Maschio, M. Sedahmed, and R. C. V. Coelho. Implementation of the lattice Boltzmann method in CUDA C, 2022.
- [39] A. J. C. Ladd. Numerical simulations of particulate suspensions via a discretized Boltzmann equation. Part 1. Theoretical foundation. *Journal of Fluid Mechanics*, 271:285–309, 1994.
- [40] R. Mei, D. Yu, W. Shyy, and L.-S. Luo. Force evaluation in the lattice Boltzmann method involving curved geometry. *Physical Review E*, 65:041203, 2002.
- [41] A. F. V. Matias, R. C. V. Coelho, J. S. Andrade, and N. A. M. Araújo. Flow through time-evolving porous media: Swelling and erosion. *Journal of Computational Science*, 53:101360, 2021.
- [42] B. Shi, Z. Guo, and C. Zheng. Non-equilibrium extrapolation method for velocity and pressure boundary conditions in the lattice Boltzmann method. *Chinese Physics B*, 11:366, 2002.
- [43] X. Shan and H. Chen. Simulation of nonideal gases and liquid-gas phase transitions by the lattice Boltzmann equation. *Physical Review E*, 49:2941–2948, 1994.
- [44] D. P. F. Silva, R. C. V. Coelho, M. M. Telo da Gama, and N. A. M. Araújo. Effect of particle deformability on shear thinning in a 3D channel, 2021.
- [45] Z.-G. Feng and E. E. Michaelides. The immersed boundary-lattice Boltzmann method for solving fluid-particles interaction problems. *Journal of Computational Physics*, 195:602–628, 2004.
- [46] D. P. F. Silva, R. C. V. Coelho, I. Pagonabarraga, S. Succi, M. M. Telo da Gama, and N. A. M. Araújo. Lattice Boltzmann simulation of deformable fluid-filled bodies: progress and perspectives. *Soft Matter*, 20:2419–2441, 2024.
- [47] Z. Guo, C. Zheng, and B. Shi. Discrete lattice effects on the forcing term in the lattice Boltzmann method. *Physical Review E*, 65:046308, 2002.
- [48] A. Tiwari and S. P. Vanka. A ghost fluid lattice Boltzmann method for complex geometries. *International Journal for Numerical Methods in Fluids*, 69:481–498, 2012.
- [49] C. K. Aidun, Y. Lu, and E.-J. Ding. Direct analysis of particulate suspensions with inertia using the discrete Boltzmann equation. *Journal of Fluid Mechanics*, 373:287–311, 1998.
- [50] C. Pan, L.-S. Luo, and C. T. Miller. An evaluation of lattice Boltzmann schemes for porous medium flow simulation. *Computers & Fluids*, 35:898–909, 2006. Proceedings of the First International Conference for Mesoscopic Methods in Engineering and Science.

- [51] I. Ginzburg and D. d’Humières. Multireflection boundary conditions for lattice Boltzmann models. *Physical Review E*, 68:066614, 2003.
- [52] R. Jäger, M. Mendoza, and H. J. Herrmann. Clogging at pore scale and pressure-induced erosion. *Physical Review Fluids*, 3:074302, 2018.
- [53] A. J. C. Ladd. Numerical simulations of particulate suspensions via a discretized Boltzmann equation. Part 2. Numerical results. *Journal of Fluid Mechanics*, 271:311–339, 1994.
- [54] X. Xue, L. Biferale, M. Sbragaglia, and F. Toschi. A lattice Boltzmann study on Brownian diffusion and friction of a particle in a confined multicomponent fluid. *Journal of Computational Science*, 47:101113, 2020.
- [55] X. He and G. D. Doolen. Lattice Boltzmann method on a curvilinear coordinate system: Vortex shedding behind a circular cylinder. *Physical Review E*, 56:434–440, 1997.

Chapter 25

Lattice Boltzmann for Active Fluids

ŽIGA KOS AND MIHA RAVNIK

Lattice Boltzmann (LB) methods are emerging as a powerful tool for modeling active matter [1, 2]. As discussed in the previous Chapter, LB methods allow for efficient implementation of different geometries and boundary conditions, are highly parallelizable, and can access fair-sized experimentally relevant simulation volumes and times, already without super-computing resources. Implementations of LB modeling of active matter often draw from the implementations used and developed for modeling analogous passive (i.e., not active) matter systems, notably including liquid crystals and binary fluids.

The major strength of using lattice Boltzmann methods for active matter is their ability to embed, solve, or get coupled to different material fields (e.g., orientation fields, concentration, multi-phase parameters) in parallel to solving the material flow field (Fig. 25.1). Chapter 24 introduced the use of lattice Boltzmann methods for solving the Navier–Stokes equation, which for active fluids gets generalized and also coupled to additional dynamic equations of different material fields. LB methods can be used to solve the full set or only a part of this set of coupled dynamic equations (for example only the Navier–Stokes equation). In comparison to simulation of isotropic fluids, LB simulations of active fluids typically use lower-order accuracy LB schemes, often for reasons of simplicity, better code stability and lower computing resources.

This Chapter gives a selected overview of: (i) the LB methods for modeling selected active liquid crystal systems; (ii) the selected LB approaches used and developed first in passive fluids but now transferred to active

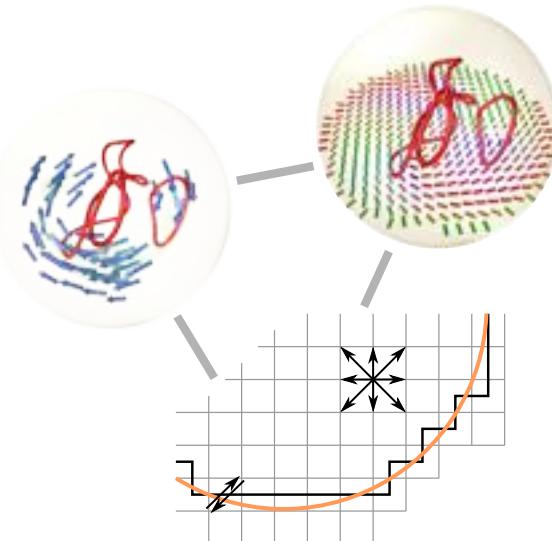


Figure 25.1: **Use of lattice Boltzmann for modeling active fluids.** Active flows (top left) and orientational order with topological defects (top right) within an active nematic confined into a spherical cavity.

fluids; (iii) the full implementation of the currently rather widely used hybrid LB method for active nematics; (iv) discussion, overview and challenges; and (v) a set of problems.

25.1 Lattice Boltzmann in active liquid crystal continuum models

LB methods are used in modeling different active matter systems, including active nematics [3, 4], active polar fluids [5, 6, 7], and active binary fluids [8, 9]. Generally, in these applications the LB method is used to solve the Navier–Stokes equation used in active models to account for the effects of hydrodynamics, as coupled to other active matter fields or variables. Alternatively, particle-resolved LB approaches are used to model individual active agents as affected by their environments [10].

25.1.1 Active nematics

Active nematics are a type of active system characterized by self-driven units with elongated shape [3], such as bacteria [11, 12, 13, 14] and cytoskeleton biopolymers [15, 16]. They are characterized by the average orientational order of their units, termed *director* \mathbf{n} , which notably exhibits head-tail symmetry, i.e., $\mathbf{n} \leftrightarrow -\mathbf{n}$. Theoretical approaches to active nematics are explained for example in Ref. [3]; here, we recall only that, at the mesoscopic continuum level, the approaches to active nematics generally require solving a set of coupled partial differential equations for the nematic order and the velocity field, and that, with the exception of few simple geometries, they need to be solved numerically. A widely used approach is to use the lattice Boltzmann method to solve the Navier–Stokes equation, which is coupled with the finite-difference approach for the orientational field or other fields, called *hybrid Lattice Boltzmann method*.

LB modeling in active nematics is used to model different 1D, 2D, quasi-2D and 3D geometries, including layers, films, tubes, spheres, droplets and bulk materials. The modeling provides particular insight into the symmetry and structure of the flow profiles [17] as well as the role and dynamics of topological defects [18, 19, 20], both in confined and non-confined geometries, as affected by the emergence of irregular chaotic motion -at low Reynolds numbers- of topological defects, called active turbulence [21, 22]. **Particular importance in the recent implementation of LB is also with the role of friction with the substrate [Check sentence, not clear]**, which was shown to affect the density and profiles of the active topological defects [23].

25.1.2 Active polar fluids

Active polar fluids are analogous to active nematic systems because they consist of self-driven units of elongated shape, but they are different because they exhibit vectorial (i.e., polar-orientational) order [24]. LB modeling was used to model the interplay between alignment and activity in active polar fluids, which was shown to lead to completely different results [Different results from what?], according to the geometry and nature of the activity [25]. Defect hydrodynamics in 2D polar active fluids was explored by means of lattice Boltzmann simulations [26], as well as fluid flow in finite droplets [7].

25.1.3 Active (nematic) binary fluids

Active binary fluids are systems that rely on the effective coexistence of two material components, such as in vesicles or droplets of active anisotropic materials in an isotropic host fluid [27, 28]. LB was used to model active–passive emulsions [29], nematic films (with active anchoring) [9], active nematic shells [30], and motile active droplets [31]. The role of coupling between the active nematic and its embedded topological defects with the actual shape of the surface was explored in Ref. [32]. Soft channel formation and symmetry breaking in exotic active emulsions was shown in Ref. [8]. LB was used to solve the *active model H*, which is a continuum theory of self-propelled particles, that notably also accounts for phase separation [33].

25.2 Lattice Boltzmann approaches

LB modeling of active matter draws from LB implementations and approaches developed and used in passive complex fluids [34, 35, 4]. Generally, modeling of complex fluids at the mesoscopic continuum level requires

solving multiple material fields (e.g., velocity, order parameters), which are each determined by their own distinct dynamic equations. From the modeling perspective, a major decision is which method or methods to use and solve these coupled sets of equations, usually in the form of partial differential equations. Past, yet well-established practices to complex fluids with LB methods have used LB approaches to solve *all* dynamic equations, whereas more recent approaches (usually called *hybrid-LB approaches*) use LB methods to solve some equation(s) (such as the Navier–Stokes equation) and other methods to solve other material degrees of freedom. A prime example of hybrid implementation includes a combination of an LB approach for material flow and a finite difference approach for order parameters.

25.2.1 Numerical implementation

We now present an overview of the classical three-dimensional implementation of the *LB-only approach* to complex fluids (specifically, passive nematic liquid crystals), as introduced in Ref [35], which was also extended and generalized for modeling active nematics [4]. Liquid crystal hydrodynamics is determined by two dynamic equations, the first for the velocity field \mathbf{v} (i.e., the Navier–Stokes equation) and the second for the order parameter tensor of the nematic Q_{ij} . To solve these two equations, two separate sets of the distribution functions [Can these distribution function be defined here? I think there is an explanation a few pages later, but it would be good to provide it at this stage when first introduced.] are introduced: the scalars f_i for \mathbf{v} and the symmetric traceless tensors \mathbf{G}_i for Q_{ij} . Each f_i and \mathbf{G}_i is associated with a lattice vector \mathbf{e}_i (for example see Fig. 25.2). Physical variables are defined as moments of the distribution functions:

$$\rho = \sum_i f_i, \quad \rho v_\alpha = \sum_i f_i e_{i\alpha}, \quad \mathbf{Q} = \sum_i \mathbf{G}_i. \quad (25.1)$$

The evolution of the distribution functions in a given timestep Δt follows a set of *streaming and collision steps*, which select the accuracy of the LB numerical scheme [36]. The implementation according to a *predictor–corrector scheme* and a *single relaxation time* can be written as (for full implementation, see [35])

$$f_i(\mathbf{x} + \mathbf{e}_i \Delta t, t + \Delta t) = f_i(\mathbf{x}, t) + \frac{1}{2} \Delta t (\mathcal{C}_{f_i}(\mathbf{x}, t, \{f_i\}) + \mathcal{C}_{f_i}(\mathbf{x} + \mathbf{e}_i \Delta t, t + \Delta t, \{f_i^*\})) \quad (25.2)$$

$$\mathbf{G}_i(\mathbf{x} + \mathbf{e}_i \Delta t, t + \Delta t), = \mathbf{G}_i(\mathbf{x}, t) + \frac{1}{2} \Delta t (\mathcal{C}_{\mathbf{G}_i}(\mathbf{x}, t, \{\mathbf{G}_i\}) + \mathcal{C}_{\mathbf{G}_i}(\mathbf{x} + \mathbf{e}_i \Delta t, t + \Delta t, \{\mathbf{G}_i^*\})), \quad (25.3)$$

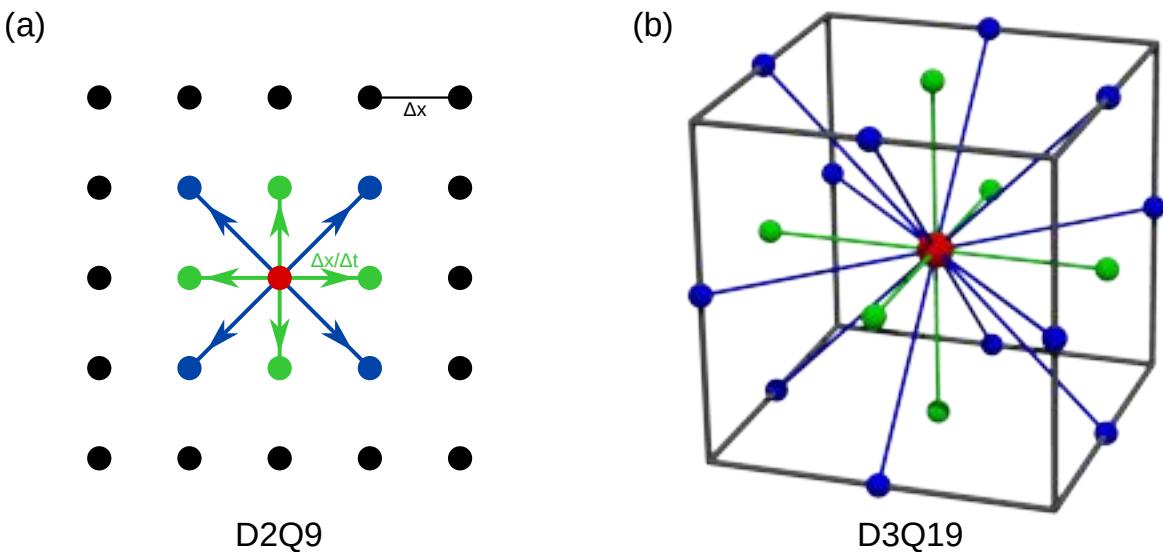


Figure 25.2: **Velocity sets in the lattice Boltzmann method.** (a) Two-dimensional set with 9 allowed velocities (one corresponds to particles at rest). (b) Three-dimensional velocity set with 19 allowed velocities. In each time step particles can stay at rest (in red), they can move to the nearest mesh point (in green) or second-nearest mesh point (in blue). Equally, the particles get streamed from every other mesh point.

where f_i^* and \mathbf{G}_i^* are first-order approximations for the distribution functions. The collision operators $\mathcal{C}_{f_i}(\mathbf{x}, t, \{f_i\})$ and $\mathcal{C}_{\mathbf{G}_i}(\mathbf{x}, t, \{\mathbf{G}_i\})$ are defined as

$$\begin{aligned}\mathcal{C}_{f_i}(\mathbf{x}, t, \{f_i\}) &= -\frac{1}{\tau_f}(f_i(\mathbf{x}, t) - f_i^{\text{eq}}(\mathbf{x}, t, \{f_i\})) + p_i(\mathbf{x}, t, \{f_i\}), \\ \mathcal{C}_{\mathbf{G}_i}(\mathbf{x}, t, \{\mathbf{G}_i\}) &= -\frac{1}{\tau_G}(\mathbf{G}_i(\mathbf{x}, t) - \mathbf{G}_i^{\text{eq}}(\mathbf{x}, t, \{\mathbf{G}_i\})) + \mathbf{M}_i(\mathbf{x}, t, \{\mathbf{G}_i\}),\end{aligned}\quad (25.4)$$

where f_i^{eq} and \mathbf{G}_i^{eq} are equilibrium distribution functions, and p_i and \mathbf{M}_i are the driving (forcing) terms for f_i and \mathbf{G}_i , respectively. The equilibrium distribution functions and the driving terms are formulated as polynomial expansions in the velocity field \vec{v} and then constrained to give the dynamics of the original dynamic equations of the complex fluid. The polynomial expansion can be formulated as

$$\begin{aligned}f_i^{\text{eq}} &= A_i + B_i v_\alpha e_{i\alpha} + C_i v^2 + D_i v_\alpha v_\beta e_{i\alpha} e_{i\beta} + E_{i\alpha\beta} e_{i\alpha} e_{i\beta}, \\ \mathbf{G}_i^{\text{eq}} &= \mathbf{J}_i + \mathbf{K}_i v_\alpha e_{i\alpha} + \mathbf{L}_i v^2 + \mathbf{N}_i v_\alpha v_\beta e_{i\alpha} e_{i\beta}, \\ p_i &= T_i \partial_\beta \tau_{\alpha\beta} e_{i\alpha}, \\ \mathbf{M}_i &= \mathbf{R}_i + \mathbf{S}_i v_\alpha e_{i\alpha},\end{aligned}\quad (25.5)$$

where $A_i, B_i, C_i, D_i, E_{i\alpha\beta}, \mathbf{J}_i, \mathbf{K}_i, \mathbf{L}_i, \mathbf{N}_i, T_i, \mathbf{R}_i$ and \mathbf{S}_i are expansion coefficients (some are matrices), summation over Greek indices is assumed, and τ is the asymmetric part of the fluid stress tensor. We note here that the presented brief overview above is a particular selected formulation of the LB approach (which was shown to be capable of solving multiple and diverse nematodynamic systems), but today also other LB formulations are used. Especially, different formulations of the implementations of the coupling between the fluid flow and the order parameter have been developed, which here it is done via equilibrium functions, but it can also be implemented via driving (forcing) terms.

25.2.2 LUDWIG code

A strong documented LB code for modeling complex fluids is LUDWIG ¹ that was developed largely at the University of Edinburgh, and is freely available under a BSD-style license [37]. The LUDWIG code has the underlying hydrodynamic model based on the lattice Boltzmann equation. It may be used to study simple (Newtonian) fluids, but it is more generally suited to complex fluids, including symmetric binary fluids and Brazovskii smectics, polar gels, liquid crystals, or charged fluids via a *Poisson–Boltzmann equation approach*. These material features are added via a free energy approach, where specific order parameters are evolved according to the appropriate coarse-grained dynamics, but are fully coupled and interact with the fluid.

25.3 Hybrid lattice Boltzmann approaches

In this section, we provide a selected complete formulation of a *hybrid lattice Boltzmann numerical method*, based on an example of active nematic fluids. The hybrid approach consists of a *finite difference method* for the *Q-tensor evolution equation* (Eq. (25.16)) and a *lattice Boltzmann method* for the *Navier–Stokes equation* and the *compressibility equations* (Eq. (25.13) and Eq. (25.14)). It is established as a main simulation method for continuous active nematic models [4]. A finite difference method for the Q-tensor evolution is chosen as it provides stable solutions, can be implemented in a straightforward fashion, and can be also expanded to include the dynamics of other scalar or polar order parameters. In this section, we discuss the details of the lattice Boltzmann time step, the meshing and the finite difference time step, the separation of scales, a choice of the numerical parameters for passive and active nematics, and typical computational requirements for the hybrid numerical algorithm.

25.3.1 Lattice Boltzmann time step

The central quantity in the lattice Boltzmann method is the distribution function, which represents the density of virtual particles with certain velocity at a specific point in space and time. The main idea behind the lattice

¹<https://ludwig.epcc.ed.ac.uk/>

Boltzmann approach is that the distribution function is discretized to be defined only in the mesh points (with resolution Δx) and in consecutive time steps (Δt apart). Additionally, only specific directions and sizes of the particle velocities are allowed (see Fig. 25.2). The allowed velocities stream the particles in one time step from one mesh point to another, which extensively simplifies the numerical procedure. Such numerical algorithm, inspired by Boltzmann's picture of gasses, can in fact be used to simulate the Navier–Stokes equation [38].

The first step in the implementation of the lattice Boltzmann method is the selection of a velocity set along which the virtual particles can propagate. In Figure 25.2, a 9-velocity set in two dimensions (D2Q9) and a 19-velocity set in three dimensions (D3Q19) are shown. For the D3Q19 set, the allowed velocities of the virtual particles are

$$\begin{aligned}\mathbf{c}_0 &= (0, 0, 0) \frac{\Delta x}{\Delta t}, \\ \mathbf{c}_{1-6} &= (\pm 1, 0, 0) \frac{\Delta x}{\Delta t}, \quad (0, \pm 1, 0) \frac{\Delta x}{\Delta t}, \quad (0, 0, \pm 1) \frac{\Delta x}{\Delta t}, \\ \mathbf{c}_{7-18} &= (0, \pm 1, \pm 1) \frac{\Delta x}{\Delta t}, \quad (\pm 1, 0, \pm 1) \frac{\Delta x}{\Delta t}, \quad (\pm 1, \pm 1, 0) \frac{\Delta x}{\Delta t}.\end{aligned}\quad (25.6)$$

Other popular choices for the velocity sets in 3D are the D3Q15 and D3Q25 sets.

The density of particles with each of the allowed velocities is measured by the *velocity distribution function* f_α . The Greek-letter indices (α, β, \dots) are used for quantities associated with the velocity lattice, while the Latin-letter indices (i, j, \dots) are used for Cartesian coordinates. [Are the Latin indices used in the text?] The density ρ and fluid velocity \mathbf{v} at each mesh point can be determined from the moments of the velocity distribution function:

$$\rho = \sum_{\alpha} f_{\alpha}, \quad (25.7)$$

$$\mathbf{v} = \frac{1}{\rho} \sum_{\alpha} f_{\alpha} \mathbf{c}_{\alpha} + \frac{\mathbf{F} \Delta t}{2\rho}, \quad (25.8)$$

where the term proportional to the force \mathbf{F} in Eq. (25.8) has been added to ensure the second-order accuracy of the force implementation [39, 38].

Each time step of the lattice Boltzmann method can be divided into two: a *collision step* (i.e., a recombination of the velocity distribution function and the application of the force) and a *streaming step* (an advection of the distribution function to the adjacent node). The collision step can be formulated by creating a new distribution function f_{α}^* :

$$f_{\alpha}^*(\mathbf{r}, t) = f_{\alpha}(\mathbf{r}, t) - \frac{\Delta t}{\tau_f} [f_{\alpha}(\mathbf{r}, t) - f_{\alpha}^{\text{eq}}] + \Delta t P_{\alpha}, \quad (25.9)$$

where $\frac{\Delta t}{\tau_f} [f_{\alpha} - f_{\alpha}^{\text{eq}}]$ is known as the Bhatnagar–Gross–Krook collision operator with τ_f the fluid relaxation time (associated with the viscosity in Eq. (25.15)), f_{α}^{eq} is the equilibrium distribution function, and P_{α} the forcing term (defined below). After the collision step, the velocity distribution functions f_{α}^* are streamed along the corresponding velocity \mathbf{c}_{α} :

$$f_{\alpha}(\mathbf{r} + \mathbf{c}_{\alpha} \Delta t, t + \Delta t) = f_{\alpha}^*(\mathbf{r}, t). \quad (25.10)$$

Collision and streaming steps can be combined into what is known as the lattice Boltzmann equation.

In the absence of forcing terms ($P_{\alpha} = 0$), f_{α}^{eq} describes the local equilibrium distribution functions at given density and velocity. The expression of f_{α}^{eq} is expanded to second-order terms in fluid velocity \mathbf{v} and can be written in the form [38]

$$f_{\alpha}^{\text{eq}} = \rho w_{\alpha} \left(1 + \frac{\mathbf{v} \cdot \mathbf{c}_{\alpha}}{c_s^2} + \frac{(\mathbf{v} \cdot \mathbf{c}_{\alpha})^2}{2c_s^4} - \frac{\mathbf{v} \cdot \mathbf{v}}{2c_s^2} \right), \quad (25.11)$$

where for the D3Q19 velocity set, $c_s = \frac{1}{\sqrt{3}} \frac{\Delta x}{\Delta t}$ is the speed of sound and w_{α} are the corresponding weights [38]:

$$\begin{aligned}w_0 &= 1/3, \\ w_{1-6} &= 1/18, \\ w_{7-18} &= 1/36.\end{aligned}$$

Similarly, the forcing term is constructed from the local force density F_i :

$$P_\alpha = \left(1 - \frac{\Delta t}{2\tau_f}\right) w_\alpha \left(\frac{c_{\alpha i} c_{\alpha j} - c_s^2 \delta_{ij}}{c_s^4} v_j \right) F_i, \quad (25.12)$$

where F_i includes the external force density (i.e., gravity or a homogeneous flow driving force) and the divergence of the stress tensor generated by the active fields. The term of the isotropic viscous stress tensor $\eta(\partial_i v_j + \partial_j v_i)$ is subtracted from the nematic stress tensor as it is already included within the lattice Boltzmann algorithm (η is the isotropic viscosity within the algorithm). Typically, the divergence of the stress tensor can be partially expressed analytically as a combination of fields already known in the finite difference step of the hybrid approach [40, 41], so that the divergence itself does not need to be fully calculated numerically.

In a macroscale continuum limit, the lattice Boltzmann equation can be mapped to the Navier–Stokes equation. The mapping is typically demonstrated in a Chapman–Enskog analysis, in which $f_\alpha - f_\alpha^{\text{eq}}$, time and space derivatives, and P_α are expanded up to second-order terms in the Knudsen number. The *Knudsen number* is a dimensionless quantity that compares an effective mean free path of (virtual) particles and a length scale at which continuum variables show considerable variations. For small Knudsen numbers, a Navier–Stokes fluid dynamics is recovered [38]:

$$\partial_t(\rho v_i) + \partial_j(\rho v_i v_j + \rho c_s^2 \delta_{ij}) = \partial_j[\eta(\partial_j v_i + \partial_i v_j)] + F_i, \quad (25.13)$$

$$\partial_t \rho + \partial_i(\rho v_i) = 0, \quad (25.14)$$

$$\eta = c_s^2 \Delta t \left(\frac{\tau_f}{\Delta t} - \frac{1}{2} \right), \quad (25.15)$$

where Eq. (25.14) is the continuity equation and Eq. (25.15) is the isotropic viscosity within the lattice–Boltzmann algorithm. The dilatational viscosity in the compressible formulation of the Navier–Stokes equation (Eq. (25.13)) equals $2\eta/3$ and the ρc_s^2 term plays the role of pressure. The described numerical model is a valid method to solve the Navier–Stokes equation only for weakly compressible flows (for which $(v/c_s) \ll 1$) for which the third-order terms proportional to $(v/c_s)^3$ in the derivation of Eq. (25.13) can be neglected. Furthermore, the hybrid lattice Boltzmann approach is typically used to simulate incompressible models of active matter. It is therefore necessary to keep the compressibility that arises in the numerical method small. These “compressibility errors” typically scale as $\mathcal{O}(v^2/c_s^2)$ [38], which is another reason that the velocity in the simulations has to be well below the speed of sound.

The lattice Boltzmann numerical algorithm has to use appropriate boundary conditions, describing the fluid interaction at the solid walls and open or periodic conditions at the end of the simulation box. Numerous methods are available to create slip, partial slip, or no-slip boundary conditions at the walls [38]. A very commonly used boundary condition is a no-slip wall implemented by a bounce-back method. The algorithm of the bounce-back method is that the distribution functions propagated in the solid wall nodes (see Fig. 25.3) are reflected back in the opposite direction. This creates a no-slip boundary at the middle between the solid node and the fluid node. This method can also be easily extended to allow for tuning the boundary position between the solid and the fluid node, thus also avoiding a zig-zag boundary shape, and also to implement a moving boundary [38, 42]. As with many lattice-based numerical methods, periodic boundary conditions are easily achieved by creating a layer of virtual nodes at each side of the periodic boundary (Fig. 25.3). Additionally, periodic boundary conditions can be adapted to include a pressure drop between the boundaries [38, 43, 44]. The pressure drop is realized by a small density difference between both sides of a channel. The generated pressure gradient then drives the fluid flow. This approach is particularly useful if channels are not straight, in which case generating flow by a constant force is not optimal. Often, ends of microchannels cannot be closed into periodic boundaries, and one needs to implement open boundary conditions by specifying either fluid velocity or fluid density at the boundary [38, 45].

The lattice Boltzmann algorithm in the hybrid approach has been written in various different forms to simulate active matter in the past. A common choice is to use a predictor–corrector scheme in the streaming step of the lattice Boltzmann method [46]. Also, an alternative approach to include the stress tensor in the lattice Boltzmann algorithm through its divergence is to incorporate the stress tensor into the expression for the equilibrium distribution functions (Eq. (25.11)) [47]. This approach works only for the symmetric part of the stress tensor, while a possible antisymmetric part (as for the case of nematic stress tensor) still needs to

be included as a body force. The formulation that we provide in this section does not depend on the specifics of the active matter order parameters and can be found also in lattice Boltzmann textbooks [38]. The detailed dynamics of the active order parameter fields is implemented only in the finite difference time step of the hybrid approach.

25.3.2 Finite difference time step

Within the implementation of the hybrid lattice Boltzmann approach, simulations are performed on a rectangular mesh, where each mesh point has a dedicated logical marker, depending on the position of the point: (i) in the bulk, (ii) at the surface, (iii) within the confining walls or colloidal particles, or (iv) at the mesh boundary, as shown by Fig. 25.3.

Additionally, a normal vector with the direction that is not bound by the mesh grid is prescribed to each surface node. A logical marker at each node point determines which step in the lattice Boltzmann and in the finite difference algorithm will be implemented at each time step of the computation.

In the bulk nodes, at each time step, consecutive steps in the lattice Boltzmann domain and in the finite difference domain are performed. For lattice Boltzmann, the equilibrium distribution functions (Eq. (25.11)) and the forcing term (Eq. (25.12)) are calculated and the streaming and collision steps are executed (Eqs. (25.9) and (25.10)). This includes the computation of the local density (Eq. (25.7)), velocity (Eq. (25.8)), and the divergence of the active nematic stress tensor calculated via central differences. In the finite difference time domain, evolution of the Q-tensor is discretized in explicit form into time steps Δt :

$$Q_{ij}(t + \Delta t) = Q_{ij}(t) + \Delta t [\Gamma H_{ij} + S_{ij} - (\mathbf{v} \cdot \nabla) Q_{ij}], \quad (25.16)$$

where the molecular field H_{ij} , the generalized advection field S_{ij} , and the Q-tensor gradients have to be calculated through central differences, and Γ is the rotational viscosity parameter in the model.

In the case where the Q-tensor at the surface is not fixed, surface nodes are evolved as

$$Q_{ij}^{\text{surf}}(t + \Delta t) = Q_{ij}^{\text{surf}}(t) - \Delta t \Gamma^{\text{surf}} \left[\frac{1}{2} (H_{ij}^{\text{surf}} + H_{ji}^{\text{surf}}) - \frac{1}{3} \delta_{ij} H_{kk}^{\text{surf}} \right], \quad (25.17)$$

where

$$H_{ij}^{\text{surf}} = -\frac{\partial f_{\text{vol}}}{\partial (\partial_k Q_{ij})} v_k - \frac{\partial f_{\text{surf}}}{\partial Q_{ij}} \quad (25.18)$$

is the surface molecular field, f_{vol} is the nematic bulk free energy density, f_{surface} is the surface free energy density, and Γ^{surf} is the surface rotational viscosity parameter. Effects of the velocity gradient are typically neglected in the surface step of the Q-tensor dynamics. Forward (or backward) difference is used to compute the surface molecular field. Streaming functions in the surface and wall nodes that come from the lattice Boltzmann implementation are reversed and bounced back into the bulk to implement the no-slip boundary. Bounce-back and surface time step of the Q-tensor are implemented to create a virtual surface in the midpoint between surface and bulk nodes. In this way, no simulation volume is associated with surface nodes, which also greatly simplifies the calculation of the bulk and surface free energy.

25.3.3 Separation of scales

Models of active hydrodynamics typically include an abundance of time and length scales, which must be carefully resolved in simulations. From the Navier–Stokes equation, a short diffusive time scale $\tau_{\text{diff}} = \frac{\rho l^2}{\eta}$ and an advective time scale $\tau_{\text{adv}} = \frac{l}{v}$ can be extracted, where l is a typical length scale, v a typical velocity, and η/ρ the kinematic viscosity. The ratio between the diffusive and the advective time scale gives the Reynolds number. Additional examples include time scales associated with the variations in the degree of order of orientational fields (τ_S), applications of external fields (τ_{field}), effective elasticity of orientational structure (τ_{elastic}), and activity (τ_{active}). All of these mechanisms typically include a characteristic time scale as well as a characteristic length scale.

The characteristic time and length scales of active hydrodynamics span over orders of magnitude, which poses a great numerical difficulty. For example, in a cell 100 mesh points thick, the diffusive time scale τ_{diff} is approximately 10^4 time steps (in accordance with data in Table 25.1). For Reynolds number of 10^{-4} , the

corresponding advective time scale τ_{adv} would take $\sim 10^8$ time steps. Given that a single simulation must span over numerous advective time scales, the whole simulation would take more than 10^9 time steps, which is typically above the available computational power. A similar problem is encountered when considering the size of the simulation box. The resolution is limited by the smallest length scale, which is typically the scale of the degree of order scale [Rephrased as not clear. Check] that dictates the thickness of defect cores. In small molecule nematic materials, the length scale of the degree of order [Rephrased as not clear. Check] is on the order of nm (see Table 25.1). In experiments, nematic cells easily reach thicknesses of $10 \mu\text{m}$ or more, which would require numerical meshes with > 1000 points in each direction. Similarly in microtubule-based active materials, the scale of the degree of the order [Rephrased as not clear. Check] is on the order of μm , while the cell sizes reach up to several mm [51]. Both in terms of time steps and mesh size, the exact experimental scale separations cannot be met numerically. The problem of large scale separation is by no means unique to active nematic hydrodynamics, it occurs also in other multiscale systems such as colloidal suspensions [52]. An approach to produce qualitatively and (up to some error) also quantitatively relevant simulations is to reduce the separation of scales while preserving the hierarchy between them [52].

Time scales in a typical simulation can be divided into short ones (diffusive scale, degree-of-order scale), medium ones (advective scale, active scale, external field time scale — depending on the strength of activity and external fields), and slow ones (orientational structure relaxation scale), as shown in Table 25.2. Dimensionless numbers that characterize active flows can be constructed as ratios between pairs of time scales and also reflect the separation of scales: the *Reynolds number* is typically in the range $0.01 < \text{Re} < 1$ and the *Ericksen number* $\text{Er} = \tau_{\text{elastic}}/\tau_{\text{adv}}$ is typically in the range of $1 < \text{Er} < 100$. The fact that the Reynolds number is much larger than in experiments (where $\text{Re} < 10^{-4}$) causes effectively larger density in numerical simulations, which in

Parameter	Value (in L , ξ_N , and Γ)	Value in SI units for nematodynamics	Value in SI units for active systems
A	$-0.190 \frac{L}{\xi_N^2}$	$-1.72 \cdot 10^5 \text{ J/m}^3$	$-3.56 \cdot 10^{-2} \text{ J/m}^3$
B	$-2.34 \frac{L}{\xi_N^2}$	$-2.12 \cdot 10^6 \text{ J/m}^3$	-0.439 J/m^3
C	$1.91 \frac{L}{\xi_N^2}$	$1.73 \cdot 10^6 \text{ J/m}^3$	0.358 J/m^3
L_2, L_3	0	0	0
$\epsilon_0 \epsilon_a^{\text{mol}} E_0^2$	$\lesssim 0.05 \frac{L}{\xi_N^2}$	$\lesssim 4.5 \cdot 10^4 \text{ J/m}^3$	$\lesssim 9.4 \cdot 10^{-3} \text{ J/m}^3$
W_{uni}	$0.48\text{--}4.8 \frac{L}{\xi_N}$	$10^{-3}\text{--}10^{-2} \text{ J/m}^2$	$3.6 \cdot 10^{-7}\text{--}3.6 \cdot 10^{-6} \text{ J/m}^2$
Γ^{surf}	$0.67 \frac{\Gamma}{\xi_N}$	$6.5 \cdot 10^9 \text{ 1/Pa s m}$	$1.7 \cdot 10^6 \text{ 1/Pa s m}$
ζ		0.95 [48]	0.7–0.9 [16, 49]
η	$\frac{1}{\Gamma}$	0.067 Pas	0.1 Pas
ρ (for $\Delta t = 0.02 \frac{\xi_N^2}{L\Gamma}$)	$1.78 \cdot 10^{-2} \frac{1}{L\Gamma^2}$		
α	$\lesssim 0.1 \frac{L}{\xi_N^2}$		0.019 N/m^2
Δx	$1.5 \xi_N$	3.45 nm	$6 \mu\text{m}$
Δt	$0.02\text{--}0.2 \frac{\xi_N^2}{L\Gamma}$	1.47–14.7 ns	10.7–107 ms
τ_f	$2\Delta t$	2.94–29.4 ns	21.4–214 ms

Table 25.1: **Typical parameters used in simulations.** Parameter values are expressed in dimensions of L , ξ_N , and Γ (second column), in SI units for nematodynamics of thermotropic nematics (third column) at $L = 4.8 \text{ pN}$, $\xi_N = 2.3 \text{ nm}$ and $\Gamma = 15 \text{ (Pas)}^{-1}$, and in SI units for active nematics (fourth column) at $L = 3 \text{ pN}$, $\xi_N = 4 \mu\text{m}$ and $\Gamma = 10 \text{ (Pas)}^{-1}$. L_2 and L_3 are additional nematic elastic constants, $\epsilon_0 \epsilon_a^{\text{mol}} E_0^2$ is the coupling term with the electric field, W_{uni} the surface anchoring strength, ζ the alignment parameter [50], and α the activity parameter. Some values are omitted as they are not relevant to the given system. The parameters in the bottom part are associated with the numerical method.

fast scales	$\tau_{\text{diff}}, \tau_s$
medium scales	$\tau_{\text{adv}}, \tau_{\text{active}}, \tau_{\text{field}}$
slow scales	τ_{elastic}

Table 25.2: **Hierarchy of time scales in numerical simulations of active nematics.** [Can the symbols be defined in the caption using the same terminology as in the text?]

turn means that the nematic flow profile takes longer to adapt to the force distribution; however, the constraint on the Reynolds number in simulations still allows for the velocity profile to fully develop during the director reorientation process.

Besides the intrinsic length scales shown in Table 25.2, there are also numerical scales (resolution and length of simulation) and emergent length scales (e.g., defect–defect separation in bulk active turbulence). For numerical stability, it is important that the resolution is lower than (or on the order of) the smallest scales, and that the size of the simulation box and the length of the simulation are much greater than the emergent length and time scales. Only so can the system reach the dynamic equilibrium states and show the predicted bulk behavior [53]. By preserving the basic hierarchy of scales, numerical simulations are able to qualitatively predict nematic structure, response, and scaling properties.

25.3.4 Numerical implementation

Numerical values for the model parameters of (active) nematodynamics are presented in Table 25.1. The parameters are expressed in terms of correlation length ξ_N , elastic constant L , and rotational viscosity parameter Γ . An alternative approach is also to give values of numerical parameters directly in simulation units of mesh resolution Δx , time resolution Δt , and elastic constant L . The results of the numerical simulations can be expressed in SI units (third and fourth columns of Table 25.1) by making a choice of ξ_N , L , and Γ that corresponds to a desired nematic material. As the values of most of the model parameters for a range of (active) nematic materials are not well known, we typically use a universal choice of parameters in the second column of Table 25.1 that is based on well-measured thermotropic liquid crystals, and only then make a choice of ξ_N , L , and Γ based on a specific system.

Passive nematic liquid crystals are a rather well-explored and -measured phase, with also rather well-known mesoscopic material parameters. The major difference to a typical active system is the size of the building blocks (nematic agents), which in thermotropic nematics is usually in the order of nanometers, whereas in active nematics it is usually in the order of microns, i.e., different by several orders of magnitude. A well-documented passive nematic material is 4-Cyano-4'-pentylbiphenyl (5CB) and its parameters are quite commonly used in various studies, including with LB modeling, and are given in the second and third columns of Table 25.1. Note that the parameters are given in units of the nematic correlation length (which is comparable to the 5CB molecule size $\xi_N = 2.3 \text{ nm}$), the single elastic constant $L = 4.8 \text{ pN}$ (computed from the average Frank elastic constant $\bar{K} = 6.1 \text{ pN}$ [54] through the relation $L = \frac{2}{9S^2} \bar{K}$), and the rotational viscosity parameter $\Gamma = 15 \text{ (Pas)}^{-1}$ (computed from $\gamma_l = 0.085 \text{ Pas}$ [48] through the relation $\Gamma = \frac{9S^2}{2\gamma_l}$). The phase parameters A (at 10 K below the supercooling temperature), B , and C [Have these parameters been defined earlier with the same language? check for consistency] are chosen such that they roughly reproduce the experimentally measured temperature dependence of the equilibrium order parameter S_{eq} of 5CB and give, in combination with $L_1 = 4.8 \text{ pN}$, the nematic correlation length of $\xi_N = 2.3 \text{ nm}$. Exact values of the phase parameters depend on the method used to extract them from experimental observations [55, 56]. For simulations at constant temperature, nematic phase behavior can also be adequately described by only two parameters [57]. The lattice Boltzmann viscosity η , density ρ and the fluid relaxation time τ_f are interconnected by Eq. (25.15). η is found to condition the stability of the LB scheme and can be often increased in simulations [41].

Model parameters for active materials are very rarely precisely known. As a rough estimation, we set the physical parameter values for active materials in the fourth column of Table 25.1 based on the estimated values of $L = 3 \text{ pN}$, $\xi_N = 4 \mu\text{m}$, and $\Gamma = 10 \text{ (Pa s)}^{-1}$, while keeping the values in the second column fixed. These values are obtained from bacterial sizes in active turbulence [58] and approximate the size of defect cores in microtubular active nematics [51] (for ξ_N), the elastic constant for bacterial systems [58], and the timescale of active nematic dynamics [51, 30] (to estimate Γ). Similar values of L , ξ_N and Γ were used also in simulations

of passive lyotropic liquid crystals [50].

The bottom part of Table 25.1 gives parameters that are linked to the hybrid lattice Boltzmann numerical implementation of the theoretical model. Typically, the mesh resolution Δx should be below $1.5\xi_N$ which allows for the defects to move around the grid and not get pinned to a mesh point [59]. The upper limit of the time step Δt is typically used for nematodynamic simulations without flow, while the lower limit is used in simulations of nematic flow with fine temporal precision and low Reynolds numbers. The time step can be in principle increased to around $0.2\xi_N^2/L\Gamma$ even for simulations with flow included, but for larger time steps the method typically becomes unstable.

25.3.5 Computational requirements

The hybrid lattice Boltzmann numerical algorithm is based on consecutive time steps in the finite difference and lattice Boltzmann domains. Between the time steps, local fields such as the stress tensor and its divergence, the molecular field, the velocity field and its gradients, and the Q-tensor gradients have to be calculated, which affects the computational speed. Typically, numerical simulations have to be parallelized on multiple CPUs or GPUs to improve the running times. Carenza *et al.* [60] report on MPI-based hybrid lattice Boltzmann simulations of active chiral nematics, parallelized to up to 256 processors. They ran simulations on $(512)^2$ mesh points in two dimensions and $(128)^3$ mesh points in three dimensions, and observed a particularly strong scaling of numerical speed-up with the number of processors in the case of three-dimensional simulations. For 256 processors, they reported a 78% speed-up of the ideal scaling for three dimensions and $\approx 40\%$ for two-dimensional simulations. We reported on the OpenMP-based numerical simulations run on a computational node with two 12-core CPUs clocked at 2.6GHz and multithreaded to 48 cores [41]. A typical three-dimensional simulation involved $(200)^3$ mesh points and took up approximately 3.5GB of memory. About 60,000 time steps were computed per day, with several hundred thousand steps typically needed to complete the simulation. With increased availability of computational nodes with a high-core count and high-clock rates, the simulation sizes can be increased further. Recently, simulations of active nematics in three dimensions were run for $(400)^3$ mesh points [22].

25.3.6 Visualization of fields

Numerical simulations of active matter typically involve several fields that can be scalar (e.g., density, concentration, degree of order), vectorial (e.g., velocity field, director, polarization), or tensorial (e.g., nematic tensor order parameter, strain tensor, stress tensor) in nature. To gain insight and present the numerical simulations, the resulting fields must be carefully visualized and possible topological information must be extracted from them. We show field visualization and topological analysis for the example of three-dimensional spherical active nematic droplets [19].

Figure 25.4 shows various visualizations of velocity and tensor order parameter fields in active nematic droplets. The velocity field is plotted both using arrows (Fig. 25.4a) and rubbing-like streamlines (Fig. 25.4b) to emphasize local self-advection of defect line segments and circular flow in the droplet. The degree of order is the main eigenvalue of the nematic tensorial order parameter and is shown as a red isosurface (Fig. 25.4a,b). Inside the red regions are the cores of the defect loops, where the degree of order falls towards zero. Another possibility to visualize defect cores would be for example to plot isosurfaces of $(\partial_k Q_{ij})^2$, the large values of which correspond to regions of high director field deformation (i.e., defect cores). The director field itself is plotted as colored rods in (Fig. 25.4c), where the color corresponds to the in-plane director orientation. The minima of the degree of order allow to locate the defect segments and also track them in time. However, the local degree of order does not reveal the local director profile of defect line segments, which can change from local twist profiles to splay-bend profiles (i.e., $\pm 1/2$ defects). One way of visualizing local director profiles of defect line segments is to plot isosurfaces of splay-bend

$$S_{SB} = \frac{\partial^2 Q_{ij}}{\partial x_i \partial x_j} \quad (25.19)$$

and twist parameter [61]

$$S_T = \epsilon_{ikl} Q_{ij} \frac{\partial Q_{lj}}{\partial x_k}, \quad (25.20)$$

which reveal the regions of high splay-bend and twist deformations, respectively (Fig. 25.4c,d). Formally, the director field profiles in close proximity do defect line segments can be classified using a phase offset and a twist angle [20]. To obtain the topological charge of individual defect loops (Fig. 25.4d), one needs to follow the continuous quaternionic transformation of local defect line director profiles around the loop [62]. In active nematics, charged and zero-charged loops are observed [19, 16]. Visualization of the parameter fields can also be performed with specialized software, in particular for three-dimensional fields, where a possible choice for an interactive visualization application is Paraview [63]. Alternatively, we commonly also use a custom-build software that provides an input for raytracing programs, where, for example, Figure 25.4 is plotted using custom code developed by Simon Čopar to generate POV-Ray [64] input data from numerical results.

25.4 Discussion, overview, and challenges

This Chapter gives a selected overview of the use of the lattice Boltzmann method in modeling active matter. We provide a brief overview of the current state of the art and also indicate analogies and links with modeling passive complex fluids, especially passive nematic liquid crystals. In section 25.3, we present a detailed formulation of a hybrid lattice Boltzmann method, which is the used method in numerous numerical papers on active nematic materials. The approach is based on two self-standing numerical toolboxes — namely, lattice Boltzmann for driven fluids and finite difference for the time evolution of the order parameter and concentration fields — that can also be used in their own right. The lattice Boltzmann and the finite difference parts are coupled together by forcing terms in the lattice Boltzmann and velocity-coupling in the order parameter dynamics. A strong advantage of the hybrid lattice Boltzmann approach is that the order parameter time evolution can be supplemented or expanded by the evolution equations of other fields, such as the polar vector order parameter or various concentration fields which can also be solved in the finite difference domain. The hybrid lattice Boltzmann method is based on a rectangular mesh that can distort the shape of boundaries, which is a disadvantage particularly in highly curved or thin geometries (e.g., shells) compared to, for example, finite element methods. A strong advantage of the hybrid lattice Boltzmann method is that its defining set of equations is very compact and can be often implemented into a numerical code without (m)any additional numerical libraries, giving flexibility for the implementation as well as a good overview and control over all the steps of the computation and parallelization.

Numerical models of active matter typically include several time and length scales, which must be carefully considered, as we discuss in section 25.3.3. Computational availability often prevents from using numerical scales that are too many orders of magnitude apart. Therefore, the separation of scales must typically be reduced, while preserving the hierarchy between them. An example of this is the Reynolds number that is in simulations of active nematics often orders of magnitude higher than in experiments, but is still below 1. Another example of bridging the separation of scales is the size of the numerical mesh, which is typically a few hundred times larger than the nematic correlation length. However, in experiments, the ratio might be more than ten times higher. It is important to consider how the rescaling of the system — which is today quite commonly and widely used in the literature — affects the observed behavior. In the case of the nematic correlation length, defects typically appear thicker than in experiments and might therefore take a different shape. In the case of the active length scale, it was shown that taking a too small simulation box (compared to the active scale) can alter the scaling of the emergent length scales (such as defect-defect separation) with respect to activity [53].

Active matter and nematodynamics in general are relevant in a range of experimental systems, from bacteria and microtubules to small molecule liquid crystals, for which the model parameters are not precisely known. We describe an approach in which a set of parameters is used across many nematic systems, and the right scales are provided by the choice of the elastic constant L , correlation length ξ_N and inverse rotational viscosity Γ . However, some systems such as lyotropic nematics [50] are distinctly characterized by highly anisotropic elasticity and viscosity. Most models of (active) nematodynamics consider using up to two elastic constants, which allows to decouple twist director modes from the splay and bend modes. However, more elastic terms that are of third order in Q are typically not considered in the theoretical models. A hybrid lattice Boltzmann numerical approach implements an isotropic viscosity tensor directly through the lattice Boltzmann step, while additional anisotropic corrections are implemented through the divergence of the nematic stress tensor. While this allows for the anisotropic viscosity coefficients to get close to the isotropic viscosity, the method becomes unstable for highly anisotropic viscosity tensors as it is expected in some lyotropic nematics. All these effects can be seen as future open challenges for both science and the implementation of the LB method.

25.5 Problems

The set of problems below is based on numerical simulations of two-dimensional active nematics. To solve the problems, you can: (i) use the hybrid lattice Boltzmann code for active nematics that we made available in [65], (ii) download already developed modeling packages such as LUDWIG [37], or (iii) even write the Lattice Boltzmann simulation yourself.

Problem 25.1: Active nematic. Simulate continuum equations of active nematic hydrodynamics.

- a. Set up your numerical code, start the simulation and let the system evolve until it reaches a dynamical steady state. Plot the velocity, the director, and the scalar degree of order fields. Try to explain and describe what your simulations show.
- b. Define and compute the dimensionless numbers, such as the Ericksen number and the Reynolds number for a typical length scale in the steady state regime.

Problem 25.2: Active nematic: role of the initial conditions. Explore the role of initial conditions.

- a. Change the director's initial condition to a homogeneous field with some small perturbation in the middle. Follow how the director's and the velocity's structure evolves at different activities.
- b. Change the Q-tensor initial condition to a $\pm 1/2$ defect pair at some distance. Plot the annihilation dynamics of a defect pair at a few different activities. [*Hint: a planar director field with defects can be initialized from a linear superposition of the director angles θ_1 and θ_2 , where θ_1 corresponds to the director profile of a defect at location \mathbf{r}_1 and θ_2 to a defect at \mathbf{r}_2 .*]

Problem 25.3: Steady state in dynamic active nematic. Generate and explore the dynamic active nematic steady state also known as active turbulence for different activities.

- a. How does the number of defects scale with activity? [*Hint: defect positions are easily recognizable as minima in the scalar degree of order field. You can estimate the number of defects from the number of mesh points with the degree of order below some threshold value. You might want to increase the time step and the simulation box to get better statistics.*]
- b. Plot the power spectrum of the kinetic energy at some high activity. Do you observe a characteristic scaling at large wavelengths? You might have to use a larger simulation box to get better results.

Bibliography

- [1] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85:1143, 2013.
- [2] Nature Research collection Active Matter. https://www.nature.com/collections/hvczfmjfzl?utm_source=twitter&utm_medium=social&utm_content=boosted&utm_campaign=NCOM_1_SZ_ActiveMatter-GRC-social. Jan 2019.
- [3] A. Doostmohammadi, J. Ignes-Mullol, J. M. Yeomans, and F. Sagues. Active nematics. *Nature Communications*, 9:045006, 2018.
- [4] D. Marenduzzo, E. Orlandini, M. E. Cates, and J. M. Yeomans. Steady-state hydrodynamic instabilities of active liquid crystals: Hybrid lattice Boltzmann simulations. *Physical Review E*, 76:031921, 2007.
- [5] R. Voituriez, J. F. Joanny, and J. Prost. Spontaneous flow transition in active polar gels. *Europhysics Letters*, 70:404, 2005.
- [6] E. Tjhung, M. E. Cates, and D. Marenduzzo. Nonequilibrium steady states in polar active fluids. *Soft Matter*, 7:7453, 2011.
- [7] C. A. Whitfield, D. Marenduzzo, R. Voituriez, and R. J. Hawkins. Active polar fluid flow in finite droplets. *The European Physical Journal E*, 37:323, 2014.
- [8] L. N. Carenza, G. Gonnella, A. Lamura, D. Marenduzzo, G. Negro, and A. Tiribocchi. Soft channel formation and symmetry breaking in exotic active emulsions. *Scientific Reports*, 10:078101, 2020.

- [9] M. L. Blow, M. Aqil, B. Liebchen, and D. Marenduzzo. Motility of active nematic films driven by "active anchoring". *Soft Matter*, 13:6137, 2017.
- [10] J. de Graaf, H. Menke, A. J. T. M. Mathijssen, M. Fabritius, C. Holm, and T. N. Shendruk. Lattice-boltzmann hydrodynamics of anisotropic active matter. *Journal of Chemical Physics*, 144:134106, 2016.
- [11] H. H. Wensink, J. Dunkel, S. Heidenreich, K. Drescher, R. E. Goldstein, H. Lowen, and J. M. Yeomans. Meso-scale turbulence in living fluids. *The Proceedings of the National Academy of Sciences (PNAS)*, 109:14308, 2012.
- [12] C. Peng, T. Turiv, Y. Guo, Q.-H. Wei, and O. D. Lavrentovich. Command of active matter by topological defects and patterns. *Science*, 354:882, 2016.
- [13] D. Dell'Arciprete, M. L. Blow, A. T. Brown, F. D. C. Farrell, J. S. Lintuvuori, A. F. McVey, D. Marenduzzo, and W. C. K. Poon. A growing bacterial colony in two dimensions as an active nematic. *Nature Communications*, 9:4828, 2018.
- [14] A. Sokolov, A. Mozaffari, R. Zhang, J. J. de Pablo, and A. Snezhko. Emergence of radial tree of bend stripes in active nematics. *Physical Review X*, 9:031014, 2019.
- [15] T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic. Spontaneous motion in hierarchically assembled active matter. *Nature*, 491:431, 2012.
- [16] G. Duclos, R. Adkins, D. Banerjee, M. S. E. Peterson, M. Varghese, I. Kolvin, A. Baskaran, R. A. Pelcovits, T. R. Powers, A. Baskaran, F. Toschi, M. F. Hagan, S. J. Streichen, V. Vitelli, D. A. Beller, and Z. Dogic. Topological structure and dynamics of three-dimensional active nematics. *Science*, 367:1120, 2020.
- [17] S. A. Edwards and J. M. Yeomans. Spontaneous flow states in active nematics: A unified picture. *EPL*, 85:18008, 2009.
- [18] A. Doostmohammadi, T. N. Shendruk, K. Thijssen, and J. M. Yeomans. Onset of meso-scale turbulence in active nematics. *Nature Communications*, 8:15326, 2017.
- [19] S. Čopar, J. Aplinc, Ž. Kos, S. Žumer, and M. Ravnik. Topology of Three-Dimensional Active Nematic Turbulence Confined to Droplets. *Physical Review X*, 9:031051, 2019.
- [20] J. Binysh, Ž. Kos, S. Čopar, M. Ravnik, and G. P. Alexander. Three-dimensional active defect loops. *Physical Review Letters*, 124:257, 2020.
- [21] J. Urzay, A. Doostmohammadi, and J. M. Yeomans. Multi-scale statistics of turbulence motorized by active matter. *Journal of Fluid Mechanics*, 822:762, 2017.
- [22] Ž. Krajnik, Ž. Kos, and M. Ravnik. Spectral energy analysis of bulk three-dimensional active nematic turbulence. *Soft Matter*, 16:9059, 2020.
- [23] S. P. Thampi, R. Golestanian, and J. M. Yeomans. Active nematic materials with substrate friction. *Physical Review E*, 90:062307, 2014.
- [24] K. Kruse, J. Joanny, F. Jüicher, J. Prost, and K. Sekimoto. Aster, Vortices, and Rotating Spirals in Active Gels of Polar Filaments. *Physical Review Letters*, 92:078101, 2004.
- [25] F. Bonelli, G. Gonnella, A. Tiribocchi, and D. Marenduzzo. Spontaneous flow in polar active fluids: the effect of a phenomenological self propulsion-like term. *The European Physical Journal E*, 39:1143, 2016.
- [26] J. Elgeti, M. E. Cates, and D. Marenduzzo. Defect hydrodynamics in 2D polar active fluids. *Soft Matter*, 7:3177, 2011.
- [27] F. C. Keber, E. Loiseau, T. Sanchez, S. J. DeCamp, L. Giomi, M. J. Bowick, M. C. Marchetti, Z. Dogic, and A. R. Bausch. Topology and dynamics of active nematic vesicles. *Science*, 345:1135, 2014.

- [28] A. Tiribocchi, N. Stella, G. Gonnella, and A. Lamura. Hybrid lattice Boltzmann model for binary fluid mixtures. *Physical Review E*, 80:026701, 2009.
- [29] P. Guillamat, Ž Kos, J. Hardoin, J. Ignes-Mullol, M. Ravnik, and F. Sagues. Active nematic emulsions. *Science Advances*, 4:eaao1470, 2018.
- [30] R. Zhang, Y. Zhou, M. Rahimi, and J. J. de Pablo. Dynamic structure of active nematic shells. *Nature Communications*, 7:13483, 2016.
- [31] E. Tjhung, D. Marenduzzo, and M. E. Cates. Spontaneous symmetry breaking in active droplets provides a generic route to motility. *The Proceedings of the National Academy of Sciences (PNAS)*, 109:12381, 2012.
- [32] L. J. Ruske and J. M. Yeomans. Morphology of active deformable 3D droplets. *Physical Review X*, 11:021001, 2021.
- [33] A. Tiribocchi, R. Wittkowski, D. Marenduzzo, and M. E. Cates. Active model H: scalar active matter in a momentum-conserving fluid. *Physical Review Letters*, 115:188302, 2015.
- [34] M. E. Cates, O. Henrich, D. Marenduzzo, and K. Stratford. Lattice Boltzmann simulations of liquid crystalline fluids: active gels and blue phases. *Soft Matter*, 5:3791, 2009.
- [35] C. Denniston, D. Marenduzzo, E. Orlandini, and J. M. Yeomans. Lattice Boltzmann algorithm for three-dimensional liquid-crystal hydrodynamics. *Philosophical Transactions of the Royal Society A*, 362:1745, 2004.
- [36] S. Chen and G. D. Doolen. Lattice Boltzmann method for fluid flows. *Annual Review of Fluid Mechanics*, 30:329, 1998.
- [37] J.-C. Desplat, I. Pagonabarraga, and P. Bladon. LUDWIG: A parallel Lattice-Boltzmann code for complex fluids. *Computer Physics Communications*, 134:273, 2001.
- [38] T. Kruger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva, and E. M. Viggen. *The Lattice Boltzmann Method*. Springer International Publishing, Switzerland, 2017.
- [39] Z. Guo, C. Zheng, and B. Shi. Discrete lattice effects on the forcing term in the lattice Boltzmann method. *Physical Review E*, 65:046308, 2002.
- [40] V. M. O. Batista, M. L. Blow, and M. M. Telo da Gama. The effect of anchoring on the nematic flow in channels. *Soft Matter*, 11:4674, 2015.
- [41] Ž. Kos. *Microfluidic structures from nematic liquid crystals*. PhD thesis, University of Ljubljana, Faculty of Mathematics and Physics, Jadranska 19, 1000 Ljubljana, Slovenia, 2019.
- [42] M. Bouzidi, M. Firdauss, and P. Lallemand. Momentum transfer of a Boltzmann-lattice fluid with boundaries. *Physics of Fluids*, 13:3452, 2001.
- [43] S. H. Kim and H. Pitsch. A generalized periodic boundary condition for lattice Boltzmann method simulation of a pressure driven flow in a periodic geometry. *Physics of Fluids*, 19:108101, 2007.
- [44] J. Zhang and D. Y. Kwok. Pressure boundary condition of the lattice Boltzmann method for fully developed periodic flows. *Physical Review E*, 73:047702, 2006.
- [45] S. Izquierdo, P. Martínez-Lera, and N. Fueyo. Analysis of open boundary effects in unsteady lattice Boltzmann simulations. *Computers & Mathematics with Applications*, 58:914, 2009.
- [46] E. Orlandini, M. E. Cates, D. Marenduzzo, L. Tubiana, and J. M. Yeomans. Hydrodynamic of active liquid crystals: A hybrid lattice Boltzmann approach. *Molecular Crystals and Liquid Crystals*, 494:293, 2008.

- [47] C. Denniston, E. Orlandini, and J. Yeomans. Lattice Boltzmann simulations of liquid crystal hydrodynamics. *Physical Review E*, 63:056702, 2001.
- [48] K. Skarp, S. T. Lagerwall, and B. Stebler. Measurements of hydrodynamic parameters for nematic 5CB. *Molecular Crystals and Liquid Crystals*, 60:215, 1980.
- [49] T. B. Saw, A. Doostmohammadi, V. Nier, L. Kocgozlu, S. Thampi, Y. Toyama, P. Marcq, T. Lim, J. M. Yeomans, and B. Ladoux. Topological defects in epithelia govern cell death and extrusion. *Nature*, 544:212, 2017.
- [50] R. Zhang, N. Kumar, J. L. Ross, M. L. Gardel, and J. J. de Pablo. Interplay of structure, elasticity, and dynamics in actin-based nematic materials. *The Proceedings of the National Academy of Sciences (PNAS)*, 115:E124, 2018.
- [51] S. J. DeCamp, G. S. Redner, A. Baskaran, M. F. Hagan, and Z. Dogic. Orientational order of motile defects in active nematics. *Nature Materials*, 14:1110, 2015.
- [52] M. E. Cates, K. Stratford, R. Adhikari, P. Stansell, J.-C. Desplat, I. Pagonabarraga, and A. J. Wagner. Simulating colloid hydrodynamics with lattice Boltzmann methods. *Journal of Physics: Condensed Matter*, 16:S3903, 2004.
- [53] E. J. Hemingway, P. Mishra, M. C. Marchetti, and S. M. Fielding. Correlation lengths in hydrodynamic models of active nematics. *Soft Matter*, 12:7943, 2016.
- [54] M. Cui and J. R. Kelly. Temperature dependence of visco-elastic properties of 5CB. *Molecular Crystals and Liquid Crystals A*, 331:49, 1999.
- [55] H. J. Coles. Laser and Electric Field Induced Birefringence Studies on the Cyanobiphenyl Homologues. *Molecular Crystals and Liquid Crystals*, 49:67, 1978.
- [56] M. L. Magnuson, B. M. Fung, and J. P. Bayle. On the temperature dependence of the order parameter of liquid crystals over a wide nematic range. *Liquid Crystals*, 19:823, 1995.
- [57] O. Henrich, D. Marenduzzo, K. Stratford, and M. E. Cates. Domain growth in cholesteric blue phases: Hybrid lattice Boltzmann simulations. *Computers & Mathematics with Applications*, 59:2360, 2010.
- [58] C. W. Wolgemuth. Collective swimming and the dynamics of bacterial turbulence. *Biophysics Journal*, 95:1564, 2008.
- [59] M. Ravnik and S. Žumer. Landau de Gennes modelling of nematic liquid crystal colloids. *Liquid Crystals*, 36:1201, 2009.
- [60] L.N. Carenza, G. Gonnella, and G. Negro. Lattice Boltzmann simulations of self-propelling chiral active droplets. *arXiv:2011.00287 [cond-mat.soft]*, 2020.
- [61] S. Čopar, T. Porenta, and S. Žumer. Visualisation methods for complex nematic fields. *Liquid Crystals*, 40:1759, 2013.
- [62] S. Čopar and S. Žumer. Quaternions and hybrid nematic disclinations. *Proceedings of the Royal Society A*, 469:20130204, 2013.
- [63] U. Ayachit. *The ParaView Guide: A Parallel Visualization Application*. Kitware, New York, 2015.
- [64] POV-Ray. <http://www.povray.org/>. Accessed: 2020-11-30.
- [65] Ž. Kos and M. Ravnik. Two-dimensional hybrid lattice Boltzmann code for active nematodynamics, 2021.

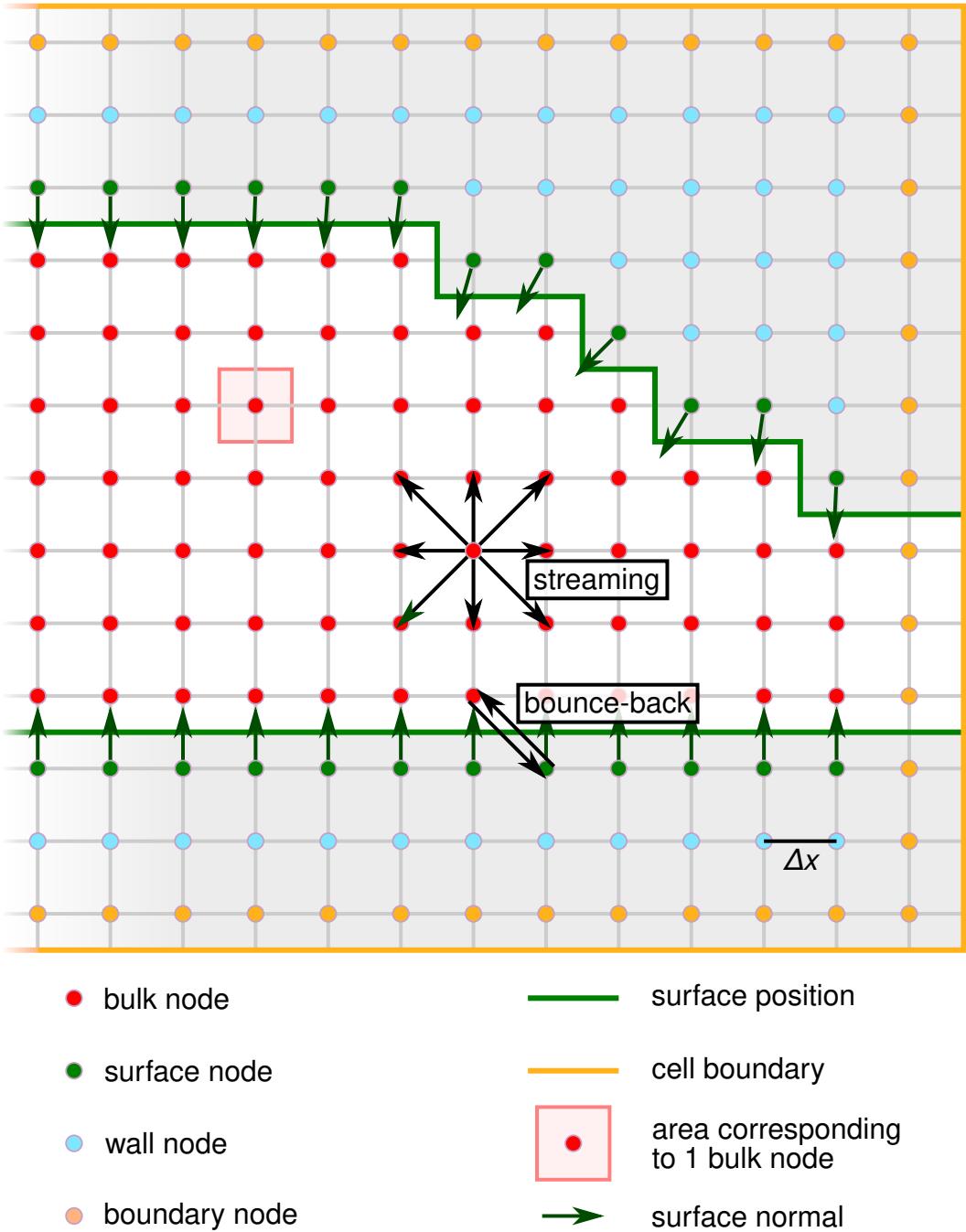


Figure 25.3: **Mesh allocation in the hybrid lattice Boltzmann method.** Red points represent bulk nodes in which streaming and collision (Eqs. (25.10) and (25.9)) of the velocity distribution functions and bulk finite difference time step (Eq. (25.16)) are performed. In the case of no-slip fluid-wall interactions, velocity distribution functions are bounced back if they reach a surface or a wall node. At surface nodes, a surface finite difference time step (Eq. (25.17)) is performed according to a surface preferred director orientation. A common anchoring type is *homotropic* with preferred orientation along the surface normal or *planar* with preferred orientation perpendicular to the surface normal. Boundary conditions are imposed at boundary nodes. For periodic boundary conditions, velocity distribution functions and Q-tensors are copied from the opposite side of the cell. To improve visibility, resolution in the mesh allocation sketch is severely reduced compared to an actual simulation.

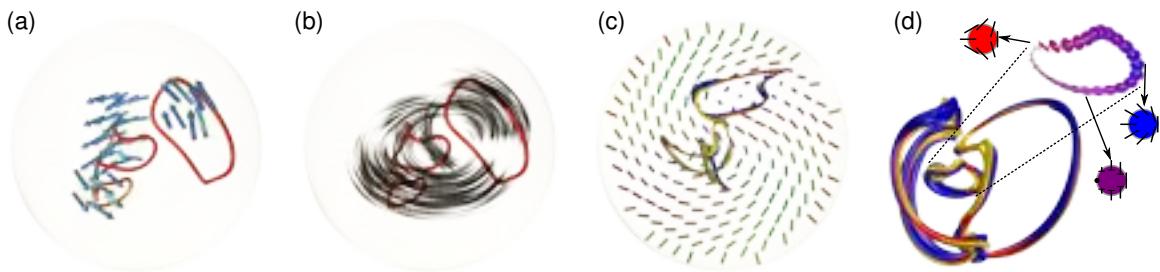


Figure 25.4: Visualization of fields in active nematic droplets [19]. Velocity and orientational fields around three defect loops in three-dimensional active droplets are represented using different visualization techniques. The defect cores are associated with low values of the scalar order parameter (the main eigenvalue of nematic tensor order parameter) and are drawn as isosurfaces in red color. The velocity field in (a) is represented by blue arrows at regions of highest magnitude, while in (b) it is represented by streamlines at selected cross-sections. In (c), a director field cross-sections is shown by rods, which are color-coded according to the in-plane orientation. Yellow and blue iso-surfaces around defect loop cores represent negative and positive extrema of the splay-bend parameter, respectively. (d) Zoom-in of a near-director field structure of a selected defect loop. Small disks represent the director field plane when circumnavigating a small defect segment. The color coding of the disks gives the angle between the disk normal and the tangent of the defect segment.

Chapter 26

Phase Field Models of Active Systems

ROMAIN MUELLER, AMIN DOOSTMOHAMMADI

In the preceding Chapters, diverse models of active matter systems were discussed in detail. While understanding pattern formation within active matter has taken important strides in the past decades [5, 6], the vast majority of active systems are characterized by dynamic interfaces between an active phase and the surrounding medium, which can itself be active or passive with significantly different properties, some examples of which are shown in Fig. 26.1. Whether it is a bacterial biofilm invading its surrounding environment [7, 8, 1], a cellular tissue closing a wound [9], or a cluster of human cancer cells migrating into extracellular matrices [10, 11], the interaction between active matter and its environment plays a crucial role in determining the behavior of

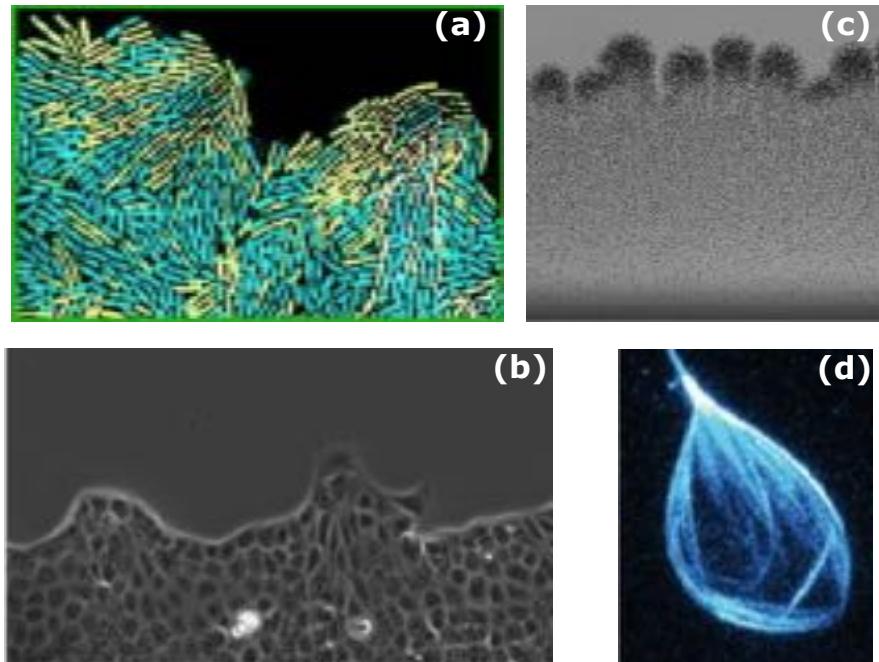


Figure 26.1: **Examples of natural and synthetic active interfaces.** (a) Bacterial competition during *Pseudomonas aeruginosa* biofilm invasion. Fast- and slow-moving bacterial strains are shown by yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP), respectively. Figure adapted from Ref. [1]. (b) Epithelial cell progression during wound healing. Figure adapted from Ref. [2]. (c) Synthetic active matter composed of colloidal rollers forming fingering patterns at the progressive front. Figure adapted from Ref. [3]. (d) Self-deforming active droplet composed of microtubule-motor protein mixtures in a fluid background. Figure adapted from Ref. [4].

prominent physiological processes.

In addition, even within the active matter itself, material properties are not always homogeneous and interfaces between active phases with different physical and chemical properties can arise. Striking examples include synthetic self-propelled particles that phase-separate because of their different self-propulsion [12], competition between different phenotypes during bacterial collective invasion [1], and cellular segregation in epithelial cells with distinct cell-cell interactions [13].

To correctly model such complex systems, it is then crucial to both capture the dynamics of each phase independently but also to accurately track the dynamic interface between them. This can represent, however, a major modeling challenge. For example, there are many relevant biophysical scenarios where active cellular materials interact with extracellular matrices with viscoelastic properties [8, 14]. Phase field methods provide a generic and versatile framework that is adaptable to such complex situations and allow for the modeling of multiple active or passive phases with their own complex dynamics and interactions.

Here, we present an overview of phase field modeling of active matter systems as a tool for capturing various aspects of complex and active interfaces. This Chapter is organized as follows: we first describe how interfaces between different phases are characterized in phase field models and provide simple fundamental governing equations that describe their evolution. For a simple model, we then show how physical properties of the interface, such as surface tension and interface thickness, can be recovered from these equations. We then explain how the phase field formulation can be coupled to various active matter realizations and discuss three particular examples of *continuum biphasic active matter*: *active nematic-isotropic interfaces*, active matter in *viscoelastic environments*, and *active shells* in fluid background. Finally, we describe how multiple phase fields can be used to model active cellular monolayers and present a general framework that can be applied to the study of tissue behavior and collective migration.

26.1 Phase field description of diffuse interfaces

The *phase field method*, originally introduced to model solidification processes [15, 16, 17], has since then seen a wide range of applications, such as the description of fingering and elastic surface instabilities, fluidization and crystallization in complex media, and the modeling of soft vesicles — see Ref. [18] for a review.

The basic idea behind this method is to describe interfaces between two distinct phases implicitly using an auxiliary field ϕ whose value distinguishes the two phases. For instance, $\phi = 0$ can denote the first phase while $\phi = 1$ the second (but this choice is arbitrary). The phase field is then subject to a mixing free energy such that the balance between mixing and demixing effects results in two distinct stable phases with a diffusive interface of small, but finite, thickness. Under complete demixing, a sharp interface separates these two phases, while the presence of a mixing free energy results in diffusion of the order parameter in the interfacial region. This method is extremely versatile and can be used to describe interfaces between a variety of different phases, such as the solid and liquid phases of a melting material, the nematic and homogeneous phases of a biphasic liquid crystal, or simply the inside and outside of a vesicle or a cell.

In what follows, we present a simple phase field model with two phases and show how it can be coupled to flow and stresses.

26.1.1 Free-energy functional

We present here the equations of a simple phase field model describing the demixing of two distinct phases and show analytically how the model parameters relate to the physical properties of the interface. For simplicity, we assume that the phase field ϕ relaxes diffusively towards the minimum of the free-energy $\mathcal{F} = \mathcal{F}[\phi, \nabla\phi]$ according to the equation of motion

$$\partial_t \phi = -\frac{\delta \mathcal{F}}{\delta \phi} = -\frac{\partial \mathcal{F}}{\partial \phi} + \nabla \cdot \frac{\partial \mathcal{F}}{\partial \nabla \phi}. \quad (26.1)$$

This equation of motion, which was first introduced in Ref. [19] for the study of binary alloys, is arguably the simplest choice of evolution for the phase field ϕ , but others are possible and can be found throughout the literature [18]. A simple free-energy exhibiting two distinct phases is the classical *Cahn–Hilliard free energy*

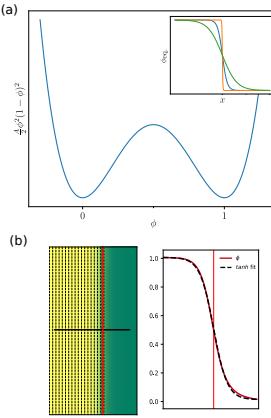


Figure 26.2: **Analysis of interfacial profile for ϕ .** (a) The double-well potential corresponding to the demixing free energy in Eq. (26.2). The inset shows the equilibrium phase field profile from Eq. (26.3) for different values of the interface width λ . The interface becomes sharp as λ goes to 0. (b) Simulation setup for measuring the interface between two phases with $\phi = 1$ (yellow) and $\phi = 0$ (green) in a complex fluid with nematic order parameter (see Section 26.2.1). The phase field profile is measured along the black solid line and, even in this, it is very close to the theoretical \tanh form of ϕ_{eq} .

first introduced in Ref. [20], which can be written as

$$\mathcal{F}_{\text{CH}} = \int d\mathbf{x} \left\{ \frac{A}{2} \phi^2 (1 - \phi)^2 + \frac{\kappa}{2} (\nabla \phi)^2 \right\}. \quad (26.2)$$

The first term is the demixing free energy and corresponds to a double-well potential with minima at $\phi = 0$ and $\phi = 1$ corresponding to the first and second phases, respectively, while the second term is the mixing part that penalizes gradients $\nabla \phi$ (see Fig. 26.2a). Note the phase field ϕ is not conserved in Eq. (26.1) but a conservation constraint can be easily introduced using Lagrange multipliers (see, e.g., Refs. [21, 22]).

In order to get a better understanding of the parameters appearing in Eq. (26.2), we search for steady state solutions of Eq. (26.1), which leads us to solve

$$\frac{\delta \mathcal{F}_{\text{CH}}}{\delta \phi} = A\phi(1 - \phi)(1 - 2\phi) - \kappa\Delta\phi = 0.$$

It is easy to verify that

$$\phi_{\text{eq}}(x) = \frac{1}{2} \tanh \left(\frac{x_0 - x}{\sqrt{4\kappa/A}} \right) + \frac{1}{2} \quad (26.3)$$

is a solution of the above equation with boundary conditions $\phi_{\text{eq}}(-\infty) = 1$ and $\phi_{\text{eq}}(+\infty) = 0$. This corresponds to a one-dimensional interface along the x -axis located at x_0 with interface width $\lambda = \sqrt{2\kappa/A}$ (see inset of Fig. 26.2a). Note that this solution is the absolute minimum of the free energy for these boundary conditions, but that infinitely many unstable minima exist where the phase field oscillates multiple times between $\phi = 0$ and $\phi = 1$ [23].

Inserting ϕ_{eq} into the free energy density, Eq. (26.2), and integrating perpendicularly to the interface allow us to directly obtain the surface tension between the two phases $\phi = 1$ and 0 as

$$\gamma = \int_{-\infty}^{\infty} dx \left\{ \frac{A}{2} \phi_{\text{eq}}^2 (1 - \phi_{\text{eq}})^2 + \frac{\kappa}{2} (\nabla \phi_{\text{eq}})^2 \right\} = \sqrt{A\kappa}/6.$$

A similar analysis can be carried out in two dimensions for the case of a radially-symmetrical droplet even though no analytical solution for the profile exists in this case. Considering a circular droplet whose radius R is much larger than the interface width λ , it is easy to see that the Laplacian in radial coordinates can be approximated by $\Delta = \partial_r^2 + 1/r\partial_r \approx \partial_r^2 + \mathcal{O}(\lambda/R)$, in which regime the hyperbolic tangent profile from above is recovered. When the radius of the droplet is somewhat similar to the interface width, this term cannot be

neglected and the interface profile will only be approximated by ϕ_{eq} . Also, note that in the non-stationary case $\partial_t \phi \neq 0$, the interface will also deviate from its equilibrium profile, but these deviations are found to be small in practice [22].

Using a free-energy-based description makes phase field models of interfaces easily adaptable for modeling different types of complex microstructured fluids, such as liquid crystals or viscoelastic fluids, where the mixing free energy for the phase field is simply complemented by the free energy of the microstructured fluid [22]. Another advantage of the phase field models compared to the other approaches for modeling fluid interfaces is that it can be proven that the energy is always conserved even in the case of describing complex fluids such as liquid crystals [24].

26.1.2 Hydrodynamics

One attractive feature of phase models is that their governing equation for the binary order parameter ϕ can be easily coupled to hydrodynamic equations of fluid flow and other order parameters such as orientation field, polarity field, and even viscoelasticity. In this section, we introduce a general framework through which such coupling is achieved, starting with a generic coupling to hydrodynamics and discussing specific examples from active nematics (where coupling to the orientation field is introduced) and active matter in viscoelastic surrounding (where coupling to viscoelasticity is described). We further discuss how the generic framework for coupling to other order parameters can be adapted for more complex setups such as active shells within fluid backgrounds.

In order to couple phase field model to fluid flow, the governing equation for the binary order parameter ϕ is simply supplemented with an advective flux of ϕ , $\mathbf{j} = \mathbf{u}\phi$, where \mathbf{u} is the fluid velocity. This gives

$$\partial_t \phi + \nabla \cdot \mathbf{j} = -\frac{\delta \mathcal{F}}{\delta \phi}. \quad (26.4)$$

As such, coupling to flow is easily achieved without introducing any additional parameters in the phase field equation. In its most general form, the fluid flow \mathbf{u} is in turn determined from the Navier-Stokes equation (see Chapter 22 [Is this reference correct?](#)):

$$\rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) = \nabla \cdot \boldsymbol{\Pi}, \quad (26.5)$$

where ρ is the fluid density and $\boldsymbol{\Pi}$ is the stress tensor. The stress tensor, in general, comprises fluid pressure p and viscous stresses $\eta \mathbf{E}$, with η the dynamic viscosity and $\mathbf{E} = (\nabla \mathbf{u})^{\text{Symm}}$ the strain rate tensor that characterizes the symmetric part of the velocity gradient tensor.

The presence of an interface between two phases gives rise to additional stresses in the flow field. To capture these additional stresses, the back-coupling from the binary order parameter ϕ to the fluid flow equation is introduced through capillary stresses [25, 26, 27, 28]:

$$\boldsymbol{\Pi}^{\text{capillary}} = (\mathcal{F} - \mu \phi) \mathbb{1} - \nabla \phi \frac{\partial \mathcal{F}}{\partial \nabla \phi}, \quad (26.6)$$

where $\mu = -\delta \mathcal{F} / \delta \phi$ is the chemical potential and $\mathbb{1}$ is the identity matrix. As such, the first term describes contributions to the fluid pressure (isotropic contributions to the stress tensor) due to particle exchange between two phases through chemical potential. The second term on the right-hand-side of Eq. (26.6) characterizes anisotropic contributions to the fluid stress due to the presence of surface tension between the binary phases. Calculating the divergence of the capillary stress, it can be shown that the corresponding force field simplifies to [25]

$$\mathbf{F}^{\text{capillary}} = \nabla \cdot \boldsymbol{\Pi}^{\text{capillary}} = \phi \nabla \mu. \quad (26.7)$$

Therefore, from Eqs. (26.1)-(26.7), the coupled set of equations for a phase field model in the presence of hydrodynamic fluid flow can be written as:

$$\begin{aligned} \partial_t \phi + \nabla \cdot (\mathbf{u}\phi) &= \mu, \\ \rho (\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) &= \nabla \cdot \boldsymbol{\Pi} + \phi \nabla \mu, \end{aligned} \quad (26.8)$$

where, as before, the stress tensor $\boldsymbol{\Pi}$ includes pressure and viscous contributions.

Equation (26.8) can be solved numerically using hybrid schemes combining a finite-difference integration method [29] for the scalar binary order parameter with the desired fluid solver, e.g. the lattice Boltzmann method [30, 31] (see Chapter 24).

The framework discussed above represents a general formulation for the coupling between the binary order parameter and hydrodynamic fluid flow, and can be easily adapted to problems where the fluid flow is generated by active particles. For active fluids, additional active stresses can easily be added to the stress tensor [27, 32, 33] and their exact form depends on the type of the active system under consideration. The important point, however, is that the effect on the evolution of the binary order parameter can be described by only considering the active flux ($\mathbf{j} = \mathbf{u}\phi$) that is solely determined from the resulting velocity field. In what follows, we discuss specific examples from active nematic models to make this more clear.

26.2 Biphasic active systems

Biphasic active systems arise naturally in a variety of biological and synthetic settings, where an active phase coexists with a second medium that can be either passive or itself active with different physical properties. Examples range from bacterial biofilms expanding into surrounding fluids, to living tissues embedded in viscoelastic matrices, and even microtubule-based active shells in a bulk fluid. In all such cases, not only must one accurately capture the bulk dynamics of each phase, but also the evolution of the interface between them. In the following sections, we show how phase field methods offer a versatile framework to achieve this by coupling the interfacial phase field to appropriate order parameters (e.g., nematic or viscoelastic fields) and fluid flow. We will illustrate, with concrete examples, how these biphasic formulations reveal the rich variety of instabilities, morphological transitions, and dynamical patterns that emerge when an active phase interacts with a second, possibly distinct, phase.

26.2.1 Active nematic interfaces

As described in detail in Chapter 12, *active nematics* represent a class of active materials that are characterized by having a local orientational order parameter. This comprises solutions of elongated particles such as bacteria [36, 1], microtubule filaments put in motion by kinesin motor proteins [37, 38] or actin filaments and myosin mixtures [39], but also spindle-shaped cells such as mouse embryonic fibroblasts [40] and neural progenitor cells [41]. More surprisingly, deformable cells, such as Madine–Derby Canine Kidney cells (MDCK) [42], human bronchial cells (HBC) [43], and human fibrosarcoma [44], can also be described by active nematic theories, where the orientational order is defined as the direction of cell elongation. In nematic systems, the orientational order is characterized by a second-rank nematic tensor \mathbf{Q} . The field associated with the orientational order of elongated particles must be represented by a tensor since it must show head-tail symmetry, i.e., be apolar. For this reason, it cannot be defined using a directional unit vector $\hat{\mathbf{n}}$ only and instead higher moments must be used to define the nematic tensor $\mathbf{Q} = q \frac{d}{d-1} (\hat{\mathbf{n}}^\top \hat{\mathbf{n}} - \mathbf{1}/d)$, where q is the magnitude of the nematic order that corresponds to the largest eigenvalue of \mathbf{Q} (see Chapter 12 for details on the nematic tensor definition).

It is well-established that for an active nematic system, the prominent contribution to active stresses is proportional to the nematic tensor \mathbf{Q} such that the active stress takes the form $\zeta \mathbf{Q}$ [45, 46, 47, 33], where ζ determines the strength of activity. Therefore, in order to account for the effect of activity, an active stress proportional to the nematic tensor \mathbf{Q} is added to the stress tensor in Eq. (26.8): $\Pi^{\text{active}} = \phi \zeta \mathbf{Q}$. Since the divergence of the stress determines the force acting on the fluid, this means that any gradients in the nematic tensor \mathbf{Q} result in additional forces and actively generate fluid flow [46]. Moreover, the proportionality of the stress to the binary order parameter ϕ ensures that the active stress is only applied in the phase with $\phi = 1$ and is zero in the other phase where $\phi = 0$. Such a form for the active stress can be used to model systems with a binary mixture of active and passive nematics (where both phases have orientational order, but only one of them is active), such as mixtures of live and dead bacteria [48] or a mixture of motor proteins-filaments in which only a part of the mixture is activated through light-activated motor proteins [49].

In several experimental conditions, however, an active phase of particles with orientational order, forms an interface with an otherwise isotropic fluid. This can be easily accommodated within the framework described here by ensuring that the passive phase $\phi = 0$ has no orientational order, i.e., $\mathbf{Q} \rightarrow \mathbf{0}$ when $\phi \rightarrow 0$. To this end,

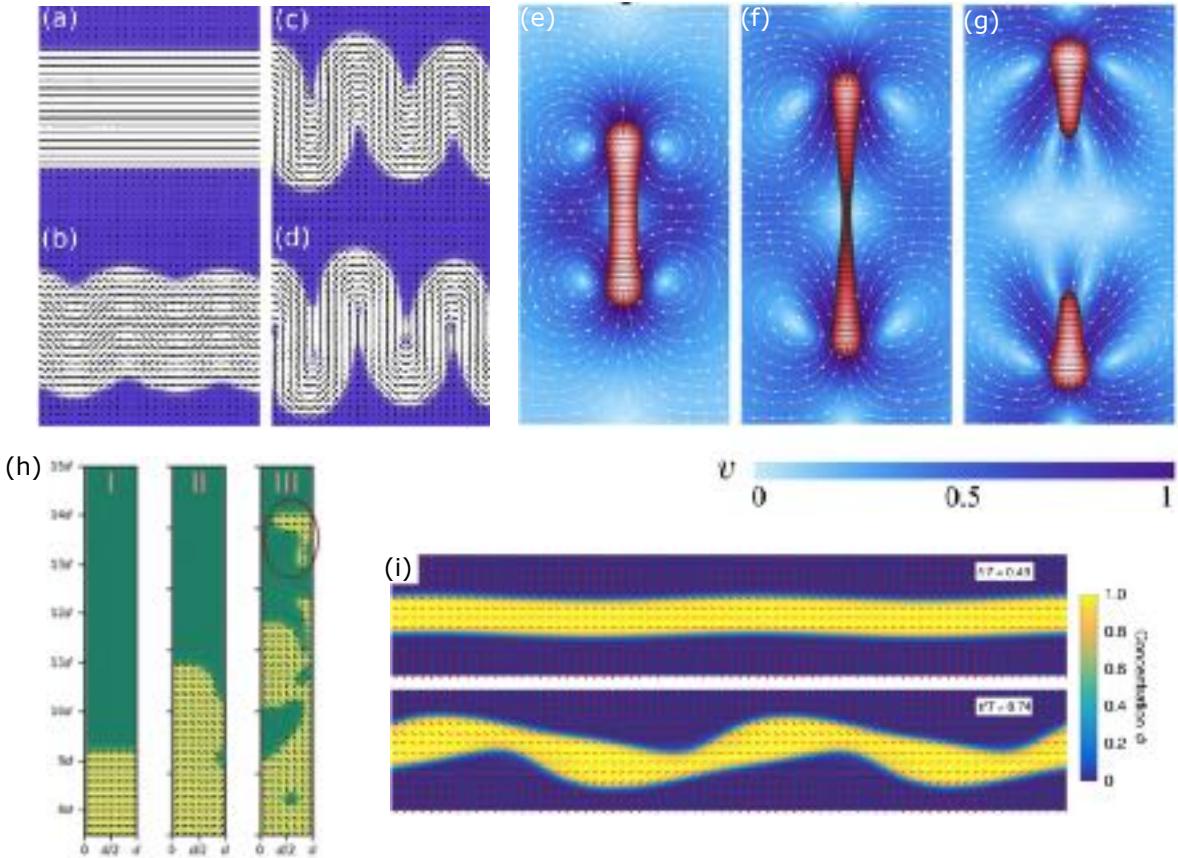


Figure 26.3: Modeling active biphasic systems composed of active nematics and isotropic phases. (a-d) Temporal evolution of the instability of an active nematic band (white region) within an isotropic fluid background (blue region). The black solid lines represent the director field of the active nematic. An initially flat interface (a) is unstable to bend deformation of the active nematics (b), which grows in amplitude (c) and generate a highly deformed interface (d). Figure adapted from [26]. (e-g) A drop of active nematic elongates (e), forms a narrow neck (f), and eventually divides (g). The black solid lines represent the director field of the active nematic and the white vectors indicate the flow streamlines. The active nematic region is marked as red, while the isotropic fluid region is shown in blue. Figure adapted from Ref. [34]. (h) Growing active nematic (yellow region) invading a surrounding isotropic fluid (green region). I-III show different patterns of invasion for increasing activities. Figure adapted from Ref. [28]. (i) Active nematic fluid (yellow region) in a background passive liquid crystal develops instabilities that do not grow beyond a certain length set by the properties of the background liquid crystal. Red solid lines represent the director field of the background liquid crystal that are initially patterned to form a splay region in the middle. Figure adapted from Ref. [35].

the Landau–De Gennes free energy for the nematic introduced in Chapter 12, $\mathcal{F}_{\text{L-G}} = \frac{1}{2}\mathbf{Q}^2 + \frac{1}{4}\mathbf{Q}^4$ [50, 51], is coupled to the binary order parameter ϕ as follows:

$$\mathcal{F}_{\text{L-G}} = \frac{1}{2}\phi\mathbf{Q}^2 + \frac{1}{4}\mathbf{Q}^4. \quad (26.9)$$

This free energy ensures that in the active nematic phase $\phi = 1$, the orientational order is retained, while in the passive phase $\phi = 0$, the magnitude of orientational order that minimizes the free energy — i.e., \mathbf{Q} for which $\delta\mathcal{F}/\delta\mathbf{Q} = 0$ — diminishes to zero, resulting in an isotropic fluid [26, 28]. In Fig. 26.2b, we compare analytical results for the one-dimensional ϕ profile to the results from numerical integration of Eq. (26.1).

An interesting feature of active nematic interfaces, compared to their passive counterpart, is the emergence of the phenomenon of *active anchoring* [26]: in the absence of any anchoring energies, the presence of activity alone results in a certain alignment of active particles at the interface. This effect has been shown for dividing

bacteria, where cell division provides a source of active force generation, that align parallel to the interface of their colony [52] and can also be seen in Fig. 26.1a. Considering the active force

$$\begin{aligned} F_\alpha^{\text{active}} &= -\zeta \partial_\beta (\phi Q_{\alpha\beta}) \\ &= -\zeta (\partial_\beta \phi S) (2n_\alpha n_\beta - \delta_{\alpha\beta}) - 2\zeta \phi S (n_\alpha (\partial_\beta n_\beta) + (\partial_\beta n_\alpha) n_\beta), \end{aligned}$$

the active anchoring can be understood by decomposing the active force parallel and perpendicular to the interface between active and passive phases [26]:

$$f_\perp = m_\alpha f_\alpha = \zeta |\nabla(\phi S)| (2(m_\alpha n_\alpha)^2 - 1), \quad (26.10)$$

$$f_\parallel = l_\alpha f_\alpha = 2\zeta |\nabla(\phi S)| (l_\alpha n_\alpha) (m_\beta n_\beta), \quad (26.11)$$

where $\mathbf{m} \equiv -\nabla S / |\nabla S| \equiv -\nabla(\phi S) / |\nabla(\phi S)|$ and \mathbf{l} define the outward surface normal and the surface tangent directions, respectively. [Check consistency of notation between equation and explanation]

Depending on the sign of the activity parameter ζ , the normal force on the interface f_\perp points outwards ($\zeta < 0$) or inwards ($\zeta > 0$) for $\mathbf{n} \perp \mathbf{m}$ and hence induces active forces pushing the interface outwards or inwards, respectively. Therefore, for $\zeta > 0$ activity favors alignment of active particles parallel to the interface, while $\zeta < 0$ favors perpendicular alignment. The case of $\zeta > 0$ describes extensile active systems such as bacteria, where particles pull the surrounding fluid from their side and push it along their tail and head. The $\zeta < 0$ case, on the other hand, corresponds to contractile active systems such as actomyosin networks, where active constituents contract along their direction of elongation.

The method presented here has been extensively used to model fundamental aspects of active interfaces, including the study of the instabilities of active nematic interfaces [26, 53] (Figs. 26.3a-d), describing dynamics of elongation, motility, and division of active nematic droplets as a model of eukaryotic cells [34, 54] (Figs. 26.3e-g), and describing wetting dynamics of active nematics on solid surfaces [55, 56]. Furthermore, the biphasic modeling of active nematic interaction with isotropic fluid has been used to describe morphologies of growing bacterial colonies [52, 57], active matter invasion into new territories [28] (Fig. 26.3h), and bacterial streams within pre-patterned medium in living liquid crystals [35] (Fig. 26.3i).

26.2.2 Active matter in a viscoelastic medium

The exact same procedure used in the previous section can be followed to introduce couplings to other order parameters. One important example is active matter in a viscoelastic environment, which has applications to the modeling of several biophysical systems of interest. For example, cells in extracellular matrices are characterized by the cross-talk between an active phase and a medium which has viscoelastic properties [14] (Fig. 26.4a). Following the framework introduced in the previous section, we extend our formalism to the case of an active nematic in contact with a fluid phase that, rather than being isotropic, has now viscoelastic properties.

To characterize a viscoelastic fluid, it is common to introduce a conformation tensor \mathbf{C} as the corresponding order parameter, which describes the orientation and elongation of polymers that are present in the fluid [59, 58]. The conformation tensor \mathbf{C} is defined such that: (i) it satisfies $\mathbf{C} = \mathbf{1}$ at equilibrium, (ii) its trace $\text{tr}[\mathbf{C}]$ characterizes the square of the polymer elongation, and (iii) the eigenvector corresponding to its largest eigenvalue represents the local polymer orientation. Therefore, as before, a coupling between the binary order parameter and the conformation tensor can be introduced such that within the active phase ($\phi = 1$) there is an orientational order characterized by the nematic tensor \mathbf{Q} but no viscoelasticity $\mathbf{C} = \mathbf{0}$, while in the passive phase ($\phi = 0$) there are no active nematic particles $\mathbf{Q} = \mathbf{0}$ but polymers are present $\mathbf{C} \neq \mathbf{0}$. To ensure this, the free energy of polymers is coupled to the binary order parameter ϕ as follows:

$$\mathcal{F}_{\text{polymer}} = \frac{\nu}{\tau} \frac{1}{2} (1 - \phi) (\text{tr}[\mathbf{C} - \mathbf{1}] - \log(\det[\mathbf{C}])), \quad (26.12)$$

where ν characterizes the polymer viscosity and τ is the polymer relaxation time. Note that the factor $1 - \phi$ in front of the polymer free energy ensures that the order parameter is retained within the passive phase $\phi = 0$, while it diminishes otherwise when $\phi = 1$. Finally, since polymers are within the passive phase, no extra

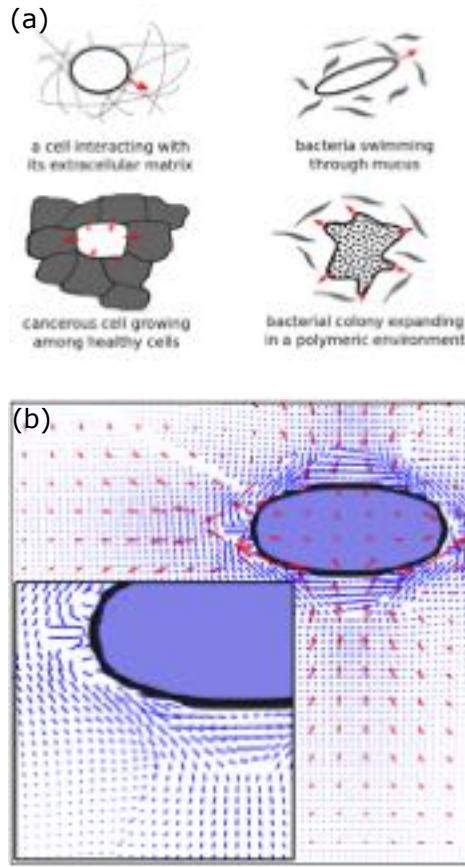


Figure 26.4: Modeling active biphasic systems composed of active nematics and passive polymeric phases. (a) A schematic of various interpretations of the two-phase active-viscoelastic model. (b) Typical flow field of a drop of active nematics elongating during the initial steps of cell division. Velocity vectors are shown as red arrows, and polymer deformations are shown as blue line segments. Inset: Zoom of the lower left part of the cell, showing polymers aligning with the stretching direction. Figure adapted from Ref. [58].

addition to active stresses is required. However, passive stresses need to be accounted for in Eq. (26.8) as follows:

$$\boldsymbol{\Pi}^{\text{polymer}} = \frac{\nu}{\tau}(1 - \phi)(\mathbf{C} - \mathbb{1}). \quad (26.13)$$

It is noteworthy that the formulation described above uses an Oldroyd-B model of polymer relaxation since it is one of the simplest and most widely-used polymer models [58]. The framework can be trivially extended to other constitutive models of viscoelastic fluids in two and three dimensions.

The important feature of the present framework that couples viscoelasticity, fluid flow, and activity through the phase field formulation is its versatility, which means that it can be applied to various biophysical examples. For instance, similar formulation has been used to study a model of cell division and motility within polymer gels with variable viscoelastic properties (Fig. 26.4b), and to study the hampering effect of polymers on generic instabilities of active matter in a viscoelastic environment [58] (Fig. 26.4).

26.2.3 Self-deforming active surfaces

So far, we have described applications of the phase field method to binary mixtures, where $\phi = 0$ is used to characterize one phase and $\phi = 1$ to describe the other, with $\phi = 0.5$ marking the interface between the two phases. It is, however, possible to extend this framework to model scenarios in which thin shells of one phase are embedded within another phase. An important example is the actomyosin cortex of eukaryotic cells that plays a vital role in shaping cell deformation and various cell functions [62]. Another interesting example is the

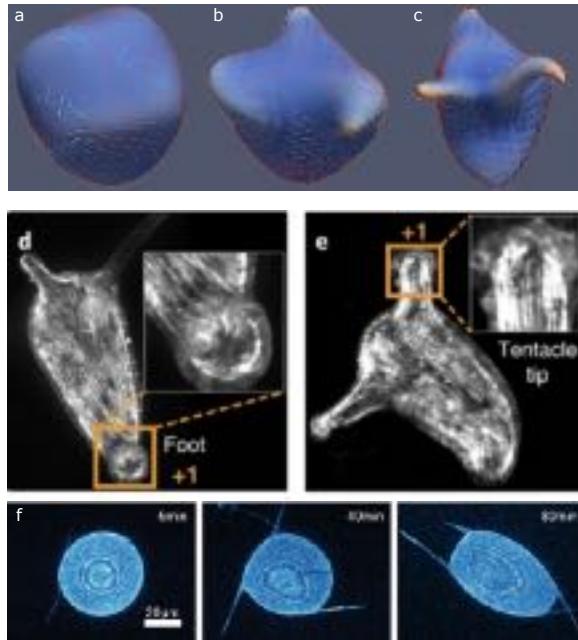


Figure 26.5: Modeling three-dimensional active self-deforming surfaces. (a-c) Temporal evolution of an active shell made of active nematics in an isotropic fluid background. The active shell develops protrusions at the position of $+1/2$ defects. The protrusions eventually develop $+1$ defects at their tip leaving behind $-1/2$ defects. The solid lines illustrate the nematic director and are colored by the magnitude of the nematic order. The shell is colored by the curvature of the surface. Figure reproduced from Ref. [60]. (d-e) Developments of protrusions and topological defects during the morphogenesis of regenerating animal *hydra*. Figure reproduced from Ref. [61]. (f) Temporal evolution of protrusions in a shell of active nematics built from microtubule-motor protein mixtures. Figure reproduced from Ref. [4].

active deformable shell formed by stabilizing microtubule-motor protein mixtures at the oil–water interface of a droplet [4, 63, 64]. The resulting active shell continuously exerts active stresses on the surrounding medium, generating dynamic patterns of motion and exotic morphologies that can be considered as examples of self-deforming and self-shaping materials (see Fig. 26.5a).

The distinguishing feature of these systems is that the shell itself is active and as such activity is generated at a thin interface between the inside and the outside of the shell. The phase field model of active interfaces described in this section can be easily adapted to model active shells by describing the coupling between the binary order parameter ϕ and the desired order parameter (in the case of active nematic this is the nematic tensor \mathbf{Q} , which is the source of active stresses) such that the active phase is only stabilized at the interface, i.e., where $\phi \approx 0.5$. For instance, starting from the active nematic system described in Section 26.2.1, it is only required to modify the free energy as follows:

$$\mathcal{F}_{\text{L-G}} = \frac{1}{2} \left(1 - 2|\phi - \frac{1}{2}| \right) \mathbf{Q}^2 + \frac{1}{4} \mathbf{Q}^4, \quad (26.14)$$

where the term $(1 - 2|\phi - \frac{1}{2}|)$ ensures that $\mathbf{Q} \rightarrow \mathbf{0}$ as both $\phi \rightarrow 0$ and $\phi \rightarrow 1$, while the orientational order is retained at the interface ($\phi \approx 0.5$). Again, this provides only a simple example of how such coupling can be achieved. The framework can be trivially extended to consider further complexities of the interface, for example, by including Helfrich type free energies that describe bending rigidity of the shell [60].

This method has been recently applied to model self-deforming active nematic shells in fluid backgrounds, reproducing topological defect dynamics on the shell surface observed in the experiments on microtubule-kinesin motor mixtures [4, 63, 60], and predicting exotic morphologies that can be obtained by tuning the activity of the shell. More importantly, such modeling provides a tool to probe the nature of cross talk between activity and self-induced surface curvature. For example, it is predicted that protrusions are initiated at the

locations of $+1/2$ defects in the orientation field, creating finger-like structures with a $+1$ topological defect at their tip, leaving behind $-1/2$ topological defects in regions with negative surface curvature (Fig. 26.5b). Similar dynamics and cross talk between activity and morphology has been recently reported in the development of regenerating animal *hydra* where actin filaments form an active shell that is capable of self-deformation [61] (Fig. 26.5c).

The examples described above provide an introduction to the applications of phase field model for modeling active systems. We focused particularly on active nematic interfaces, where the dynamics of the phase field is coupled to a tensor order parameter and which describes the orientational order of active particles. Importantly, phase field models have been applied to various realizations of active systems including scalar phase field models describing the phenomenon of motility-induced phase separation [65] (see [27] for a recent review), as well as polar active matter, where phase field dynamics is coupled to a vector order parameter describing the polarity of active particles [66, 67]. The latter has been particularly successful in modeling cellular motility, where polar active matter is used to describe the dynamics of actomyosin networks inside the cell, governing cell deformation and its morphology [68, 69, 70, 54]

26.3 Cell monolayers

In the previous sections, we discussed applications of phase field models to continuum representations of active interfaces. Interestingly, the same framework can be adapted to agent-based modeling of active systems where each agent is represented by an individual phase. An important example is the phase field modeling of cell monolayers which allows to model the deformations of individual cells and describe their physical properties, as well as introducing cell-cell interactions easily through the interfaces of the individual phase fields. In this section, we discuss how such models are constructed and applied to the understanding of the dynamics of cellular layers.

26.3.1 Cells as active deformable droplets

In silico models of cell motility have an important role to play in unravelling the interplay between single cell properties and their collective dynamics [71, 72]. There exist a breadth of numerical approaches to tackle this problem, such as cellular Potts models [73], vertex models [74, 75, 76] (see Chapter 20), continuum models [77, 32], and phase field models [78] (see the extensive reviews in Refs. [72, 79, 80]). The phase field approach is particularly interesting as it allows to describe the collective behavior of cells while modeling their physical properties at the individual level as well as their interactions explicitly. It has been widely applied to problems involving single cells [81, 68, 82, 69], few migrating cells [83], as well as systems of

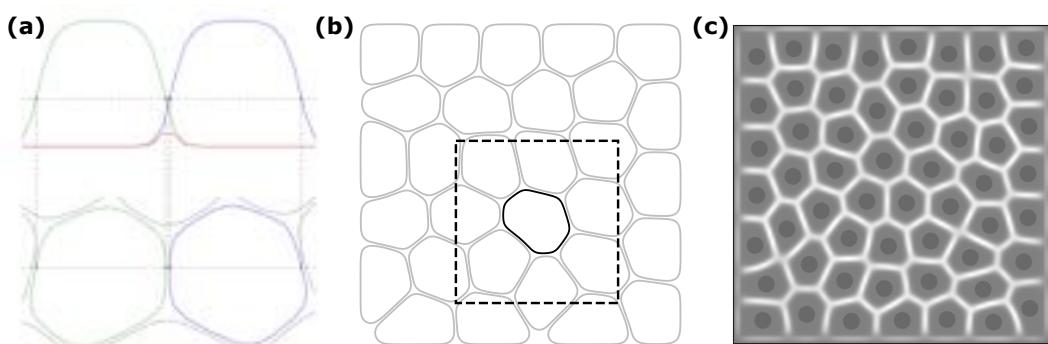


Figure 26.6: Modeling a cell monolayer composed of multiple active phase fields representing each cell. (a) Schematic representation of the phase fields of two cells (green and blue) as well as their overlap (red) seen from the side. (b) A phase field representation of a cell monolayer as a collection of active deformable drops seen from the top. The square around the cell represents the sub-domain on which the equations of motion are solved, see Section 26.3.3. (c) A confluent monolayer in confinement where the overlap between cells and their centres of mass are shown instead of the cell contours.

colliding binary cells [84, 78]. More recent works [70, 85, 86, 87, 88, 89] have concentrated on the collective dynamics of a large number of cells and have uncovered interesting phenomena such as oscillatory patterns of cells under confinement [86], as well as solid–liquid and flocking transitions [90]. However, since this approach is computationally quite intensive, it is necessary to reduce the description of cells to their bare minimum by abstracting away their internal machinery while retaining their relevant physical properties. Here, we concentrate on a simple model that represents cells on a substrate in two-dimensions and assumes simple physical properties for the cells.

A fruitful approach, first pioneered by Ref. [85], is to separate the description of the physical properties of the cells from their dynamics. In this model, each cell is described as a deformable active particle whose shape is defined using an individual phase field (Figs. 26.6a-b), while its velocity is described separately and is given by a force-balance equation. A cellular monolayer consisting of N cells can then be described by the equations of motion

$$\partial_t \phi_i + \mathbf{v}_i \cdot \nabla \phi_i = -\frac{\delta \mathcal{F}}{\delta \phi_i}, \quad i = 1, \dots, N, \quad (26.15)$$

where \mathcal{F} is a free energy describing the physical properties of the cells, and ϕ_i and \mathbf{v}_i are the phase field and the center-of-mass velocity of cell i , respectively. The phase fields are defined such that $\phi = 0$ denotes the exterior of the cell i , while $\phi = 1$ denotes its interior. This mirrors Eq. (26.1) with the difference that there is a different phase field for each cell and that each individual cell is driven by a simple advection term. The velocity \mathbf{v}_i for each cell can be obtained from a force balance equation: since Reynolds numbers are typically of the order of $\sim 10^{-4}$ for cell monolayers [91], we can safely assume overdamped dynamics and write

$$\xi \mathbf{v}_i = \mathbf{F}_i^{\text{tot}} = \mathbf{F}_i^{\text{pol}} + \mathbf{F}_i^{\text{inter}}, \quad (26.16)$$

where ξ is a friction coefficient and $\mathbf{F}_i^{\text{tot}}$ is the total force exerted on cell i , which will be defined below. Note that \mathbf{v}_i is a single vector for each cell and does not depend on \mathbf{x} . The definition of the total force $\mathbf{F}_i^{\text{tot}}$ depends on the model and can include contributions from different sources such as passive interface forces, interactions with other cells or walls, or active forces such as polar or nematic driving. Note that we have assumed here for simplicity that the free energy is the same for all cells but this does not need to be the case in general and it is straightforward to extend the above equation to cases where different cells have different physical properties, see for example [13]. Equations (26.15) and (26.16) are our master equations and define a general framework where only the free-energy \mathcal{F} and the total force $\mathbf{F}_i^{\text{tot}}$ on each cell need to be specified.

26.3.2 Minimal model of cell monolayers

Let us now define a minimal model that allows the successful description of cell monolayers following Refs. [85, 88] and write the free energy as a sum of three distinct contributions as $\mathcal{F} = \mathcal{F}_{\text{CH}} + \mathcal{F}_{\text{area}} + \mathcal{F}_{\text{rep}}$. The first term is the *Cahn–Hilliard free energy* (26.2), which we rewrite as

$$\mathcal{F}_{\text{CH}} = \sum_i \frac{\gamma}{\lambda} \int d\mathbf{x} \{ 4\phi_i^2(1-\phi_i)^2 + \lambda^2(\nabla \phi_i)^2 \}, \quad (26.17)$$

using the definitions of the interface width $\lambda = 2\sqrt{\kappa/A}$ and surface tension $\gamma = \sqrt{A\kappa}/6$ from Section 26.1 (up to an overall factor of 3). As shown in Section 26.1, this term creates and stabilises an interface between the interior ($\phi_i = 1$) and the exterior ($\phi_i = 0$) of each cell and results in a equilibrium shape for the interface that follows an approximate hyperbolic tangent profile of size λ . The second term is a soft area constraint defined by

$$\mathcal{F}_{\text{area}} = \sum_i \mu \left(1 - \frac{1}{\pi R^2} \int d\mathbf{x} \phi_i^2 \right)^2, \quad (26.18)$$

which is a square potential of strength μ ensuring the cells areas $A_i = \int d\mathbf{x} \phi_i^2$ are close to πR^2 .¹ Note that the phase field ϕ_i is not conserved in Eq. (26.15) and that even though cells are mostly incompressible in three dimensions, the area that individual cells occupy within a monolayer can dramatically change as they are

¹The area is proportional to the *square* in order to ensure that it is always positive even when the phase field is slightly negative which can happen during simulations.

squeezed by their neighbors and expand in the direction perpendicular to the substrate [92]. The final term discourages overlap between cells and is simply given by

$$\mathcal{F}_{\text{rep}} = \sum_i \sum_{j \neq i} \frac{\kappa}{\lambda} \int d\mathbf{x} \phi_i^2 \phi_j^2, \quad (26.19)$$

where κ is the strength of the repulsion. Note that this term implements a coupling between the cells and is proportional to the cell overlaps, see an illustration on Fig. 26.6c. The physical properties of the cells are parameterized by λ , γ , μ , and κ which set the interface width, the surface tension, the strength of the elastic restoring force for the area, and the repulsive force between cells, respectively. Normalization is chosen such that the properties of the cells are roughly preserved when the interface width λ is rescaled (see Ref. [88]).

Cellular deformations such as changes of shape or area should lead to forces at the cell boundaries and contribute to the overall force balance. Such interface forces can be constructed in a thermodynamically consistent way using the free energy as

$$\mathbf{F}_i^{\text{inter}} = \int d\mathbf{x} \sum_j \left(\frac{\delta \mathcal{F}_{\text{rep}}}{\delta \phi_j} - \frac{\delta \mathcal{F}_{\text{CH}}}{\delta \phi_j} - \frac{\delta \mathcal{F}_{\text{area}}}{\delta \phi_j} \right) \nabla \phi_i, \quad (26.20)$$

see Refs. [85, 88] for more details. Since $\nabla \phi_i$ is only non-zero at the cell boundary and pointing towards the cells center, this expression can be interpreted as the integral over the interface of the cell i of the total force density generated by changes of the free energy. Contributions from interactions with other cells come with a positive sign (repulsion), while self-contributions come with a negative sign (restoring forces). With this and in the absence of any active contribution, a confluent cell monolayer will relax to a minimum of the total free energy and favor hexagonal lattice arrangement (Fig. 26.6c).

The great strength of this formulation is its versatility: it can be easily extended to more complex models that can include forces such as adhesion between cells [93] or viscous friction [86]. Moreover, as shown in [86], it is also easy to model non-trivial boundary conditions (such as walls) by introducing a static phase field ϕ_{walls} that interacts with the cells with a simple repulsion term.

26.3.3 Numerical implementation

The coupled system of differential equations defined by Eqs. (26.15) and (26.16) can be solved using standard finite difference schemes [29]. The main difficulty is the large number of phase fields ϕ_i that are required when simulating a monolayer with many cells. In fact, it is easy to see that a naive simulation will have at least quadratic computational and memory requirements with respect to the number of cells due to each phase field extending over the whole domain. This can be mitigated by realising that the individual phase fields are non-zero only in a well-defined region around the center of the corresponding cell, which allows to simulate each phase field only in a sub-domain centered around its center-of-mass [94, 88] (see Fig. 26.6b).

A common choice is to select a square of lattice points that tracks the center of mass of each cell and use Dirichlet boundary conditions at the boundary of the sub-domain [89]. It is important to note, however, that because the centre of mass of the cell can lie between lattice points, the location of the sub-domain changes abruptly every time the cell center moves by a lattice length along any axis. This can lead to numerical artifacts when the size of the sub-domain is not large enough compared to the radius of the cell. A more refined approach is to use periodic boundary conditions instead [88], which has the advantage of being completely insensitive to the discrete hopping of the sub-domain and allows smaller sub-domain sizes compared to Dirichlet boundary conditions.

With such techniques, both the computational and memory costs of the algorithm scale linearly with respect to the number of cells when simulating a multicellular monolayer. Note that even though cells are only simulated around their center of mass, the computation of the total force acting on a cell interface will still require the computation of some quantities over the whole domain, such as the sum of all phase fields $\sum_i \phi_i$ and their squares $\sum_i \phi_i^2$, but the number of such quantities is fixed and does not scale with the number of cells. An implementation by the authors of the model presented in Section 26.3.2 using these techniques can be found online at <https://github.com/rhomu/celadro>.

26.3.4 Active polar driving

The driving forces behind the movements of single cells on a flat substrate are well understood. Directional actin filaments, which continuously polymerize and depolymerize, allow the cell to create a pushing force against the substrate via focal adhesion, which are mechanical links between internal actin bundles and the external surface [95]. In doing so, the cell polarizes its actin filaments and tends to elongate in the direction of motion, creating protrusions (lamellipodia) [96, 97]. This suggests a minimal physical picture of cellular motility where each cell experiences a net active force

$$\mathbf{F}_i^{\text{pol}} = \alpha \mathbf{p}_i$$

in the direction of its polarity \mathbf{p}_i , where α is a parameter denoting the strength of the focal adhesion (Fig. 26.7a). This force contributes to the force-balance Eq. (26.16) and propels the phase field ϕ_i uniformly in the direction of \mathbf{p}_i . A simple non-trivial implementation can be obtained by introducing for each cell a new internal degree of freedom θ_i^{pol} describing the direction of its polarity $\mathbf{p}_i = (\cos \theta_i^{\text{pol}}, \sin \theta_i^{\text{pol}})$ and assume that the polarization

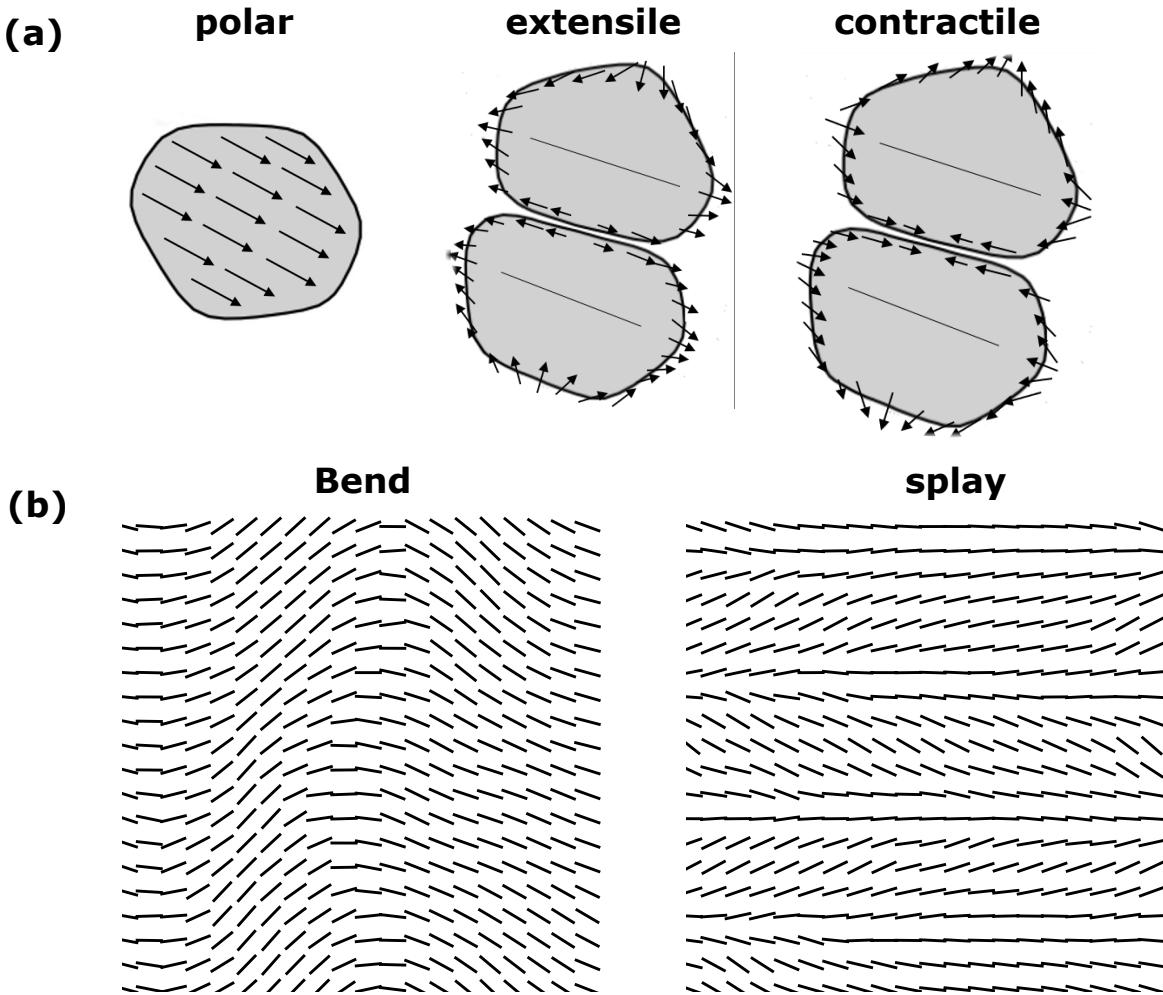


Figure 26.7: **Active forces and instabilities in cell monolayers.** (a) Schematic representation of different sources of active force. Figure adapted from Ref. [93]. (b) Coarse-grained nematic field of the cells showing the emergence of multicellular scale bend instability for extensile active forces and splay instability for contractile active forces acting at the individual cell level. The cells are driven out of equilibrium by an internal nematic degree of freedom.

angle relaxes diffusively towards an arbitrary vector \mathbf{d}_i [86]. This leads to the following equation of motion

$$\partial_t \theta_i^{\text{pol}} = -J^{\text{pol}} |\mathbf{d}_i| \angle(\mathbf{p}_i, \mathbf{d}_i) + (2D^{\text{pol}})^{\frac{1}{2}} \eta_i, \quad (26.21)$$

where $\angle(\mathbf{p}_i, \mathbf{d}_i)$ is the angle between \mathbf{p}_i and \mathbf{d} , η_i is Gaussian white noise, and the positive constants J^{pol} and D^{pol} are the strength of the alignment torque and of the rotational diffusivity, respectively. This defines a dynamics similar to an Ornstein–Uhlenbeck process and has been well studied in the context of active systems [98, 90].

There are multiple interesting choices for the definition of the aligning direction \mathbf{d}_i and there is as of today no clear microscopic or dynamical basis favoring one of them [72]. A simple choice that is consistent with the mechanism of *contact inhibition of locomotion* (CIL) [99] is to assume that the polarity aligns to the direction of the total force exerted on a cell’s interface, namely $\mathbf{d}_i = \mathbf{F}_i^{\text{int}}$. Such a coupling was shown to describe accurately patterns of sustained oscillations observed experimentally in systems of confined MDCK cells [86]. Other interesting possibilities are to assume that the polarity aligns either to the total velocity \mathbf{v}_i of each cell or to its main axis of elongation, and an extensive evaluation of these two choices is presented in [72, 93]. At high enough alignment strengths, the former case shows a Vicsek-type phase transition where all cells move in the same direction, while the latter exhibits unjamming to a liquid-like state [93]. Finally, the purely diffusive dynamics of the system ($J^{\text{pol}} = 0$) is interesting in its own right and it has been shown that densely packed cells can show “bursts” of motions as they quickly relax from their deformed shape in this case [85], while confluent monolayers exhibit a solid–liquid phase transition as D^{pol} is increased [89, 93]. Further research should investigate potential microscopic bases for such alignment dynamics of the polarization and put it on a firmer footing.

26.3.5 Active nematic driving

While polar driving has been extensively studied theoretically in the context of single cells, there is strong evidence that descriptions based on active nematic liquid crystals offer a compelling framework to understand the collective behavior of cell monolayers [42, 43, 88]. Nematic driving can be introduced following Ref. [88] by rewriting the interface force (Eq. (26.20)) as

$$\mathbf{F}_i^{\text{inter}} = \int d\mathbf{x} \phi_i \nabla \boldsymbol{\Pi}^{\text{tissue}} = - \int d\mathbf{x} \boldsymbol{\Pi}^{\text{tissue}} \nabla \phi_i, \quad (26.22)$$

where $\boldsymbol{\Pi}^{\text{tissue}}$ is a tissue stress tensor that can be written in the usual fashion as

$$\boldsymbol{\Pi}^{\text{tissue}} = -p \mathbb{1} - \zeta \mathbf{Q}. \quad (26.23)$$

The pressure $p = \delta(\mathcal{F}_{\text{rep}} - \mathcal{F}_{\text{CH}} - \mathcal{F}_{\text{area}})/\delta\phi_i$ can be read directly from Eq. (26.20), while $\mathbf{Q} = \sum_i \phi_i \mathbf{Q}_i$ is a newly introduced tissue nematic tensor written as a weighted sum over the contributions of the individual cells. Such a definition has two main advantages: (i) it bridges the gap between local (cell level) and global (tissue level) properties and (ii) allows for naturally introducing the usual active term $-\zeta \mathbf{Q}$ found in continuum theories of active liquid crystals [47, 33] (see Section 26.2.1 and Chapter 12). It can be interpreted as creating a dipolar force density distributed along the cells interfaces such that each cell pushes or pulls its neighbors depending on the direction of their contact area with respect to the nematic tensor (Fig. 26.7a).

The simplest choice for \mathbf{Q}_i is to introduce an internal nematic degree of freedom for each cell and define its dynamics similarly to the polar case. In the continuum theory, most of the phenomenology of active liquid crystals depends on a balance between the restoring elastic forces and the flow alignment of the nematic tensor. Writing $\mathbf{Q}_i = 2(\hat{n}_i^T \hat{n}_i - \mathbb{1}/2)$ with $\hat{n}_i = (\cos \theta_i^{\text{nem}}, \sin \theta_i^{\text{nem}})$, we can mirror these two components in our model by defining the following dynamics of the angle θ_i :

$$\partial_t \theta_i^{\text{nem.}} = K^{\text{nem.}} \tau_i + J^{\text{nem.}} \omega_i, \quad (26.24)$$

where the torques are given by

$$\tau_i = \frac{1}{\lambda} \int d\mathbf{x} \phi_i \mathbf{Q} \wedge \mathbf{Q}_i, \quad \omega_i = \int d\mathbf{x} \mathbf{v} \wedge \nabla \phi_i. \quad (26.25)$$

We have defined the tissue velocity as $\mathbf{v} = \sum_i \phi_i \mathbf{v}_i$ and $\mathbf{A} \wedge \mathbf{B} = A_{xx}B_{xy} - A_{xy}B_{xx}$ for symmetric traceless matrices \mathbf{A} and \mathbf{B} . The first torque τ_i aligns \mathbf{Q}_i to the tissue nematic tensor \mathbf{Q} and induces an elastic restoring force

favoring the homogeneous state. The second torque ω_i rotates \mathbf{Q}_i with the local vorticity computed as the integral of the neighboring cells velocity projected on the cell interface. Together with the active term, these torques are able to drive the cells out of equilibrium and reproduce the bend and splay instabilities observed in continuum theories of active nematic liquid crystals (Fig. 26.7b). In a confluent epithelium, one can also check that the total force is approximately zero and the system does not develop any total velocity under periodic boundary conditions. In particular this means the transition to collective movement is different from the Vicsek-type phase transitions observed with polar driving [90].

Another interesting possibility explored in Ref. [88] is to avoid introducing supplementary degrees of free-

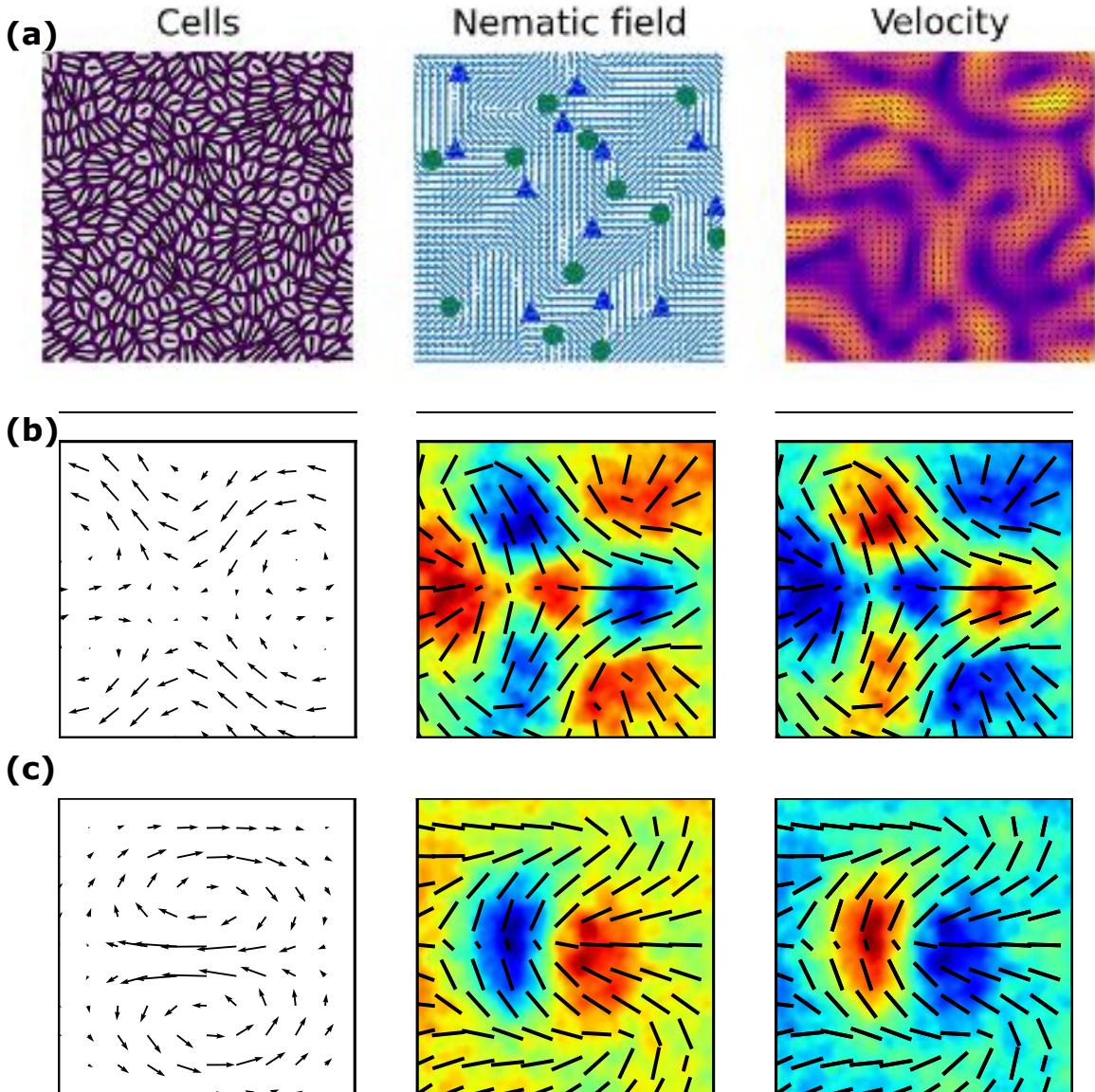


Figure 26.8: Emergent features in a phase field model of cell monolayers. **a** Emergence of active turbulence. Shown here are the cells in confluence, their associated coarse-grained nematic field, and the corresponding velocity field. The nematic field describes the local orientation of the cells defined as their elongation direction. The singularities in the orientation field are marked by green circles for $+1/2$ and blue triangles for $-1/2$ topological defects. The velocity field is colored by the normalized magnitude of velocity ranging from 0 (purple) to 1 (yellow). Average velocity field, isotropic stress, and pressure around $-1/2$ (b) and $+1/2$ topological defects (c). Figures adapted from Ref. [88].

dom altogether and define an active coupling that is directly proportional to the shape deformations of the individual cells. This is motivated by recent studies indicating that the local deformation of cells provides a suitable nematic order parameter that allows the description of the dynamics of epithelial cells using theories of active liquid crystals [2, 100, 42, 40, 43]. Such a connection is quite surprising because individual epithelial cells on a substrate are typically not elongated and have a well-defined direction of movement, suggesting polar rather than nematic driving. In this setting, the tissue nematic tensor is defined as $\mathbf{Q} = \sum_i \phi_i \mathbf{S}_i$, where \mathbf{S}_i is the deformation tensor of cell i given by

$$\mathbf{S}_i = - \int d\mathbf{x} \left((\nabla \phi_i)^\top \nabla \phi_i - \frac{1}{2} \text{tr} [(\nabla \phi_i)^\top \nabla \phi_i] \right). \quad (26.26)$$

Equation (26.26) defines a 2×2 matrix whose eigenvalues and eigenvectors describe the strength and orientation of the main deformation axes of each cell, see Refs. [88, 101] for details. This corresponds to a nematic tensor with order parameter and director given by the largest eigenvalue and its associated eigenvector. With this simple definition, the system shows an activity-driven transition to non-zero nematic order and flows for high enough activity strengths ζ . The spontaneous creation of defects in the coarse grained nematic field as well as observed patterns of flows and mechanical stresses around topological defects are accurately predicted, see Fig. 26.8. This points to a strong connection between the shape deformation of cells and their active behavior, irrespective of the corresponding microscopic mechanism.

Example 26.1: 1D implementation of the phase field model. To consolidate the basic features of the phase field model that are introduced in this section, let's us work through an exercise to implement a version of the above phase-field model that describes a single cell in a box in one dimension. An example solution that uses python and numpy can be found at <https://github.com/rhomu/cell-in-a-box>.

a. As a first step, we need to compute $\delta \mathcal{F} / \delta \phi_i$ explicitly. Show that

$$\begin{aligned} \frac{\delta \mathcal{F}_{\text{CH}}}{\delta \phi_i} &= \frac{\gamma}{\lambda} (8\phi_i(1-\phi_i)(1-2\phi_i) - 2\lambda^2 \delta \phi_i), \\ \frac{\delta \mathcal{F}_{\text{area}}}{\delta \phi_i} &= -\frac{4\mu}{\pi R^2} \left(1 - \frac{1}{\pi R^2} \int d\vec{x} \phi_i^2 \right), \\ \frac{\delta \mathcal{F}_{\text{rep}}}{\delta \phi_i} &= \frac{2\kappa}{\lambda} \sum_{j \neq i} \phi_j^2 \phi_i. \end{aligned}$$

Note that because we are in one dimension, we will need to modify slightly our expression for the area free energy $\mathcal{F}_{\text{area}}$ by replacing the target area πR^2 by a target length L .

b. We will implement the simulation using a simple finite difference scheme where we discretise fields $f(\vec{x})$ at equally spaced lattice sites $f_i = f(\vec{x}_i)$, where $x_i = i \cdot h$. Note that the spacing h is arbitrary and can be set to be unity in the simulation for simplicity. From the above expression for the derivative of the free energy, we see that we will need to compute first and second derivatives of the phase field. Show that the following finite difference coefficients approximate the first and second derivatives:

$$\begin{aligned} \frac{f_{i-2} - 8f_{i-1} + 8f_{i+1} - f_{i+2}}{12h} &= f'_i + \mathcal{O}(h^4), \\ \frac{-f_{i-2} + 16f_{i-1} - 30f_i + 16f_{i+1} - f_{i+2}}{12h^2} &= f''_i + \mathcal{O}(h^4). \end{aligned}$$

c. At first, let us neglect all forces and implement Eq. (26.15) without the advection term. Considering a single cell and writing the time derivative as a time difference, we get

$$\phi|_{t+\Delta t} - \phi|_t = -\frac{\delta \mathcal{F}}{\delta \phi} \Delta t.$$

Implement the above evolution on a finite domain of size 300 with periodic boundary conditions and use as initial conditions $\phi(x_i)|_{t=0} = 1$ if $i \in [100, 200]$ and $\phi(x_i)|_{t=0} = 0$ otherwise. Periodic boundary conditions can be easily implemented introducing two supplementary lattice sites on both sides of the

domain and set them after each time increment to be $\phi(x_{-2}) = \phi(x_{297})$, $\phi(x_{-1}) = \phi(x_{299})$, $\phi(x_{300}) = \phi(x_0)$, $\phi(x_{301}) = \phi(x_1)$ when using zero-indexing. Note that the finite-difference derivatives can be easily implemented using one-dimensional convolutions which are widely available in numerical packages. Find a range of parameters such that the cell relaxes to its equilibrium shape and observe that its profile takes a hyperbolic tangent shape. The shape of the profile can be tuned using the parameters γ and λ .

d. As a final step, let us implement active driving, forces, and walls. This part is optional and we will leave out most of the details but please refer to the example solution if needed. Cell polarization can be introduced as explained in the above sections, with the difference that in one dimension polarity becomes a scalar quantity, which means that Eq. (26.21) cannot be used in this case. Instead one could introduce the following coupling between the polarity and the interaction force $F^{int.}$,

$$\partial_t p = J|F^{int.}|(F^{int.} - p) + (2D)^{\frac{1}{2}}\eta,$$

which can be implemented in finite time form using the Euler-Maruyama scheme. The instantaneous velocity of the cell can then be obtained as $v = (\alpha p + F^{int.})/\xi$ and the advection term can be reintroduced in Eq. (26.15). Finally, in order to generate non-trivial forces one can add walls at the boundary of the domain by introducing another phase field ϕ_{walls} that does not evolve with time but contributes to the repulsion free energy \mathcal{F}_{rep} . Putting everything together, find a parameter regime where the simulation is numerically stable and observe how the cell moves in the box and interacts with the walls. Note that when forces are present, it is usually good to start the simulation with a warm up phase where all forces are disabled which allows the cell to relax to its equilibrium phase without creating numerical instabilities.

26.4 Summary

We have presented an introduction to phase field modeling of active systems, providing examples from both continuum biphasic active materials and agent-based models of cellular monolayers. A general framework that allows the introduction of various forms of coupling to the phase field was developed. In particular, it was shown how hydrodynamic effects can be included in phase field models and how tensor order parameters such as active particles orientation field or polymer conformation (describing viscoelastic effects) can be coupled to the equations of motions of the phase field in order to construct complex models of active interfaces. Furthermore, we have described how using multiple phase fields can be used to construct a versatile model of cellular monolayers, where each cell is described as an active deformable droplet using an individual phase field. Applications of this approach to modeling collective cell motion and capturing cell-cell interaction forces were described.

As we gain more knowledge of complex spatiotemporal features of active systems, understanding the dynamics of physical interfaces between different active materials as well as between active matter and its surrounding media becomes more and more important. Such understanding will be of prime importance to tackle biophysical problems such as bacterial biofilm development, cellular invasion, and morphogenesis. Building versatile and predictive computational models of active interfaces will be, together with experimental advances, the determining factor in advancing our understanding of such systems, both by helping to explain the mechanisms behind experimental observations and, even more importantly, by providing predictive tools for experimentally inaccessible scenarios.

26.5 Problems

Problem 26.1: Band of active nematic within an isotropic fluid. In this problem you will build on the code you used in Chapter 24 to model a two dimensional active nematics, and on Example 26.1, a finite difference implementation of phase field.

- a.** Numerically setup a band of active nematic within an isotropic fluid to provide a flat interface between an active phase and a passive phase, also using the simple modification introduced in Eq. (26.12).
- b.** By varying the activity parameter, and providing representative snapshots from the simulations, describe qualitatively the physics that you observe and the difference between the above activity levels.

Problem 26.2: Average interface profile. From the simulations in Problem 26.1, extract the average interface profile numerically, estimate the interface width and surface tension, and plot the estimated values as a function of activity.

Problem 26.3: Average angle of the nematic directors with respect to the interface. From the simulations in Problem 26.1, numerically calculate the average angle (averaged over interface width and over the last 100 simulation time frames) that nematic directors make with respect to the interface (0 corresponds to the nematic director being completely aligned parallel to the interface). Plot the distribution of angles across the interface (averaged over the last 100 time frames) for different values of activity. How does this fit with the *active anchoring* phenomenon [26, 52]?

Problem 26.4: Interface instabilities as a function of activity. From the simulations in Problem 26.1, numerically measure the wavelength and the growth rate of interface instabilities as a function of activity. How does this compare with the theoretical predictions of hydrodynamic instabilities in bulk active nematics [102]?

Bibliography

- [1] O. J. Meacock, A. Doostmohammadi, K. R. Foster, J. M. Yeomans, and W. M Durham. Bacteria solve the problem of crowding by moving slowly. *Nature Physics*, pages 1–6, 2020.
- [2] A. Doostmohammadi, S. P. Thampi, T. B. Saw, C. T. Lim, B. Ladoux, and J. M. Yeomans. Celebrating Soft Matter’s 10th Anniversary: Cell division: a source of active stress in cellular monolayers. *Soft Matter*, 11:7328–7336, 2015.
- [3] M. Driscoll, B. Delmotte, M. Youssef, S. Sacanna, A. Donev, and P. Chaikin. Unstable fronts and motile structures formed by microrollers. *Nature Physics*, 13:375–379, 2017.
- [4] F. C. Keber, E. Loiseau, T. Sanchez, S. J. DeCamp, L. Giomi, M. J. Bowick, M. C. Marchetti, Z. Dogic, and A. R. Bausch. Topology and dynamics of active nematic vesicles. *Science*, 345:1135–1139, 2014.
- [5] M. C. Marchetti, J.-F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85:1143, 2013.
- [6] D. Needleman and Z. Dogic. Active matter at the interface between materials science and cell biology. *Nature Reviews Materials*, 2:1–14, 2017.
- [7] M. Burmølle, J. S. Webb, D. Rao, L. H. Hansen, S. J. Sørensen, and S. Kjelleberg. Enhanced biofilm formation and increased resistance to antimicrobial agents and bacterial invasion are caused by synergistic interactions in multispecies biofilms. *Applied and environmental microbiology*, 72:3916–3923, 2006.
- [8] C. D. Nadell, K. Drescher, N. S. Wingreen, and B. L. Bassler. Extracellular matrix structure governs invasion resistance in bacterial biofilms. *The ISME journal*, 9:1700–1709, 2015.
- [9] A. Brugués, E. Anon, V. Conte, J. H. Veldhuis, M. Gupta, J. Colombelli, J. J. Muñoz, G. W. Brodland, B. Ladoux, and X. Trepat. Forces driving epithelial wound healing. *Nature Physics*, 10:684, 2014.
- [10] P. Friedl and D. Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology*, 10:445–457, 2009.
- [11] Peter Friedl, Joseph Locker, Erik Sahai, and Jeffrey E Segall. Classifying collective cancer cell invasion. *Nature Cell Biology*, 14(8):777, 2012.
- [12] J. Zhang, R. Alert, J. Yan, N. S. Wingreen, and S. Granick. Active phase separation by turning towards regions of higher density. *Nature Physics*, 17:961–967, 2021.

- [13] L. Balasubramaniam, A. Doostmohammadi, T. B. Saw, G. Sankara, R. Mueller, T. Dang, M. Thomas, S. Gupta, S. Sonam, Y. Toyama, R.-M. Mège, J. M. Yeomans, and B. Ladoux. Investigating the nature of active forces in tissues reveals how contractile cells can form extensile monolayers. *Nature*, 20:1156–1166, 2021.
- [14] O. Chaudhuri, J. Cooper-White, P. A. Janmey, D. J. Mooney, and V. B. Shenoy. Effects of extracellular matrix viscoelasticity on cellular behaviour. *Nature*, 584:535–546, 2020.
- [15] P. C. Hohenberg and B. I. Halperin. Theory of dynamic critical phenomena. *Reviews of Modern Physics*, 49:435–479, 1977.
- [16] G. J. Fix. Phase field methods for free boundary problems. 1982.
- [17] J. S. Langer. *Models of pattern formation in first-order phase transitions*, pages 165–186. World Scientific, 1986.
- [18] E. Emmerich. Advances of and by phase-field modelling in condensed-matter physics. *Advances in Physics*, 57:1–87, 2008.
- [19] S. M. Allen and J. W. Cahn. Ground state structures in ordered binary alloys with second neighbor interactions. *Acta Metallurgica*, 20:423 – 433, 1972.
- [20] J. W. Cahn and J. E. Hilliard. Free energy of a nonuniform system. I. Interfacial free energy. *The Journal of Chemical Physics*, 28:258–267, 1958.
- [21] D. Jacqmin. Calculation of two-phase Navier–Stokes flows using phase-field modeling. *Journal of Computational Physics*, 155:96–127, 1999.
- [22] P. Yue, J. J. Feng, C. Liu, and J. Shen. A diffuse-interface method for simulating two-phase flows of complex fluids. *Journal of Fluid Mechanics*, 515:293, 2004.
- [23] R. Mauri, R. Shinnar, and G. Triantafyllou. Spinodal decomposition in binary mixtures. *Physical Review E*, 53:2613–2623, 1996.
- [24] F.-H. Lin and C. Liu. Nonparabolic dissipative systems modeling the flow of liquid crystals. *Communications on Pure and Applied Mathematics*, 48:501–537, 1995.
- [25] N. Sulaiman, D. Marenduzzo, and J. M. Yeomans. Lattice Boltzmann algorithm to simulate isotropic-nematic emulsions. *Physical Review E*, 74:041708, 2006.
- [26] M. L. Blow, S. P. Thampi, and J. M. Yeomans. Biphasic, lyotropic, active nematics. *Physical Review Letters*, 113:248303, 2014.
- [27] M. E. Cates and E. Tjhung. Theories of binary fluid mixtures: from phase-separation kinetics to active emulsions. *Journal of Fluid Mechanics*, 836:P1, 2018.
- [28] F. Kempf, R. Mueller, E. Frey, J. M. Yeomans, and A. Doostmohammadi. Active matter invasion. *Soft Matter*, 15:7538–7546, 2019.
- [29] R. J. LeVeque. Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems. *SIAM*, 2007.
- [30] D. Marenduzzo, E. Orlandini, M. E. Cates, and J. M. Yeomans. Steady-state hydrodynamic instabilities of active liquid crystals: Hybrid lattice Boltzmann simulations. *Physical Review E*, 76:031921, 2007.
- [31] J.-C. Desplat, I. Pagonabarraga, and P. Bladon. Ludwig: A parallel lattice-Boltzmann code for complex fluids. *Computer Physics Communications*, 134:273–290, 2001.
- [32] F. Jülicher, S. W. Grill, and G. Salbreux. Hydrodynamic theory of active matter. *Reports on Progress in Physics*, 81:076601, 2018.

- [33] A. Doostmohammadi, J. Ignés-Mullol, J. M. Yeomans, and F. Sagués. Active nematics. *Nature Communications*, 9:1–13, 2018.
- [34] L. Giomi and A. DeSimone. Spontaneous division and motility in active nematic droplets. *Physical Review Letters*, 112:147802, 2014.
- [35] T. Turiv, R. Koizumi, K. Thijssen, M. M. Genkin, H. Yu, C. Peng, Q.-H. Wei, J. M. Yeomans, I. S. Aranson, A. Doostmohammadi, and O. D. Lavrentovich. Polar jets of swimming bacteria condensed by a patterned liquid crystal. *Nature Physics*, 16:481–487, 2020.
- [36] D. Volfson, S. Cookson, J. Hasty, and L. S. Tsimring. Biomechanical ordering of dense cell populations. *Proceedings of the National Academy of Sciences*, 105:15346–15351, 2008.
- [37] T. Sanchez, D. T. N. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic. Spontaneous motion in hierarchically assembled active matter. *Nature*, 491:431, 2012.
- [38] P. Guillamat, J. Ignés-Mullol, and F. Sagués. Control of active liquid crystals with a magnetic field. *Proceedings of the National Academy of Sciences*, 113:5498–5502, 2016.
- [39] N. Kumar, R. Zhang, J. J. De Pablo, and M. L. Gardel. Tunable structure and dynamics of active liquid crystals. *Science Advances*, 4:eaat7779, 2018.
- [40] G. Duclos, C. Blanch-Mercader, V. Yashunsky, G. Salbreux, J.-F. Joanny, J. Prost, and P. Silberzan. Spontaneous shear flow in confined cellular nematics. *Nature Physics*, 14:728–732, 2018.
- [41] K. Kawaguchi, R. Kageyama, and M. Sano. Topological defects control collective dynamics in neural progenitor cell cultures. *Nature*, 545:327–331, 2017.
- [42] T. B. Saw, A. Doostmohammadi, V. Nier, L. Kocgozlu, S. Thampi, Y. Toyama, P. Marcq, C. T. Lim, J. M. Yeomans, and B. Ladoux. Topological defects in epithelia govern cell death and extrusion. *Nature*, 544:212–216, 2017.
- [43] C. Blanch-Mercader, V. Yashunsky, S. Garcia, G. Duclos, L. Giomi, and P. Silberzan. Turbulent dynamics of epithelial cell cultures. *Physical Review Letters*, 120:208101, 2018.
- [44] V. Yashunsky, D. J. G. Pearce, C. Blanch-Mercader, F. Ascione, P. Silberzan, and L. Giomi. Chiral edge currents in nematic cell monolayers. *Physical Review X*, 12:041017, 2022.
- [45] R. Voituriez, J.-F. Joanny, and J. Prost. Spontaneous flow transition in active polar gels. *Europhysics Letters (EPL)*, 70:404, 2005.
- [46] S. Ramaswamy. The mechanics and statistics of active matter. *Annual Reviews in Condensed Matter Physics*, 1:323–345, 2010.
- [47] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao, and R. A. Simha. Hydrodynamics of soft active matter. *Reviews of Modern Physics*, 85:1143–1189, 2013.
- [48] A. E. Pateson, A. Gopinath, and P. E. Arratia. The propagation of active-passive interfaces in bacterial swarms. *Nature Communications*, 9:1–10, 2018.
- [49] R. Zhang, S. A. Redford, P. V. Ruijgrok, N. Kumar, A. Mozaffari, S. Zemsky, A. R. Dinner, V. Vitelli, Z. Bryant, M. L. Gardel, and J. J. de Pablo. Spatiotemporal control of liquid crystal structure and dynamics through activity patterning. *Nature Materials*, 20:875–882, 2021.
- [50] P.-G. De Gennes and J. Prost. *The physics of liquid crystals*, volume 83. Oxford university press, 1993.
- [51] A. N. Beris and B. J. Edwards. *Thermodynamics of Flowing Systems*. Oxford University Press, 1994.
- [52] A. Doostmohammadi, S. P. Thampi, and J. M. Yeomans. Defect-mediated morphologies in growing cell colonies. *Physical Review Letters*, 117:048102, 2016.

- [53] M. L. Blow, M. Aqil, B. Liebchen, and D. Marenduzzo. Motility of active nematic films driven by “active anchoring”. *Soft Matter*, 13:6137–6144, 2017.
- [54] L. N. Carenza, G. Gonnella, D. Marenduzzo, and G. Negro. Rotation and propulsion in 3d active chiral droplets. *Proceedings of the National Academy of Sciences*, 116:22065–22070, 2019.
- [55] R. C. V. Coelho, N. A. M. Araújo, and M. M. Telo da Gama. Propagation of active nematic–isotropic interfaces on substrates. *Soft Matter*, 16:4256–4266, 2020.
- [56] P. Neta, M. Tasinkevych, M. Telo da Gama, and C. Dias. Wetting of a solid surface by active matter. *Soft Matter*, 17:2468–2478, 2021.
- [57] D. Dell’Arciprete, M. L. Blow, A. T. Brown, F. D. C. Farrell, J. S. Lintuvuori, A. F. McVey, D. Marenduzzo, and W. C. K. Poon. A growing bacterial colony in two dimensions as an active nematic. *Nature communications*, 9:1–9, 2018.
- [58] L. C. Emmanuel, J. M. Yeomans, and A. Doostmohammadi. Active matter in a viscoelastic environment. *Physical Review Fluids*, 5:023102, 2020.
- [59] E. J. Hemingway, A. Maitra, S. Banerjee, M. C. Marchetti, S. Ramaswamy, S. M. Fielding, and M. E. Cates. Active viscoelastic matter: from bacterial drag reduction to turbulent solids. *Physical review letters*, 114:098302, 2015.
- [60] L. Metselaar, J. M. Yeomans, and A. Doostmohammadi. Topology and morphology of self-deforming active shells. *Physical Review Letters*, 123(20):208001, 2019.
- [61] Y. Maroudas-Sacks, L. Garion, L. Shani-Zerbib, A. Livshits, E. Braun, and K. Keren. Topological defects in the nematic order of actin fibers as organization centers of *Hydra* morphogenesis. *Nature Physics*, 17:251–259, 2021.
- [62] S. R. Naganathan, S. Fürthauer, M. Nishikawa, F. Jülicher, and S. W. Grill. Active torque generation by the actomyosin cell cortex drives left–right symmetry breaking. *Elife*, 3:e04165, 2014.
- [63] R. Zhang, Y. Zhou, M. Rahimi, and J. J. De Pablo. Dynamic structure of active nematic shells. *Nature communications*, 7:1–9, 2016.
- [64] P. Guillamat, Ž. Kos, J. Hardouin, J. Ignés-Mullol, M. Ravnik, and F. Sagués. Active nematic emulsions. *Science advances*, 4:eaao1470, 2018.
- [65] R. Wittkowski, A. Tiribocchi, J. Stenhammar, R. J. Allen, D. Marenduzzo, and M. E. Cates. Scalar φ field theory for active-particle phase separation. *Nature Communications*, 5:1–9, 2014.
- [66] G. Negro, L. N. Carenza, A. Lamura, A. Tiribocchi, and G. Gonnella. Rheology of active polar emulsions: from linear to unidirectional and inviscid flow, and intermittent viscosity. *Soft Matter*, 15(41):8251–8265, 2019.
- [67] L. N. Carenza, G. Gonnella, A. Lamura, G. Negro, and A. Tiribocchi. Lattice boltzmann methods and active fluids. *The European Physical Journal E*, 42:81, 2019.
- [68] F. Ziebert, S. Swaminathan, and I. S. Aranson. Model for self-polarization and motility of keratocyte fragments. *Journal of the Royal Society Interface*, 9:1084–1092, 2011.
- [69] E. Tjhung, D. Marenduzzo, and M. E. Cates. Spontaneous symmetry breaking in active droplets provides a generic route to motility. *Proceedings of the National Academy of Sciences*, 109:12381–12386, 2012.
- [70] W. Marth and A. Voigt. Collective migration under hydrodynamic interactions: a computational approach. *Interface Focus*, 6:20160037, 2016.
- [71] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning. A density-independent rigidity transition in biological tissues. *Nature Physics*, 11:1074–1079, 2015.

- [72] R. Alert and X. Trepaut. Physical models of collective cell migration. *Annual Review of Condensed Matter Physics*, 11:77–101, 2020.
- [73] F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Physical Review Letters*, 69:2013–2016, 1992.
- [74] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17:2095–2104, 2007.
- [75] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning. Motility-driven glass and jamming transitions in biological tissues. *Physical Review X*, 6:021011, 2016.
- [76] D. M. Sussman, M. Paoluzzi, M. C. Marchetti, and M. L. Manning. Anomalous glassy dynamics in simple models of dense biological tissue. *EPL (Europhysics Letters)*, 121:36001, 2018.
- [77] J. Prost, F. Jülicher, and J.-F. Joanny. Active gel physics. *Nature Physics*, 11:111–117, 2015.
- [78] J. Löber, F. Ziebert, and I. S. Aranson. Collisions of deformable cells lead to collective migration. *Scientific Reports*, 5:9172, 2015.
- [79] B. A. Camley and W.-J. Rappel. Physical models of collective cell motility: from cell to tissue. *Journal of Physics D: Applied Physics*, 50:113002, 2017.
- [80] M. R. Shaebani, A. Wysocki, R. G. Winkler, G. Gompper, and H. Rieger. Computational models for active matter. *Nature Reviews Physics*, 2:1–19, 2020.
- [81] J. Löber, F. Ziebert, and I. S. Aranson. Modeling crawling cell movement on soft engineered substrates. *Soft Matter*, 10:1365–1373, 2014.
- [82] F. Ziebert and I. S. Aranson. Effects of adhesion dynamics and substrate compliance on the shape and motility of crawling cells. *PLoS one*, 8:e64511, 2013.
- [83] S. Najem and M. Grant. Phase-field model for collective cell migration. *Physical Review E*, 93:052405, 2016.
- [84] B. A Camley, Y. Zhang, Y. Zhao, B. Li, E. Ben-Jacob, H. Levine, and W.-J. Rappel. Polarity mechanisms such as contact inhibition of locomotion regulate persistent rotational motion of mammalian cells on micropatterns. *Proceedings of the National Academy of Sciences*, 111:14770–14775, 2014.
- [85] B. Palmieri, Y. Bresler, D. Wirtz, and M. Grant. Multiple scale model for cell migration in monolayers: Elastic mismatch between cells enhances motility. *Scientific Reports*, 5:11745, 2015.
- [86] G. Peyret, R. Mueller, J. d’Alessandro, S. Begnaud, P. Marcq, R.-M. Mège, J. M. Yeomans, A. Doostmohammadi, and B. Ladoux. Sustained oscillations of epithelial cell sheets. *Biophysical Journal*, 117:464–478, 2019.
- [87] D. Wenzel, S. Praetorius, and A. Voigt. Topological and geometrical quantities in active cellular structures. *The Journal of Chemical Physics*, 150:164108, 2019.
- [88] R. Mueller, J. M. Yeomans, and A. Doostmohammadi. Emergence of active nematic behavior in monolayers of isotropic cells. *Physical Review Letters*, 122:048004, 2019.
- [89] B. Loewe, M. Chiang, D. Marenduzzo, and M. C. Marchetti. Solid-liquid transition of deformable and overlapping active particles. *Physical Review Letters*, 125:038003, 2020.
- [90] F. Giavazzi, M. Paoluzzi, M. Macchi, D. Bi, G. Scita, M. L. Manning, R. Cerbino, and M. C. Marchetti. Flocking transitions in confluent tissues. *Soft Matter*, 14:3471–3477, 2018.
- [91] J. H. Kim, X. Serra-Picamal, D. T. Tambe, E. H. Zhou, C. Y. Park, M. Sadati, J.-A. Park, R. Krishnan, B. Gweon, E. Millet, X. Trepaut J. P. Butler, and J. J. Fredberg. Propulsion and navigation within the advancing monolayer sheet. *Nature Materials*, 12:856–863, 2013.

- [92] M. Deforet, V. Hakim, H. G. Yevick, G. Duclos, and P. Silberzan. Emergence of collective modes and tri-dimensional structures from epithelial confinement. *Nature Communications*, 5:3747, 2014.
- [93] G. Zhang, R. Mueller, A. Doostmohammadi, and J. M. Yeomans. Active inter-cellular forces in collective cell motility. *Journal of The Royal Society Interface*, 17:20200312, 2020.
- [94] M. Nonomura. Study on multicellular systems using a phase field model. *PLOS ONE*, 7:1–9, 2012.
- [95] B. R. Sarangi, M. Gupta, B. L. Doss, N. Tissot, F. Lam, R.-M. Mège, N. Borghi, and B. Ladoux. Coordination between intra-and extracellular forces regulates focal adhesion dynamics. *Nano Letters*, 17:399–406, 2017.
- [96] T. J. Mitchison and L. P. Cramer. Actin-based cell motility and cell locomotion. *Cell*, 84:371–379, 1996.
- [97] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York, 6th edition, 2015.
- [98] B. Szabó, G. J. Szöllösi, B. Gönci, Zs. Jurányi, D. Selmeczi, and T. Vicsek. Phase transition in the collective migration of tissue cells: Experiment and model. *Physical Review E*, 74:061908, 2006.
- [99] B. Smeets, R. Alert, J. Pešek, I. Pagonabarraga, H. Ramon, and R. Vincent. Emergent structures and dynamics of cell colonies by contact inhibition of locomotion. *Proceedings of the National Academy of Sciences*, 113:14621–14626, 2016.
- [100] G. Duclos, C. Erlenkämper, J.-F. Joanny, and P. Silberzan. Topological defects in confined populations of spindle-shaped cells. *Nature Physics*, 13:58–62, 2017.
- [101] M. Asipauskas, M. Aubouy, J. A. Glazier, F. Graner, and Y. Jiang. A texture tensor to quantify deformations: the example of two-dimensional flowing foams. *Granular Matter*, 5:71–74, 2003.
- [102] R. A. Simha and S. Ramaswamy. Hydrodynamic fluctuations and instabilities in ordered suspensions of self-propelled particles. *Physical Review Letters*, 89:058101, 2002.

Part III

Experiments

Chapter 27

Chemistry of Micromotors

JULIANE SIMMCHEN

Living systems constantly dissipate energy, thus avoid to reach an equilibrium with their environment. Reaching an equilibrium indeed means death for a living organism. The energy required to fuel life generally comes from metabolic activity, i.e., the whole set of chemical reactions that is required to sustain all functions of an organism, be it to supply energy, to provide building blocks for the body, or to eliminate waste products. These chemical reactions differ strongly between species and environments, and I refer the interested reader to designated literature for further details on this [1, 2]. All living entities use the energy that they gained from their metabolism for different processes, including their own displacement. Motion can have different functions, such as foraging for food, shelter, or mates; few living creatures do not move at different stages of their life cycle. Biological microscopic entities such as algae, bacteria, or spermatozoa move mostly by body deformations. Nature has developed different strategies that typically lead to high swimming speeds and efficient locomotion: for ciliates, an effective stroke followed by a recovery stroke makes for an asymmetric

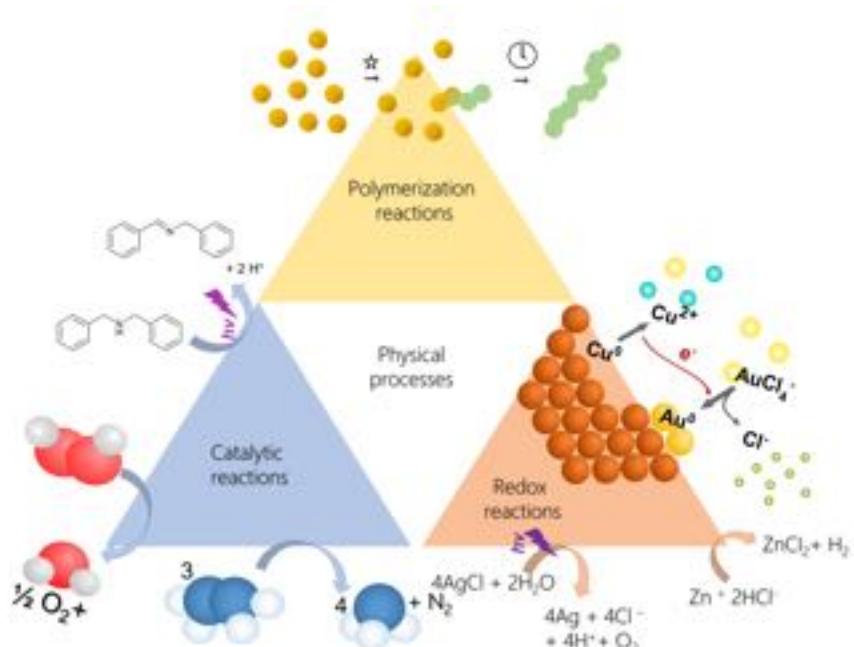


Figure 27.1: **Reactions for synthetic micromotors.** When active propulsion stems from chemical reactions, we can divide the previously presented examples in three main groups: catalytic reactions, polymerization reactions, and redox reactions.

pattern that generates a net thrust on the cell's body; prokaryotic flagella are generally rotating relative to the cell body — resembling a cork screw; finally, the actively bending eukaryotic flagella can move either through planar or helical waves. Even though synthetic systems are becoming more sophisticated and increasingly gain functionalities and flexibility, efficient body-deforming synthetic swimmers are still rare. Most synthetic micromotors are driven by hydrodynamic interactions often directly or indirectly induced by a chemical gradient. In this Chapter, I will introduce general chemistry concepts useful to describe and understand the motion of many synthetic micromotors. To keep this chapter concise, I will not discuss active matter driven by external fields.

27.1 Basic physical chemistry of micromotors

In general, micromotors are energy dissipating systems, that, just like living systems, constantly transform one form of energy into another in order to keep certain processes running. The famous textbook *Atkins Physical Chemistry* [3] defines energy as *the ability to do work or produce heat*, specifying that many processes depend on the transmission or transformation of energy. Here, it becomes obvious that the basic laws of thermodynamics will form a general framework for all considerations, namely:

1. The *law of conservation of energy* states that energy can only be transformed between states, neither created nor destroyed, i.e., an isolated system has a constant energy, so that

$$\Delta U = Q - W \quad (27.1)$$

with ΔU being the change in internal energy of the system, Q the quantity of energy supplied to the system as heat, and W the work the system can do.

2. The *second law of thermodynamics* has many expressions, made by early physicists such as Carnot, Planck, Lord Kelvin, or Clausius. At its core, it describes the understanding that all processes can be divided into reversible and irreversible processes. To do so, a state variable called entropy — often referred to as the statistical disorder of the system — is used. The change in entropy is equal to the heat transfer divided by the temperature T , or

$$\Delta S = \Delta Q/T. \quad (27.2)$$

In case of a reversible process, the combined entropy of both system and environment remains constant, while it increases for an irreversible system. This holds true for all processes, whether they restrict to physical transformations (e.g., melting or mixing) or include chemical reactions. Since a multitude of such processes have been used to propel microobjects, we want to consider them individually.

27.1.1 Single-component phase transitions

A phase within a single-component material is characterized by uniform physical properties. Phase transitions in single component materials are one of the simpler answers to environmental changes (such as pressure or temperature) and are typically characterized using phase diagrams. These typically display the conversion and coexistence of different phases (e.g., solid, liquid and vapor) depending on temperature (x -axis) and pressure (y -axis). I am unaware of any such phase transitions being directly exploited for active matter propulsion.

Traditionally, most micromotor bodies or active colloids are made out of solid phase materials. Even within a single state of matter, such as the solid phase, different phases can co-exist. An example frequently encountered in micromotor research are anatase and rutile phases that are formed via calcination of amorphous TiO_2 and have important consequences for the catalytic properties of the micromotors. Recent exceptions to solid bodies are liquid metal based micromotors [4, 5] and active droplets, both sharing the common advantage of an increased deformability with respect to their solid micromotor counterparts.

Example 27.1: Phase transitions within a single-component material. Phase transitions within a single material are generally displayed in diagrams with *pressure* (P) and *temperature* (T) on the axes, which describe how both, pressure and temperature, influence the state of matter.

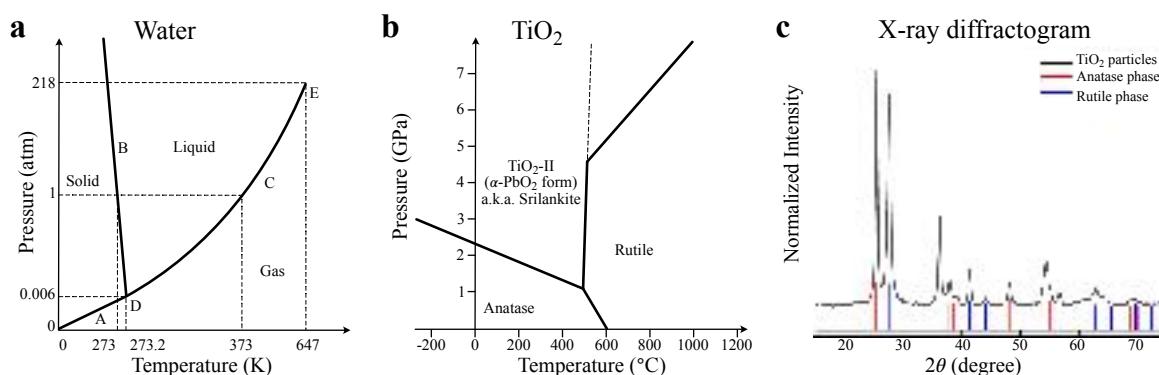


Figure 27.2: Single-component materials and their phase diagrams. (a) Phase diagram of water. (b) Phase diagram of titanium oxide (TiO_2) and (c) X-ray diffractogram (XRD) of TiO_2 particles and of the two different phases (Anatase and Rutile).

a. Traditionally, phase diagrams consider the transition between different phases, e.g., melting or evaporation of H_2O . The increase of T per se only affects the phase state if it reaches a point or a line in the diagram. See, for example, the phase diagram of H_2O in Fig. 27.2a.

b. Different phases within a single material also influence catalytic properties and thereby swimming behavior (see TiO_2). See, for example, the phase diagram of TiO_2 in Fig. 27.2b.

27.1.2 Binary mixtures

Mixtures are materials that consist of more than one component — still without including chemical reactions. To describe simple, ideal mixtures from a thermodynamic perspective our best tool is the chemical potential μ , or the partial molar Gibbs energy. The difference of the Gibbs energies $\Delta_M G$ before and after mixing components A and B is the free mixing energy:

$$\Delta_M G = n \bar{R} T (x_A \ln x_A + x_B \ln x_B), \quad (27.3)$$

where $n = n_A + n_B$ is the total number of moles, $\bar{R} = 8.314 \text{ J K}^{-1} \text{ mol}^{-1}$ is the universal gas constant, T is the temperature, and x_A , x_B are the molar fractions n_A/n , n_B/n of the component A and B of the mixture. The driving force of mixing is generally the gain in entropy. This relation holds, when we consider ideal mixtures, i.e., when both types of molecules present in the mixture are rather similar in their chemical structures and subsequently also in their general behaviors and mix without restrictions. An example would be the mixture of toluene and benzene, whose formula and the corresponding liquid vapour equilibrium diagram can be seen in Fig. 27.3. The molecular structures resemble each other strongly, so that both components are miscible over the whole range of compositions.

For artificial micromotors, a different type of binary mixtures is of far higher significance: *partially miscible binary mixtures*. Here, the molecular structure is so different that certain conditions have to apply for the molecules to be well-mixed together. Often, such influences are temperature dependent: for small amounts of B in a different liquid A, the B molecules manage to distribute in the A phase, resulting in a homogeneous mixture. Upon further addition of B, the mixture reaches a composition, where it is thermodynamically more beneficial to separate into two distinct phases, one richer in A and one richer in B. Statistically, we encounter an upper critical mixing temperature more frequently, because the thermal movement of the molecules helps overcome the adversity between different molecules.

For micromotors, one system that presents a lower critical mixing temperature is of particular interest: water and 2,6-lutidine [7]. This kind of system is rather rare, because the components are miscible below a certain temperature, but after reaching a critical temperature the mixture demixes into two separate phases. The reason for such behaviors lies usually in highly specific chemical interactions (e.g., complex formation) that are temperature dependent and increase the miscibility. This principle has been used first by Volpe *et al.* [7] by suspending Janus particles in mixture of water and 2,6-lutidine which is maintained just below the

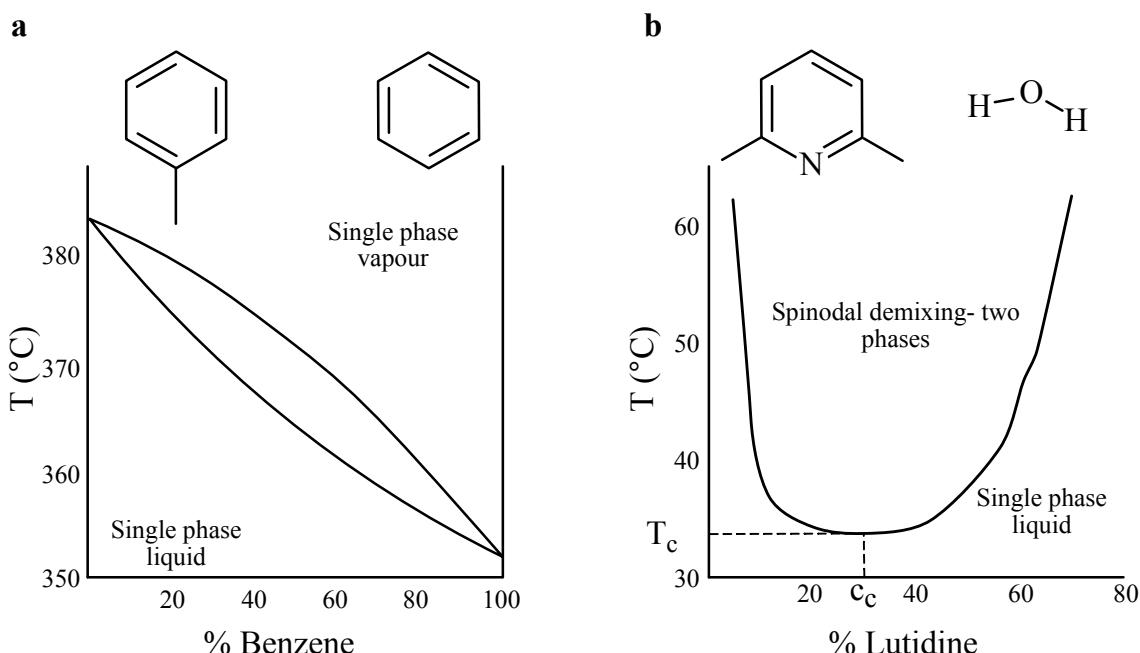


Figure 27.3: **Temperature-composition phase diagrams of binary mixtures.** (a) Phase diagram of the binary toluene–benzene mixture. Reproduced from Ref. [3]. (b) Phase diagram of the binary mixture water–2,6-lutidine showing the fully miscible and the demixing areas, separated by a lower critical mixing temperature. Reproduced from Ref. [6].

critical temperature. In this state the entire system is irradiated with light at a wavelength of 532 nm, which is preferentially absorbed by the gold cap of the Janus particles and leads to very local heating. This is followed by a local spinodal decomposition and demixing of the binary solution, which finally leads to a harvestable energy that can propel the particles.

Example 27.2: Critical concentration of a mixture of water and 2,6-lutidine. The lutidine (2,6-dimethyldpyridine)–water mixture shows a critical lower solution temperature of 34.1°C for a mixture of 28 wt% lutidine and 72 wt% water. Considering the densities of $\rho_{\text{lutidine}} = 925 \text{ kg m}^{-3}$ and $\rho_{\text{water}} = 1000 \text{ kg m}^{-3}$, that results in a volume ratio lutidine : water of 302.7 : 720. Below 34.1°C, any proportion of the mixture results in a single phase, while above that temperature we obtain two phases with different compositions according to the position on the binodal line.

27.1.3 Chemical reactions

IUPAC defines a chemical reaction as *a process that results in the interconversion of chemical species* [8], which generally leads the reactive mixture to a state of minimal Gibbs energy. This parameter is critical to evaluate any chemical reaction and to calculate equilibrium constants and can be used as an indicator of the energy available from a certain reaction. Any such process proceeds in the direction that minimizes the Gibbs free energy until the standard reaction Gibbs energy $\Delta_R G = 0$, i.e., the equilibrium state. This quantity is defined as

$$\Delta_R G = \sum v G_m(\text{products}) - \sum v G_m(\text{reactants}) \quad (27.4)$$

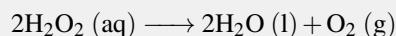
or the difference in standard molar Gibbs energies G_m weighted by the stoichiometric coefficients v . Usually, the standard molar Gibbs energies are not known, but we can calculate $\Delta_R G$ from the standard reaction enthalpy $\Delta_R H^\ominus$ and the standard reaction entropy $\Delta_R S^\ominus$:

$$\Delta_R G^\ominus = \Delta_R H^\ominus - T \Delta_R S^\ominus \quad (27.5)$$

For mere completeness, we briefly want to consider what these quantities are: the standard enthalpy $\Delta_R H^\ominus$, or also the standard heat content of a compound, is accessible through the measurable (or calculable) standard enthalpies of formation; the standard entropy $\Delta_R S^\ominus$ is a measurable physical quantity that is often referred to as the degree of disorder in a system. The entropy of a reaction system and its environment determines whether a reaction is reversible or irreversible. Deeper insight into the conditions of reactions and reaction parameters can be found in any physical chemistry textbook [3, 9].

Many reactions are thermodynamically favorable but still do not occur in our normal environment because when we look at a graph displaying the potential energy over the reaction coordinate we see that the reactants' energy initially rises. This rise that has to be overcome for the reaction to happen is termed *activation energy*. One strategy is to use a substance referred to as *catalyst*, that has certain properties that accelerate a reaction without taking part in the overall chemical equation. This means it provides a faster route from reactants to products by lowering the activation energy without influencing the standard molar Gibbs energies of reactants and products. Catalysts can be divided into homogeneous and heterogeneous catalysts, depending on whether they are in the same or in a different phase as the reaction medium, respectively. For microswimmers, the catalysts are typically colloids or areas on the surface of the former and therefore they all belong to the class of heterogeneous catalysts.

Example 27.3: Changing activation energies for H_2O_2 (hydrogen peroxide). H_2O_2 is a rather unstable molecule and disporportionates at room temperature and in UV light:



The activation energy E_a is about 75 kJ mol^{-1} in absence of any catalyst. The presence of a catalyst enables alternative reaction pathways, which require lower activation energies. The reaction above in presence of a Pt catalyst only requires 49 kJ mol^{-1} , leading to an increased proportion of molecules that can form products: the Arrhenius equation allows us to calculate which part of molecules have an energy larger than the activation energy ($\exp(-E_a/\bar{R}T)$).

- a. Calculate the proportion of H_2O_2 molecules in presence and absence of catalyst. By comparing the ratio between the catalytic and the uncatalyzed reaction, you obtain the relative increase in reaction rate.
- b. Curious fact: The enzyme catalase (= biocatalyst) exists in many different varieties, which can lower the activation energy even further.

27.2 Chemical reactions driving active motion

In what follows, we will look at some of the main chemical reactions that have been employed in the literature to propel micromotors. We will deepen our understanding of the conditions for the chemical reactions involved and we will discuss specifications for the catalysts used. This part does not intend to be comprehensive. For a complete overview, the readers are directed to specific review articles on the topic [10, 11, 12].

27.2.1 Catalytic micromotors

All catalytic micromotors share the fact that their propulsion mechanism draws energy from the local conversion of a fuel (the surrounding medium), and they therefore lower the free energy of the system. The group of reactions which is dominating the field are degradation reactions. These reactions have the advantage of being dependent only on the attachment of a single species to the catalytic surface. The use of highly reactive species is preferred, with hydrogen peroxide being the most abundantly used fuel.

The degradation of hydrogen peroxide (H_2O_2)

The degradation of **hydrogen peroxide** (H_2O_2) is an exothermic reaction, but due to an activation energy of 76 kJ mol^{-1} [3] solutions of hydrogen peroxide are rather stable under ambient conditions. Since the activation energy can be drastically reduced when using a catalyst, this exothermic reaction with a standard enthalpy of about -190 kJ mol^{-1} makes hydrogen peroxyde an excellent fuel, which is for example used in rockets due to

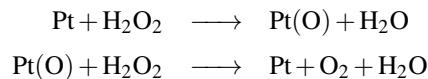
its high energy content per weight. By far, peroxide is also the most abundantly used fuel in the micromotors' literature and, even though there is a general route for the peroxide degradation leading to water and oxygen as reaction products, the specific conditions on the micromotor can result in drastically different propulsion mechanisms and have therefore important consequences for the motion.

The general reaction is given by



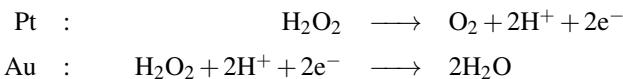
The oxygen within the peroxide molecule is characterized by an oxidation number of +1, which is rather atypical for oxygen, as it is more frequently found in the oxidation numbers 0 or -2. The hydrogen peroxide molecule can then serve both as oxidant and as reducing agent.

The most commonly used catalyst is **Platinum** (Pt), which decomposes hydrogen peroxide via a so called peroxide-oxide mechanism [13]:



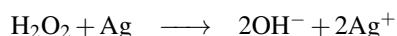
It is assumed that after attaching to the surface, the peroxide reacts with the surface of Pt to form Pt(O). In this intermediate step, the Pt is oxidized which allows the electrons to reduce the oxygen in water to an oxidation number of -2, leading to the release of one molecule of H₂O. Subsequently, another H₂O₂ molecule reduces the Pt(O) back to its initial state (metallic Pt), while forming a second molecule of water and one molecule of O₂. Pt can be evaporated, prepared chemically or electrodeposited, thus leading to a large variety of micromotors.

The first use of Pt as catalyst in a micromotor coincides with one of the first reports on synthetic micromotors [14]. In this work, the authors designed asymmetric polydimethylsiloxane (PDMS) plates with a catalytic Pt pin that, when placed at the liquid-air interface of a hydrogen peroxide solution, caused O₂ evolution and movement from the recoil of bubbles. Conditions that favor the formation of bubbles have been created for many other Platinum-peroxide systems, often involving cavities that constrict the diffusion of peroxide, such as in hollow tubes (rolled up or electrodeposited) [15, 16] or shell structures [17]. Chronologically, the next use of Pt to drive microscale objects happened soon thereafter using a microscale bimetallic Au/Pt rod, where again the Pt side degrades the peroxide [18]. Even though the overall catalytic reaction is the same (see Eq. (27.6)), the mechanism that leads to active propulsion is a very different one. Here, without going into too much details (see Chapter 8), an electrocatalytic mechanism comes into play and the peroxide degradation is split into two half-reactions:



In the anodic half reaction, the H⁺ ions are produced on the Pt-side. The protons migrate through the fluid towards the Au-side of the rod, where they are consumed by the cathodic half reaction. This results in an electrokinetic flow around the body, while the electrons pass through the metallic body, creating a self-imposed electric field. Later, Howse *et al.* developed micromotors (later referred to as Janus particles) by covering a hemisphere of non-conductive colloids with a thin Pt layer. Here, even though the average reaction remained the same (Eq. (27.6)), the lack of a conductive body led to another different mechanism: diffusiophoresis, which relies solely on a solute gradient, while the influence of any charged species is still disputed [19, 20]. When considering the catalyst properties more in detail, studies showed that the roughness of Pt increases the surface area, which is followed by an enhanced oxygen evolution and leads to a more efficient propulsion [21]. However, Esplandiu *et al.* studied the roughness influence on Pt patches in combination with semiconducting or insulating substrates and demonstrated how diverse the influences can actually be [22]. The catalytic activity of Platinum can also be influenced by other factors, such as the use of co-catalysts, that can enhance its catalytic activity, as has been shown by the team of Joe Wang [23].

Also other noble metals and metal oxides catalyze the degradation of hydrogen peroxide and have been occasionally used to propel micromotors. **Silver metal** (Ag) has long been used in jewellery and catalytic applications, including the degradation of hydrogen peroxide which triggered its use for catalytic micromotors [24]. While Eq. (27.6) remains valid, Silver is less stable than Platinum catalysts and takes part in the reaction, below, forming charged species:



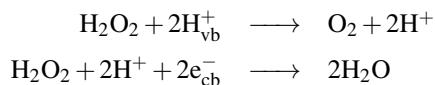
This effect increases with peroxide concentration and diminishes the micromotor swimming velocity.

Since peroxide and its degradation are well studied, it is known that in neutral solution, the Pt group metals can be ordered as follows, according to their specific activity: Pt > Os > Ir > Pd > Ru > Rh > Au. Alloys instead often show intermediate activities [25]. Some of these noble metals have been also employed as catalytic materials in peroxide driven micromotors. Examples include Au [26], Pd [27], and Cu [28] using **pH dependent peroxide degradation**. In an alkaline environment H_2O_2 is less stable, i.e., its self-decomposition rates are strongly enhanced. It has been shown that there is an optimal pH range in alkaline media which maximizes this degradation [29, 25]. This increased tendency to form reactive oxidic intermediates with noble metal surfaces was used by Dey *et al.* to catalyze peroxide on gold nanoparticles immobilized on polymeric microbeads leading to vertical (bubble driven) motion. The creation of bubbles occurred only at pH values > 9 [26]. The same team used this increased activity to create a micromotor capable of undergoing pH taxis, i.e., motion directed by a gradient of proton concentration. Nonetheless, exchanging the catalytic material to Palladium increased the overall reactivity and swimming activity [27].

Different **Manganese** (Mn) compounds have traditionally been used to degrade peroxide, with simple Mn(+IV) oxide MnO_2 being most frequently employed in micromotors [30, 31, 32]. MnO_2 exists in different crystalline polymorphic forms, which also influence its catalytic behavior. As was the case with Pt, the MnO_2 can be evaporated, prepared chemically or electrodeposited, thus leading to a large variety of micromotors. Differently from Pt, however, the mechanisms of motion for MnO_2 do not differ so much: even particle micromotors are propelled via bubble propulsion. The reasons are scarcely discussed in the literature, but we can assume that, besides the rather high catalytic rates, a major reason is Manganese's rough surface structures that offer nucleation points for bubbles.

With hydrogen peroxide being a very strong oxidative agent which easily forms hydroxyl radicals that damage proteins, DNA, whole cells and tissues as well as being a side product of various reactions, nature developed efficient strategies to deactivate it. **Enzymes** that degrade peroxide come from two main enzyme families: catalases and peroxidases. According to the Atkins Physical Chemistry [3], catalase reduces the activation energy for hydrogen peroxide decomposition to 8 kJ mol^{-1} , leading to a 10^{15} fold increase in reaction rates. This high throughput generally leads to high concentrations of peroxide and thereby fast nucleation of bubbles and subsequent propulsion [3]. However, enzymes are generally *engineered* by nature to efficiently degrade peroxides, not for stability in high peroxide concentrations, which means that these micromotors often suffer from short lifetimes caused by enzyme denaturation. Protective shelling is one possible solution to this problem [33].

Finally, **photocatalytic materials** can be employed for the decomposition of hydrogen peroxide. Before moving to concrete examples in the micromotors' literature, we should first have a brief look at the processes happening in general photocatalysis. The activation of the catalysts using light of certain wavelengths is often connected to semiconductors. This class of materials is characterized by its electronic properties, leading to a conductivity between those of a conducting material and of an insulator. The quantum states of electrons within the material can be associated with a band structure, that, depending on the filling, leads to certain electron transport properties (Fig. 27.4). During the process of photoactivation, the material absorbs a photon which provides the energy to promote an electron from the valence band (vb) to the conduction band (cb). This process creates two mobile charge carriers: an electron (e^-) in the conduction band and a hole (H^+) in the valence band. Depending on the charge carrier mobility in the specific material and the recombination tendencies between these charge pairs, the charge carriers can diffuse to the surface and cause reactions with the chemicals present in the solution. Degradation of hydrogen peroxide via photocatalytic materials is relatively straightforward because the energy needed to overcome the activation barrier is reached by a large number of photocatalytic materials. This process follows the reaction



Therefore, a large variety of materials has been produced into asymmetric morphologies. Early on, Palacci *et al.* used Fe_2O_3 embedded into a TPM sphere [35]; TiO_2 [36] and C_3N_4 [37] are frequently used after asymmetrization by a metal hemisphere. Another elegant strategy is using materials such as BiVO_4 , which have an inherent crystal asymmetry, enabling swimmers to propel without an extra asymmetrization step [38, 39].

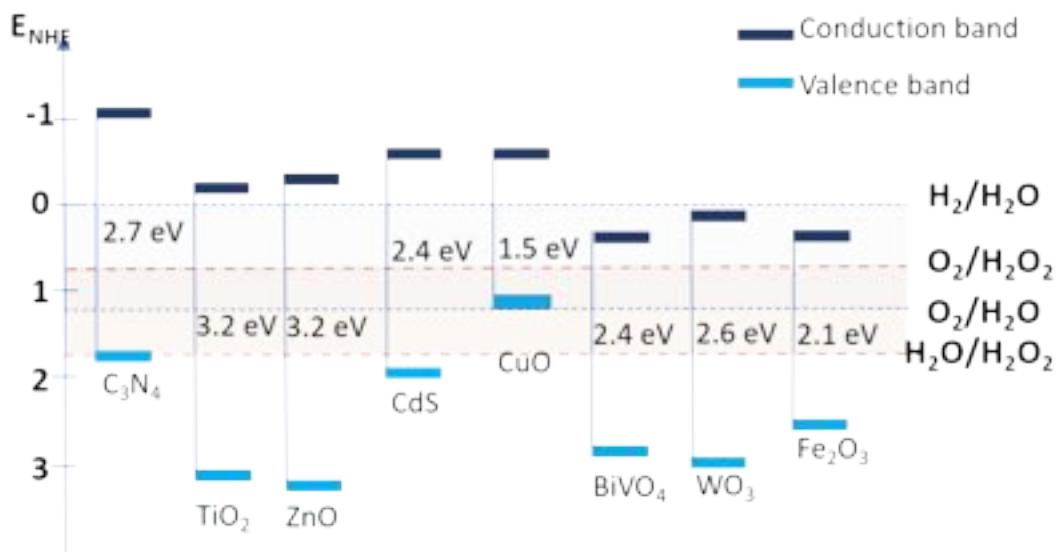
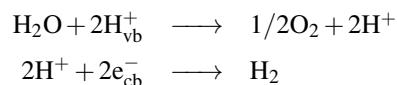


Figure 27.4: **Photocatalytic materials and their band gaps.** Band gaps of photocatalytic materials commonly employed in micromotor research. Data from Ref. [34].

Water splitting (H_2O)

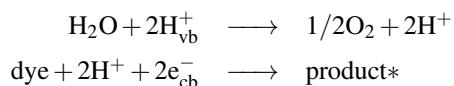
Splitting of water H_2O is one of the most promising reactions for generating green hydrogen by splitting one molecule of water into one molecule of hydrogen H_2 and half of an oxygen molecule O_2 :



There are different means of promoting this reaction. While the most widely used strategy is electro- or photoelectrochemical water splitting, only pure photocatalytic reactions have been employed to propel micromotors. The major limitation is due to the difficulty (or impossibility) of connecting external power sources to the catalytic material on the micromotors. This reaction typically results in less efficient propulsion than peroxide degradation. In a few publications, the same materials have been compared, obtaining generally much slower swimming velocities for water fuelled reactions [40].

Degradation of dyes

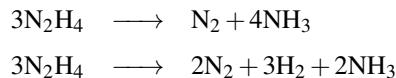
Using a photocatalytic strategy as explained above, organic materials can be also decomposed. When dissolved in water, the organic dyes methyl blue, cresol red and methyl orange were shown to enhance the velocities of Au@TiO_2 micromotors (i.e., a Janus particle where a gold shell covers a TiO_2 -based particle) by 1.7, 1.5, and 1.4 times, respectively [41], compared to swimming in pure water. The authors indicated the following mechanism:



However, the authors only studied the degradation using the decrease in absorbance intensity. This information does not allow conclusions how the chemical degradation of the dye proceeds. Before the motion experiments, motors and dyes were stirred for 30 minutes in the dark to favor adsorption of dyes on the surface of the photocatalyst, a prerequisite for efficient electrochemical redox reactions, especially in case of larger molecules that diffuse slowly.

Decomposition of hydrazine (N_2H_4)

As another rocket fuel, hydrazine (N_2H_4) has also been used to propel micromotors. Similarly to hydrogen peroxide, hydrazine is a colorless liquid which is degraded in an exothermic reaction. The molecular structure also has certain features in common with hydrogen peroxide: due to the free electron pair, a gauche conformation is favorable and the rotation around the N-N bond is limited by a strong rotational barrier. Generally, despite the toxicity of the molecule itself, the advantage of hydrazine compared to peroxide is the avoidance of expensive Pt-catalysts. The most frequently used catalyst is iridium, often in combination with oxide carrier materials, enhancing the catalytic activity. Some years ago Wang's group studied the use of hydrazine intensely [42], mainly for the very low fuel concentrations that could be used. The overall degradation at an Iridium surface mechanism has been reported as



Another feature that is worth mentioning is that the micromotors swim most efficiently with hydrazine concentration of 0.001%. An increase of the concentration beyond this value decreases the speed which the authors explained with an increased ionic strength, leading to inhibiting effects. The protonation of ammonia leads to high pH values in the solution, which enables the following reaction:



In an extreme example, even the hydrazine vapor accumulating in a neighboring droplet was sufficient to fuel micromotor propulsion [43].

Hydrolysis of urea ($(\text{NH}_2)_2\text{CO}$)

The degradation of urea via hydrolysis into carbon dioxide and ammonia is a reaction frequently found in nature, as urea is one of the main ways of the body to excrete nitrogen. Even though the molecule itself is not toxic, high concentrations in body fluids can become dangerous, so many living being had to find strategies to dismantle it. This is mainly done via the enzyme urease and the following reaction



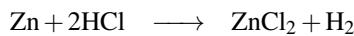
The inherent nonuniformity of urease coverage of microbeads, joint with the non-homogeneous functioning of individual enzymes was sufficiently asymmetric to propel micromotors [44]. Compared to catalase, for example, urease has a lower turn-over-number, so that the accumulation of reaction products is less favored. So despite the fact that CO_2 diffuses slower than O_2 and accumulates more, urease driven micromotors rarely generate bubbles (because of the slower functioning of urease) and therefore move via diffusiophoresis [45].

27.2.2 Redox reactions

Redox reactions are very common in general chemistry and are characterized by a change of oxidation state of the atoms taking part in it, involving a transfer of electrons. They are characterized by a dependence on the amount of the reactants, meaning that something is being consumed (typically part of the micromotor). Therefore, these reactions are limited to a certain time frame in which the compound is available in sufficient quantity.

Acids

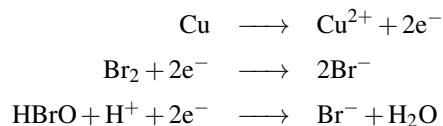
When strong acids are coming in contact with certain non-noble metals, the metals are able to displace the hydrogen, generating hydrogen gas. This reaction is frequently used to produce small amounts of H_2 . Using the example of zinc, this reaction was used in several micromotors by Wang's group [46]. The employed reaction reads as



The spontaneous ejection of hydrogen bubbles leads to an efficient jet force, which is frequently observed also for other gases, such as oxygen. An analogous concept has also been used for Aluminium metal [47] as well as for Magnesium [48].

Halogens

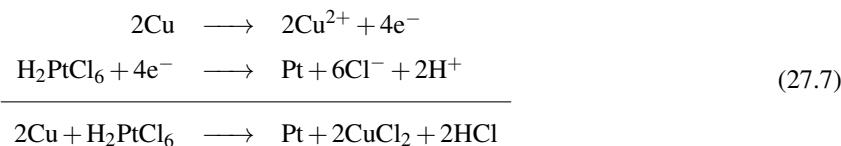
Another redox reaction, which we know from batteries as well, is the motion of copper-Platinum rods in solutions of different halogens, such as bromine or iodine. The movement of such a *nanobattery* is self-electrophoretic: the rods form a short-circuited galvanic cell. The electrons resulting from the reduction of Br_2 or I_2 (in dilute aqueous solutions) pass inside the rod creating an electric field [49]; the copper part of the rod is oxidized by the halogens, providing electrons that can further reduce either the bromine or its dissociated form in water HBrO , according to



This reaction happens analogously for I_2 . All the redox potentials of the respective reactions are in good agreement with the micromotor speed, establishing self-electrophoresis as a very likely mechanism for this reaction. The observed time of achievable motility was proportional to the rod length and, even without the Pt part, the motors were able to move if made into asymmetric shapes [49].

Noble metal ions

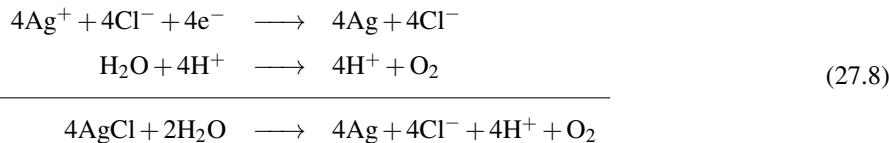
Metals are not only part of the motor body during a propelling redox reaction, but they can also be used as salts in their soluble form to generate propelling forces: in particular, noble metal salts or dissolved complexes in contact with less noble metals can generate electromotive forces that can be used to propel microobjects. The different standard potentials in the Galvanic series are known to be the driving force of a process called Galvanic exchange, which can be observed frequently on the macroscale. Using the same principle on microscale devices has also lead to active propulsion [50]. The mentioned example uses copper Janus particles in combination with hexachloroplatinic acid. The following reactions compensate for the potential difference present in the system and thereby create an electromotive force that directly propels the micromotors:



One useful feature of this system is that, after the galvanic reaction ends, Pt-Janus particles are obtained, that can be employed as normal, catalysis-driven micromotors. This principle can also be expanded to different noble metals, e.g., Gold [51].

Photosensitive silver salts

The inherent property of some specific materials to absorb and transform light energy for chemical processes has been used in photography over centuries. Sen *et al.* pioneered the use of AgCl for propelling micromotors after light exposure [52]. Specifically, the absorbed energy induces a surface modification on the particles from ionic silver (Ag^+) within the salt crystal to metallic silver (Ag):



This reaction creates a localized electrolyte gradient around the particle that results in self-diffusiophoresis [52, 53]. Later, strategies to produce Janus structures enabled more deep studies of the influences these ions have on diffusiophoresis and their respective environments [54, 55].

27.2.3 Polymerization reactions

In biological systems motility is mostly caused by bending of flexible body parts, but, in certain occasions, other strategies are also employed. *Listeria monocytogenes* have flagella for swimming outside of a host body; however, after entering a cell, the bacteria have been observed to use a polymerization-powered strategy. Living inside of host cells, *Listeria* can use the polymerization of the cell's actin to delocate its own body [56]. Inspired by this strategy, Sen's group immobilized a highly efficient Grubbs' catalyst on the silica side of gold–silica Janus microspheres. Norbornene (see Fig. 27.5) was selected as a monomer and dissolved in 1,1,1-trichloroethane, a solvent which also solubilized the resulting polymer.

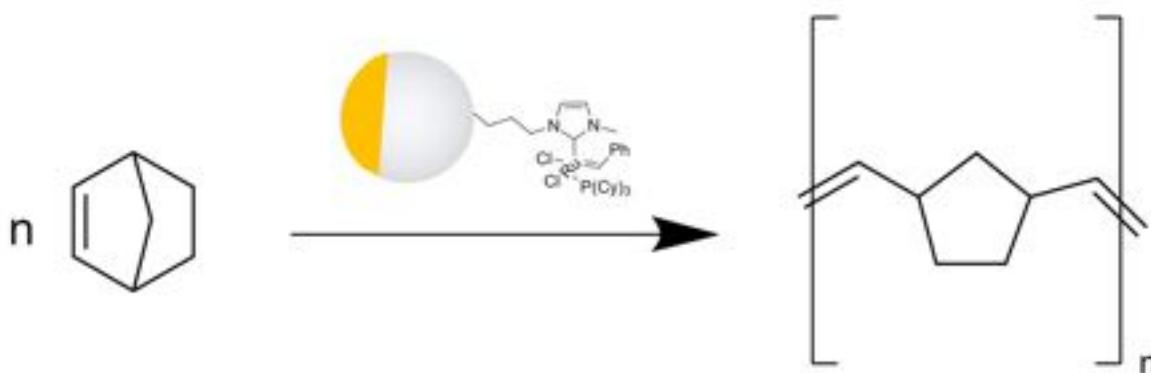


Figure 27.5: Polymerization-powered mobility. Polymerization of norbornene catalyzed by a Grubbs' catalyst immobilized on the silica hemisphere of a Au@SiO₂ Janus particle.

Unlike in *Listeria*, the particles are not pushed forward by the polymer chain, but the consumption of the monomer, combined with the production of a slowly diffusing product leads to an osmotic force across this chemical gradient. The force causes a net fluid flow, resulting in an active propulsion of the particle in the opposite direction.

27.3 Energy conversion efficiency

From the wide variety of reactions described above, one might expect that the underlying reaction determines directly (via ΔG) how efficiently a micromotor can move. Unfortunately, this simple correlation does not apply, as we can see from the broad range of different speeds and mechanisms we can obtain for peroxide degradation. In general, the efficiency η , i.e., the ratio of work output of the motor to the expended energy, is not a frequently investigated parameter on the nano- and microscale. For some biological systems, it is known that they harvest mechanical work rather efficiently from the conversion of chemical species, a specific ATPase is known to reach almost 100% energy conversion [57].

Wang and Mallouk's article in 2013 [58] is one of the few in-depth considerations of the efficiencies of artificial micromotors. They compared the efficiency of different micromotors by calculating the ratio

$$\text{efficiency} = \frac{\text{power output}}{\text{power input}}$$

from the observed speed values and fuel consumptions [58]. Thereby they find that most micromotors have rather low energy efficiencies, due to the fact that energy is lost to the environment at multiple stages (Fig. 27.6):

- Background catalytic consumption of fuel.
- Losses in conversion (e.g., chemical energy to electric potential).
- Opposing flows.

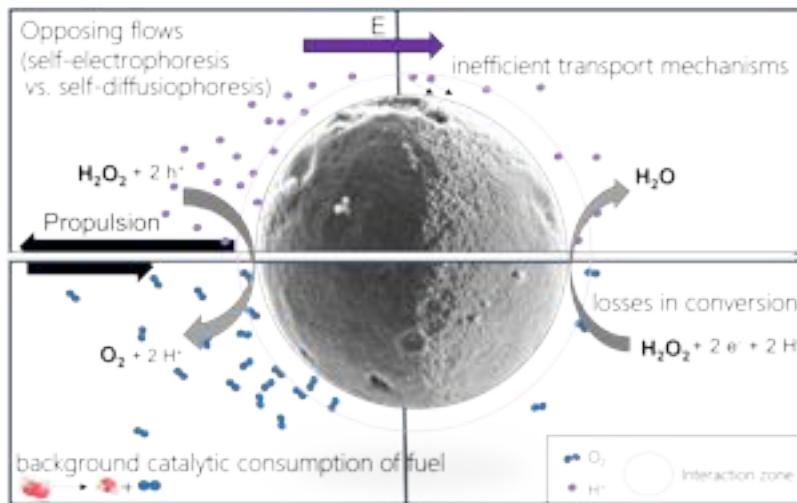


Figure 27.6: Energy conversion efficiency. Energy efficiency of active matter is an intricate topic. Considering two different mechanisms for $\text{Au}@\text{TiO}_2$ micromotors reveals that different mechanistic contributions can lead to propulsion towards opposite directions and other contributions that lower the output speed.

- Inefficient transport mechanisms.

In an attempt to understand where these losses in energy conversion originate, we look at theoretical considerations evaluating the thermodynamic efficiency under idealized conditions: Tao and Kapral studied self-propelled catalytic-inert nanodymers propelled by concentration gradients as well as unequal affinities of the reaction components to the solution with an efficiency of 10^{-4} [59], about 4-5 orders of magnitude larger than the experimental systems studied by Mallouk's group and Paxton's estimate for experimental bimetallic nanorod [60].

Calculations yielding the efficiency generally consider the ratio of mechanical power dissipated by the rod to the input power calculated from the oxygen generation rate multiplied with the free energy change of the reaction. While the mechanical power is straightforward, obtained as the product of velocity and drag force, the quantification of products on the microscale is rarely attempted experimentally. Wang *et al.* developed an approach where the oxygen evolution rate was determined for a well-known amount of active $\text{Au}@\text{TiO}_2$ colloids. The comparison with coarse estimates of velocities confirmed that the motion mechanism relies mainly on the ionic mechanism [61], which implies that the motion involving charged species is significantly more efficient, compared to neutral diffusiophoresis. This is likely due to better atomic-level mechanical coupling of the catalyst with solutes and the improved localization of fuel degradation at the catalytic site. While looking at the specific conditions of chemical reactions is important, without considering also the fluid dynamic conditions of each motion principle, the involved energy conversions and the possibilities of energy loss, understanding of motion on the microscale will not be complete.

Example 27.4: Energy conversion efficiency of a catalytic Janus swimmer. In Ref. [51], we calculated the efficiencies of galvanophoretic micromotors using the ratio of power output to power input. Here, we adapt this approach for catalytic Janus swimmers:

$$\eta = \frac{P_{\text{output}}}{P_{\text{input}}} = \frac{P_{\text{mech}}}{P_{\text{chem}}}. \quad (27.9)$$

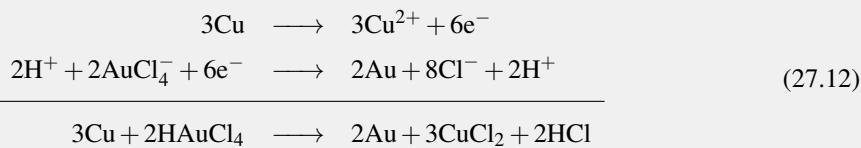
Following Mallouk's paper, the power output of a micromotor equals the energy that is dissipated, which can be approximated using the drag force and the motor velocity:

$$P_{\text{output}} = P_{\text{mech}} = 6\pi\mu rv^2. \quad (27.10)$$

The power input of the micromotor is estimated using the galvanic exchange which drives the microswimmers. Similarly to earlier approaches, we use the total reaction rate multiplied by the Gibbs free energy of Eq. (27.12).

$$P_{\text{input}} = P_{\text{chem}} = k_S |\Delta G_{\text{free}}| = k_S | -zFE |. \quad (27.11)$$

For the reactions involved, z , the number of exchanged electrons, is equal to 6, as can be seen here:



The *chemical power input* is estimated using the Galvanic exchange which drives the microswimmers. The cell potential E equals 0.66 V:

$$\Delta G_{\text{free}} = -zFE = -6 \times 96485 \text{ A s mol}^{-1} \times 0.66 \text{ V} \quad (27.13)$$

resulting in $\Delta G_{\text{free}} = -382 \text{ kJ mol}^{-1}$.

The total reaction rate k_S is often the value that is most difficult to obtain. In the example in Ref. [51], we assume the entire surface to be involved and the reaction to be limited by the availability of the AuCl_4^- complexes. We can then use a theoretically obtained diffusion-based upper bound of the reaction rate.

With the help of simulations, a theoretical upper limit to k_S can be determined by exploiting the fact that the fuel AuCl_4^- can only diffuse to the particle's surface at a finite rate, as one can see from the table of the maximally possible reaction rates for galvanophoretic micromotors (from the SI of Ref. [51]):

$c(\text{AuCl}_4^-) [\mu\text{mol}]$	$k_S [\text{mol s}^{-1}]$	
	$r = 1 \mu\text{m}$	$r = 1.5 \mu\text{m}$
1.6	1.20×10^{-17}	1.89×10^{-17}
3.2	2.83×10^{-17}	4.38×10^{-17}
6.4	6.42×10^{-17}	9.83×10^{-17}
13	1.42×10^{-16}	2.16×10^{-16}

Results for the swimmer efficiencies lie in the range of 10^{-7} [51].

27.4 A note on nanomotors and enzymes

Looking towards applications in biomedical engineering, they often demand the shift towards smaller active units. One frequently cited example is drug delivery, where many published works rely on the Enhanced Permeability and Retention effect (EPR), which was demonstrated for particles up to $\approx 180 \text{ nm}$. However, moving from micro- to nanoscale motion causes several additional challenges. While from the synthetic side, nanoscale options for most materials are available, and oftentimes even more numerous compared to their larger cousins, the physics of motion are not that straight forward. The prevalence of laminar flow at the nanoscale is even more dominant than at the microscale, so influences originating from the Navier-Stokes equation, such as the lack of time asymmetry and the resulting requirement of a constant energy throughput, must be considered. Compared to microscale systems, the influence of Brownian motion increases and clearly dominates over ballistic self-propulsion. Because any propulsion must be observed on top of the diffusive random walk, directed behaviour is hard to achieve. Generally, with the exception of externally guided systems, only enhanced diffusion can be observed. Due to the Abbe limitation, the detection and characterization of nanoscale motion is significantly more difficult compared to the microscale. Because in most experiments, visibility in optical microscopy is not given, other methods for evaluation are required.

While a number of works have measured significantly increased diffusion coefficients in the presence of substrate, Günther *et al.* highlight that Fluorescence Correlation Spectroscopy (FCS) is an intricate technique and suggests a number of control experiments (including further purification and characterization of the labelled enzymes) and analysis of complementary data (such as the intensity time trace) to avoid misinterpretation of

signals and clarify whether, and to which extent, enzymes can show active self-propulsion. However, enzymes have not been limited to motile entities themselves, but they have even been used as biocatalytic units grafted onto microscale particles [62, 63]. Synthetic difficulties, for example the reduction of enzyme activity during the chemical coupling to the colloid [33], restricted the practical use of this type of systems drastically, with the exception of the Sanchez' group. Focusing mainly on the enzyme urease [44, 64], they could show the importance of the ionic nature of the generated product. Recent reviews on this research area are available [65, 66, 67].

27.5 Problems

Problem 27.1: Chemical composition of fuels and catalysts.

- a. Revise the literature described above (and active matter articles beyond that) and mark the chemical composition of the fuels in the periodic table.
- b. Reconsider the articles describing (photo-)catalytic reactions and mark the materials of the catalysts in the periodic table. Compare to the periodic table in a, how do fuels and catalysts differ?

Problem 27.2: Phase identification for TiO₂.

Compare the XRD diffractogram of the photocatalyst TiO₂ to the two different phases (Anatase and Rutile). The XRD diffractogram is given in Fig. 27.2b.

Bibliography

- [1] D. Voet and J. G. Voet. *Biochemistry*. Wiley, 2011.
- [2] J. M. Berg, J. L. Tymoczko, and L. Stryer. *Biochemistry*, 2002.
- [3] P. W. Atkins and J. De Paula. *Physical chemistry*, 1998.
- [4] X. Wang, R. Guo, and J. Liu. Liquid metal based soft robotics: materials, designs, and applications. *Advanced Materials Technologies*, 4:1800549, 2019.
- [5] M. Liu, Y. Wang, Y. Kuai, J. Cong, Y. Xu, H.-G. Piao, L. Pan, and Y. Liu. Magnetically powered shape-transformable liquid metal micromotors. *Small*, 15:1905446, 2019.
- [6] C. A. Grattoni, R. A. Dawe, C. Y. Seah, and J. D. Gray. Lower critical solution coexistence curve and physical properties (density, viscosity, surface tension, and interfacial tension) of 2,6-lutidine+water. *Journal of Chemical and Engineering Data*, 38:516–519, 1993.
- [7] G. Volpe, I. Buttinoni, D. Vogt, H.-J. Kümmerer, and C. Bechinger. Microswimmers in patterned environments. *Soft Matter*, 7:8810–8815, 2011.
- [8] M. Nic, L. Hovorka, J. Jirat, B. Kosata, and J. Znamenacek. Iupac. compendium of chemical terminology 2nd ed.(the gold book). international union of pure and applied chemistry, 2005.
- [9] A. Adamson. *A textbook of physical chemistry*. Elsevier, 2012.
- [10] L. Sánchez, S. and Soler and J. Katuri. Chemically powered micro- and nanomotors. *Angewandte Chemie International Edition*, 54:1414–1444, 2015.
- [11] M. Safdar, J. Simmchen, and J. Jänis. Light-driven micro- and nanomotors for environmental remediation. *Environmental Science: Nano*, 4:1602–1616, 2017.
- [12] W. Wang, W. Duan, S. Ahmed, A. Sen, and T. E. Mallouk. From one to many: Dynamic assembly and collective behavior of self-propelled colloidal motors. *Accounts of Chemical Research*, 48(7):1938–1946, 2015.

- [13] R. Serra-Maia, M. Bellier, S. Chastka, K. Tranhuu, A. Subowo, J. D. Rimstidt, P. M. Usov, A. J. Morris, and F. M. Michel. Mechanism and kinetics of hydrogen peroxide decomposition on platinum nanocatalysts. *ACS applied materials & interfaces*, 10:21224–21234, 2018.
- [14] R. F. Ismagilov, A. Schwartz, N. Bowden, and G. M. Whitesides. Autonomous movement and self-assembly. *Angewandte Chemie International Edition*, 41:652–654, 2002.
- [15] A. A. Solovev, Y. Mei, E. Bermúdez Ureña, G. Huang, and O. G. Schmidt. Catalytic microtubular jet engines self-propelled by accumulated gas bubbles. *Small*, 5:1688–1692, 2009.
- [16] W. Gao, S. Sattayasamitsathit, J. Orozco, and J. Wang. Highly efficient catalytic microengines: Template electrosynthesis of polyaniline/platinum microtubes. *Journal of the American Chemical Society*, 133:11862–11864, 2011.
- [17] M. Saifdar, T. Ikonen, and J. Jänis. Bubble-propelled trimetallic microcaps as functional catalytic micromotors. *RSC Advances*, 5:13171–13174, 2015.
- [18] W. F. Paxton, K. C. Kistler, C. C. Olmeda, A. Sen, S. K. St. Angelo, Y. Cao, T. E. Mallouk, P. E. Lammert, and V. H. Crespi. Catalytic nanomotors: autonomous movement of striped nanorods. *Journal of the American Chemical Society*, 126:13424–13431, 2004.
- [19] A. Brown and W. Poon. Ionic effects in self-propelled pt-coated Janus swimmers. *Soft Matter*, 10:4016–4027, 2014.
- [20] A. T. Brown, W. C. K. Poon, C. Holm, and J. de Graaf. Ionic screening and dissociation are crucial for understanding chemical self-propulsion in polar solvents. *Soft Matter*, 13:1200–1222, 2017.
- [21] U. Choudhury, L. Soler, J. G. Gibbs, S. Sanchez, and P. Fischer. Surface roughness-induced speed increase for active Janus micromotors. *Chemical Communications*, 51:8660–8663, 2015.
- [22] K. Zhang, J. Fraxedas, B. Sepulveda, and M. J. Esplandiu. Photochemically activated motors: from electrokinetic to diffusion motion control. *ACS applied materials & interfaces*, 9:44948–44953, 2017.
- [23] C. Chen, X. Chang, H. Teymourian, D. E. Ramírez-Herrera, E. Esteban-Fernández de Ávila, X. Lu, J. Li, S. He, C. Fang, Y. Liang, F. Mou, J. Guan, and J. Wang. Bioinspired chemical communication between synthetic nanomotors. *Angewandte Chemie International Edition*, 57:241–245, 2018.
- [24] Z. H. Shah, S. Wang, L. Xian, X. Zhou, Y. Chen, G. Lin, and Y. Gao. Highly efficient chemically-driven micromotors with controlled snowman-like morphology. *Chemical Communications*, 56:15301–15304, 2020.
- [25] D. W. McKee. Catalytic decomposition of hydrogen peroxide by metals and alloys of the platinum group. *Journal of Catalysis*, 14(4):355–364, 1969.
- [26] K. K. Dey, B. Ranjan Panda, A. Paul, S. Basu, and A. Chattopadhyay. Catalytic gold nanoparticle driven ph specific chemical locomotion. *Journal of Colloid and Interface Science*, 348:335–341, 2010.
- [27] K. K. Dey, S. Bhandari, D. Bandyopadhyay, S. Basu, and A. Chattopadhyay. The ph taxis of an intelligent catalytic microbot. *Small*, 9:1916–1920, 2013.
- [28] P. Chattopadhyay and J. Simmchen. Interactions of different Janus particles with passive tracers. 2019.
- [29] Joseph Weiss. The free radical mechanism in the reactions of hydrogen peroxide. volume 4 of *Advances in Catalysis*, pages 343–365. Academic Press, 1952.
- [30] M. Saifdar, O. M. Wani, and J. Janis. Manganese oxide-based chemically powered micromotors. *ACS applied materials & interfaces*, 7:25580–25585, 2015.
- [31] H. Wang, G. Zhao, and M. Pumera. Beyond platinum: Bubble-propelled micromotors based on Ag and MnO₂ catalysts. *Journal of the American Chemical Society*, 136(7):2719–2722, 2014.

- [32] L. Wang, J. Chen, X. Feng, W. Zeng, R. Liu, Xujing L., Y. Ma, and L. Wang. Self-propelled manganese oxide-based catalytic micromotors for drug delivery. *RSC Advances*, 6:65624–65630, 2016.
- [33] J. Simmchen, A. Baeza, D. Ruiz-Molina, and M. Vallet-Regí. Improving catalase-based propelled motor endurance by enzyme encapsulation. *Nanoscale*, 6:8907–8913, 2014.
- [34] D. Xia, Q. Chen, Z. Li, M. Luo, and P. K. Wong. 6 - band gap engineered chalcogenide nanomaterials for visible light-induced photocatalysis. In Mohammad Mansoob Khan, editor, *Chalcogenide-Based Nanomaterials as Photocatalysts*, Micro and Nano Technologies, pages 135–172. Elsevier, 2021.
- [35] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339:936–940, 2013.
- [36] L. Wang, A. Kaepller, D. Fischer, and J. Simmchen. Photocatalytic TiO₂ micromotors for removal of microplastics and suspended matter. *ACS applied materials & interfaces*, 11:32937–32944, 2019.
- [37] V. Sridhar, F. Podjaski, J. Kröger, A. Jiménez-Solano, B.-W. Park, B. V. Lotsch, and M. Sitti. Carbon nitride-based light-driven microswimmers with intrinsic photocharging ability. *Proceedings of the National Academy of Sciences*, 117:24748–24756, 2020.
- [38] S. Heckel and J. Simmchen. Photocatalytic BiVO₄ microswimmers with bimodal swimming strategies. *Advanced Intelligent Systems*, 1:1900093, 2019.
- [39] S. Heckel, J. Grauer, M. Semmler, T. Gemming, H. Löwen, B. Liebchen, and J. Simmchen. Active assembly of spheroidal photocatalytic bivo₄ microswimmers. *Langmuir*, 36:12473–12480, 2020.
- [40] L. Wang, M. N. Popescu, F. Stavale, A. Ali, T. Gemming, and J. Simmchen. Cu@ TiO₂ Janus microswimmers with a versatile motion mechanism. *Soft Matter*, 14:6969–6973, 2018.
- [41] Y. Wu, R. Dong, Q. Zhang, and B. Ren. Dye-enhanced self-electrophoretic propulsion of light-driven TiO₂–Au Janus micromotors. *Nano-micro letters*, 9:30, 2017.
- [42] W. Gao, A. Pei, R. Dong, and J. Wang. Catalytic iridium-based janus micromotors powered by ultralow levels of chemical fuels. *Journal of the American Chemical Society*, 136:2276–2279, 2014.
- [43] R. Dong, J. Li, I. Rozen, B. Ezhilan, T. Xu, C. Christianson, W. Gao, D. Saintillan, B. Ren, and J. Wang. Vapor-driven propulsion of catalytic micromotors. *Scientific reports*, 5:1–7, 2015.
- [44] X. Arqué, A. Romero-Rivera, F. Feixas, T. Patiño, S. Osuna, and S. Sánchez. Intrinsic enzymatic properties modulate the self-propulsion of micromotors. *Nature Communications*, 10:1–12, 2019.
- [45] K. K. Dey, X. Zhao, B. M. Tansi, W. J. Méndez-Ortiz, U. M. Córdova-Figueroa, R. Golestanian, and A. Sen. Micromotors powered by enzyme catalysis. *Nano Letters*, 15:8311–8315, 2015.
- [46] W. Gao, A. Uygun, and J. Wang. Hydrogen-bubble-propelled zinc-based microrockets in strongly acidic media. *Journal of the American Chemical Society*, 134:897–900, 2012.
- [47] W. Gao, M. D’Agostino, V. Garcia-Gradilla, J. Orozco, and J. Wang. Multi-fuel driven Janus micromotors. *Small*, 9:467–471, 2013.
- [48] W. Gao, X. Feng, A. Pei, Y. Gu, J. Li, and J. Wang. Seawater-driven magnesium based Janus micromotors for environmental remediation. *Nanoscale*, 5:4696–4700, 2013.
- [49] R. Liu and A. Sen. Autonomous nanomotor based on copper–platinum segmented nanobattery. *Journal of the American Chemical Society*, 133:20064–20067, 2011.
- [50] J. Bastos-Arrieta, C. Bauer, A. Eychmüller, and J. Simmchen. Galvanic replacement induced electromotive force to propel Janus micromotors. *The Journal of Chemical Physics*, 150:144902, 2019.
- [51] L. Feuerstein, C. G. Biermann, Z. Xiao, C. Holm, and J. Simmchen. Highly efficient active colloids driven by galvanic exchange reactions. *Journal of the American Chemical Society*, 143:17015–17022, 2021.

- [52] M. Ibele, T. E. Mallouk, and A. Sen. Schooling behavior of light-powered autonomous micromotors in water. *Angewandte Chemie*, 121:3358–3362, 2009.
- [53] J. Simmchen, A. Baeza, A. Miguel-Lopez, M. M. Stanton, M. Vallet-Regi, D. Ruiz-Molina, and S. Sánchez. Dynamics of novel photoactive AgCl microstars and their environmental applications. *Chem-NanoMat*, 3:65–71, 2017.
- [54] C. Zhou, H. P. Zhang, J. Tang, and W. Wang. Photochemically powered AgCl Janus micromotors as a model system to understand ionic self-diffusiophoresis. *Langmuir*, 34:3289–3295, 2018.
- [55] T. Huang, V. R. Misko, S. Gobeil, X. Wang, F. Nori, J. Schütt, J. Fassbender, G. Cuniberti, D. Makarov, and L. Baraban. Inverse solidification induced by active Janus particles. *Advanced Functional Materials*, 30:2003851, 2020.
- [56] J. M. Stevens, E. E. Galyov, and M. P. Stevens. Actin-dependent movement of bacterial pathogens. *Nature Reviews Microbiology*, 4:91–101, 2006.
- [57] A. F. Huxley, R. M. Simmons, K. Kinoshita, R. Yasuda, H. Noji, and K. Adachi. A rotary molecular motor that can work at near 100% efficiency. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355:473–489, 2000.
- [58] W. Wang, T.-Y. Chiang, D. Velegol, and T. E. Mallouk. Understanding the efficiency of autonomous nano-and microscale motors. *Journal of the American Chemical Society*, 135(28):10557–10565, 2013.
- [59] Y.-G. Tao and R. Kapral. Dynamics of chemically powered nanodimer motors subject to an external force. *The Journal of Chemical Physics*, 131:024113, 2009.
- [60] W. F. Paxton, A. Sen, and T. E. Mallouk. Motility of catalytic nanoparticles through self-generated forces. *Chemistry—A European Journal*, 11:6462–6470, 2005.
- [61] L. Wang and J. Simmchen. Determination of the swimming mechanism of au@tio₂ active matter and implications on active–passive interactions. *Soft Matter*, 19:540–549, 2023.
- [62] D. Pantarotto, W. R. Browne, and B. L. Feringa. Autonomous propulsion of carbon nanotubes powered by a multienzyme ensemble. *Chemical Communications*, pages 1533–1535, 2008.
- [63] J. Simmchen, A. Baeza, D. Ruiz, M. J. Esplandiu, and M. Vallet-Regí. Asymmetric hybrid silica nanomotors for capture and cargo transport: towards a novel motion-based DNA sensor. *Small*, 8:2053–2059, 2012.
- [64] A. C. Hortelao, C. Simó, M. Guix, S. Guallar-Garrido, E. Julián, D. Vilela, L. Rejc, P. Ramos-Cabrera, U. Cossío, V. Gómez-Vallejo, T. Patiño, J. Llop, and S. Sánchez. Swarming behavior and in vivo monitoring of enzymatic nanomotors within the bladder. *Science Robotics*, 6:eabd2823, 2021.
- [65] T. Patiño, X. Arqué, R. Mestre, L. Palacios, and S. Sánchez. Fundamental aspects of enzyme-powered micro- and nanoswimmers. *Accounts of Chemical Research*, 51:2662–2671, 2018.
- [66] H. Yuan, X. Liu, L. Wang, and X. Ma. Fundamentals and applications of enzyme powered micro/nano-motors. *Bioactive Materials*, 6:1727–1749, 2021.
- [67] G. Salinas, S. M. Beladi-Mousavi, and A. Kuhn. Recent progress in enzyme-driven micro/nanoswimmers: From fundamentals to potential applications. *Current Opinion in Electrochemistry*, 32:100887, 2022.

Chapter 28

Experiments with Artificial Active Particles

CHUN-JEN CHEN, FÉLIX GINOT

In this Chapter, we discuss how to perform experiments with active Brownian particles and how to analyze data from these experiments. As an emblematic example of experimental active Brownian particles, we consider the case of Janus particles. The first part of the Chapter starts with a general introduction to experiments with Janus particles with instructions and explanations regarding experimental systems at hand; this includes not only systems with self-propulsion but also active reorientation (torque) and individual controls of active motion.

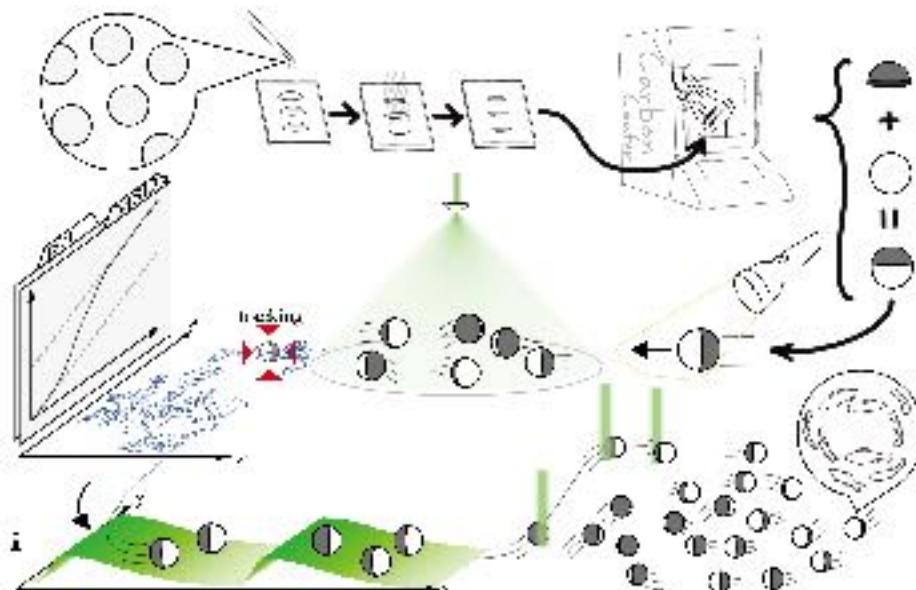


Figure 28.1: **Experiments with Janus particles.** Flowchart from fabrication of carbon coated spherical Janus particles to experiments with motion control and analysis. Starting from the upper left corner of the image and proceeding clockwise: Janus particles are fabricated from spherical colloidal particles by coating a dry monolayer with carbon. When suspended in a critical mixture of water–2,6-lutidine, the Janus particles self-propel themselves when activated by light. They can be used as a model system in experiments to study the collective behavior displayed by active matter systems thanks to the possibility to control their motion in speed and orientation. Computer-assisted tracking of the experiment can happen also during the experiment if real-time control of the particles behavior is needed. [@Chun-Jen could you check the caption?]

The second part introduces the basics for experimental analysis, as well as the description of some typical observables that are meaningful in active matter. The topics cover from particle detection to characterization of active motion with global drift removal, correlation functions, mean square displacement, and some commonly defined order parameters.

28.1 Janus particles as active particles

Named after the two-faced Roman god, a Janus particle consists of two parts in one body. Such a structure is anisotropic, making it easy for a direction of motion to arise. In general, a Janus particle can be any particle that satisfies the above definition, in any size and of any arbitrary geometry. In this Chapter, we will consider spherical micrometer-sized Janus particles.

Janus particles are one of the most representative synthetic motile active particles (also called microswimmers or self-propelled particles) due to their simple and straight-forward design. When designing a self-propelled particle, two questions must be answered: what is the energy source of the particles and how to convert such energy into motion. Answers to the first question can be, for example, chemical reactions, absorption of light, interaction with electric or magnetic fields [1]. The specific choice highly depends on the application. As for the answer to the second question, the key is to create the vector of propulsion velocity. Janus particles are good solutions because of the well-defined orientation direction across the two parts. For example, in hydrogen-peroxide solution, platinum is an active material since it acts as a catalyst to the spontaneous exothermic reaction that releases oxygen and water. If we attach a platinum surface to an inert, non-catalytic material, say polystyrene, the difference in reaction rates between the two sides results in a net propulsion pointing from the platinum to the other material [2, 3].

Janus particles are considered as the closest realistic counterpart of the active Brownian particle (ABP) model. Particles in the ABP model only experience translational diffusion, rotational diffusion, and a constant speed propulsion always aligned with the particle orientation [1]. The propulsion of the Janus particles is usually, if not always, aligned with the axis across the two different parts owing to their propulsion mechanism.

In this section, we briefly review the key ideas for the fabrication and propulsion mechanisms of Janus particles. Afterwards, we take the example of light-absorbing Janus particles that propel in a near-critical-point binary mixture. This specific type of self-propelled particles provides additional control through the manipulation of the light intensity, which we introduce in the next section.

28.1.1 Fabrication of Janus particles

The structure of the Janus particles leads to three major kinds of fabrication strategies: combination, synthesis, and modification. One can easily imagine that the combination of two particles, say an A particle and a B particle, can generate a Janus particle. In fact, this method does work [4, 5]. However, the technique to arrange one single A particle and one single B particle together and trigger their fusion is not always straightforward, or even possible.

Synthesis of Janus particles from raw materials is also possible through generation of a biphasic pattern. For example, a microfluidic device produces two parallel streams of photocurable fluids, containing the ingredients for the A and B parts of the Janus particles, respectively. Either a fast photo-curing process creates Janus particle that captures such a pattern [6], or a second microfluidic design breaks the co-flowing streams into droplets before photo-curing [7, 8, 9]. The biphasic pattern can also form within rather homogeneous droplets due to phase separation [10, 11]. These synthesis methods require more knowledge and work regarding particle synthesis and the manipulation of mixing and demixing of the materials compared to the other approaches, but if applicable, they allow more control and engineering on the resulting Janus particles.

The last approach is to modify only one side of an A particle with the characteristic of the B material [12, 3]. Such an approach is often easily applicable and suitable for mass production due to various coating and surface modification techniques. The resulting Janus particles are, from the surface point of view, composed of two parts, but inside the bulk, they are majorly only one single type of particle. Luckily, most of the self-propulsion mechanisms depend on the interaction between the particle surface and the surroundings, making the surface modification a common approach for such kind of experiments. In the following, we provide an example of the preparation of Janus particles by surface coating.

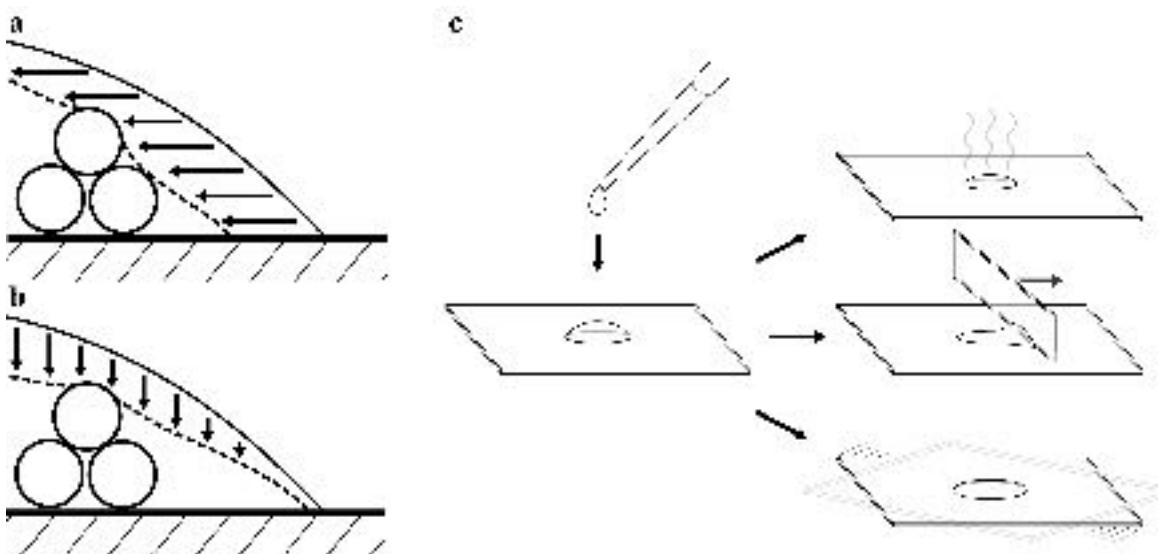


Figure 28.2: Strategies for monolayer preparation with surface tension. **a,b** Schematics of a liquid–air interface moving through a double-layer from the side (**a**) and from the top (**b**). **c** Monolayer preparation with drop casting (top), blade coating (middle), and spin coating (bottom).

We take the example of carbon-coated silica Janus particles. These particles absorb more visible light on the dark (carbon-coated) hemisphere than on the transparent (non-coated) hemisphere. You can fabricate such particles from colloids of various sizes. For the sake of simplicity, in this Chapter, we focus on monodisperse colloids with a diameter of $6.2\text{ }\mu\text{m}$. In the first step, a suspension of particles in water is spread over the surface of a microscope slide and left to dry. It is important to ensure that in the drying process, a monolayer of particles is formed, i.e., that no particles sit on top of others, otherwise not all particles receive the same coating. The preparation of a monolayer on a substrate can be helped by the surface tension at the water–air interface. As shown in Figs. 28.2a and 28.2b, particles inside the layer of water get pushed away from a position on top of the monolayer while the water layer collapses vertically or shrinks horizontally. Making use of such an effect due to surface tension, techniques like drop-casting, blade-coating, and spin-coating can produce monolayers of particles with the right parameters. *Drop-casting* works by dropping the suspension onto the substrate and waiting for the water to evaporate in a controlled condition. The water–air interface travels through the sedimented particles as the droplet shrinks due to evaporation and drags the particles on top of others down to the substrate. *Blade-coating* replaces the evaporation with a hydrophilic deposition blade that wets on the droplet. As the blade moves, it drags the droplet containing the particle suspension through the coating area and leaves the sedimented particles behind. In *spin-coating*, the substrate with a droplet containing the particle suspension spins on a rotor. The spinning rate is high enough so that the centrifugal force flattens the droplet until the thin film of liquid evaporates. Subsequently, a layer of carbon (here around 80 nm) is sputtered perpendicular to the monolayer, casting a hemispherical coating on each exposed particle. Afterwards, the coated particles are removed from the substrate (e.g., a microscope slide) by dipping it into water. The prepared Janus particles are then stored in water for later use.

28.1.2 Propulsion mechanisms of Janus particles

Due to the fact that there are no moving parts, a Janus particle does not propel due to the Newton's third law. Instead, it generates local flow field that results in particle motion.

In the famous case of Pt-coated Janus particles in a hydrogen-peroxide solution, self-diffusiophoresis and/or self-electrophoresis due to the different reaction rates of hydrogen-peroxide decomposition induce a surface flow that propels the particle. In self-diffusiophoresis, the reaction on the particle surface results in a concentration gradient of the chemicals involved. Such chemical gradient causes a concentration flow that imposes particle motion (diffusiophoresis). In self-electrophoresis, charge (e^- and ions) transfer and flux in the chem-

ical reaction are taken into account. These movements of charges generate electric fields that drive the liquid flow at the surface.

As for our carbon-coated Janus particles with different absorbance on each hemisphere, they induce concentration flows when suspended in a water–2,6-lutidine mixture and heated close to the mixture critical point with light. This type of self-propulsion by local demixing was first presented in 2011 using gold-coated Janus particles in a critical mixture of water and 2,6-lutidine (2,6-dimethylpyridine) [13]. When illuminated with a laser beam (e.g., of wavelength $\lambda = 532\text{ nm}$), these Janus particles propel toward the non-coated hemisphere with a propulsion velocity adjustable by the illumination intensity. Such active motion is achieved by exploiting the phase-behavior of a binary liquid which can either exist in a homogeneously mixed or in a separated phase. Depending on the concentrations, the transition from the mixed to the separated phase can either be first order (binodal decomposition) or, at the critical point, second order (spinodal decomposition). The phase diagram of water–2,6-lutidine has an upper critical point (UCP) at a temperature of $\approx 230^\circ\text{C}$ [14] and a lower critical point (LCP) with critical temperature $T_c = 34.1^\circ\text{C}$ at the lutidine concentration $c_{\text{lut}} = 28\%\text{wt}$ (Fig. 28.3a). If a critical composition is heated from $T < T_c$ to $T \gtrsim T_c$, it phase separates by spinodal decomposition (Fig. 28.3b).

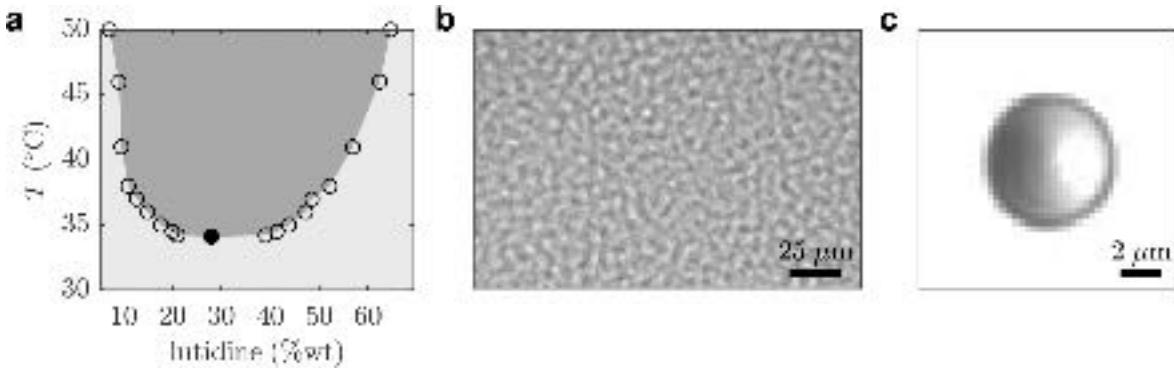


Figure 28.3: Phase behavior of water–2,6-lutidine. (a) Phase diagram of the water–2,6-lutidine mixture in the vicinity of the lower critical point (filled symbol). Data taken from [15]. (b) Spinodal decomposition of the binary mixture when heated above the critical point. (c) Snapshot of a carbon-coated silica Janus particle acquired with brightfield microscopy.

When the critical mixture is kept in the mixed phase with a temperature $T < T_c$, selective heating on one side of a suspended Janus particle above T_c induces local demixing. The Janus particles that generate such temperature difference are particles with a coating of a light-absorbing material on one hemisphere (Fig. 28.3c). In early experiments [13, 16], gold was used for the coating and later replaced by carbon to reduce van der Waals forces between particles. When particles are illuminated with a highly defocused laser beam of intensity I , light absorption on the coating leads to a heating $\Delta T \propto I$ [17]. The resulting temperature profile has its maximum on the coated surface and decays asymmetrically. If the laser beam intensity I , and thus ΔT is large enough to reach the critical temperature T_c , the liquid demixes within the isotherm $T = T_c$. Such asymmetric demixed zone aligns with the intrinsic axis of the Janus particle and induces a concentration flow that propels the particle in the direction also parallel to the intrinsic axis of the particle. The details of the propulsion mechanism are further explained theoretically and experimentally in Refs. [18] and [17] (the latter in a critical binary mixture of water and PnP (propylene glycol n-propyl ether), which shows a similar behavior as water–2,6-lutidine).

28.2 Janus particles in critical mixtures

A Janus particle with different absorbance between the two hemispheres propels by inducing a temperature gradient across the critical point of a binary mixture. Such mechanism is discussed in the previous section. Particles propelled by absorbing light can be manipulated by controlling the light intensity. In this section, we continue to introduce the experiment of active motion with the $6.2\text{ }\mu\text{m}$ carbon-coated Janus particles suspended in a water–2,6-lutidine mixture. These colloidal samples are prepared by mixing the coated particles with the

binary water–2,6-lutidine mixture and filling the suspension into the sample cell with a thickness of $200\text{ }\mu\text{m}$, which is then sealed with parafilm for higher stability and enhanced lifetime. Later, the sample cell is kept on a copper stage with temperature control inside an optical setup. The temperature of the sample and the laser intensity projected through the optics ensures that ΔT is sufficiently large for self-propulsion.

In the following sections, we describe three optical setups that each provide different levels of active motion control, i.e., ABP-like motion, active aligning torque, and individual steering of such Janus particles.

28.2.1 Simple active motion

When illuminated with a homogeneous light field with sufficient intensity, the carbon-coated silica Janus colloids in a near-LCP water–2,6-lutidine solution start self-propelling. In Fig. 28.4a, we introduce one realization of such an experiment in a commercial microscope with minimal modification. Here, we make use of the Gaussian profile of a typical laser beam. By expanding the beam, we obtain a rather flat intensity distribution near the center of the beam. With this intensity field, the resulting active motion of the Janus particles is ABP-like since the particles experience no active torque.

Compared with other ABP-like particle systems, our Janus colloids allow the control of their speed through changing the light intensity. With the sample properly sealed, the active motion of the Janus particles only depends on the temperature of the system and the provided light intensity. Because of that, it is possible to perform experiments with constant particle speed for hours or days with the undisturbed colloidal sample, which is tremendously hard for, e.g., Pt-coated Janus particles in hydrogen-peroxide solutions, where hydrogen-peroxide concentration decreases with time. Moreover, the manipulation of the propulsion speed is possible with the control of laser intensity [19]. We can also turn on and off the active motion at any instance, which might provide interesting features in, for example, viscoelastic environments.

The propulsion speed as a function of illumination intensity in Fig. 28.5 [17] shows three regimes: linear, saturated, and reversed regime. For experiments with best stability, the saturated regime provides best resistance

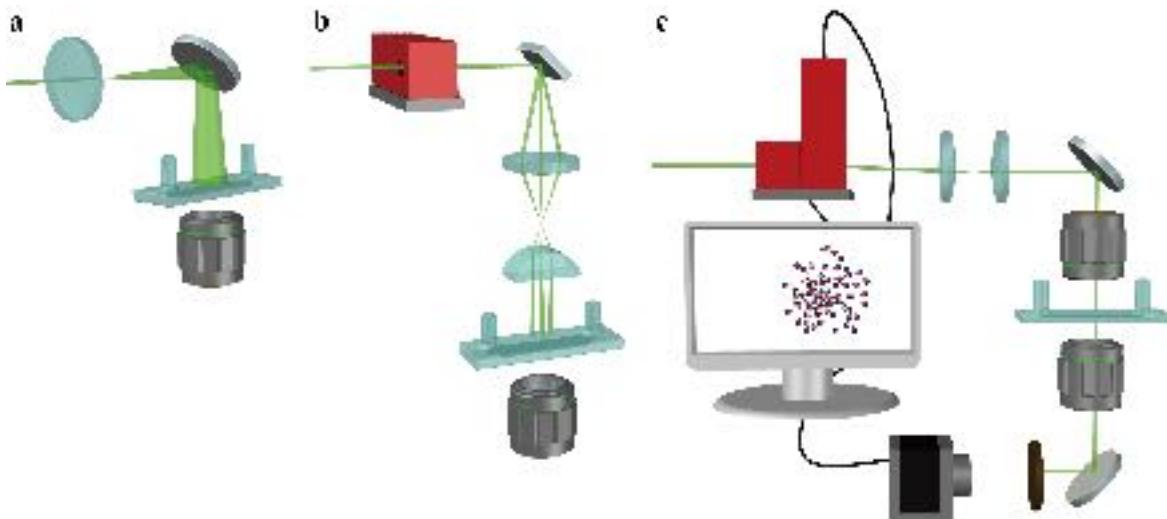


Figure 28.4: [the figure needs to be labelled with the various elements specified] **Experimental setups for Janus particles activated by light.** Sketch of three possible experimental setups for experiments with Janus particles suspended in a water–2,6-lutidine mixture at the critical concentration, in increasing order of complexity. **a** Experimental setup for ABP-like motion, where a homogeneous illumination is generated in the sample. The homogeneous illumination activates the Janus particles and sets the self-propulsion speed. The particle orientation is subject to rotational diffusion. **b** Experimental setup for active aligning torque, where a gradient of intensity is generated in order to set a preferential orientation of motion for the particles. **c** Experimental setup for individual control of the Janus particles. Each particle's speed and orientation of motion is where the light intensity is localized at each Janus particle thanks to a real-time feedback loop. [Check last sentence for clarity]

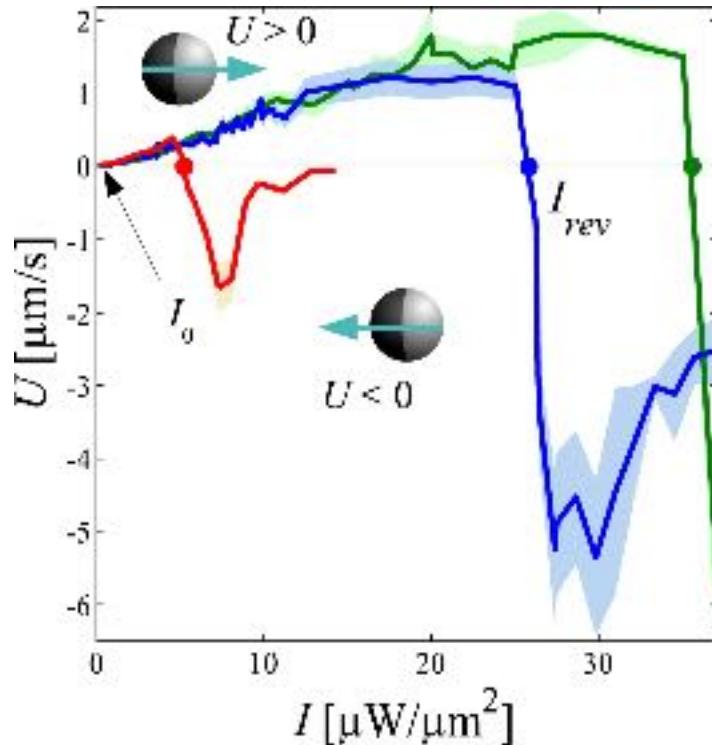


Figure 28.5: Propulsion speed vs optical intensity. [@Agnese: change the notation in the figure accordingly.] Dependence of the propulsion speed v as a function of the heating intensity I for particles of radius $R = 3.14 \mu\text{m}$ (green line), $R = 3.88 \mu\text{m}$ (blue line), and $R = 11.78 \mu\text{m}$ (red line) in water–PnP (propylene glycol n-propyl ether) solution at bath temperature $T = 25^\circ\text{C}$. The shaded areas represent the corresponding errors bars, computed over different particles in the same sample.

to fluctuations of intensity. As for experiments in need of speed variation, the linear regime can be handy when adjusting the speed proportionally. If, for some reasons, the maximum speed in the saturated regime is not sufficient, an even faster propulsion can occur in the reversed regime. The relation between the intensity and propulsion speed also depends on other parameters in the system, e.g., the diameter of the Janus particle. Experimental results and discussions regarding the reversal of propulsion direction are present in Ref. [17].

28.2.2 Active alignment and phototaxis

Reorientation occurs when a Janus particle experiences unbalanced propulsion across the axis of orientation; in our case, a light field with an intensity gradient provides such a condition. When a carbon-coated Janus particle in a water–2,6-lutidine solution is not aligned with the light gradient ∇I , this Janus particle features active reorientation [20]. Due to the gradient, the absorbed intensity (and hence the heating) is not homogeneous on the coated hemisphere, making the demixed zone asymmetric even in the perpendicular axis of the intrinsic axis of the particle. This results in a viscous torque Γ acting on the particle which is given by

$$\Gamma(\Theta) = -\Gamma_{\max} \sin \Theta, \quad (28.1)$$

where Θ is the angle between the particle orientation θ and the negative direction of the gradient $-\nabla I$, and Γ_{\max} is the maximum torque when $\Theta = \pm\pi/2$, which depends on the intensity I and the steepness of the gradient $|\nabla I|$. According to Eq. 28.1, the Janus particles reorient towards $\Theta = 0$ where the restoring torque vanishes, i.e., $\Gamma(0) = 0$. In Fig. 28.6 the active alignment is shown with a particle anti-aligned with the preferred direction $(-\nabla I)$ and reoriented through time. The maximum angular velocity ω_{\max} , hence Γ_{\max} , occurs at an orientation θ perpendicular to ∇I .

In summary, these Janus particles under light gradient perform negative phototactic motion as they reorient towards the direction of lower intensity. Fig. 28.4b shows a schematic of optical setup that was used in Ref. [20].

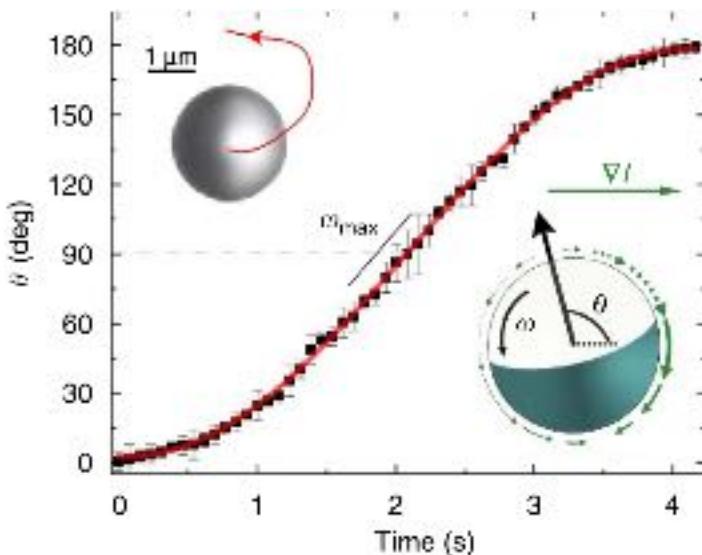


Figure 28.6: Orientation in a gradient of intensity. Time evolution of the orientation θ of an active Janus particle (upper inset) in the presence of an intensity gradient. Initially, the particle is oriented as represented in the inset, with its coated cap on the left. The initial orientation corresponds to $\theta = 0$. The gradient of intensity ∇I is directed towards the right (i.e., the intensity increases from left to right). The particle, subject to the intensity gradient, performs the motion indicated by the red trajectory, re-orienting itself in the opposite direction, i.e., with its coated end pointing towards the direction of increasing intensity. The constant intensity gradient magnitude is $\nabla I = 0.02 \mu\text{W} \mu\text{m}^{-3}$ and the peak intensity is $I = 0.55 \mu\text{W} \mu\text{m}^{-2}$. The data are averaged over 10 experimental runs. Upper inset: a trajectory of a single experiment. Lower inset: sketch of an active colloid in a non-uniform light field with gradient ∇I . The slip velocity of the fluid near the particle surface (green arrows) becomes axially asymmetric, when the particle orientation is at an angle θ with respect to the gradient, resulting in an angular velocity ω .

The synchronized galvano mirror and intensity modulator create a scanning line with linear gradient which expands after the cylindrical lens into a sheet of one-dimensional gradient. It is important that the galvano mirror scans with a sufficiently high frequency, so the intensity at the sample can be considered to be quasi-static.

28.2.3 Individual control

With localized light intensity at each Janus particle and a real-time feedback loop controlling such intensity fields, individual control of particle motion is possible. When we illuminate a particle at its geometrical center with a laser beam (Gaussian intensity profile) that only covers that individual particle, this particle propels with no aligning torque as the intensity gradient ∇I is symmetric to the particle orientation [21]. By adjusting the intensity of the laser beam that illuminates each particle, we can achieve individual particle control in the amplitude of the active motion. As the particle moves, each laser beam also needs to be reconfigured to illuminate the correct position regarding the new particle position. This requires real-time particle detection and a feedback loop of optical control. The particle detection and corresponding image processing techniques are discussed in Section 28.3.

One example of such experiment [21] uses a single laser beam that scan through every particle. As long as the period of one scan through all particles is sufficiently short, the active motion of the particles is equivalent to that by continuous illumination. In Fig. 28.4c, the schematics shows a two-axial acousto-optical deflector (AOD) that tunes the position and intensity of the laser beam at the sample. A computer acquires and analyzes the image of the particles in real time and reconfigures the AOD driver with the new laser positions.

The same setup, in Fig. 28.4c, can also achieve individual particle steering [22]. The Gaussian profile of the laser beam can also provide an intensity gradient ∇I to generate an aligning torque by simply offsetting

the position of the beam from the particle center. The resulting particle motion fits to the negative phototactic behavior of such a particle under a one-dimensional gradient. When implementing a steering direction with the laser offset to the particle center, from Eq. 28.1, an effective aligning potential

$$V(\Theta) = - \int \Gamma(\Theta) d\Theta = -\Gamma_{\max} \cos \Theta \quad (28.2)$$

can be defined. Due to rotational diffusion, the particle orientation Θ fluctuates even after the particle has reoriented into the steering direction, i.e., $\Theta = 0$. According to the Boltzmann distribution, the probability for the misalignment Θ is then

$$P(\Theta) \propto \exp\left(-\frac{V(\Theta)}{k_B T}\right) = \exp\left(\frac{\Gamma_{\max} \cos \Theta}{k_B T}\right), \quad (28.3)$$

which corresponds to the particle orientation under such active steering. Combining the two kinds of individual control experiments, the typical feedback control in Ref. [22] consists of the following three steps:

- *Image acquisition and particle detection:* An image of the Janus particles is recorded by video microscopy. Particle positions and orientations are obtained from this image (Section 28.3).
- *Calculation of next-step particle motion:* Depending on the experiment design, the desired motion for each active particle needs to be calculated according to the instantaneous condition. Based on the result, the propulsion velocity $v_{i,t}$ and torque $\Gamma_{i,t}$ are assigned to individual particles and then converted into corresponding laser powers and positions.
- *Illumination sequence:* A defocused laser beam scans through all the laser positions with the corresponding powers for each active particle. The power and position of the laser beam in the colloidal sample is tuned by an AOD, which manipulates the laser beam with RF-input signals and can scan through all particles sufficiently fast to ensure at all effects the simultaneous motions of all the particles.

The control of both propulsion speed and reorientation torque allows explorations of much complex interactions between particles, e.g., Vicsek-like models [23] and other social interaction rules [24], or virtual landscapes, e.g., a repulsive confinement [25].

28.3 Image processing and active particle detection

The detection and analysis for the positions and orientations of the active Janus particles is the very first step to quantify the experimental system. In this section, we focus on the image processing and object detection techniques that work especially for Janus particles and our application. Notably, instantaneous particle positions and orientations are information not only essentially for most post-experimental analysis, but also for feedback-controlled experiments (see Section 28.2.3).

In feedback-controlled experiments, the applicability of a particle detection method depends on two competing features: its precision and robustness on one side, and the speed of the algorithm on the other. In general, the particle detection gets more complicated and time-consuming as the number of particles in view increases. Especially when particles collide, i.e., when the distance between their contours vanishes, not all detection methods reliably detect the particles. Here, we introduce the method implemented in Ref. [22], which considered both speed and robustness of the detection. This method consists of, first, particle detection for the positions and, second, computation for orientations at the positions of the particles.

28.3.1 Position detection

The Janus particle has non-symmetric profile across the particle body in the video microscopy images (Fig. 28.7a) and hence requires detection by its geometrical shape. The typical fast particle tracking by pinpointing the maximum (or minimum) point of the images for emulsion or homogeneous transparent microspheres does not apply in such a case. However, the detection through particle shape is still applicable. Under the experimental conditions, particles have a dark, circular ring around them in the video microscopy images (Fig. 28.7a). This feature allows the position detection by a circular Hough transform based on phase-coding [26]. Similar approaches

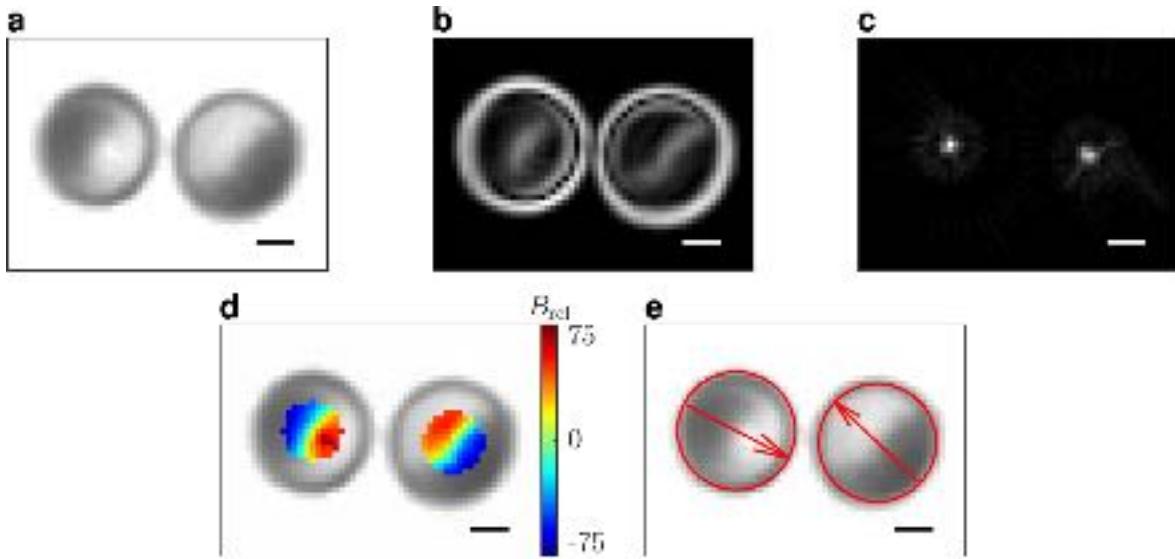


Figure 28.7: Particle detection. (a) Snapshot of two half-coated Janus particles. (b) Gradient image calculated from the snapshot in a. (c) Accumulator resulting from the circular Hough transformation. From the accumulation, one obtains the position of the particles’ centers. (d) The central region of the half-coated Janus particles (colored area) is used for the orientation detection. The color code corresponds to the relative brightness (B_{rel}) of the colored pixels. [©Chun-Jen what is the scale measuring? what brightness does it represent? The intensity of the image in pixels? Or another quantity?] (e) Result of the particle detection. The arrows indicate the Janus particles’ orientation. The scale bars are $2 \mu\text{m}$.

are embedded in, e.g., the `imfindcircles` function of MATLAB [27] and the `HoughCircles` function in the OpenCV Python package. Here, we take our adapted `imfindcircles` function for example: in the first step, the image is convolved with a Sobel filter in the x - and y -directions and the resulting directional gradients are combined into a gradient image (Fig. 28.7b). A threshold is used on the gradient image to include only pixels where the gradient is sufficiently large for speeding up the computation. Each of the remaining pixels contributes in the calculation of the so-called accumulator (Fig. 28.7c). Based on the direction of the gradient at each pixel, a circle center is estimated for a reasonable range of radii, e.g., the radius of the Janus particles R with tolerance. The accumulator then accumulates the coordinates of these centers. This allows for the detection of circles with different radii, which can be present due to polydispersity of the particle size. Finally, the localization of the peaks in the accumulator gives the center positions of the Janus particles.

28.3.2 Orientation detection

Close to the centre of a carbon-coated particle, the brightness of the image increases from the coated side to the uncoated side. Using the pixels that are closer than approximately half of the particle radius to the particle centre (colored area in Fig. 28.7d), the particle orientation can be obtained by doing a principle component analysis of the brightness with respect to the x - and y -position. Projecting the eigenvector of the largest eigenvalue to the xy -plane gives the orientation of the particle (Fig. 28.7e).

28.4 Introduction to the motion analysis of active particles

[©Felix could you check the following short paragraph?] In the previous section, we have seen how to detect the particle position and orientation in single frames of an experimental acquisition. The detected positions and orientations are then linked together, particle by particle, to obtain trajectories. These have to be regarded as raw data that can be affected by noise and artifacts. In this section, we show how to extract useful information from these raw trajectories.

In the following we assume that the active particles have a radius R , a fixed propulsion speed v_0 , a translational diffusion coefficient D , and that their orientation is bound into the plane (2D), so that they diffuse with a rotational diffusion coefficient D_R . In practice, most experimental systems have a 2.5D particle orientation, meaning that it is not completely locked parallel into the experimental plan, but not fully 3D either.

28.4.1 Trajectory and velocity

The first step is to get the self-propulsion of the particle from the raw trajectories. This is done through the following three steps:

1. Getting the quantity $\dot{\mathbf{r}}(t)$ from the raw trajectories (see Box 28.4.1).
2. Removing the drift from the raw trajectories to obtain trajectories without drift (see Box 28.4.1) [I am not sure I agree with this messaging. The sample should be as much as possible drift free in the making].
3. Smoothing the trajectories to obtain the particles propulsion speed (see Box 28.4.1).

[@Agnese, shouldn't the boxes be labelled differently and the label also be included in the box?]

Getting $\dot{\mathbf{r}}(t)$: The cleanest way to obtain $\dot{\mathbf{r}}(t)$ is to use a 3-point derivative of the trajectory:

$$\dot{\mathbf{r}}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t}, \quad (28.4)$$

where Δt is typically the acquisition time-step. In this way, the positions and displacements of the active particles are both well defined, and you make sure there is no *delay* effect, that can be introduced with a simpler 2-point derivative.

Removing the drift: Drift is one of the biggest enemy of anybody working with active particles. For some measurements, even a very small amount of drift — negligible compared to the propulsion speed — can have some tremendous impact. For a system with a single particle, there is no easy way to remove the drift, since we cannot differentiate the displacement that comes from the diffusion of the particle from the drift advection. However, if one uses many active particles, we know that we should have, in the absence of external forces (drift),

$$\langle \dot{\mathbf{r}}(t) \rangle_i = 0. \quad (28.5)$$

We can thus compute a *drift trajectory* by integrating the average velocity of the system for every time-steps:

$$\mathbf{r}_{\text{drift}}(t) = \int_0^t \langle \dot{\mathbf{r}}(t') \rangle_i dt'. \quad (28.6)$$

And we then remove the drift contribution from the trajectories:

$$\mathbf{r}_i^{\text{nodrift}}(t) = \mathbf{r}_i(t) - \mathbf{r}_{\text{drift}}(t). \quad (28.7)$$

Smoothing the trajectories: Active Brownian particles exhibit both ballistic (active) and diffusive (Brownian) motion. As a result, the instantaneous velocity $\mathbf{v}(t)$ of the particle is complicated to measure, i.e., $\dot{\mathbf{r}}(t) \neq \mathbf{v}(t)$. To deal with this, one usually applies some smoothing on the trajectories. At short timescales ($\Delta t \approx \tau$, where $\tau = 1/D_R$) [This is correct for $\Delta t \approx \tau$, it is Brownian diffusive at short timescales], the mean square displacement of an active Brownian particle can be averaged to

$$\langle \Delta \mathbf{r}^2 \rangle = 4D\Delta t + v^2\Delta t^2, \quad (28.8)$$

Dividing everything by Δt^2 leads to

$$\frac{\langle \Delta \mathbf{r}^2 \rangle}{\Delta t^2} = \frac{4D}{\Delta t} + v_0^2. \quad (28.9)$$

As a result, we can see that to get a well-defined measure of v_0 , the self-propulsion speed of the particle,

we need to have $\frac{4D}{\Delta t} \ll v_0^2$. In practice, it means that we need to smooth the trajectories over a time t_{sh} so that

$$\tau \gg t_{\text{sh}} \gg \frac{4D}{v_0^2}. \quad (28.10)$$

Also, since D_R and D are physically related to each other, the previous condition is only possible if

$$v_0 \gg \frac{D}{R}. \quad (28.11)$$

This is mostly the case for all active micro-scale swimmers, but might become not true at the nano-scale, or for very small velocities. If this condition is not fulfilled, diffusive motion dominates ballistic motion in the system, and therefore instantaneous velocity has no real meaning. Now that we know the strength of our smoothing, we can then apply a Gaussian filtering on the trajectories to obtain the velocity:

$$\mathbf{v}(t) = \sum_{i=-n}^n \dot{\mathbf{r}}(t + i\Delta t) G(i), \quad (28.12)$$

where $G(i)$ is a Gaussian kernel with standard deviation $\sigma = \frac{t_{\text{sh}}}{2\Delta t}$. Since the filtering will result in cutting out data-points at the beginning and at the end of the trajectories, we do not want to have $n \rightarrow \infty$. Thus we will limit the size of our Gaussian kernel, typically so that $2n = 4\sigma$, as the weights become negligible afterwards. Note that the kernel needs to be normalized so that $\sum_{i=-n}^n G(i) = 1$. Then we should have

$$\langle v \rangle_{i,t} \sim v_0, \quad (28.13)$$

where v_0 is the particles propulsion speed.

28.4.2 Velocity and orientation autocorrelation

Now that we have our trajectories and velocities, we can start to characterize the system. First, we measure $\tau = \frac{1}{D_R}$, the rotational diffusion time. The easiest method is to compute the velocity autocorrelation function $C_v(\Delta t)$:

$$C_v(\Delta t) = \langle \mathbf{v}(t) \cdot \mathbf{v}(t - \Delta t) \rangle_{i,t} = v_0^2 \exp(-D_R^v \Delta t). \quad (28.14)$$

In practice, reorientation of self-propelled particles can also be *active*, meaning that the propulsion mechanism can affect the rotational diffusion of the particle. Also the orientation itself is usually not purely bound in the 2D plane, adding another layer of complexity to the problem of determining D_R . Overall, while the theoretical value of D_R and the experimental one you measure with the autocorrelation function D_R^v can be different, they still should have the same order of magnitude (if not, it is a sign of strong active reorientation). In the following, D_R^v is the value you should use with other observables.

If one has a direct access to the particle orientation $\theta(t)$, it is then possible to compute the orientation self-correlation:

$$C_\theta(\Delta t) = \exp(-D_R^\theta \Delta t). \quad (28.15)$$

Interestingly, comparing D_R^v and D_R^θ gives us information regarding the ideality of the experimental system. On the one hand, since θ provides only 2D information, we should have $D_R^v = D_R^\theta$ if the particle's orientation is bound to the 2D plane. On the other hand, if the orientation is completely free in 3D, then we should have $D_R^v = 2D_R^\theta$. As mentioned previously, real experimental systems usually lie between these two extreme cases.

28.4.3 Mean square displacement

To characterize the particles displacement, the easiest tool is generally the time mean square displacement (MSD):

$$\text{MSD}(\Delta t) = \langle \Delta r^2(\Delta t) \rangle = \frac{1}{N} \sum_{i=1}^N (\Delta \mathbf{r}_i(t + \Delta t) - \Delta \mathbf{r}_i(t))^2. \quad (28.16)$$

What makes the MSD so powerful is that it allows to characterize the (average) motion of the particles at every times and length scale. For a typical active Brownian particle [28], the MSD reads:

$$\text{MSD}(\Delta t) = \left(4D + \frac{2v_0^2}{D_R} \right) \Delta t + \frac{2v_0^2}{D_R^2} (\exp(-D_R \Delta t) - 1). \quad (28.17)$$

The particle's motion is then typically diffusive at very short time scales, where the Brownian diffusion term,

$$\text{MSD}(\Delta t) \sim 4D\Delta t, \quad (28.18)$$

dominates. At intermediate time scales, the motion becomes ballistic and the active propulsion,

$$\text{MSD}(\Delta t) \sim v_0^2 \Delta t^2, \quad (28.19)$$

dominates. Finally, at large time scales, the particle can reorient and the motion becomes diffusive again:

$$\text{MSD}(\Delta t) \sim \left(4D + \frac{2v_0^2}{D_R} \right) \Delta t. \quad (28.20)$$

This last regime is often named *effective diffusion*, with a coefficient

$$D_{\text{eff}} = D + \frac{v_0^2}{2D_R}. \quad (28.21)$$

Fitting the full MSD curve is usually complicated, but in theory allows to characterize D , D_R , and v_0 at once.

28.4.4 Pair correlation function

After time correlations, we can investigate also spatial correlations. While the previous functions give information on the particles displacement and dynamics, spatial correlations aim to measure the organization of the system. The most basic tool for that is the pair correlation function $g(r)$, which measures the probability to find a particle at a distance r from another:

$$g(r_{ij}) = \frac{1}{\rho} \left\langle \sum_{i \neq j} \delta(r_i - r_j) \right\rangle_{i,j}. \quad (28.22)$$

[@Chun-Jen, @Felix: how is ρ defined?] Could you double check the equation above? Shouldn't it be something like:

$$g(r) = \frac{1}{\rho} \left\langle \sum_{i \neq j} \delta(|\mathbf{r}_i - \mathbf{r}_j|) \right\rangle$$

?

The pair-correlation function is directly linked (through a Fourier transform) to the structure factor of the system:

$$S(q) = 1 + \rho \int_r e^{-iqr} g(r) dr, \quad (28.23)$$

which better highlights correlations at short length scales (while the structure factor does the opposite). Since active systems generally do not exhibit long-range order, the former is usually favored. In a semi-dilute system, the first peak of $g(r)$ allows to measure an effective radius $R_{\text{eff}} > R_{\text{phys}}$, larger than the physical radius of the particles [How to you estimate the effective radius from the first peak of g\(r\)?](#). This effective radius corresponds to the minimum distance between two particles, and is particularly relevant when computing the area fraction of the system (especially for density-sensitive systems such as a glass). The height of the first peak and the presence (or absence) of secondary peaks allow to highlight the presence of short-range order (such as clusters) and its length-scale. Finally, in the dilute limit (and at equilibrium), $g(r)$ allows to measure the pair-interaction $U(r)$ using

$$U(r) = -\log(g(r)). \quad (28.24)$$

For active systems, this is normally wrong, since the system is out of equilibrium and does not follow Boltzmann statistics. One can still define a $U_{\text{eff}}(r)$ function, that typically depends on the system's activity (v_0 and D_R) and density. Note that for active systems, the combination of steric interaction and finite reorientation time generally leads to an effective attraction [29], which can lead to clustering or even motility induced phase separation (MIPS) [30] (see Chapter 11).

28.4.5 Polar and rotational order parameters

Active systems are known to exhibit collective motion, such as clustering, flocking, and active turbulence. To characterize global motion, we need to measure the order of these collective structures. We will highlight two important active order parameters amongst the many possible ones. First is the *polar order parameter*:

$$O_P = \frac{1}{N} \left| \sum_{i=1}^N \mathbf{v}_i \right|, \quad (28.25)$$

which measures how *flocky* the system is. This notably makes sense for Vicsek-like [23] systems, with particles' interactions that lead to speed alignment. Note that one can easily modify this order parameter to create a nematic one.

Conversely, one can define a *rotational order parameter*:

$$O_R = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i \times \mathbf{v}_i) \cdot \mathbf{e}_z, \quad (28.26)$$

which measures how *swirly* the system is, and the sense of rotation. Such swirls are often observed in experimental systems, as it is the easiest way to have collective motion without any global translation. Generally, collective behavior can exhibit a combination of both effects.

28.5 Problems

[@Chun-Jen, @Felix: the exercise sheet has some lines of code in Matlab. Would you provide scripts for reference?]

[@Chun-Jen, @Felix: could you provide the movies?]

Problem 28.1: Dual well. [@Chun-Jen, @Felix: could you describe shortly what the movie represents?] For this problem, you need the following information: particle size $R = 1.36 \mu\text{m}$; [size or radius?] pixel ratio 41.84 nm px^{-1} ; acquisition rate 200 fps; temperature $T = 298.15 \text{ K}$.

- a. Load the movie `DualWell.avi`.
- b. Detect the particle position. [Hint: You can use the `pkfnd` and `cntrd2` functions to detect the particle position.]
- c. Plot the $x(t)$ and $y(t)$ trajectories obtained from the tracking.
- d. Compute the probability density function $P(x, y)$ of the system.
- e. What do you notice from $P(x, y)$? Try to measure the trap stiffness and the barrier height of the system.
- f. Compute the MSD of this single trajectory. Try to measure the trap stiffness using the MSD.

Problem 28.2: Active Brownian motion. [@Chun-Jen, @Felix: could you describe shortly what the movie represents?] For this problem, you need the following information: particle size $2R = 6.2 \mu\text{m}$; pixel ratio 244 nm px^{-1} ; acquisition rate 4 fps; fluid viscosity $\eta = 4 \times 10^{-3} \text{ Pas}$; temperature $T = 300.65 \text{ K}$.

- a. Load the movie `ABP.avi`.
- b. Detect the particle position. [Hint: you can use the `ABP_detectEXP.m` function (provided at the link: <http://>) to detect the particle position.]
- c. Detect the particle orientation. [Hint: you can use the `get_polarity.m` function (provided at the link: <http://>) to detect the particle orientation.]
- d. Reconstruct the particle trajectories. [Hint: You can use the provided track `[trace?]` function `slice.m` (provided at the link: <http://>) to reconstruct the trajectories.]
- e. Smooth the trajectories as explained in Subsection 28.4.1 to measure the average propulsion speed.
- f. Compute the velocity autocorrelation function (Eq. (28.14)) to measure D_R^v , the value of the rotational diffusion coefficient.
- g. Compute the orientation autocorrelation function (Eq. (28.15)) to measure D_R^θ . What can you conclude regarding the orientation of the particles? [Hint: See discussion in Subsection 28.4.2.]
- h. Compare the values D_R^v and D_R^θ with the theoretical value of D_R^v from the theory: $D_R^v = \frac{k_B T}{8\pi\eta R^3}$.

- i. Compute the MSD and measure D_{eff} (see Eqs. (28.16)-(28.21)). Check if it is consistent with the theoretical prediction.

Problem 28.3: Strong drift. [@Chun-Jen, @Felix: could you describe shortly what the movie represents?]

For this problem, you need the following information: particle size $2R = 6.2 \mu\text{m}$; pixel ratio 301 nmpx^{-1} ; acquisition rate 5 fps; temperature $T = 300.65 \text{ K}$.

- a. Load the movie `Drift.mp4`.
- b. Detect the particle position. [Hint: You can use the `ABP_detectEXP.m` function (provided at the link: <http://>) to detect the particle position.]
- c. Reconstruct the particle trajectories. [Hint: You can use the provided track `[trace?]` function `slice.m` (provided at the link: <http://>) to reconstruct the trajectories.]
- d. Compute the average velocity of the system: $\langle \dot{\mathbf{r}}(t) \rangle_i$ (Eq. (28.4)).
- e. Integrate the average velocity of the system to obtain the drift trajectory $\mathbf{r}_{\text{drift}}(t)$ (Eq. (28.6)).
- f. Remove the drift contribution from the particles trajectories and obtain $\mathbf{r}_i^{\text{nodrift}}(t)$ (Eq. (28.7)).
- g. Compare the particles' trajectories with and without drift.
- h. Compute and compare the MSD with and without drift.
- i. Measure the effective diffusion constant D_{eff} (see Eqs. (28.16)-(28.21)).

Problem 28.4: Swirl. [@Chun-Jen, @Felix: could you describe shortly what the movie represents?]

For this problem, you need the following information: particle size $2R = 6.3 \mu\text{m}$; pixel ratio 315 nmpx^{-1} ; acquisition rate 5 fps; temperature $T = 301.15 \text{ K}$.

- a. Load the movie `Swirl.mp4`.
- b. Detect the particle position. [Hint: You can use the `ABP_detectEXP.m` function (provided at the link: <http://>) to detect the particle position.]
- c. Reconstruct the particle trajectories. [Hint: You can use the provided track `[trace?]` function `slice.m` (provided at the link: <http://>) to reconstruct the trajectories.]
- d. Plot the particles trajectories.
- e. Measure the polar O_P and rotational O_R parameter through time (see Eqs. (28.25) and (28.26)).

Problem 28.5: Swirl. can this title of the problem be different from the previous one? [@Chun-Jen, @Felix: could you describe shortly what the movie represents?]

For this problem, you need the following information: particle size $2R = 1.1 \mu\text{m}$; pixel ratio 171 nmpx^{-1} ; acquisition rate 20fps; temperature $T = 293.15 \text{ K}$.

- a. Load the datafiles `MultiPart01.mat`, `MultiPart02.mat`, and `MultiPart03.mat`. The data are in the format of a trajectory matrix $[XYti]$, where X is the horizontal position (in pixels), Y is the vertical position (in pixels), t is the time, and i represent the particle index within the frame.
- b. Compute the pair correlation function $g(r)$ (Eq. (28.22)). What can you conclude about the structure of these systems?
- c. Compute the MSD (you should do it with a lot time array) [@Chun-Jen, @Felix: could you clarify?]. Which systems are active?
- d. Compute the local hexatic order parameter Ψ_6 and the orientation for one frame of the data in the file `MultiPart03.mat`. [@Chun-Jen, @Felix: could you clarify? You should also define the Ψ_6 . What do you mean by orientation?]
- e. Plot its modulus and its orientation. What can you say about the structure? [@Chun-Jen, @Felix: could you clarify?]

Bibliography

- [1] C. Bechinger, R. Di Leonardo, C. Löwen, H. and Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [2] R. Golestanian, T. B. Liverpool, and A. Ajdari. Propulsion of a molecular machine by asymmetric distribution of reaction products. *Physical Review Letters*, 94:220801, 2005.

- [3] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. Self-motile colloidal particles: From directed propulsion to random walk. *Physical Review Letters*, 99:048102, 2007.
- [4] S. Ni, J. Leemann, H. Wolf, and L. Isa. Insights into mechanisms of capillary assembly. *Faraday Discussions*, 181:225–242, 2015.
- [5] L. Alvarez, M. A. Fernandez-Rodriguez, A. Alegria, S. Arrese-Igor, K. Zhao, M. Kröger, and L. Isa. Reconfigurable artificial microswimmers with internal feedback. *Nature Communications*, 12:4762, 2021.
- [6] D. Dendukuri, D. C. Pregibon, J. Collins, T. A. Hatton, and P. S. Doyle. Continuous-flow lithography for high-throughput microparticle synthesis. *Nature Materials*, 5:365–369, 2006.
- [7] Z. Nie, W. Li, M. Seo, S. Xu, and E. Kumacheva. Janus and ternary particles generated by microfluidic synthesis: design, synthesis, and self-assembly. *Journal of the American Chemical Society*, 128:9408–9412, 2006.
- [8] N. Prasad, J. Perumal, C.-H. Choi, C.-S. Lee, and D.-P. Kim. Generation of monodisperse inorganic–organic Janus microspheres in a microfluidic device. *Advanced Functional Materials*, 19:1656–1662, 2009.
- [9] K. P. Yuet, D. K. Hwang, R. Haghgooie, and P. S. Doyle. Multifunctional superparamagnetic Janus particles. *Langmuir*, 26:4281–4287, 2010.
- [10] N. G. Min, T. M. Choi, and S.-H. Kim. Bicolored Janus microparticles created by phase separation in emulsion drops. *Macromolecular Chemistry and Physics*, 218:1600265, 2016.
- [11] Z.-Y. Feng, T.-T. Liu, Z.-T. Sang, Z.-S. Lin, X. Su, X.-T. Sun, H.-Z. Yang, T. Wang, and S. Guo. Microfluidic preparation of Janus microparticles with temperature and pH triggered degradation properties. *Frontiers in Bioengineering and Biotechnology*, 9:756758, 2021.
- [12] C. Casagrande, P. Fabre, E. Raphaël, and M. Veyssié. Janus beads: Realization and behaviour at water/oil interfaces. *Europhysics Letters*, 9:251, 1989.
- [13] G. Volpe, I. Buttinoni, D. Vogt, H.-J. Kümmeler, and C. Bechinger. Microswimmers in patterned environments. *Soft Matter*, 7:8810–8815, 2011.
- [14] R. J. L. Andon and J. D. Cox. 896. Phase relationships in the pyridine series. Part I. The miscibility of some pyridine homologues with water. *Journal of the Chemical Society*, pages 4601–4606, 1952.
- [15] C. A. Grattoni, R. A. Dawe, C. Y. Seah, and J. D. Gray. Lower critical solution coexistence curve and physical properties (density, viscosity, surface tension, and interfacial tension) of 2,6-lutidine+water. *Journal of Chemical and Engineering Data*, 38:516–519, 1993.
- [16] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger. Active Brownian motion tunable by light. *Journal of Physics: Condensed Matter*, 24:284129, 2012.
- [17] J. R. Gomez-Solano, S. Samin, C. Lozano, P. Ruedas-Batuecas, R. van Roij, and C. Bechinger. Tuning the motility and directionality of self-propelled colloids. *Scientific Reports*, 7:14891, 2017.
- [18] S. Samin and R. van Roij. Self-propulsion mechanism of active Janus particles in near-critical binary mixtures. *Physical Review Letters*, 115:188305, 2015.
- [19] N. Narinder, C. Bechinger, and J. R. Gomez-Solano. Memory-induced transition from a persistent random walk to circular motion for achiral microswimmers. *Physical Review Letters*, 121:078003, 2018.
- [20] C. Lozano, B. Ten Hagen, H. Löwen, and C. Bechinger. Phototaxis of synthetic microswimmers in optical landscapes. *Nature Communications*, 7:12828, 2016.
- [21] T. Bäuerle, A. Fischer, T. Speck, and C. Bechinger. Self-organization of active particles by quorum sensing rules. *Nature Communications*, 9:3232, 2018.

- [22] T. Bäuerle, R. C. Löffler, and C. Bechinger. Formation of stable and responsive collective states in suspensions of active colloids. *Nature Communications*, 11:2547, 2020.
- [23] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75:1226–1230, 1995.
- [24] C.-J. Chen and C. Bechinger. Collective response of microrobotic swarms to external threats. *New Journal of Physics*, 24:033001, 2022.
- [25] T. Knippenberg, A. Lüders, C. Lozano, P. Nielaba, and C. Bechinger. Role of cohesion in the flow of active particles through bottlenecks. *Scientific Reports*, 12:11525, 2022.
- [26] T. J. Atherton and D.J. Kerbyson. Size invariant circle detection. *Image and Vision Computing*, 17:795–803, 1999.
- [27] Inc. The Mathworks. Matlab, 2019. version 9.6.0.1150989 (R2019a). Natick, Massachusetts.
- [28] J. R. Howse, R. A. L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. Self-motile colloidal particles: From directed propulsion to random walk. *Physical Review Letters*, 99:048102, 2007.
- [29] F. Ginot, I. Theurkauff, D. Levis, C. Ybert, L. Bocquet, L. Berthier, and C. Cottin-Bizonne. Nonequilibrium equation of state in suspensions of active colloids. *Physical Review X*, 5:011004, 2015.
- [30] M. E. Cates and J. Tailleur. Motility-induced phase separation. *Annual Review of Condensed Matter Physics*, 6:219–244, 2015.

Chapter 29

Active Particles with Feedback Interactions

FRANK CICHOS

In the vibrant field of active particle systems, *feedback control* introduces a fascinating dimension by infusing elements of information exchange and system responsiveness, concepts inspired by natural feedback mechanisms in biological entities and commonplace technologies. In essence, every living organism is a complex conglomerate of feedback processes that commence at the single-cell level, as illustrated in Fig. 29.1. This Chapter presents the intricacies of feedback control in active particles, beginning with an examination of switchable microswimmers as a platform for a deeper exploration of the technical details behind the experimental setup and the governing equations of feedback-controlled active particles. The Chapter subsequently continues by

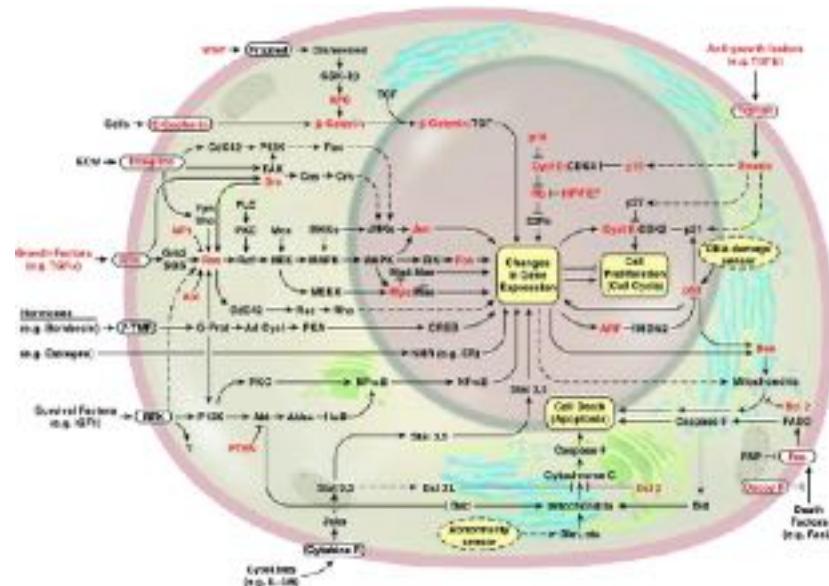


Figure 29.1: **Emergent integrated circuit within a cell.** The intricate signalling pathways within the cell mimic the complexity of electronic integrated circuits, with proteins such as kinases and phosphatases replacing transistors, and phosphates and lipids standing in for electrons. Highlighted pathways include growth signaling centered around Ras **What is this?**, antigrowth and differentiation signals, as well as commands regulating apoptosis. Genes known to be functionally altered in cancer cells are highlighted in red. Reproduced from Ref. [1].

discussing promising applications, such as studying active particle pressure, orchestrating information-based self-organization of active particles, and combining active particles with machine-learning procedures. The discussion illuminates the burgeoning area of thermophoretic control as a potent approach to implement fine control in active particles, underscoring the immense potential of feedback control in bringing about novel behaviors and applications in active particle systems.

Agnese: renumber figures. Also check and add infor about permissions.

Agnese: Please, add some exercises and problems based on the tutorial materials provided by Frank.

Editors: This is a specific form of control and other implementations are discussed in the Felix Ginot's chapter, for example. How do we discuss this or take into account?

29.1 Why feedback control of active particles

Why would one want to feedback control active particles? Feedback systems are everywhere in our daily life. No matter whether you drive a car or use your cell phone, many everyday objects are equipped with feedback systems, including sensors for the processing of information and for responding to it. The weather is an example of a natural feedback system, where different phenomena influence each other strongly and globally.

The term ‘feedback’ describes a scenario in which at least two dynamical systems are intimately connected, mutually influencing each other, typically in a non-linear manner. This potent coupling often obfuscates the causality, blurring the distinction of which system acts first. Thus, the traditional demarcation between cause and effect no longer holds. In many cases, these systems can also be mapped to critical systems (or even self-organized critical systems [2]): for example, brain activity [3, 4], earthquakes [5] or even climate change [6, 7] exhibit signatures of critical systems. We will not classify all different types of feedback, but the interested reader can find more details in the book by Murray and Åström [8].

Biological systems leverage feedback mechanisms extensively and in remarkably diverse ways. A prime example is *homeostasis* — the self-regulation of biological systems to maintain a specific state or function. This phenomenon spans various scales, from molecules and cells to entire organisms and ecosystems. For instance, the concentration of glucose in the blood is tightly regulated by insulin and glucagon in the pancreas, showcasing an intricate feedback loop. Numerous other examples abound, including those that manage gene expression, regulate body temperature, and orchestrate muscle activity to maintain posture while sitting and studying.

Returning to the context of active particles, our inspiration for creating synthetic microswimmers is rooted in biology. Cells, bacteria, insects, animals, and humans all serve as active entities that effectively transform energy into mechanical motion. This critical characteristic arguably positions them as the simplest forms of non-equilibrium systems — those with an ongoing energy flux. Biological feedback systems liberate entities from the constraints of Newton’s third law, as their capacity for sensing and reacting to environmental signals enables them to respond in a manner that is not symmetrically matched with the response of the signal-emitting object. This nuanced interaction is an essential feature in orchestrating complex biological processes. On the cellular level, feedback circuits are related to chemical signals, which induce physio-chemical reactions in complex chemical networks. At larger length scales, more complex signals, e.g., vision, sound or smell, are available and are processed by highly developed information processing units, e.g., neurons and brains. These are the interactions which enable flocks of birds, swarms of bacteria, swirls of ants, and even whole societies. Being a bit more physical, this natural invention is very important as it allows to control an ensemble of active objects, with a minimum amount of energy. Those low energy excitations are termed *Nambu–Goldstone modes* [9, 10, 11] in solid state physics and quantum field theory, and are responsible for symmetry breaking and the appearance of phase transitions. Many fundamental questions arise when looking at collective systems. Surprisingly, when characterizing swarms, for example, one finds that their structure is related to topological rules (number of neighbors vs neighboring distance) and their dynamics seem to follow simple scaling laws. All such analyses of collective biological states show that the systems are close to a state with a maximum in the susceptibility, i.e., they are close to a critical state, which is often interpreted as a high level of awareness.

In summary, there are a lot of reasons to introduce feedback control to active particles. It introduces the aspect of information exchange to the response of active particles. And, while we are not yet able to create complex responsive active particles, simple examples already exist, where we can introduce feedback with the help of an external processing unit (a computer) as already done in a number of experiments [12, 13, 14, 15, 16].

All of these experiments will bring us a bit closer to new artificial systems which are very much differently constructed than the systems we can currently manufacture.

29.2 Switchable thermophoretic swimmers

The basic ingredient for the feedback control of active particles is, of course, an active particle which can be controlled. There are a multitude of ways to control active particles in an ensemble with external fields [17]. The control we are referring to in this Chapter is acting on each particle separately. It may give each particle an identity, a specific behavior, and even a memory of what has happened. From this, it becomes obvious that the propulsion mechanism needs to be switchable, and one of the best ways to implement such level of control is by using light [18, 12, 19, 20]. There are again multiple pathways to use this feature. In the following, we will focus on thermophoretic control and, in the next section, we will briefly introduce the fundamentals behind this mechanism.

29.2.1 Thermo-osmosis vs thermophoresis

The most common physical phenomena driving microswimmers are phoretic effects. The one in which we are interested here is *thermophoresis*. Active particles will undergo self-thermophoresis as they carry the heat source (i.e., the energy source) with them. Yet, it is not phoresis which drives them. The underlying effect is actually called *thermo-osmosis*. Similarly, there are equivalent osmotic effects for other phoretic phenomena. Let us have a quick look at the difference between thermo-osmosis and thermophoresis.

Fig. 29.2a depicts a common osmotic pressure setup. A solute (small dots) is dissolved in a liquid with different concentrations on both sides of a semipermeable membrane (dashed line). This membrane will let the solvent pass but not the solute. As the system tends to equilibrate the concentrations, liquid will flow

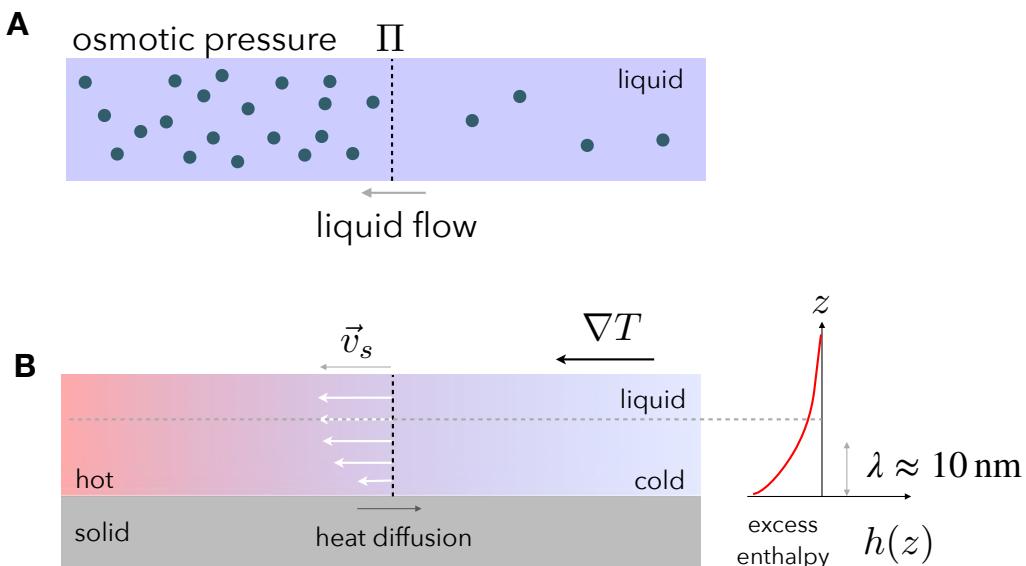


Figure 29.2: **Osmosis vs thermo-osmosis.** (a) *Osmotic pressure* is illustrated with a setup involving a semipermeable membrane. Here, a solute (represented by small dots) is dissolved in a liquid. The different concentrations of the solute on either side of the membrane (indicated by the dashed line) lead to osmotic pressure. The membrane permits the passage of the solvent but restricts the solute. (b) *Thermo-osmotic flow*, on the other hand, occurs along a solid–liquid interface and plays a crucial role in the motion of microswimmers. This is particularly relevant in the case of self-thermophoresis, where the active particles carry the heat source, their energy source, with them. The process driving this is thermo-osmosis. [Agnese:Change labels in figure.](#)

through the membrane and cause an increased osmotic pressure on the left side of the membrane, which could eventually equalize the concentrations on both sides.

Imagine now that we replace the solutes by heat, where the “concentration” of heat is different on the left and the right sides of a liquid film close to a solid surface, as shown in Fig. 29.2b. Heat would, therefore, also diffuse against the temperature gradient to the cold to equilibrate the temperature at each side. If there were now a semipermeable membrane that allows the solvent to pass but not the heat, the liquid would flow from the cold to the hot to equilibrate the temperature. Although there is no such membrane, the same effect is provided by a solid–liquid interface as the liquid has, in a tiny region close to the solid, an additional interaction with the solid, quantified by the so-called *excess enthalpy*. Taking a liquid volume element close to the solid from the cold side and exchanging it with one at the hot side would not only transport heat since the liquid volumes have different temperatures, but also because of additional free energy as the liquid has a different interaction with the solid in these regions. This excess enthalpy is causing a liquid flow at the interface, which is called *thermo-osmotic flow* [21, 22]. The thermo-osmotic flow velocity saturates at a distance from the boundary in the above drawing, which corresponds to the characteristic interaction length λ . This is due to the fact that we did not set any hydrodynamic boundary condition at infinity. Since this interaction range is typically very small, it is contracted to a velocity jump, that is called a quasi-slip velocity v_s given by

$$v_s = -\frac{1}{\eta} \int z h(z) dz \frac{\nabla T}{T} = \mu \nabla T, \quad (29.1)$$

where η is the viscosity of the fluid (assumed constant), $h(z)$ the excess enthalpy density, and T the temperature. The integral can be contracted to yield a mobility coefficient μ . This mobility coefficient, therefore, contains all the information about the interfacial interaction between liquid and solid, which could be quite complex. If $h(z) > 0$, the liquid is driven to the cold, while if $h(z) < 0$, the liquid is moving to the hot, which is the more common case.

These thermo-osmotic flows exist on all interfaces that carry a temperature gradient and have been directly measured only recently [21], even though they have been known for quite a while [22]. The interface could be either the substrate of a sample or the surface of a particle. In case the interface belongs to a mobile particle in the liquid and the interfacial flow is directed to the hot, a force balance between the particle and the liquid (including its hydrodynamic boundary condition at infinity) causes the particle to move to the cold. This migration of the particle is now called *thermophoresis* and drives the motion of our *self-thermophoretic particles*. *Thermo-osmosis transports the solvent, while thermophoresis transports the solute*. In both cases, the system is force-free, i.e., all forces are already balanced, so that there is no need to overcome friction. Everything moves because of friction.

The thermophoretic velocity of a particle in a temperature field is related to the above-defined mobility coefficient by Add reference for the derivation for the next equation

$$v_{th} = -\frac{2}{3} \mu \nabla T \quad (29.2)$$

as long as the temperature field is not distorted by the presence of the particle. In general, the prefactor in front of the temperature gradient is called thermo-diffusion coefficient D_T , which is a bit odd, since it does not denote the thermal diffusion but the temperature-induced phoretic mobility of the colloid. The factor 2/3 stems from the integration over a spherical surface.

29.2.2 Self-thermophoretic swimmer

Fig. 29.3 shows how this principle can be used to create self-thermophoretic swimmers by incorporating the heat sources into particles controlled by laser illumination. There are two main ways to do this. The first is to coat a colloidal particle hemispherically with a metal layer (or any other light-absorbing, heat-releasing material), as shown in Fig. 29.3a, so that the asymmetry of the structure provides the means to break the time symmetry of the hydrodynamic equations at low Reynolds numbers [23, 24]. Fig. 29.3b shows the second approach, which allows a more advanced control, where the particle is symmetric but the energy release is asymmetric due to an asymmetric illumination of the particle [25]. In both cases, the particle can be as small as several hundred nanometers to several micrometers.

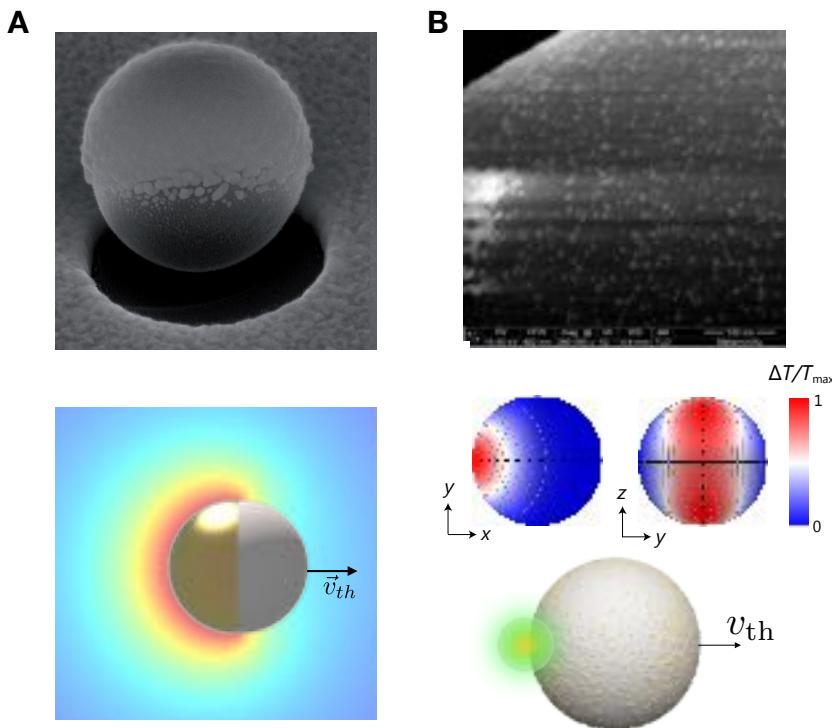


Figure 29.3: Two types of self-thermophoretic particles. (a) Janus particle consisting of a polymer core and a 50-nm-thick hemispherical gold cap. The temperature profile is qualitatively indicated below. (b) Melamine-resin particle with 10-nm gold nanoparticles decorating the whole surface. The surface temperature when illuminated asymmetrically with a focused laser is shown below. [Change labels in figure.](#)

29.3 Feedback control

Feedback control of microswimmers refers to the fact that we can observe the state of the swimmer (e.g., its orientation, speed, or position) and respond by modifying some of its properties (e.g., its propulsion speed, orientation, angular speed, noise). In the simplest case, feedback control requires three parts: an *input signal*, a *processing* of that input signal to determine an error signal, and a *coupling* of the input signal back to the original system.

29.3.1 Photon nudging

One of the simplest versions of active particle control is called *photon nudging*. In photon nudging, we can correlate the orientation of the particle with its propulsion velocity in a very simple way, as illustrated in Fig. 29.4 [12]. Assume that we define a target position in space to which the particle shall be propelled. We activate the propulsion whenever the orientation of the swimmers axis is towards the target and switch it off whenever it is pointing in other directions. [Agnew: invert panels for figure and change text accordingly.](#) To a certain extent, this resembles run-and-tumble motion, but there are also considerable differences to the motion observed for bacteria. As a particle essentially never points precisely in the direction of the target, photon nudging works with an acceptance angle (β_0). So the simplest photon nudging procedure is connected to a simple on/off switching of the laser if the target is in the acceptance angle range. Rotational diffusion with the diffusion coefficient $D_r = 1/\tau_r$ is doing the rest of the work by causing a reorientation of the particle axis. One may think of the whole procedure being a tiny Maxwell daemon sitting on the particle and enabling its dynamics.

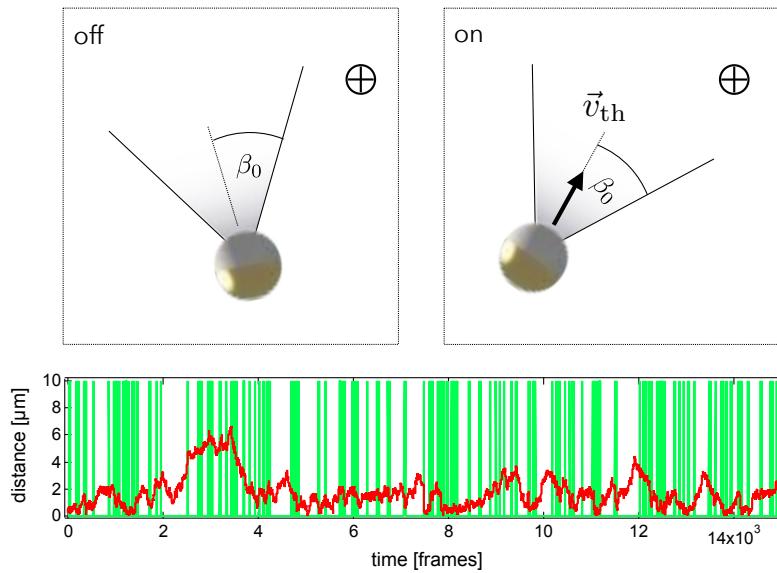


Figure 29.4: Principle of photon nudging. The particle is passive with rotational diffusion as long as the target position is not in a direction within the acceptance angle β_0 (*off state*, top left inset). Whenever the target position is within the acceptance angle, the propulsion is switched on (*on state*, top right inset). The stochastic switching of the laser propulsion is shown in the time trace below (green lines) together with a trajectory of an active particle showing the distance from the target position (red line).

29.3.2 Experimental setup

Fig. 29.5 shows a typical photon-nudging setup, where the input signal of the feedback loop stems from a CCD taking images of the sample in a microscopy setup with an inverse framerate of Δt . The processing involves the image analysis, which can be just centroid-based, or more complex if the orientation of the particle is also required. Even deep-learning models can be employed [26, 27]. For feedback control to work, the image analysis (including all additional processing steps) has to be faster than Δt , otherwise one might run into problems with, e.g., dropped frames. The response is carried out by steering and modulating a laser, e.g., with an *acousto-optic deflector* (AOD) or any other beam steering device (a *digital mirror device* or a *spatial light modulator*). The technical details depend very much on your particular application, but here are a few hints on beam steering and image processing.

Beam steering can be carried out with different devices:

Acousto-optic deflector (AOD). The AOD is a device driven with acoustic waves in crystals, where the frequency of the acoustic wave (around 100 MHz) creates an optical grating deflecting the beam to different diffraction orders. By tuning the acoustic frequency, one tunes the grating constant and thus the deflection of the beam. For 2D beam steering, one needs two orthogonal deflectors, where deflection angles of several degrees are achieved. Their advantage is a very fast response (microseconds). The disadvantage is that for multi-particle control, the particles have to be multiplexed, meaning that only one particle is active at a time.

Spatial light modulator (SLM). An SLM is typically a liquid crystal display allowing to introduce controlled phase shifts to the light wave and to modify the intensity pattern in some image plane. The desired phase pattern can be generated by an iterative procedure (e.g., using the *Gerchberg–Saxton algorithm*), which could be in complex cases quite time consuming and require a GPU. An SLM can create hundreds of foci to control active particles. These foci exist at the same time, so all particles can be propelled simultaneously. The speed of an SLM is comparable to typical frame rates (up to 100 Hz for fast ones).

Digital micromirror device (DMD). A DMD essentially contains an array of bi-stable mirrors. The light

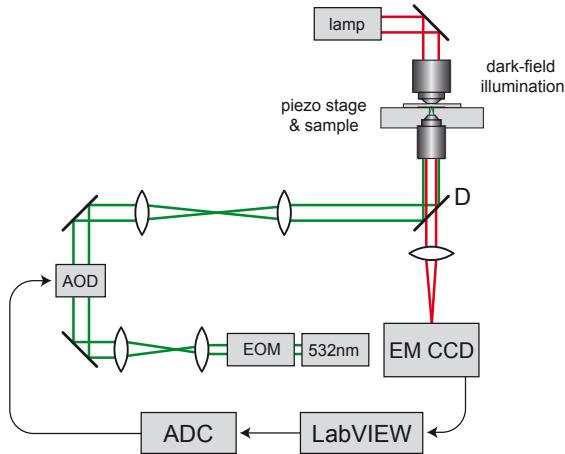


Figure 29.5: Photon-nudging experimental setup. Photon-nudging setup, featuring an intensified CCD camera capturing real-time images of active particles in a dark-field microscope. The images undergo real-time analysis in LabVIEW, which can range from a simple centroid-based approach to complex deep-learning models. A crucial aspect is that image analysis must occur faster than the camera's inverse framerate, Δt , to prevent frame dropping. The feedback response is executed by modulating and steering a laser with an acousto-optic deflector (AOD). The data from image processing allows for the introduction of correlations between the input and output signals, with the potential to correlate the propulsion of a particle with its orientation or velocity.

power falling on the mirror surface is, therefore, split into the power contributions of each individual pixel, which are spread across the image area in the focal plane of the objective lens. Compared to an SLM, where all incident light power is mainly distributed across the focal spots, the light of the pixels of the DMD in the off state is just lost. Nevertheless, the DMD are much faster than typical SLMs.

Image processing requires the accurate detection and identification of the positions and/or orientations of all active particles within a sample, a task that can be achieved through a multitude of methods. This computational undertaking can be carried out on specialized framebuffer cards, graphics cards, or, in more straightforward scenarios, directly on the computer's central processing unit (CPU). The choice of hardware depends largely on the complexity of the image analysis and the computational resources available. Hardware control is often carried out with LabVIEW, which, with NI Vision, provides a powerful module for image processing using *centroid-based particle localization* methods and also more advanced techniques, like the *Hough transforms*. When images contain more complex objects and are heterogeneous in the objects' intensity more advanced techniques, e.g., involving *deep learning* [26, 27] approaches are advantageous, but may need optimization to match the speed of your experiment. Especially methods like *deep reinforcement learning* with real-time experience replay can be rather demanding in data storage and computational effort and certainly require a well thought interface to GPUs. As benchmark numbers, the system described in Ref. [27] [Check reference](#) is capable of taking images at a framerate of 100Hz, but is typically run around 60Hz. Simple centroid-based particle detection can be done on timescales below 10ms for a 512px \times 512px image, even for large numbers of particles. The deep-learning approach in Ref. [27] [Check reference](#) is capable of localization and classification with a framerate larger than 50Hz with a custom made interface to a GPU.

The processing of the image information provides the freedom to introduce any type of correlation of the input signal with the output signal. For example, a correlation of the propulsion of the particle with the orientation of the particle in the lab frame can be easily achieved. But one can also correlate the propulsion velocity with the position in the lab frame. As the brain of the particle is the computer, one may also give each particle a memory to carry out actions based on past observations or future extrapolations. Therefore, one has to keep the identity of the particle over the experiment. Using a trajectory-linking algorithm, it is possible to provide each particle with a long memory to introduce delays or predictions to perform on the microscale

experiments similar to those presented in Ref. [28]. The system may even remember all its previous moves, which becomes important for a technique called “experience replay” during deep reinforcement learning. There is an immense freedom in this choice, which makes the control of active particles a very powerful method to study various physical processes.

29.3.3 Photon-nudging theory

To understand a bit better the results of feedback control and its relevance, we can have a closer look at a simple theoretical description of the photon-nudging process. We can understand the trajectories of single particles in photon nudging with the help of two coupled Langevin equations, where the particle’s speed v_{th} depends on time

$$\begin{cases} \frac{d\mathbf{r}(t)}{dt} = v_{\text{th}}(t) \hat{\mathbf{n}}(t) + \sqrt{2D} \boldsymbol{\zeta}(t) \\ \frac{d\hat{\mathbf{n}}(t)}{dt} = \sqrt{2D_r} \hat{\boldsymbol{\xi}}(t) \times \hat{\mathbf{n}}. \end{cases} \quad (29.3)$$

The first equation refers to the position \mathbf{r} of the particle: it contains the propulsion term with the direction of motion (unit vector $\hat{\mathbf{n}}$) and a term responsible for Brownian motion with the diffusion coefficient D and the white noise of unitary variance $\boldsymbol{\zeta}(t)$. The second equation is the Langevin equation for the orientation of the particle $\hat{\mathbf{n}}$, which is also performing Brownian motion with a rotational diffusion coefficient D_r , driven by another white noise term of unitary variance $\hat{\boldsymbol{\xi}}(t)$ [This term should be different from the rotational diffusion].

Agnese: Add as series of exercises to simulate a photon nudged particle based on this Jupyter notebook <https://mybinder.org/v2/gh/Molecular-Nanophotonics/Photon-Nudging/master?filepath=Photon%20Nudging.ipynb>. You can also add references to the relevant numerical chapters.

To quantify the quality of the control, we can however proceed in a different way that also provides some physical insight [29, 30]. Let us assume that we want to nudge a Janus particle to a specific target position ($x = y = 0$) in the plane. If we measure the stationary density distribution of trajectory points of the swimmer with a distance ρ from the target position, we can obtain a measure of the accuracy of the control with its variance $\sigma^2 = \sqrt{\langle \rho \rangle}$.

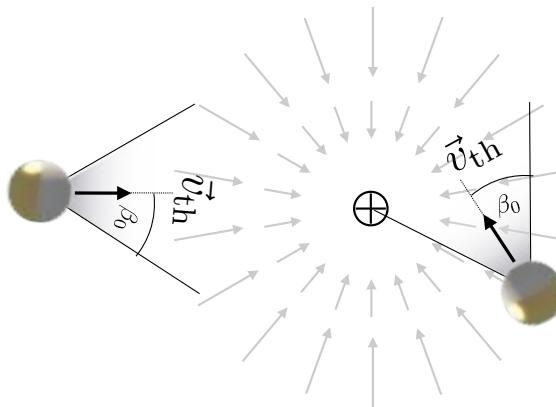


Figure 29.6: **Working principle of photon nudging for Janus particles.** The process involves a correlation between the particle’s orientation and propulsion towards a single target, which generates an inward drift field. During “on” periods, propulsion directs the microswimmers towards the target, while in “off” periods, they move diffusively. This cycle induces an inward radial drift, contributing to a tug-of-war dynamic where the particle repeatedly approaches and diffuses away from the target. [The labels in the figures should be all horizontal for ease of reading them]

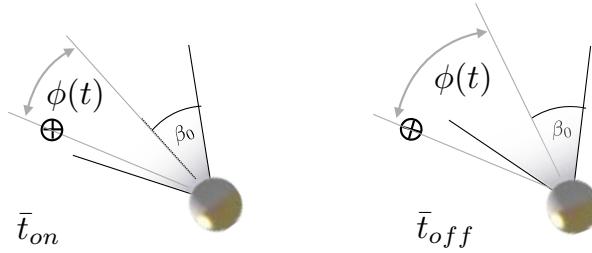


Figure 29.7: **First passage problems defining the effective drift velocity during photon nudging.** (a) Whenever the target gets inside the acceptance angle, an “on”-time starts which ends with the first passage of the angle ϕ through the cone defined by the acceptance angle β_0 . (b) An equivalent first passage problem can be defined for the “off”-time. [Agnese: Combine this with the previous figure? You can have multiple panels.](#)

Fig. 29.6 displays the effective working principle of photon nudging of Janus particles. During the “on” periods, the microswimmers are driven by their propulsion towards the target position, while they just move diffusively during the “off”-periods. This creates an effective inward radial drift with a velocity \mathbf{v}_ρ , which depends on the mean time period during which the particle is in the “on” state (its duty cycle). In general, the speed may depend on the distance ρ from the center. Once the particle has reached the target and has entered an “off” state, it diffuses away from the target position until the next “on” time starts.

In a steady state, the diffusive drift away from the target given by $\mathbf{j}_{\text{diff}} = -D\nabla p$ during the “off” times is, therefore, compensated by the advective drift $\mathbf{j}_{\text{nudge}} = p(\rho)\mathbf{v}_\rho$ towards the target, i.e.,

$$\mathbf{j}_{\text{diff}} + \mathbf{j}_{\text{nudge}} = -D\nabla p + p(\rho)\mathbf{v}_\rho = 0. \quad (29.4)$$

This is a steady-state version of the overdamped Fokker–Planck equation (or the Smoluchowski equation), where $p(\rho)$ denotes the probability density to find a particle at a distance $\rho, \rho + d\rho$ [Should this be denoted as a set?] from the target. The effective drift speed of the particle \mathbf{v}_ρ is governed by the Janus particle propulsion speed v_{th} and two additional factors, the duty factor f_t and a geometric factor f_r ,

$$v_\rho = f_t f_r v_{\text{th}}. \quad (29.5)$$

The duty cycle factor f_t is determined by the ratio between the average “on” time and the sum of average “on” and “off” times:

$$f_t = \frac{\bar{t}_{\text{on}}}{\bar{t}_{\text{on}} + \bar{t}_{\text{off}}} = \frac{\beta_0}{\pi}. \quad (29.6)$$

The expression on the right side, which is valid for nudging in 2D only, depends on the mean first passage times of the stochastic reorientation process (Fig. 29.7) for a given acceptance angle [30, 29].

The geometric factor f_r is determined by the fact that an acceptance angle β_0 (called vision cone in other experiments [14]) is required, as the probability for an exact orientation towards the target is essentially zero. Due to this finite acceptance angle/vision cone, the radial approach speed is smaller than the propulsion velocity v_{th} . This yields

$$f_r = \frac{1}{2\beta_0} \int_{-\beta_0}^{\beta_0} \cos(\phi) d\phi = \frac{\sin(\beta_0)}{\beta_0}, \quad (29.7)$$

where ϕ is the angle between the direction to the target and the swimmer’s body axis.

Eq. (29.4) describes a sedimentation problem towards a point-like target when assuming that $v_\rho = \text{const}$, i.e., $f_r = \text{const}$ and $f_t = \text{const}$, so that

$$p(\rho) = \frac{v_\rho^2}{D^2} \exp\left(-\frac{\rho v_\rho}{D}\right) \quad (29.8)$$

in 2 dimensions. From this equation, one can directly determine the sedimentation length $\rho_s = \frac{D}{v_\rho}$, which indicates the decay length of the probability density distribution $p(\rho)$ around the target. This is the accuracy with which you can localize an active particle at a target, and it corresponds to the accuracy of your control.

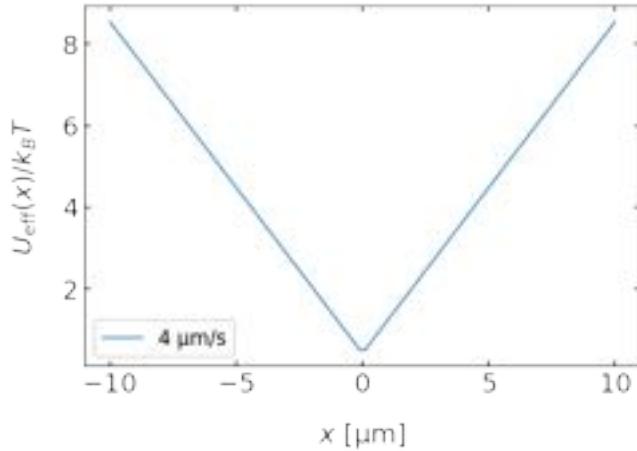


Figure 29.8: **Effective potential generated by photon nudging.** Effective potential, mirroring a gravitational field in sedimentation problems, caused by photon nudging with an infinite bandwidth of the feedback loop. This potential, purely a result of the feedback loop counteracting Brownian motion, linearly increases with distance from the target. It demonstrates how finite time-resolution and the particle's speed influence the radial drift and the critical orbit radius around the target. [Agnese: Redraw figure \(just style adjustments\).](#)

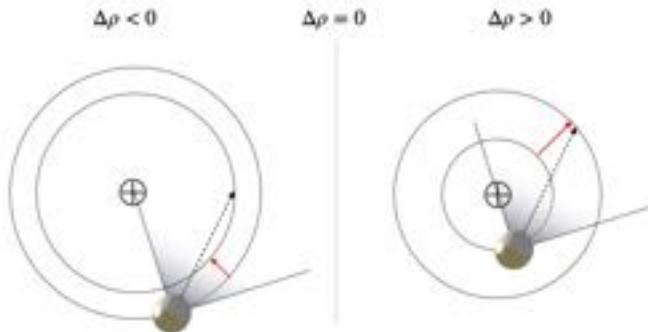


Figure 29.9: **Radial drift velocity in a photon nudging system with finite bandwidth** Distance-dependent radial drift velocity in a photon nudging system. At larger distances (a), particles effectively approach the target, while at smaller distances (b), they move away, a phenomenon influenced by the finite bandwidth of the system. [Agnese: combine with previous figure. Use panels.](#)

As the solution corresponds to a sedimentation problem, one may interpret Eq. (29.8) also in terms of a virtual potential where the swimmer is confined, illustrated in Fig. 29.8. This potential increases linearly with the distance from the target as in the case of the original sedimentation problem. However, there is no real potential and the motion is solely created by the feedback loop, which constantly tries to undo the spreading caused by Brownian motion. The best confinement is obtained when the velocity is infinitely high. This is, however, only true if one considers a feedback loop with infinite time-resolution $\Delta t = 0$. Only for this case, we can assume v_p constant. In the case of finite time-resolution ($\Delta t > 0$), the radial drift velocity will depend on the distance of the swimmer from the target, as illustrated in Fig. 29.9. In both cases, the particle travels with a speed v_{th} in the direction of the body axis (dashed arrow) for a time period Δt . If the particle is still far from the target (Fig. 29.9a), there is an effective radial drift inwards ($\Delta\rho < 0$). However, if the particle is rather close to the target (Fig. 29.9b), there is a radial drift outwards ($\Delta\rho > 0$) and the swimmer moves away from the target.

In between, there is a critical radius where the radial drift is zero ($\Delta\rho = 0$) and the swimmer orbits around the target. This radius R_c is given by

$$R_c = \frac{v_{\text{th}}\Delta t}{2\cos(\beta_0)} \quad (29.9)$$

and sets a stable orbit on which the particle is moving for large velocities or large sampling time Δt . Therefore, the finite time resolution (or bandwidth) of the feedback system defines the minimum distance (or the actual effective potential minimum), which is now no longer at $\rho = 0$ but at $\rho = R_c$. It is also interesting to note that the motion naturally gives a rotation around the target position. This is again, a natural consequence of the finite time resolution and the finite acceptance angle β_0 .

We can summarise the behavior of an active particle controlled by the photon nudging procedure as follow: in the limit of small particle speeds (or low Péclet number [29]), the particle diffuses during the “off” times away from the target, so the localization error is given by the sedimentation length, while for larger velocities (high Péclet number), the radius of the stable orbit defined by this length limits the particle’s approach to the target. This can be interpreted as an apparent repulsive component in the effective potential. Overall, there is an optimum localization accuracy as displayed in Fig. 29.10.

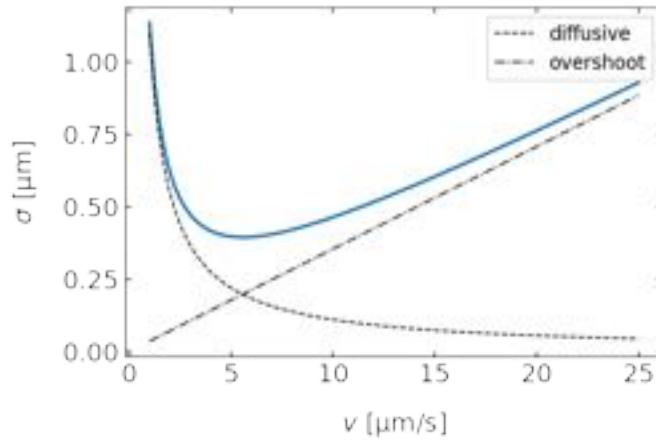


Figure 29.10: **Particle speed impact on localization accuracy.** Influence of the active particle’s speed ($1\,\mu\text{m}$ in diameter) on the localization accuracy at a target position. At lower speeds, diffusive motion leading the particle away from the target dominates. Conversely, at higher speeds, overshooting arises as active particles form a stable orbit around the target, indicating the interplay of diffusive and active movements in determining the control accuracy. [Agnese: redraw figure \(adjust to book style\)](#).

The previous discussion already gives some intuitive insights into the control of Janus particles by feedback control. A more detailed theoretical analysis also yields the optimal parameters that lead to a perfect localization and also to the shortest time to travel a certain distance. They can be expressed with the help of two unitless quantities, the *revolution number* $\lambda = \Delta t / \tau_r$ and the *propulsion number* $\kappa = v_{\text{th}} \tau_r / R$, where τ_r is the rotation diffusion time and R the radius of the active particle. The optimal localization of the active particle is thus given by

$$\sigma_{\min} \approx 6.9R\sqrt{\lambda}, \quad (29.10)$$

while the optimal acceptance angle is given by $\beta_{0,\text{opt}} \approx \pi/3$ in 2D. This suggests that, if one keeps the revolution number constant, one can localize smaller particles much better than larger ones. This comes from the fact that the rotational diffusion coefficient is scaling differently with the size of the particle R than its diffusion coefficient D : the rotational diffusion coefficient scales with R^{-3} , while the diffusion coefficient scales as R^{-1} , so smaller particles do reorient much faster towards the target than they can diffuse away compared to larger particles. Brownian motion helps, but, of course, it would be even better, if it was not there.

29.3.4 Advanced control with symmetric active particles

The preceding section illustrates a key limitation in the photon nudging of Janus particles, rooted in their rotational motion. Such motion is typically a significant factor for all microswimmers as it randomizes the direction of motion in the lab frame, leading to effective diffusive movement at larger time scales. This effect arises from the microswimmers' reliance on asymmetry to counteract the time-symmetry of the Stokes equation, which dominates in low Reynolds number environments. While this asymmetry is often linked to the microswimmer's shape, as seen with Janus particles, it can be uncoupled from the particle's form. By creating an asymmetric surface temperature, thereby inducing a thermo-osmotic flow, we can achieve the necessary asymmetry simply by making the heat release pattern asymmetric rather than altering the entire structure.

This idea can be implemented in an active particle where the surface is homogeneously covered with gold nanoparticles, as shown earlier. Such a particle is symmetric and has no preferential direction, but, when we asymmetrically illuminate the particle with a focused laser beam, we can create an asymmetric temperature profile, thus leading to propulsion [25]. The microswimmer propulsion is, therefore, completely equivalent to that of a Janus particle with the fundamental difference that we can now decide on the propulsion direction with the help of the illumination. The control of the particle is, therefore, much more accurate [25], but also requires a higher experimental precision on the control of the laser illumination. Fig. 29.11a shows a schematics for such a microswimmer which is illuminated by a tightly focused laser beam at a distance δ from the center. Fig. 29.11b shows the position r_h at which heating is required to propel towards a target r_t . As we program that in our computerized feedback mechanism, we only have to take care of defining the rule once to obtain the desired result, e.g., for the motion of the particle between two target positions, as illustrated in Fig. 29.11c. The advantage of such a system is that now we can not only program the translational motion of the microswimmer, but we can also superimpose any rotational motion as well: we can use a single particle and modify its rotational diffusion without ever changing its size.

Also this form of control is limited by the time resolution of our feedback mechanism. Since one detects

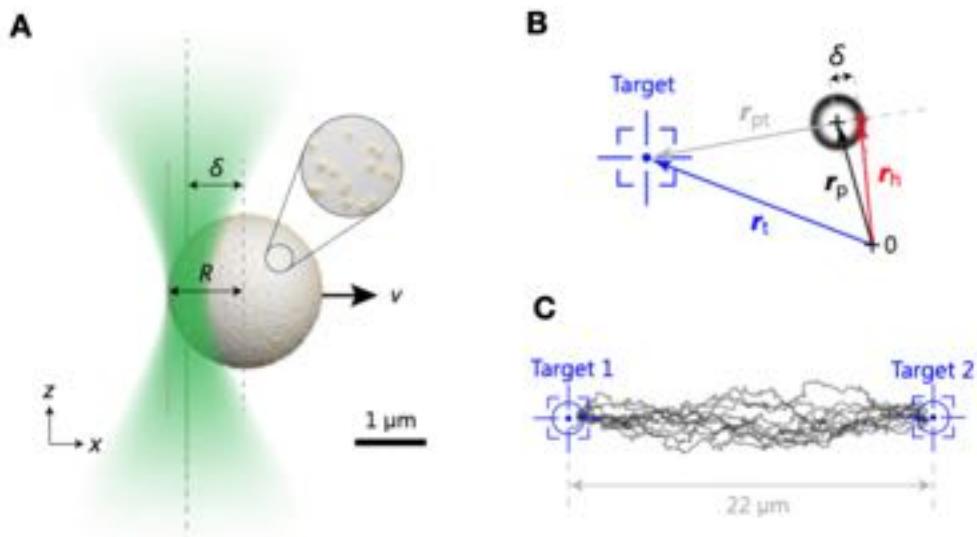


Figure 29.11: Advanced steering of a symmetric swimmer by photon nudging. [The figure looks pixelated. Is it possible to have a higher resolution version?] (a) A symmetric swimmer, homogeneously covered with gold nanoparticles, lacks a preferential direction. However, by applying asymmetrical illumination with a focused laser beam, an uneven temperature profile is created, inducing propulsion similar to that of a Janus particle. (b) Unlike a Janus particle, the propulsion direction can be altered using illumination, demanding greater experimental precision for laser control. (c) The system's benefit is that it allows the programming of both translational and rotational motion of the microswimmer, offering the possibility to modify rotational diffusion without altering its size.

a particle at a certain place in a given camera image, the particle will have moved a certain distance due to the previous actuation and its Brownian motion by the time that the image has been processed and the correction via the feedback mechanism is applied. At the moment that one is applying the new direction of motion, the particle position is wrong and one obtains a distribution of directions (given by the angle θ) around the intended motion direction as

$$p(\theta) \approx (2\pi\sigma_\theta^2)^{-1/2} e^{-\theta^2/(2\sigma_\theta^2)}, \quad (29.11)$$

with $\sigma_\theta^2 = 2D\tau/R^2$. Check this formula. Should it be D_τ instead? Or what is τ ? This uncertainty in the propulsion direction has important consequences. If one measures the mean squared distance traveled by the particle along the desired and the perpendicular direction, one obtains a contribution to the noise which is coming from the active propulsion and the random direction of propulsion. These mean squared distances can be calculated by

$$\sigma_\parallel^2 = (1 - \sigma_{y\theta}^2) v_\theta^2 \tau^2 - e^{-\sigma_\theta^2} v_\theta^2 \tau^2 + 2D\tau, \quad (29.12)$$

$$\sigma_\perp^2 = \sigma_{y\theta}^2 v_\theta^2 \tau^2 + 2D\tau, \quad (29.13)$$

with $\sigma_{y\theta}^2 = (1 - e^{-2\sigma_\theta^2})/2$ and $v_\theta = e^{\sigma_\theta^2/2} v_\parallel$. There is now a contribution to the square displacement which grows with the velocity. This means that, for objects which are propelled by feedback loops, large velocities reduce the level of control, and it actually provides a speed limit for active objects with non-zero feedback latency, which, for example, could also be valid for bacteria.

29.4 Applications of feedback control

Having discussed some of the basic principles behind the feedback control of active particles, we can now look into a few examples where to use those techniques. These examples are not exhaustive, and there are many other great examples in the literature, for example, by the groups of Bechinger [31, 14], Bevan [32, 33], and Zheng [34]. Check latter reference.

29.4.1 Active particle pressure and activity landscapes

As a first example, we address a very fundamental issue for active matter, which concerns the pressure an ensemble of active particles is generating when confined by walls or other objects. This pressure plays an essential role for observations such as motility induced phase separations (MIPS) from a gaseous phase to a crystalline phase of active particles [20, 19]. The understanding of pressure for active particles in conjunction to its equilibrium counterpart is more complex (some excellent literature for a deeper physics understanding is provided by Refs. [35, 36, 37, 38]).

For a system at equilibrium, the pressure is defined by the particle density ρ and the thermal energy $k_B T$ as $P = \rho k_B T$. This expression depends on the momentum transfer at the boundaries. At equilibrium, the pressure is a state function and is independent of the interactions of the particles with the wall.

For an active particle system, thermal energy does not define the speed v of the particles, which will keep pushing against the wall until they rotate away from it. Referring to the systems shown in Fig. 29.12, the starting point for a mathematical description of the pressure of an active particle is the overdamped Langevin equation (29.3), where we just have to add an additional force term $F_w \mathbf{e}_x$ on the right-hand side for the wall-particle interaction [39, 40, 35, 41]. Here, \mathbf{e}_x is the unit vector in the direction perpendicular to the wall. In the following, we will neglect any interactions between active particles, which, in principle, would need to be included too. Integrating over all orientations of the particles and averaging over the noise term yields a final expression for the current densities, which is, in one dimension,

$$-F_w \rho = vp(x) - D \frac{\partial \rho}{\partial x}. \quad (29.14)$$

According to this expression, the wall (left-hand side) is balancing the current created by an excess of particles pointing towards it ($p(x)$) and by a diffusion current (right-hand side). This excess $p(x)$ of particle orientations pointing towards the wall is a polarization density of active particles, which, in one dimension, is given by

$$p(x) = \langle \mathbf{n} \cdot \mathbf{e}_x \rangle \rho(x), \quad (29.15)$$

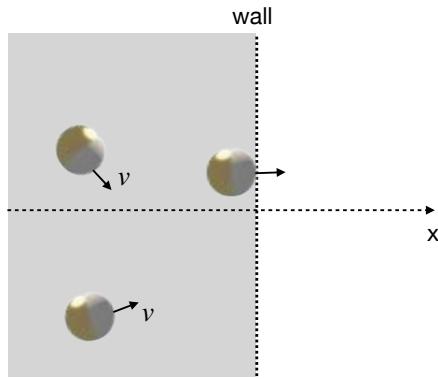


Figure 29.12: Geometry for active particle pressure calculation. The figure illustrates the concept underlying the mathematical modeling of pressure in an active particle system. Such systems are not defined by thermal energy, but by the persistent motion of the particles against a wall until they rotate away from it. The diagram shows a system where an additional force term, $F_w \mathbf{e}_x$, has been included to represent the wall-particle interaction. Here, \mathbf{e}_x is the unit vector perpendicular to the wall. This addition allows the derivation of an equation for the current densities and, consequently, the active particle pressure. [Check caption and figure. Agnese: Improve the quality of this figure and its relevance within the model described in the text.](#)

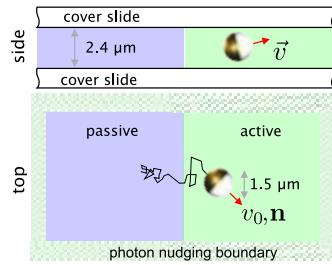


Figure 29.13: Experimental setup to confine an active particle to a virtual arena. Experimental setup demonstrating a single active particle confined within a virtual arena. The arena is partitioned into an active (right) and passive (left) region, implemented via feedback control. [Agnese: Combine with previous figure making panels.](#)

where \mathbf{n} is the particle orientation and the averaging $\langle \dots \rangle$ is over time. The active particle pressure is, therefore,

$$\Pi = - \int F_w(x) \rho(x) dx = v \int p(x) dx + D\rho. \quad (29.16)$$

The first term on the right-hand side corresponds to the active pressure exerted by the particle and the second term is the normal “ideal gas” pressure of the colloidal particles confined in a volume. We will neglect this contribution in the following as it does not relate to activity. The total polarization is then given by

$$P = \int_{-\infty}^{\infty} p(x) dx \quad (29.17)$$

and we can write the active particle pressure as

$$\Pi = vP. \quad (29.18)$$

If we could measure the polarization $p(x)$ at a wall, we would then measure the active particle pressure. A real-world measurement, however, would involve hydrodynamic interactions of the active particles with the

walls, which is not included in this theory. This shows that the specific interactions with a wall modifies the active particle pressure, which is not the case at equilibrium. A question arises: how to measure this pressure without an influencing wall?

Photon nudging offers an innovative solution for creating interaction-free boundaries for an active particle within an arena. As depicted in Fig. 29.13, a single active particle can be confined within a rectangular arena using photon nudging. The hashed lines represent the photon nudging boundaries - upon reaching these borders, the propulsion of the particle is switched off, waiting until it reorients towards the center of the arena before reigniting propulsion. This technique allows for long-term confinement and observation of a single particle. To gain insights into the active pressure, a boundary can be introduced within the arena to distinguish between active and passive regions, as also depicted in Fig. 29.13. Real-time tracking of the particle position allows for dynamic toggling of global illumination, effectively creating and modifying the passive and active zones. The boundaries created through this feedback control are sharper and more precise than those achievable through conventional diffractive optics.

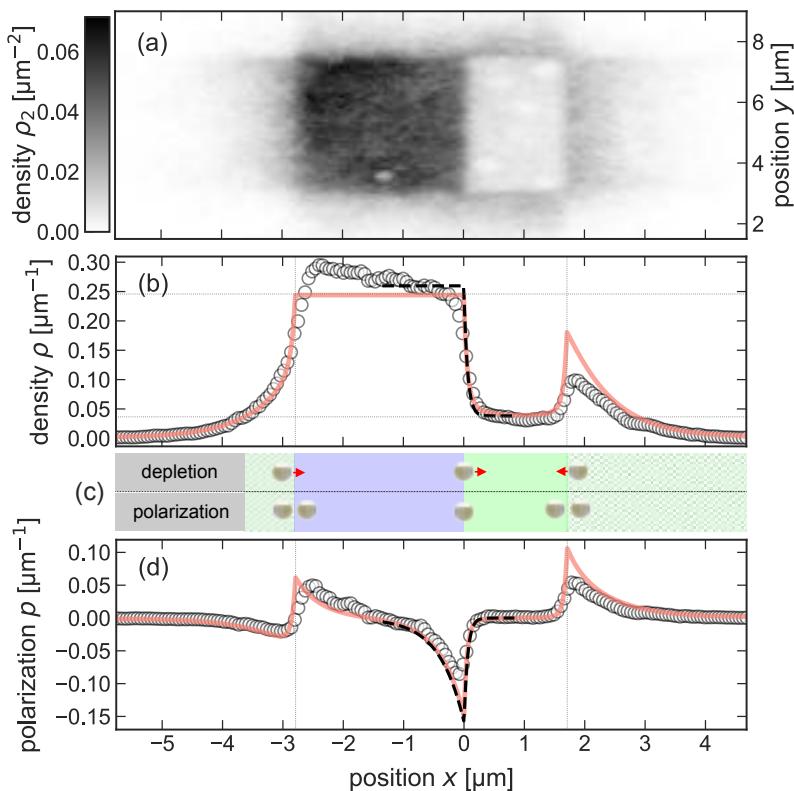


Figure 29.14: Measurement of active pressure. (a) Density of active particle trajectories in the virtual arena with the activity step. (b) Density profile obtained from (a) by integrating over the y -direction. (c) Sketch of the dynamics leading to polarization at the individual boundaries. (d) Active particle polarization density defining the pressure at the individual boundaries. The solid and dashed lines denote theoretical predictions. [Agnese: Change labels in figure and text.](#)

The result of such an experiment is shown in Fig. 29.14. Fig. 29.14a displays the density of trajectory points in either the passive (left/dark) or the active (right/bright) region. Fig. 29.14b displays this density integrated over the y -position. The density ratio between both sides follows an extension of Schnitzer's derivation [42], which states that the inverse velocity determines the density, i.e., $\rho = 1/v$ in the limit of zero Brownian motion. The final expression is given by

$$\frac{\rho_a}{\rho_p} = \sqrt{\frac{D + v^2 \tau_p / 2}{D + v^2 \tau_a / 2}}, \quad (29.19)$$

where τ_a, τ_p are the inverse rotational diffusion coefficients in the active and passive region, respectively. Fig. 29.14d displays the excess polarization $p(x)$ together with a theoretical prediction. Integrating over this polarization density gives the active particle pressure.

29.4.2 Information-controlled self-assembly

Interactions grounded in signal reception are vital across all scales pertinent to living matter. Despite these signals originating from the physical world (e.g, molecular diffusion, or the transmission of sound or light), they liberate living entities from the confines of Newton's third law. For instance, we need not exert a physical force to alter a bird's flight direction; instead, it suffices to communicate with the bird, prompting it to willingly change course. In essence, this suggests that, in active systems, order can be induced from information. While altering the rules guiding birds' flight direction change is currently still challenging, we have the means to manipulate these information-based influences in the realm of man-made active particles.

A demonstration of feedback and information-controlled self-organization can be carried out with the symmetric microswimmers detailed before. For example, Fig. 29.15a schematically represent a simple rule that tells two swimmers to propel towards each other when they are farther than a distance r_{eq} and away from each other when they are closer. This pairwise interaction can be mapped onto an effective force, as shown in Fig. 29.15b, or an effective potential, as depicted in Fig. 29.15c. Mathematically, this pairwise interaction is given by the following two equations:

$$\mathbf{e}_i(t) = \frac{\sum_{j \neq i}^N \text{sign}(r_{ij} - r_{eq}) \mathbf{e}_{ij}}{\left| \sum_{j \neq i}^N \text{sign}(r_{ij} - r_{eq}) \mathbf{e}_{ij} \right|}, \quad (29.20)$$

$$\mathbf{v}_i(t) = -v_{th} \mathbf{e}_i, \quad (29.21)$$

ensure all quantities are defined. Together they define the velocity of the particle i . Note that the interaction strength between particles is independent of the distance. In these systems, all particles consistently move at a speed of v_{th} . When such rules are applied to active particles, they result in self-organized structures, as demonstrated in Fig. 29.16. For groups larger than four, these structures tend to be unusual; they display frustration because the processes unfold in two dimensions, and the resulting configurations do not align with the expected formation of triangular assemblies. This deviation can be attributed to the absence of a cut-off in

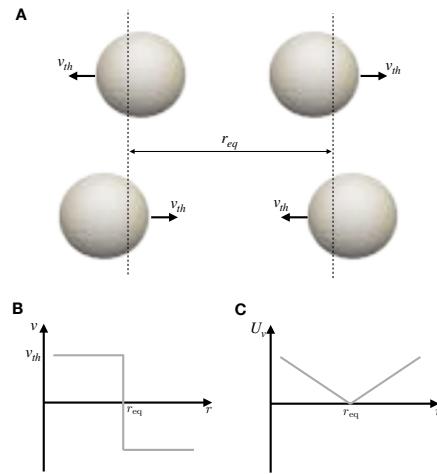


Figure 29.15: **Information-based self-organisation of active particles.** (a) Pairwise interaction rule for the feedback controlled self-organization of active particles. (b) Equivalent “force” acting on the particles. (c) Effective potential experienced by the active particles for the feedback controlled self-organization. [Agnese: Arrange in a line](#)

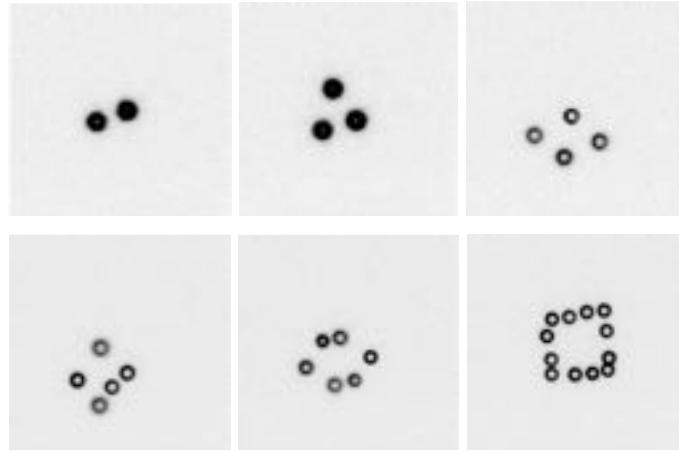


Figure 29.16: Self-organized structures of active particles generated by feedback control. Diverse configurations of active particles under the influence of feedback control rules. The observed structures, particularly for groups larger than four, tend to be non-trivial, showcasing the complex dynamics inherent in these systems. **Agnese: this can be combined with the previous figure.**

interactions — each particle is equally attracted to others. Moreover, these structures are not static; dimers and trimers exhibit discernible oscillations. The orthogonal modes of these oscillations can be ascertained through principal component analysis (PCA). Fig. 29.17, for instance, illustrates the outcome of PCA for a dimer.

When subtracting the motion of the center-of-mass (Fig. 29.17a), two residual modes of motion are retained (Fig. 29.17b-d), which are a “bond” stretch (Fig. 29.17b) and a “bond” rotation (Fig. 29.17c). Both these modes exhibit oscillations. This oscillatory behavior can be visualized by projecting the movement of the particle, as seen in the center-of-mass frame, onto these two orthogonal modes (Figs. 29.17e and (Fig. 29.17f). These oscillations reveal something very important in every feedback system, which is a feedback latency or delay. Sensing information and responding to it are always separated by a time δt . This is true for man-made devices as well as for any living systems [28]. In our experimental system, this delay is given by the inverse frame rate if not specified otherwise. Considering the dimer with the “bond” length $\mathbf{r}_{12}(t) = \mathbf{r}_1(t) - \mathbf{r}_2(t)$, we can thus describe the dynamics of this bond length with the following Langevin equation:

$$\frac{dr_{12}}{dt}(t) = 2v_1(t) + \sqrt{4D_0}\eta_{12} = -2v_{\text{th}} \text{sign}(|\mathbf{r}_{12}(t - \delta t)| - r_{\text{eq}}) \mathbf{e}_{12}(t - \delta t) + \sqrt{4D_0}\eta_{12}, \quad (29.22)$$

which is solved by

$$r_{12}(t) = r_{\text{eq}} + Ax_{\text{tr}}(t + \phi), \quad (29.23)$$

where $A = 2v_{\text{th}}\delta t$ is the amplitude of oscillations and

$$x_{\text{tr}}(t) = \frac{8}{\pi^2} \sum_{k=0}^{\infty} (-1)^k \frac{\sin(2\pi(2k+1)ft)}{(2k+1)^2} \quad (29.24)$$

denotes a triangular wave with amplitude 1 and period $\tau = 1/f = 4\delta t$. The phase shift ϕ is determined by the initial condition, which is assumed to be drawn from the interval $[r_{\text{eq}} - A, r_{\text{eq}} + A]$ obtained by Eq. (29.23) in the steady state. As explained in Ref. [43], the dynamics of the system are governed by the activity v_{th} and the delay. The oscillation period is $T = 4\delta t$, the amplitude of the oscillation $A = 4v_{\text{th}}\delta t$ and the Brownian motion destroys the coherence of the oscillation after a time $T_2 \approx A^2/2D$.

We can highlight two main results. First, the feedback loop controlling the particles is generating structural order in the system. Particles propel continuously with the same speed and therefore the entropy production in the system is in principle constant. Yet, the information gained about the system is used to steer the particles and, in this way, the entropy of the system is reduced without being generated by any direct physical interaction. Second, feedback loops have a latency that is the source of the oscillating dynamics in this system. Living

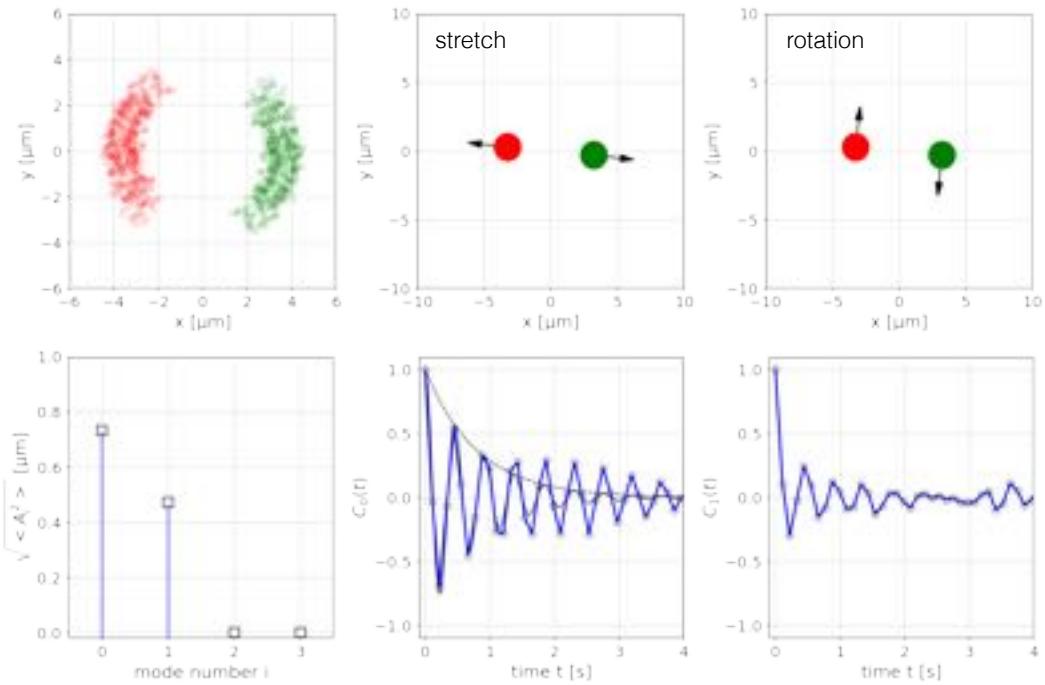


Figure 29.17: Dynamic behavior of a feedback-controlled self-organized active particle dimer. (a) Movement of the active particles in the center-of-mass frame, illustrating their continuous displacement. (b-c) The first two principal component modes extracted from a principal component analysis (PCA), revealing the dominant patterns of fluctuation in the system. (d) Amplitude of these principal components, correlating with the magnitude of radial and tangential variations. (e-f) Autocorrelations of these radial and tangential fluctuations, offering insight into the temporal persistence of these variations in the active particle dynamics.

matter dynamics develop from many of such feedback loops, which are coupled with each other and introduce different timescales. Living systems must therefore function in their presence [44], and probably living matter is robust because of the spread of different delays and timescales involved. With the help of feedback control, synthetic active particles could contribute to understand these dynamics of living systems.

29.4.3 Machine learning with microswimmers

So far we have decided the rules that define the action of the particles. In living matter, similar rules are shaped by the environment in a learning process, which may also include evolution at very long time scales. The next example introduces a similar learning procedure into the feedback control of active particles. This learning procedure is based on an abstraction of actual learning processes and is called *reinforcement learning* [45], which has meanwhile also found its way into a number of theoretical studies of active systems [46, 47, 48]. Reinforcement learning is a semi-supervised learning process [49]. It is based on rewards that are awarded whenever the agent reaches a specific goal. The reward it earns during an episode of the learning process will modify its actions in the future. This type of learning can be squeezed into iterative algorithms that keep track of the total reward, which is maximized under optimal behavior. There are several algorithms available and the simplest one is the so-called *Q-learning*, which is based on the following mathematical procedure:

$$Q_{t+\Delta t}(s, a) = Q_t(s, a) + \alpha \left[R(s') + \gamma \max_{a'} Q_t(s', a') - Q_t(s, a) \right], \quad (29.25)$$

where Q is a matrix which represents the memory of the object. In this matrix, each individual state s in which the object can be has an entry. In each state, the object can take one out of a set of possible actions a . Not all

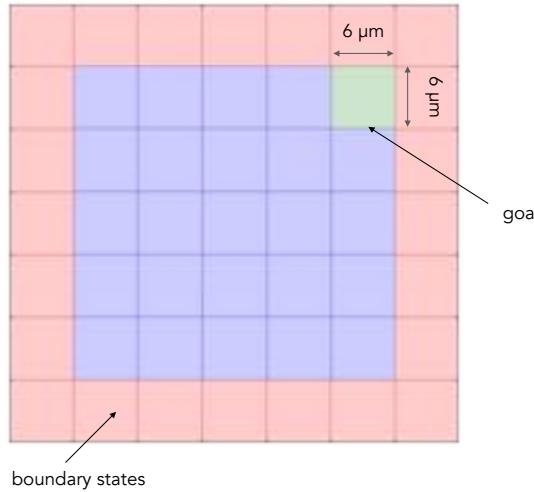


Figure 29.18: Representation of the reinforcement learning arena used for the microswimmer. This gridworld illustrates a 5×5 matrix of states navigated by the microswimmer. Each cell symbolizes a state, and the microswimmer moves among these through different actions, accruing rewards and penalties based on its location. The top right state is the intended goal, while the red-bordered squares represent confining boundaries. Penalties and rewards are applied accordingly: a penalty of $R = -1$ is given in all blue squares, $R = -100$ for leaving the blue area, and a reward $R = 10$ for reaching the goal. This scenario allows for the observation of the microswimmer's adaptation to its virtual world.

the actions will be good, so we need to store a value associated to each action in the Q-matrix. Thus, if there are 8 possible actions, each state has 8 Q-matrix entries. The above equation states that the Q-matrix entry of the state s and the action a at the next step will depend on the current Q-matrix entry plus some quantity that is multiplied by the so-called learning rate α . The learning rate defines the speed at which new information is incorporated into the system memory. The terms in brackets consist of the reward R that is obtained when carrying out the action a to the next state s' and a prediction term $\max_{a'} Q_t(s', a')$ providing the value of the best action in the next state. This last term is multiplied by a scaling factor γ as current rewards count more than future ones.

Agnese: add material from the workbooks as exercises.

For our microswimmer, we can tackle a very simple problem of navigating in a grid of 5×5 states towards a goal. This is shown in Fig. 29.18. The top right state is the goal and the red states around the 25-state grid resemble some confining boundaries. In each of the states, the microswimmer can go in 8 directions (2 horizontal, 2 vertical and 4 diagonal) with the help of our feedback control. Each action is applied until the state changes, i.e., until a new square is reached. In all blue squares, the microswimmer obtains a penalty of $R = -1$, while it gets a penalty of $R = -100$ when leaving the blue area and a reward $R = 10$ when reaching the goal. Applying this set of rewards and penalties to a single microswimmer allows one to observe the adaptation of the microswimmer to the defined (virtual) world. After about 4800 episodes (1 episode is starting at a random point and ending whenever the particle leaves the board or reaches the goal), the microswimmer finds its way straight towards the goal state, no matter its initial position (Fig. 29.19). There is therefore a specific action in each state that is optimal for reaching the goal. The set of actions for each state, which provides the behavior of the microswimmer, is called policy. The policy for the previous case is depicted in Fig. 29.20 for the two times before the learning starts (lower left) and after the learning procedure (lower right) and clearly shows that the microswimmer has figured out the right path towards the target.

There are a number of additional issues connected with this simple model system. For example, the details about the importance of noise during this learning process is discussed in Ref. [50]. Besides this, learning in a real environment opens the way to develop adaptable microswimmers, which can learn from their environment, for example, by also learning from neighboring particles.

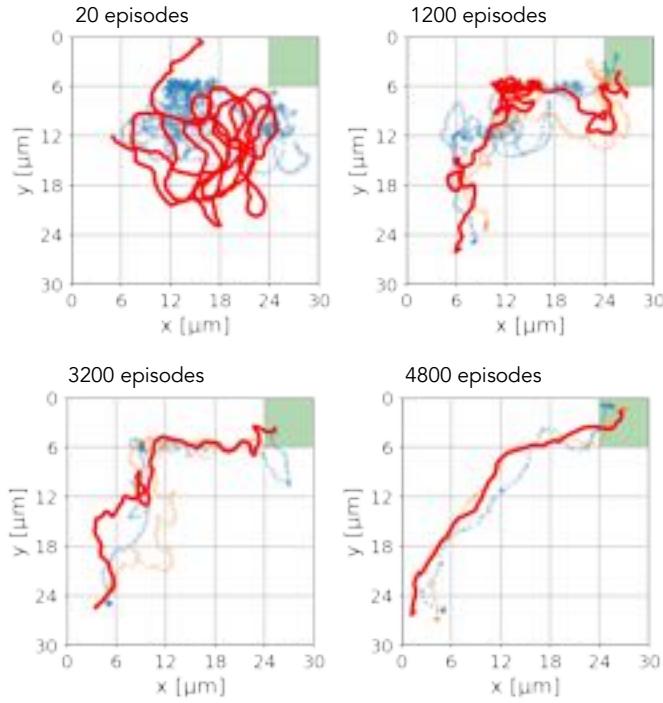


Figure 29.19: Evolution of microswimmer trajectories throughout the reinforcement learning process. Distinct instances of a single microswimmer’s behavior at various stages of learning. The first instance shows a novice swimmer embarking from a random point in the gridworld, characterized by multiple loops in its trajectory. As the learning episodes accumulate, the swimmer’s trajectories exhibit growing persistence towards the goal, culminating in a direct path regardless of initial position after approximately 4800 episodes. [Agnese: Arrange panels in a line](#)

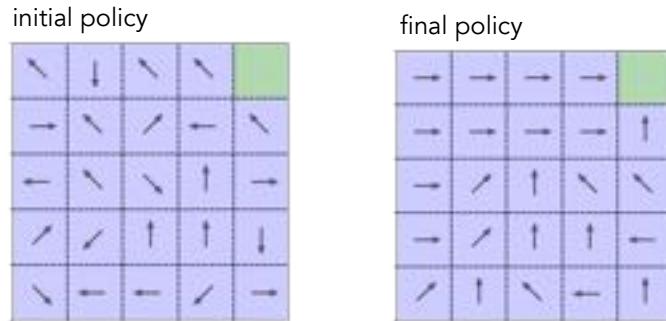


Figure 29.20: Policies defined by the Q-matrix before and after the convergence of the learning process. Actions (policy) taken by the microswimmer in each state before (left) and after (right) the learning process. Post-learning, the policy clearly demonstrates that the microswimmer has discerned the most effective path to the target. [Agnese: combine with previous figure by adding the intial policy at the beginning and the final one at the end.](#)

Bibliography

- [1] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

- [2] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: An explanation of the $1/f$ noise. *Physical Review Letters*, 59:381–384, 1987.
- [3] D. Dahmen, S. Grün, M. Diesmann, and M. Helias. Second type of criticality in the brain uncovers rich multiple-neuron dynamics. *Proceedings of the National Academy of Sciences*, 116(26):13051–13060, 2019.
- [4] M. A. Muñoz. Colloquium: Criticality and dynamical scaling in living systems. *Reviews of Modern Physics*, 90(3):031001, 2018.
- [5] A. Sornette and D. Sornette. Self-organized criticality and earthquakes. *Europhysics Letters*, 9(3):197–202, 1989.
- [6] V. Dakos, M. Scheffer, E. H. van Nes, V. Brovkin, V. Petoukhov, and H. Held. Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, 105(38):14308–14312, 2008.
- [7] T. M. Lenton, V. N. Livina, V. Dakos, and M. Scheffer. Climate bifurcation during the last deglaciation? *Climate of the Past*, 8(4):1127–1139, 2012.
- [8] K. J. Åström and R. M. Murray. *Feedback Systems*. Princeton University Press, 2010.
- [9] J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Physical Review E*, 58(4):4828 – 4858, 1998.
- [10] T. Mora and W. Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.
- [11] J. Goldstone, A. Salam, and S. Weinberg. Broken symmetries. *Physical Review*, 127(3):965–970, 1962.
- [12] B. Qian, D. Montiel, A. Bregulla, F. Cichos, and H. Yang. Harnessing thermal fluctuations for purposeful activities: the manipulation of single micro-swimmers by adaptive photon nudging. *Chemical Science*, 4(4):1420–1429, 2013.
- [13] A. P. Bregulla, H. Yang, and F. Cichos. Stochastic localization of microswimmers by photon nudging. *ACS Nano*, 8(7):6542 – 6550, 2014.
- [14] T. Bäuerle, R. C. Löffler, and C. Bechinger. Formation of stable and responsive collective states in suspensions of active colloids. *Nature Communications*, 11(1):2547 – 9, 2020.
- [15] T. Mano, J.-B. Delfau, J. Iwasawa, and M. Sano. Correction for mano et al., optimal run-and-tumble-based transportation of a Janus particle with active steering. *Proceedings of the National Academy of Sciences*, 115(8):201801089, 2018.
- [16] H. Xie, M. Sun, X. Fan, Z. Lin, W. Chen, L. Wang, L. Dong, and Q. He. Reconfigurable magnetic microrobot swarm: Multimode transformation, locomotion, and manipulation. *Science Robotics*, 4(28):eaav8006, 2019.
- [17] J. Yan, M. Han, J. Zhang, C. Xu, E. Luijten, and S. Granick. Reconfiguring active particles by electrostatic imbalance. *Nature Materials*, 15(10):1095 – 1099, 2016.
- [18] H.-R. Jiang, N. Yoshinaga, and M. Sano. Active motion of a Janus particle by self-thermophoresis in a defocused laser beam. *Physical Review Letters*, 105(26):268302, 2010.
- [19] I. Buttinoni, G. Volpe, F. Kümmel, G. Volpe, and C. Bechinger. Active brownian motion tunable by light. *Journal of Physics: Condensed Matter*, 24(28):284129, 2012.
- [20] J. Palacci, S. Sacanna, A. P. Steinberg, D. J. Pine, and P. M. Chaikin. Living crystals of light-activated colloidal surfers. *Science*, 339(6122):936 – 940, 2013.

- [21] A. P. Bregulla, A. Würger, K. Günther, M. Mertig, and F. Cichos. Thermo-osmotic flow in thin films. *Physical Review Letters*, 116(18):188303, 2016.
- [22] B. V. Derjaguin and G. P. Sidorenkov. Thermoosmosis at ordinary temperatures and its analogy with the thermomechanical effect in Helium II. *Doklady Akademii Nauk SSSR (Proceedings of the USSR Academy of Sciences)*, 32:622, 1941.
- [23] E. M. Purcell. Life at low reynolds number. *American Journal of Physics*, 45(1):3–11, 1977.
- [24] J. Happel and H. Brenner. *Low Reynolds number hydrodynamics, with special applications to particulate media*. Mechanics of fluids and transport processes. Martinus Nijhoff Publishers, a member of the Kluwer Academic Publishers Group, 1983.
- [25] M. Fränzl, S. Muñoz Landin, V. Holubec, and F. Cichos. Fully steerable symmetric thermoplasmonic microswimmers. *ACS Nano*, 15(2):3434–3440, 2021.
- [26] S. Helgadottir, A. Argun, and G. Volpe. Digital video microscopy enhanced by deep learning. *Optica*, 6(4):506, 2019.
- [27] M. Fraenzl and F. Cichos. Active particle feedback control with a single-shot detection convolutional neural network. *Scientific Reports*, 10(1), 2020.
- [28] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe. Engineering sensorial delay to control phototaxis and emergent collective behaviors. *Physical Review X*, 6(1):011008, 2016.
- [29] M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, and H. Yang. Theory for controlling individual self-propelled micro-swimmers by photon nudging II: confinement. *Physical Chemistry Chemical Physics*, 20(15):10521 – 10532, 2018.
- [30] M. Selmke, U. Khadka, A. P. Bregulla, F. Cichos, and H. Yang. Theory for controlling individual self-propelled micro-swimmers by photon nudging I: directed transport. *Physical Chemistry Chemical Physics*, 20(15):10502 – 10520, 2018.
- [31] T. Bäuerle, A. Fischer, T. Speck, and C. Bechinger. Self-organization of active particles by quorum sensing rules. *Nature Communications*, 9(1):3232, 2018.
- [32] J. Zhang, J. Yang, Y. Zhang, and M. A. Bevan. Controlling colloidal crystals via morphing energy landscapes and reinforcement learning. *Science Advances*, 6(48):eabd6716, 2020.
- [33] Y. Yang and M. A. Bevan. Cargo capture and transport by colloidal swarms. *Science Advances*, 6(4):eaay7679, 2020.
- [34] L. Lin, J. Zhang, X. Peng, Z. Wu, A. C. H. Coughlan, Z. Mao, M. A. Bevan, and Y. Zheng. Opto-thermophoretic assembly of colloidal matter. *Science Advances*, 3(9):e1700458, 2017.
- [35] A. P. Solon, Y. Fily, A. Baskaran, M. E. Cates, and Y. Kafri. Pressure is not a state function for generic active fluids. *Nature Physics*, 11(8):673 – 678, 2015.
- [36] S. C. Takatori, R. De Dier, J. Vermant, and J. F. Brady. Acoustic trapping of active matter. *Nature Communications*, 7(1):10694, 2016.
- [37] S. C. Takatori, W. Yan, and J. F. Brady. Swim pressure: Stress generation in active matter. *Physical Review Letters*, 113(2), 2014.
- [38] H. Row and J. F. Brady. Reverse osmotic effect in active matter. *Physical Review E*, 101(6), 2020.
- [39] A. P. Solon, J. Stenhammar, R. Wittkowski, M. Kardar, Y. Kafri, M. E. Cates, and J. Tailleur. Pressure and phase equilibria in interacting active brownian spheres. *Physical Review Letters*, 114(19):198301, 2015.
- [40] A. K. Omar, Z.-G. Wang, and J. F. Brady. Microscopic origins of the swim pressure and the anomalous surface tension of active matter. *Physical Review E*, 101(1):012604, 2020.

- [41] S. Hermann and M. Schmidt. Active interface polarization as a state function. *Physical Review Research*, 2(2):022003, 2020.
- [42] M. J. Schnitzer. Theory of continuum random walks and application to chemotaxis. *Physical Review E*, 48(4):2553 – 2568, 1993.
- [43] U. Khadka, V. Holubec, H. Yang, and F. Cichos. Active particles bound by information flows. *Nature Communications*, 9(1):3864, 2018.
- [44] H. L. More, J. R. Hutchinson, D. F. Collins, D. J. Weber, S. K. H. Aung, and J. M. Donelan. Scaling of sensorimotor control in terrestrial mammals. *Proceedings of the Royal Society B: Biological Sciences*, 277(1700):3563–3568, 2010.
- [45] R. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.
- [46] K. Gustavsson, L. Biferale, A. Celani, and S. Colabrese. Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning. *European Physical Journal E*, 40(12), 2017.
- [47] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale. Flow navigation by smart microswimmers via reinforcement learning. *Physical Review Letters*, 118(15):1–5, 2017.
- [48] E. Schneider and H. Stark. Optimal steering of a smart active particle. *Europhysics Letters*, 127(6):64003, 2019.
- [49] F. Cichos, K. Gustavsson, B. Mehlig, and G. Volpe. Machine learning for active matter. *Nature Machine Intelligence*, 2(2):94 – 130, 2020.
- [50] S. Muiños Landin, A. Fischer, V. Holubec, and F. Cichos. Reinforcement learning with artificial microswimmers. *Science Robotics*, 6(52):eabd9285, 2021.

Chapter 30

Active Colloids at Fluid Interfaces

LUCIO ISA

Even though the first successful demonstration of self-propelled objects used millimetre-sized swimmers confined at a water-air interface [1], the intriguing realm of active colloids confined at fluid interfaces has been less explored than their bulk counterparts. Some examples of these swimmers are provided in Fig. 30.1. This Chapter provides a comprehensive guide on key concepts including the energetics of particle adsorption at fluid interfaces, with special emphasis on Janus particles. Additionally, it introduces experimental techniques for measuring the contact angle of single particles at fluid interfaces and the interfacial tension between two fluids. It also discusses the unique propulsion strategies possible for particles absorbed at interfaces, such as surface tension gradients, and the novel interactions absent in bulk systems like capillary forces. To illustrate these principles, two studies on catalytic swimmers [5] and Marangoni surfers [6] are presented. The discussion, while limited to colloidal particles crossing the interface as opposed to those near a fluid interface, underscores the unique characteristics and implications of irreversible trapping. [Agnese: renumber figure files](#)

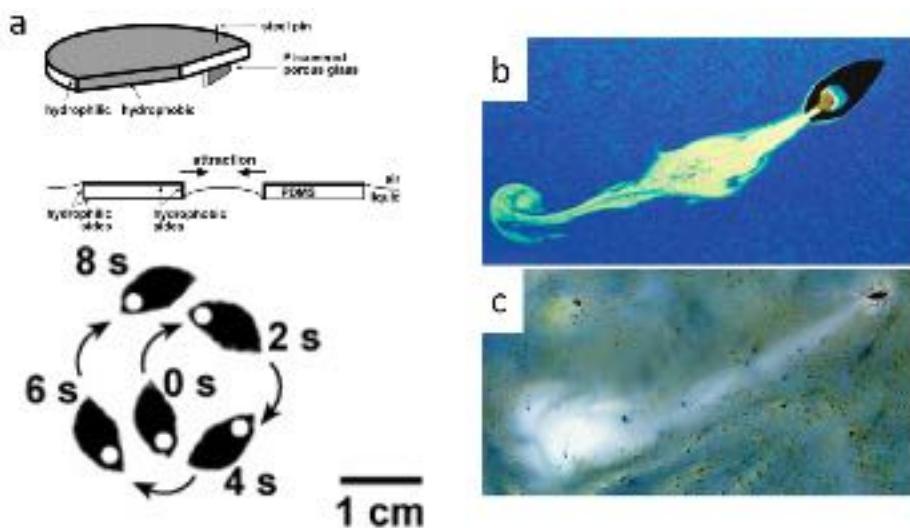


Figure 30.1: **Self-propelled objects at water-air interfaces.** Examples of self-propelled objects at water-air interfaces: (a) a silicon elastomer (PDMS) catalytic swimmers [1], (b) a soap-powered boat [2, 3], and (c) an insect (genus *Microvelia*) self-propelling thanks to Marangoni stresses [4]. [Agnese: polish figure, maybe three squares in a row?](#)

30.1 Why active colloids at fluid interfaces

Most of the discussion on active colloids in this book is devoted to the description of their experimental behaviors in bulk aqueous environments. In particular, the case of self-propelling catalytic microswimmers powered by the decomposition of hydrogen peroxide (H_2O_2) by means of a platinum (Pt) cap has emerged as the benchmark case to study the dynamics of active Brownian particles. However, not all may know that the first demonstration of catalytic self-propelled objects actually involved millimetre-sized swimmers confined at a water-air interface [1]. In this pioneering study, Ismagilov *et al.* took silicon elastomer (PDMS) platelets equipped with a Pt-doped porous glass *motor* and showed that they self-propelled at velocities of the order of 1 cm s^{-1} due to the Pt-catalyzed decomposition of hydrogen peroxide present in the water (Fig. 30.1a). Active propulsion of macroscopic objects confined at the water–air interface has also been popularized by experiments with camphor- or *soap-powered boats* releasing surfactant at one end [2, 3] (Fig. 30.1b). Finally, propulsion at interfaces is even exploited by insects to move efficiently over the surface of water (Fig. 30.1c) [4].

Downscaling the size of the swimmers to the microscale and investigating the behavior of active Brownian particles confined at a fluid interface presents exciting opportunities for interesting new physics and applications [7]. First of all, the case of active colloids at fluid interfaces has been much less explored than the bulk case. Moreover, being confined at the interface offers some unique distinctive traits. In the rest of this Chapter, we will refer to a *fluid interface* as the interface between an aqueous phase and an apolar (or less polar) phase (e.g., oil or air). The most important features, which makes the study of these systems interesting, include:

Confinement within the interface plane. As described later, the presence of a fluid interface leads to the irreversible trapping of microscale objects, including active colloids. Adsorption to the interface implies that the particles move in a quasi-2D world defined by the interface plane and are localized without the need of confining walls, as it is instead the case for swimmers in bulk. Even if vertically trapped at the interface, the particles are still free to move within the interface plane and experience, in general, a complex viscoelastic environment, which they navigate by a combination of diffusive and active processes.

Access to unique propulsion strategies. In addition to propulsion strategies also present in bulk systems (e.g., catalytic reaction), the presence of the interface enables unique modes of locomotion (e.g., exploiting self-induced gradients of surface tension).

Access to additional interactions. Even though this Chapter focuses on the behavior of single swimmers, the presence of fluid interfaces induces the manifestation of additional interactions absent in bulk. Examples are capillary forces or long-ranged electrostatic forces of dipolar nature, which can be used to direct the formation of interesting structures, which in turn can be coupled to self-propulsion in different ways.

The interested reader can find an excellent review of the recent progress on active particles at interfaces in Ref. [8], which in particular describes different propulsion mechanisms viable at the interface. Further reviews discuss colloidal interactions at fluid interfaces in detail [9, 10].

As a last note, in the following we limit our description to the case of colloidal particles that actually cross the interface and have parts of their surfaces exposed to both fluids. This case is different from that of particles close to a fluid interface, which has been studied by Dietrich and coworkers [11]. Even though some of the phenomenology is similar, the presence of irreversible trapping leads to very distinctive characteristics, as described below.

30.2 Particle adsorption at fluid interfaces

We begin by quickly recapitulating the concepts of interfacial free energy and interfacial tension. Each interface between two different materials is characterized by an *interfacial free energy* F . The differential change of F relative to a variation of the interfacial area A defines the *interfacial tension* γ as

$$\frac{\partial F}{\partial A} = \gamma. \quad (30.1)$$

Each interface thus has an “energy cost” and the interfacial tension is a representation of what is the energy cost to increase the area of the interface per unit area. Because increasing the area of an interface increases the

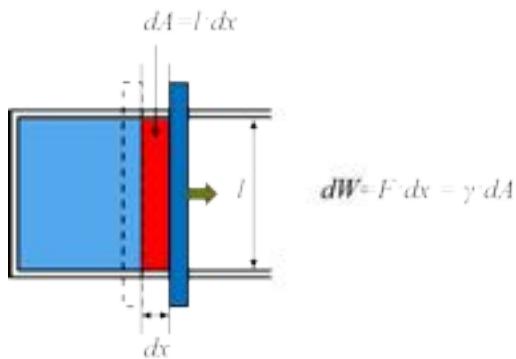


Figure 30.2: Mechanical definition of interfacial tension. Liquid film in a rectangular frame with a movable barrier. As force F shifts the barrier rightward, the liquid film's area increases. This work dW , representing the energy spent, is proportional to the area change dA , with surface tension γ as the proportionality constant. The unit of γ is force per unit length (commonly mN m^{-1}). [check caption Agnese: redraw/uniform?](#)

associated free energy, γ is a positive quantity and can be mechanically represented as the driving force for the minimization of the interfacial area, and thus of interfacial energy.

For fluid–fluid interfaces, this can be easily represented by imagining to have a liquid film freely suspended in a rectangular frame that has a movable barrier (Fig. 30.2). If the barrier is moved to the right by applying a force F , then the area of the liquid film is increased by an amount directly proportional to the displacement of the frame. We can thus calculate the energy spent (the work done) in this process as indicated in Fig. 30.2. The work can also be written as proportional to the area change, where the surface tension γ is the proportionality constant. Given such definition, γ has the units of a force per unit length and it is typically measured in mN m^{-1} .

This basic concept is at the core of the energetic description of the adsorption of colloidal particles at fluid interfaces. In order to understand why colloidal particles can be trapped at a fluid interface and to estimate how

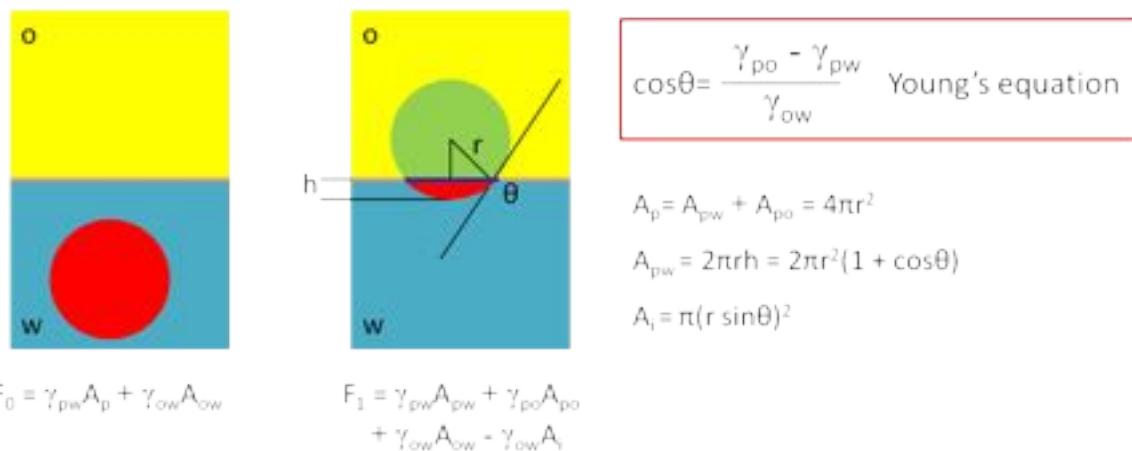


Figure 30.3: Free-energy calculation for the adsorption of a colloidal sphere at liquid–liquid interfaces. Free-energy calculations when a spherical colloidal particle adsorbs at a liquid–liquid interface, visualizing the various interfacial energy considerations, changes in surface areas, and the role of contact angles during the adsorption process. The equations depicted derive the change in free energy ($\Delta F = F_1 - F_0$), offering insights into the energetics of particle adsorption and the equilibrium state of the particle at the interface. [check caption Agnese: redraw/uniform?](#)

strong this trapping is, we need to calculate the free energy difference caused by the adsorption of a spherical particle to the interface. In order to do so, we have to consider the free energy of the particle-plus-interface system before and after adsorption.

We start with the case of a particle initially dispersed in water and examine what happens as it adsorbs to an oil–water interface, illustrated by Fig. 30.3 (analogous considerations can be done for air–water interfaces or for particles starting from the non-aqueous phase). In this case, the initial interfacial free energy of the system F_0 is given by the sum of the energies of the oil–water interface, obtained by multiplying the area of the interface A_{ow} by its interfacial tension γ_{ow} , and the energy of the particle–water interface, given by $\gamma_{\text{pw}}A_p$. If the particle adsorbs to the interface and partly protrudes into the oil phase, two additional contributions have to be considered to the final free energy F_1 : the newly created particle–oil interface of energy $\gamma_{\text{po}}A_{\text{po}}$ and the contribution stemming from the fact that an area equal to the particle cross-section at the interface A_i has been removed from the oil–water interface, which has then an energy contribution of $-\gamma_{\text{ow}}A_i$. Moreover, partial exposure to the oil implies that the area of the particle exposed to the water changes from its total surface area A_p to A_{pw} . The free-energy difference upon particle adsorption is therefore given by

$$\Delta F = F_1 - F_0 = \gamma_{\text{pw}}A_{\text{pw}} + \gamma_{\text{po}}A_{\text{po}} + \gamma_{\text{ow}}A_{\text{ow}} - \gamma_{\text{ow}}A_i - \gamma_{\text{pw}}A_p - \gamma_{\text{ow}}A_{\text{ow}}. \quad (30.2)$$

For spherical particles, simple trigonometry allows us to easily define the surface areas based on the particle radius r and its protrusion height into the water h , as illustrated in Fig. 30.3. Taking the definition of the particle contact angle θ given by Young's equation, we can rewrite Eq. (30.2) as [12]

$$\Delta F = -\gamma_{\text{ow}}\pi r^2(1 - |\cos \theta|)^2. \quad (30.3)$$

Here, the contact angle determines the equilibrium position of the particle relative to the interface. If the contact angle is lower than 90° , the particle is mostly immersed in the water (*hydrophilic particle*). If instead the contact angle is larger than 90° , most of the particle surface is exposed to the oil (*hydrophobic particle*). For $\theta = 90^\circ$, the particle sits in the middle (*neutrally wetting*).

From Eq. (30.3) we can deduce three main observations, which will be important for future considerations:

The free energy difference is always negative for any value of the contact angle. This means that it is always favorable to adsorb a particle at a liquid interface, as the free energy is always reduced. This is true under the conditions that a contact angle can be defined, and thus that $|\gamma_{\text{po}} - \gamma_{\text{pw}}| \leq \gamma_{\text{ow}}$, following Young's equation. If this condition is not verified, then it is not energetically favorable for a particle to adsorb to the interface. Most materials colloidal particles are made of, e.g., silica or polystyrene as described in the following, obey this condition.

The free energy of adsorption depends on the square of the particle size. This implies that by increasing the dimensions of the objects adsorbed at the interface, the adsorption energy grows quadratically.

The free energy of adsorption depends on the contact angle. In particular, it vanishes for contact angles approaching 0° and 180° , while it is maximized when $\theta = 90^\circ$. This is directly related to the fact that for neutrally-wetting conditions, the cross section of the particle at the interface is maximized. The particle contact angle depends on the surface properties and on the two liquid phases.

These considerations indicate that the mechanism driving the adsorption of particles to liquid interfaces is the reduction of the interfacial area between the two fluids. If we calculate the typical values of the adsorption free energies for oil–water interfaces and microparticles, we immediately see that adsorption is irreversible. In fact, taking neutrally wetting particles of $1\text{ }\mu\text{m}$ radius at the water–toluene interface ($\gamma_{\text{ow}} = 36\text{ mN m}^{-1}$) we obtain that $\Delta F \approx -3 \times 10^7 k_{\text{B}}T$ [13].¹

As a last note, the calculations reported above are purely energetic and do not take into consideration any kinetic effects related to how the particles adsorb to the interface. In particular, the presence of surface charges can create energetic barriers for adsorption, which can be lowered by screening them. Another commonly

¹The huge values of the trapping energy for solid particles at fluid interfaces are at the basis of their use for the stabilization of emulsions and foams, also called *Pickering emulsions* and *Pickering foams*, from the name of one of the two scientists who first described them (the other being Ramsden) at the beginning of the twentieth century [14, 15]. Essentially, if the surface of a droplet or bubble is coated by a layer of solid particles, they will never spontaneously desorb to allow the coalescence with neighboring droplets or bubbles. Interested readers can refer to Ref. [16].

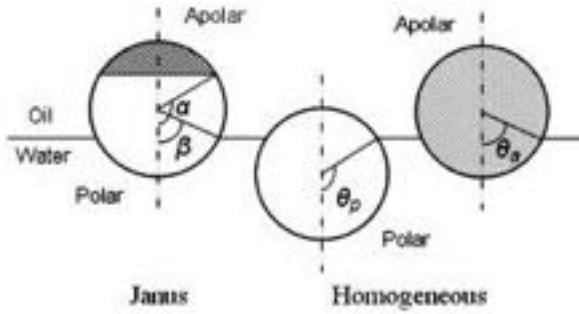


Figure 30.4: **Janus vs homogeneous particles at a oil–water interface.** Comparison between a Janus particle at the oil–water interface and homogeneous particles of materials corresponding to its hydrophilic (polar) and hydrophobic (apolar) lobes with contact angles θ_p and θ_A , respectively. The relative areas of the different surface regions are parametrized by the angle α and the position of the Janus particle relative to the interface by the angle β [17]. Agnese: redraw/uniform?

used strategy to promote adsorption is to add the right amount of surfactants to the system to modify the particle wettability. However, for the purpose of the experiments shown in this Chapter, the focus is not on the adsorption kinetics and microparticles can be directly placed at the interface by spreading them from an alcoholic mixture, e.g., using water–ethanol or water–isopropanol mixtures. In practice, a small droplet of the mixture containing the particles is brought into contact with the interface and the presence of a low-surface tension alcohol allows the effective spreading of the droplet onto the interface, depositing particles which are then irreversibly trapped.

These conclusions are valid for particles with a homogeneous surface chemistry, but the situation is more complex for Janus particles. Here, we focus on the case of Janus particles that present contrasts in wettability between the two surface regions, as is naturally the case for the types of Janus particles considered in this book. In particular, the case of the adsorption of Janus particles of various shapes has been treated in the literature and is reviewed by Ref. [18]. In essence, the determination of the equilibrium position of a Janus particle depends both on the contact angle of the more polar region exposed to the water θ_p and to that of the more apolar region exposed to the oil θ_A , as well as on an angle α that represents the size of the apolar cap. The overall position of the particle relative to the interface can be defined by the angle β , as shown in Fig. (30.4). The same geometrical and energetic calculations that we have carried out for the homogeneous case can now be extended to the Janus case, as first calculated by Ondarçuhu *et al.* [19], leading to the following equations:

$$\text{for } \beta \leq \alpha : \quad F(\beta) = 2\pi r^2 \left[\gamma_{ao}(1 + \cos \alpha) + \gamma_{po}(\cos \beta - \cos \alpha) + \gamma_{pw}(1 - \cos \beta) - \frac{1}{2} \gamma_{ow} \sin^2 \beta \right], \quad (30.4)$$

$$\text{for } \beta \geq \alpha : \quad F(\beta) = 2\pi r^2 \left[\gamma_{ao}(1 + \cos \beta) + \gamma_{aw}(\cos \alpha - \cos \beta) + \gamma_{pw}(1 - \cos \alpha) - \frac{1}{2} \gamma_{ow} \sin^2 \beta \right], \quad (30.5)$$

where γ_{ao} , γ_{po} , γ_{aw} , γ_{pw} , and γ_{ow} are the interfacial energies of the apolar-region–oil, polar-region–oil, apolar-region–water, polar-region–water, and oil–water interfaces, respectively [20]. Minimization of the free energies in Eqs. (30.4) and (30.5) leads to the three scenarios shown in Fig. 30.5. We see that, for either very large or very small values of α , the particles tend to behave as homogeneous ones and their contact angle corresponds to the respective contact angle of the polar or apolar part. However, for intermediate values of α and for large wettability contrasts $\Delta\theta = (\theta_p - \theta_A)/2$, the equilibrium contact angle corresponds to α leading to the fact that the contact line at the interface corresponds to the location of the Janus boundary. The trapping energy is maximized for $\Delta\theta = 90^\circ$ and $\alpha = 90^\circ$ and can become up to three times larger than the maximal energy for a homogeneous particle of the same size [20]. This indicates that for particles that have 50% surface coverage the latter orientation is the preferred one, but for different cap sizes and small wettability contrasts, other configurations may be favored [21]. In this case, it is worth mentioning that various cap orientations can be degenerate until the cap touches the interface or the energy minimum can be very shallow so that multiple orientations can be accessed by thermal fluctuations. This has strong implications for the breaking of symmetry

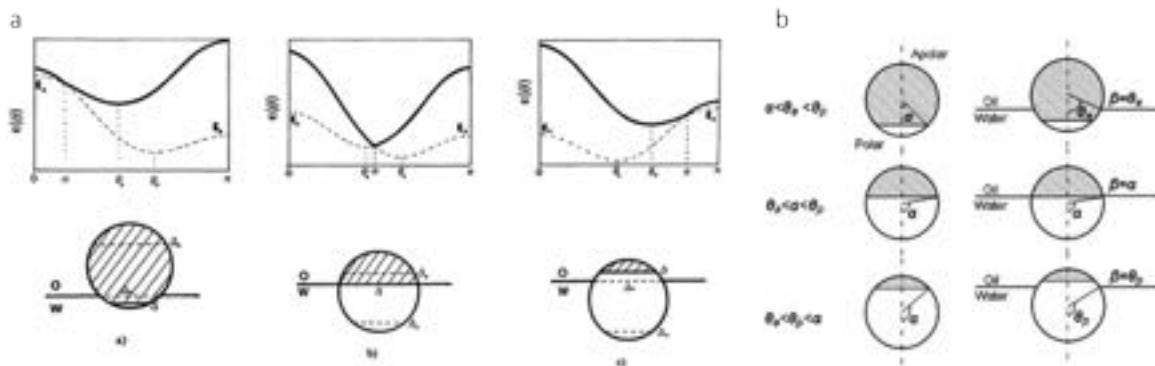


Figure 30.5: Interfacial behavior of Janus vs homogeneous particles. Free energy landscapes of a Janus particle at a water–oil interface, visualizing three possible equilibrium positions based on different contact angles and cap sizes. The total surface free energy changes based on the angle between the interface and the particle axis (β). The configurations (from left to right) represent scenarios where $\alpha < \theta_A < \theta_P$, $\theta_A < \alpha < \theta_P$, and $\theta_A < \theta_P < \alpha$. This highlights how particle surface characteristics influence interface behavior. [check caption Agnese: redraw and simplify as described in the revised caption.](#)

in the plane of the interface, as it will be discussed later.

All of these equilibrium calculations have been carried out under the assumption that the contact line is able to slide smoothly over the surface of the particle to reach the minimum energy configuration and that, consequently, the particle is free to rotate in all directions with respect to the interface. In particular, rotation within the interface plane does not necessarily imply relative motion between the particle surface and the contact line; the contact line can move rigidly with the particle. Nonetheless, out-of-plane rotations without changes of the particle contact angle require motion of the contact line. Typically, the caps of Janus particles, e.g., as obtained by metal evaporation, are not perfectly smooth and may have irregular shapes, especially at the edge. Consequently, the contact line may find topographical and chemical heterogeneities on its path and may not be able to move freely. In fact, if Janus particles either adsorb spontaneously to a fluid interface by diffusion or other directed transport mechanisms (e.g., sedimentation) or are spread at the interface using a solvent, they will be at first randomly oriented and positioned at the interface. Energy minimization then dictates that they need to reorient and move across the interface in order to reach the equilibrium position defined by the above-mentioned equations. However, if the contact line gets pinned during this process with large energy barriers, then the particle may not be able to move and reorient. As a result, the Janus particle will be trapped in a metastable configuration at the interface, where both its vertical position relative to the interface plane and its orientation may not coincide with the minimum energy ones. This fact is more common than one might think (it might even be the norm and an unobstructed motion of the contact line on the surface of a Janus particle is to be consider an exception). Fig. 30.6 shows very clear optical microscopy images of Janus glass beads adsorbed at a water-air interface with any orientation and while they could freely rotate within the plane, their orientation relative to the interface was frozen [22]. As schematically represented, out-of-plane rotations can be hindered by contact line pinning and we will show later that this pinning has strong consequences on active motion. A very interesting comparison on the effect of contact line pinning on wetting of colloidal particles and solid surfaces can be found here [23]. To add some numbers to these considerations, one can estimate that the energy barrier to dewet a single surface asperity with the shape of a spherical cap with height h , i.e., corresponding to the energy barrier seen by the contact line that has to move over that asperity, can be written as $\delta f = \pi h^2 \gamma_{\text{ow}} \cos \theta$, where θ is the contact angle of the spherical asperity. For asperities in the 10 nm-range with moderately hydrophilic surfaces, these energies can easily be of the order of $100 k_B T$, leading to irreversible pinning [24].

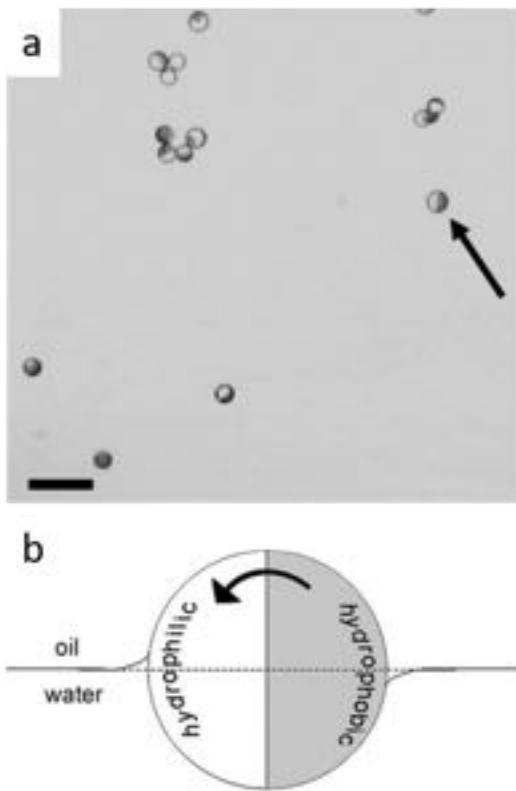


Figure 30.6: Orientation of Janus particles at an air–water interface. (a) Optical image of Janus particles (glass beads coated by a Cr–Au layer) at the interface. The particles feature any orientation with respect to the interface, including one perpendicular to the interface (indicated by the arrow). [Agnese: add value above scale bar 100 \$\mu\text{m}\$](#) (b) Simplified schematic of a Janus particle orientated perpendicular to a water–oil interface, indicating the drive for reorientation. [Agnese: Change figure structure to horizontal](#) Reproduced from Ref. [22].

30.3 Contact angle measurement

In order to calculate the adsorption energies mentioned in the previous section, one needs to measure the contact angle of individual particles adsorbed at water–oil or water–air interfaces. Extensive descriptions of various available techniques have been recently reviewed by Refs. [25, 26, 27]. Here, we focus on two different methods: the gel trapping technique (GTT), and the freeze-fracture and shadow-casting cryo-SEM (FreSca cryo-SEM).

Before we look at these techniques, the contact angle does not only determine the adsorption energy, but it also influences directly the way in which the particle moves at the interface, i.e., by simple diffusion or by active Brownian motion. The viscous drag coefficient experienced by a particle at a water–oil interface is a function $f(A_{\text{pw}}, A_{\text{po}}, A_i, \eta_w, \eta_o, \eta_i)$ [28, 29], where in addition to the bulk viscosities of the two phases η_w and η_o , the interface itself may present its own complex viscosity η_i (e.g., due to the presence of monolayers of surface-active species, such as surfactant molecules or polymers). Finally, the contact angle also strongly affects the interactions between particles at the interface (e.g., electrostatics [30]).

Going back to the measurement techniques, the existing methods can be classified based on whether they can directly measure the contact angle of individual particles or whether an average value for θ is instead extracted from macroscopic measurements on particle monolayers at a fluid interface. In general, single-particle methods are applicable to micron-sized colloids, like those often of interest in this book. Conversely, ensemble measurements are also applicable to smaller objects, down to nanoparticles. In this latter case, the wetting of the single particles cannot be resolved and only an average contact angle is extracted from the macroscopic properties of the interface. In addition to the difference in length scales, being able to resolve the

single particle is very important for Janus colloids at interfaces. As we mentioned above and as we will describe again later in more detail, the swimming behavior of Janus colloids at interfaces is closely connected to their position and orientation at the interface. Therefore, heterogeneity in the swimming behavior can be expected to derive from a distribution of contact angles and orientations, which requires single-particle resolution to be properly described.

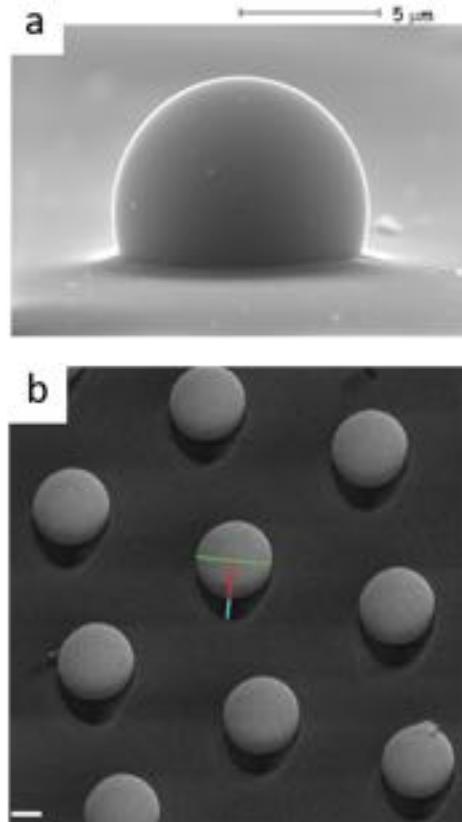


Figure 30.7: Contact angle measurement of colloidal particles at an interface. (a) SEM image showcasing sulfate PS particles, with an average diameter of $9.6 \mu\text{m}$, that have been successfully trapped on a PDMS surface via the *gel trapping technique* (GTT). The particles in this image were initially positioned at either a water–decane or a water–air interface prior to the GTT procedure. **carefully check caption, especially the last sentence** Reproduced from Ref. [31]. (b) Measurement of the contact angle of a 500-nm amidine latex particle at a water–decane interface, accomplished using the *freeze-fracture and shadow-casting cryo-SEM* (FreSCa Cryo-SEM) technique. The essential quantities necessary for extracting the contact angle from SEM images are highlighted: particle diameter (shown in green), projected height (marked in red+cyan), and shadow length (represented in cyan). This cryo-SEM image provides a direct visualization of the immobilized particles protruding from the interface, their shadows aiding in the determination of the contact angle. **Agnese: Add to figure value of scale bar 200 nm.** Reproduced from Ref. [32]. **Agnese: Change figure structure to horizontal.**

Referring to the sketch reported in Fig. 30.3, there exists a very simple geometrical relation that connects the contact angle of a spherical particle of radius r to its protrusion height h into the water $\cos \theta = [h/r - 1]$. In practice, measuring θ often is done by measuring the protrusion height h (or its complementary protrusion into the oil $2r - h$). In the case of the two techniques described below, this is done in a two-step process: first the particles are immobilized at the interface and then their protrusion is measured by means of electron or atomic force microscopy.

The first and most popular technique is the so-called *gel trapping technique* (GTT), which makes use of gellan gum as a gelling agent to *freeze* the interface upon cooling. The practical implementation of this

technique is showed in the recorded experiments Agnese: Add recorded experiment to repository and add here the reference to it.. Can these experiments be provided as an exercise? Briefly, the process consists in spreading the particles at the interface between the desired medium and a water solution of gellan at a temperature above the gellan's gelling temperature ($\approx 45^\circ\text{C}$ for a 2wt% system). Upon cooling, the gellan sets and traps the particles at the interface. The trapped monolayer is then transferred to a PDMS or glue (either superglue or UV glue) film and the position of the particles with respect to the PDMS can be measured directly by means of SEM [31]. The contact angle is determined after geometrical analysis of the SEM images (Fig. 30.7a). This technique allows measuring the contact angle of particles ranging from 100 nm to 100 μm at air–water and oil–water interfaces and it has been applied to a broad range of colloids of different shapes and materials [33, 34, 35, 36, 37]. As an alternative to SEM, the height of the particle can also be measured with an AFM, allowing the extension of the technique to smaller particles, albeit with some notes of caution on single-particle resolution [38]. GTT is an easy technique with a relatively good accuracy (5° to 10°), and is well-suited for single-particle analysis. Nevertheless, it presents several limitations in terms of system's choice and resolution. Gellan is a non-surface active polysaccharide whose structural units have a length scale of the order of 10 nm [39] and therefore inhomogeneities in the trapping and problems in defining the contact lines are to be expected for particles with cross sections below 100 nm at the interface. Moreover, the choice of the liquid phases is limited to non-volatile oils due to the high temperatures involved in the gel processing, and changes to the aqueous phase, such as ionic strength, pH or addition of surfactants, may affect the gelling. The high temperature may also induce particle deformation [40] and the impact of the slow cooling rate on contact angle evolutions is not known. GTT is moreover typically performed using a spreading solvent to ensure sufficient trapping upon particle injection at the interface and the use of spreading solvents has been demonstrated to affect strongly the particle wetting properties, even turning particles form hydrophilic to hydrophobic [41, 42]. However, if a spreading protocol is used to localize the Janus particles at the interface, the contact angle measurements should be carried under the same conditions. can you clarify the previous sentence?

In order to overcome some of the limitations of GTT, in 2011, Isa *et al.* [32] proposed an alternative to immobilize and image particles at liquid–liquid interfaces based on *freeze-fracture and shadow-casting cryo-SEM* (FreSCa cryo-SEM). In this case, the immobilization is obtained by shock-freezing the samples, with cooling rates so high ($\sim 10^3 \text{ K}^{-1}$) that water vitrifies with negligible thermal expansion to rapidly trap the particles at the interface. The samples are subsequently fractured and unidirectionally metal-coated with a well-defined tilt angle before imaging in a cryo-SEM. The presence of a particle-laden, macroscopically flat interface acts as a weak fracture plane and exposes the particles immobilized in the water after removal of the oil. Because of the oblique metal deposition, immobilized particles protruding from the interface cast a shadow. By measuring the shadow length and the projected height, and knowing the shadowing angle and the particle shape, the protrusion height into the oil can be calculated, and thus the contact angle of individual objects (Fig. 30.7b). Can you add an exercise about this, providing some data for SI? The major advantages of this method are its high accuracy and flexibility. The resolution of the technique is limited by the pixel resolution of the cryo-SEM, leading to the measurement of contact angles of individual particles as small as 10 nm in radius, which is the state of the art in terms of single-particle resolution. Additionally, this technique can be applied to virtually any particle material or shape [43, 44, 45, 46] and to any water–liquid interface, e.g., also in the presence of surfactants [47]. It has also been applied to Janus spheres, confirming the energetic predictions mentioned above [21]. Its main disadvantages are the complex machinery required for the process and its limitation to water–liquid interfaces.

30.4 Measurement of interfacial tension

Before moving on to the description of active colloids at fluid interfaces, we dive into a short excursus on the measurement of interfacial tension between liquids, as the measurement of this quantity has important consequences on the adsorption and propulsion of particles at interfaces. Very often values of interfacial tension are simply taken from tabulated literature data, but this quantity is highly sensitive to the measurement conditions, in particular the presence of surface-active species. It is therefore very useful to understand how it can be robustly measured in relation to the desired experiments. As we saw, the adsorption energy of a particle scales linearly with the oil–water (or air–water) interfacial tension and the latter quantity also affects the contact angle of the particles. Finally, we will also see later, that one can manipulate interfacial tension to drive the propulsion of colloids at the interface. Given the importance of measuring interfacial tension, the second ex-

perimental module of this lecture deals with its measurement under different conditions to highlight important experimental factors. [This sentence needs to be adapted for the book based on how the data are included/pre-sented] Agnese: Add recorded experiment/materials to repository and add here the reference to it.. Can these experiments be provided as an exercise?

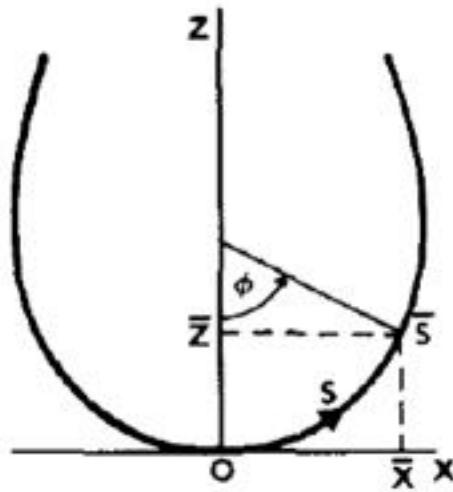


Figure 30.8: **Geometry of a pendant drop.** The figure shows a pendant drop of a denser fluid (e.g., water) hanging from a fine needle and surrounded by a lighter fluid (e.g., oil or air). This situation corresponds to the solution of the Young-Laplace equation in the pendant drop geometry, with the pressure difference across the drop and the two principal radii of curvature taken into account. [check caption Agnese: redraw](#)

There are many ways in which the surface tension of a liquid can be measured. Here, we will focus on the *pendant drop method*, which has several technical advantages that will be discussed later. Most methods for surface tension measurements rely on the solution of the Young-Laplace (YL) equation in different geometries corresponding to the experimental conditions. The YL equation

$$\gamma \left(\frac{1}{R_1} + \frac{1}{R_2} \right) = \Delta P \quad (30.6)$$

relates the surface tension to the pressure drop ΔP across a surface with the two principal radii of curvature R_1 and R_2 .

For the case of a pendant drop of a denser fluid (e.g., water) created at the tip of a fine needle and surrounded by a lighter fluid (e.g., oil or air), the only force acting on the droplet is gravity, and the pressure drop is a linear function of height from a reference plane and can be written as

$$\Delta P = \Delta P_0 + \Delta \rho g z, \quad (30.7)$$

where ΔP_0 is the pressure difference in the reference plane, $\Delta \rho$ is the density difference between the two fluids, g is the gravitational acceleration and z is the vertical coordinate, as show in Fig. 30.8. By looking at the coordinate system reported in Fig. 30.8, we can combine Eqs. (30.6) and (30.7) into

$$\gamma \left(\frac{1}{R_1} + \frac{\sin \phi}{x} \right) = \frac{2\gamma}{R_0} + \Delta \rho g z, \quad (30.8)$$

where R_1 turns in the plane of the paper and $R_2 = x/\sin \phi$ rotates around the symmetry axis of the drop; R_0 is the radius of curvature at the origin (the drop's apex) where $R_1 = R_2 = R_0$ and the angle ϕ is defined as in Fig. 30.8. Given the axial symmetry, the drop profile can be represented in the form of $x = x(s)$ and $z = z(s)$,

where s is a coordinate along it. In particular, this yields the following equations:

$$\begin{cases} \frac{dx}{ds} = \cos \phi, \\ \frac{dz}{ds} = \sin \phi. \end{cases} \quad (30.9)$$

Using the definition that $\frac{1}{R_1} = \frac{d\phi}{ds}$, Eq. (30.6) can be rewritten as

$$\frac{d\phi}{ds} = \frac{2}{R_0} + \frac{\Delta\rho g}{\gamma} z - \frac{\sin \phi}{x}. \quad (30.10)$$

The shape of a pendant drop is then found by solving Eqs. (30.9) and (30.10) with the boundary conditions $x(0) = z(0) = \phi(0) = 0$. In an experiment, the value of γ is measured by extracting the experimental drop profile from images and knowing $\Delta\rho$ between the two fluids. The drop profile extraction is done digitally. [Agnese: Add recorded experiment/materials to repository and add here the reference to it. Can these experiments be provided as an exercise?](#) This technique has many advantages relative to competing ones. In particular, it enables the user to work with a very broad range of fluids, it makes it possible to look at dynamical changes of interfacial tension by simply measuring the drop shape as a function of time and thus enables one to assess the role played by surface-active species and impurities in a direct way.

30.5 Self-propulsion at fluid interfaces

This section discusses some recent investigations, focusing primarily on the behavior of self-propelled catalytic Janus particles at fluid interfaces, namely oil–water interfaces. A brief overview of their interaction at the interfaces under varying concentrations of hydrogen peroxide is presented [5, 48]. Additionally, we shed light on the intriguing phenomenon of *Marangoni propulsion* and how surface tension gradients can foster highly efficient propulsion at the microscale [6]. A profound understanding of these self-propelled systems can open up new avenues in a multitude of fields such as drug delivery, environmental remediation, and microscale robotics.

30.5.1 Catalytic particles at fluid interfaces

The properties and behavior of catalytic Janus swimmers in bulk aqueous environments have already been extensively covered in this book [Agnese: add reference to specific chapters](#). Here, we briefly report on some recent work that investigates their behavior at oil–water interfaces, and directly connects with the notions introduced above [5]. The microswimmers consist of fluorescent, platinum-coated Janus polystyrene (PS) microspheres (radius $R = 1.4 \mu\text{m}$, thickness of the Pt-layer = 2 nm), which self-propel in H_2O_2 -enriched aqueous solutions thanks to the catalytic reaction promoted by the Pt-patch [48]. We confine these particles at a water-hexadecane interface using a water-isopropanol mixture in the presence of different amounts of hydrogen peroxide in the water and study their motion.

Upon adsorption, the particles are irreversibly bound to the interface and self-propel within the interface plane with a velocity that depends on the concentration of hydrogen peroxide, as shown in Fig. 30.9a. Here, we see that the data at the interface (red points) approximately give a velocity 2–3 times smaller than the same catalytic particles in bulk aqueous environments (blue data). This difference can be partly explained by the fact that a portion of the swimmers is exposed to the more viscous hexadecane. Contact angle measurements performed using both GTT and FrSCa show that the particles have a contact angle of approximately 100° , implying that two regions of approximately the same size are exposed to either fluid. Macroscopic contact angle measurements on bare PS surfaces and PS surfaces coated by 2 nm of Pt show that the contrast in wettability is rather small, i.e., $\Delta\theta \approx 10^\circ$, which indicates that there may be two different orientations of the particles at the interface with relatively similar energy values. In particular, the deepest minimum corresponds to the orientation with the Pt cap facing the water and the PS cap fully exposed to the oil. GTT and FrSCa measurements confirm the existence of these two orientations, as we will see below.

However, a perfect symmetry of the Pt cap in the interface plane would not lead to any propulsion, and the fact that the particles propel along the interface plane indicates that the chemical symmetry needs to be

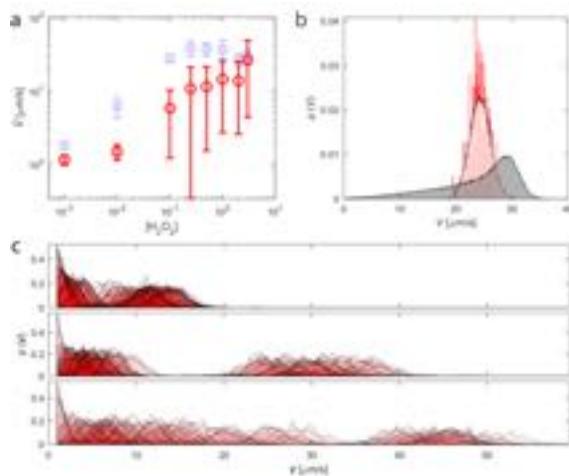


Figure 30.9: Propulsion velocity of catalytic Janus particles at a fluid interface. (a) Mean propulsion velocity as a function of hydrogen peroxide concentration in bulk (blue) and at the water-hexadecane interface (red). (b) Probability distribution of swimming velocities. Red data: experimental data for particles at the interface. White-shaded data: Brownian dynamics simulations for ABM in 2D. Grey-shaded data: 2D projection of velocities from Brownian dynamics simulations of 3D ABM. (c) Velocity distributions of approximately 100 active particles at the water-oil interface for H_2O_2 concentrations of 0.1% v/v (top), 1.0% v/v (center), and 3.0% v/v (bottom), showing the presence of two main populations corresponding to fast and slow swimmers. Reproduced from Ref. [5]. Agnese: Make into a single line with a, b, c

broken within the interface plane and that a non-centrosymmetric distribution of Pt is exposed to the aqueous phase. The chemical asymmetry in the interface plane is then defining the propulsion direction, given that the particles are vertically confined within the interface. By further examining the particle trajectories, we observe the presence of two different particle populations swimming with distinct velocities (Fig. 30.9b), which we can hypothesize belong to the two different main particle orientations. A close inspection of the velocity distributions and a comparison with numerical simulations confirms that the rotation of the particle is confined within the 2D interface plane and that the particles do not rotate out of plane (Fig. 30.9b). This observation is again suggestive of the presence of two metastable orientations separated by a barrier with a height greater than thermal energy. By employing fluorescent particles and imaging them with an inverted microscope, we verified that the two populations of velocities are associated with the two orientations (Fig. 30.10). In particular, and perhaps at first sight counter-intuitively, we see that the faster particles are the ones with most of the Pt cap exposed to the oil and only a small portion of it into aqueous phase. This is due to the fact that propulsion is not proportional to the total amount of decomposed peroxide, but to the chemical gradient, which is steepest for the configuration shown in Fig. 30.10b. We also see that the cap needs to cross the interface at an angle in order to generate an asymmetric gradient.

Finally, we examined the values of the rotational reorientation times at the interface and found that they were, on average, significantly greater than the ones for the particles in bulk (Fig. 30.11). This observation is aligned with the fact that the roughness of the Pt cap indices contact line pinning and subsequent interface deformations that increase the effective size of the Janus colloids at the interface and thus their rotational drag coefficient [49]. We remark that other studies with catalytic swimmers at air-water interfaces showed similar enhancement of persistence in the trajectories [50] and that, for thicker Pt caps, the heterogeneity stemming from contact-line pinning was large enough that an effective misalignment between the distribution of the catalytic material and the particle shape at the interface arose, causing particles to perform curvilinear trajectories [51] similar to the ones reported in bulk for L-shaped colloids [52].

Is it possible to add some exercises based on the experimental materials?

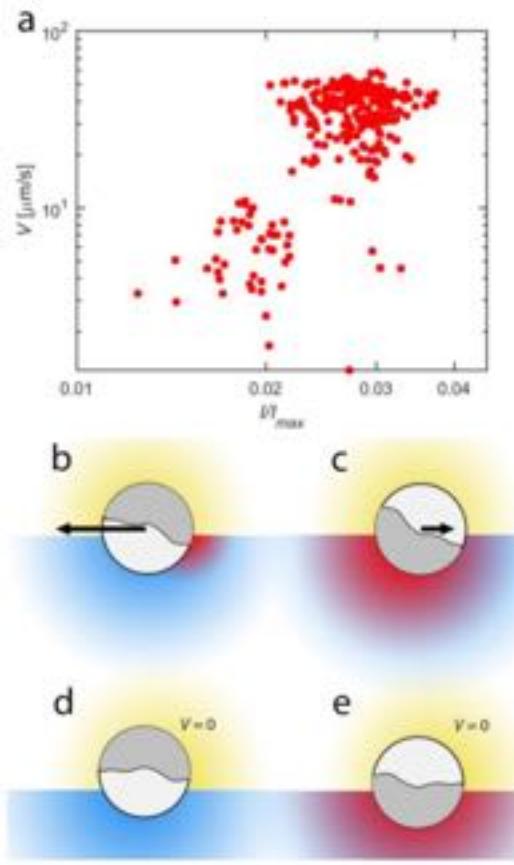


Figure 30.10: **Orientation and velocity of catalytic Janus particles at a fluid interface.** (a) Swimming speed V as a function of normalized brightness I/I_{max}) for a set of active swimmers at the water-hexadecane interface (3.0% v/v H₂O₂). High (low) fluorescence intensity corresponds to particles whose cap is mostly in the oil (water) phase. (b) Fast and (c) slow swimmers have the Pt cap preferentially wetted by oil and water, respectively. (d-e) Particles that have only one hemisphere in water do not swim because no gradient is established parallel to the interface. Reproduced from Ref. [5]. Agnese: make it in a single line.

30.5.2 Marangoni propulsion

In this second example, we experimentally demonstrate that surface tension gradients can generate very efficient propulsion at the microscale [6]. In particular, we have used light-absorbing Janus particles, which upon heating, generate asymmetric temperature gradients, and, as a consequence, controlled surface tension differences across the particle that lead to self-propulsion.

In the presence of surface tension gradients, momentum conservation at the interface prescribes the existence of tangential stresses, called Marangoni stresses, defined as:

$$\nabla_s \sigma(\Gamma, T) = \frac{\partial \sigma(\Gamma, T)}{\partial T} \nabla_s T + \frac{\partial \sigma(\Gamma, T)}{\partial \Gamma} \nabla_s \Gamma, \quad (30.11)$$

where $\sigma(\Gamma, T)$ is the interfacial tension, which is a function of temperature T and surface excess concentration of a surface-active species Γ .² $\nabla_s = (\vec{I} - \vec{n}\vec{n}) \cdot \nabla$ is the surface gradient operator, with \vec{I} the unit tensor and \vec{n} the normal to the interface. Here, we identify two sources of stress: temperature and surface excess concentration gradients, whose magnitude is set by $\partial \sigma(\Gamma, T)/\partial T = \beta$ and $\partial \sigma(\Gamma, T)/\partial \Gamma = \delta$, respectively [53]. Imposing a

²We use σ instead of γ to denote the surface tension here in order to avoid confusion with Γ , which historically denotes the interfacial concentration.

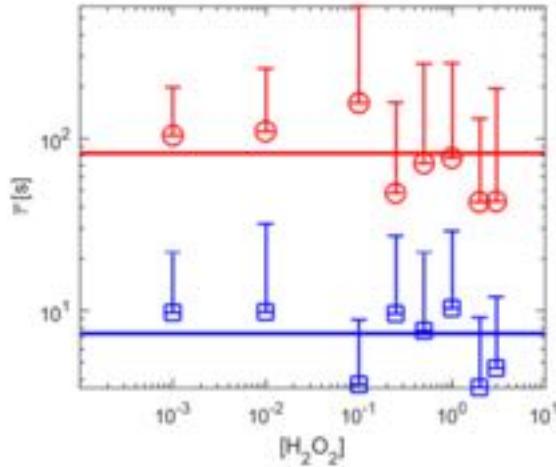


Figure 30.11: **Slowing down of the reorientation dynamics of Janus particles at a fluid interface.** Average reorientation time $\bar{\tau}$ for Janus particles in bulk (blue) and at the water–hexadecane interface (red), showing significant slowing down of rotational dynamics at the interface. Reproduced from Ref. [5].

force balance on the particle’s surface ∂P and contact line L yields

$$\int_{\partial P} \boldsymbol{\sigma} \cdot \mathbf{n}_p \, dS = \int_L \boldsymbol{\sigma}(\Gamma, T) \mathbf{t} \, dl, \quad (30.12)$$

where $\boldsymbol{\sigma}$ is the bulk stress tensor, \mathbf{n}_p is the unit vector normal to the particle surface, and \mathbf{t} is the unit vector tangential to the interface and normal to the contact line. Together with a no-slip boundary condition at the particle surface, Eq. (30.12) allows solving for the particle velocity \mathbf{V} imposed by the Marangoni stress (see Ref. [6] for more details). For a characteristic interfacial tension difference $\Delta\sigma$, simple dimensional arguments lead to predicting a propulsion speed $V \propto \Delta\sigma/\eta$, where η is an effective viscosity experienced by the particle straddling the interface. Considering thermal Marangoni effects (a.k.a. thermocapillarity) alone, the predicted self-propulsion speed is given by $V \approx (\partial\sigma/\partial T)\Delta T/(10\eta)$ [54]. Typical values of $\partial\sigma/\partial T$ for oil–water interfaces are $\mathcal{O}(10^{-4} \text{ N m}^{-1} \text{ K}^{-1})$, leading to speeds $V = \mathcal{O}(1 \text{ cm s}^{-1})$, independent of particle size and indeed corresponding to 10^4 body-lengths per second for microparticles, for $\Delta T = 1 \text{ K}$ and $\eta = \mathcal{O}(10^{-3} \text{ Pas})$ [54]. The magnitude of $\partial\sigma/\partial T$ is then able to set macroscopic objects in motion, as for instance shown for the propulsion of centimeter-sized objects [55] and the rotation of micro-gears suspended at a water–air interface [56]. Similar considerations can be made for solutal Marangoni propulsion [57, 58], as popularized by camphor or soap boats releasing surfactant at one end [2, 3], or for the motion of active droplets [59].

Our interfacial microswimmers, or *Marangoni surfers*, are Janus silica microparticles (radius $R = 3.15 \mu\text{m}$) sputter-coated with a 100-nm-thick hemisphere of gold and confined at an interface between MilliQ water and dodecane. The surface heterogeneity generated by the thick metallic caps effectively pins the Janus particles in random orientations with respect to the interface [60, 61], leading to caps typically crossing the interface (Fig. 30.12). Asymmetric heating of the particles is achieved by illuminating them with green laser light. Light absorption by the gold cap creates an asymmetric temperature profile around the particles, thus generating Marangoni stresses that propel them with velocity V and the gold cap oriented towards the back [62]. This thermally-induced motion consequently can induce an asymmetric distribution of surfactants in the vicinity of the particles, in turns generating solutal Marangoni stresses. The balance between these two components drives the propulsion.

We measure the particle speeds as a function of incident laser power, which we convert into a ΔT by carefully calibrating the local heating of the fluids induced by the gold caps relative to the critical temperature of a water–lutidine mixture. In Fig. 30.13 we show the velocity V as a function of a dimensionless Marangoni stress Π define, given by the ratio of thermal-to-solutal stresses and identify the presence of two linear regimes. At low values of Π , both solutal and thermal stresses contribute to determining the propulsion velocity, which depends linearly on ΔT . For values of $\Pi > 1$, the solutal stresses saturate to their maximum value given by

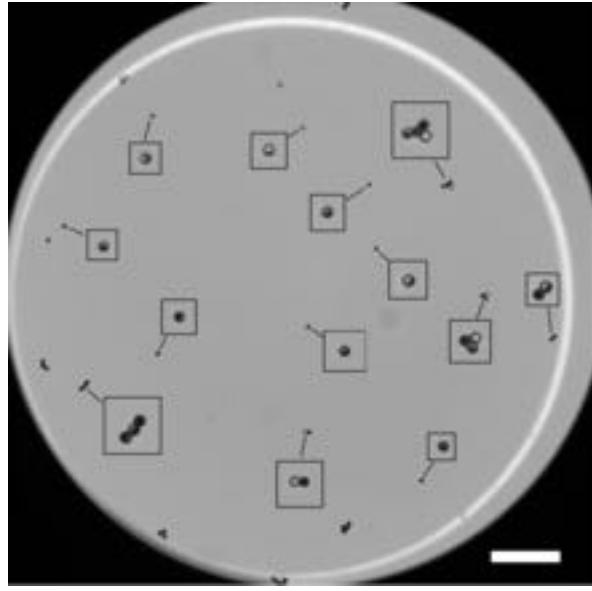


Figure 30.12: **Image of the silica Janus particles at the water–dodecane interface.** Each box shows a zoomed-in image of the particles, demonstrating the presence of random orientations of the cap relative to the interface. [Agnese: The scale bar is 100 μm on the figure.](#) Reproduced from Ref. [6]. [check reference](#)

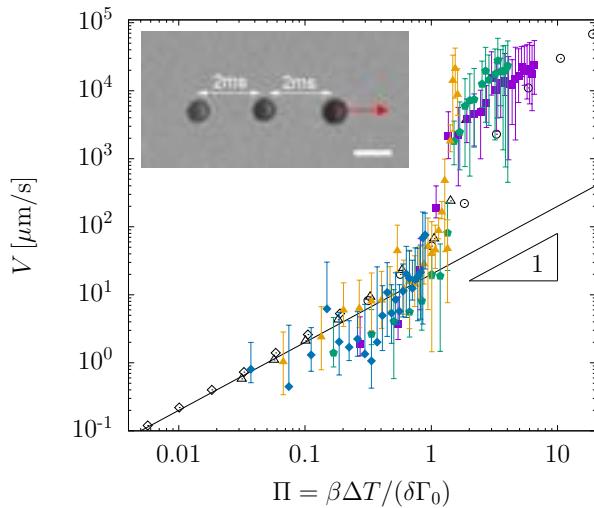


Figure 30.13: **Propulsion speed V as a function of the dimensionless Marangoni stress Π .** Allegedly pristine interface (purple squares) and increasing concentrations of SDS in the water phase: $C = 10^{-7} \text{ mol L}^{-1}$ (green pentagons), $10^{-5} \text{ mol L}^{-1}$ (orange triangles), and $10^{-3} \text{ mol L}^{-1}$ (blue diamonds). Black open symbols are corresponding simulated data. The error bars represent the standard deviations of particle velocities. (inset) Image overlay of a particle propelling with $V \approx 10 \text{ mm s}^{-1}$. [Agnese: add scale bar = 10 μm value to figure.](#)

$\delta\Gamma_0$ (with Γ_0 the concentration of surface-active species at rest, away from the moving Janus particle [\[Define δ too\]](#)) and the propulsion is dominated by purely thermal stresses, which also scale linearly with ΔT .

Of particular importance for anyone interested in performing these experiments and connected to the previous discussions on the measurements of interfacial tension, the purple data in Fig. 30.13 show that the effect of surface contamination cannot be truly eliminated in the experiments. In fact, they refer to the case of an allegedly pristine water-dodecane interface, where only thermal stresses should be present. However, in spite

of all efforts for cleanliness, solutal effects are always present. For comparison, the simulation results are also plotted (empty symbols) and the experimental behavior can be reconciled by introducing a surface excess concentration of order $\Gamma_0 = 10^{-7} \text{ mol m}^{-2}$. These minute values of Γ_0 correspond to unavoidable environmental trace contaminations [63, 64, 65], which have a hardly measurable effect on the absolute level of the surface tension. However, as low as the absolute levels are, gradients of the surface excess concentration can still significantly alter the hydrodynamics in sensitive experiments, especially at small length scales, where Marangoni stresses become increasingly important [66].

To confirm the role played by surface-active species, we purposely add controlled amounts of a water-soluble surfactant (sodium dodecyl sulfate, SDS). The choice of SDS is motivated by the fact that it has a negligible surface viscosity, and thus we do not expect surface rheology to affect particle motion [67]. Consistently with the balance between the different components of Marangoni stresses, we observe that an increased amount of SDS causes a shift of the transition towards higher values of ΔT and an overall reduced particle velocity over the same ΔT window (Fig. 30.13).

Is it possible to add some exercises based on the experimental materials?

30.6 Problems

Problem 30.1: ?? ?? Can you add some problems 2-3?

Bibliography

- [1] R.F. Ismagilov, A. Schwartz, N. Bowden, and G.M. Whitesides. Autonomous movement and self-assembly. *Angewandte Chemie - International Edition*, 41(4):652–654, 2002.
- [2] R. J. Strutt. IV. Measurements of the amount of oil necessary in order to check the motions of camphor upon water. *Proceedings of the Royal Society of London*, 47(286-291):364–367, 1890.
- [3] S. Sur, H. Masoud, and J. P. Rothstein. Translational and rotational motion of disk-shaped marangoni surfers. *Physics of Fluids*, 31(10):102101, 2019.
- [4] J. W. M. Bush and D. L. Hu. Walking on water: Biocomotion at the interface. *Annual Review of Fluid Mechanics*, 38(1):339–369, 2006.
- [5] K. Dietrich, D. Renggli, M. Zanini, G. Volpe, I. Buttinoni, and L. Isa. Two-dimensional nature of the active Brownian motion of catalytic microswimmers at solid and liquid interfaces. *New Journal of Physics*, 19(6):065008, 2017.
- [6] K. Dietrich, N. Jaansson, I. Buttinoni, G. Volpe, and L. Isa. Microscale marangoni surfers. *Physical Review Letters*, 125:098001, 2020.
- [7] T. Yao, N. G. Chisholm, E. B. Steager, and K. J. Stebe. Directed assembly and micro-manipulation of passive particles at fluid interfaces via capillarity using a magnetic micro-robot. *Applied Physics Letters*, 116(4):043702, 2020.
- [8] W. Fei, Y. Gu, and K.J.M. Bishop. Active colloidal particles at fluid-fluid interfaces. *Current Opinion in Colloid and Interface Science*, 32:57–68, 2017.
- [9] M. Oettel and S. Dietrich. Colloidal interactions at fluid interfaces. *Langmuir*, 24(4):1425–1441, 2008.
- [10] V. Garbin, J.C. Crocker, and K.J. Stebe. Nanoparticles at fluid interfaces: Exploiting capping ligands to control adsorption, stability and dynamics. *Journal of Colloid and Interface Science*, 387(1):1–11, 2012.
- [11] A. Domínguez, P. Malgaretti, M. N. Popescu, and S. Dietrich. Effective interaction between active colloids and fluid interfaces induced by marangoni flows. *Physical Review Letters*, 116:078301, 2016.

- [12] P.I. Pieranski. Two-dimensional interfacial colloidal crystals. *Physical Review Letters*, 45(7):569–572, 1980.
- [13] B.P. Binks and S.O. Lumsdon. Influence of particle wettability on the type and stability of surfactant-free emulsions. *Langmuir*, 16(23):8622–8631, 2000.
- [14] S.U. Pickering. Cxvi.-emulsions. *Journal of the Chemical Society, Transactions*, 91(0):2001–2021, 1907.
- [15] W. Ramsden. Separation of solids in the surface-layers of solutions and 'suspensions' (observations on surface-membranes, bubbles, emulsions, and mechanical coagulation). – Preliminary account. *Proceedings of the Royal Society of London*, 72(477-486):156–164, 1903.
- [16] B.P. Binks and T.S. Horozov. Colloidal particles at liquid interfaces: an introduction. *Colloidal Particles at Liquid Interfaces*, Chapt, 1:1–74, 2006.
- [17] S. Jiang and S. Granick. Janus balance of amphiphilic colloidal particles. *The Journal of Chemical Physics*, 127(16):161102, 2007.
- [18] A. Kumar, B.J. Park, F. Tu, and D. Lee. Amphiphilic Janus particles at fluid interfaces. *Soft Matter*, 9:6604–6617, 2013.
- [19] Ondarçuhu, T., Fabre, P., Raphaël, E., and Veyssié, M. Specific properties of amphiphilic particles at fluid interfaces. *Journal de Physique*, 51(14):1527–1536, 1990.
- [20] B.P. Binks and P.D.I. Fletcher. Particles adsorbed at the oil-water interface: A theoretical comparison between spheres of uniform wettability and "Janus" particles. *Langmuir*, 17(16):4708–4710, 2001.
- [21] A. Synytska, A. Kirillova, and L. Isa. Synthesis and contact angle measurements of Janus particles. *ChemPlusChem*, 79(5):656–661, 2014.
- [22] D.J. Adams, S. Adams, J. Melrose, and A.C. Weaver. Influence of particle surface roughness on the behaviour of Janus particles at interfaces. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 317(1-3):360–365, 2008.
- [23] A. Stocco and M. Nobili. A comparison between liquid drops and solid particles in partial wetting. *Advances in Colloid and Interface Science*, 247:223–233, 2017. Dominique Langevin Festschrift: Four Decades Opening Gates in Colloid and Interface Science.
- [24] M. Zanini, C. Marschelke, S. E. Anachkov, E. Marini, A. Synytska, and L. Isa. Universal emulsion stabilization from the arrested adsorption of rough particles at liquid-liquid interfaces. *Nature Communications*, 8(1):15701, 2017.
- [25] A. Maestro, E. Guzman, F. Ortega, and R.G. Rubio. Contact angle of micro-and nanoparticles at fluid interfaces. *Current Opinion in Colloid and Interface Science*, 19(4):355–367, 2014.
- [26] S. Cappelli, Q. Xie, J. Harting, A.M. de Jong, and M.W.J. Prins. Dynamic wetting: status and prospective of single particle based experiments and simulations. *New Biotechnology*, 32(5):420 – 432, 2015.
- [27] M. Zanini and L. Isa. Particle contact angles at fluid interfaces: pushing the boundary beyond hard uniform spherical colloids. *Journal of Physics: Condensed Matter*, 28(31):313002, 2016.
- [28] T.M. Fischer, P. Dhar, and P. Heinig. The viscous drag of spheres and filaments moving in membranes or monolayers. *Journal of Fluid Mechanics*, 558:451–475, 7 2006.
- [29] Y. Peng, W. Chen, T.M. Fischer, D.A. Weitz, and P. Tong. Short-time self-diffusion of nearly hard spheres at an oil-water interface. *Journal of Fluid Mechanics*, 618:243–261, 2009. Times Cited: 15.
- [30] T.S. Horozov, R. Aveyard, J.H. Clint, and B.P. Binks. Order-disorder transition in monolayers of modified monodisperse silica particles at the octane-water interface. *Langmuir*, 19(7):2822–2829, 2003.

- [31] V.N. Paunov. Novel method for determining the three-phase contact angle of colloid particles adsorbed at air-water and oil-water interfaces. *Langmuir*, 19(19):7970–7976, 2003.
- [32] L. Isa, F. Lucas, R. Wepf, and E. Reimhult. Measuring single-nanoparticle wetting properties by freeze-fracture shadow-casting cryo-scanning electron microscopy. *Nature Communications*, 2:438, 2011.
- [33] O.J. Cayre and V.N. Paunov. Contact angles of colloid silica and gold particles at air-water and oil-water interfaces determined with the gel trapping technique. *Langmuir*, 20(22):9594–9599, 2004.
- [34] B. Madivala, J. Fransaer, and J. Vermant. Self-assembly and rheology of ellipsoidal particles at interfaces. *Langmuir*, 25(5):2718–2728, 2009.
- [35] E.P. Lewandowski, M. Cavallaro, L. Botto, J.C. Bernate, V. Garbin, and K.J. Stebe. Orientation and self-assembly of cylindrical particles by anisotropic capillary interactions. *Langmuir*, 26(19):15142–15154, 2010.
- [36] E. L. Sharp, H. Al-Shehri, T.S. Horozov, S.D. Stoyanov, and V.N. Paunov. Adsorption of shape-anisotropic and porous particles at the air-water and the decane-water interface studied by the gel trapping technique. *RSC Advances*, 4(5):2205–2213, 2014.
- [37] Y.H. Lee, W. Shi, H.K. Lee, R. Jiang, I.Y. Phang, Y. Cui, L. Isa, Y. Yang, J. Wang, S. Li, and X.Y. Ling. Nanoscale surface chemistry directs the tunable assembly of silver octahedra into three two-dimensional plasmonic superlattices. *Nature Communications*, 6:6990, 2015.
- [38] L.N. Arnaudov, O.J. Cayre, M.A. Cohen Stuart, S.D. Stoyanov, and V.N. Paunov. Measuring the three-phase contact angle of nanoparticles at fluid interfaces. *Physical Chemistry Chemical Physics*, 12(2):328–331, 2010.
- [39] K. Nakajima, T. Ikebara, and T. Nishi. Observation of gellan gum by scanning tunneling microscopy. *Carbohydrate Polymers*, 30(2-3):77–81, 1996.
- [40] V.N Paunov and O.J Cayre. Supraparticles and “Janus” particles fabricated by replication of particle monolayers at liquid surfaces using a gel trapping technique. *Advanced Materials*, 16(9-10):788–791, 2004.
- [41] A. Maestro, L.J. Bonales, H. Ritacco, R.G. Rubio, and F. Ortega. Effect of the spreading solvent on the three-phase contact angle of microparticles attached at fluid interfaces. *Physical Chemistry Chemical Physics*, 12(42):14115–14120, 2010.
- [42] M. A. Ray and L. Jia. Micropatterning by non-densely packed interfacial colloidal crystals. *Advanced Materials*, 19(15):2020–2022, 2007.
- [43] S. Coertjens, P. Moldenaers, J. Vermant, and L. Isa. Contact angles of microellipsoids at fluid interfaces. *Langmuir*, 30(15):4289–4300, 2014.
- [44] L. Isa, N. Samudrala, and E.R. Dufresne. Adsorption of sub-micron amphiphilic dumbbells to fluid interfaces. *Langmuir*, 30(18):5057–5063, 2014.
- [45] S. Jordens, L. Isa, I. Usov, and R. Mezzenga. Non-equilibrium nature of two-dimensional isotropic and nematic coexistence in amyloid fibrils at liquid interfaces. *Nature Communications*, 4:1917, 2013.
- [46] K. Geisel, L. Isa, and W. Richtering. Unraveling the 3d localization and deformation of responsive microgels at oil/water interfaces: A step forward in understanding soft emulsion stabilizers. *Langmuir*, 28(45):15770–15776, 2012.
- [47] B.P. Binks, L. Isa, and A.T. Tyowua. Direct measurement of contact angles of silica particles in relation to double inversion of pickering emulsions. *Langmuir*, 29(16):4923–4927, 2013.
- [48] J. R. Howse, R. A.L. Jones, A. J. Ryan, T. Gough, R. Vafabakhsh, and R. Golestanian. Self-motile colloidal particles: from directed propulsion to random walk. *Physical Review Letters*, 99(4):048102, 2007.

- [49] G. Boniello, C. Blanc, D. Fedorenko, M. Medfai, N.B. Mbarek, M. In, M. Gross, A. Stocco, and M. Nobili. Brownian diffusion of a partially wetted colloid. *Nature Materials*, 14(9):908–911, 2015.
- [50] X. Wang, M. In, C. Blanc, M. Nobili, and A. Stocco. Enhanced active motion of Janus colloids at the water surface. *Soft Matter*, 11:7376–7384, 2015.
- [51] X. Wang, M. In, C. Blanc, A. Würger, M. Nobili, and A. Stocco. Janus colloids actively rotating on the surface of water. *Langmuir*, 33(48):13766–13773, 2017.
- [52] F. Kümmel, B. ten Hagen, R. Wittkowski, I. Buttinoni, R. Eichhorn, G. Volpe, H. Löwen, and C. Bechinger. Circular motion of asymmetric self-propelling particles. *Physical Review Letters*, 110:198302, 2013.
- [53] V.-M. Ha and C.-L. Lai. The onset of stationary marangoni instability of an evaporating droplet. *Proceedings: Mathematical, Physical and Engineering Sciences*, 457(2008):885–909, 2001.
- [54] A. Würger. Thermally driven marangoni surfers. *Journal of Fluid Mechanics*, 752:589–601, 2014.
- [55] D. Okawa, S.J. Pastine, A. Zettl, and Fréchet. Surface tension mediated conversion of light to work. *Journal of the American Chemical Society*, 131(15):5396–5398, 2009.
- [56] C. Maggi, F. Saglimbeni, M. Dipalo, F. De Angelis, and R. Di Leonardo. Micromotors with asymmetric shape that efficiently convert light into work by thermocapillary effects. *Nature Communications*, 6:7855, 2015.
- [57] E. Lauga and A.M.J. Davis. Viscous Marangoni propulsion. *Journal of Fluid Mechanics*, 705:120–133, 2012.
- [58] H. Masoud and H.A. Stone. A reciprocal theorem for Marangoni propulsion. *Journal of Fluid Mechanics*, 741:1–7, 2014.
- [59] C.C. Maass, C. Krüger, S. Herminghaus, and C. Bahr. Swimming Droplets. *Annual Review of Condensed Matter Physics*, 7(1):171–193, 2016.
- [60] D. J. Adams, S. Adams, J. Melrose, and A. C. Weaver. Influence of particle surface roughness on the behaviour of Janus particles at interfaces. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 317(1):360–365, 2008.
- [61] X. Wang, M. In, C. Blanc, P. Malgaretti, M. Nobili, and A. Stocco. Wetting and orientation of catalytic Janus colloids at the surface of water. *Faraday Discussions*, 191:305–324, 2016.
- [62] T. Bickel, A. Majee, and A. Würger. Flow pattern in the vicinity of self-propelling hot Janus particles. *Physical Review E*, 88:012301, 2013.
- [63] B. Liu, R. Manica, Q. Liu, E. Klaseboer, Z. Xu, and G. Xie. Coalescence of Bubbles with Mobile Interfaces in Water. *Physical Review Letters*, 122(19):194501, 2019.
- [64] A. Maali, R. Boisgard, H. Chraibi, Z. Zhang, H. Kellay, and A. Würger. Viscoelastic Drag Forces and Crossover from No-Slip to Slip Boundary Conditions for Flow near Air-Water Interfaces. *Physical Review Letters*, 118(8):1–5, 2017.
- [65] O. Manor, I.U. Vakarelski, X. Tang, S.J. O’Shea, G.W. Stevens, F. Grieser, R.R. Dagastine, and D.Y.C. Chan. Hydrodynamic Boundary Conditions and Dynamic Forces between Bubbles and Surfaces. *Physical Review Letters*, 101(2):024501, 2008.
- [66] C.E. Brennen. *Cavitation and Bubble Dynamics*. Cambridge University Press, 2014.
- [67] Z. A. Zell, A. Nowbar, V. Mansard, L. G. Leal, S. S. Deshmukh, J. M. Mecca, C. J. Tucker, and T. M. Squires. Surface shear inviscidity of soluble surfactants. *Proceedings of the National Academy of Sciences*, 111(10):3677–3682, 2014.

Chapter 31

Optical Tweezers

DAVID BRONTE CIRIZA, MARIA G. DONATO, ONOFRIO M. MARAGÒ

The ability of light to exert a force on matter was recognized as early as 1619 by Johannes Kepler who first described the deflection of comet tails by the rays of the Sun [2]. However, only in 1970, and thanks to the advent of the laser [3], it has become possible to concentrate enough optical power in a small enough area to significantly affect the motion of microscopic particles [4, 5]. Thanks to this development, Arthur Ashkin was able to invent the *optical tweezers* [6], i.e., a tightly focused laser beam that can hold and manipulate a microscopic particle near its focal spot. The range of objects that have been optically manipulated is broad and goes from single atoms to mammalian cells (Fig. 31.1). By being able to apply and to measure nanometer scale forces, optical tweezers have heralded a revolution in the study of microscopic systems. Optical forces have already contributed significantly to fields like biology, physics, and environmental sciences [7, 8]. Their potential in advancing active matter lies in accurately measuring forces, propelling microscale particles, and controlling the motion of active systems, driving exciting progress in this area [9].

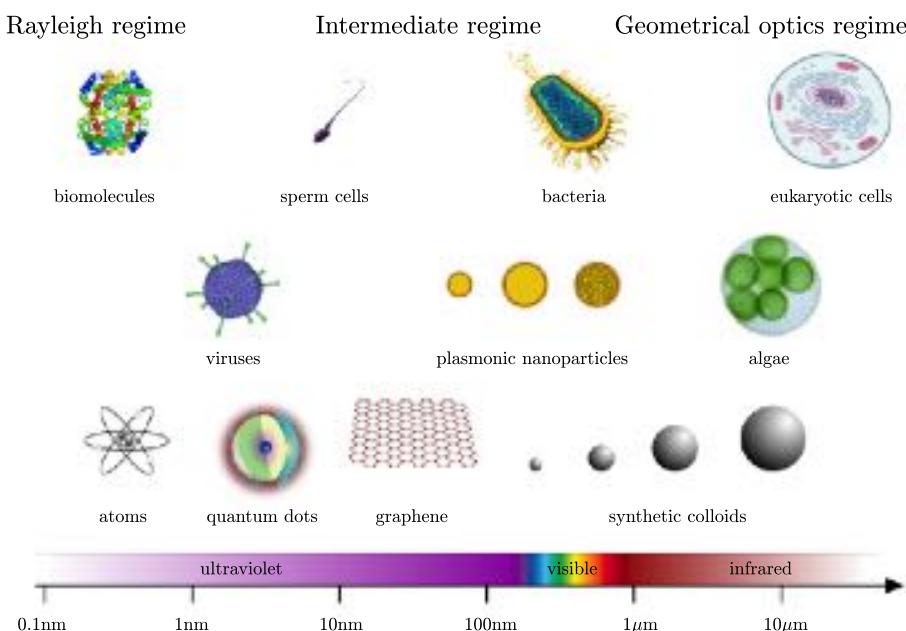


Figure 31.1: **Optically trappable objects.** Microscopic objects that are typically trapped in optical manipulation experiments. The trapping wavelengths is usually in the visible or near-infrared spectral region. Reproduced from Ref. [1].

31.1 Theory of optical trapping

The radiation force exerted by light on matter stems from the conservation of electromagnetic momentum during the scattering process. Historically, optical forces have been generally understood within strong approximations based on limiting size regimes. The parameter used to determine the range of validity of these approaches is the *size parameter*

$$\xi = k_m a = \frac{2\pi n_m}{\lambda_0} a, \quad (31.1)$$

where k_m is the light wavenumber in the medium surrounding the particle, a is the characteristic dimension of the particle (e.g., the radius of a microsphere), λ_0 is the trapping wavelength in vacuum, and n_m is the refractive index of the surrounding medium. For $k_m a \gg 1$, *geometrical optics* can be used to calculate optical forces [10, 11]. For $k_m a \ll 1$, the *Rayleigh approximation* can be employed [12]; since this approximates the particle as a dipole and assumes the electromagnetic fields homogeneous inside the particle, one further condition on the size parameter is required, i.e.,

$$\left| \frac{n_p}{n_m} \right| k_m a \ll 1, \quad (31.2)$$

where n_p is the refractive index of the particle. Finally, in the *intermediate size regime*, the calculation of the optical forces requires to consider the full electromagnetic theory of the light-matter interaction [7].

31.1.1 Geometrical optics

In the *geometrical optics approximation*, the optical field may be described as a collection of N rays, each of which is associated with a portion, P_j , of the incident power $P = \sum_j P_j$. Each ray carries a linear momentum per unit time corresponding to $n_m P_j / c$, where c is the speed of light.

To understand the forces acting on a trapped microscopic particle, we start with a minimalistic model: the force due to a single incident ray \mathbf{r}_i hitting a dielectric sphere at an angle of incidence θ_i (Figs. 31.2a and 31.2b). When \mathbf{r}_i strikes the sphere, a small part of the power is diverted into the reflected ray $\mathbf{r}_{r,0}$, while most of the power is carried by the transmitted ray $\mathbf{r}_{t,0}$. This transmitted ray $\mathbf{r}_{t,0}$ crosses the sphere until it reaches the next surface, where again it is largely transmitted outside the sphere into the ray $\mathbf{r}_{t,1}$, while a further small amount is reflected inside the sphere into the ray $\mathbf{r}_{r,1}$. This process continues until all light has escaped from the sphere. At each scattering event, the change in momentum of the ray causes a reaction force on the center of mass of the particle. By considering these multiple reflection and refraction events the optical force can be calculated

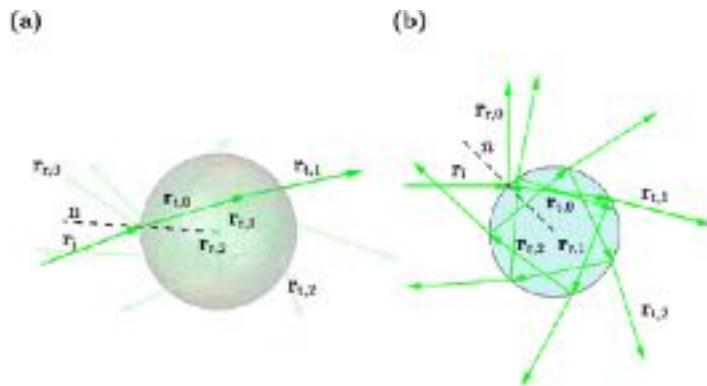


Figure 31.2: Scattering and optical forces associated with a ray impinging on a sphere. Multiple scattering of a light ray impinging on a sphere (a) in 3D and (b) in the incident plane. All the reflected and transmitted rays, as well as the optical force acting on the sphere lie in the plane of incidence. Reproduced from Ref. [1]. [Improve figure quality as pixelated]

as [11]:

$$\mathbf{F}_{\text{ray}} = \frac{n_m P_i}{c} \hat{\mathbf{r}}_i - \frac{n_m P_r}{c} \hat{\mathbf{r}}_{r,0} - \sum_{j=1}^{+\infty} \frac{n_m P_{t,j}}{c} \hat{\mathbf{r}}_{t,j}, \quad (31.3)$$

where $\hat{\mathbf{r}}_i$, $\hat{\mathbf{r}}_{r,j}$ and $\hat{\mathbf{r}}_{t,j}$ are unit vectors representing the direction of the incident ray and the j^{th} reflected and transmitted rays, respectively, calculated using Fresnel's reflection and transmission coefficients. Generally, most of the momentum transferred from the ray to the particle is only due to the first two scattering events, especially for small angles of incidence.

The force \mathbf{F}_{ray} in Eq. (31.3) has components only in the plane of incidence (Fig. 31.2b) and can be split into two perpendicular components. The component in the direction of the incoming ray $\hat{\mathbf{r}}_i$ represents the *scattering force*, $\mathbf{F}_{\text{ray},s}$, that pushes the particle in the direction of the incoming ray ($\hat{\mathbf{r}}_i$). The component perpendicular to the incoming ray is the *gradient force*, $\mathbf{F}_{\text{ray},g}$, that pulls the particle in a direction perpendicular to that of the incoming ray ($\hat{\mathbf{r}}_{\perp}$). Studying the dependence of the trapping efficiency of a ray with the angle of incidence, it is observed that independently of the size of the sphere, the maximum trapping efficiencies are obtained for large angles of incidence ($\approx 80^\circ$).

To model an optical trap, we must not only consider a single incident ray but all the rays constituting a highly-focused laser beam, i.e., a set of multiple rays that converge over a very large solid angle. This means that the total force acting on the particle is the sum of all the contributions from each ray forming the beam. For a single-beam optical trap, the focused rays will produce a restoring force proportional to the particle's displacement from an equilibrium position that lies close to the beam waist and generates a harmonic potential.

A geometrical-optics approach can also be employed to study more complex non-spherical geometries, as long as all the characteristic dimensions of the object under study are significantly larger than the light wavelength. However, going away from sphericity introduces two major differences with the case of a sphere. First, in the case of non-spherical objects a significant torque can also appear and induce the rotation of the object. This effect is known as the *windmill effect* because of its analogy to the motion of a windmill, where the wind in this case is the flow of momentum due to the electromagnetic field [13]. The torque due to a single ray can be calculated as the difference of the angular momentum associated with the incoming ray and that associated with the outgoing rays. The total torque on the object can then be calculated as the vector sum of the torques due to each ray. Second, while for a spherical particle the radiation pressure of a plane wave, i.e., a set of parallel rays, is always directed along the propagation direction because of symmetry, for particles of anisotropic shapes the radiation pressure has a transverse component that is responsible for an *optical lift effect*, i.e., non-spherical particles can move transversally with respect to the incident light propagation direction [14].

31.1.2 Rayleigh approximation

The *Rayleigh approximation* is based on the assumptions that a particle can be approximated as a small dipole and that fields are homogeneous inside the particle. An incident electromagnetic field \mathbf{E}_i induces an electric dipole moment \mathbf{p} that, for sufficiently small fields, can be expressed in terms of the particle polarizability as

$$\mathbf{p} = \alpha_p(\omega) \mathbf{E}_i, \quad (31.4)$$

where α_p is the complex polarizability of the particle relative to the surrounding medium given by [15]

$$\alpha_p = \frac{\alpha_{\text{CM}}}{1 - i\alpha_{\text{CM}}k_m^3/(6\pi\epsilon_m)} \quad (31.5)$$

with α_{CM} the static Clausius-Mossotti polarizability, i.e.,

$$\alpha_{\text{CM}} = 3V\epsilon_m \left(\frac{\epsilon_p - \epsilon_m}{\epsilon_p + 2\epsilon_m} \right), \quad (31.6)$$

ϵ_m and ϵ_p the permittivities of the medium and particle, respectively, and V the particle volume [You also need to define k_m]. The complex polarisability α_p , which typically depends on the frequency of the electromagnetic field ω , has a real part, α'_p (which represents the oscillation of the dipole in phase with the field) and an imaginary part α''_p (which represents its oscillation in phase quadrature). The polarizability in the dipole regime is also linked to the extinction (σ_{ext}), scattering (σ_{scat}), and absorption ($\sigma_{\text{abs}} = \sigma_{\text{ext}} - \sigma_{\text{scat}}$) cross sections, which

describe the light–particle interaction. For a small particle, we can write the extinction and scattering cross-sections as $\sigma_{\text{ext}} = k_m \text{Im}\{\alpha_p\}/\epsilon_m$ and $\sigma_{\text{scat}} = \frac{k_m^4}{6\pi\epsilon_m^2} |\alpha_p|^2$.

We can consider the time-averaged optical force experienced by a small particle when illuminated by a time-varying electromagnetic field [12, 16, 17]:

$$\mathbf{F}_{\text{DA}} = \frac{1}{2} \text{Re} \left\{ \sum_i \alpha_p E_i \nabla E_i^* \right\}, \quad (31.7)$$

where E_i are the electric field components. Starting from this expression, one can explicitly write the optical force in terms of extinction cross-section and particle's polarizability as

$$\mathbf{F}_{\text{DA}} = \frac{1}{4} \alpha'_p \nabla |E_i|^2 + \frac{n_m \sigma_{\text{ext}}}{c} \mathbf{S}_i - \frac{1}{2} n_m \sigma_{\text{ext}} c \nabla \times \mathbf{s}_d, \quad (31.8)$$

where \mathbf{S}_i is the time-averaged Poynting vector of the incident electromagnetic field, and \mathbf{s}_d is the time-averaged spin density [18, 17].

The first term in Eq. (31.8) represents the gradient force and is responsible for particle confinement in optical tweezers. Particles with a positive α'_p will be attracted towards the high-intensity region of the optical field, while particles with a negative α'_p will be repelled. As an example, for an incident laser beam with a typical Gaussian profile which propagates along the z axis, the trap stiffnesses in the polarization plane κ_p and along the direction of propagation κ_z related to the gradient force are [19]:

$$\kappa_p = 2 \frac{\alpha'_p}{cn_m} \frac{I_0}{w_0^2} \quad \text{and} \quad \kappa_z = 2 \frac{\alpha'_p}{cn_m} \frac{I_0}{z_0^2}, \quad (31.9)$$

where I_0 is the maximum intensity at the center of the beam, w_0 is the Gaussian beam waist, and z_0 is the beam Rayleigh range.

The second term in Eq. (31.8) corresponds to the scattering force, which is responsible for the radiation pressure and directed along the propagation direction of the laser.

The last term in Eq. (31.8) is a spin-dependant force [20]. While this term can be generated by polarization gradients in the electromagnetic field, it is usually negligible compared to the other terms for optical trapping. However, it plays a more significant role when trapping with beams of higher order with inhomogeneous polarization patterns, such us cylindrical vector beams [21] or superpositions of Hermite-Gauss beams [22].

31.1.3 Electromagnetic theory

The *intermediate regime* is characterized by a size of the particle that is comparable to the optical wavelength. In this regime, the dipole and geometrical optics approximations are no longer valid and a full modelling of the light–particle interaction is required. It is possible to derive the radiation force \mathbf{F}_{rad} and the radiation torque \mathbf{T}_{rad} using the conservation of linear and angular momentum. The time-averaged force and torque exerted by monochromatic light on a particle are [23, 24, 25]:

$$\mathbf{F}_{\text{rad}} = \int_S \bar{\mathbf{T}}_M \cdot \hat{\mathbf{n}} dS, \quad \text{and} \quad \mathbf{T}_{\text{rad}} = - \int_S (\bar{\mathbf{T}}_M \times \mathbf{r}) \cdot \hat{\mathbf{n}} dS, \quad (31.10)$$

where the integration is carried out over a surface S surrounding the particle, the vector $\hat{\mathbf{n}}$ is the outward unit vector normal to S , \mathbf{r} is the position vector, and $\bar{\mathbf{T}}_M$ is the time-averaged Maxwell stress tensor that describes the mechanical interaction of light and matter [26]. The general expression of $\bar{\mathbf{T}}_M$ in the Minkowski form [7] can be simplified when considering harmonic fields at angular frequency ω in a homogeneous, linear, and non-dispersive medium. In fact, writing the real physical fields, e.g., $\mathbf{E}(\mathbf{r}, t) = \text{Re}\{\mathbf{E}(\mathbf{r})e^{-i\omega t}\}$, in terms of the complex amplitudes, e.g., $\mathbf{E} = \mathbf{E}(\mathbf{r})$, the time-averaged Maxwell stress tensor simplifies as [7]:

$$\bar{\mathbf{T}}_M = \frac{1}{2} \epsilon_m \text{Re} \left[\mathbf{E}_t \otimes \mathbf{E}_t^* + \frac{c^2}{n_m^2} \mathbf{B}_t \otimes \mathbf{B}_t^* - \frac{1}{2} \left(|\mathbf{E}_t|^2 + \frac{c^2}{n_m^2} |\mathbf{B}_t|^2 \right) \mathbf{I} \right], \quad (31.11)$$

where \otimes represents the dyadic (outer) product, \mathbf{I} is the dyadic unit, and the fields $\mathbf{E}_t = \mathbf{E}_i + \mathbf{E}_s$ and $\mathbf{B}_t = \mathbf{B}_i + \mathbf{B}_s$ are the total electric and magnetic fields resulting from the superposition of the incident ($\mathbf{E}_i, \mathbf{B}_i$) and scattered ($\mathbf{E}_s, \mathbf{B}_s$) fields.

Several techniques have been proposed to compute electromagnetic scattering by non spherical particles, such as the discrete dipole approximation [27], the finite-difference time-domain [28], and the T-matrix [29]. In fact, there is no single universal method that provides the best results in all situations. Depending on the specific parameters, different techniques may prove to be the most appropriate in terms of efficiency, accuracy and applicability [30]. The *discrete dipole approximation* is a finite element method in which a particle is split into a series of dipoles. Each of these dipoles interacts with the incident wave and with the re-radiated waves by all the other dipoles. The *finite-difference time-domain* instead is based on the numerical integration of the Maxwell equations in the time domain. The fields are sampled at discrete times and positions and, therefore, do not assume a harmonic time-dependence. Differently, *T-matrix methods* provide a compact formalism based on the multipole expansion of the fields. The transformation properties under rotation and translation of the T-matrix make it possible to apply several simplifications in the force calculations permitting the T-matrix to be computed only once, making this approach the fastest one. On the other hand, even though the discrete dipole approximation and the finite-difference time-domain can be more computationally intensive, they can be slightly more versatile as they can be applied to particles of any shape and composition and to any light field configuration.

31.1.4 Enhanced simulations using deep learning

Numerically simulating the dynamics of a particle in an optical field involve many repetitions of the optical force calculation. While these simulations can be very useful to plan experiments and understand their results, all the methods discussed (except the dipole approximation, which is only valid when the conditions in Eq. (31.2) are satisfied) are computationally demanding, limiting the calculation speed and, therefore, the range of situations that can be explored.

Different approaches aim to increase the calculation speed. In some cases, the force generated by the optical tweezers can be approximated by a harmonic potential. Even though this works well for particles that are very stably trapped, there are many situations where this approximation is insufficient. Another approach to increase the calculation speed is based on interpolation. This involves calculating the optical forces at given points in the parameters space in advance, and then applying an interpolation algorithm to obtain the forces at intermediate locations. This interpolation approach works very well for problems with a few degrees of freedom, but since the memory required to store the data grows exponentially with the number of degrees of freedom, it becomes unfeasible to apply it for higher-dimensional problems. This is a limitation for problems where not only the position, but also the orientation or shape are involved.

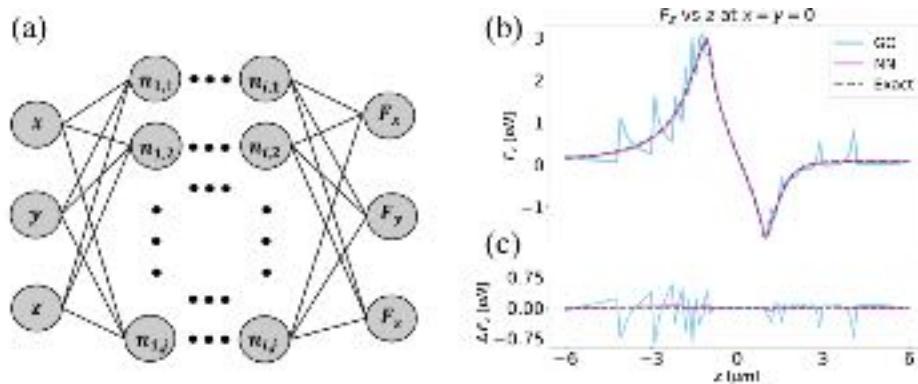


Figure 31.3: **Deep-learning-enhanced calculations.** (a) Architecture of a neural network to calculate the optical force given the particle position: the input layer has three neurons for the 3 coordinates (x, y, z) of the position. This network has i hidden layers and j neurons in each of these layers. The output layer contains 3 neurons, one for each of the components of the force (F_x, F_y, F_z). (b) Comparison between the z component of the optical force calculated using Geometrical Optics (GO) with 100 rays (solid cyan line), a neural network (NN) trained using the GO results (solid pink line), and the exact analytical solution (dashed black line). (c) The difference between the exact model (dashed black line) and the GO (solid cyan line) and NN (solid pink line) calculations shows that the NN is more accurate than GO. Reproduced from Ref. [31].

Recently, deep learning (i.e., neural networks) have been proposed to overcome these difficulties [32, 31]. As a simple example, we can consider a spherical particle (with fixed radius and refractive index) in an optical trap. As shown in Fig. 31.3a, it is possible to use a neural network that takes as inputs the three components of the position (x, y, z) and calculates as outputs the three components of the optical force (F_x, F_y, F_z). It is possible to train this neural network with examples of the inputs (positions) and the corresponding desired outputs (forces) calculated with one of the previously described methods. For example, by using a neural network trained with data generated using the T-Matrix formalism, an increase in the speed of around two orders of magnitude has been achieved [32]. Similar results have been obtained using training data generated with the geometrical optics approximation [31]. Furthermore, in the latter case, the neural network was also able to increase the accuracy in the determination of the optical forces by “smoothing out” the inaccuracies induced by the limited number of rays considered in the geometrical optics calculation used to generate the training data (Figs. 31.3b and 31.3c). The potential of this approach in terms of speed, accuracy, and simplicity holds promise to expand the applicability of optical forces simulations.

31.2 Experimental setups

In the simplest configuration, an optical tweezers is generated by focusing a laser beam to a diffraction-limited spot using a high-numerical-aperture objective lens. This objective serves the dual purpose of focusing the trapping light and imaging the trapped object. Generally, optical tweezers require little optical power (down to a few milliwatts) so that the risk of photodamage is relatively small. The position of the optical trap can be controlled using, e.g., steerable mirrors. It is also possible to generate multiple optical traps by rapidly deflecting a single beam through several positions by using, e.g., an acousto-optic deflector.

Even though a basic optical tweezers requires only a single strongly-focused laser beam to trap and manipulate a microscopic particle, more complex experimental setups have been developed to perform novel and challenging experiments. Some examples of different trapping schemes are depicted in Fig. 31.4 and some examples of optically manipulated particles are shown in Fig. 31.5. Detailed instructions on how to build advanced optical tweezers setups are provided in Ref. [33, 7] (other useful references are Ref. [34, 35, 36, 37, 38, 39]). In the following, we briefly review the main building blocks of an optical tweezers.

31.2.1 Microscope

The most convenient choice to built an optical tweezers setup is to use a conventional commercial light microscope. Commercial microscopes can be modified to host a dichroic mirror before the objective to deflect the trapping laser beam into the objective and at the same time to transmit the illumination light to a camera to image the sample. They are easy to use and optimized to reduce optical aberrations, but they may be relatively difficult to customize. Moreover, the optics are usually optimized for visible light and, therefore, some issues can arise if infrared lasers are employed. Finally, commercial microscopes do not offer the level of mechanical stability required for the most sensitive nanometer-scale experiments.

Despite the difficulties in the design and construction of a home-made microscope, this option is becoming increasingly popular. These microscopes can be built directly with standard optomechanical components. In these microscopes, the access to any part is very easy and with the proper choice of materials and design the mechanical stability is extremely high. It is also worth noting that home-made microscopes are often significantly cheaper than commercial ones.

31.2.2 Laser

The quality of the laser beam is critical to achieve the tightly-focused spot required for optical trapping, which must be as close to the diffraction limit as possible. A good pointing stability is required to maintain the position of the optical trap steady and low fluctuations are needed in order to keep the strength of the optical trap constant. It is worth mentioning that laser sources based on a monolithic non-planar ring oscillator Nd:YAG crystal exhibit optical properties unmatched by any other product, such as an output tunable over 30 GHz with the extraordinarily narrow line width of about 1 kHz and extremely low noise. These sources are perfect for experiments in which very weak forces (down to a few femtonewtons) are involved.

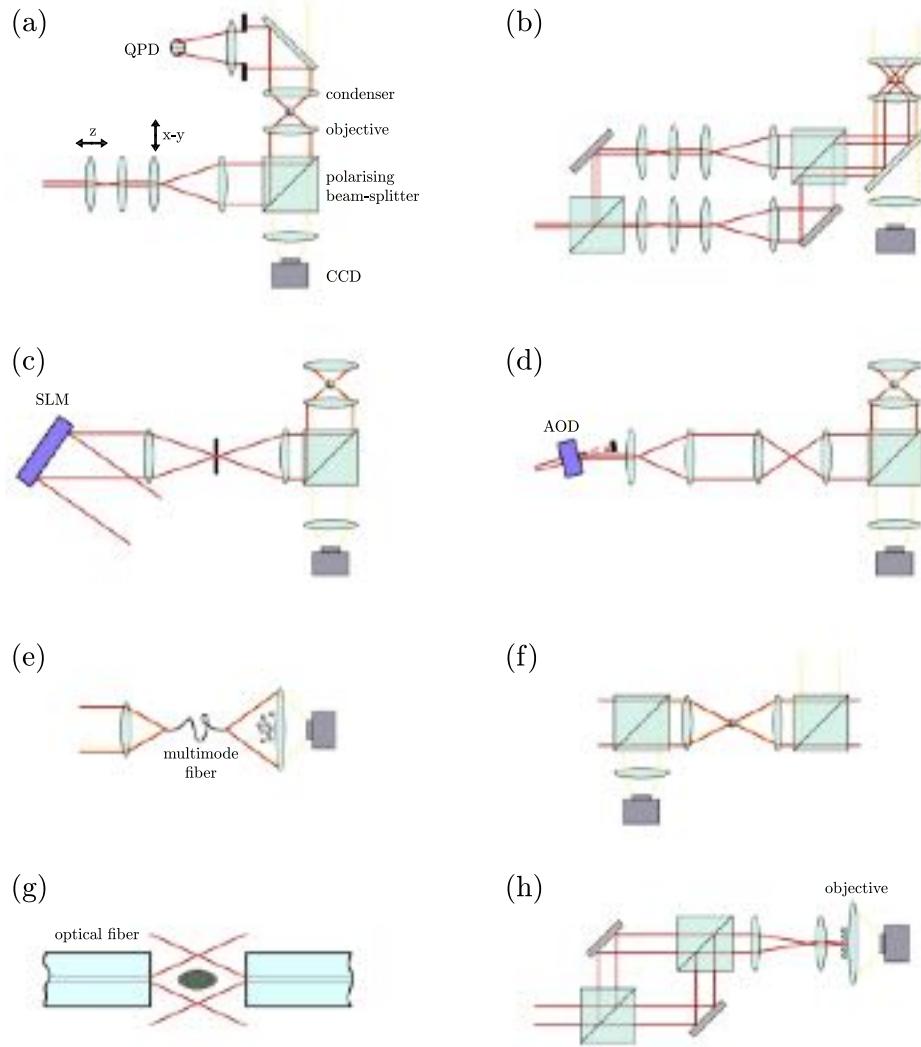


Figure 31.4: Typical optical tweezers setups. (a) A single-beam optical tweezers with quadrant photodiode. (b) A double-beam optical tweezers with mechanically steerable trapping beams. (c) A holographic optical tweezers incorporating a spatial light modulator (SLM). (d) A time-sharing optical tweezers using an acousto-optic deflector (AOD). (e) A speckle optical tweezers generating the speckle light field using a multimode fiber. (f) A counter-propagating-beam optical tweezers. (g) An optical stretcher. (h) An interferometric optical tweezers to generate large-scale optical potentials. Reproduced from Ref. [1]

31.2.3 Particle tracking

Most measurements performed with optical tweezers rely on knowing the particle position. There are two possibilities for measuring the particle position: the first is to image the trapped particle using a CCD or CMOS camera, while the second is to use detectors capable of measuring the spatial distribution of intensity in the interference pattern that occurs between the light scattered by the trapped particle and the unscattered laser light. Since a typical optical tweezers comes already equipped with a digital camera, the most straightforward means to measure the motion of a Brownian particle is to record a video of its position and then to track the position of the particle frame by frame. This technique, known as *digital video microscopy* [40, 41], has found widespread application in several fields and, in particular, in colloidal studies. It is especially well-suited to study systems where multiple particles are present, but it is relatively slow being limited by the camera frame rate, which typically goes up to a few thousands frames per second. An alternative to digital video microscopy

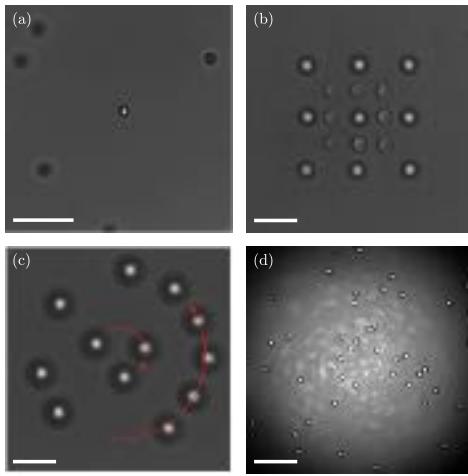


Figure 31.5: **Optical trapped particles.** (a) A particle trapped in a single-beam optical tweezers. (b) 18 particles held in three dimensions in a multi-trap holographic optical tweezers based on the use of a spatial light modulator. (c) Several particles set in rotation by the transfer of orbital angular momentum from a high order Laguerre-Gaussian beam generated by a spatial light modulator. (d) Particles optically trapped using a speckle light field. Reproduced from Ref. [33] [Agnese: make in a single line to spare space](#).

is the use of the interference pattern arising from the interference between the incoming and scattered fields [42]. The condenser collects such a pattern and a photodetector located on the condenser back-focal plane records the resulting signals. Thus, by tracking the change of the intensity distribution of the interference pattern, it is possible to measure the particle position.

31.3 Calibration

An optical tweezers creates a three-dimensional potential well that, close to the equilibrium position, can be approximated by a harmonic potential. Here, we will present three methods to calibrate an optical tweezers starting from experimental data, namely, the *power spectrum analysis*, the *autocorrelation function analysis*, and *FORMA* (Force Reconstruction via Maximum-Likelihood Estimator Analysis). For simplicity, we will consider only spherical particles and Newtonian fluids with known viscosity. However, the presented methods can be expanded to non-spherical particles [43, 44, 45] and media with unknown viscosity [46, 47].

Example 31.1: Simulation of an optically trapped particle trajectory. Simulate the trajectory of an optically trapped particle by finite differences (see Chapter ?? and Ref. [48]) starting from the Langevin equation (31.12) describing its motion. Compare these trajectories with the experimental ones provided in the book repository. [Agnese: Add David's trajectories to book repository, add here name of the files](#). The experimental data were acquired using a spherical polystyrene particle with a radius of $1\text{ }\mu\text{m}$ in water at a temperature of 293 K . At this temperature, the viscosity of water is approximately 0.001 Pas . The experimental trajectories are expressed in units of volts as they are acquired from a photodetector and, therefore, need to be converted into meters. *Hint: To convert the data from volts to meters, there are different strategies that we will explore in the next sections. One possibility is to match the experimentally estimated D to its nominal value obtained from the radius of the particle and the viscosity of the medium.*

31.3.1 Power spectrum analysis

The *power spectrum analysis* is the most widely used approach for the calibration of an optical trap, at least for spherical particles [49]. In the overdamped regime, the dynamics of a trapped spherical particle (along each

direction) follows the Langevin equation

$$\frac{dx(t)}{dt} = -\frac{k_x}{\gamma} + \sqrt{2D}W_x(t), \quad (31.12)$$

where $x(t)$ is the particle position as a function of time, k_x is the trap spring constant, γ is the friction coefficient of the particle, D is its diffusion coefficient, and W_x is a Wiener process.

Taking the Fourier transform of Eq. (31.12) and calculating the square modulus of both sides we obtain [7]:

$$P(f) = 2|\hat{X}(f)|^2 = \frac{D}{\pi^2(f_c^2 + f^2)}, \quad (31.13)$$

where $f_c = \frac{k_x}{2\pi\gamma}$ is the corner frequency. Eq. (31.13) is the *power spectral density* of the Brownian fluctuations of the optically trapped particle. Fitting it with a Lorentzian curve provides experimental estimations for f_c and D , from which the value k_x can be obtained.

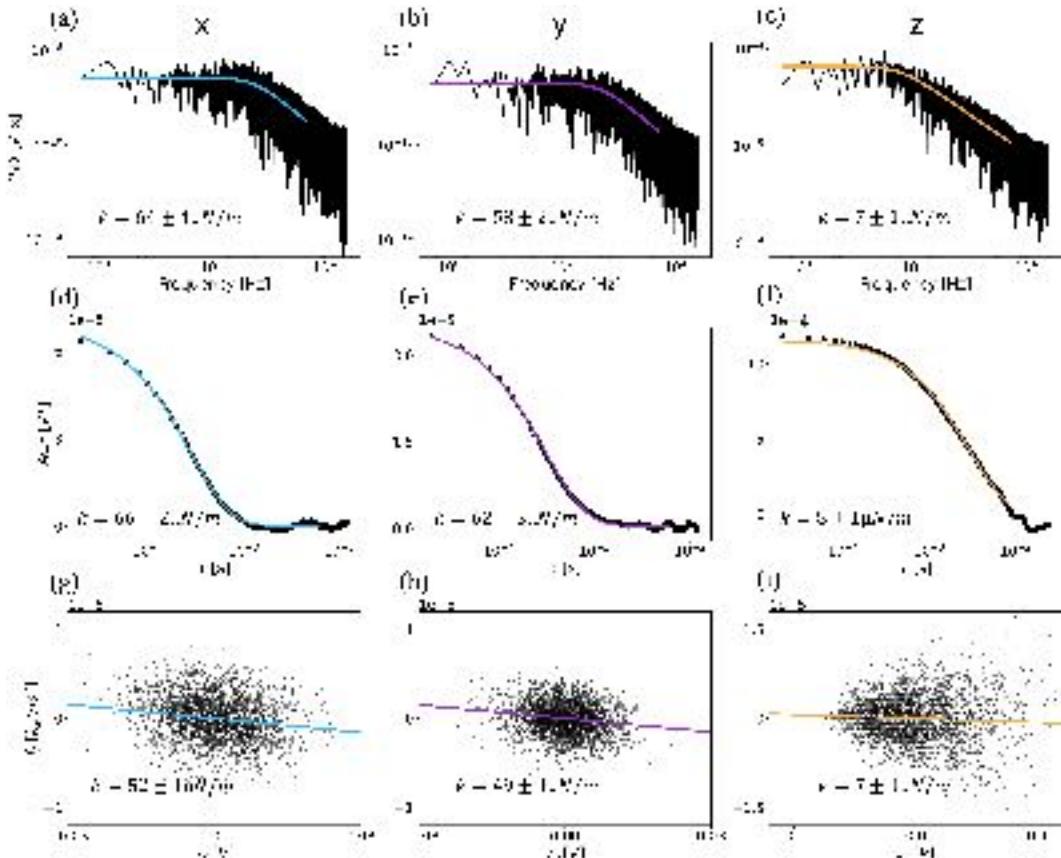


Figure 31.6: **Optical tweezers calibration.** (a–c) Power spectrum analysis, (d–f) autocorrelation function analysis, and (g–i) FORMA along the x , y , and z axes for an optically trapped particle. The plots represent the values obtained from a single experimental trajectory, but the values of the stiffness have been averaged over 10 measurements. The trajectories were recorded for a duration of 2 s with a sampling rate of 50 kHz. The experimental data were obtained using a spherical polystyrene particle with a radius of 1 μm and refractive index 1.58 immersed in water at a temperature of 293 K (water viscosity ≈ 0.001 Pas). The laser has wavelength 830 nm, is focused by a 1.3 NA objective, and has power 35 mW (after the objective). [Agnese: reproduce figure from David's experimental data.](#) Also, split it into 3 figures, one for each calibration method and update the text and caption accordingly.

Example 31.2: Calibration of optical tweezers using the power spectrum.

a. Generate a trajectory of an optically trapped particle as described in Exercise 31.1. Calculate its power spectrum and fit it with the Lorentzian model in Eq. (31.13). Verify that the fitted k and D are consistent with the values used in the simulation.

b. Now proceed to analyze the experimental trajectories provided in the book repository. [Agnese: add file names](#). Compare your results with those shown in Figs. 31.6a-c.

[Hint: Determine the conversion factor from volts to meters by matching the experimentally estimated D to its nominal value obtained from the radius of the particle and the viscosity of the medium.]

31.3.2 Autocorrelation function analysis

The *autocorrelation function analysis* is an alternative approach for the calibration of optical tweezers. Given a quantity A which is a function of the coordinates \mathbf{r}^N and momenta \mathbf{p}^N of N particles in a thermodynamic systems, its autocorrelation function is a measure of how much A at time $(t + \tau)$ depends on its value at time t , i.e.:

$$C_{AA}(\tau) = \langle A(t)A(t + \tau) \rangle. \quad (31.14)$$

In the case of optical trapping, we are interested in the autocorrelation function of the particle position, indicated as C_{xx} for the x direction:

$$C_{xx}(\tau) = \langle x(t)x(t + \tau) \rangle. \quad (31.15)$$

By using the overdamped Langevin equation, we find that

$$\langle x(t)x(t + \tau) \rangle = C_{xx}(0)e^{-\omega_x \tau}, \quad (31.16)$$

where $\omega_x = \frac{k_x}{\gamma}$ is the autocorrelation function relaxation frequency and $C_{xx}(0) = \langle x^2(t) \rangle$ can be obtained by the equipartition theorem ($\frac{1}{2}k_x\langle x^2(t) \rangle = \frac{1}{2}k_B T$, where $k_B T$ is the thermal energy with k_B being the Boltzmann constant and T the absolute temperature) [7]. This leads to the final expression for the autocorrelation function as

$$C_{xx}(\tau) = \frac{k_B T}{k_x} e^{-\omega_x \tau} \quad (31.17)$$

By fitting the autocorrelation function with an inverted exponential, we estimate the trap spring constant k_x .

The power spectral density and the autocorrelation function approaches are equivalent, as the power spectral density of a signal is the Fourier transform of its autocorrelation function. Thus, the results obtained using the two approaches must be consistent. The choice of the calibration approach is made on the basis of the experimental situation. The power spectral density approach is particularly useful when low frequency periodic noises (e.g., those due to electronic equipment or mechanical vibrations) must be removed, as they are easily recognized as peaks in the power spectral density. The autocorrelation function approach has the advantage of being easily adaptable to the study of the complex dynamics of non-spherical particles [43, 45, 50].

Example 31.3: Calibration of optical tweezers using the autocorrelation function.

a. Generate a trajectory of an optically trapped particle as described in Exercise 31.1. Calculate its autocorrelation function and fit it with the inverted exponential model in Eq. (31.17). Verify that the fitted k and D are consistent with the values used in the simulation.

b. Now proceed to analyze the experimental trajectories provided in the book repository. [Agnese: add file names](#). Compare your results with those shown in Figs. 31.6d-f.

[Hint: Determine the conversion factor from volts to meters by matching the value of the correlation at $t = 0$ to its expected value for the estimated stiffness.]

31.3.3 FORMA

Power spectrum and autocorrelation function analysis have traditionally been the most used methods to calibrate optical tweezers. However, they present some limitations as they require sampling at regular time intervals

and the stiffness obtained depends on the choice of the parameters that are used for fitting the experimental data. FORMA [51] overcomes these constraints by exploiting the fact that in the proximity of an equilibrium position the force field can be linearized and then estimated using a maximum-likelihood estimator. Starting from the overdamped Langevin equation (31.12), it is possible to discretize it and express the viscous friction force at a given n -th time interval as

$$f_n = \gamma \frac{\Delta x_n}{\Delta t_n} = -kx_n + \sigma w_n, \quad (31.18)$$

where $\sigma = \sqrt{\frac{2D\gamma^2}{\Delta t_n}}$ and w_n is a random number normally distributed with zero mean and unit variance. Given that $\frac{\Delta x_n}{\Delta t_n}$ and x_n can be obtained from measurements as experimental data sets, and that Eq. (31.18) is written as a linear regression model, it is possible to calculate the parameters k and σ by means of maximum-likelihood estimators. In the case of the trap stiffness k , the estimator can be expressed as

$$k^* = \frac{\sum_n x_n f_n}{\sum_n x_n^2}, \quad (31.19)$$

which can be solved very fast as its computational solution benefits from the existence of highly optimized linear algebra libraries. To provide a quantitative consistency check of the quality of the estimation, it is possible to compare the theoretical diffusion coefficient with the estimated one that can be computed from Eq. 31.18:

$$D^* = \frac{1}{N} \sum_n \frac{\Delta t_n}{2\gamma^2} (f_n + k^* x_n)^2. \quad (31.20)$$

Example 31.4: Calibration of optical tweezers using FORMA.

- a. Generate a trajectory of an optically trapped particle as described in Exercise 31.1. Fit k and D using FORMA and verify that the fitted values are consistent with those used in the simulation.
- b. Now proceed to analyze the experimental trajectories provided in the book repository. [Agnese: add file names](#). Compare your results with those shown in Figs. 31.6g-i.

[Hint: Determine the conversion factor from volts to meters by matching the experimentally estimated D to its nominal value obtained from the radius of the particle and the viscosity of the medium.]

- c. Compare the three calibration methods (power spectrum analysis, autocorrelation function analysis and FORMA). What are the strengths and weaknesses of each of them?

31.4 Applications

In this section we give an overview of some recent applications of optical tweezers and optical forces. We describe a selection of systems where optical trapping has enabled advances in soft and active matter.

31.4.1 Mechanical properties of red blood cells

Optical tweezers have numerous applications in biological sciences. One particularly successful application is the investigation and determination of the mechanical and elastic properties of red blood cells. Altered mechanical properties of these cells have been linked with several pathological conditions [52, 53]. Very early in the history of optical tweezers Ashkin and coworkers demonstrated that red blood cells (and many other biological species) could be trapped without optical damage using an infrared laser beam [54]. Subsequent experiments aimed at determining their mechanical properties have both directly trapped red blood cells and indirectly applied forces using optically trapped microbeads bound to the cell as “handles”. Nowadays, optical tweezers have been used to test several diseases that affect red blood cells including malaria (*plasmodium falciparum*) [55, 56], diabetic retinopathy [57], and birdshot chorioretinopathy [58]. Similarly, optical tweezers have been used to show the effects of drugs on the mechanical properties of cells membranes [59].

31.4.2 Micromachines

During the last few decades much effort has gone into the miniaturization of machines down to microscopic scale, often inspired by biological systems [60]. In this context, optical tweezers are a powerful tool to assemble micro- and nanodevices thanks to their ability of contactless manipulation. Moreover, they are capable of applying and detecting extremely small forces (down to a few femtonewtons) and torques (down to a few femtonewton nanometers) [61, 62, 63, 64, 65].

Nanodevices or micro-engines need power to operate and to be controlled. A solution to this demand can be provided by structured optical beams, carrying orbital and spin angular momentum, generated by holographic optical tweezers or similar techniques [66]. Another approach to power nanodevices is to emulate the working principles of heat engines. The nucleation of vapor bubbles inside silicon micro-cavities has been used to realize several microscopic heat engines with a working volume of only 0.6 mm^3 [67, 68] [Agnese: check this value in the references](#). Also, a microparticle has been employed as a piston in an optical tweezers, realizing a microscopic steam engine powered by the periodic generation of cavitation bubbles [69]. Microscopic versions of the Stirling and Carnot cycles have also been realized using optically trapped particles to study their stochastic thermodynamic properties [70, 71].

For light-absorbing particles, not only momentum transfer but also energy absorption and consequent heating play a key role in their dynamics, giving rise to more complex behaviors. Because of the combination of optical and thermal effects, microengines can feature elevator-like motion [72], elliptical [73], trochoidal [74], and circular orbits [75, 76, 65], higher rotation velocities [77], and reconfigurable assembly of multiple particles [78]. This shows that the integration of optical and thermal effects can induce a diverse range of dynamic behaviors. However, the controllability over these dynamic behaviours is still limited and represents a challenge in the field.

31.4.3 Stochastic thermodynamics

Optical tweezers can also be employed to unveil and characterize the statistical properties of micro- and nanoscopic systems. Brownian noise and large thermal fluctuations play a crucial role by introducing stochasticity [79, 80]. In particular, the dynamics of optically trapped particles result from the interplay between deterministic optical forces and Brownian motion, which introduces a well-defined noisy background. Therefore, optically trapped particles can be employed as probes to investigate statistical physics phenomena (from biomolecules and nanodevices to financial markets and human organizations), whose dynamics are driven by both random and deterministic forces.

Stochastic thermodynamics can be successfully applied to different systems, such as living matter [81, 82]. For example, biomolecules are often coupled with active baths, due to molecular motors inside the cytoplasm. This coupling is thought to lead to anomalous diffusion within the cytoplasm, a phenomenon that is largely not yet understood [81]. Recently, the behavior of an optically trapped particle, immersed in an active bath, has been investigated by digital video microscopy to show a transition from Boltzmann to non-Boltzmann statistic [83]. This transition takes place whenever the characteristic scale of an optical trap becomes comparable to the characteristic correlation length of the active noise. A consequence of this transition is that non-equilibrium relations such as the Jarzynski equality [84] and Crooks fluctuation theorem [85] cannot be applied in active baths according to their classical formulation [83]. Although this behavior is unexpected in active systems, its investigation is crucial to develop better models for living and far-from-equilibrium systems.

31.5 Problems

[Agnese: Revise the problems to ensure that they are clearly written and all necessary materials are available on the book web.](#)

Problem 31.1: Active particle in an optical potential. Consider a spherical particle (diameter of $2\text{ }\mu\text{m}$, viscosity 0.001 Pa s) in an optical potential.

- a. Start by approximating the optical trap as a harmonic potential (stiffness $5\text{ }\mu\text{N m}^{-1}$). Simulate the dynamics of the spherical particle in this harmonic trap.

b. Now add some activity to the particle — you can assume that the particle is moving only in the xy -plane, while z remains constant. Hold the orientation of the particle (and, therefore its swimming velocity) fixed. What do you observe? Can the particle escape the optical trap?

c. Use the neural network that you can find at <https://github.com/brontecir/ExerciseOpticalTweezers> [Agnewe: make this available in the github of the book. Update the reference accordingly.](#) to compute a more realistic estimation of the optical forces. What differences do you observe? Can the particle escape the optical trap in this situation?

d. Study now the situation where the particle orientation is always pointing away from the trap. Imagine that you are able to transfer angular momentum from the light to the particle producing a slight tilt in the particle. What do you obtain? [What answer are you exactly after here? Clarify the question]

Problem 31.2: Active particles in a complex optical potential. Consider the scenario where an active particle is capable of escaping an optical potential. You have two options to compute the optical forces: either utilize the neural network available at <https://github.com/brontecir/ExerciseOpticalTweezers> or assume a harmonic potential with a maximum range of $1\text{ }\mu\text{m}$, beyond which particles remain unaffected.

a. Investigate the motion of an active particle traversing a landscape containing different optical potentials randomly distributed.

b. Now, consider a scenario where the swimming speed of the active particle depends on the intensity of the light. Study also the two extreme cases (optical forces dominate or swimming speed dominates).

Problem 31.3: Active particles in a complex optical potential that affects the orientation. In this problem, we explore the behavior of active particles in a complex optical potential that influences their orientation. Similar to Problem 31.2, we consider particles capable of escaping the optical potential, and we examine different optical potentials as before. However, in this case, the optical potentials apply a torque towards a specific direction, which is randomly determined but remains constant for a given optical potential.

a. Investigate the motion of an active particle in this scenario where optical potentials exert a torque, aligning the particle towards a specific direction. Analyze the relationship between the particle's size and the distance between the optical potentials.

b. Now, consider the scenario where multiple particles coexist in the same landscape, subject to the orientation-inducing optical potentials. Observe and analyze the changes that occur when several particles interact within this complex optical potential. Investigate collective behaviors, such as alignment, clustering, or other emergent phenomena, and discuss the impact of particle interactions on the system's dynamics.

Bibliography

- [1] G. Pesce, P. H. Jones, O. M. Maragò, and G. Volpe. Optical tweezers: theory and practice. *The European Physical Journal Plus*, 135:1–38, 2020.
- [2] J. Kepler. *De Cometi libelli tres*. Augustae Vindelicorum, Augsburg, 1619.
- [3] C. H. Townes. *How the laser happened: Adventures of a scientist*. Oxford University Press, 1999.
- [4] A. Ashkin. Acceleration and trapping of particles by radiation pressure. *Physical Review Letters*, 24:156–159, 1970.
- [5] A. Ashkin. Atomic-beam deflection by resonance-radiation pressure. *Physical Review Letters*, 25:1321–1324, 1970.
- [6] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu. Observation of a single-beam gradient optical trap for dielectric particles. *Optics Letters*, 11:288–290, 1986.
- [7] P. H. Jones, O. M. Maragò, and G. Volpe. *Optical Tweezers: Principles and Applications*. Cambridge University Press, 2015.
- [8] G. Volpe, O. M. Maragò, H. Rubinsztein-Dunlop, G. Pesce, A. B. Stilgoe, G. Volpe, G. Tkachenko, V. G. Truong, S. N. Chormaic, F. Kalantarifard, et al. Roadmap for optical tweezers. *Journal of Physics: Photonics*, 5:022501, 2023.

- [9] M. Rey, G. Volpe, and G. Volpe. Light, matter, action: Shining light on active matter. *ACS Photonics*, 10:1188–1201, 2023.
- [10] A. Ashkin. Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime. *Biophysical Journal*, 61:569–582, 1992.
- [11] A. Callegari, M. Mijalkov, A. B. Gököz, and G. Volpe. Computational toolbox for optical tweezers in geometrical optics. *Journal of the Optical Society of America B*, 32:B11–B19, 2015.
- [12] P. C. Chaumet and M. Nieto-Vesperinas. Time-averaged total force on a dipolar sphere in an electromagnetic field. *Optics letters*, 25:1065–1067, 2000.
- [13] E. Higurashi, H. Ukita, H. Tanaka, and O. Ohguchi. Optically induced rotation of anisotropic micro-objects fabricated by surface micromachining. *Applied Physics Letters*, 64:2209–2210, 1994.
- [14] G. A. Swartzlander Jr, T. J. Peterson, A. B. Artusio-Glimpse, and A. D. Raisanen. Stable optical lift. *Nature Photonics*, 5:48–51, 2010.
- [15] B. T. Draine and J. Goodman. Beyond Clausius-Mosotti: Wave propagation on a polarizable point lattice and the discrete dipole approximation. *Astrophys. J.*, 405:685–697, 1993.
- [16] J. R. Arias-González and M. Nieto-Vesperinas. Optical forces on small particles: Attractive and repulsive nature and plasmon-resonance conditions. *Journal of the Optical Society of America A*, 20:1201–1209, 2003.
- [17] D. Gao, W. Ding, M. Nieto-Vesperinas, X. Ding, M. Rahman, T. Zhang, C.-T. Lim, and C.-W. Qiu. Optical manipulation from the microscale to the nanoscale: Fundamentals, advances and prospects. *Light: Science & Applications*, 6:e17039, 2017.
- [18] S. Albaladejo, M. I. Marqués, F. Scheffold, and J. J. Sáenz. Giant enhanced diffusion of gold nanoparticles in optical vortex fields. *Nano Letters*, 9:3527–3531, 2009.
- [19] P. Zemánek, A. Jonáš, P. Jákl, M. Šerý, J. Ježek, and M. Liška. Theoretical comparison of optical traps created by standing wave and single beam. *Optics Communications*, 220:401–412, 2003.
- [20] S. Albaladejo, M. I. Marqués, M. Laroche, and J. J. Sáenz. Scattering forces from the curl of the spin angular momentum of a light field. *Physical Review Letters*, 102:113602, 2009.
- [21] M. G. Donato, S. Vasi, R. Sayed, P. H. Jones, F. Bonaccorso, A. C. Ferrari, P. G. Gucciardi, and O. M. Maragò. Optical trapping of nanotubes with cylindrical vector beams. *Optics Letters*, 37:3381–3383, 2012.
- [22] M. I. Marqués. Beam configuration proposal to verify that scattering forces come from the orbital part of the poynting vector. *Optics Letters*, 39:5122–5125, 2014.
- [23] M. I. Mishchenko. Radiation force caused by scattering, absorption, and emission of light by nonspherical particles. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 70:811–816, 2001.
- [24] R. Saija, M. A. Iatì, A. Giusto, P. Denti, and F. Borghese. Transverse components of the radiation force on nonspherical particles in the T-matrix formalism. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 94:163 – 179, 2005.
- [25] F. Borghese, P. Denti, R. Saija, and M. A. Iatì. Radiation torque on nonspherical particles in the transition matrix formalism. *Optics Express*, 14:9508–9521, 2006.
- [26] R. N. C. Pfeifer, T. A. Nieminan, N. R. Heckenberg, and H. Rubinsztein-Dunlop. Momentum of an electromagnetic wave in dielectric media. *Reviews of Modern Physics*, 79:1197–1216, 2007.
- [27] M. A. Yurkin and A. G. Hoekstra. The discrete dipole approximation: An overview and recent developments. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 106:558–589, 2007.

- [28] K. Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14:302–307, 1966.
- [29] P. C. Waterman. Symmetry, unitarity and geometry in electromagnetic scattering. *Physical Review D*, 3:825–839, 1971.
- [30] M. I. Mishchenko, L. D. Travis, and A. A. Lacis. *Scattering, absorption, and emission of light by small particles*. Cambridge University Press, Cambridge, United Kingdom, 2002.
- [31] D. Bronte Ciriza, A. Magazzù, A. Callegari, G. Barbosa, A. A. R. Neves, M. A. Iatì, G. Volpe, and O. M. Maragò. Faster and more accurate geometrical-optics optical force calculation using neural networks. *ACS Photonics*, 10:234–241, 2023.
- [32] I. C. D. Lenton, G. Volpe, A. B. Stilgoe, T. A. Nieminen, and H. Rubinsztein-Dunlop. Machine learning reveals complex behaviours in optically trapped particles. *Machine Learning: Science and Technology*, 1:045009, 2020.
- [33] G. Pesce, G. Volpe, O. M. Maragò, P. H. Jones, S. Gigan, A. Sasso, and G. Volpe. Step-by-step guide to the realization of advanced optical tweezers. *Journal of the Optical Society of America B*, 32:B84–B98, 2015.
- [34] S. P. Smith, S. R. Bhalotra, A. L. Brody, B. L. Brown, E. K. Boyda, and M. Prentiss. Inexpensive optical tweezers for undergraduate laboratories. *American Journal of Physics*, 67:26–35, 1999.
- [35] J. Bechhoefer and S. Wilson. Faster, cheaper, safer optical tweezers for the undergraduate laboratory. *American Journal of Physics*, 70:393–400, 2002.
- [36] A. S. Mellish and A. C. Wilson. A simple laser cooling and trapping apparatus for undergraduate laboratories. *American Journal of Physics*, 70:965–971, 2002.
- [37] D. C. Appleyard, K. Y. Vandermeulen, H. Lee, and M. J. Lang. Optical trapping for undergraduates. *American Journal of Physics*, 75:5–14, 2007.
- [38] W. M. Lee, P. J. Reece, R. F. Marchington, N. K. Metzger, and K. Dholakia. Construction and calibration of an optical trap on a fluorescence optical microscope. *Nature Protocols*, 2:3225–3238, 2007.
- [39] M. Mathew, S. I. C. O. Santos, D. Zalvidea, and P. Loza-Alvarez. Multimodal optical workstation for simultaneous linear, nonlinear microscopy and nanomanipulation: Upgrading a commercial confocal inverted microscope. *Review of Scientific Instruments*, 80:073701, 2009.
- [40] J. C. Crocker and D. G. Grier. Methods of digital video microscopy for colloidal studies. *Journal of Colloid and Interface Science*, 179:298–310, 1996.
- [41] I. F. Sbalzarini and P. Koumoutsakos. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology*, 151:182–195, 2005.
- [42] F. Gittes and C. H. F. Schmidt. Interference model for back-focal plane displacement detection in optical tweezers. *Optics Letters*, 23:7–9, 1998.
- [43] O. M. Maragò, P. H. Jones, F. Bonaccorso, V. Scardaci, P. G. Gucciardi, A. G. Rozhin, and A. C. Ferrari. Femtonewton force sensing with optically trapped nanotubes. *Nano Letters*, 8:3211–3216, 2008.
- [44] M. G. Donato, M. A. Monaca, G. Faggio, L. De Stefano, P. H. Jones, P. G. Gucciardi, and O. M. Maragò. Optical trapping of porous silicon nanoparticles. *Nanotechnology*, 22:505704, 2011.
- [45] O. M. Maragò, F. Bonaccorso, R. Saija, G. Privitera, P. G. Gucciardi, M. A. Iatì, G. Calogero, P. H. Jones, F. Borghese, P. Denti, V. Nicolosi, and A. C. Ferrari. Brownian motion of graphene. *ACS Nano*, 4:7515–7523, 2010.

- [46] S. F. Tolić-Nørrelykke, E. Schäffer, J. Howard, F. S. Pavone, F. Jülicher, and H. Flyvbjerg. Calibration of optical tweezers with positional detection in the back focal plane. *Review of Scientific Instruments*, 77:103101, 2006.
- [47] A. Buosciolo, G. Pesce, and A. Sasso. New calibration method for position detector for simultaneous measurements of force constants and local viscosity in optical tweezers. *Optics Communications*, 230:357–368, 2004.
- [48] G. Volpe and G. Volpe. Simulation of a brownian particle in an optical trap. *American Journal of Physics*, 81:224–230, 2013.
- [49] K. Berg-Sorensen and H. Flyvbjerg. Power spectrum analysis for optical tweezers. *Review of Scientific Instruments*, 75:594, 2004.
- [50] M. G. Donato, E. Messina, A. Foti, T. J. Smart, P. H. Jones, M. A. Iatì, R. Saija, P. G. Gucciardi, and O. M. Maragò. Optical trapping and optical force positioning of two-dimensional materials. *Nanoscale*, 10:1245–1255, 2018.
- [51] L. Pérez García, J. Donlucas Pérez, G. Volpe, A. V. Arzola, and G. Volpe. High-performance reconstruction of microscopic force fields from brownian trajectories. *Nature Communications*, 9:1–9, 2018.
- [52] S. Suresh. Mechanical response of human red blood cells in health and disease: Some structure-property-function relationships. *Journal of Materials Research*, 21:1871–1877, 2006.
- [53] R. Agrawal, J. Sherwood, J. Chhablani, A. Ricchariya, S. Kim, P. H. Jones, S. Balabani, and D. Shima. Red blood cells in retinal vascular disorders. *Blood Cells, Molecules, and Diseases*, 56:53 – 61, 2016.
- [54] A. Ashkin and J. M. Dziedzic. Optical trapping and manipulation of viruses and bacteria. *Science*, 235:1517–1520, 1987.
- [55] S. Suresh, J. Spatz, J.P. Mills, A. Micoulet, M. Dao, C.T. Lim, M. Beil, and T. Seufferlein. Single-cell biomechanics and human disease states: Gastrointestinal cancer and malaria. *Acta Biomaterialia*, 1:16, 2005.
- [56] J. M. A. Mauritz, T. Tiffert, R. Seear, F. Lautenschläger, A. Esposito, V. L Lew, J. R. Guck, and C. F. Kaminski. Detection of plasmodium falciparum-infected red blood cells by optical stretching. *Journal of Biomedical Optics*, 15:030517, 2010.
- [57] R. Agrawal, T. J .Smart, J. Nobre-Cardoso, C. J. Richards, R. Bhatnagar, A. Tufail, D. Shima, P. H. Jones, and C. Pavesio. Assessment of red blood cell deformability in type 2 diabetes mellitus and diabetic retinopathy by dual optical tweezers stretching technique. *Scientific Reports*, 6:15873, 2016.
- [58] R. Agrawal, B. Ang, P. Kumar Balne, C. J. Richards, T. J. Smart, J. Cardoso, D. Shima, P. H. Jones, and C. Pavesio. Non-occlusive retinal vascular inflammation and role of red blood cell deformability in birdshot chorioretinopathy. *Ocular Immunology and Inflammation*, 27:978–986, 2019.
- [59] V. Sheikh-Hasani, M. Babaei, A. Azadbakht, H. Pazoki-Toroudi, A. Mashaghi, A. A. Moosavi-Movahedi, and S. N. S. Reihani. Atorvastatin treatment softens human red blood cells: an optical tweezers study. *Biomedical Optics Express*, 9:1256–1261, 2018.
- [60] L. K. E. A. Abdelmohsen, F. Peng, Y. Tu, and D. A. Wilson. Micro-and nano-motors for biomedical applications. *Journal of Materials Chemistry B*, 2:2395–2408, 2014.
- [61] P. H. Jones, F. Palmisano, F. Bonaccorso, P. G. Gucciardi, G. Calogero, A. C. Ferrari, and O. M. Maragò. Rotation detection in light-driven nanorotors. *ACS Nano*, 3:3077–3084, 2009.
- [62] A. A. R. Neves, A. Camposeo, S. Pagliara, R. Saija, F. Borghese, P. Denti, M. A. Iatì, R. Cingolani, O. M. Maragò, and D. Pisignano. Rotational dynamics of optically trapped nanofibers. *Optics Express*, 18:822–830, 2010.

- [63] A. Lehmuskero, R. Ogier, T. Gschneidtner, P. Johansson, and M. Kall. Ultrafast spinning of gold nanoparticles in water using circularly polarized light. *Nano Letters*, 13:3129–3134, 2013.
- [64] A. Lehmuskero, P. Johansson, H. Rubinsztein-Dunlop, L. Tong, and M. Kall. Laser trapping of colloidal metal nanoparticles. *ACS Nano*, 9:3453–3469, 2015.
- [65] F. Schmidt, A. Magazzù, A. Callegari, L. Biancofiore, F. Cichos, and G. Volpe. Microscopic engine powered by critical demixing. *Physical Review Letters*, 120:068004, 2018.
- [66] M. Padgett and R. Di Leonardo. Holographic optical tweezers and their relevance to lab on chip devices. *Lab on a Chip*, 11:1196–1205, 2011.
- [67] J. Kao, X. Wang, J. Warren, J. Xu, and D. Attinger. A bubble-powered micro-rotor: conception, manufacturing, assembly and characterization. *Journal of Micromechanics and Microengineering*, 17:2454, 2007.
- [68] C. Lee, M. Liamini, and L. G. Frechette. A silicon microturbopump for a rankine-cycle power-generation microsystem - Part II: Fabrication and characterization. *Journal of Microelectromechanical Systems*, 20:326–338, 2011.
- [69] P. A. Quinto-Su. A microscopic steam engine implemented in an optical tweezer. *Nature Communications*, 5:5889, 2014.
- [70] V. Bickle and C. Bechinger. Realization of a micrometre-sized stochastic heat engine. *Nature Physics*, 8:143, 2012.
- [71] I. A. Martínez, É. Roldán, L. Dinis, D. Petrov, J. M. R. Parrondo, and R. A. Rica. Brownian Carnot engine. *Nature Physics*, 12:67, 2016.
- [72] S. Nedev, S. Carretero-Palacios, P. Kühler, T. Lohmüller, A. S. Urban, L. J. E. Anderson, and J. Feldmann. An optically controlled microscale elevator using plasmonic Janus particles. *ACS Photonics*, 2:491–496, 2015.
- [73] J. Lui, H.-L. Guo, and Z.-Y. Li. Self-propelled round-trip motion of Janus particles in static line optical tweezers. *Nanoscale*, 8:19894–19900, 2016.
- [74] H. Moyses, J. Palacci, S. Sacanna, and D. G. Grier. Trochoidal trajectories of self-propelled Janus particles in a diverging laser beam. *Soft Matter*, 12:6357–6364, 2016.
- [75] F. S. Merkt, A. Erbe, and P. Leiderer. Capped colloids as light-mills in optical traps. *New Journal of Physics*, 8:216, 2006.
- [76] A. Girot, N. Danné, A. Würger, T. Bickel, F. Ren, J. C. Loudet, and B. Pouliquen. Motion of optically heated spheres at the water–air interface. *Langmuir*, 32:2687–2697, 2016.
- [77] C. Maggi, F. Saglimbeni, M. Dipalo, F. De Angelis, and R. Di Leonardo. Micromotors with asymmetric shape that efficiently convert light into work by thermocapillary effects. *Nature Communications*, 6:7855, 2015.
- [78] D. Paul, R. Chand, and G. V. P. Kumar. Optothermal evolution of active colloidal matter in a defocused laser trap. *ACS Photonics*, 9:3440–3449, 2022.
- [79] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. Probability of second law violations in shearing steady states. *Physical Review Letters*, 71:2401, 1993.
- [80] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. *Physical Review Letters*, 89:050601, 2002.
- [81] E. Barkai, Y. Garini, and R. Metzler. Strange kinetics of single molecules in living cells. *Physics Today*, 65:29, 2012.

- [82] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.
- [83] A. Argun, A.-R. Moradi, E. Pinçē, G. B. Bagci, A. Imparato, and G. Volpe. Non-Boltzmann stationary distributions and nonequilibrium relations in active baths. *Physical Review E*, 94:062150, 2016.
- [84] C. Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78:2690, 1997.
- [85] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60:2721, 1999.

Chapter 32

Atomic Force Microscopy

SHIVAPRAKASH N. RAMAKRISHNA, LUCIO ISA

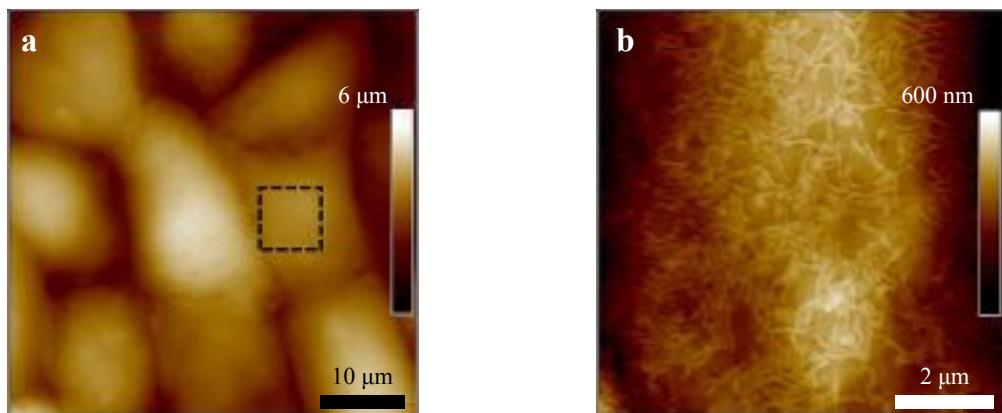


Figure 32.1: **PeakForce Tapping image of glutaraldehyde-fixed MDCK C11 cells.** (a) Overview scan ($50 \times 50 \mu\text{m}$) of several cells and (b) a higher resolved image ($8 \times 8 \mu\text{m}$) in the area indicated by the black outline. Reproduced from Ref. [1].

Atomic force microscopy (AFM) is a most commonly used technique for surface topography measurements with a sub-nanometer resolution. Along with topographical measurements, over the years, AFM has also been extensively applied to the mechanical characterization of materials, and to the measurement of frictional and functional properties of surfaces [2, 3, 4]. The recent development of new modes of operation, in particular, facilitates the imaging of soft biological samples (see, for example, Fig. 32.1) under liquid environment [5, 6], while progress in the acquisition rate enables high-speed AFM imaging and precise control of the force acting on the sample to explore and visualize of many dynamic phenomena, such as biomolecular processes [7, 8]. This chapter provides a concise description of the functioning principles of AFM and its developments over the years, with a focus to its applications in active matter systems. It starts with brief history of scanning probe microscopy (SPM) and focuses on AFM, describing the modes of operation and the physics behind the tip-sample interaction. In a later section, it introduces surface roughness quantification and different existing types of roughness parameters, which are relevant to characterize the surfaces of active colloids. It also provides practical elements to guide the reader in the image processing of experimentally acquired AFM images, and shows how to extract the roughness parameters using the open source software Gwyddion. The last section gives an insight into the use of an AFM in biological active matter.

32.1 Surface analysis in active matter systems

Active matter spans a broad range of different systems, encompassing self-propelling colloids, motile microorganisms, living cells and biomolecular entities. For all of those systems, characterization of mechanics and surface properties at the micro and nanoscale is essential. In particular, as the dimensions of the objects of interest are reduced, the relative importance of surface-to-volume ratios increases and, at the microscale, interactions are dominated by colloidal and surface forces. Among different techniques, atomic force microscopy (AFM) has become the tool of choice to characterize the topography and mechanics of surfaces at the small scale. Before dwelling on the historical developments of this powerful experimental technique and on the basics of its workings, we introduce here a few examples that demonstrate how surface analysis is essential to provide a comprehensive description of many active matter phenomena at a small scale.

The most widely used system of artificial microswimmers consists of half-coated, Janus particles having a catalytically active hemisphere capable of generating local chemical or electrochemical gradients, which are responsible for self-phoretic propulsion [9, 10]. In most cases, this is a platinum hemisphere and the chemical fuel is hydrogen peroxide [11, 12], but different catalysts, e.g. titania or iron oxide [13], have also been demonstrated. Catalytic propulsion can also be achieved via selective coating with enzymes [14]. In all of these cases, the topography of the active surface plays an important role. Even for the simplest case of thin Pt layers, significant differences in the propulsion velocity are reported by using different coating methods, e.g. sputter-coating versus evaporation, connected to a difference in the catalytically active surface [15]. In the extreme cases of thicker coatings, non-uniform distributions of the catalytic material can even generate shape asymmetries, which lead to a qualitative modification of the swimming trajectories [16]. A systematic characterization of surface properties, in particular surface roughness and chemistry is therefore important when comparing swimming velocities and efficiency of different systems. Moreover, recent reports showed that the propulsion velocity of catalytic Janus particles close to a substrate is also strongly affected by the surface properties of the substrate itself, with the degree of hydrophobicity/hydrophilicity correlating with swimming speed [17]. Finally, in experiments, active colloids are often found to adhere to interfaces and become immobilized at solid-liquid interfaces. In this case, a precise quantification of the interaction forces between particles and surfaces, such as the one provided by AFM, is an important tool to control particle motion. The characterization of surfaces is therefore important beyond the characterization of the surface of the active colloids, but it extends to the surfaces that the colloids are in contact with. Understanding this latter aspect is perhaps even more crucial for the case of biological systems, where controlling, harnessing and unveiling adhesion between cells or bacteria is at the basis of the study of their individual and collective dynamics [18, 19, 20]. When it comes to biological systems, AFM also offers the extraordinary capability to be used both as a manipulation tool and as a local mechanical probe [21, 22, 23, 24, 25]. As it will be described later, modifications of AFM tips enable the user to trap, move and characterize single cells and even single biomolecules, such as proteins or DNA [21, 22, 26, 27, 28]. The fact that an AFM essentially measures the interactions between a tip and a substrate, allows it to be operated in the so-called force-spectroscopy mode, meaning that one can use the AFM tip to measure forces upon approaching and retracting two surfaces but also after the surfaces have been put into contact, as in indentation measurements [29, 30, 31]. These types of measurements are highly interesting for biologically-relevant active systems. In fact, as described later, there are evidences of connections between a cell's metabolism and pathogenic activity and its mechanical properties [32, 33, 34, 35]. Finally, high-speed AFM images enables the direct visualization of dynamic biomolecular phenomena, such as the motion of molecular motors on myosin fibers [7, 8, 36, 37].

32.1.1 A brief history of scanning probe microscopy (SPM)

SPM is based on the principle that a sharp tip scans over a sample surface to construct the surface topography three-dimensionally [2]. The operational principle of SPM uses feedback-controlled piezoelectric scanners to control the probe-sample separation distance (Fig. 32.2a). The development of the family of SPMs started with the invention of scanning tunneling microscopy (STM) at IBM Zurich by Gerd Binning and Heinrich Rohrer during 1981, a discovery for which they won the 1986 Nobel Prize in Physics. The STM (Fig. 32.2) provides information about the electron density distributions, the electronic work function, and the surface topography. The main principle of STM is that of electrons quantum-mechanically tunneling between an atomically sharp conducting tip and the sample surface. These are separated from each other by a potential barrier. Tunneling can

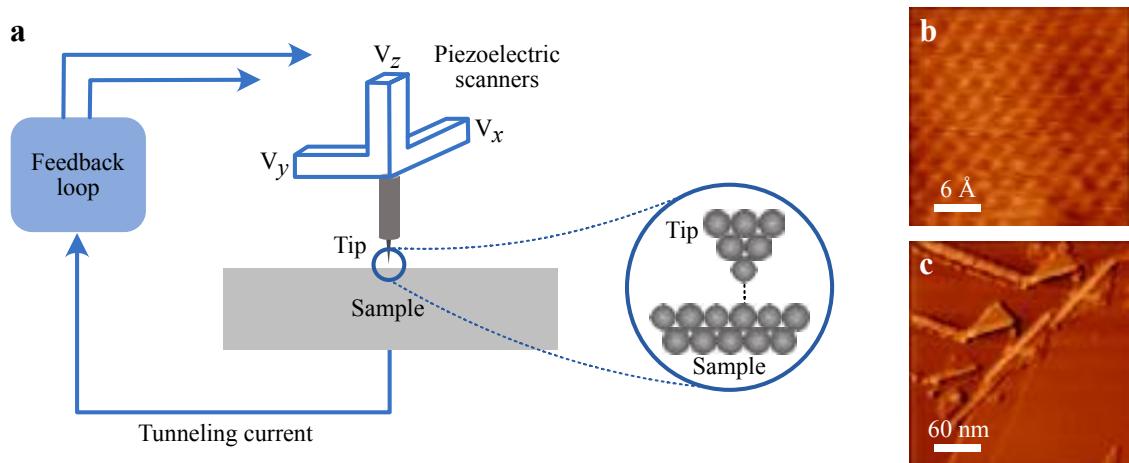


Figure 32.2: **Scanning tunneling microscopy.** (a) Schematic diagram of a scanning tunneling microscope. The zoomed-in regions depicts an ideal, atomic sharp tip scanning over the individual atoms of the surface. The atoms are represented with grey spheres. STM image with atomic resolution of a (b) MoS₂ crystal obtained in ambient condition and of (c) MoS₂ nanotubes deposited on HOPG surface.

take place from tip to surface or surface to tip, simply depending on the potential applied [38]. The tunneling current density j can be described as:

$$j \propto \frac{\sqrt{\Phi}}{s} \cdot V \cdot \exp(-k\sqrt{\Phi}s), \quad (32.1)$$

where the effective electron affinity is $\Phi = 1/2(\Phi_1 + \Phi_2)$, the applied potential is V , the tunneling distance is s , and k is a constant [39]. Since j varies exponentially with the distance s the STM very sensitive to small height variations and vertical resolutions up to 0.01 nm can be achieved. The lateral resolution is about 0.1 nm.

32.2 Atomic force microscopy (AFM)

Though STM is a powerful tool for imaging surface topography, it has a serious disadvantage, namely that the samples need to be conductive. To overcome this drawback, in 1986, Binnig, Quate, and Gerber developed the AFM, also known as the scanning force microscope (SFM) [40]. Unlike in STM where the imaging parameter is the tunneling current, the AFM measures the forces between the tip and the surface as the tip moves over the sample (Fig. 32.3). A cantilever with a known spring constant with a sharp tip at its end is mounted over the sample surface. The cantilever is deflected by an amount proportional to the force and its deflection is usually measured by an optical detection method.

AFM has not only the advantage (over, e.g., SEM and STM) that non-conductive samples can be analyzed with near-atomic resolution, but also that the analysis can be performed in air, in vacuum, or even in a liquid. The method has become more refined over the last twenty years, so that today cells, DNA fragments, ceramics, polymers, colloidal particles, catalysts, and many other very different samples can be analyzed by AFM. The large flexibility of the AFM is one of the reasons why AFM is so frequently used. Another is its very reasonable cost, compared to instruments that require ultrahigh vacuum.

When using AFM, a tip attached to the end of a cantilever is brought close to the surface while scanning the sample along the x- and y-direction. A feedback mechanism enables the piezo-electric scanners to maintain the cantilever either at constant deflection or constant height between the tip and the sample. During scanning over a sample surface, a laser beam is reflected from the backside of the reflective cantilever to a four-quadrant, position-sensitive detector (PSD). The PSD converts the voltage differences due to the movements of the cantilever to a three-dimensional image, which corresponds to the topography of the sample surface. The schematic representation of the working scheme of AFM is shown in Fig. 32.3.

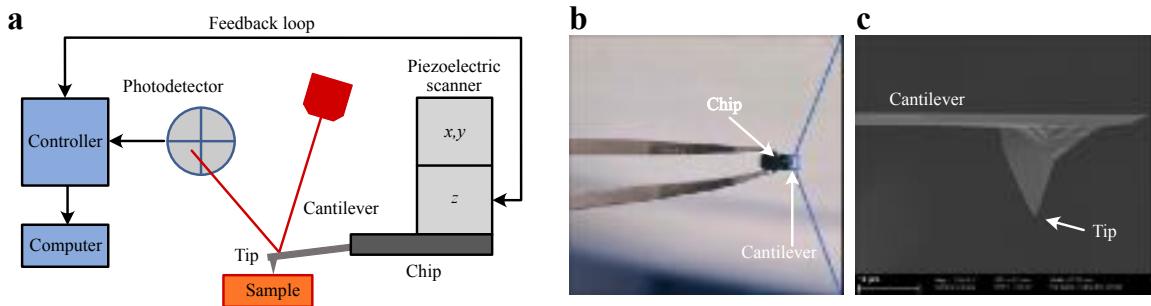


Figure 32.3: Schematic diagram of an atomic force microscopy setup. (a) A tip attached to the end of a cantilever is brought close to the surface while scanning the sample along the x - and y -direction. A feedback mechanism enables the piezo-electric scanners to maintain the cantilever either at constant deflection or constant height between the tip and the sample. During scanning over a sample surface, a laser beam is reflected from the backside of the reflective cantilever to a four-quadrant, position-sensitive detector (PSD). The PSD converts the voltage differences due to the movements of the cantilever to a three-dimensional image, which corresponds to the topography of the sample surface. (b-c) An optical (b) and SEM (c) image of an AFM chip holding the cantilever and a cantilever holding the tip.

32.2.1 Forces between the tip and the sample surface

The most important tip-sample force in the AFM is the *van der Waals interaction*. It plays a major role in all phenomena that involve intermolecular forces. For two electrically neutral macroscopic objects interacting in a medium, the van der Waals forces play a dominant role in the total adhesion force. The origin of van der Waals forces is believed to be at the atomic or molecular level. Three important interactions contribute to the total van der Waals force: (i) the *Keesom force*, due to dipole-dipole interaction (Fig. 32.4a), (ii) the *Debye force*, due to the interaction between a dipole and the induced dipole (Fig. 32.4b), and (iii) the *London force*, also called dispersion force, due to the interaction between an instant dipole (emerging from fluctuations of the electron cloud density in non-polar atoms) and the induced dipole (Fig. 32.4c). All the three contributions to the van der Waals force are discussed below. Other important forces which acts between AFM tip and sample are summarized in Table 32.1 and the reader is referred, for details, to the relevant literature, for example Refs. [41, 42].

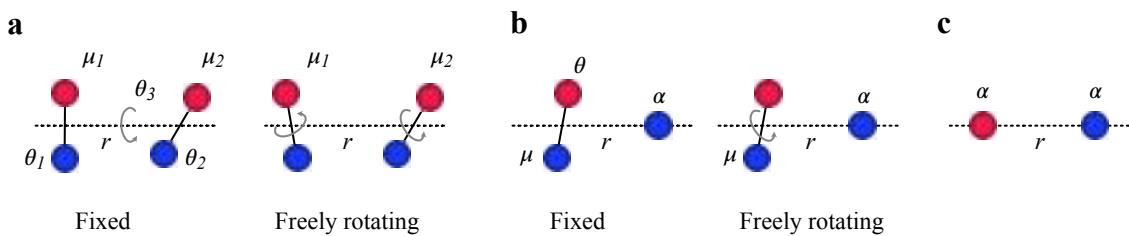


Figure 32.4: Interaction components in the van der Waals force. (a) Dipole-dipole interaction (*Keesom force*): Interaction between two dipoles, when the relative orientation angles are fixed and when the dipoles can freely rotate. (b) Dipole-induced-dipole interaction (*Debye force*): Interaction between a dipole and a non-polar atom or molecule, when the angular orientation of the dipole is fixed and freely rotating. (c) Dispersion interaction (*London force*): Interaction between two non-polar atoms.

Dipole-dipole interaction (Keesom force): These forces exist between molecules having permanent dipoles. The interaction depends on the relative orientation of the two dipole moments. The total interaction is

a function of distance between the two dipoles and the relative angle between them:

$$V_{\text{Keesom}}(r, \theta_1, \theta_2, \theta_3) = -\frac{\mu_1 \mu_2}{4\pi\epsilon_0 \epsilon r^3} (2 \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \cos \theta_3), \quad (32.2)$$

where μ_1 and μ_2 are the dipole moments, r is the distance between the dipoles, and θ_1 , θ_2 , and θ_3 are the three angular orientations as represented in Fig. 32.4. The maximum interaction will occur when two dipoles are lying in the same line and $\theta_1 = \theta_2 = 0$:

$$V_{\text{Keesom}}(r, 0, 0, \theta_3) = -\frac{\mu_1 \mu_2}{2\pi\epsilon_0 \epsilon r^3}. \quad (32.3)$$

At larger separations, or in a medium with high dielectric constant, when the angular dependency of the interaction energy is small versus the thermal energy ($k_B T$), the dipoles can rotate quasi-freely. In this case, the Keesom interaction becomes zero due to angular averaging. However, due to the finite temperatures, there always exists a Boltzmann weighting factor, which favors angles resulting in lower interaction energy (i.e., attractive forces). The angle-averaged dipole–dipole interaction, which is called Keesom interaction potential is given as

$$V_{\text{Keesom}}(r) = -\frac{\mu_1^2 \mu_2^2}{3(4\pi\epsilon_0 \epsilon)^2 k_B T r^6}, \quad \text{for } k_B T > \frac{\mu_1 \mu_2}{4\pi\epsilon_0 \epsilon r^3}. \quad (32.4)$$

Dipole-induced-dipole interaction (Debye force): All atoms and molecules are polarizable. Dipole–induced-dipole interactions arise when a polar atom or molecule polarizes a nearby non-polar atom or molecule and induces a dipole moment on it. For a dipole moment μ oriented at an angle θ to the line joining the dipole, non-polar atom or molecule, the electric field E acting on non-polar molecule is given as

$$E(r) = \frac{\mu \sqrt{1 + 3 \cos^2 \theta}}{4\pi\epsilon_0 \epsilon r^3}. \quad (32.5)$$

The interaction potential can be written as:

$$V_{\text{Debye}}(r, \theta) = -\frac{1}{2} \alpha E^2 = -\frac{\mu^2 \alpha (1 + 3 \cos^2 \theta)}{2(4\pi\epsilon_0 \epsilon)^2 r^6}, \quad (32.6)$$

where α is the polarizability of the non-polar molecule. The angle-averaged interaction potential between a polar and non-polar atom or molecule is given as

$$V_{\text{Debye}}(r) = -\frac{\mu^2 \alpha}{(4\pi\epsilon_0 \epsilon)^2 r^6}. \quad (32.7)$$

This form is known as Debye energy or induction form of the van der Waals interaction.

Dispersion interaction (London force): The dispersion interaction is the third and the most abundant $1/r^6$ contribution to the van der Waals forces, and has quantum mechanical origin. It arises due to the charge fluctuations of the atoms. For a non-polar atom or molecule, the time average of its dipole moment is zero. However, at any instant of time there can exist a transient dipole moment due to the fluctuations in the electron cloud density of the atom or molecule. This transient dipole can generate an electric field and polarize any nearby neutral atom, and thus induce a dipole moment. The two dipoles interact with each other resulting in an attractive force between them. The expression for attractive part of the dispersion interaction potential between two atoms or molecule is

$$V_{\text{London}}(r) = -\frac{3h\nu\alpha^2}{4(4\pi\epsilon_0 \epsilon)^2 r^6}, \quad (32.8)$$

where h is the Planck's constant ($h = 6.62607015 \times 10^{-34}$ Js), ν is the characteristic frequency related to the first ionization potential of the atom or molecule, and r is the distance between the two atom or molecules having the polarizability of α .

Table 32.1: **Surface forces.** Different types of surface forces, specifications and their characteristics. For details, the reader is referred to Ref. [41, 42].

Type of forces	Alternative names/sub-classes	Main characteristics
Electrostatic force	Ionic, coulombic force	Strong, long-ranged forces acting between charged surfaces.
van der Waals forces	Keesom force, Debye force, London force (dispersion force)	Origin of these forces between atoms and molecules, acts over $\approx 10\text{ nm}$ between surfaces, can be attractive or repulsive depending on the medium.
Capillary force	Meniscus force	Acts when bridging of adsorbed water layer occurs between two surfaces.
Double layer and Derjaguin, Landau, Verwey and Overbeek (DLVO) forces	EDL force and DLVO force	Acts between charged surfaces in electrolyte solutions. Combination of van der Waals and double layer force yields DLVO forces.
Solvation force	Hydration force (in water), structural forces, non-DLVO forces	Short range forces, occurs between ions and molecules.
Hydrophobic force	Non-DLVO forces	Acts between two hydrophobic surfaces.
Steric force	Entropic force	Acts between polymer-modified surfaces, range comparable to molecular size.

32.2.2 Operational modes of the AFM

AFM can work in a *contact mode* (using the deflection or force of the cantilever as a feedback), a *non-contact mode* (using the amplitude or characteristic frequency of oscillating cantilever as a feedback), or intermittent-contact modes (*tapping mode*).

In order to understand in which regime of forces different modes of AFM works, let us take the Lennard-Jones potential $U_{\text{LJ}}(r)$, which is attractive at long range (large r) and repulsive at short range (small r). The

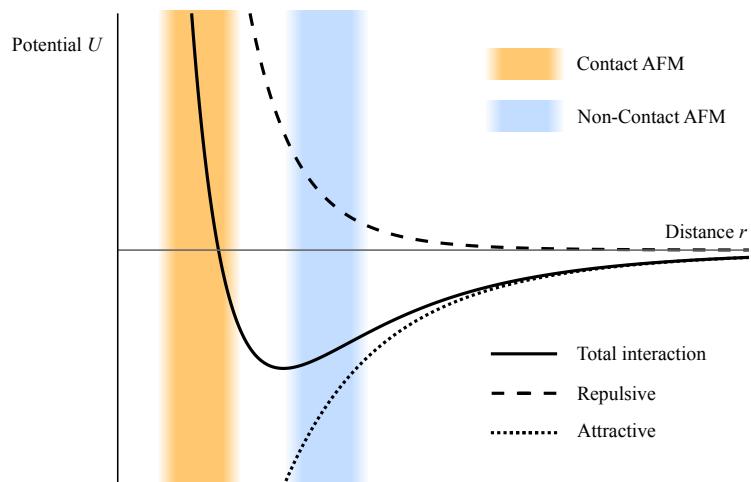


Figure 32.5: **AFM ranges for contact and non-contact modes.** Lennard-Jones potential (continuous line) and its constituent attractive (dotted line) and repulsive (dashed line) components. The ranges for contact AFM and non-contact AFM are represented with different color nuances (orange and light blue, respectively).

Lennard–Jones potential (Fig. 32.5) can be written as

$$U_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (32.9)$$

It describes the potential between the two atoms (or molecules) set at a distance r from one another. The term $(\frac{\sigma}{r})^{12}$ is repulsive and is responsible of the repulsion at short range. The term $-(\frac{\sigma}{r})^6$ is attractive and is responsible of the attraction at long range. ϵ is the measure of how strongly the two atoms attract each other, σ is the shortest distance at which the potential between the two atom or molecules is zero (see also Chapter 16 for more insights about Lennard–Jones potential).

Contact mode: The *contact mode* (Fig. 32.6) involves analyzing the sample after the contact between tip and surface has occurred, i.e., keeping the tip is always in contact with the sample surface (repulsive interaction regime) throughout the scanning. This is the most frequently used operational mode. The contact mode has the disadvantage that the tip–sample interaction force can have destructive effects on the sample surface. While with inorganic samples these effects are generally insignificant (Fig. 32.6a), with organic thin films, soft polymers, or, in particular, biological samples, the surface texture can be destroyed by the moving tip. AFM under water (liquid cell) can somewhat improve the situation, since the meniscus forces present under ambient conditions add to the forces exerted on the sample.

Apart from being used as an imaging tool, contact mode AFM is also used for measuring friction forces, often known as friction force microscopy (FFM) or lateral force microscopy (LFM). While scanning in contact-mode the frictional forces between the tip and surface cause twisting of the cantilever. Both the vertical and lateral movements of the cantilever are detected by the laser beam, which is reflected off the rear side of the cantilever to the PSD. The photo-diode signal (Fig. 32.3) $\{(A + B) - (C + D)\}$ gives the topographical

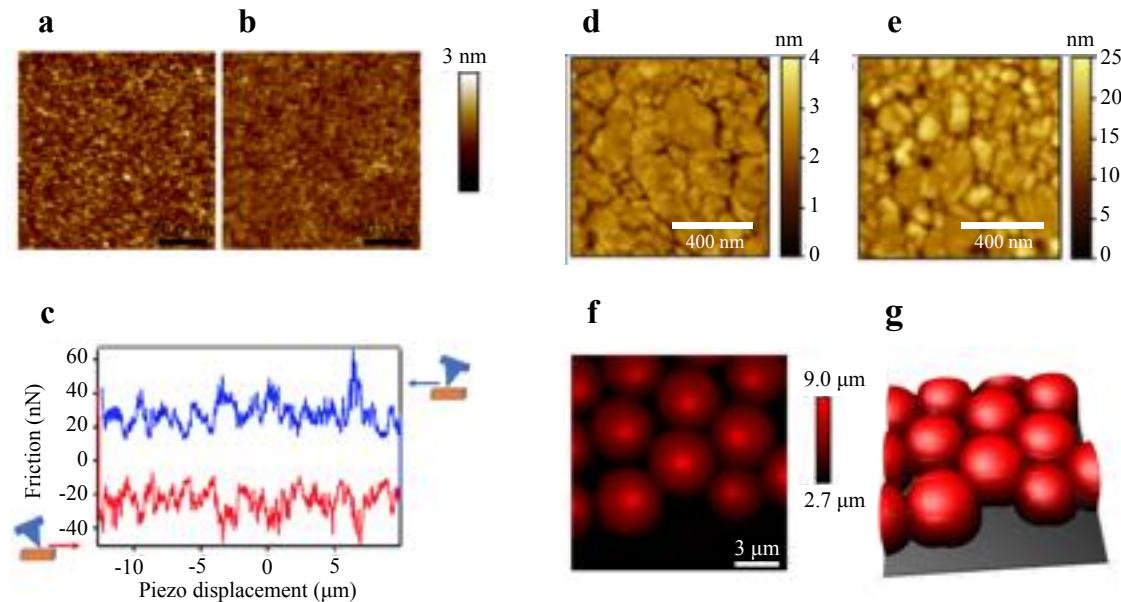


Figure 32.6: **AFM's different acquisition modes.** *Contact mode:* Contact mode AFM image of 2 nm platinum film deposited on a silicon wafer by (a) thermal evaporation and (b) sputter coating techniques. (c) Friction loop in contact mode AFM between a tip and a rough silicon surface. While scanning in contact-mode the frictional forces between the tip and surface cause twisting of the cantilever, which is recorded by the photodiode and converted into force. The area enclosed in the friction loop gives the amount of energy dissipated while scanning the cantilever across the surface. *Tapping mode:* AFM tapping mode images of (d) ultra-flat gold and (e) silver surfaces prepared by thermal evaporation coating. (f, g) Two-dimensional and three-dimensional tapping mode images of polystyrene Janus particles (4.5 μm particle and 25 nm of gold coating) deposited on a poly-L-lysine (PLL)-coated glass surface.

information whereas the signal $\{(A+C) - (B+D)\}$ gives the information on the lateral twist in the cantilever due to the friction. Fig. 32.6c shows a representative friction loop obtained during the back and forth scanning of the cantilever (trace and retrace). The area inside the loop gives the amount of energy dissipated while scanning the cantilever across the surface.

Non-contact mode: A partial solution to the sample-damage problem is to let the AFM run in the *non-contact mode*. In this mode, the system vibrates a stiff cantilever near its resonant frequency (typically 100 to 400 kHz) with amplitude of a few tens to hundreds of Ångstroms. The system detects changes in the cantilever's resonance frequency or vibration amplitude. With this method, the instrument relies on the weak attractive forces in the region before the contact occurs. Although it is possible, in principle, to use this small force in the automatic control loop, as one would use the repulsive force in the contact mode, this procedure is rarely used due to the instability of this approach (large sensitivity to thermal effects). In order to operate successfully in the non-contact mode, the cantilever is oscillated at a high frequency with a small amplitude ($< 5 \text{ nm}$). The force gradient is then measured by modifications in the amplitude, frequency, and phase of the vibrating cantilevers. This method has been successfully used, especially in the case of very flat samples.

Tapping mode: A further variant is the so-called *tapping mode* (Fig. 32.6d-g). The cantilever is vibrated close to its resonant frequency with a much larger amplitude (20 to 100 nm). With each cycle the tip comes briefly into contact with the sample. Here the forces are smaller than with the contact mode and always normal to the surface, rather than the sliding (and scraping) movement that is characteristic of contact mode. This method is much less destructive to sensitive sample than contact mode and allows to non-destructively characterize soft bio-interfaces at the nanoscale.

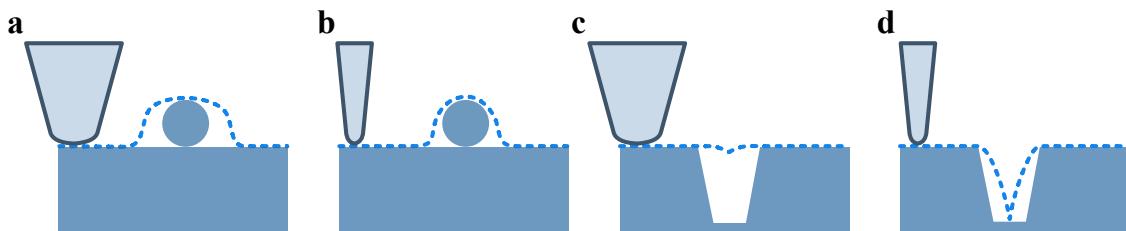


Figure 32.7: **Artifacts in AFM imaging.** Examples of artifacts during AFM imaging: (a) A tip radius larger than the sample feature will produce reverse imaging effect. (b) Even with sharper tips, the shadows of the spherical features are masked due to the tip shape. (c) Trenches smaller than the tip size cannot be resolved. (d) With relatively sharper tip compared to the trench, the depth can be resolved, however, the walls of the trenches cannot be resolved. Adapted from Ref. [43].

With all methods it is important to note that AFM images consist of a convolution of the sample morphology and the tip morphology. This issue is particularly significant if the characteristic features of the sample surface are of a similar size of the radius of the tip (imagine trying to recognize the shape of somebody's face in the dark with your finger, your elbow, or your knee). Typical artifacts (Fig. 32.7) are, for example, a rounding of sharp edges. If the shape of the tip is determined independently, the image can be mathematically corrected to a certain extent, such that feature sizes can be corrected. Any information that is lost because of the tip shape, however, cannot be recovered.

32.3 Force spectroscopy in AFM

Together with topography, in contact mode one can also measure the forces between the tip and the sample. The measurement of force as a function of the tip-sample distance is called *force spectroscopy mode*. This technique can also be used to analyze surface contaminants, viscosity, film thickness, and local variations in the elastic properties of the surface. Fig. 32.8 shows a schematic force-distance curve obtained during a single approach-retraction cycle of the AFM tip.

The obtained force-distance curve corresponds to the cantilever (having a normal spring constant, k_N) deflection with respect to the piezo position (d). By calibrating the deflection (x) obtained from the photodetector, the force F can be calculated using Hooke's law: $F = k_N x$.

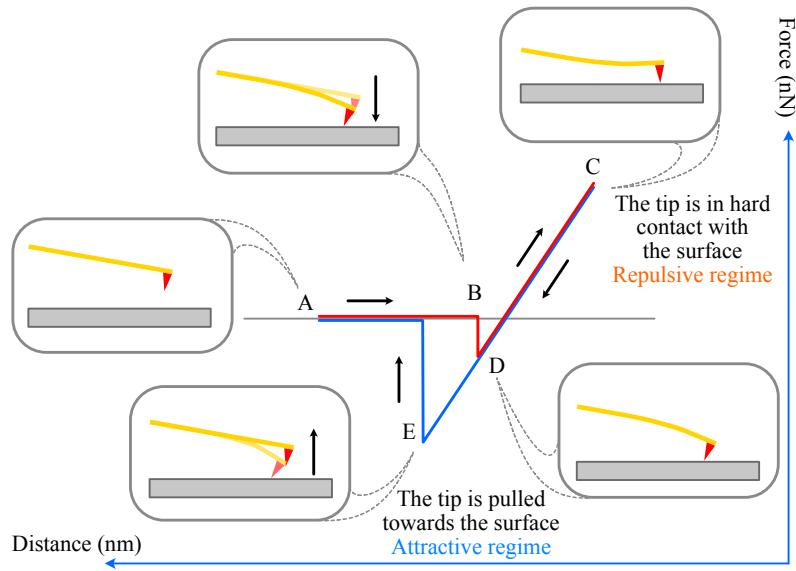


Figure 32.8: **Force–distance curve.** Schematic of the force–distance curve obtained during a single approach–retraction cycle of the AFM tip. The sketches represent the deflection of the cantilever for the various positions in the cycle.

Figure 32.8 shows the schematic of one such force versus distance measurement. Initially, when the tip is far away from the surface, no forces between the tip and sample are measured (position A). As the tip approaches the sample surface, the initial contact between the tip and the surface is made by the attractive van der Waals forces (position B), which deflects the cantilever downwards and leads to a jump to contact. The tip later applies a constant force upon the surface that leads to sample indentation and cantilever deflection (position C). Once a pre-defined trigger voltage is reached, the tip tries to retract from the surface (position D). Various adhesive forces between the sample and the AFM tip cause the cantilever to bend prior to lift-off. Finally, the tip withdraws and loses contact from the surface upon overcoming the adhesive forces and retains its original position (A). The portion E gives the amount of adhesive force acting between sample and the tip.

32.3.1 Nanoindentation

AFM force-distance-curve-based nanoindentation measurements are frequently used to study the mechanical behavior of soft solids, like polymers, hydrogels and biological samples [45, 46, 47, 48, 49]. Figure 32.9 illustrates the principle of a stiffness measurement on a single cell. At first, deflection-vs-distance curves are recorded on the cells by ramping the Z piezo, as described above. The obtained deflection is converted in to a force by multiplying it by the spring constant of the cantilever. The scanner position is then converted into indentation (δ) by subtracting the cantilever deflection. The stiffness values were obtained by fitting the force vs indentation curves to the suitable models, generally the Hertz model, when there is no adhesion between the tip and the sample [50]:

$$F = (4\sqrt{R})/(3(1 - \nu^2))E \cdot \delta^{1.5} \quad (32.10)$$

where F is the applied force, R is the radius of the tip used. By knowing the tip's elastic properties and sample's Poisson's ratio ν , the Young's moduli E of the cells are obtained. During the indentation measurement, it is important to select AFM cantilevers with spring constants in line with the sample stiffness. If the cantilever is too stiff, it could easily damage the cell, on the other hand, if the cantilever is too soft, it will not indent to the cell sufficiently to acquire a reliable stiffness value [44]. Often, other contact mechanics models like Sneddon model, Johnson, Kendall, and Roberts (JKR) model, or Derjaguin, Müller, and Toporov (DMT) model are also used for extracting the sample modulus depending upon the tip shape, tip-sample interactions and the deformations [51, 52, 53].

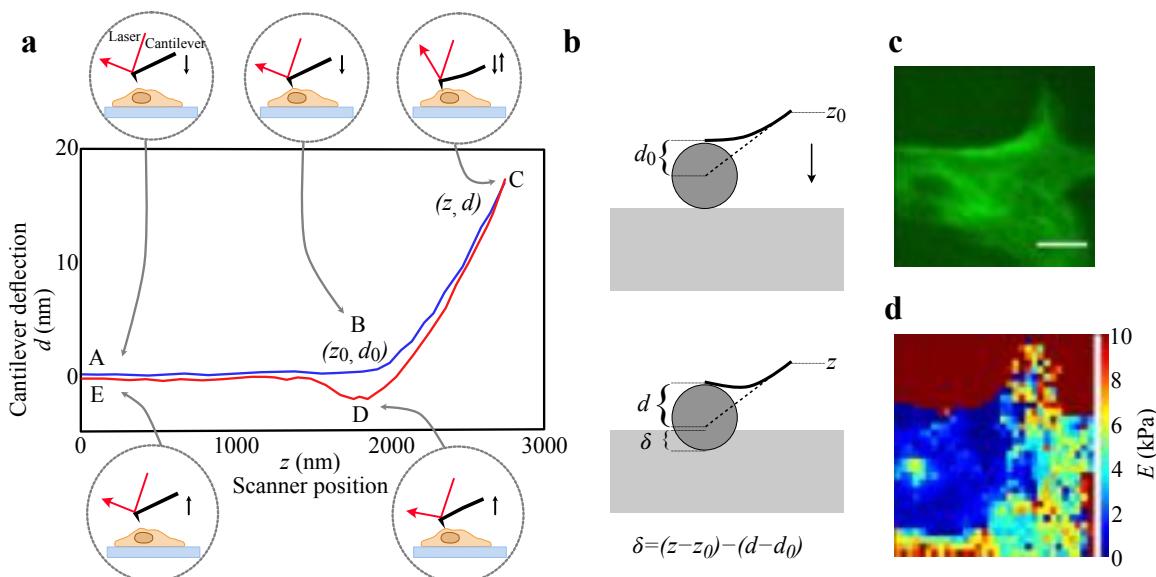


Figure 32.9: Measurement of cellular stiffness using AFM nanoindentation. (a) The cantilever begins at point A, above the cell surface, at an initial distance of a few microns, and starts approaching the cell. During the approach, the sample indentation remains zero until it reaches point B, where the tip makes contact with the cell. Between points B and C, the cantilever indents into the cell until its deflection reaches a set point. The cantilever is then withdrawn from the cell to point D, where it is often pulled downwards due to tip-sample adhesive interaction. Eventually, the tip detaches from the cell and the cantilever returns to its initial location at E. (b) Relationship between the indentation and the recorded z and d signal is shown. (c) Fluorescence image of a 3T3 fibroblast transfected with GFP vimentin, showing only a portion of the cell. Scale bar represents 20 μm . (d) Recorded 32 x 32 pixel stiffness map of the same area. Adapted from Ref. [44].

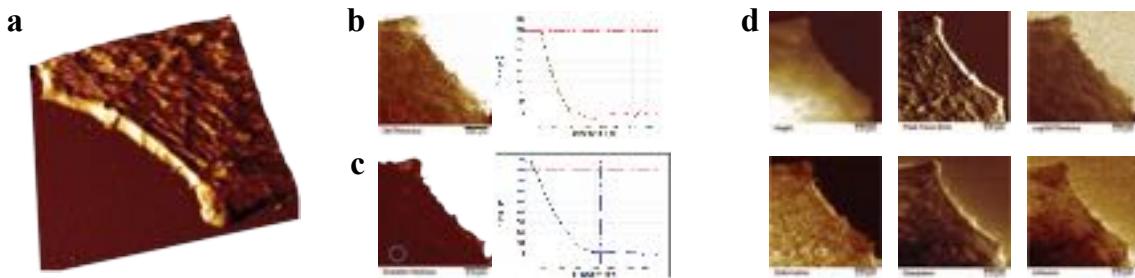


Figure 32.10: Mapping mechanical property in QNM mode. (a) Lamellipodium of a mouse B16 cell imaged under buffer in QNM mode. 3D rendering of the lamellipodium topography showing actin fibrils. (b,c) An example curve fit and comparison of modulus maps between DMT and Sneddon models. (d) Additional images collected simultaneously, mapping the deformation, dissipation, and adhesion properties of the cell. Adapted from Ref. [54].

32.3.2 Force-distance curve-based imaging

The force-distance curves are also capable of producing topographical images in combination with the local physical/chemical properties of the sample such as modulus of elasticity, deformation and adhesion. This is mode of imaging in which manufacturers tend to have their own trade mark names, such as, PeakForce Tapping® (PFT), Fast Force Mapping® (FFM), and Quantitative imaging® (QI) mode.

This mode is similar to the Tapping Mode AFM, however, unlike in the tapping mode where the cantilever oscillation is near to the resonance frequency, here the oscillation frequency is much lower (typically between 0.25-2KHz). The oscillation combines the benefits of contact- and tapping-mode imaging by having direct force control and avoiding lateral forces. A map of force-distance curves is produced and by keeping the force as a constant trigger, and images, where each pixel corresponds to the local values of the modulus, adhesion force, and deformation depth, are obtained along with the height image. PeakForce Quantitative Nanomechanical Mapping (QNM®) (Bruker) is a recent and powerful technique that provides quantitative mechanical characterization of surfaces at the nanoscale level. Figure 32.10 presents the topographical image and mechanical property maps acquired in QNM mode, illustrating the lamellipodium of a mouse B16 cell immersed in buffer solution.

32.4 Surface roughness

As mentioned at the beginning of this chapter, characterizing the topography of surfaces has strong implications on their application in active matter systems. One way to reduce the complexity of a surface's topography into a single quantity to compare with other materials is to measure surface roughness. Roughness is often a good predictor of the performance of many mechanical components but it is also related to the accessible surface area, e.g. for catalytic materials, such as the caps of Janus microswimmers. Most real surfaces have roughness over many length scales. Even when two thoroughly polished surfaces come into contact, due to the existence of the micro/nano scale roughness on the surface, the actual contact occurs at discrete points or *asperities*, which effectively lowers the area of contact to a small fraction of its apparent value. For example, in Refs. [55, 56], the increase in roughness on the substrate as nanoparticle density decreases the pull-off (adhesion) and friction forces between a polyethylene bead and the silica particle covered surface. The adhesion and the friction forces measured by AFM here is directly proportional to the number of asperities which comes under contact (see Fig. 32.11).

The roughness of a surface can be defined using many parameters, but only a few parameters are relevant in practice. For most engineering surfaces, the roughness is a mixture of shapes that dominate at different length scales. Fig. 32.12 shows a textured surface, which is regular on large length scales and irregular on small length scales. The regular, wave-like structure (with a length scale of a couple of millimeters) over the surface usually results from a machining process. If a line is drawn orthogonally in the direction of the lay, the spacing and wave height can be determined. Further magnification of the structure yields a finer irregular structure of

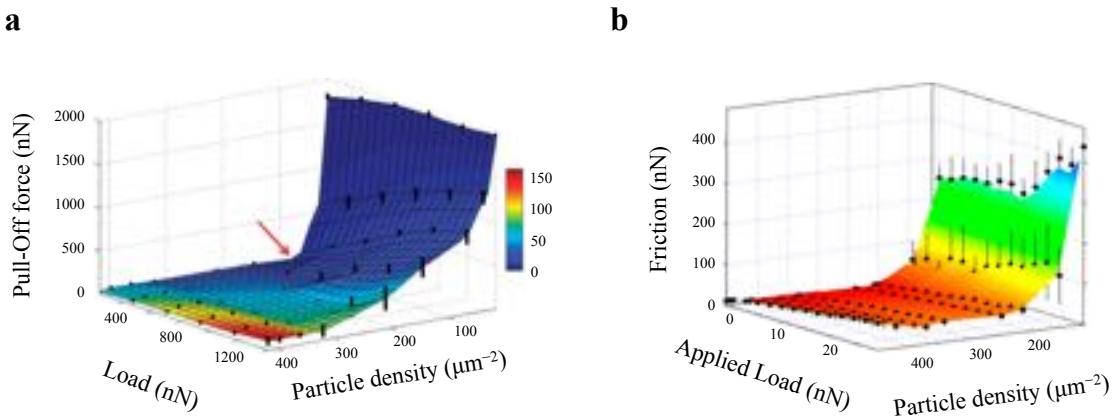


Figure 32.11: **Effect of roughness on adhesion and friction.** (a) Pull-off force (adhesion force) and (b) Friction force measured as a function of applied load on different roughness as nanoparticle density. The adhesion and friction forces here are affected by the number of asperities that come into contact, which in turn alters the real area of contact. The friction force recorded here is directly influenced by the adhesion force. Reproduced from Ref. [55, 56].

asperities and valleys (with length scales usually ranging from a couple of micrometers to nanometers), which is superimposed on the wavy structure. Further magnification of the asperities and valleys ultimately reveals the individual atoms on the surface [57].

The surface shown in Fig. 32.12 has macro- and microscopic details that can be broadly categorized into:

- **Atomic-scale roughness:** Roughness due to the atoms on the surface.
- **Roughness:** Small waviness in the irregularities described by hills and valleys, which have random spacing, depths and heights.
- **Waviness:** Much larger spacing than the roughness with quite regular structures. For engineering applications, materials are often subjected to heat treatment in order to alter their mechanical properties. This involves heating and cooling the materials — often to extreme temperatures. During this process, waviness of the surfaces is often imparted. Additionally, during manufacturing processes, vibrations and polishing processes can lead to surface waviness.

32.4.1 Measuring roughness

Surface topographical measurement of engineering materials is essential to confirm a surface's suitability for its function. The variations in z -height for a single profile are determined, and a center line or mean height located such that the summation of areas above and below this line are equal. Here z is defined as the height of the profile relative to the mean height of the profile. In 3D surfaces, a mean surface plane, instead of mean height line, is defined.

Several amplitude parameters can be defined to characterize the surface based on the vertical deviation of the roughness profile from the above-described mean height line (Fig. 32.13a).

Mean or average roughness: The mean (or average) roughness R_a (sometimes called S_a) is the arithmetic average of the absolute values of z along the roughness profile:

$$R_a = \frac{1}{L} \int_0^L |z(x)| \, dx. \quad (32.11)$$

R_a is one of the most commonly used roughness parameters in engineering applications. It gives a good general description of the height variations in the surface. It is also called CLA (center line average).

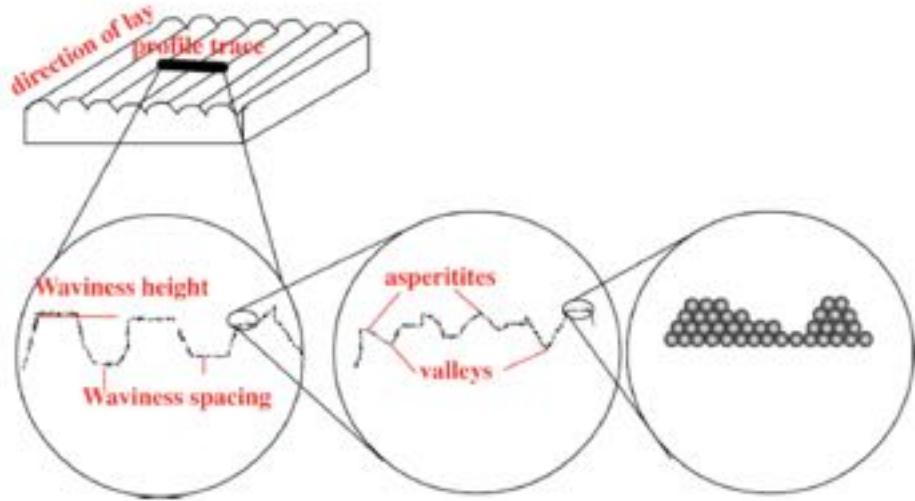


Figure 32.12: **Hierarchical structural patterns on a regular surface.** [@Agnese: redraw figure.](#) Illustration describing the hierarchical structural patterns on a regular surface. The magnification of a regular wavy surface texture shows fine-scale roughness with micron-size features.

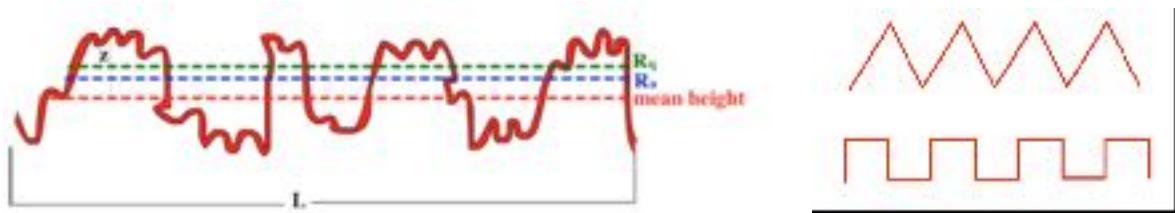


Figure 32.13: **Profile on a regular surface.** [@Agnese: redraw figure.](#) (a) Single vertical profile trace along the x -direction. First, from the vertical profile trace the mean height line (dashed red line) is determined. The vertical deviation from the mean height line is named z . From z , the parameters R_a and R_s are calculated according to Eqs. (32.11)-(32.12). (b) Surface profiles with same RMS values: therefore, surface profiles with the same value can actually represent different structural features, which can result in different contact areas upon many contact situations.

RMS roughness: The root-mean-square roughness R_a (called also S_a or the RMS) is the standard deviation value of z about the mean height. It is a statistically more meaningful parameter and is defined by

$$R_q = \text{RMS} = \sigma = \sqrt{\frac{1}{L} \int_0^L z^2(x) dx}. \quad (32.12)$$

Peak-to-valley roughness: The peak-to-valley roughness R_t is defined as the vertical height from the highest peak to the deepest valley:

$$R_t = \max z(x) - \min z(x). \quad (32.13)$$

R_p roughness: The roughness R_p is defined as the vertical height from the highest peak to the mean height:

$$R_p = \max z(x). \quad (32.14)$$

R_z roughness: The roughness R_z is defined as the distance between the average of the five lowest points to the five highest points.

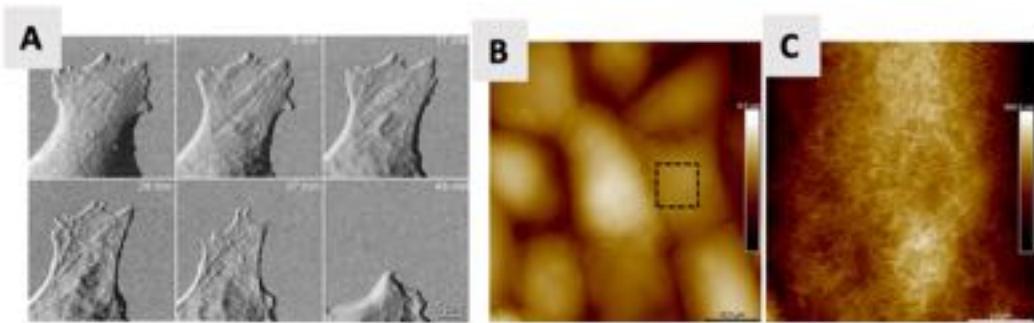


Figure 32.14: **Figure title missing. [FIGURE NOT MENTIONED IN THE TEXT.]** (A) Imaging of cells past and present. (A) Time-lapse AFM imaging of the movement of a live MDCK cell. Reproduced from Ref. [60]. (B) AFM image of glutaraldehyde-fixed MDCK C11 cells. Overview scan ($50\ \mu\text{m} \times 50\ \mu\text{m}$) of several cells and (C) a higher resolved image ($8\ \mu\text{m} \times 8\ \mu\text{m}$) in the area indicated by the black outline. Together with PeakForce Tapping mode and a specially designed live cell imaging probe Schillers et al., were able to visualize the very soft and flexible structures on living cells under physiological conditions. Reproduced from Ref. [1].

The above-mentioned roughness parameters consider only the variation of the z -height about the center line of the profile and do not consider the correlation between the neighboring z values. The lateral variations of the topography are ignored here, and can be best described using slope and spacing parameters. For example, Fig. 32.13b shows that surface profiles with the same value can actually represent different structural features, which can result in different contact areas upon many contact situations.

32.5 Applications of AFM to active matter

In the context of active matter, AFM has mostly been employed to characterize biological materials in liquid, including the investigation of many dynamic cellular processes in real time [58]. One of the early attempts made by Binning et al. in the early 90s, successfully showed the use of AFM in the study of kinematics of biological processes [59]. Dynamic AFM imaging of living cells infected by virus highlighted a change in the cell's stiffness after the addition of the virus, which would cause the tip to penetrate through the cell membrane during the measurement. Moreover, a change in the cell membrane's structure was observed by scanning a few hours after infection.

Although, many attempts have been made during early 90s to use AFM to investigate cellular behavior, the existence, at that time, of contact scanning only made it difficult to overcome the stress applied by the tip on the cells during the imaging. However, the development of new modes of AFM during the last decades has made it possible to overcome these limitations and use AFM for imaging live cells as an alternative and a complement to optical microscopy [60].

AFM is often used to characterize cell roughness, in which the topographical images are used to quantify the surface roughness of cell membranes in detail. The roughness of a cell's membrane is an indicator of cell's health state, and it is involved in several cellular mechanisms such as cell motility, adhesion and intracellular contact [61, 62, 63, 64, 65]. Monitoring roughness enables investigating the damages caused to the erythrocytes by artificial organs [66], the effect of antimicrobial peptide PGLa on bacteria [67], the effect of aldosterone on cardiomyocytes [68] and the aging of red blood cell [69], to name a few examples.

In more detail, an example is given by a study by Ohta et al., who observed mechanically stressed red blood cells via AFM and showed morphological changes because of shear stress. In the case of normal, unstressed cells, fine protrusions were observed on the membrane, while, on the stressed cells, the protrusions disappeared and the RMS roughness values increased, which was associated to the liberated hemoglobin concentration [66].

In addition to imaging the cell surfaces, AFM has been comprehensively employed to study the mechanical properties of the cell, particularly, via the stiffness measurements [23, 44, 70]. Measuring stiffness is critical in understanding many biological processes including cell growth, motility, division, differentiation, and tissue homeostasis [71]. The stiffness of live cells is quite often considered as an index to assess the cytoskeletal

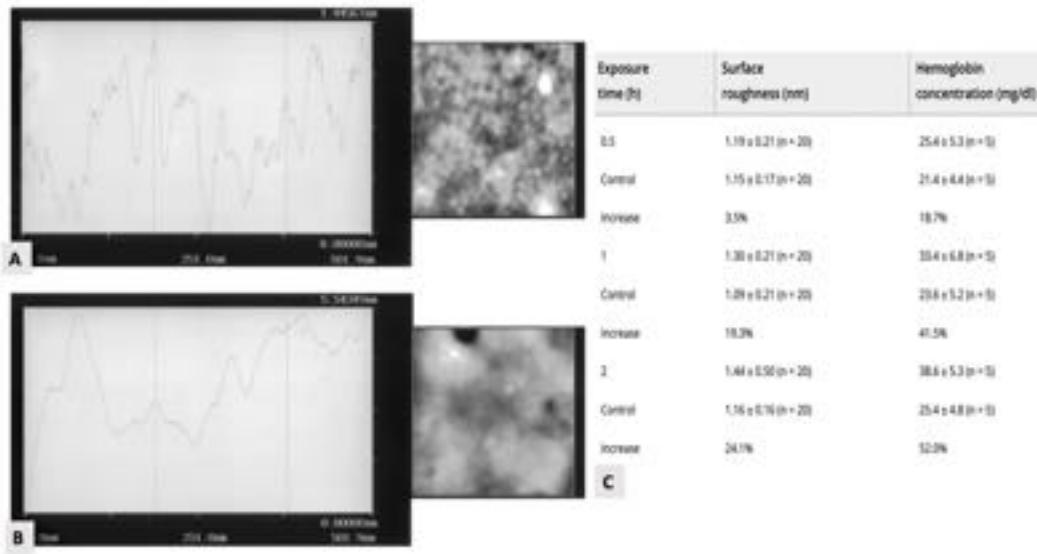


Figure 32.15: **Figure title missing. [FIGURE NOT MENTIONED IN THE TEXT.]** (A) AFM surface images and section profiles of (A) normal and (B) stressed sheep erythrocyte. (C) Table of RMS roughness values and hemoglobin concentration after shear stress. Reproduced from Ref. [66].

structure, myosin activity and other cellular processes. Furthermore, changes in cell stiffness are frequently found to be linked with several diseases such as tumor formation, cancer and metastasis [32, 34, 35, 72, 73].

As examples, for instance, Oberleithner et al., investigated the effect of salt concentration on the physical properties of the endothelial cells to comprehend the role of concentration of plasma sodium on the blood pressure [33]. To measure their stiffness, the living cells were incubated for 3 days in a culture medium containing aldosterone at a physiological concentration (0.45 nM). The results showed that the stiffness of the endothelial cell was unchanged by acute changes in sodium concentration <135 mM but increased sharply and quickly between 135 and 145 mM. The study suggested that the variations in plasma sodium concentration may affect endothelial function and thus control vascular tone.

Together with the elastic modulus of cells, AFM have also been applied to measure the viscoelastic properties of living cells [74, 75, 76], where the use of special probes has also been proposed [77].

Force vs distance measurements by AFM not only enable the evaluation of sample stiffness, but additionally allow manipulating individual biomolecules, making AFM an ideal tool for measuring the interactions between many active systems [78, 79]. Biomolecules such as nucleic acids and proteins are associated with a variety of cellular processes during which they undergo conformational changes. The use of AFM in single-molecule force spectroscopy has emerged as a powerful technique for investigating in real time the conformational dynamics of biomolecules [26, 80]. Figure. 32.17 shows a schematic representation of an AFM as a molecular force probe (MFP) where a DNA molecule is attached to the substrate on the piezoelectric stage and the cantilever tip. As the piezo scanner retracts in the vertical direction, the separation between the tip and the sample surface increases. The cantilever deflection is a measure of the force acting on the DNA. The extension of DNA is calculated from the distance between the AFM tip and the substrate [26, 81].

The development of the colloidal probe technique [83] furthermore enables single-cell force spectroscopy (SCFS) studies, which involve the immobilization of a single living cell onto the AFM cantilever and measuring the interaction forces between the probe and another cell or a substrate [82].

The process of gluing a cell to an AFM cantilever is however is time-consuming, limiting the throughput to few cells per day. A recent development, called FluidFM, offers single-cell force spectroscopy at a 10 times higher throughput and with a much easier manipulation of the cells of interest [84]. FluidFM uses AFM principles to manipulate living biological cells by stimulation through delivery of active agents directly from a solution. Overall, it offers a dramatic improvement, due to the possibility of reversibly immobilizing a cell to a FluidFM probe by suction, and subsequently releasing it with pressure.

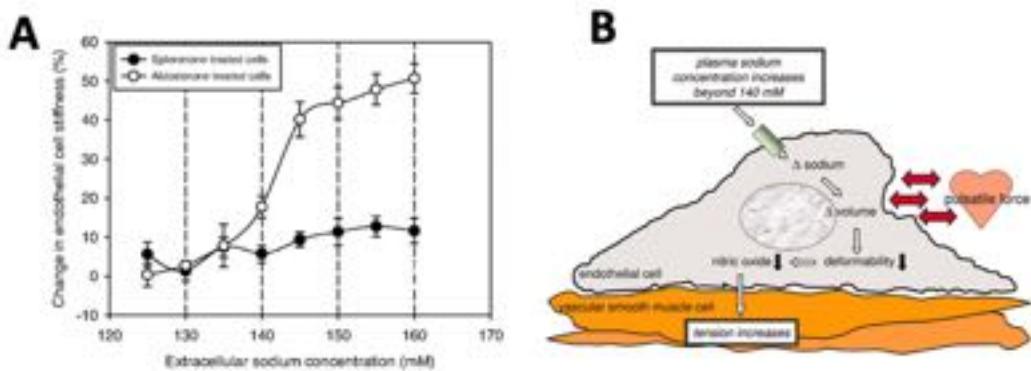


Figure 32.16: **Figure title missing.** [FIGURE NOT MENTIONED IN THE TEXT.] (A) Changes in endothelial cell stiffness measured by AFM in response to increasing sodium concentration measured in buffered electrolyte solution. The comparisons between the mean values for eplerenone-treated endothelium and aldosterone-treated endothelium, obtained at the same sodium concentrations, show the statistical significances ($P < 0.01$) at 140 mM sodium concentration and higher. (B) Schematic picture representing the role of extracellular sodium in the regulation of vascular tone. Sodium enters the endothelial cell through ENaC. After a transient increase in cell volume, the endothelium stiffens. Thus, vascular pulse pressure exerted by the working heart has a reduced impact on endothelial cell shape. The reduced rhythmic deformation of the endothelial cells decreases nitric oxide synthesis and release. As a result, vascular smooth muscle tone increases. Reproduced from Ref. [33].

Further, recent progress has improved the time resolution of AFM scanning units, from minutes to tens of milliseconds, meaning that it is now possible to watch single biomolecules in action in real time, with a technique called high-speed AFM (HS-AFM) [85]. HS-AFM is used to observe topographical changes of living cells and dynamic processes occurring on their surfaces [86]. Also, HS-AFM is used to observe dynamic processes occurring in many systems such as, synthetic polymer chains [87], detergents [88], and corrosion reactions at solid-liquid interfaces [89], nanobubble formation [90] and biomolecular processes [91]. HS-AFM required progress in realizing are ultrasmall cantilevers, fast actuators, fast amplitude detectors and an adaptive/dynamic feedback mechanism [36].

Imaging of dynamic molecular movement on living cells is now possible by using HS-AFM. Yamashita et al. studied the molecular dynamics on a living cell surface [37], where the successive AFM images of the outer cell surface in liquid medium are recorded. Fig. 32.21 shows still images of a successive AFM movie with a frame time of 0.5 s.

32.6 Outlook

The use of AFM to characterize various aspects of biologically active entities in liquid has already been greatly explored and will continue to be a key endeavor in the future, also thanks to numerous technical advantages, as the ones described above. However, the full potential of surface characterization using AFM for synthetic microswimmers and their interactions with the surfaces around them has so far received little exploitation. Starting from the topographical description of the surfaces of microswimmers and the correlation with their self-propulsion and the direct measurement of interactions with interfaces and with other microswimmers, we expect that nanoscale surface characterization methods will take on a more prominent role in active matter in the near future.

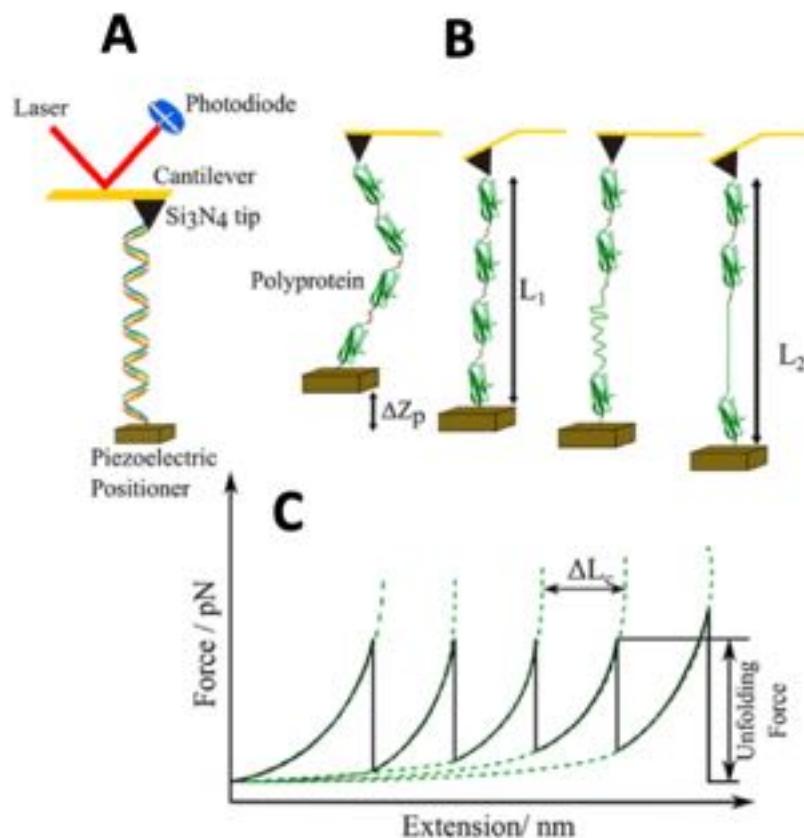


Figure 32.17: **Figure title missing.** (A) Schematic representation of an atomic force microscope as a molecular force probe (MFP). (B) AFM measurements of a polyprotein construct. The movement of the piezoelectric positioner is represented by ΔZ_p . Initially, the protein is in a relaxed state. Stretching this protein to near its folded contour length, L_1 , requires a force that is measured as a deflection of the cantilever. Stretching further increases the applied force, which triggers the unfolding of a domain, increasing the contour length of the protein and relaxing the cantilever back to its resting position. Further stretching removes the slack and brings the protein to its new contour length L_2 . (C) Characteristic force–extension sawtooth pattern curve resulting from stretching a polyprotein. Each sawtooth peak corresponds to the unfolding of one of the domains, while the last peak arises from the detachment of the molecule from the substrate or AFM tip. The amplitude of the sawtooth unfolding force peak measures the force at which the protein domain unfolds. Dotted lines correspond to fits of the worm-like chain model of polymer elasticity to the experimental data. The contour length increment, ΔL_c , measures the length increment upon protein unfolding. Reproduced from Ref. [26, 81].

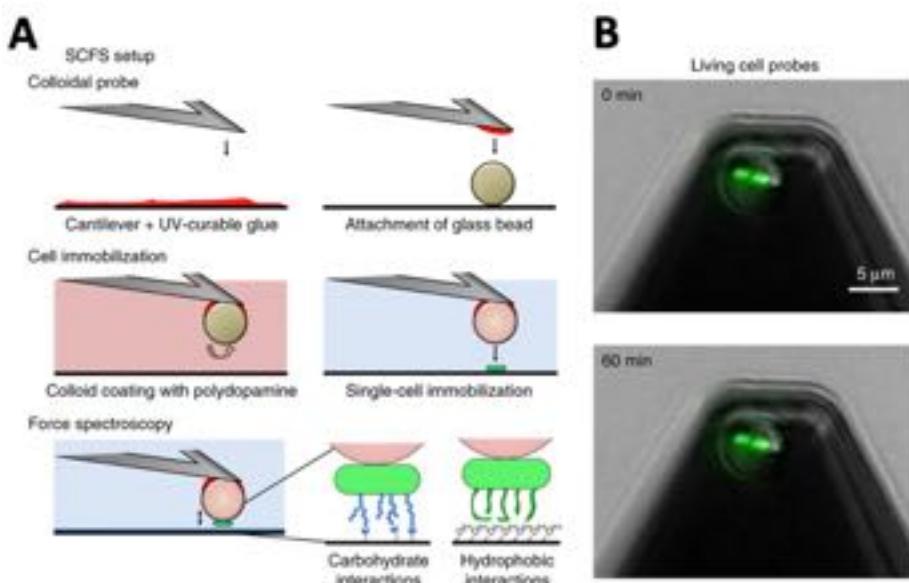


Figure 32.18: **Figure title missing.** [FIGURE NOT MENTIONED IN THE TEXT.] (A) The SCFS protocol involves the use of colloidal probe cantilevers combined with bioinspired polydopamine polymers, and it consists of three steps: preparation of the colloidal probe, controlled attachment of single cells and force-distance curve measurements. (B) Checking that the probed cells are still alive by using fluorescence stains: fluorescence images of bacterial cells labeled with the Baclight LIVE/DEAD stain, attached on polydopamine-coated cantilevers and imaged either immediately (0 min, top) or after 60 min of force measurements (bottom). Reproduced from Ref. [28, 82].

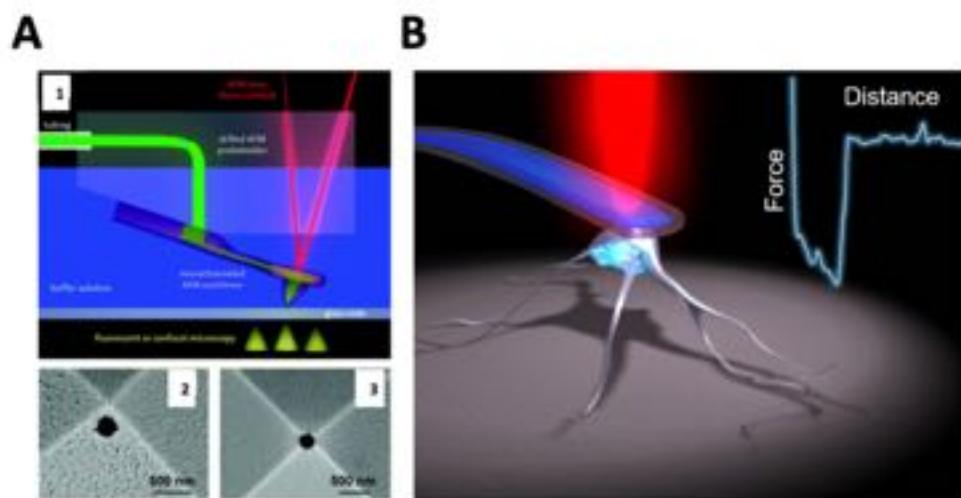


Figure 32.19: **Figure title missing.** [FIGURE NOT MENTIONED IN THE TEXT.] The fluidFM. (A) Diagram showing a microchanneled cantilever chip fixed to an AFM probe holder. The system was shown to be watertight up to 4 bar internal overpressure and overpressures of the order of 10 mbars are usually applied. The FluidFM can be operated in air or with the whole system (probe holder and chip) immersed in a liquid. (B) Scanning electron micrograph of the aperture beside the apex of the pyramidal AFM tip for the used as microfluidic pipette. Reproduced from Ref. [84]. B): Schematic of an experiment illustrating a cell being detached from a substrate by FluidFM. The inset shows an example of force-vs-distance curve which can be used to directly measure adhesion strength, energy, distance as well as the involved bio-chemical bonds between the cell and the substrate (adopted from Cytosurge AG, Copyright © 2020 all rights reserved).

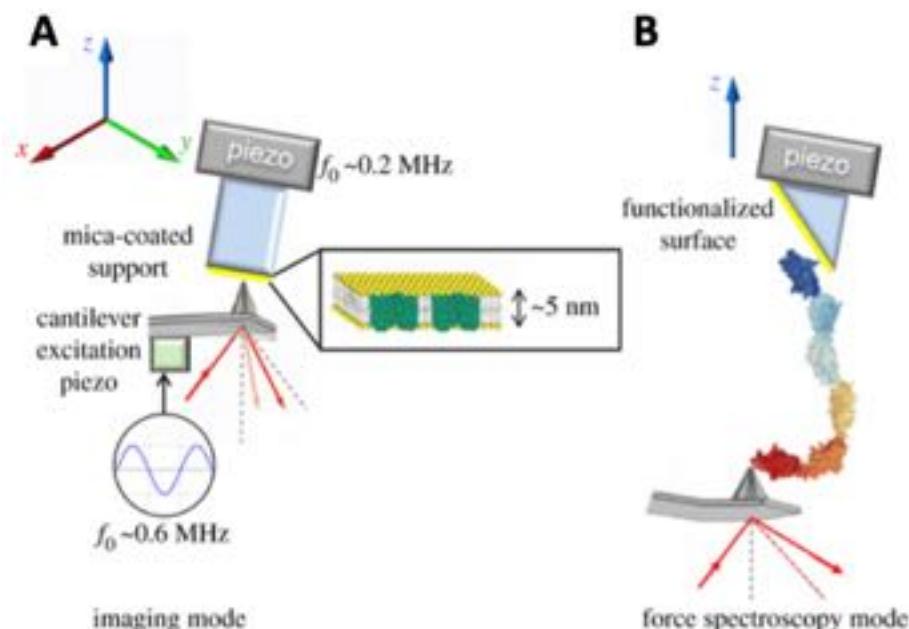


Figure 32.20: **Figure title missing. [FIGURE NOT MENTIONED IN THE TEXT.]** Schematic representation of HS-AFM systems for (A) imaging and (B) force spectroscopy mode. For imaging, the cantilever is excited through a small piezo element near its resonance frequency. For flat surfaces such as biomembranes (schematized in the inset) the sample is scanned in the horizontal plane (xy), the images are obtained by keeping the amplitude of oscillation constant by using the feedback loop. In the force spectroscopy mode, the Z piezo is approached onto the surface and retracted in the vertical direction. The samples are often tilted to further reduce the effective viscous drag coefficient of the cantilever near the surface. Functionalized surfaces and cantilever tips are frequently used to probe the mechanics of single molecules. Reproduced from Ref. [36].

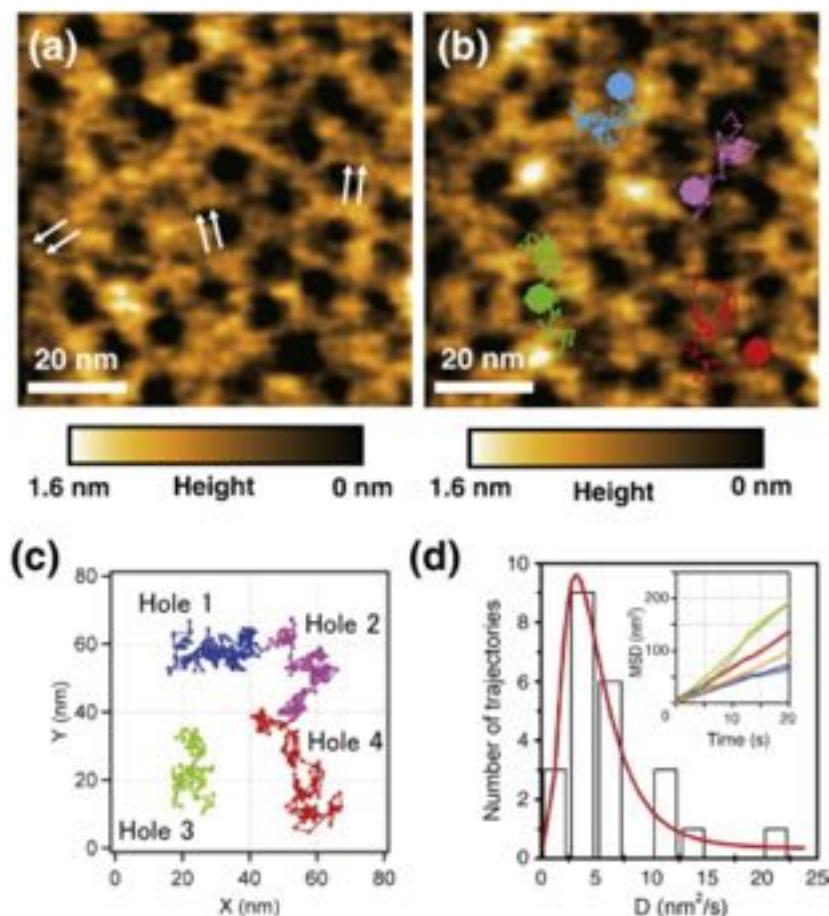


Figure 32.21: Observation of the molecular dynamics on a living cell surface by means of HS-AFM. (A) High-resolution still image of cell surface obtained from a continuous AFM movie. (B) AFM image showing trajectories of four net-like structures. These trajectories were drawn by tracing the movement of each hole. (C) Total trajectories of the two-dimensional diffusion for four net-like structures. The analyzed time spans were 82 s for holes 1, 2, and 4, and 47.5 s for hole 3. (D) Distribution of diffusion constants for individual trajectories. The data follow a log-normal distribution characterized by a mean $D = (3.2 \pm 0.4) \text{ nm}^2/\text{s}$ (continuous line). Inset shows mean square displacements of individual trajectories. AFM images were recorded at imaging rates of 0.5 s/frame with 200×200 pixels. Reproduced from Ref. [37].

32.7 Problems

For the following problem, you need to download and install the software Gwyddion <http://gwyddion.net> (make sure you download the most updated stable version) and the dataset from the following address: https://www.dropbox.com/sh/nxzipx1wyyb5y5g/AAC6-B-iB0NFR-bH1p25tn_6Xa?dl=0. we should move this to github or figshare, ok?

Problem 32.1: Raw image processing of a recorded AFM image with Gwyddion. Launch Gwyddion and open (File > Open) the image file named structure. The samples may have different level of tilt. To correct the tilt, try Data Process > Level > Plane Level or use the icon on Gwyddion's front panel. Note that even if the background has almost a uniform color, it is still slightly distorted. If the image appears in gray scale, you might not see the distortion easily. You can change the color scale by right clicking on the data scale on the right end of the image.

- a. Change the color scale of the image such that the image does not appear in grey scale. You can change the color scale by right clicking on the data scale on the right end of the image.
- b. Try to reduce the background distortion by going to Data Process > Level > Polynomial Background (or selecting its icon from the menu). Usually a polynomial degree of 2 or 3 in both the horizontal and vertical directions is enough. Also, in the Level menu you find Fix Zero, which shifts the lowest Z-value to zero. This option also show if you right click on the image or choose it from the front panel.
- c. Play around by increasing the polynomial degree to set how much the background is removed. Does a higher polynomial degree of background removal make sense?

Problem 32.2: Extracting the roughness values measuring feature sizes with Gwyddion. In Problem 32.1 you leveled the image. After leveling the image, in the Tools section of Gwyddion's front panel, select the Measure Distances icon.

- a. After leveling the image, in the Tools section of Gwyddion's front panel, select the Measure Distances icon. Click and drag on the image to create a line and measure the distance between adjacent points in both directions of repeating features. Make note of the distances, and then clear your lines.
- b. Now, in the Tools section of Gwyddion's front panel, select the Extract Profiles icon. Follow the same indications of point a and click and drag a line across multiple features of the image. You will see the cross-section of the line appear in the window that open when you open Extract Profiles.
- c. Choose Apply, the window that now opens has an icon in the lower left that will allow you to measure distances in the graph in the cross section by clicking in the plot (Method: Intersection). You can save the profile by using the right click option.
- d. Now can you say what is the height and period of the features in the structure sample? Also, do you obtain a better measurement by examining adjacent features or measuring across multiple ones? Why?
- e. Compare the simplicity and accuracy of the approaches above.

Problem 32.3: Measuring R_a and R_q with Gwyddion. In this problem, you will have to compare two AFM images representing the shiny side and the matte side of an aluminum foil.

- a. Open the images obtained for two sides of an aluminum foil, A11 and A12. Level the images following the procedure described in Problem 32.1.
- b. After leveling the images, click on the icon to calculate the roughness parameters. Draw a line across each image and see how the profile look like. The roughness values are shown on the left side.
- c. You can drag the line across different area of the two images and see how the roughness values changes. Save a section profile and the corresponding roughness values for the two images.
- d. The RMS value is the most commonly reported roughness parameter. Can you say now which one is the shiny and which one is the matte side of the aluminum foil?
- e. You can also obtain the roughness values for the entire image by clicking statistical quantities in the tools section: you can also select only an area of interest for roughness values in the image by holding and dragging a box inside the image. Compare the roughness values obtained from extracting the line profile and the entire image: which way is better for reporting the roughness of your sample?

Problem 32.4: Explore Gwyddion's options.

- a. Watch the series of tutorials on how to use Gwyddion <http://gwyddion.net/presentations/tutorials.php>.

b. At the same address, there are also images obtained for gold and silver surfaces: compare the roughness values. Also, visualize them using the 3D view of the data option.

Bibliography

- [1] H. Schillers, I. Medalsy, S. Hu, A. L. Slade, and J. E. Shaw. Peakforce tapping resolves individual microvilli on living cells. *Journal of Molecular Recognition*, 29:95–101, 2016.
- [2] E. Meyer, R. Bennewitz, and H. J. Hug. *Scanning Probe Microscopy, The Lab on a Tip*. Graduate Texts in Physics. Springer Nature, 2021.
- [3] V. D. Lyles, W. K. Serem, J.-J. Yu, and J. C. Garno. Surface science techniques. In *Surface Science Techniques*, Springer Series in Surface Sciences, pages 599–620. Surface Science Techniques, 2013.
- [4] G. Stan and S. W. King. Atomic force microscopy for nanoscale mechanical property characterization. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, 38:060801, 2020.
- [5] Y. F. Dufrêne, T. Ando, R. Garcia, D. Alsteens, D. Martinez-Martin, A. Engel, C. Gerber, and D. J. Müller. Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nature Nanotechnology*, 12:295–307, 2017.
- [6] C.-W. Yang, I.-S. Hwang, Y. F. Chen, C. S. Chang, and D. P. Tsai. Imaging of soft matter with tapping-mode atomic force microscopy and non-contact-mode atomic force microscopy. *Nanotechnology*, 18:084009, 2007.
- [7] M. Shibata, H. Watanabe, T. Uchihashi, T. Ando, and R. Yasuda. High-speed atomic force microscopy imaging of live mammalian cells. *Biophysics and Physicobiology*, 14:127–135, 2017.
- [8] N. Kodera, D. Yamamoto, R. Ishikawa, and T. Ando. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature*, 468:72–76, 2010.
- [9] S. J. Ebbens and D. A. Gregory. Catalytic janus colloids: Controlling trajectories of chemical microswimmers. *Accounts of Chemical Research*, 51:1931–1939, 2018.
- [10] M. N. Popescu, W. E. Uspal, and S. Dietrich. Self-diffusiophoresis of chemically active colloids. *The European Physical Journal Special Topics*, 225:2189–2206, 2016.
- [11] S. J. Ebbens and J. R. Howse. Direct observation of the direction of motion for spherical catalytic swimmers. *Langmuir*, 27:12293–12296, 2011.
- [12] S. Ebbens, M.-H. Tu, H. R. Howse, and R. Golestanian. Size dependence of the propulsion velocity for catalytic janus-sphere swimmers. *Physical Review E*, 85:020401, 2012.
- [13] J. Palacci, S. Sacanna, S.-H. Kim, G.-R. Yi, D. J. Pine, and P. M. Chaikin. Light-activated self-propelled colloids. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372:20130372, 2014.
- [14] X. Ma, A. C. Hortelão, T. Patiño, and S. Sánchez. Enzyme catalysis to power micro/nanomachines. *ACS Nano*, 10:9111–9122, 2016.
- [15] K. Dietrich, G. Volpe, M. N. Sulaiman, D. Renggli, I. Buttinoni, and L. Isa. Active atoms and interstitials in two-dimensional colloidal crystals. *Physical Review Letters*, 120:268004, 2018.
- [16] X. Wang, M. In, C. Blanc, A. Würger, M. Nobili, and A. Stocco. Janus colloids actively rotating on the surface of water. *Langmuir*, 33:13766–13773, 2017.
- [17] S. Ketzeltzi, J. de Graaf, R. P. Doherty, and D. J. Kraft. Slip length dependent propulsion speed of catalytic colloidal swimmers near walls. *Physical Review Letters*, 124:048002, 2020.

- [18] J. T. Parsons, A. R. Horwitz, and M. A. Schwartz. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. *Nature Reviews Molecular Cell Biology*, 11:633–643, 2010.
- [19] C. Berne, C. K. Ellison, A. Ducret, and Y. V. Brun. Bacterial adhesion at the single-cell level. *Nature Reviews Microbiology*, 16:616–627, 2018.
- [20] N. J. Hallab, K. J. Bundy, K. O'Connor, R. Clark, and R. L. Moses. Cell adhesion to biomaterials: correlations between surface charge, surface roughness, adsorbed protein, and cell morphology. *Journal of long-term effects of medical implants*, 5:209–31, 1995.
- [21] O. Guillaume-Gentil, E. Pothoff, D. Ossola, C. M. Franz, T. Zambelli, and J. A. Vorholt. Force-controlled manipulation of single cells: from AFM to FluidFM. *Trends in Biotechnology*, 32:381–388, 2014.
- [22] R. Afrin, U. S. Zohora, H. Uehara, T. Watanabe-Nakayama, and A. Ikai. Atomic force microscopy for cellular level manipulation: imaging intracellular structures and DNA delivery through a membrane hole. *Journal of Molecular Recognition*, 22:363–372, 2009.
- [23] K. Haase and A. E. Pelling. Investigating cell mechanics with atomic force microscopy. *Journal of The Royal Society Interface*, 12:20140970, 2015.
- [24] S. Kasas, G. Longo, and G. Dietler. Mechanical properties of biological specimens explored by atomic force microscopy. *Journal of Physics D: Applied Physics*, 46:133001, 2013.
- [25] J. Hu, S. Chen, D. Huang, Y. Zhang, S. Lü, and M. Long. Global mapping of live cell mechanical features using peakforce qnm AFM. *Biophysics Reports*, 6:9–18, 2020.
- [26] S. S. Mandal. Force spectroscopy on single molecules of life. *ACS Omega*, 5:11271–11278, 2020.
- [27] Z. N. Scholl and P. E. Marszalek. Nanoscale imaging, methods and protocols. In *Nanoscale Imaging*, volume 1814 of *Methods in Molecular Biology*, pages 35–47. Nanoscale Imaging, 2018.
- [28] A. Beaussart, S. El-Kirat-Chatel, P. Herman, D. Alsteens, J. Mahillon, P. Hols, and Y. F. Dufrêne. Single-cell force spectroscopy of probiotic bacteria. *Biophysical Journal*, 104:1886–1892, 2013.
- [29] H.-J. Butt, B. Cappella, and M. Kappl. Force measurements with the atomic force microscope: Technique, interpretation and applications. *Surface Science Reports*, 59:1–152, 2005.
- [30] Y. Ding, G.-K. Xu, and G.-F. Wang. On the determination of elastic moduli of cells by AFM based indentation. *Scientific Reports*, 7:45575, 2017.
- [31] L. Sirghi, J. Ponti, F. Broggi, and F. Rossi. Probing elasticity and adhesion of live cells by atomic force microscopy indentation. *European Biophysics Journal*, 37:935–945, 2008.
- [32] M. Plodinec, M. Loparic, C. A. Monnier, E. C. Obermann, R. Zanetti-Dallenbach, P. Oertle, J. T. Hyotyla, U. Aebi, M. Bentires-Alj, R. Y. H. Lim, and C.-A. Schoenenberger. The nanomechanical signature of breast cancer. *Nature Nanotechnology*, 7:757–765, 2012.
- [33] H. Oberleithner, C. Riethmüller, H. Schillers, G. A. MacGregor, H. E. de Wardener, and M. Hausberg. Plasma sodium stiffens vascular endothelium and reduces nitric oxide release. *Proceedings of the National Academy of Sciences*, 104:16281–16286, 2007.
- [34] S. E. Cross, Y.-S. Jin, J. Rao, and J. K. Gimzewski. Nanomechanical analysis of cells from cancer patients. *Nature Nanotechnology*, 2:780–783, 2007.
- [35] M. M. Yallapu, K. S. Katti, D. R. Katti, S. R. Mishra, S. Khan, M. Jaggi, and S. C. Chauhan. The roles of cellular nanomechanics in cancer. *Medicinal Research Reviews*, 35:198–223, 2015.
- [36] I. Casuso, L. Redondo-Morata, and F. Rico. Biological physics by high-speed atomic force microscopy. *Philosophical Transactions of the Royal Society A*, 378:20190604, 2020.

- [37] H. Yamashita, A. Taoka, T. Uchihashi, T. Asano, T. Ando, and Y. Fukumori. Single-molecule imaging on living bacterial cell surface by high-speed AFM. *Journal of Molecular Biology*, 422:300–309, 2012.
- [38] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel. Surface studies by scanning tunneling microscopy. *Physical Review Letters*, 49:57–61, 1982.
- [39] G. Binnig and H. Rohrer. Scanning tunneling microscopy. *Helvetica Physica Acta*, 5:726, 1982.
- [40] G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56:930–933, 1986.
- [41] J. Israelachvili. *Intermolecular & Surface Forces*. Academic Press, London, 2000.
- [42] C. Mathew Mate. *Tribology on the Small Scale: A Bottom Up Approach to Friction, Lubrication, and Wear*. Oxford University Press, Oxford, 2008.
- [43] P. Eaton and P. West. *Atomic Force Microscopy*. Oxford University Press, 2010.
- [44] G. Thomas, N. A. Burnham, T. A. Camesano, and Q. Wen. Measuring the mechanical properties of living cells using atomic force microscopy. *Journal of Visualized Experiments*, 76, 2013.
- [45] M. D. A. Norman, S. A. Ferreira, G. M. Jowett, L. Bozec, and E. Gentleman. Measuring the elastic modulus of soft culture surfaces and three-dimensional hydrogels using atomic force microscopy. *Nature Protocols*, 16:2418–2449, 2021.
- [46] R. Garcia. Nanomechanical mapping of soft materials with the atomic force microscope: methods, theory and applications. *Chemical Society Reviews*, 49:5850–5884, 2020.
- [47] D. Tranchida, S. Piccarolo, and M. Soliman. Nanoscale mechanical characterization of polymers by AFM nanoindentations: Critical approach to the elastic characterization. *Macromolecules*, 39:4547–4556, 2006.
- [48] R. Ferencz, J. Sanchez, B. Blümich, and W. Herrmann. AFM nanoindentation to determine Young's modulus for different EPDM elastomers. *Polymer Testing*, pages 425–432, 2012.
- [49] J. Zemla, J. Bobrowska, A. Kubiak, T. Zieliński, J. Pabijan, K. Pogoda, P. Bobrowski, and M. Lekka. Indenting soft samples (hydrogels and cells) with cantilevers possessing various shapes of probing tip. *European Biophysics Journal*, 49:485–495, 2020.
- [50] K. L. Johnson. *Contact Mechanics*. Cambridge University Press, Cambridge University Press, 1987.
- [51] K. L. Johnson, K. Kendall, and A. D. Roberts. Surface energy and the contact of elastic solids. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 324:301–313, 1971.
- [52] B. V. Derjaguin, V. M. Muller, and Yu. P. Toporov. Effect of contact deformations on the adhesion of particles. *Progress in Surface Science*, 45:131–143, 1994.
- [53] I. N. Sneddon. The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *International Journal of Engineering Science*, 3:47–57, 1965.
- [54] B. Pittenger, A. Slade, A. Berquand, P. Milani, A. Boudaoud, and O. Hamant. Toward quantitative nanomechanical measurements on live cells with PeakForce QPN, 2013. (Bruker Application Note #141).
- [55] S. N. Ramakrishna, L. Y. Clasohm, A. Rao, and N. D. Spencer. Controlling adhesion force by means of nanoscale surface roughness. *Langmuir*, 27:9972–9978, 2011.
- [56] S. N. Ramakrishna, P. C. Nalam, L. Y. Clasohm, and N. D. Spencer. Study of adhesion and friction properties on a nanoparticle gradient surface: Transition from JKR to DMT contact mechanics. *Langmuir*, 29:175–182, 2013.
- [57] C. M. Mate and R. W. Carpick. *Tribology on the Small Scale*. Oxford University Press, USA. Oxford University Press, USA, 2019.

- [58] E. Henderson, P. G. Haydon, and D. S. Sakaguchi. Actin filament dynamics in living glial cells imaged by atomic force microscopy. *Science*, 257:1944–1946, 1992.
- [59] W. Häberle, J. K. H. Hörber, F. Ohnesorge, D. P. E. Smith, and G. Binnig. In situ investigations of single living cells infected by viruses. *Ultramicroscopy*, 42:1161–1167, 1992.
- [60] C. A. Schoenenberger and J. H. Hoh. Slow cellular dynamics in MDCK and R5 cells monitored by time-lapse atomic force microscopy. *Biophysical Journal*, 67:929–936, 1994.
- [61] P. D. Antonio, M. Lasalvia, G. Perna, and V. Capozzi. Scale-independent roughness value of cell membranes studied by means of AFM technique. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818:3141–3148, 2012.
- [62] Y. Wang, C. Xu, N. Jiang, L. Zheng, J. Zeng, C. Qiu, H. Yang, and S. Xie. Quantitative analysis of the cell-surface roughness and viscoelasticity for breast cancer cells discrimination using atomic force microscopy. *Scanning*, 38:558–563, 2016.
- [63] M. Girasole, G. Pompeo, A. Cricenti, A. Congiu-Castellano, F. Andreola, A. Serafino, B. H. Frazer, G. Boumis, and G. Amiconi. Roughness of the plasma membrane as an independent morphological parameter to study RBCs: A quantitative atomic force microscopy investigation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768:1268–1276, 2007.
- [64] K. Keren, Z. Pincus, G. M. Allen, E. L. Barnhart, G. Marriott, A. Mogilner, and J. A. Theriot. Mechanism of shape determination in motile cells. *Nature*, 453:475–480, 2008.
- [65] E. Reister, T. Bihr, U. Seifert, and A.-S. Smith. Two intertwined facets of adherent membranes: membrane roughness and correlations between ligand-receptors bonds. *New Journal of Physics*, 13:025003, 2011.
- [66] Y. Ohta, H. Okamoto, M. Kanno, and T. Okuda. Atomic force microscopic observation of mechanically traumatized erythrocytes. *Artificial Organs*, 26:10–17, 2002.
- [67] A. J. da Silva and O. Teschke. Dynamics of the antimicrobial peptide PGLa action on Escherichia coli monitored by atomic force microscopy. *World Journal of Microbiology and Biotechnology*, 21:1103–1110, 2005.
- [68] K. Kliche, M. Kuhn, U. Hillebrand, Y. Ludwig, C. Stock, and H. Oberleithner. Direct aldosterone action on mouse cardiomyocytes detected with atomic force microscopy. *Cellular Physiology and Biochemistry*, 18:265–274, 2006.
- [69] M. Girasole, G. Pompeo, A. Cricenti, G. Longo, G. Boumis, A. Bellelli, and S. Amiconi. The how, when, and why of the aging signals appearing on the human erythrocyte membrane: an atomic force microscopy study of surface roughness. *Nanomedicine: Nanotechnology, Biology and Medicine*, 6:760–768, 2010.
- [70] C. Roduit, S. Sekatski, G. Dietler, S. Catsicas, F. Lafont, and S. Kasas. Stiffness tomography by atomic force microscopy. *Biophysical Journal*, 97:674–677, 2009.
- [71] D. E. Discher, P. Janmey, and Y.-I. Wang. Tissue cells feel and respond to the stiffness of their substrate. *Science*, 310:1139–1143, 2005.
- [72] J. Guck, S. Schinkinger, B. Lincoln, F. Wottawah, S. Ebert, M. Romeyke, D. Lenz, H. M. Erickson, R. Ananthakrishnan, D. Mitchell, J. Käs, S. Ulwick, and C. Bilby. Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence. *Biophysical Journal*, 88:3689–3698, 2005.
- [73] M. J. Rosenbluth, W. A. Lam, and D. A. Fletcher. Force microscopy of nonadherent cells: A comparison of leukemia cell deformability. *Biophysical Journal*, 90:2994–3003, 2006.
- [74] Y. M. Efremov, W.-H. Wang, S. D. Hardy, R. L. Geahlen, and A. Raman. Measuring nanoscale viscoelastic parameters of cells directly from AFM force-displacement curves. *Scientific Reports*, 7:1541, 2017.

- [75] Y. H. Chim, L. M. Mason, N. Rath, M. F. Olson, M. Tassieri, and H. Yin. A one-step procedure to probe the viscoelastic properties of cells by atomic force microscopy. *Scientific Reports*, 8:14462, 2018.
- [76] A. Cartagena and A. Raman. Local viscoelastic properties of live cells investigated using dynamic and quasi-static atomic force microscopy methods. *Biophysical Journal*, 106:1033–1043, 2014.
- [77] D. Guan, E. Charlaix, R. Z. Qi, and P. Tong. Noncontact viscoelastic imaging of living cells using a long-needle atomic force microscope with dual-frequency modulation. *Physical Review Applied*, 8:044010, 2017.
- [78] W. F Heinz and J. H. Hoh. Spatially resolved force spectroscopy of biological surfaces using the atomic force microscope. *Trends in Biotechnology*, 17:143–150, 1999.
- [79] Y. F. Dufrêne, D. Martínez-Martín, I. Medalsy, D. Alsteens, and D. J. Müller. Multiparametric imaging of biological systems by force-distance curve-based AFM. *Nature Methods*, 10:847–854, 2013.
- [80] L. Wang, Y. Qian, Y. Sun, B. Liu, and G. Wei. Single-molecule force spectroscopy: A facile technique for studying the interactions between biomolecules and materials interfaces. *Reviews in Analytical Chemistry*, 39:116–129, 2020.
- [81] T. Hoffmann and L. Dougan. Single molecule force spectroscopy using polyproteins. *Chemical Society Reviews*, 41:4781–4796, 2012.
- [82] A. Beaussart, S. El-Kirat-Chatel, R. M. A Sullan, D. Alsteens, P. Herman, S. Derclaye, and Y. F. Dufrêne. Quantifying the forces guiding microbial cell adhesion using single-cell force spectroscopy. *Nature Protocols*, 9:1049–1055, 2014.
- [83] W. A. Ducker, T. J. Senden, and R. M. Pashley. Direct measurement of colloidal forces using an atomic force microscope. *Nature*, 353:239–241, 1991.
- [84] A. Meister, M. Gabi, P. Behr, P. Studer, J. Vörös, P. Niedermann, J. Bitterli, J. Polesel-Maris, M. Liley, H. Heinzelmann, and T. Zambelli. Fluidfm: Combining atomic force microscopy and nanofluidics in a universal liquid delivery system for single cell applications and beyond. *Nano Letters*, 9:2501–2507, 2009.
- [85] T. Ando. High-speed atomic force microscopy coming of age. *Nanotechnology*, 23:062001, 2012.
- [86] T. Uchihashi and S. Scheuring. Applications of high-speed atomic force microscopy to real-time visualization of dynamic biomolecular processes. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1862:229–240, 2018.
- [87] K.-I. Shinohara and Y. Makida. Direct observation of dynamic interaction between a functional group in a single sbr chain and an inorganic matter surface. *Scientific Reports*, 8:13982, 2018.
- [88] S. Inoue, T. Uchihashi, D. Yamamoto, and T. Ando. Direct observation of surfactant aggregate behavior on a mica surface using high-speed atomic force microscopy. *Chemical Communications*, 47:4974–4976, 2011.
- [89] J. J. Santillan and T. Itani. Dissolution characteristics of euv resist by high speed AFM. *Journal of Photopolymer Science and Technology*, 25:95–100, 2012.
- [90] H.-S. Liao, C.-W. Yang, H.-C. Ko, E.-T. Hwu, and I.-S. Hwang. Imaging initial formation processes of nanobubbles at the graphite-water interface through high-speed atomic force microscopy. *Applied Surface Science*, 434:913–917, 2018.
- [91] T. Ando, T. Uchihashi, and S. Scheuring. Filming biomolecular processes by high-speed atomic force microscopy. *Chemical Reviews*, 114:3120–3188, 2014.

Chapter 33

Spectroscopy

PIETRO GIUSEPPE GUCCIARDI, ANTONINO FOTI

[The Chapter reads well but the topic is an outlier and doesn't fit with the rest of the book. I suggest either making it more relevant to active matter or merging a much shorter version 1-3 pages with the OT Chapter (e.g. Section 2.3).] Optical spectroscopy permits the study of the energy levels of physical systems using light. It can provide information on sample's chemical composition, temperature, as well as optical and structural properties. It can be performed in a wide portion of the electromagnetic spectrum, allowing one to probe different physical processes occurring when light and matter interact, such as light absorption, emission, and scattering. *Ultraviolet (UV) and visible spectroscopy* generally provides information on the electronic transitions in atomic, molecular and solid systems. *Infrared (IR) spectroscopy* is used to study the atomic vibrations inside molecules and crystals. *Microwave spectroscopy* can be used to observe rotational transitions of molecules. *Extinction/absorption spectroscopy* allows one to get information on the sample's shape, size, electronic and roto-vibrational transitions. Different chemical substances absorb radiation at different wavelengths, giving a specific and unique fingerprint of the object analyzed. In a complementary way, emission spectroscopy (e.g., fluorescence and phosphorescence) is based on the analysis of the spectral lines or bands emitted from an object when it decays from an excited state into its ground state. Photons emitted at different frequencies carry

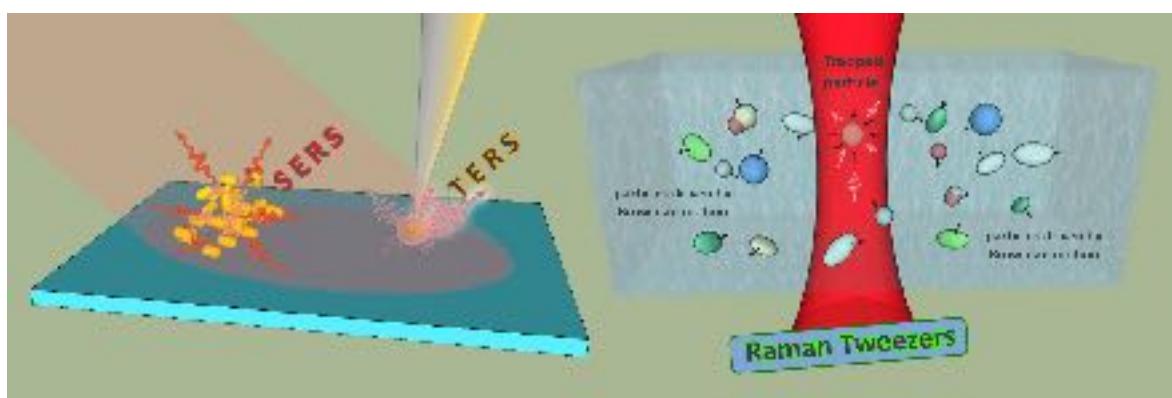


Figure 33.1: **SERS/TERS spectroscopy and Raman tweezers.** Raman scattering of molecules in close proximity to plasmonic nanoparticles can experience a giant amplification when the excitation beam is resonant with the localized surface plasmon of the metal nanostructures, which can be either aggregated nanoparticles as it is common in SERS (left) or individual nano-tips like in TERS (center). Apart from enhancing the Raman signal by several orders of magnitude, this latter approach allows also to push the spatial resolution of the optical imaging at the nanoscale. Finally, Raman Tweezers (right) enables the analysis of nano-objects in liquid media and at the single particle level. In fact, it is able to trap nanoparticles in the laser focus region and therefore to limit their Brownian diffusion, allowing to acquire enough signal for the Raman analysis.

information on the energy levels and radiative relaxation process.

Among the optical spectroscopies, IR and Raman spectroscopies (Fig. 33.1) are the most widely used techniques to study the vibrational properties of molecules and solids. Raman spectroscopy, in particular, is becoming an increasingly used analytical tool with applications spanning from pharmacy to security, health, food safety, and environment. When coupled with microscopy, Raman spectroscopy enables chemical mapping. With the development of solid state lasers and CCD cameras, the required equipment has become much more sensitive and even portable, with several models of handheld spectrometers present in the market for easy deployment in field measurements campaigns. Raman spectroscopy, however, suffers several limitations, intrinsic to the Raman scattering process and the way it is implemented, such as the low signal levels (up to 16 orders of magnitude smaller than fluorescence), the diffraction-limited spatial resolution (250 nm), and the difficulties to detect sparse particulate dispersed in liquid media (e.g., micro-plastics in water).

The main goal of this Chapter is to illustrate some of the most promising research trends in Raman spectroscopy aimed at overcoming the above mentioned limitations. The Chapter is organized in 4 sections. In Section 33.1 we introduce the fundamentals of vibrational IR and Raman spectroscopies; in Section 33.2 we describe Surface-Enhanced Raman Spectroscopy (SERS) as a way to enhance the Raman signal levels by several orders of magnitude; in Section 33.3 we show the combination of Raman spectroscopy and optical trapping (Raman Tweezers, RT) to analyze individual particles dispersed in liquids; finally, in Section 33.4 we introduce tip-enhanced Raman spectroscopy (TERS) as a tool to achieve sub-diffraction, nanometer-scale spatial resolution in chemical imaging.

33.1 IR and Raman spectroscopy

IR and Raman spectroscopy are powerful techniques used to study molecular vibrations and gain insights into the structural and chemical properties of molecules. Molecules, composed of atoms connected by chemical bonds, exhibit vibrations due to the thermal energy that causes atomic displacements. These vibrations can be approximated as the motion of masses connected by springs, where each atom in a molecule moves around its equilibrium position along its three Cartesian coordinates. For a molecule with N atoms, there are $3N$ degrees of freedom, including translational, rotational, and vibrational modes. The vibrational modes, specific to the molecule's structure, create a unique vibrational fingerprint.

Vibrational transitions occur within the IR spectrum, spanning wavelengths from 800 nm to 1 mm, and are typically expressed in wavenumbers (cm^{-1}). Fundamental vibrational modes fall primarily within the mid-IR (MIR) region, which is rich in molecular information. The IR absorption process involves the interaction of the electromagnetic field with the molecular dipole moment, resulting in specific absorption frequencies corresponding to molecular vibrations.

Raman spectroscopy, on the other hand, uses monochromatic light to excite the sample and measures the inelastic scattering of photons. The energy difference between incident and scattered photons provides information about the vibrational states. Unlike IR spectroscopy, Raman spectroscopy is sensitive to changes in the molecular polarizability during vibrations. This complementary nature of IR and Raman spectroscopy allows for a comprehensive analysis of molecular vibrations and structural characteristics.

33.1.1 Molecular vibrations

Molecules are made of atoms connected by chemical bonds [1]. The length of the chemical bonds is not fixed, but changes with time as a consequence of thermal energy. Hence, atoms in molecules do collectively vibrate. In a first approximation, atomic vibrations in a molecule can be modeled as the motion of point-like masses connected by springs. Each atom undergoes translational displacement around its equilibrium position (x_0, y_0, z_0) , along its three degrees of freedom in the x , y , z Cartesian coordinates. A molecule composed by N atoms will have $3N$ degrees of freedom: 3 degrees of freedom belonging to the translation of the center of mass, the remaining $3N - 3$ degrees of freedom accounting for rotations and vibrations. The number of vibrational degrees of freedom depends on the structure of the molecule. A non-linear polyatomic molecule has three rotational degrees of freedom, leaving $3N - 6$ degrees of freedom for interatomic vibrations, i.e., $3N - 6$ vibrational modes. Linear molecules have only two degrees of rotational freedom, leaving $3N - 5$ vibrational modes. Every vibrational mode exhibits its own pattern of atomic displacements around the equilibrium position $(x_0 \pm \Delta x, y_0 \pm \Delta y, z_0 \pm \Delta z)$. In these so-called *normal modes* all atoms of a molecule move collectively at the

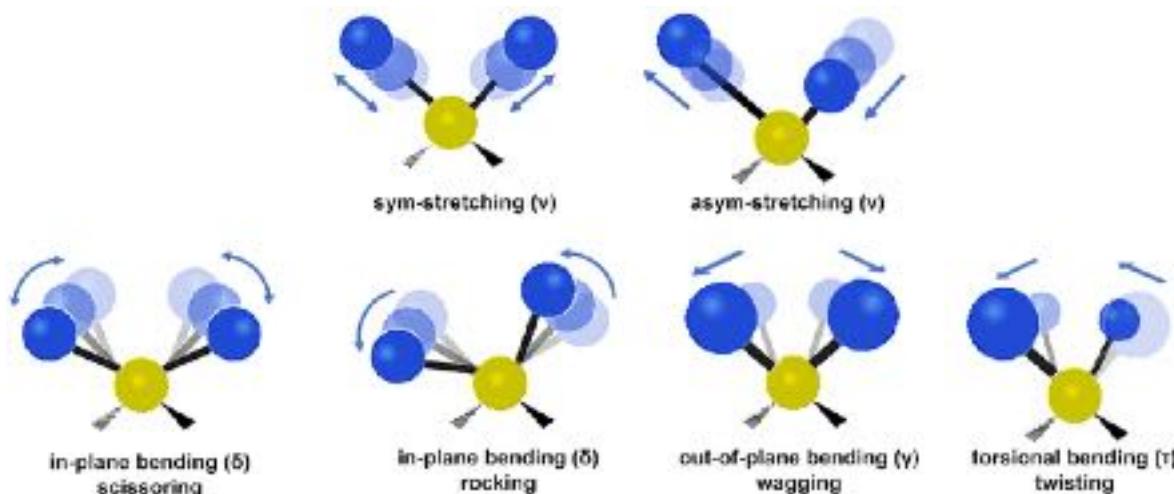


Figure 33.2: **Molecular vibration modes.** Stretching vibrations (top row) and bending vibrations (bottom row).

same frequency and in phase around the equilibrium positions, while the molecular center of mass remains unchanged. Vibrations cause a distortion of the molecular structure (Fig. 33.2). Stretching vibrations alter the bond lengths that elongate and contract periodically, in a symmetric or asymmetric fashion. In valence-angle bending vibrations, the bond angle of the molecule changes. In out-of-plane bending vibrations, the angle between a bond and a molecular plane defined by two other bonds varies. In torsional vibrations, the dihedral angle of a four-atomic fragment of the molecule oscillates.

The determination of the vibrational modes in a molecule is carried out based on its symmetry properties, using group theory, irreducible representations and character tables [2]. In the harmonic oscillator picture, the normal vibrations do not interfere with each other (orthogonality of the vibrational eigenvectors) and the energies associated to different modes are generally distinct. The ensemble (spectrum) of frequencies (or energies) associated to the vibrational normal modes is generally specific of the molecule and vary from molecule to molecule, being related to its chemical structure and conformation, like a fingerprint (so-called *vibrational fingerprint*). While the vibrational modes of simple molecules like water (3 atoms, non-linear molecule) only consist of 3 characteristic energies, polymers feature between 10 and 10² vibrational modes, an average protein has circa 10⁴ normal modes and this number ramps up to 10⁹ for large macromolecules such as *E. coli* DNA.

33.1.2 Vibrational transitions and spectroscopies

The most simple theoretical picture to describe molecular vibrations is a model harmonic oscillator with a quadratic energy potential (Fig. 33.3a). For a two-atom oscillator with nuclei masses m_1 and m_2 connected by a bond with force constant k , a characteristic vibrational frequency exists, that is $\nu_{\text{vib}} = (2\pi)^{-1} \sqrt{k/m_r}$, where $m_r = (m_1^{-1} + m_2^{-1})^{-1}$ is the reduced mass of the system. The stretching of a molecular bond meets a stronger force constant than a bending movement. Therefore, the frequency of stretching vibrations is typically three times larger than that of the bending vibrations. The frequency is also dependent on the mass of the vibrating atoms. The smaller the mass, the larger the vibration frequency. Vibrations involving hydrogen atoms, for example, have the highest frequencies.

In a quantum mechanical picture, the harmonic oscillator has a spectrum of discrete, equally spaced, energy levels $E = (n + 1/2) \hbar \nu_{\text{vib}}$ where $n = 0, 1, 2, \dots$ is the vibrational quantum number, ν_{vib} the frequency and $\hbar = 6.63 \times 10^{-34} \text{ m}^2 \text{kg s}^{-1}$ the Planck constant. Electromagnetic radiation can induce transitions among the energy levels. The selection rule for the harmonic oscillator ($\Delta n = \pm 1$), only permits transitions between adjacent levels, characterized by an energy jump $\Delta E = \hbar \nu_{\text{vib}}$. Real molecules deviate from the ideal harmonic behavior. Anharmonicity is caused by cubic and higher order terms in the potential energy. The anharmonic terms describe the fact that the vibrational energy levels are no longer equally spaced and the spacing decreases with increasing energy. Anharmonic terms, moreover, lift the selection rules of the electromagnetic transitions and

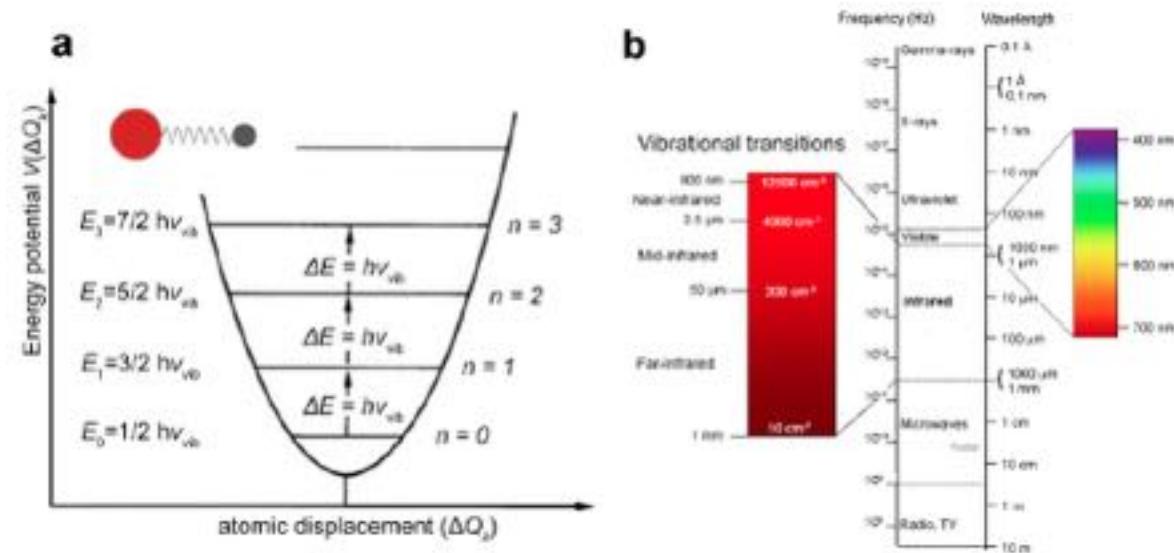


Figure 33.3: Harmonic oscillator model for molecular vibration. (a) Harmonic vibrational potential and transition energies. (b) Frequency, wavelength and energy scales of vibrational transitions in the electromagnetic spectral range.

couple the normal modes. Transitions over several levels (overtones, $\Delta n = \pm 2, \pm 3, \dots$) as well as simultaneous transitions of several normal modes (combination transitions) occur, although they have much lower transition probability. Energy transfer may occur from one excited normal mode to another.

Vibrational transitions occur in the IR part of the electromagnetic spectrum (Fig. 33.3b), at wavelengths between 800 nm (near-IR) and 1 mm (far-IR), corresponding to energies comprised between 1.24 meV to 1.55 eV. Vibrational energies are more commonly expressed in cm^{-1} units, where E (in cm^{-1}) = $10^7 / \lambda$ (in nm), resulting in a range between 10 and 12500 cm^{-1} . Low energy vibrations in the far-IR (FIR, $10 - 200 \text{ cm}^{-1}$), or terahertz region, are generally due to collective oscillations in crystalline or amorphous solids. The mid-IR range (MIR, $200 - 4000 \text{ cm}^{-1}$) is where most of the fundamental vibrational transitions of crystals and molecules occur, and is by far the most interesting and rich of information, in particular the region up to 1700 cm^{-1} (also called the *fingerprint region*). Finally, in the near-IR (NIR, $4000 - 12500 \text{ cm}^{-1}$) we find combinations and overtones modes, whose energies are multiples and/or sums of the fundamental transitions' ones.

Fundamental transitions proceed from the vibrational ground state level ($n = 0$) to the excited levels because oscillators are in the ground state at room temperature ($k_B T \simeq 25 \text{ meV}$). Hot transitions starting from an excited energy level, may occur at higher temperatures, due to increasing level population, dictated by Boltzmann's statistics. Anharmonicities cause hot transitions to have lower energy than the corresponding fundamental ones.

Vibrational spectroscopy probes the transitions between normal modes using light [3]. The main techniques are IR and Raman spectroscopy. In IR spectroscopy, the sample is irradiated with broadband polychromatic light in the IR. Whenever the frequency of the incident light (ν_{ph}) coincides with the one of a vibrational mode (ν_{vib}), photons are absorbed, inducing a transition from the ground state to an excited level separated by a quantum of energy $\Delta E_{\text{vib}} = \hbar\nu_{\text{vib}}$. In the IR absorption process the electromagnetic field couples to the molecular dipole moment. IR transitions are related to a change of the molecular dipole moment induced by the atomic vibrations and, as a consequence (selection rule in IR transitions), an IR absorption transition is allowed only if accompanied by a net change of the net dipole moment when a molecule passes from the initial vibrational state to the final one. In IR spectra the transmitted intensity is plotted against the wavenumber (Fig. 33.4, top line), which is the inverse of the wavelength expressed in cm^{-1} . Dips in the spectrum indicate the occurrence of a vibrational transition at a wavenumber proportional to its energy (frequency).

Raman spectroscopy [4] uses monochromatic light in the UV, visible or near-IR, to excite the sample. Photons that are inelastically scattered by the molecule loose (Stokes scattering) or gain (anti-Stokes scattering)

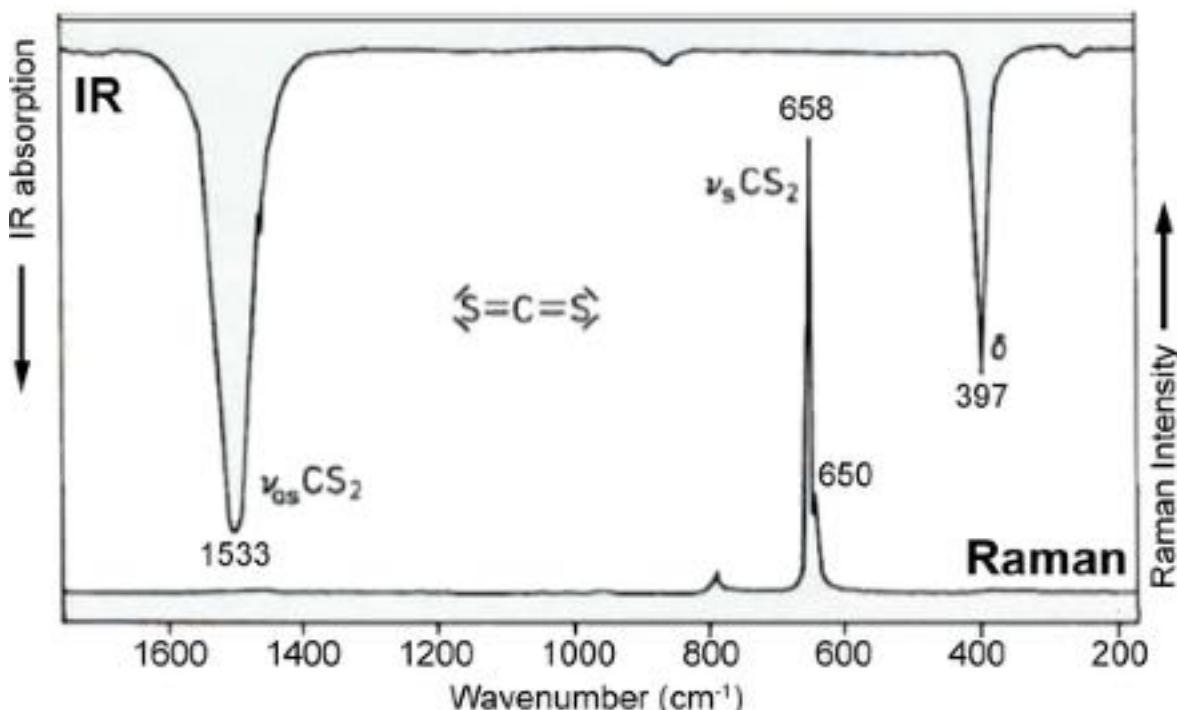


Figure 33.4: **IR and Raman spectra of CS_2 molecules.** Reproduced from Ref. [3].

tiny amounts of energy corresponding to the energy difference between the vibrational states, ΔE_{vib} . Raman is a two-photons transition in which the electromagnetic field first induces an electric dipole in the molecule, through the molecular polarizability α , and then it interacts with the induced dipole in a scattering process. The difference between the energy of the incident photons ($h\nu_{\text{ph}}$) and that of the scattered ones ($h\nu_S$ for Stokes and $h\nu_{\text{AS}}$ for the anti-Stokes), $\Delta E_{\text{vib}} = |h\nu_{\text{ph}} - h\nu_{S/\text{AS}}|$, contains information on the vibrational levels separation. Laser sources are nowadays election tools for Raman spectroscopy due to their extreme monochromaticity, directionality and brilliance. In the Raman spectra the scattered light intensity is plotted against the wavenumber (Fig. 33.4, bottom line), or Raman shift, that is the energy difference between the incident and the scattered photons expressed in cm^{-1} . The presence of sharp peaks in the spectrum highlights the presence of the individual vibrational modes. While IR spectroscopy is sensitive to the molecular dipole moment changes, Raman scattering probes the molecular polarizability (*vide infra*), that is the dipolar moment induced by the radiation itself. The selection rule for Raman scattering requires, therefore, that the polarizability changes during the transition. For highly symmetric molecules, the different selection rules for Raman and IR imply that both techniques give complementary information: strong IR absorbing vibrations are often weak Raman scatterers, and vice versa (Fig. 33.4).

Example 33.1: Selection rules. Explain why in CS_2 , shown in Fig. 33.4, the antisymmetric stretching and the bending are only IR active and the symmetric stretching is only Raman active.

Vibrational spectra encode information on the energy of the vibrations, transition oscillator strength and symmetry of the molecules, the number of molecules; the width of the peaks is in turn are related to electronic properties (e.g., metallic or semiconducting nanostructures), structural heterogeneity and defects, interaction with the environment. In the most general terms vibrational spectroscopy can be used to gain insight on:

- The chemical structure and conformation of a compound.
- The presence of specific compounds in an unknown sample.
- The properties of the chemical bonds and interaction of a compound with the environment.
- The electronic structure, presence of strain or stress in solids.
- The occurrence of new species and changes of concentrations during chemical reactions.

33.2 Surface enhanced Raman spectroscopy

Surface enhanced Raman scattering (SERS) [5, 6] is a surface-assisted spectroscopy in which molecules are probed after deposition on nano-textured surfaces. Signal amplifications in the 10^4 to 10^{10} range [7] are obtained as a result of the surface–molecule interaction. This huge enhancement can tailor the molecular sensitivity of Raman spectroscopy down to the atto-molar range [8] — and even reach the single-molecule level [9].

Two conceptually distinct mechanisms are believed to be at the origin of the SERS amplification: (a) an *electromagnetic enhancement*, that amplifies the electromagnetic fields, both incident and re-radiated [10], and a *chemical enhancement* increasing of the process cross-section [11, 12].

33.2.1 Electromagnetic enhancement

The electromagnetic enhancement [13, 14], which accounts for Raman signal amplifications of up to 10^{10} , results from the excitation of localized surface plasmons (LSP) within the metal nanoparticles (MNPs) which serve as SERS substrates on which molecules are absorbed. MNPs are typically much smaller in size than the wavelength λ_{ph} of the excitation light, ranging from few tens to few hundreds of nanometers. MNPs act as polarizable dipoles in which the collective oscillation of electrons in the conduction band produce fields localized at the surface, in the so called *hot-spots* (Fig. 33.5). The resonant excitation of the such oscillations, the localized surface plasmons resonances (LSPR), enhances greatly the electromagnetic fields at the hot spots. In SERS the MNPs provide a twofold amplification of the fields [10]. If we call $\mathbf{E}_0(\mathbf{r}_m, \omega_{ph})$ the incident field at frequency ω_{ph} at the molecule position \mathbf{r}_m , as a first process, this will be amplified by the LSPR, giving rise to a local field $\mathbf{E}_{loc}(\mathbf{r}_m, \omega_{ph})$ equal to:

$$\mathbf{E}_{loc}(\mathbf{r}_m, \omega_{ph}) = \Gamma_{exc}(\mathbf{r}_m, \omega_{ph}) \cdot \mathbf{E}_0(\mathbf{r}_m, \omega_{ph}), \quad (33.1)$$

where $\Gamma_{exc}(\mathbf{r}_m, \omega_{ph})$ is a tensor that describes the MNP-orientation-dependent enhancement factor of the incident field. For isotropic structures Γ_{exc} is a scalar. The resulting power enhancement factor experienced by the molecule will be $M_{exc}(\mathbf{r}_m, \omega_{ph}) = |\mathbf{E}_{loc}(\mathbf{r}_m, \omega_{ph})|^2 / |\mathbf{E}_0(\mathbf{r}_m, \omega_{ph})|^2$. The local field acting on the molecule, will induce a Raman dipole $\mathbf{P}(\mathbf{r}_m, \omega_r) = \alpha(\omega_r, \omega_{ph}) \cdot \mathbf{E}_{loc}(\mathbf{r}_m, \omega_{ph})$ oscillating at the Stokes frequency $\omega_r = \omega_{ph} \pm \omega_{vib}$, where $\alpha(\omega_r, \omega_{ph})$ is the polarizability tensor of the molecule. The radiation characteristics of the oscillating dipole $\mathbf{P}(\mathbf{r}_m, \omega_r)$ will be significantly affected by the dielectric properties of the MNP on which it is deposited and, in particular, by the LSPRs. Expressing the SERS enhancement as a two-step process leads to conceptual difficulties related to understanding where the power comes from for the second

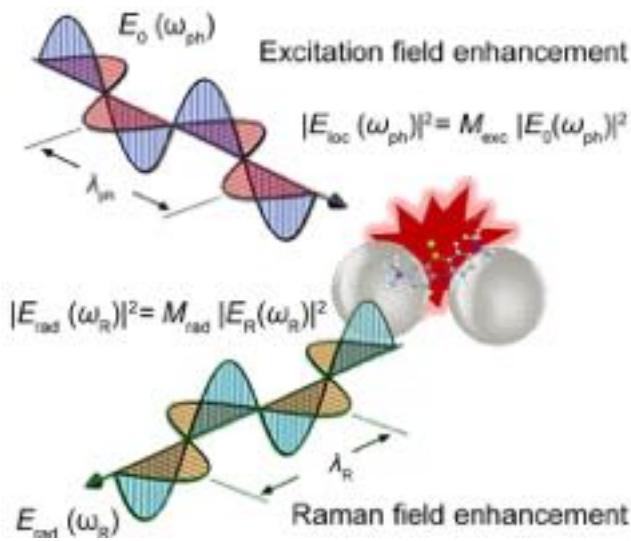


Figure 33.5: **Electromagnetic enhancement.** Schematic of the two-step SERS enhancement mechanism: local excitation field enhancement and re-radiation enhancement.

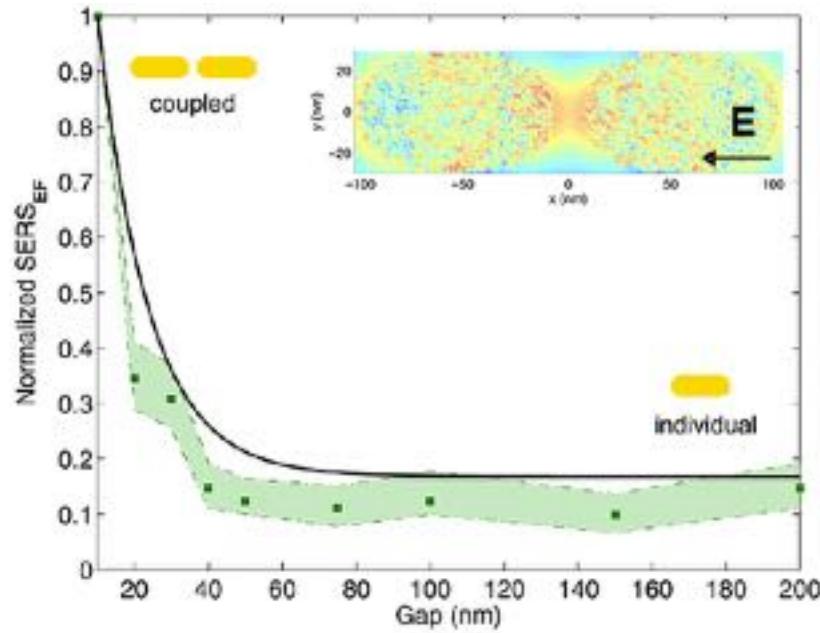


Figure 33.6: **Comparison of experimental and theoretical SERS enhancement.** Experimental (dots) and theoretical (line) values of the SERS enhancement measured on coupled nano-rods (100×60 nm) when the gap is decreased. The inset shows the map of the enhanced field for small gaps. Reproduced from Ref. [15].

enhancement step. Solving the problem as a coupled electromagnetic problem of a radiating dipole in the presence of a plasmonic nanoparticle in the presence of an oscillating electromagnetic field gives the correct result. The Raman radiation field $\mathbf{E}_{\text{rad}}(\mathbf{R}, \omega_r)$ at the observer coordinate \mathbf{R} in the far field is composed by the field emitted by the molecular oscillating dipole and by the dipole re-radiation field. The total power at the Raman-scattered frequency is dissipated in two ways: some is radiated into the far field, the Raman radiative power, while some is dissipated as heat (absorbed) in the plasmonic nanostructures. The radiated Raman power is the integral of the Poynting vector associated to $\mathbf{E}_{\text{rad}}(\mathbf{R}, \omega_r)$ over all the solid angles. The radiation enhancement $M_{\text{rad}}(\mathbf{r}_m, \omega_r)$ can be rigorously calculated as the ratio between the radiated power in presence, and in absence of the MNP. $M_{\text{rad}}(\mathbf{r}_m, \omega_r)$ is usually approximated using the optical reciprocity theorem as $M_{\text{rad}}(\mathbf{r}_m, \omega_r) = |\mathbf{E}_{\text{loc}}(\mathbf{r}_m, \omega_r)|^2 / |\mathbf{E}_0(\mathbf{r}_m, \omega_r)|^2$, where $\mathbf{E}_{\text{loc}}(\mathbf{r}_m, \omega_r)$ is the local field produced by a plane wave field oscillating at the Raman frequency. Finally, the SERS enhancement factor at the molecule's position can be approximated as

$$EF(\mathbf{r}_m, \omega_0, \omega_r) = M_{\text{exc}}(\mathbf{r}_m, \omega_{\text{ph}}) \cdot M_{\text{rad}}(\mathbf{r}_m, \omega_r), \quad (33.2)$$

that is, the product of the excitation and re-radiation field enhancements factors. Since ω_r is generally close to ω_{ph} , compared to the plasmon resonance width, we can approximate $\mathbf{E}_{\text{loc}}(\mathbf{r}_m, \omega_r) \simeq \mathbf{E}_{\text{loc}}(\mathbf{r}_m, \omega_{\text{ph}})$, and therefore:

$$EF \simeq \left(\frac{|\mathbf{E}_{\text{loc}}(\mathbf{r}_m, \omega_{\text{ph}})|}{|\mathbf{E}_0(\mathbf{r}_m, \omega_{\text{ph}})|} \right)^4, \quad (33.3)$$

which is the well known $|E|^4$ approximation of the SERS enhancement factor, according to which the SERS field enhancement is proportional to the fourth power of the local field amplification. Just as an example, an amplification of a local field of a factor ~ 3.2 , is expected to produce a two orders of magnitude amplification of the Raman signal.

Achieving intense SERS signals largely relies on devising substrates that possess efficient SERS *hotspots*. Most efficient SERS hotspots arise from coupled nanostructures with nanometric inter-particle gaps. Typical examples are the MNPs dimers. The local EM field in the gap between Au or Ag nanoparticle dimers and oligomers with inter-particle nano-gaps is very intense due to the strong EM coupling. Fig. 33.6 shows how the SERS signal increases by 1 order of magnitude when 100×60 nm gold nano-rods are approached to one another

from 200 nm (tip-by-tip distance) to 15 nm. The SERS enhancement in the hotspots can be estimated in relation to that of uncoupled antennas by an order of magnitude, a calculation that establishes some upper bounds for the SERS EF [13]. Calling V_A and V_{HS} the effective interaction volume of incident light, respectively, with the nanoantenna and the volume of hotspots region, the averaged *EF* (*AEF*) over the nanostructures, i.e., assuming a uniform molecular coverage of the structures, turns out to be $AEF = (V_A/V_{HS})^{4/3}$. Assuming reasonable values for $V_A = 100 \times 60 \times 60 \text{ nm}^3$ and $V_{HS} \sim 10^3 \text{ nm}^3$, values for $AEF \sim 10^3$ are expected. This values ramp up to 10^6 if we assume hot spots of a volume $V_{HS} \sim 10 \text{ nm}^3$, typical of nano-rods at a few nm distance. The above model is probably too optimistic, but provides an upper level to the gain that can be expected by the near-field coupling of MNPs in SERS.

33.2.2 Chemical enhancement

The electromagnetic model is not fully satisfactory, since it implies that any molecular compound close to a hot spot should be subject to SERS amplification, and that the SERS spectrum should be an exact enhanced replica of its ordinary Raman spectrum. Experiments show, instead, that this is not the case, suggesting additional or perhaps alternative molecule-specific contributions to the SERS enhancement that induce selective amplification of specific vibrational modes. The so-called chemical enhancement is a short range effect (few Angstroms) that crucially depends on the electronic states of the molecules and of the metal substrate on which they are absorbed. The chemical enhancement factor is defined as $G_{\text{chem}} = \sigma_k^{\text{ads}}/\sigma_k^{\text{free}}$, i.e., as the ratio between the Raman cross-section of the k -th vibrational mode of a molecule adsorbed on a surface, and the cross-section of a free molecule. It generally does not exceed 10^3 , and that is why it is generally neglected. Chemisorption leads to formation of strong molecule-surface chemical bonds and modifies the Raman polarizability of the molecule through the creation of new metal-particle charge-transfer (CT) states at intermediate energies that can be resonantly excited [11]. Therefore, the chemical enhancement is understood in terms of a surface-induced *Resonance Raman Scattering*. A model describing the electromagnetic and the chemical enhancement of a molecule chemisorbed on a metallic substrate has been developed by Lombardi et al. [12].

33.3 Raman tweezers

Raman tweezers (RT) combine the optical trapping and manipulation capabilities of optical tweezers (OT) with the analytical potential of Raman, enabling chemical analysis (Raman, but also photoluminescence) of micro and nanostructures individually trapped in liquid dispersions. The setup is basically an inverted microscope coupled to a Raman spectrometer that uses high numerical aperture (NA) objectives to tightly focus a laser beam. The high field intensity gradient thus created in the focal region, guarantees both the trapping and the high power density needed to achieve Raman detection (some $\text{mW } \mu\text{m}^{-2}$ are generally sufficient). Figure 33.7 shows a typical single beam RT setup, in which the same laser is used for trapping and Raman excitation. This is the simplest experimental configuration possible. Dual-beam setups can be used when high powers are needed for trapping of very small (nano-metric) objects. In these case, an intense laser at 1064 nm serves for trapping (power up to 1 W), and a visible/NIR laser with powers of 1 mW to 100 mW is used for spectroscopy. Both commercial microscopes and home-assembled structures can be used for focussing, sample positioning and imaging of the trapping process. RT work in back-scattering, i.e., the back-reflected Raman/fluorescence signal is collected through the same objective used for excitation. Generally oil immersion 100X objectives with NA = 1.2 to 1.4, or apochromat water immersion objectives (60 \times , NA = 1.2) are used. Compensation of chromatic aberrations is necessary in dual-beams traps, when exploiting lasers at different wavelengths, to insure the correspondence of the Raman laser focus with the center of the optical trap. In order to achieve a diffraction limited spot of diameter, $d \sim \lambda/\text{NA}$, the laser beams need to be expanded and overfill the central objective pupil (diameter of few cm²). A telescope composed by two lenses whose focal distance ratio is equal to the desired magnification, is used for beam expansion. RT have been demonstrated using different lasers, such as gas lasers, including Ar⁺⁺ (lines at 454, 476, 488, 515 nm) or HeNe (632.8 nm, powers up to 30 mW), Diode-Pumped Solid State (DPSS) lasers (532, 561 nm), or diode lasers (405, 638, 785, 830 nm).

Lasers in the NIR (785, 830 nm) are preferred when fluorescence is an issue. Single longitudinal mode emission is required, together with narrow line-width (< 100 MHz), high power stability (< 1%) and wavelength stability (for laser diodes). Interferential filters are used to isolate the central laser emission and cut out plasma lines (in gas lasers) or residual fluorescence (in diode lasers) that could interfere with the spectroscopic

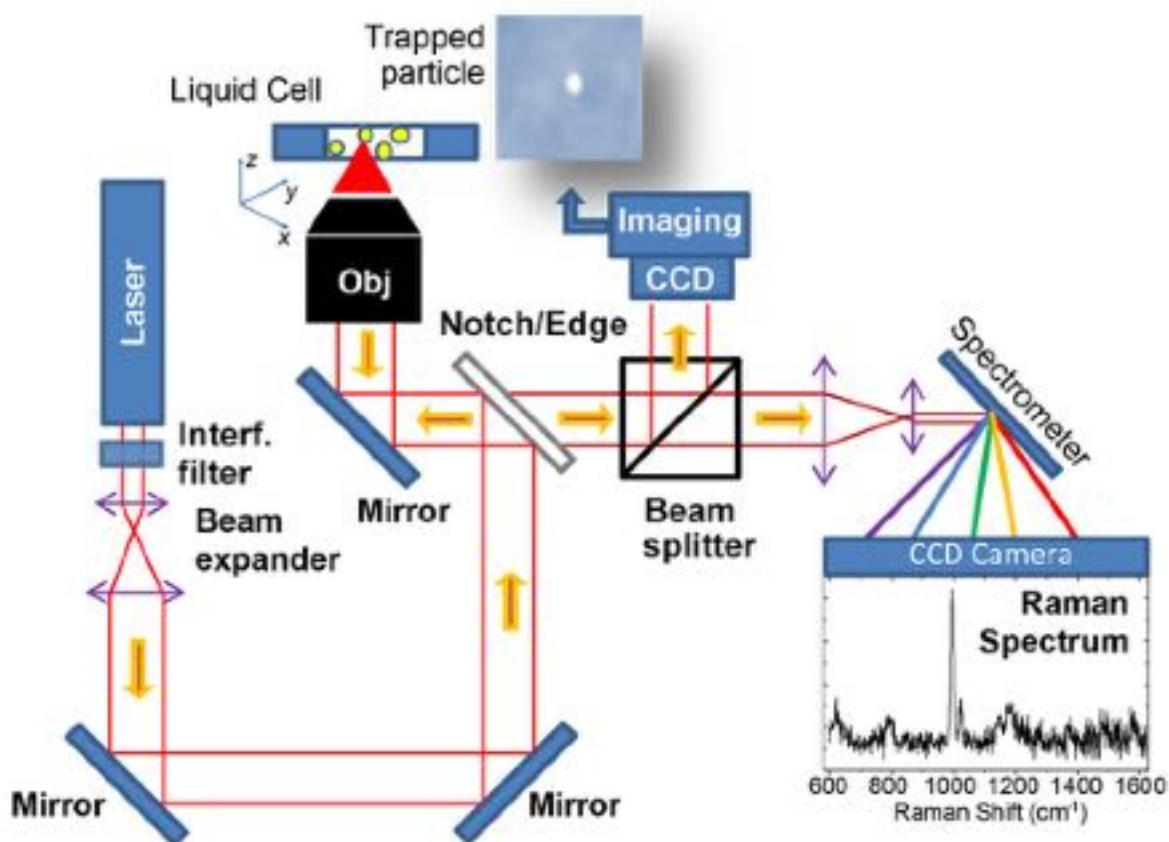


Figure 33.7: Schematic of a single-beam Raman Tweezers setup. Configuration of a single-beam Raman tweezers setup, where a single laser is utilized for both optical trapping and Raman excitation. The setup features an inverted microscope with high numerical aperture objectives to focus the laser beam, creating a high-intensity gradient for effective trapping and Raman signal detection. Key components include the laser source, beam expansion optics, microscope objective, sample chamber, and detection system for capturing the Raman or fluorescence signal in a back-scattering geometry. The inset highlights the expanded laser beam path and the precise focusing necessary for optimal performance.

signal. The expanded laser beam is conveyed into the microscope objective, and then focussed onto the sample, after reflection from a notch or edge filter centered at laser emission wavelength. The laser light is focussed on glass micro-chambers containing the particles dispersions. The micro-chamber is illuminated from the top with white light sources for optical imaging purposes. To this aim, part of the backscattered beam is sent to a CCD camera through a beam splitter. The identification of the specific particle to be trapped is carried out by moving the sample's micro-chamber under the laser spot with the aid of a motorized stages or, when nano-metric resolution is needed, with the help of piezoelectric tables. Monochromators are used for spectral analysis of the light. Grating and focal lengths should be chosen in order to obtain spectral resolutions of a some cm^{-1} . Short focal length monochromators, e.g., 19 cm, with 1200 grooves mm^{-1} gratings and Peltier-cooled CCD cameras offer a good compromise in terms of luminosity, spectral resolution and speed of detection. As microcells for sample hosting, microscope slides with hemispherical cavities (15 – 18 mm diameter, 0.5 – 0.8 mm depth) can be used, covered with soda-lime coverslips and sealed with nail polish. Alternatively, microcells can be assembled from two soda-lime coverslips separated by a 30 μm thick, 1 cm diameter ring of bioadhesive tape. This enables one to focus on both sides of the cell and use extremely small volumes of sample, ca. 10 μL .

33.3.1 Applications of Raman tweezers for Raman and SERS

Raman tweezers (RT) offer a versatile and powerful method for trapping and analyzing micro- and nano-scale materials in liquid environments. This combination of optical trapping and Raman spectroscopy enables precise chemical identification and structural analysis of various samples, including microplastics and biomolecules.

Trapping and analysis of micro- and nano- plastics in water

RT represent an enabling tool for micro-plastics analysis in water. Microplastics (MPL) are small fragments of plastic materials smaller than 5 mm. Primary sources are cosmetics, cleaning products, by-products of the industrial manufacturing processes (e.g., thermal cutting, 3D printing, air-blasting). Fragmentation of larger particles into micrometric debris, due to weathering or mechanical friction, constitutes an important source of secondary MPL. Plastics have refractive index from 1.45 to 1.65 and, unless treated, have negligible absorption in the visible-NIR. For this reason in water, MPL are easily trapped. For nano-plastics, the size is important. 300 nm PMMA particles are stably trapped, whereas to trap sub-100 nm MPL laser powers of several hundreds of mW are needed. Raman spectroscopy provides means for unambiguous chemical discrimination of the particles. Fig. 33.8 shows the optical images and Raman spectra of MPL made of different materials (PET, PA6, PVC, PPMA, PP), optically trapped in water and analyzed by RT. Particles from 2 μm (PVC) to $\sim 1 \mu\text{m}$ (PET, PA6), and sub-micron scale (PMMA, PP) are observed and their size be measured within the spatial resolution of the microscope. The chemical nature of the particles is determined by comparing the Raman fingerprint with reference spectral libraries.

Example 33.2: Trapping power. Calculate and plot the power threshold needed to trap PE, PS, PVC, PMMA as a function of their diameter in the 10 - 100nm range using several laser lines commonly available, from the NUV to the NIR, assuming a $NA = 1.3$. Calculate the gradient-to-scattering force ratio objective and comment.

SERS detection of biomolecules in liquid via optical aggregation of metal colloids

SERS has enormous application potentials in label-free detection of biomolecules and proteins [17, 18, 19]. The vibrational spectrum of proteins can give insight on their conformation, structure, and functional state [20]. Different concepts of SERS biosensors have been demonstrated [21]. SERS detection of proteins in liquid environment has the advantage of keeping them in their natural aqueous habitat, being extremely sensitive at the same time. Optically-induced aggregation of metal nanoparticles can offer a viable solution to achieve high sensitivity SERS and tailor the limit of detection of Raman spectroscopy, generally 1 mM, to the pico-molar regime, although it must be taken into account that the interaction with the nanoparticles can alter their native state [22]. The idea is to embed the target molecules at the nanoscale interstices among the nanoparticles during the aggregation process, and exploit the local field enhancement at the *hot spots* (reaching factors of 10^5 , as

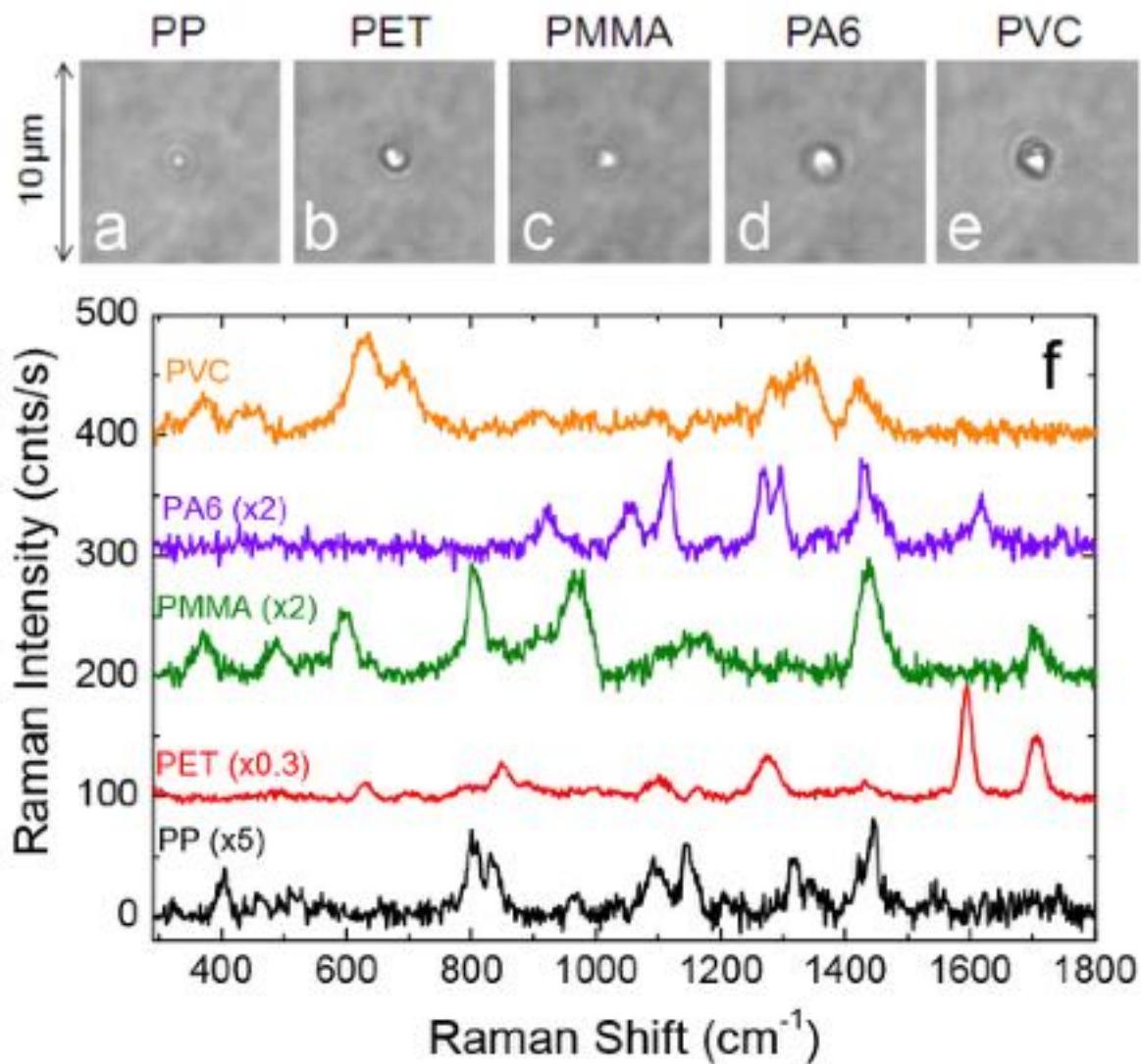


Figure 33.8: **Raman spectra of optically trapped microplastics.** (a)-(e) Optical images of optically trapped micro- and nano-plastics made of different materials and dispersed in seawater. The dimensions range from 2 μm (PVC) to 1 μm (PET, PA6), down to sub-micron regime (PMMA, PP). (f) Raman spectra of the optically trapped particles. Integration times vary from 2 to 60 s. Laser power is 11 mW. Reproduced from Ref. [16].

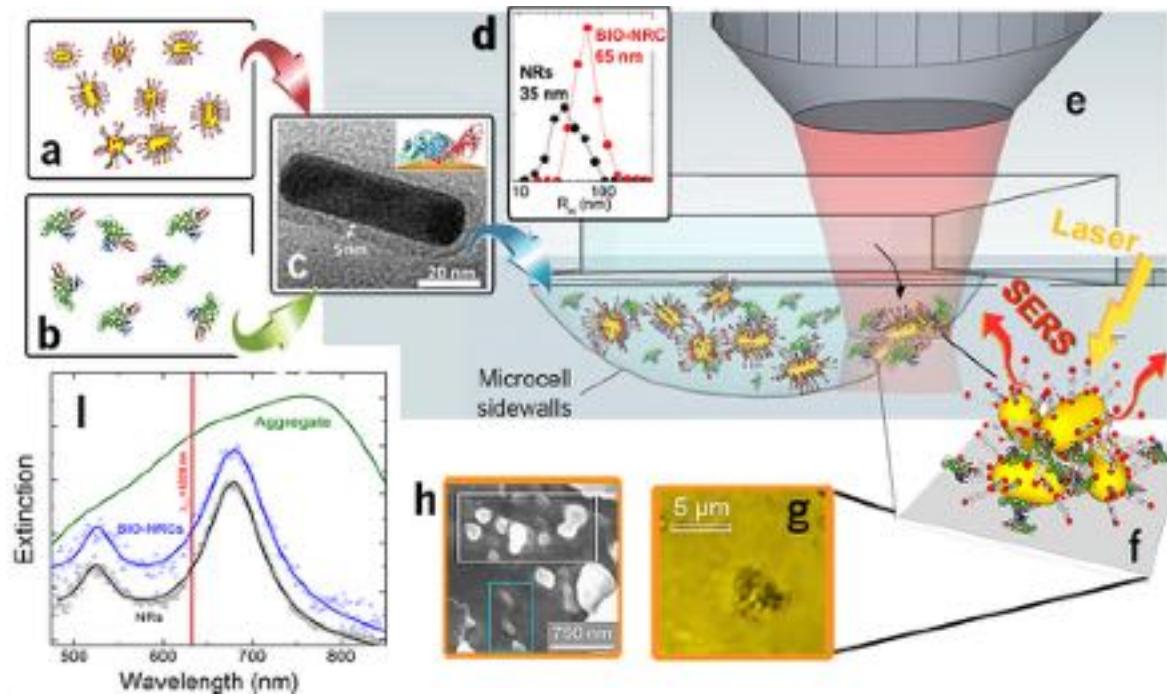


Figure 33.9: Sketch of the LIQUISOR. The steps include: (a) Gold nano-rods (NRs) stabilized with cetyltrimethylammonium bromide in deionized water. (b) Mixing NRs with a biomolecule solution in phosphate buffered saline (PBS) to maintain pH stability. (c) Formation of biomolecule-NRs complexes (BIO-NRCs) as biomolecules bind to NRs. (d) Increase in the hydrodynamic radius of NRs upon binding with biomolecules, detected by dynamic light scattering. (e) Optical aggregation of BIO-NRCs in a glass microcell using a focused laser beam. (f) and (g) Aggregation of BIO-NRCs around the laser focus, forming micron-sized clusters. (h) Random orientation and mixing of gold NRs with biomolecules within the aggregate. (i) Extinction spectrum of the aggregate showing a broadening and red shift of the plasmon resonance, indicative of cluster formation. Adapted from Ref. [18]. The sub-figure (c) is reproduced from Ref. [22].

shown in Ref. [23]) to detect their vibrational fingerprint by SERS. Radiation pressure from a focused laser beam can be used to push and aggregate metal NPs in liquid and, at the same time, to excite SERS.

The various steps of this methodology, referred to as LIQUISOR (from LIQUId SERS sensOR) by Fazio et al. [18, 24], are sketched in Fig. 33.9. Gold nano-rods covered by a stabilizing surfactant layer of cetyltrimethylammonium bromide and dispersed in deionized water (Fig. 33.9a), are mixed to a solution of biomolecules dispersed in phosphate buffered saline (PBS, Fig. 33.9b) to preserve the pH. Upon mixing, the biomolecules bind to the gold NRs (Fig. 33.9c), due to destabilization of the surfactant bilayer and intercalation of the amino acid residues of the protein. This yields the formation of biomolecules-NRs complexes (BIO-NRCs) [22] dispersed in the solution, in which the biomolecule stabilizes the NRs. Dynamic light scattering (Fig. 33.9d) evidences the increase of the mean hydrodynamic radius of the NRs from 35 to 65 nm when bound with a protein such as Bovine Serum Albumin (BSA). The BIO-NRCs solutions are pipetted into glass microcells and placed under a Raman spectrometer [18, 24] or a RT setup [25] (Fig. 33.9e), where optically induced aggregation is achieved by focusing a laser beam into the sample compartment. Radiation pressure enables the dynamic accumulation of the BIO-NRCs onto the bottom surface of the microcell, around the laser focus, where they stuck and aggregate. NRs aggregates reach the size of several microns (Fig. 33.9g). Inside the aggregate (Fig. 33.9h) gold NRs are generally randomly oriented and mixed with the biomolecules. The extinction spectrum of the aggregate (Fig. 33.9i, green line) shows a broadening and red shift of the plasmon resonance, as expected when the NRs form clusters.

Fig. 33.10a shows the SERS spectra of BSA acquired by LIQUISOR at decreasing concentrations from 0.1 mM, (red line), down to 50 nM (orange line). The spectra show very intense peaks in the same spectral ranges of the BSA vibrations measured by Raman in solution phase and lyophilized state, although shifts and

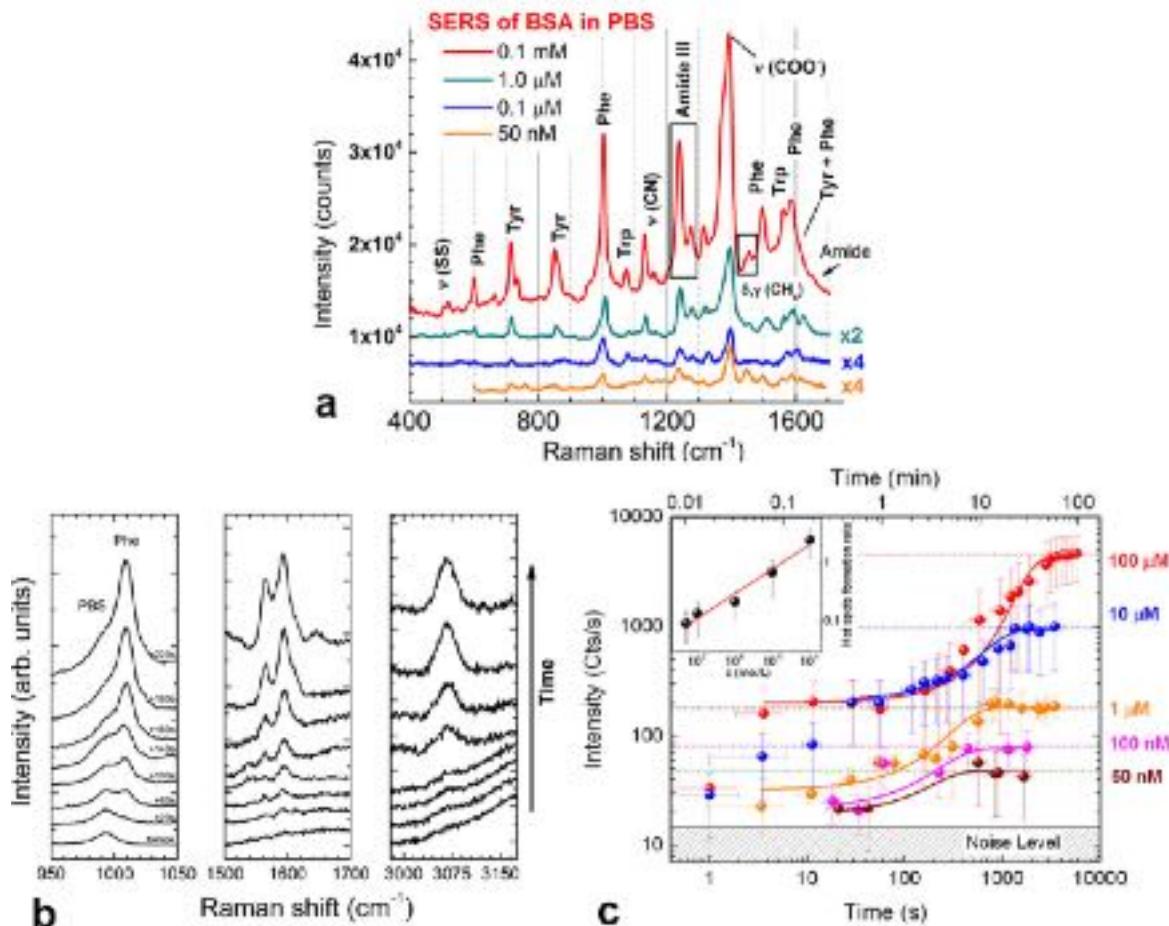


Figure 33.10: Concentration-dependent SERS analysis of BSA and temporal evolution during aggregation process. (a) SERS of BSA at concentrations of 0.1 mM (red), 1 μM (green), 0.1 μM (blue), and 50 nM (orange). Spectra are offset for clarity. Experiments are carried out at a laser wavelength of 632.8 nm and power 6.7 mW. (b) SERS spectra of BSA acquired during the aggregation process in selected spectral ranges: 950 – 1050 cm^{-1} (Phe ring breathing peak), 1500 – 1650 cm^{-1} (aromatic amino acids and Amide I), and 3000 – 3150 cm^{-1} (aromatic amino acids CH stretching). Reproduced from Ref. [18]. (c) Time evolution of the SERS intensity at different BSA concentrations. $T = 0$ is set at the onset of the process, when the first weak SERS signal is detected. Solid lines are fits of data using a Boltzmann growth kinetics. (Inset) Plot of the SERS intensity at saturation Vs BSA concentration. The solid red line is a power law fit with exponent 0.40 ± 0.04 . Reproduced from Ref. [24].

different peaks intensity ratios are observed. The modes were attributed to vibrations of the side-chain aromatic amino acids (Phe, Tyr, Trp) of the protein, characterized by a high Raman cross-section. Such mode assignment must be taken with care, since spectral shifts and intensity changes occur as a consequence of the interaction of the protein with the residual surfactant layer and the gold surface. The measurements confirmed a picture in which the protein at the hot spot was aggregated and strongly interacting with the NRs surface, featuring a somehow altered secondary structure. In these experiments, 50 nM was the limit of detection of BSA in PBS, to be compared to the value of 1 mM found in conventional Raman experiments [18]. Fig. 33.10b displays the temporal evolution of the SERS spectra, evidencing how the enhanced BSA vibrational peaks emerge from the PBS background during the laser irradiation and aggregation process. Fig. 33.10c displays the time evolution of the SERS intensity for different protein concentrations, starting from the onset of the process ($T = 0$), up to the saturation of the signal when the BIO-NRC aggregates cover the entire laser spot. The SERS intensity curves follow a Boltzmann growth kinetics profile (solid lines in Fig. 33.10c) [24]. Intensity fluctuations from one aggregate to another, caused by non-uniformity of the enhancement, are represented by the error bars. Although large fluctuations are observed, a clear discrimination among the different concentrations is visible when plotting the signal intensity at saturation (inset of Fig. 33.10c).

33.4 Tip-enhanced Raman spectroscopy

Tip-enhanced Raman spectroscopy (TERS) combines the atomic spatial resolution of scanning probe microscopy (SPM), with the analytic information provided by Raman spectroscopy and the ultrahigh sensitivity of plasmon enhanced spectroscopies, in a unique tool capable of morphological and chemical imaging of nanostructures and molecules absorbed on surfaces with extreme sub-diffraction spatial resolution. This is achieved by exploiting the large amplification and the nanoscale spatial confinement of the electromagnetic field when a sharp metallic tips is illuminated with light [26, 27, 28]. Metallic (or metallized) tips act as optical nano-antennas. By combining the lightning rod effect with plasmonic resonances in the apical region, metallic tips feature strong oscillating charge densities at their apex that efficiently transform the optical far-field radiation into extremely spatially confined near-fields, making the tip act as a highly confined light source, analogous to *hot-spots* in SERS [29, 30]. The integration of metallic TERS tips on SPM instruments, like scanning tunneling microscopy (STM) [31], atomic force microscopy (AFM) [32] or shear-force microscopy (ShFM) [33], permits a precise spatial positioning and scan of a surface and to gain, in addition to the morphological information, also a fast characterization of the physical /chemical properties of nano-structured surfaces. The resolution can reach the nanoscale (10 – 20 nm) or even the atomic level under ultrahigh vacuum and low temperature conditions [34, 35, 36]. In this section, after an introduction on diffraction and spatial resolution, we will illustrate the experimental implementations and applications of TERS.

33.4.1 Field propagation, diffraction, and spatial resolution

The limited spatial resolution of conventional optical microscopy (200 nm to 300 nm in the visible range) is intrinsic to the wave nature of light. When a propagating beam is focused by a lens or when light emitted from a point size source, e.g., a point dipole, is collected by an objective lens, the minimum size of the corresponding optical image will have a finite spatial extension, dictated by the diffraction of light.

Optical diffraction arises because of the wave nature of electromagnetic field and the way it propagates in the medium, as described by the Huygens–Fresnel principle. This states that the wavefront of the electromagnetic field can be considered as an infinite ensemble of point sources emitting spherical waves (secondary waves) and that the electromagnetic field at an arbitrary point of space is given by the superposition of all the secondary waves at this point. A powerful way to understand these concepts is going through the *angular spectrum representation* of optical fields [30]. The electromagnetic field at a given point $z = 0$ is assumed monochromatic at frequency ω and wavelength λ , and expanded into a summation of plane waves in the Fourier space with different spatial frequency (or angular) components of amplitude $A(k_x, k_y) = 1/4\pi^2 \iint E(x, y, z = 0) \exp[-i(k_x x + k_y y)] dk_x dk_y$. To calculate the field at a generic position $E(x, y, z)$ in the $z > 0$ half-plane, the field propagator $\exp[ik_z(k_x, k_y)z]$ is applied (multiplied) to each spatial frequency component $A(k_x, k_y)$ and then a sum of all spectral components is carried out at position z , i.e., $E(x, y, z) = \iint \{A(k_x, k_y) \exp[ik_z(k_x, k_y)z]\} \exp[i(k_x x + k_y y)] dk_x dk_y$. Since $k_z(k_x, k_y) = \sqrt{\omega^2/c^2 - k_x^2 - k_y^2}$, for the low spa-

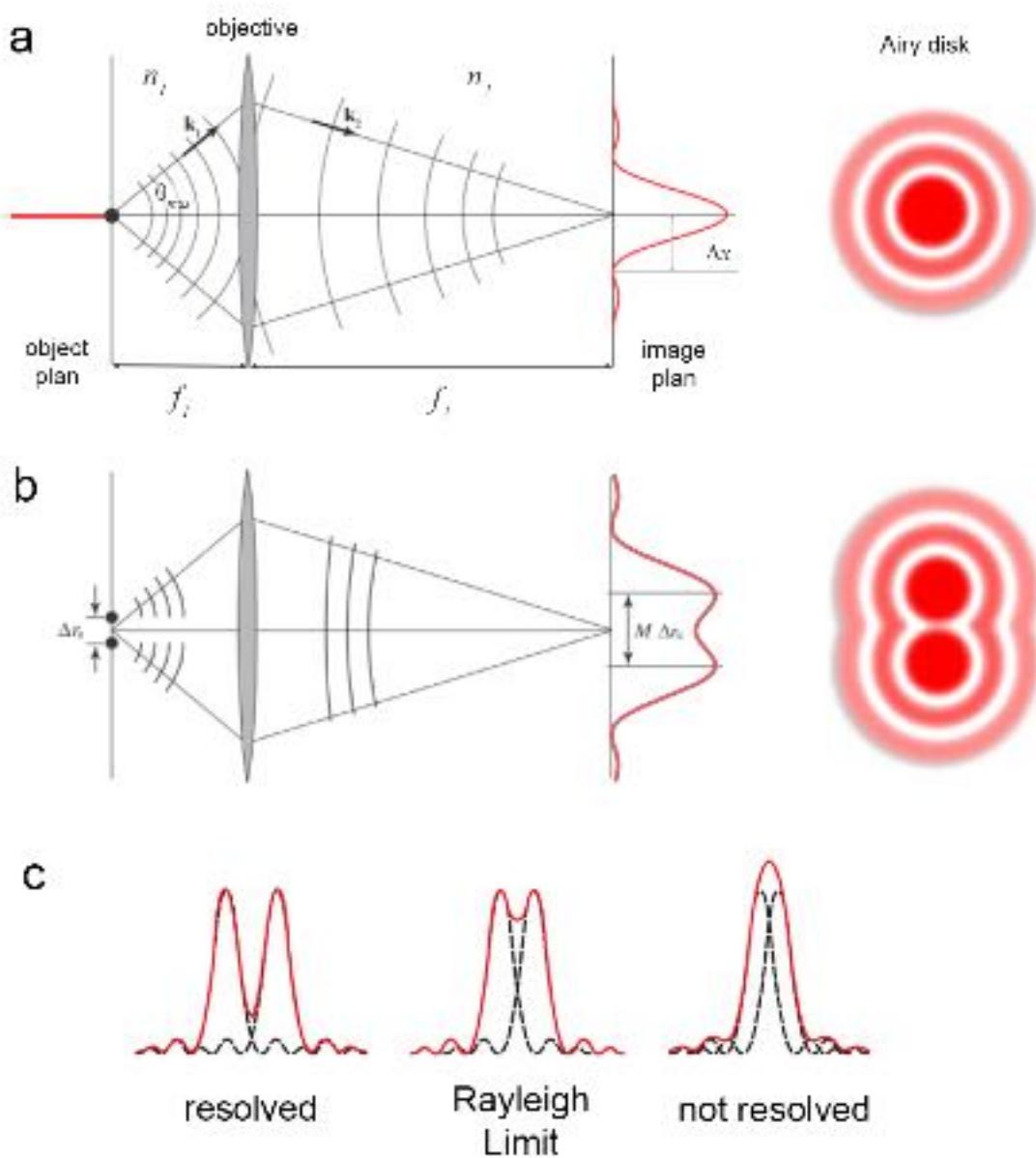


Figure 33.11: **Point spread function (PSF) of a point source.** (a) PSF (Airy disk) of a point source imaged by a lens system of focal distances f_1 and f_2 . (b) PSF of two closely spaced point sources. (c) Line profile of the PSF of two closely spaced point sources and illustration of the Rayleigh criterion in the definition of the spatial resolution. Adapted from Ref. [30].

tial frequencies of the optical field, $k_x^2 + k_y^2 \leq \omega^2/c^2$, k_z is real and the field propagator is an oscillating phase function. The corresponding angular components will consequently propagate to distances $z \gg \lambda$, in the the *far-field*. For the high spatial frequencies, $k_x^2 + k_y^2 > \omega^2/c^2$, k_z turns imaginary and the field propagator will cause an exponential decay of the corresponding angular components, giving rise to *evanescent waves*. Consequently, the high spatial frequencies of the field will remain confined at distances from the source comparable to the wavelength ($z \sim \lambda$), in the *near-field*, being vanishingly small in the *far-field*. Such a low-pass filtering effect, induced by propagation [30], causes the spread of the spatial distribution of the optical field in the plane at $z > 0$, due to the missing high spatial frequency components in the field angular spectrum. This phenomenon causes the spread of the optical image of a point source, limiting in the resolving power of any optical imaging system, i.e., the ability to distinguish two separated point-size objects.

The field distribution of a point source upon propagation in the far-field is described by the so-called *point-spread function* (PSF). The PSF of a point dipole is an *Airy function* (Fig. 33.11a), characterized by an oscillating pattern with a central maximum (84% of the total intensity) and an infinity of less intense side bands. If we consider an optical imaging system with f_1 and f_2 representing the focal distances and n_1 and n_2 the refractive indexes in the object and image space respectively (Fig. 33.11a), the distance between the maximum and the first minimum, namely the Airy disk radius, is $0.61M\lambda/NA$ [30], where $M = (n_1f_2)/(n_2f_1)$ is the magnification of the system, $NA = n_2 \sin \theta_{\max}$ is the numerical aperture, where θ_{\max} is the maximum collection angle (Fig. 33.11a), which defines the maximum spatial frequency that can be detected. The higher the spatial frequency sampled, the smaller the extension Δx of the PSF. The spatial resolution of an optical system is defined as the minimum distance $\Delta r_{\parallel} = \sqrt{\Delta x^2 + \Delta y^2}$ in the object plane at which two point objects produce distinguishable images (Fig. 33.11b). The *Rayleigh criterion* is generally used to calculate $\min(\Delta r_{\parallel})$, which states that two objects are resolved if the maximum of the Airy disk of the first object lies on the first minimum of the Airy disk of the second object (Fig. 33.11c). From the discussion above, it turns out the spatial resolution of lens system is:

$$\min(\Delta r_{\parallel}) = 0.61 \frac{\lambda}{NA}. \quad (33.4)$$

Therefore, in order to increase the resolution, we can either use lower excitation wavelength and/or employ objectives with higher NA. As an example, for red light (633 nm) imaging systems with conventional 100 \times , NA = 0.9 objectives, a spatial resolution of ~ 430 nm is expected, whereas for blue light (405 nm), in combination with a high-NA objective (NA = 1.4), the expected resolution is ~ 175 nm.

Despite the presence of the diffraction limit, optical resolution can be improved taking advantage of innovative far-field excitation/detection schemes [37], as well as exploiting the interaction with evanescent optical fields [27]. This latter approach relies on the high spatial information encoded in evanescent waves, which can be collected and converted into propagating optical fields by placing a nano-structured probe, i.e., a *nanoantenna*, at few nanometers from the emitter [30]. By raster scanning the nano-probe on the sample surface, a nanometric distances, sub-diffraction resolution is achieved in optical imaging. Scanning near-field optical microscopies (SNOM) are techniques that use nano-apertures to push the resolution deep in the nanoscale regime [38, 39]. Apertureless-SNOM (or scattering-SNOM) approaches use a metallic tip illuminated by laser radiation for near-field imaging. First ideas were introduced by Wessel in 1985 [40] and developed to even commercial instruments (see Ref. [41] for a review), with most interesting applications in IR-nanoimaging and Nano-FTIR spectroscopy [42]. While first SNOM implementations were focused on the analysis of the elastic scattering (dielectric contrast), photoluminescence (nanoPL or *tip-enhanced photo-luminescence*, TEPL) and Raman analysis (nanoRaman or TERS) soon followed, allowing a chemical characterization of surfaces and interfaces at the nanoscale. The tip in near-field microscopy plays a twofold role: (i) in optical dielectric imaging it acts as a nanoantenna, scattering in the far-field the evanescent high spatial frequency components of the optical fields generated by nanoscale objects; (ii) in photoluminescence and Raman imaging, it acts as a resonant nanoantenna, locally enhancing the local fields, and improving the detection sensitivity by orders of magnitude.

33.4.2 Field enhancement and spatial confinement in TERS

Plasmonic nano-tips (made of gold or silver) can provide the strongest field enhancement whenever the LSP resonances at the tip apex are resonantly excited and nano-structures are imaged after deposition on metallic surfaces. This is generally achieved in the *gap-mode* configuration where TERS is combined with STM. The

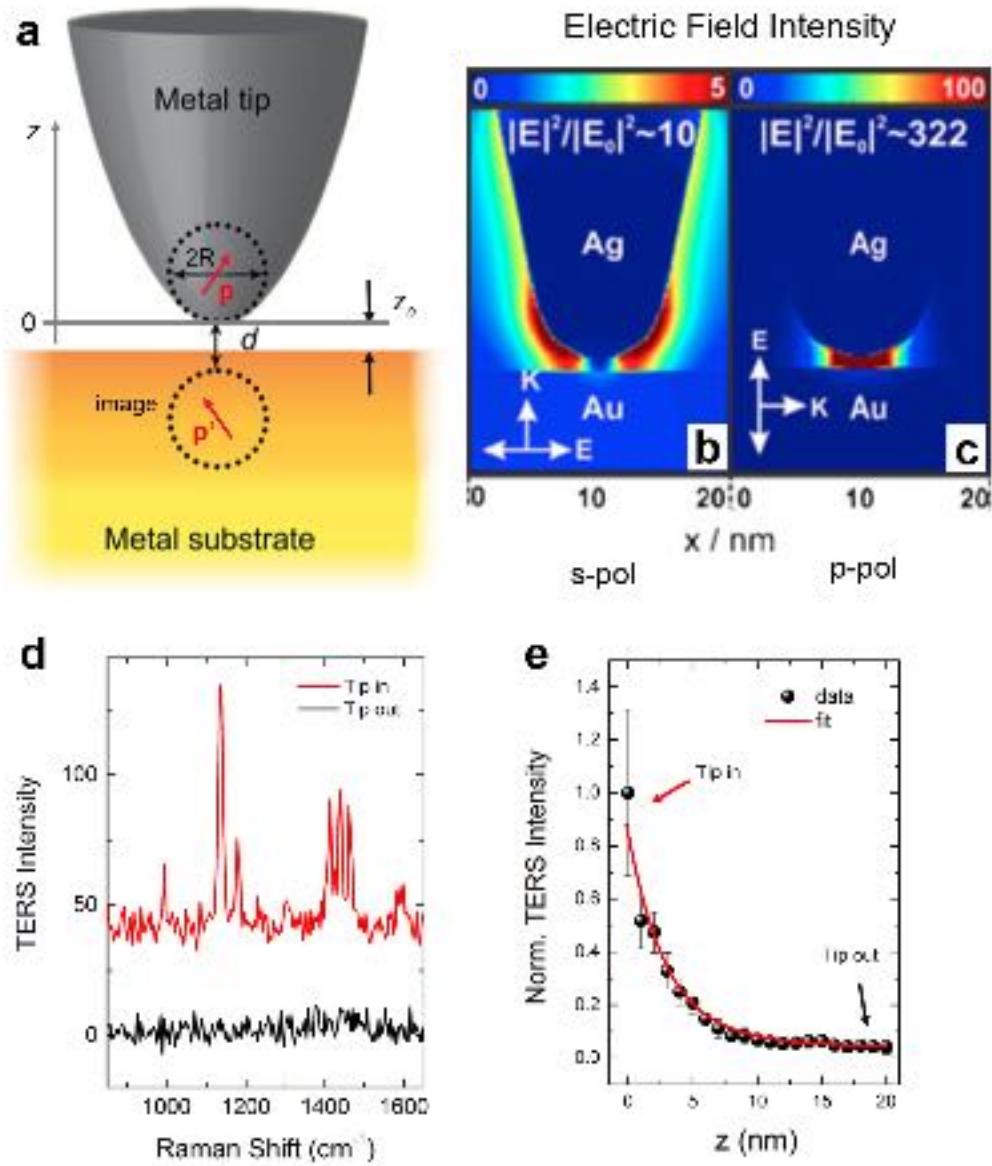


Figure 33.12: Field enhancement and spatial confinement in TERS in the gap-mode configuration. (a) Illustration of TERS in the gap-mode configuration. The field enhancement and polarization-dependent properties of the system are qualitatively described by the polarizability of the dimer constituted by the tip-dipole (\mathbf{p}) interacting with its image dipole (\mathbf{p}'). Adapted from Ref. [28]. (b), (c) Electric field distribution at the apex of a 10 nm silver tip placed at 1 nm from a gold substrate and illuminated at 532 nm by linearly polarized field \mathbf{E} perpendicular (b) or parallel (c) to the tip axis. Adapted from Ref. [43]. (d) TERS spectrum (red line) of azobenzene on a gold surface with the tip at about 1 nm, compared to the Raman signal (black line) acquired when the tip-sample distance is increased to ~ 20 nm (tip out). (e) TERS intensity vs tip-sample distance from azobenzene molecules on a gold film. Red line is a model calculation in the dipole approximation of the experimental data (black dots).

amplification and charge localization is extremely sensitive to the tip shape and to the polarization of the excitation field [30, 44]. A simple theoretical picture of *gap-mode* TERS is obtained using the dipole image-

dipole model [45, 41]. The tip is approximated as a plasmonic nano-sphere with radius R that, when excited, generates an electric dipole \mathbf{p} which will in turn induce an image dipole \mathbf{p}' in the metal (Fig. 33.12a). This creates a virtual plasmonic dimer separated by a distance d and featuring in the gap a very intense hot-spot. This model notably describes, at least qualitatively, the strong dependence of the local field on the excitation field polarization. When the excitation field is oriented parallel to the tip axis (p -polarized), the generated dipoles are parallel and their fields add up; when the polarization is orthogonal to the tip axis (s -polarized), the dipoles are anti-parallel and the fields tend to cancel out [41]. More sophisticated finite-elements theoretical simulations, based on the $|E|^4$ approximation and taking into account the entire tip geometry [43], allow us to visualize that when the incident beam is s -polarized the hot-spot regions are located at the shaft of the nano-tip (Fig. 33.12b), while if the incident beam is p -polarized the hot spot region is at the apex and the corresponding local field enhancement is ca. 10 times stronger (Fig. 33.12c). In general, TERS can reach signal enhancement values up to 10^6 [46].

To quantify the TERS enhancement the contrast function $C = (I_{\text{NF}} - I_{\text{FF}}) / I_{\text{FF}}$ is generally used, that compares the enhanced TERS near-field signal, I_{NF} , measured when the tip is in contact with the sample surface (Tip-in, Fig. 33.12d, red line), with the Raman far-field signal, I_{FF} , measured when the tip is at some microns distance (Tip-out, Fig. 33.12d, black line). In order to properly evaluate the enhancement, the volume of the sample effectively probed in each case must also be taken into account. If we call V_{NF} the near-field volume interested by the enhanced field in the tip-in configuration, V_{NF} is expected to be far smaller compared to the diffraction limited volume, V_{FF} , illuminated in the tip-out configuration by the focused laser beam. An estimation of V_{FF} can be obtained assuming $V_{\text{FF}} \sim R_{\text{focus}}^2 \cdot h_{\text{focus}}$, where $R_{\text{focus}} \approx 0.61\lambda/NA$ and $h_{\text{focus}} \approx 2\lambda/NA^2$ are the radius and the height of the focused laser beam, given by PSF extension within the focal plane and along the beam propagation axis, respectively [30]. On the other hand, $V_{\text{NF}} \sim R_{\text{NF}}^2 \cdot h_{\text{NF}}$ where R_{NF} is the lateral extension of the near-field and can be considered in a first approximation as $\sim R$, i.e., dictated by the radius of curvature of the tip apex, while h_{NF} is the depth of the near-field region. The EF can be calculated normalizing the contrast function to the probed volumes as follows [46, 41]:

$$EF = C \frac{V_{\text{FF}}}{V_{\text{NF}}} = \left(\frac{I_{\text{NF}}}{I_{\text{FF}}} - 1 \right) \frac{R_{\text{focus}}^2}{R_{\text{NF}}^2} \frac{h_{\text{focus}}}{h_{\text{NF}}}. \quad (33.5)$$

The estimation of h_{NF} is strongly dependent on the characteristics of the tip and of the tip-sample interaction [47]. However, if the sample thickness is of few nanometers, as for molecular monolayers, nanotubes or 2D materials on surfaces, or more generally when $h_{\text{sample}} \lesssim h_{\text{NF}} \ll h_{\text{focus}}$, the sample can be considered bidimensional and only the surface illuminated has to be taken into account, yielding $EF \sim C \cdot (R_{\text{focus}}/R_{\text{tip}})^2$. Eq. (33.5) provides meaningful results if $I_{\text{FF}} \neq 0$. When no signal is detected in far-field conditions, like in Fig. 33.12d (black line), we can use the signal noise level (RMS) as a reference to estimate a lower bound of the EF . An experimental approach for retrieving the extension h_{NF} of the near-field region is the analysis of the fast increase of TERS intensity as the distance d between tip and sample surface decreases to contact (Fig. 33.12e). In the dipole model [48, 49] the TERS intensity depends on d as a power law $(1 + d/R)^{-q}$, where the exponent q is 8 or 10 for fully coherent or incoherent scattering, respectively [49]. The approach curve in Fig. 33.12e shows an example of the TERS signal increase when approaching the tip to the sample surface. The TERS signal starts appearing at tip-sample distances $d < 5 \text{ nm}$, proving that the enhanced field is confined at few nanometers from the tip apex.

33.4.3 Experimental configurations for TERS

Due to the extreme spatial localization of the field enhancement at the tip apex, a fundamental requirement for TERS imaging is that the tip must be kept at nanometric distance from the sample surface during all the surface scan. In order to fulfill this need, experimental TERS setups take advantage of the tip-sample distance control systems used in scanning probe microscopes such as STM, AFM or ShFM. The control signal S_{control} sensitive to the tip-sample distance is the tunneling current flowing between tip and sample in STM, the mechanical deflection of a tip attached to a cantilever in AFM, the oscillation amplitude of a quartz tuning fork to which of a tapered metal wire is glued in ShFM [31, 32]. Piezoelectric stages are generally used to scan the sample (or tip) in the three spatial directions with the required precision. While the sample is raster scanned in the xy plane, a proportional-integral-differential (PID) feedback loop automatically adjusts the vertical z position of the sample in order to keep S_{control} equal to some predefined set point S_{set} that, in turn, will keep constant

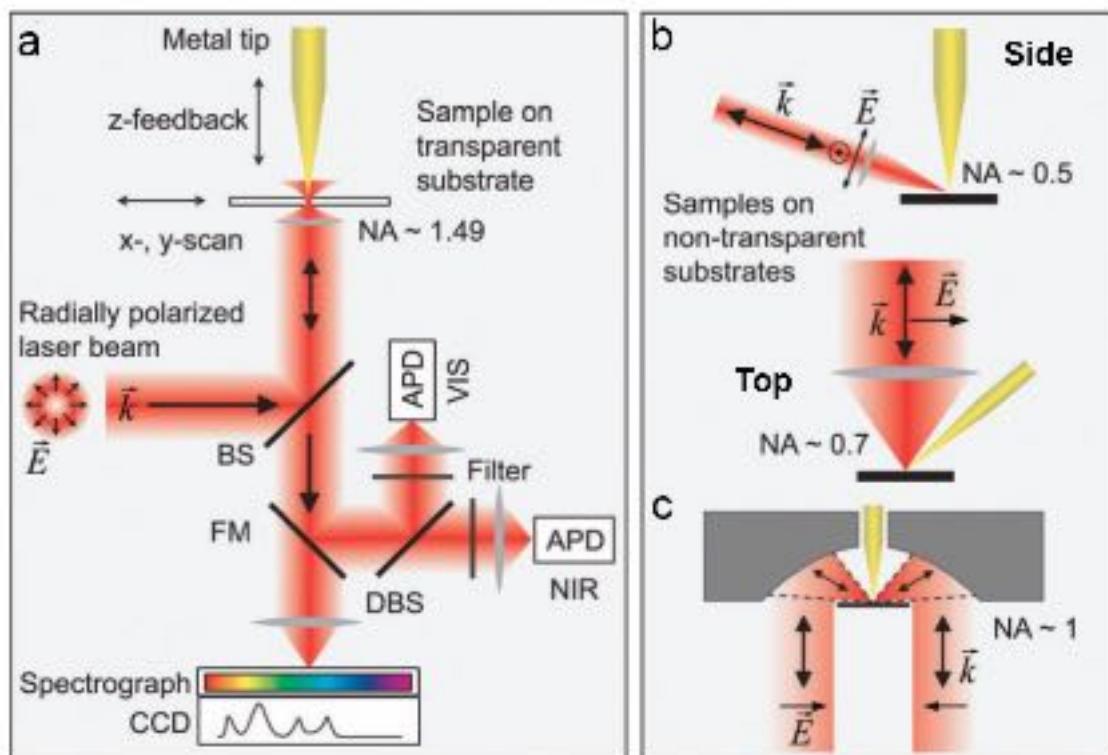


Figure 33.13: **Experimental configurations for TERS.** (a) On-axis *transmission-mode* illumination, suitable for transparent samples or samples deposited on transparent substrates. The sharp metal tip is positioned in the center of a tightly focused radially polarized laser beam. The Raman signal is detected either by avalanche photodiodes (APDs) for the VIS and NIR spectral range or by a combination of a spectrograph and a CCD. (b) Side- and Top- illumination configurations, suitable for non-transparent substrates. The Raman signal is collected in backscattering. (c) Parabolic mirrors with $NA \sim 1$ can be used in place of glass optics to focus and collect laser, maximizing the collection efficiency. To generate a strong field component parallel to the tip axis required for efficient field enhancement, scheme (a) and (c) need a radially polarized laser modes. Reproduced from Ref. [50].

the tip-surface distance. By monitoring $z(x,y)$, the topographic profile of the sample surface is reconstructed, together with the spectroscopic optical/chemical information.

Illumination schemes for TERS can be divided in two classes depending on whether the substrate on which the sample is deposited (or the sample itself) is transparent or opaque (Fig. 33.13). On transparent samples, on-axis illumination in *transmission-mode* can be carried out using an inverted microscope configuration, i.e., illuminating the plasmonic tip with a laser beam that propagates parallel to the tip axis, tightly focussed from the bottom (Fig. 33.13a). This configuration permits the use of high numerical aperture oil/water immersion objectives ($NA \sim 1.4$), that maximizes the power density at the tip apex, decreases the size of the confocal volume contributing to the far-field background. This configuration provides spectra with the highest signal-to-background ratio [27, 51]. In order to benefit of an optimal field amplification, however, it is important that part of the electric field is parallel to the tip axis. Higher order laser modes are thus preferred, using radially or azimuthally polarized beams with a donut shape, such as the Hermite-Gaussian modes HG_{10} [44, 52].

Opaque samples requires alternative TERS excitation geometries. The most widely employed configuration is illustrated in Fig. 33.13b (*side-illumination*). The tip axis forms an angle α with the propagating direction of the laser beam [53] which is focussed from one side with the aid of long working distance objectives ($WD \sim 5 - 10$ mm), working in air and characterized by a numerical aperture ($NA \sim 0.5 - 0.7$) with is smaller with respect to immersion objectives. TERS shows an optimal enhancement when the angle α is about 60 degree [54]. A further configuration introduced for opaque samples investigation is the one in which light is focused from the top, perpendicularly to the sample plane as illustrated in Fig. 33.13b (*top-illumination*). Top-illumination works in combination with bent or *nose-tips* that form with the sample surface an angle of some tens of degrees. As for the transmission- and side-illumination configuration, the backscattered signal is collected also here. A configuration enabling the investigation of non-transparent samples, that maximizes the illumination/collection angle, is the one that employs a parabolic mirror, instead of glass optics, as a focusing element (Fig. 33.13c). Radially polarized laser beams are preferred, as in *transmission-mode*, to provide a strong component of the electric field parallel to the tip axis. While this illumination/detection scheme is expected to provide optimal signal collection efficiency on opaques surfaces, it requires critical handling and alignment procedures to perform at its best [55, 56]. *Side-illumination* configurations are currently the most adopted ones in laboratory and commercial devices.

The core element in a TERS experiment is indeed the tip. It directly influences spatial resolution, reproducibility, and signal enhancement [47]. It is well consolidated that silver and gold are good choices for achieving good TERS signal with an illumination in the visible region [47, 57]. Silver is considered optimal for plasmonic field enhancement in the visible range, since the imaginary part of the dielectric constant in silver is quite small and the absorption lies in the ultraviolet, below 350 nm [30]. Degradation makes, however, more difficult to achieve reproducible results obtained with silver tips. When exposed to air, tips feature a rapid decrease of the signal enhancement performances, due to oxidation and sulfurization processes. As well, they are more subject to contamination from the sample itself, already after a limited number of uses [58, 59]. For this reasons gold tips, which are chemical inert in a laboratory environment, are the most widely used, despite of their lower signal enhancement with respect to silver. Gold features interband absorption at wavelengths shorter than 500 nm, making Au tips suitable for applications in the visible range from 600 nm towards the lower energies. The conformation and fabrication procedure of TERS tips is related to the SPM microscope on which experiments are implemented. AFM-TERS tips are usually produced by evaporation or electrodeposition of plasmonic metals (silver, gold and even aluminum) onto commercial AFM silicon or silicon nitride tips [28, 47]. In order to obtain TERS enhancement, a single plasmonic nanoparticle (NP) placed at the end of a conventional AFM tip would be required. While examples have been shown, this fabrication approach is highly time consuming and scarcely reproducible. In the most common case, AFM-TERS tips are represented by a normal AFM tip homogeneously coated by a continuous metallic film or with several spatially separated metal nanoparticles occurring on the tip surface. STM-TERS tips are tapered metal wires fabricated by means of chemical [60] or electrochemical [61, 62, 63] etching techniques, leading to sharp bulky structures that can be mounted and duly polarized in a STM framework. This kind of TERS tips is more desirable compared to metal-coated tips since they can have lower tip radius ($\lesssim 10$ nm against 20 nm to 30 nm) and higher field enhancement [60]. Other smart ways to excite LSPRs confined at the tip apex involve the possibility to launch surface plasmon polaritons (SPP) along the smooth surface of a tapered metal wire. Propagation towards the wire apex, finally leads to adiabatically compressed upon approaching the tip apical region to generate a strong charge localization which amplify the local field [64, 65]. SPPs can be launched either exploiting a grating

nano-fabricated on the tip shaft [64] or using a photonic crystal [65]. The big advantage of this excitation configurations is the absence of the far-field background that increases the signal-to-noise ratio. No matter the kind of tip used, a precise focusing of the light on the tip apex is necessary to efficiently excite the tip-induced amplification. Nowadays, this is accomplished by precisely raster scanning the laser beam focus across the tip apical region with the help of piezoelectric scanners, collecting the weak continuum optical emission from the metallic tip. Maximization of this signal leads to optimal TERS enhancement when the probe is approached to the sample.

33.4.4 TERS applications

TERS has rapidly become an appealing technique in many application fields, including materials science, catalysis, biology, because it goes beyond a simple localization of small objects at the nanoscale. In fact, thanks to the high sensitivity and spatial resolution, it can provide a nano-resolved structural characterization of surfaces and interfaces also enabling the study and the control of chemical-physical processes at a local level [67]. We provide here few examples to highlight the extreme sensitivity of the technique. Mapping the surface structural composition of carboxyl-modified graphene oxide with a resolution of about 10 nm is shown in Figs. 33.14a-d. Here TERS, by providing simultaneous topography and chemical maps, shows the potential to systematically study point defects in the materials [66], as well to monitor strain in carbon nanotubes with sub-nanometer resolution [68, 69].

Other applications are related to the mapping of how dye molecules are absorbed on metal films. Figs. 33.14e-i shows the distribution of Rhodamine 6G (R6G) on gold crystalline terraces (topography in Fig. 33.14e, chemical map in Fig. 33.14g), separated by nanometric tranches, with a spatial resolution of 10 nm (line profiles in Fig. 33.14f) and Fig. 33.14h, and spectra acquired at 5 nm distance in Fig. 33.14i). Gold tip fabricated by electrochemical etching [62] are effective for the characterization of inorganic nanowires [71], carbon-based nanomaterials [50, 72, 73] or self-assembled monolayers of molecules grafted on gold surfaces [74, 75]. TERS

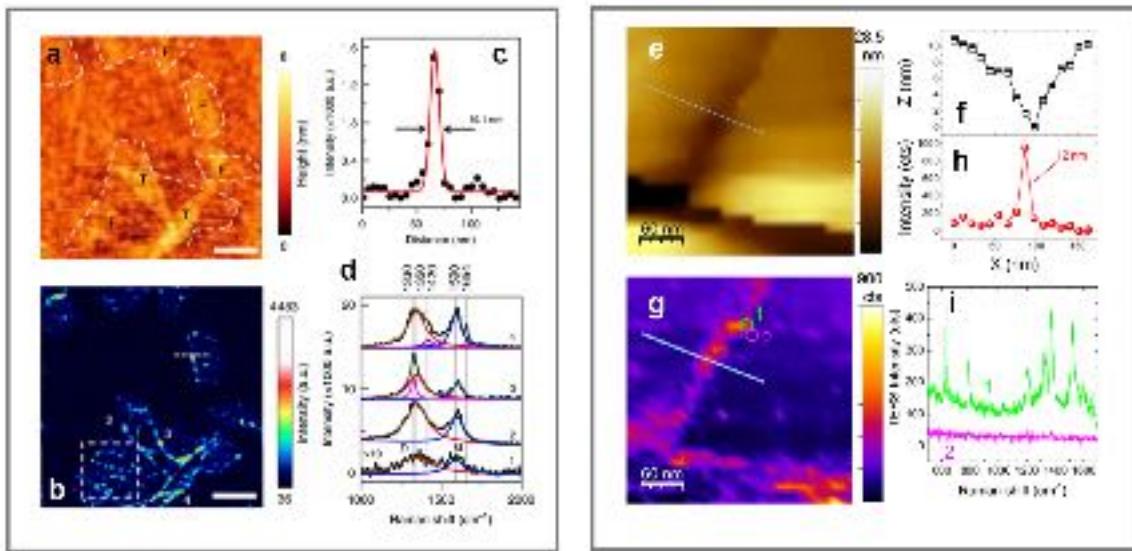


Figure 33.14: **High sensitivity and spatial resolution in TERS.** (a) Topography and (b) TERS map at the D band (1350cm^{-1}) emission of a multilayer GO-COOH sample. (c) TERS intensity profile along the white dashed line marked in b. (d) TERS spectra measured at the locations marked 1-4 in (b). Scale bars: 200 nm. Reproduced from Ref. [66]. (e) Topography image of Au(111) terraces on which R6G 10^{-4}M is adsorbed ($\Delta V = 0.05\text{V}$ tip positive, current set-point is 80pA). The gray line indicates the zone where the line profile plotted in (f) is drawn. (g) Simultaneous TERS image of the emission at 1524cm^{-1} . The gray line indicates the zone where the line profile plotted in (h) is drawn. (i) TERS spectra acquired in correspondence of the circled areas in (c) distant only one scanning step (10 nm). Reproduced from Ref. [62].

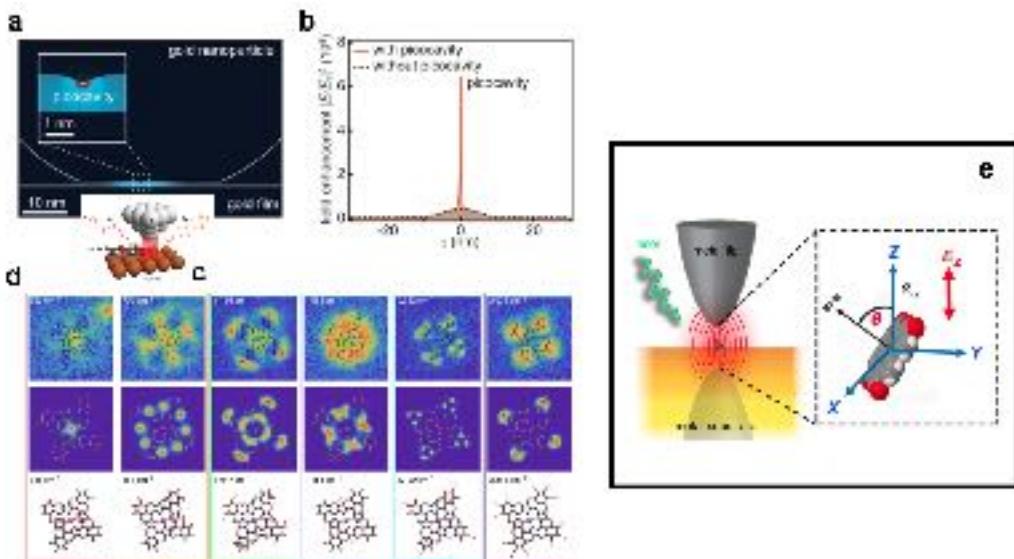


Figure 33.15: Sub-atomic resolution in TERS. (a) Near-field map of a faceted gold nanoparticle with and without atomic protrusion at $x = 0\text{ nm}$. The inset is an enlarged view of the pico-cavity showing sub-nanometer localization of optical field. (b) Near-field intensity across the gap. (From Ref. [70]) (c) Schematic of TERS of the CoTPP/Cu(100) system. (d) Top, experimental maps of vibrational normal modes of CoTPP ($29\text{ \AA} \times 29\text{ \AA}$, 64×64 pixels) with overlaid molecular frames. Middle, simulated maps of vibrational normal modes. Each image is $20\text{ \AA} \times 20\text{ \AA}$ in size. Bottom, assigned vibrational normal modes. Reproduced from Ref. [36]. (e) The incident far-field radiation induces a strong and confined plasmon field within the gap nano-cavity. The dashed box indicates the frame of reference of the molecule and the coordinates of the plasmonic field, in which is mainly confined and enhanced in the z direction. The black axis arrow (labeled *axis*) indicates the defined normal direction of the molecule. Reproduced from Ref. [28].

has also been used to spatially resolve ferroelectric domains in crystals [76] and to characterize artworks in-situ in a non-destructive way permitting the chemical identification of different pigments [77]. Particularly interesting is the application of the technique in life science [78]. The possibility to have spectral information open new routes for the study of biomolecules structure and conformation, closely related to their functionalities in living organisms permitting, for example, to distinguish toxic from non-toxic conformations of protein oligomers [79]. Biological systems like RNA, DNA, proteins, collagen, bacteria and viruses, were successfully studied [78, 80, 81, 82, 83] and imaged [84] by TERS, which can be also used to study *in-vivo* the protein uptake by yeast cell envelope [85].

Very recently the possibility to achieve sub-nm spatial resolution has been demonstrated, thanks to the formation of *pico-cavities* [70] between a plasmonic flat substrate and the atomic protrusions of a TERS tip (Fig. 33.15a). Pico-cavities push to the limit the optical coupling, confining the largest part of field enhancement in a region of few Ångstrom (Fig. 33.15b) [70]. This has allowed the feasibility of imaging single molecules [35] and even isolated atoms in a molecule [86] with sub-atomic resolution. Indeed, these results are possible only in Ultra High Vacuum and low temperature conditions, with a precise control of thermal drift and immobilization of the molecules on atomically flat substrates. In such extreme conditions, TERS offers the outstanding possibility to selectively image the different vibrational normal modes of a single molecule (Fig. 33.15c-d) [36].

TERS has the potential to probe the orientation of molecules in air [74, 87] and in liquid during electrochemical experiments [88], as well as to image the absorption geometry of 2D polymers, lipid membranes or even single molecules (see [89] for a review). Orientational information is encoded in the angular dependence of the Raman tensor when a molecule is rotated [4, 2]. The most straightforward way to get orientation information is, therefore, to compare the intensity ratio among selected Raman modes, featuring different symmetry, with predictions carried out within the Density Functional Theory ([89] and references therein). A more *phys-*

ical model can be derived assuming the configuration in Fig. 33.15e, in which an axial-symmetric molecule, oriented at an angle θ , lays in a tip-sample gap aligned in the z -direction. By making some reasonable assumptions on the z -orientation of the tip-enhanced excitation and re-radiated Raman fields, and using simple geometric rules to calculate the Raman tensor of the rotated molecule, the orientation-dependent TERS signal can be calculated analytically [87]. This model was used in STM-TERS experiments on azo-benzene chains of different lengths, to evidence the bending of the molecules in the gap when subject to electric fields [87], and to describe this phenomenon as an interplay between the electrostatic interaction of the molecular dipoles with the electric field in the gap, and a mechanical bending likely due to a physical contact between tip and molecular layer [74].

Bibliography

- [1] W. Demtröder. *Molecular physics: theoretical principles and experimental methods*. John Wiley & Sons, 2008.
- [2] B. S. Tsukerblat. *Group theory in chemistry and spectroscopy: A simple guide to advanced usage*. Courier Corporation, 2006.
- [3] A. Fadini and F.-M. Schnepel. Vibrational spectroscopy: methods and applications. *Ellis Horwood Ltd, Chichester, UK*, 1989.
- [4] D. A. Long. *The Raman effect: A unified treatment of the theory of Raman scattering by molecules*, volume 8. Wiley Chichester, 2002.
- [5] E. Le Ru and P. Etchegoin. *Principles of Surface-Enhanced Raman Spectroscopy: and related plasmonic effects*. Elsevier, 2008.
- [6] R. Pilot, R. Signorini, C. Durante, L. Orian, M. Bhamidipati, and L. Fabris. A review on surface-enhanced Raman scattering. *Biosensors*, 9:57, 2019.
- [7] E. C. Le Ru, E. Blackie, M. Meyer, and P. G. Etchegoin. Surface enhanced Raman scattering enhancement factors: A comprehensive study. *The Journal of Physical Chemistry C*, 111:13794–13803, 2007.
- [8] F. De Angelis, F. Gentile, F. Mecarini, G. Das, M. Moretti, P. Candeloro, M. L. Coluccio, G. Cojoc, A. Accardo, C. Liberale, R. P. Zaccaria, G. Perozziello, L. Tirinato, A. Toma, G. Cuda, R. Cingolani, and E. Di Fabrizio. Breaking the diffusion limit with super-hydrophobic delivery of molecules to plasmonic nanofocusing SERS structures. *Nature Photonics*, 5:682–687, 2011.
- [9] S. Nie and S. R. Emory. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *science*, 275:1102–1106, 1997.
- [10] B. Fazio, C. D’Andrea, F. Bonaccorso, A. Irrera, G. Calogero, C. Vasi, P. G. Gucciardi, M. Allegrini, A. Toma, D. Chiappe, C. Martella, and F. Buatier de Mongeot. Re-radiation enhancement in polarized surface-enhanced resonant Raman scattering of randomly oriented molecules on self-organized gold nanowires. *ACS Nano*, 5:5945–5956, 2011.
- [11] A. Otto. The ‘chemical’(electronic) contribution to surface-enhanced Raman scattering. *Journal of Raman Spectroscopy*, 36:497–509, 2005.
- [12] J. R. Lombardi and R. L. Birke. A unified view of surface-enhanced Raman scattering. *Accounts of chemical research*, 42:734–742, 2009.
- [13] S.-Y. Ding, E.-M. You, Z.-Q. Tian, and M. Moskovits. Electromagnetic theories of surface-enhanced Raman spectroscopy. *Chemical Society Reviews*, 46:4042–4076, 2017.
- [14] C. Noguez. Surface plasmons on metal nanoparticles: the influence of shape and physical environment. *The Journal of Physical Chemistry C*, 111:3806–3819, 2007.

- [15] S. Kessentini, D. Barchiesi, C. D'Andrea, A. Toma, N. Guillot, E. Di Fabrizio, B. Fazio, O. M. Marago, P. G. Gucciardi, and M. Lamy de la Chapelle. Gold dimer nanoantenna with slanted gap for tunable LSPR and improved SERS. *The Journal of Physical Chemistry C*, 118:3209–3219, 2014.
- [16] R. Gillibert, G. Balakrishnan, Q. Deshoules, M. Tardivel, A. Magazzù, M. G. Donato, O. M. Maragò, M. Lamy de la Chapelle, F. Colas, F. Lagarde, and P. G. Gucciardi. Raman tweezers for small microplastics and nanoplastics identification in seawater. *Environmental science & technology*, 53:9003–9013, 2019.
- [17] W. Xie and S. Schlücker. Medical applications of surface-enhanced Raman scattering. *Physical Chemistry Chemical Physics*, 15:5329–5344, 2013.
- [18] B. Fazio, C. D'Andrea, A. Foti, E. Messina, A. Irrera, M. G. Donato, V. Villari, N. Micali, O. M. Maragò, and P. G. Gucciardi. Sers detection of biomolecules at physiological ph via aggregation of gold nanorods mediated by optical forces and plasmonic heating. *Scientific reports*, 6:1–13, 2016.
- [19] M. Cottat, C. D'Andrea, R. Yasukuni, N. Malashikhina, R. Grinyte, N. Lidgi-Guigui, B. Fazio, A. Sutton, O. Oudar, N. Charnaux, V. Pavlov, A. Toma, P. G. Gucciardi, and M. Lamy de la Chapelle. High sensitivity, high selectivity SERS detection of MnSOD using optical nanoantennas functionalized with aptamers. *The Journal of Physical Chemistry C*, 119:15532–15540, 2015.
- [20] R. Tuma. Raman spectroscopy of proteins: from peptides to large assemblies. *Journal of Raman Spectroscopy*, 36:307–319, 2005.
- [21] T. J. Moore, A. S. Moody, T. D. Payne, G. M. Sarabia, A. R. Daniel, and B. Sharma. In vitro and in vivo SERS biosensing for disease diagnosis. *Biosensors*, 8:46, 2018.
- [22] L. Wang, J. Li, J. Pan, X. Jiang, Y. Ji, Y. Li, Y. Qu, Y. Zhao, X. Wu, and C. Chen. Revealing the binding structure of the protein corona on gold nanorods using synchrotron radiation-based techniques: understanding the reduced damage in cell membranes. *Journal of the American Chemical Society*, 135:17359–17368, 2013.
- [23] H. Liu, Z. Yang, L. Meng, Y. Sun, J. Wang, L. Yang, J. Liu, and Z. Tian. Three-dimensional and time-ordered surface-enhanced Raman scattering hotspot matrix. *Journal of the American Chemical Society*, 136:5332–5341, 2014.
- [24] A. Foti, C. D'Andrea, V. Villari, N. Micali, M. G. Donato, B. Fazio, O. M. Maragò, R. Gillibert, M. Lamy de la Chapelle, and P. G. Gucciardi. Optical aggregation of gold nanoparticles for SERS detection of proteins and toxins in liquid environment: Towards ultrasensitive and selective detection. *Materials*, 11:440, 2018.
- [25] S. Bernatova, M. G. Donato, J. Jezek, Z. Pilat, O. Samek, A. Magazzù, O. M. Marago, P. Zemanek, and P. G. Gucciardi. Wavelength-dependent optical force aggregation of gold nanorods for SERS in a microfluidic chip. *The Journal of Physical Chemistry C*, 123:5608–5615, 2019.
- [26] R. M. Stöckle, Y. D. Suh, V. Deckert, and R. Zenobi. Nanoscale chemical analysis by tip-enhanced Raman spectroscopy. *Chemical Physics Letters*, 318:131–136, 2000.
- [27] A. Hartschuh. Tip-enhanced near-field optical microscopy. *Angewandte Chemie International Edition*, 47:8178–8191, 2008.
- [28] F. Shao and R. Zenobi. Tip-enhanced Raman spectroscopy: principles, practice, and applications to nanospectroscopic imaging of 2d materials. *Analytical and bioanalytical chemistry*, 411:37–61, 2019.
- [29] L. Novotny and N. Van Hulst. Antennas for light. *Nature photonics*, 5:83–90, 2011.
- [30] L. Novotny and B. Hecht. *Principles of nano-optics*. Cambridge University Press, 2012.
- [31] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel. Surface studies by scanning tunneling microscopy. *Physical Review Letters*, 49:57, 1982.

- [32] G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Physical review letters*, 56:930, 1986.
- [33] E. Betzig, P. L. Finn, and J. S. Weiner. Combined shear force and near-field scanning optical microscopy. *Applied physics letters*, 60:2484–2486, 1992.
- [34] M. D. Sonntag, J. M. Klingsporn, L. K. Garibay, J. M. Roberts, J. A. Dieringer, T. Seideman, K. A. Scheidt, L. Jensen, G. C. Schatz, and R. P. Van Duyne. Single-molecule tip-enhanced Raman spectroscopy. *The Journal of Physical Chemistry C*, 116:478–483, 2012.
- [35] S. Jiang, Y. Zhang, R. Zhang, C. Hu, M. Liao, Y. Luo, J. Yang, Z. Dong, and J. G. Hou. Distinguishing adjacent molecules on a surface using plasmon-enhanced Raman scattering. *Nature nanotechnology*, 10:865–869, 2015.
- [36] J. Lee, K. T. Crampton, N. Tallarida, and V. A. Apkarian. Visualizing vibrational normal modes of a single molecule with atomically confined light. *Nature*, 568:78–82, 2019.
- [37] S. W. Hell. Far-field optical nanoscopy. *Science*, 316:1153–1158, 2007.
- [38] B. Hecht, B. Sick, U. P. Wild, V. Deckert, R. Zenobi, O. J. F. Martin, and D. W. Pohl. Scanning near-field optical microscopy with aperture probes: Fundamentals and applications. *The Journal of Chemical Physics*, 112:7761–7774, 2000.
- [39] P. G. Gucciardi, S. Trusso, C. Vasi, S. Patanè, and M. Allegrini. Near-field Raman spectroscopy and imaging. In *Applied Scanning Probe Methods V*, pages 287–329. Springer, 2007.
- [40] J. Wessel. Surface-enhanced optical microscopy. *JOSA B*, 2:1538–1541, 1985.
- [41] S. Patane, P. G. Gucciardi, M. Labardi, and M. Allegrini. Apertureless near-field optical microscopy. *Rivista del nuovo cimento*, 27:1–46, 2004.
- [42] F. Huth, A. Govyadinov, S. Amarie, W. Nuansing, F. Keilmann, and R. Hillenbrand. Nano-FTIR Absorption Spectroscopy of Molecular Fingerprints at 20 nm Spatial Resolution. *Nano Letters*, 12:3973–3978, 2012.
- [43] N. Kazemi-Zanjani, S. Vedraine, and F. Lagugné-Labarthet. Localized enhancement of electric field in tip-enhanced Raman spectroscopy using radially and linearly polarized light. *Optics Express*, 21:25271–25276, 2013.
- [44] L. Novotny, R. J. Sánchez, and X. S. Xie. Near-field optical imaging using metal tips illuminated by higher-order hermite–gaussian beams. *Ultramicroscopy*, 71:21–29, 1998.
- [45] B. Pettinger, K. F. Domke, D. Zhang, G. Picardi, and R. Schuster. Tip-enhanced Raman scattering: influence of the tip-surface geometry on optical resonance and enhancement. *Surface Science*, 603:1335–1341, 2009.
- [46] S. Kawata and V. M. Shalaev. *Tip enhancement*. Elsevier, 2011.
- [47] T.-X. Huang, S.-C. Huang, M.-H. Li, Z.-C. Zeng, X. Wang, and B. Ren. Tip-enhanced Raman spectroscopy: tip-related issues. *Analytical and bioanalytical chemistry*, 407:8177–8195, 2015.
- [48] L. G. Cançado, A. Jorio, A. Ismach, E. Joselevich, A. Hartschuh, and L. Novotny. Mechanism of near-field Raman enhancement in one-dimensional systems. *Physical Review Letters*, 103:186101, 2009.
- [49] R. V. Maximiano, R. Beams, L. Novotny, A. Jorio, and L. G. Cançado. Mechanism of near-field Raman enhancement in two-dimensional systems. *Physical Review B*, 85:235434, 2012.
- [50] N. Mauser and A. Hartschuh. Tip-enhanced near-field optical microscopy. *Chemical Society Reviews*, 43:1248–1262, 2014.
- [51] N. Hayazawa, Y. Saito, and S. Kawata. Detection and characterization of longitudinal field for tip-enhanced Raman spectroscopy. *Applied Physics Letters*, 85:6239–6241, 2004.

- [52] S. Quabis, R. Dorn, M. Eberler, O. Glöckl, and G. Leuchs. Focusing light to a tighter spot. *Optics communications*, 179:1–7, 2000.
- [53] D. Mehtani, N. Lee, R. D. Hartschuh, A. Kisliuk, M. D. Foster, A. P. Sokolov, and J. F. Maguire. Nano-raman spectroscopy with side-illumination optics. *Journal of Raman Spectroscopy*, 36:1068–1075, 2005.
- [54] A. Downes, D. Salter, and A. Elfick. Finite element simulations of tip-enhanced Raman and fluorescence spectroscopy. *The Journal of Physical Chemistry B*, 110:6692–6698, 2006.
- [55] M. A. Lieb and A. J. Meixner. A high numerical aperture parabolic mirror as imaging device for confocal microscopy. *Optics express*, 8:458–474, 2001.
- [56] J. Steidtner and B. Pettinger. High-resolution microscope for tip-enhanced optical processes in ultrahigh vacuum. *Review of Scientific Instruments*, 78:103104, 2007.
- [57] Z. Yang, J. Aizpurua, and H. Xu. Electromagnetic field enhancement in tera configurations. *Journal of Raman Spectroscopy*, 40:1343–1348, 2009.
- [58] R. D. Rodriguez, E. Sheremet, S. Müller, O. D. Gordan, A. Villabona, S. Schulze, M. Hietschold, and D. R. T. Zahn. Compact metal probes: A solution for atomic force microscopy based tip-enhanced Raman spectroscopy. *Review of scientific instruments*, 83:123708, 2012.
- [59] M. D. McMahon, R. Lopez, H. M. Meyer, L. C. Feldman, and R. F. Haglund. Rapid tarnishing of silver nanoparticles in ambient laboratory air. *Applied Physics B*, 80:915–921, 2005.
- [60] F. Bonaccorso, G. Calogero, G. Di Marco, O. M. Marago, P. G. Gucciardi, U. Giorgianni, K. Channon, and G. Sabatino. Fabrication of gold tips by chemical etching in aqua regia. *Review of Scientific Instruments*, 78:103702, 2007.
- [61] B. Ren, G. Picardi, and B. Pettinger. Preparation of gold tips suitable for tip-enhanced Raman spectroscopy and light emission by electrochemical etching. *Review of Scientific Instruments*, 75:837–841, 2004.
- [62] A. Foti, F. Barreca, E. Fazio, C. D’Andrea, P. Matteini, O. M. Maragò, and P. G. Gucciardi. Low cost tips for tip-enhanced Raman spectroscopy fabricated by two-step electrochemical etching of 125 μm diameter gold wires. *Beilstein journal of nanotechnology*, 9:2718–2729, 2018.
- [63] M. Lopes, T. Toury, M. Lamy De La Chapelle, F. Bonaccorso, and P. G. Gucciardi. Fast and reliable fabrication of gold tips with sub-50 nm radius of curvature for tip-enhanced Raman spectroscopy. *Review of Scientific Instruments*, 84:073702, 2013.
- [64] S. Berweger, J. M. Atkin, R. L. Olmon, and M. B. Raschke. Adiabatic tip-plasmon focusing for nano-Raman spectroscopy. *The Journal of Physical Chemistry Letters*, 1:3427–3432, 2010.
- [65] F. De Angelis, G. Das, P. Candeloro, M. Patrini, M. Galli, A. Bek, M. Lazzarino, I. Maksymov, C. Liberale, L. C. Andreani, and E. Di Fabrizio. Nanoscale chemical mapping using three-dimensional adiabatic compression of surface plasmon polaritons. *Nature Nanotechnology*, 5:67–72, 2010.
- [66] W. Su, N. Kumar, A. Krayev, and M. Chaigneau. In situ topographical chemical and electrical imaging of carboxyl graphene oxide at the nanoscale. *Nature communications*, 9:1–7, 2018.
- [67] M. M. Sartin, H.-S. Su, X. Wang, and B. Ren. Tip-enhanced Raman spectroscopy for nanoscale probing of dynamic chemical systems. *The Journal of Chemical Physics*, 153:170901, 2020.
- [68] M. Liao, S. Jiang, C. Hu, R. Zhang, Y. Kuang, J. Zhu, Y. Zhang, and Z. Dong. Tip-enhanced Raman spectroscopic imaging of individual carbon nanotubes with subnanometer resolution. *Nano letters*, 16:4040–4046, 2016.
- [69] A. Foti, S. Venkatesan, B. Lebental, G. Zucchi, and R. Ossikovski. Comparing commercial metal-coated AFM tips and home-made bulk gold tips for tip-enhanced Raman spectroscopy of polymer functionalized multiwalled carbon nanotubes. *Nanomaterials*, 12:451, 2022.

- [70] F. Benz, M. K. Schmidt, A. Dreismann, R. Chikkaraddy, Y. Zhang, A. Demetriadou, C. Carnegie, H. Ohadi, B. De Nijs, R. Esteban, J. Aizpurua, and J. J. Baumberg. Single-molecule optomechanics in “picocavities”. *Science*, 354:726–729, 2016.
- [71] M. Böhmler, Z. Wang, A. Myalitsin, A. Mews, and A. Hartschuh. Optical imaging of CdSe nanowires with nanoscale resolution. *Angewandte Chemie International Edition*, 50:11536–11538, 2011.
- [72] L. G. Cançado, A. Hartschuh, and L. Novotny. Tip-enhanced Raman spectroscopy of carbon nanotubes. *Journal of Raman Spectroscopy*, 40:1420–1426, 2009.
- [73] F. Pashaee, F. Sharifi, G. Fanchini, and F. Lagugné-Labarthet. Tip-enhanced Raman spectroscopy of graphene-like and graphitic platelets on ultraflat gold nanoplates. *Physical Chemistry Chemical Physics*, 17:21315–21322, 2015.
- [74] A. Foti, C. Toccafondi, and R. Ossikovski. Study of the molecular bending in azobenzene self-assembled monolayers observed by tip-enhanced Raman spectroscopy in scanning tunneling mode. *The Journal of Physical Chemistry C*, 123:26554–26563, 2019.
- [75] H. K. Wickramasinghe, M. Chaigneau, R. Yasukuni, G. Picardi, and R. Ossikovski. Billion-fold increase in tip-enhanced Raman signal. *ACS nano*, 8:3421–3426, 2014.
- [76] S. Berweger, C. C. Neacsu, Y. Mao, H. Zhou, S. S. Wong, and M. B Raschke. Optical nanocrystallography with tip-enhanced phonon Raman spectroscopy. *Nature nanotechnology*, 4:496–499, 2009.
- [77] D. Kurouski, S. Zaleski, F. Casadio, R. P. Van Duyne, and N. C. Shah. Tip-enhanced Raman spectroscopy (TERS) for in situ identification of indigo and iron gall ink on paper. *Journal of the American Chemical Society*, 136:8677–8684, 2014.
- [78] G. Sharma, T. Deckert-Gaudig, and V. Deckert. Tip-enhanced Raman scattering—targeting structure-specific surface characterization for biomedical samples. *Advanced Drug Delivery Reviews*, 89:42–56, 2015.
- [79] C. D’Andrea, A. Foti, M. Cottat, M. Banchelli, C. Capitini, F. Barreca, C. Canale, M. de Angelis, A. Relini, O. M. Marago, R. Pini, F. Chiti, P. G. Gucciardi, and P. Matteini. Nanoscale discrimination between toxic and nontoxic protein misfolded oligomers with tip-enhanced Raman spectroscopy. *Small*, 14:1800890, 2018.
- [80] B.-S. Yeo, J. Stadler, T. Schmid, R. Zenobi, and W. Zhang. Tip-enhanced Raman spectroscopy—its status, challenges and future directions. *Chemical Physics Letters*, 472:1–13, 2009.
- [81] M. Lucas and E. Riedo. Combining scanning probe microscopy with optical spectroscopy for applications in biology and materials science. *Review of Scientific Instruments*, 83:061101, 2012.
- [82] B. R. Wood, M. Asghari-Khiavi, E. Bailo, D. McNaughton, and V. Deckert. Detection of nano-oxidation sites on the surface of hemoglobin crystals using tip-enhanced Raman scattering. *Nano letters*, 12:1555–1560, 2012.
- [83] K. F. Domke, D. Zhang, and B. Pettinger. Tip-enhanced Raman spectra of picomole quantities of DNA nucleobases at Au (111). *Journal of the American Chemical Society*, 129:6708–6709, 2007.
- [84] M. Paulite, C. Blum, T. Schmid, L. Opilik, K. Eyer, G. C. Walker, and R. Zenobi. Full spectroscopic tip-enhanced Raman imaging of single nanotapes formed from β -amyloid (1–40) peptide fragments. *ACS nano*, 7:911–920, 2013.
- [85] D. Naumenko, V. Snitka, E. Serviene, I. Bruzaite, and B. Snopok. In vivo characterization of protein uptake by yeast cell envelope: single cell AFM imaging and μ -tip-enhanced Raman scattering study. *Analyst*, 138:5371–5383, 2013.
- [86] N. Tallarida, J. Lee, and V. A. Apkarian. Tip-enhanced Raman spectromicroscopy on the angstrom scale: Bare and CO-terminated Ag tips. *ACS nano*, 11:11393–11401, 2017.

- [87] C. Toccafondi, G. Picardi, and R. Ossikovski. Molecular bending at the nanoscale evidenced by tip-enhanced Raman spectroscopy in tunneling mode on thiol self-assembled monolayers. *The Journal of Physical Chemistry C*, 120:18209–18219, 2016.
- [88] N. Martin Sabanes, T. Ohto, D. Andrienko, Y. Nagata, and K. F. Domke. Electrochemical teras elucidates potential-induced molecular reorientation of adenine/au(111). *Angewandte Chemie International Edition*, 56:9796–9801, 2017.
- [89] D. Mrdenovic, Z.-F. Cai, Y. Pandey, G. L. Bartolomeo, R. Zenobi, and N. Kumar. Nanoscale chemical analysis of 2d molecular materials using tip-enhanced Raman spectroscopy. *Nanoscale*, 15:963–974, 2023.

Chapter 34

Experimental Active Nematics

BERTA MARTÍNEZ-PRAT, JORDI IGNÉS-MULLOL, FRANCESC SAGUÉS

This Chapter is devoted to active fluids that are prepared from dispersions of filamentary proteins polymerized from *tubulin*, with their complementary motor protein *kinesin* (Fig. 34.1). Mimicking the cytoskeletal milieu, these preparations are driven far from equilibrium, and thus kept active, when fueled with adenosine triphosphate (ATP). Central attention will be devoted to their performance as interfacial dense suspensions to highlight their *nematic symmetry*, which qualifies them as *active nematic fluids*, or in short *active nematics*. The first section contains a detailed account of the experimental protocol we follow to prepare the interfaced active nematic. An active nematic is basically characterized in terms of two fields: the *orientational field*, intrinsic to any nematic preparation, and the *velocity field*, which follows from its active nature. The second section is dedicated to review the basic experimental techniques employed to extracting these fields. Finally, the last section reports on the way we perform the analysis of the data to account for characteristic observables of the

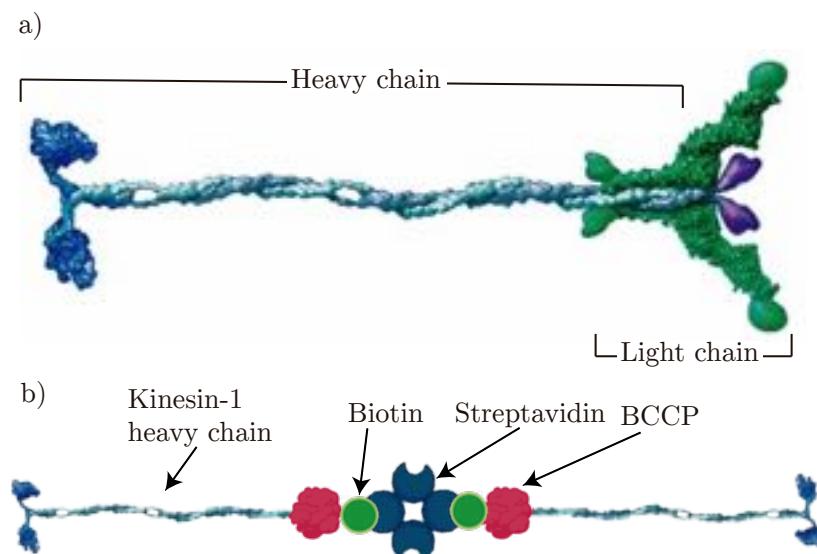


Figure 34.1: **Kinesin-1 and K401-BIO-H6.** (a) Sketch of kinesin-1 where the two chains of the molecular motor are indicated. The heavy chain allows the protein to move and the light chain can attach to cargoes to actively transport them inside the cell. (b) Cluster of two kinesins. The genetically-engineered kinesin is formed by the heavy chain of kinesin-1 (kinesin-1 truncated at the residue 401) and the biotin carboxyl carrier protein (BCCP) which binds to one biotin. A streptavidin molecule then links two kinesins through the biotins. The sketch of kinesin-1 in panel (a) is reproduced from Ref. [1]. The BCCP-biotin-streptavidin complex was created with [BioRender.com](#).

system, principally those related with its flowing nature, i.e., the spectra of kinetic energy, the size distribution of vortices, or the correlation function of the velocity and vorticity fields.

34.1 Preparation of a 2D active nematic

The preparation of the active nematic, in the way we work with it here, requires following a precise protocol that includes a number of successive steps, from dealing with the handling of the biological materials to their assembly as an active gel. The protocol is finalized when the active material is depleted to a stabilized water–oil interface. These steps are described in detail on what follows.

34.1.1 Kinesin expression

The kinesin used in the experiments discussed here is the K401-BIO-H6, which is a dimeric kinesin containing the first 401 residues of kinesin-1 heavy chain (Fig. 34.1a) of *Drosophila melanogaster* and fused to a region of *Escherichia coli* biotin carboxyl carrier protein (BCCP) and six histidines to enable protein purification [2]. The BCCP domain allows the formation of kinesin clusters because it can attach to a biotin molecule, which can in turn link to streptavidin molecules. Each streptavidin, as a tetrameric protein [3] with a molecular weight ~ 52 kDa, can bind to up to four biotins (Fig. 34.1b).

The K401-BIO-H6 protein is expressed with *Escherichia coli* (BL21(DE3)pLysS cells, resistant to chloramphenicol and transformed with the pWC2 plasmid (Addgene 15960) from Gelles Laboratory (Brandeis University) that contains the gene of the protein of interest and confers resistance against ampicillin [2].

Briefly, we harvest the *E. coli* in Luria-Bertani (LB) broth (Sigma, L3022) with $2.5 \mu\text{g mL}^{-1}$ of chloramphenicol, $50 \mu\text{g mL}^{-1}$ of ampicillin, and $24.4 \mu\text{g mL}^{-1}$ of biotin at 37°C and 120 rpm until reaching OD₆₀₀ (optical density at 600 nm) equal to 0.6, when 1 mM of Isopropyl β -D-1-thiogalactopyranoside (IPTG) is added to induce protein over-expression. Cells are left overnight at 22°C and 140 rpm. The next day, the culture is centrifuged at 4,000 rpm and 4°C for 30 min. Pellets are re-suspended with HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer (50 mM HEPES, 4 mM MgCl₂, 10 mM β -mercaptoethanol and 50 μM of ATP at pH 7.2) supplemented with protease inhibitor cocktails (PIC) and phenylmethylsulfonyl fluoride (PMSF).

The suspension is left at -80°C for 30 min. Then, after adding 1 mg mL⁻¹ of lysozyme, it is sonicated with a probe-type sonicator to promote cell lysis. Solid residues are removed via ultra-centrifugation at 20,000 rpm for 20 min at 4°C . The protein is then loaded into a 1 mL nickel column (1 mL HiFiQ® Ni-Nitrilotriacetic acid fast protein liquid chromatography column), washed with 20 mM imidazole in HEPES buffer, and eluted with 500 mM imidazole in HEPES buffer. A PD10 desalting column (GE Healthcare, GE17-0851-01) is used to exchange the buffer and thus remove the imidazole. We measure the resulting protein concentration by means of UV-Vis spectrophotometry (A_{280} , $\epsilon = 30370^{-1} \text{ cm}^{-1}$). The kinesin is finally stored with 40 % (w/w) of sucrose at -80°C after flash freezing with liquid nitrogen.

34.1.2 Microtubules polymerization

Microtubules are polymerized at 37°C for 30 min in M2B buffer (80 mM piperazine-N,N'-bis(2-ethanesulfonic acid), 1 mM egtazic acid, 2 mM MgCl₂) (Sigma, P1851, E3889 and M4880, respectively) with α and β tubulin from bovine brain (obtained, in our case, from Brandeis University Biological Materials facility, Waltham, MA; alternatively, lyophilized tubulin can be purchased from Cytoskeleton, Inc.) and Guanosine-5'-[(α,β) -methyleno]triphosphate (GMPCPP) (Jena Biosciences, NU-405), a non-hydrolyzable analogue of Guanosine Triphosphate (GTP). GMPCPP catalyzes the polymerization of microtubules at a similar rate to the one of GTP. Nevertheless, the depolymerization rate of the former is much smaller than that of the latter. Hence, with GMPCPP, it is possible to obtain stable microtubules with a fixed length [4]. By controlling the concentration of GMPCPP, we can obtain microtubules with a mean length of $\sim 1.5 \mu\text{m}$. To enable the imaging of the active gel through fluorescence microscopy, part of the tubulin is labeled with Alexa-647 (Sigma, A20006). Microtubules are stored in small aliquots ($\sim 2 \mu\text{L}$) at -80°C .

34.1.3 Active gel preparation

To assemble the active gel, we use different stock solutions (listed in Table 34.1), which we can store for several months under the appropriate conditions (also included in Table 34.1). The stock solutions that need to be stored at -20°C or -80°C are split into smaller aliquotes with a volume $\sim 10\%$ larger than the volume used to prepare the KSV (Kinesin-Streptavidin complex), the PS (Pre-solution), and the final active suspension. All the stocks are prepared either in M2B buffer, or phosphate buffer, or are kept in the aqueous solution as purchased. The pH of all the stock solutions is adjusted with either 1M HCl (if it is too basic, $\text{pH} >$ desired pH) or with 1M KOH (if it is too acid, $\text{pH} <$ desired pH). For small volumes of stock solutions, it is difficult to use a pH meter. Thus, we use pH indicator strips with a pH range between 6 and 8. It is important to use such a small range of pH to have enough sensitivity to adjust properly the pH at ~ 6.8 because small changes in the pH can significantly impact the function of the proteins.

Compound	Concentration	Buffer	Storing conditions
ATP	50 mM	M2B (pH 6.8)	-20°C
Catalase	3.5 mg mL $^{-1}$	20 mM K ₂ HPO ₄ (pH 7.5)	-20°C
DTT	500 mM	M2B (pH 6.8)	-20°C
Glucose	300 mg mL $^{-1}$	20 mM K ₂ HPO ₄ + 70 mM KCl (pH 7.2)	-20°C
Glucose oxidase	20 mg mL $^{-1}$	20 mM K ₂ HPO ₄ (pH 7.4)	-20°C
GMPCPP	10 mM	Aqueous solution, as provided by the manufacturer	-20°C
MgCl ₂	67 mM	M2B (pH 6.8)	4°C
microtubules	8 mg mL $^{-1}$	M2B (pH 6.8)	-80°C
PEG, 20 kDa	12% (w/w)	M2B (pH 6.8)	-20°C
PEP	200 mM	M2B (pH 6.8)	-20°C
PKLDH	917 units mL $^{-1}$ PK	Aqueous buffered glycerol solution, as provided by the manufacturer	-20°C
Pluronic	17% (w/w)	M2B (pH 6.8)	4°C
Streptavidin	0.352 mg mL $^{-1}$	M2B (pH 6.8)	-20°C
Trolox	20 mM	20 mM Phosphate (pH 7.48)	-20°C

Table 34.1: **Stock solutions for the preparation of the active gel.** List of the solutions used to prepare the active solution. See text for the meaning of acronyms.

We prepare the active gel in four main steps that are explained in detail on what follows. The diagram in Fig. 34.4 shows schematically the protocol we follow to prepare the active gel.

Step 1 Preparation of Kinesin-Streptavidin complex (KSV)

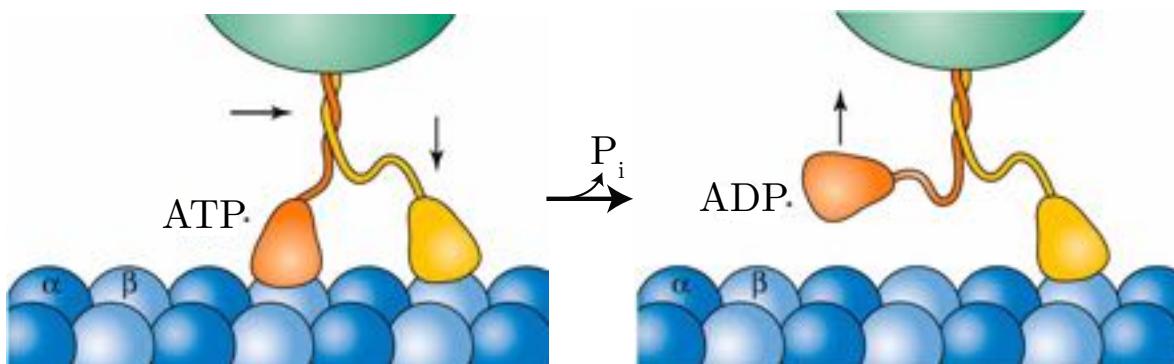


Figure 34.2: **Kinesins obtain energy from ATP and release ADP and phosphoric acid.** Kinesins obtain energy from adenosine triphosphate (ATP) and release adenosine diphosphate (ADP) and phosphoric acid. The chemical energy stored in the phosphate-phosphate bond fuels kinesins, which take the ATP, make one step forward, and release ADP and phosphoric acid (P_i). Reproduced from Ref. [5].

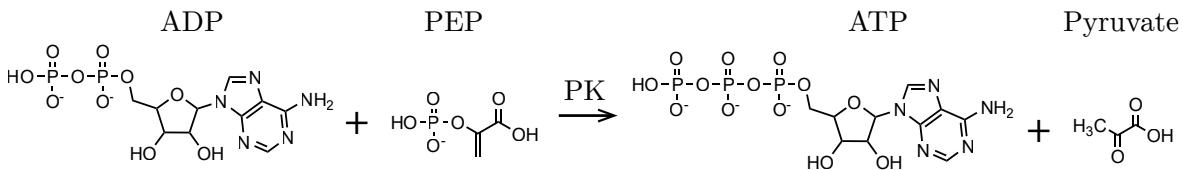


Figure 34.3: ATP regenerating system. ADP is converted back to ATP thanks to PK that catalyzes the reaction of ADP with Phosphoenol pyruvate (PEP).

K401-BIO-H6 motors (kinesins) are mixed at a 2:1 stoichiometric ratio with tetrameric streptavidin (In-vitrogen, 43-4301) and incubated at 4°C for 30 min. This makes the KSV suspension. Here, we must remark that the expression of kinesin as described in Subsection 34.1.1 can produce non-active proteins. As a consequence, the activity of the active gel can be lower than the one expected for the concentration obtained through UV-Vis spectroscopy. One could try to do activity assays to extract the real activity of the protein suspension. Nevertheless, we do it empirically. Starting from the theoretical concentration of kinesin, we follow the protocol here described to prepare the active gel and screen different concentrations of molecular motors keeping the other compound concentrations constant until we obtain a high activity active gel without seeing cross-linking. At low concentrations of kinesins, streptavidin proteins are only bound to one molecular motor, giving a low activity active gel. As the concentration is increased, the activity rises until a point where streptavidins start being bound to more than two kinesins. Consequently, activity starts dropping, and crosslinking can be observed. Maximum activity is reached at a kinesin/streptavidin ratio of ~1:2 [6]. In our case, we find a good preparation to be: 5 μL stock Kinesin + 1 μL stock streptavidin solution + 0.5 μL stock DTT (Dithiothreitol) diluted 100 times.

Step 2 Preparation of the pre-solution (PS)

To prepare the PS suspension, we mix the non-adsorbing polymer Polyethylene glycol, PEG (20kDa) (Sigma, 95172) to promote the aggregation of the microtubules into bundles through depletion forces, ATP to fuel the kinesins (Fig. 34.2), and an ATP-regenerating system (PK/LDH - Pyruvate Kinase/Lactic Dehydrogenase - and phosphoenol-pyruvate) (Fig. 34.3) to maintain the ATP concentration constant. The presence of this regenerating system does not limit the rate of ATP hydrolysis by kinesins [7]. In addition to such compounds, we also add MgCl₂ to reach an adequate strength force and anti-oxidants (Trolox, glucose, glucose oxidase, catalase, and DTT (Sigma, 238813, G8270, G2133, C1345, 43815, respectively)) to ensure the correct functionality of the proteins by preventing photo-bleaching, the formation of sulfur bonds, and the oxidation of the proteins by oxygen species. Finally, in the cases when we want to prepare an active nematic that forms at an oil–water interface, we supplement the microtubules aqueous suspension with the PEGylated surfactant pluronic (Sigma P2443). This compound assures the bio-compatibility of the proteins with the interface. The volumes of stock solutions we mix are: 8 μL PEG + 8 μL PEP (phosphoenolpyruvate) + 6 μL Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) + 2.9 μL MgCl₂ + 1.7 μL ATP + 1.7 μL PKLDH (Pyruvate Kinase/Lactic Dehydrogenase) + 1.5 μL Pluronic + 0.66 μL Catalase + 0.66 μL Glucose + 0.66 μL DTT + 0.66 μL Glucose Oxidase.

Step 3 Mix of KSV+PS+M2B

Once we have waited for 30 minutes to ensure the formation of the kinesin clusters, we can mix the KSV with the PS. We supplement this mixture with M2B to attain suitable concentrations. For the kinesin we have expressed, we mix 10 μL PS + 2.75 μL KSV + 2.25 μL M2B. These ratios should be adapted to the activity of the kinesin.

Step 4 Preparation of the active solution or active gel

Finally, we mix a small volume of the motors and ATP solution with the microtubules at a ratio of 1:5. Depending on the sample we want to make (Subsection 34.1.4), we prepare a volume of 2 μL or 8 μL. With this mixture, the microtubules self-assemble into a percolating active network.

The final concentrations of all compounds are listed in Table 34.2.

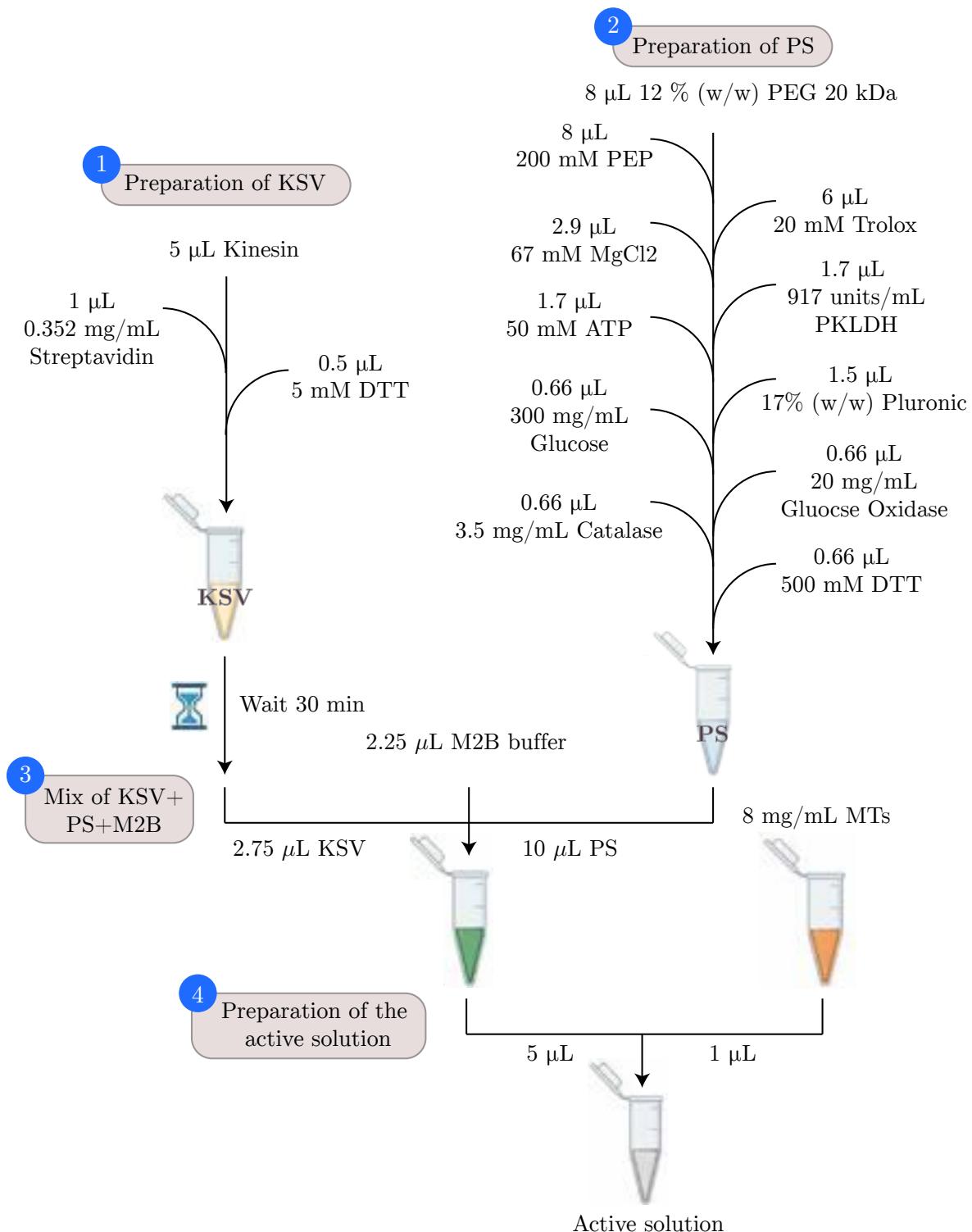


Figure 34.4: **Protocol diagram of the preparation of the active gel.** All the volumes are of the stock solutions. See text for acronyms.

Active nematic composition			
Compound	Concentration	Compound	Concentration
Streptavidin	0.16 μM	PK/LDH	27 IU mL^{-1}
Kinesin	0.32 μM	Pluronic	0.44% (w/w)
DTT	5.8 mM	Glucose	3.4 mg mL^{-1}
PEG (20kDa)	1.6% w/w	Catalase	0.040 mg mL^{-1}
PEP	27 mM	Glucose oxidase	0.23 mg mL^{-1}
Trolox	2.1 mM	microtubules	1.3 mg mL^{-1}
MgCl_2	3.3 mM	GMPCPP	0.1 mM
ATP	1.5 mM		

Table 34.2: **Typical composition of the active nematic.** Typical final concentrations of active nematic solutions.

34.1.4 Experimental setup for the assembly of the 2D active nematic

In the presence of an oil–water interface, the network of active bundled-filaments forms a 2D active layer with nematic order, the so-called active nematic. This process is driven by depletion forces due to the PEG that promotes the adsorption of the microtubule bundles onto the fluid interface. To obtain such interface, we use two different approaches depending on the experimental requirements.

The first one consists of a chamber built between two glass slides with two sheets of double-sided tape (thicknesses from 25 to 100 μm , 3M) as spacers. One of the slides is functionalized with a polyacrylamide brush (Subsection 34.1.5) to hydrophylize the substrate and prevent the proteins from sticking to the glass and denaturalizing. The other slide is functionalized with Aquapel (Subsection 34.1.5), obtaining in this way a very hydrophobic surface (Figs. 34.5a and 34.5b). A fluorinated oil (HFE 7500, 3M Novec 7500 Engineered Fluid) with 2% fluorosurfactant (008-FluoroSurfactant, RAN Biotechnology, or, alternatively, Fluosurf from EmulSEO) is then flowed into the chamber (Fig. 34.5c), followed by the microtubules aqueous suspension that displaces the oil (Fig. 34.5d). However, due to the difference in the hydrophobicity of the two slides, a stable thin oil layer remains (Fig. 34.5e).

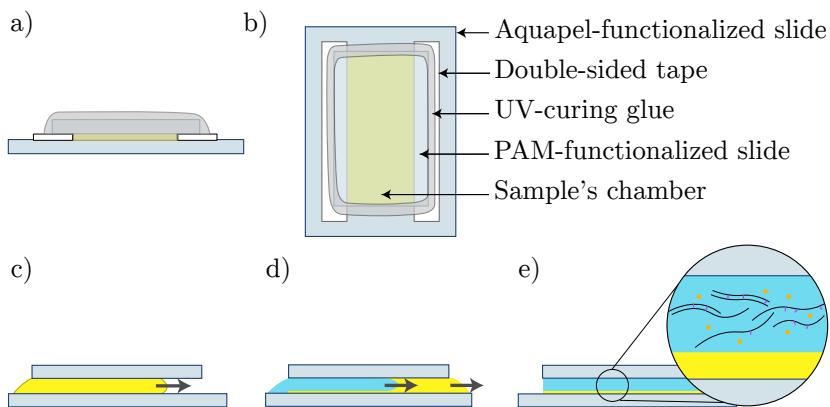


Figure 34.5: **Flow-cell experimental setup.** (a) Front and (b) top view of a sketched flow cell. (c)–(e) Filling of the cell. (c) The oil is first flowed through the chamber and then (d) replaced by the aqueous suspension of proteins — (e) a thin oil layer remains wetting the hydrophobic slide (coated with Aquapel).

The alternative approach consists of the assembly of an open cell with the sample contained inside a PDMS (polydimethylsiloxane) block with a well in the center. To prepare the block, we cure PDMS at 70°C for at least 4 hours inside a 3D-printed poly-lactic acid (PLA) mold (Fig. 34.6). Afterwards, we glue the block onto a hydrophilic and bioinert PAM (polyacrylamide) coated glass with a UV-curing adhesive (Norland, NOA81). We finally introduce in the pool 1.5 μL of the aqueous suspension and 300 μL of silicon oil (Rhodorsil Oils 47 from Bluestar Silicones) on top of it. Since the active nematic needs to be protected from the air, we use an

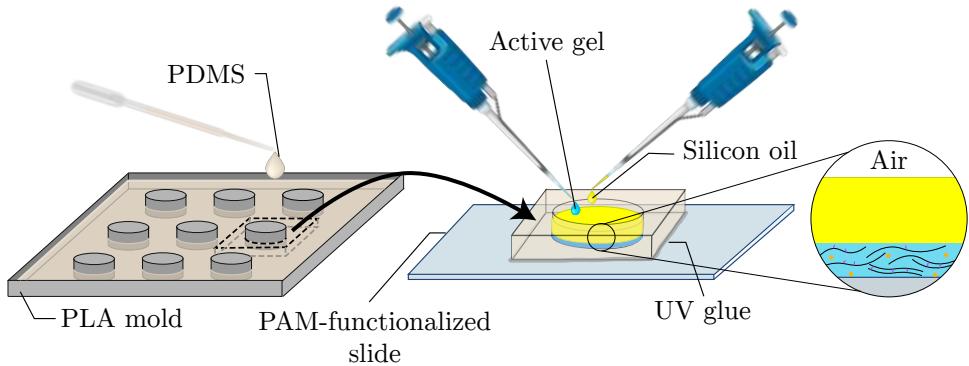


Figure 34.6: Open cell experimental setup The PDMS is polymerized within a PLA mold to create pools. A block containing a well is then cut and glued onto a PAM-functionalized glass slide using UV-curing adhesive. The microtubule-aqueous suspension (active gel) is introduced within the pool and covered by the silicon oil that is in contact with the air and with the active gel droplet. The latter interface is where the microtubules condense and form the active nematic.

oil less dense than water, thus gravity forces drag the aqueous suspension to the bottom leaving the oil layer covering it. After $\sim 20\text{min}$ the active nematic is already formed.

34.1.5 Glass-slides functionalization

Proteins tend to stick onto glass. Hence, it is imperative to functionalize glass slides and make them biocompatible.

Hydrophilic glass slides using polyacrylamide brush

One of the most common surface treatments for biological samples is the attachment of a polyacrylamide brush onto the substrate to avoid the adsorption of proteins thanks to steric repulsion. The coating of the slides is prepared by first cleaning and activating the glasses with an alkaline solution (0.1 M NaOH). Then, we silanize the glasses inside an acid ethanolic solution of 3-(trimethoxysilyl)propyl methacrylate (Sigma, 440159) to create polymerization seeds. The slides are subsequently rinsed with ethanol followed by Milli-Q water and then immersed in a degassed 2%(v/v) acrylamide solution with the initiator ammonium persulfate (APS, Sigma, A3678) and the catalyst N,N,N',N'-Tetramethylethylenediamine (TEMED, Sigma, T9281) for at least 2 hours. The substrates are stored in the acrylamide solution until their use.

Hydrophobic glass slides

Hydrophobic glass slides are obtained with Aquapel® functionalization. We place a clean glass slide to be functionalized on top of a glass slide with a small volume ($\sim 20\text{ }\mu\text{L}$) droplet of Aquapel. We let it sit for 30 seconds and then rinse the glass slide with Milli-Q water.

34.1.6 Active nematics for particle image velocimetry

To extract the flow field of the active nematic through particle image velocimetry (PIV) (Subsection 34.2.1), we use unlabeled microtubules and dope them with fluorescent microtubules (200:1) to obtain a speckled pattern. Both types of microtubules are prepared as explained above, but using either non-fluorescent or highly-fluorescent tubulin ($\sim 75\%$ labeled tubulin), respectively. We remark that multiple freezing and unfreezing cycles of the microtubules induces their aggregation (Fig. 34.7). This clustering can be useful when working at small magnifications, when more light intensity is required, although this may compromise the spatial resolution.

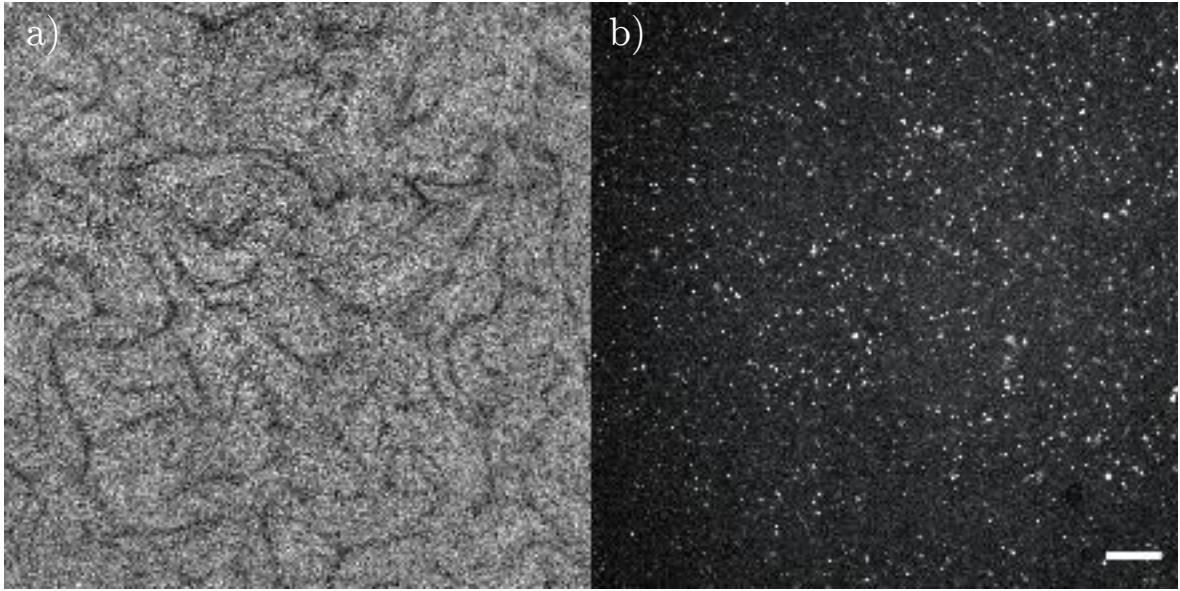


Figure 34.7: **Fluorescent microtubules tend to aggregate during freezing/unfreezing cycles.** Fluorescence micrographs of the active nematic using highly fluorescent microtubules that have been frozen (a) once and (b) twice. Scale bar, 100 μm .

34.2 Imaging and data acquisition

The relevant observables to analyze the structure and dynamics of the active nematic layers are the orientational field (local orientation of the active filaments) and the velocity field. These fields are extracted through suitable processing of the fluorescence images.

34.2.1 Particle image velocimetry

PIV is a non-invasive technique used to quantitatively determine the velocity field of a fluid [8, 9, 10, 11]. The idea of this technique is to seed the fluid with particles to create a speckled pattern in an otherwise homogeneous media. The particles are then tracked to extract the flows within the fluid. These tracers must be small enough and have density similar to the fluid's to assure they follow the streams at the same velocity as the fluid. In our experiments, we use fluorescent-labeled microtubules (Subsection 34.1.6) that we follow using fluorescence microscopy (Fig. 34.8). To extract the flows, fluorescence images are divided into sub-areas called interrogation windows, and the most probable displacement of the particles within an interrogation window is then determined by cross-correlating a pair of subsequent interrogation windows A and B with sizes $I \times J$ and $M \times N$, respectively:

$$C(m, n) = \sum_{i=1}^I \sum_{j=1}^J A(i, j)B(i - m, j - n) \quad (34.1)$$

with $-(M - 1) \leq m \leq I - 1$ and $-(N - 1) \leq n \leq J - 1$. Cross-correlation gives the degree of matching between two matrices (in this case two interrogation windows) after one of them is spatially shifted. Therefore, the position of the peak in $C(m, n)$ corresponds to the most probable displacement of the particles from A to B . Note that the result of the PIV is highly sensitive to the interrogation window size.

In practice, the computation of cross-correlation in the real space using Eq. (34.1) is highly expensive computationally speaking. Hence, $C(m, n)$ is usually obtained in the Fourier space by applying the cross-correlation theorem [12].

For our analysis, we use the off-the-shelf software PIVlab implemented in Matlab [9, 10]. Post-processing of the PIV results is also carried out with such program, which can filter out the outlying data and smooth the flow field.

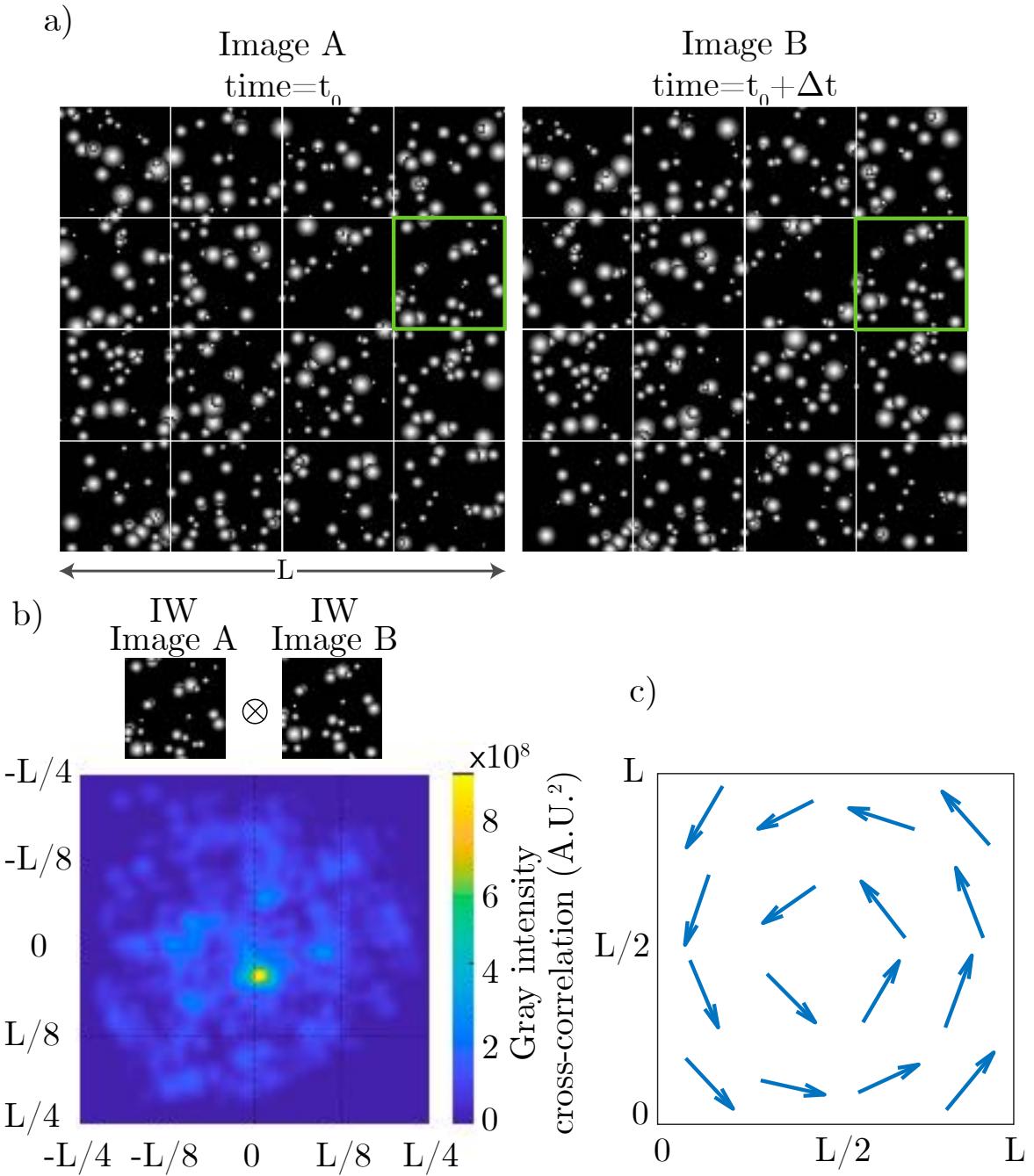


Figure 34.8: **The PIV algorithm.** (a) Synthetic image divided into 16 interrogation windows. (b) Cross-correlation calculation of the interrogation window framed in green in a. (c) Resultant flow field of the image in a. [In a, the images cover part of the text]

Depending on the fluorescence images, the outcome from the PIV analysis can be very noisy. To ensure that the obtained velocity field is the result of the signal of the fluorescent microtubules and not of the background noise, we calculate the temporal autocorrelation of the computed velocity field (Subsection 34.3.6). If the correlation drastically drops from one frame to the next, this suggests that the resulting velocity field is too noisy and further filtering of the fluorescence images should be performed.

34.2.2 Director determination

In order to obtain the nematic orientation of the active nematic, we use the method proposed by Ellis *et al.* [13, 14] from fluorescence microscopy images. In this case, the director is obtained using *coherent-enhanced diffusion filtering* (CEDF) [15], a technique employed to complete and enhance interrupted lines or flow-like structures, like the ones in fingerprints or the ones formed by the fluorescent microtubules in the active nematic. To this aim, CEDF infers the direction with the weakest spatial intensity fluctuations at a pixel level, which for the active nematic corresponds to the local alignment of the microtubules.

First of all, a Gaussian blur filter of standard deviation σ and side length $6\sigma - 1$ is applied to a raw image I to get the blurry image I_σ . Then, we obtain the tensor

$$(\nabla I_\sigma)(\nabla I_\sigma)^T = \begin{bmatrix} (\nabla_x I_\sigma)(\nabla_x I_\sigma) & (\nabla_x I_\sigma)(\nabla_y I_\sigma) \\ (\nabla_y I_\sigma)(\nabla_x I_\sigma) & (\nabla_y I_\sigma)(\nabla_y I_\sigma) \end{bmatrix}. \quad (34.2)$$

We need to work with this tensor instead of the gradient vector, ∇I_σ , to preserve the head-to-tail symmetry of microtubules. Another Gaussian blur filter of standard deviation ρ is applied to the gradient tensor to get rid of the small-scale fluctuations in the coherence direction. The value of ρ must be of the order of the size of the coherent domains in the images (i.e., regions where filaments are aligned). The eigenvector of the resultant tensor with the smallest eigenvalue, \mathbf{u} , gives the orientation where the intensity fluctuates the least. Due to the nematic symmetry of the problem, the angle associated to \mathbf{u} must be within the interval $[0^\circ, 180^\circ]$. From \mathbf{u} , we can construct the 2D tensor nematic parameter, \mathbf{Q} :

$$\mathbf{Q} = \left\langle \mathbf{u}\mathbf{u}^T - \frac{1}{2}\mathbf{I} \right\rangle, \quad (34.3)$$

where $\langle \cdot \rangle$ indicates an average over a disk of size β , and \mathbf{I} the identity matrix. Finally, we have to diagonalize \mathbf{Q} to fetch for each pixel S and \mathbf{n} (Example 34.1):

$$\mathbf{Q} = S \left(\mathbf{n}\mathbf{n}^T - \frac{1}{2}\mathbf{I} \right), \quad (34.4)$$

where $S \in [0, 1]$ is the scalar order parameter, which provides information about the degree of alignment of \mathbf{u} (local molecular orientation) relative to \mathbf{n} (mean molecular orientation inside a disk of radius β). Note that this last step is only an averaging process. Actually, if β was set to 0, we would obtain $S = 1$ and $\mathbf{n} = \mathbf{u}$.

The choice of parameters σ , ρ , and β depends on each experiment because the formulation of the active nematic and the image acquisition process significantly affect the size of the relevant features in the experiment.

Example 34.1: How to identify the eigenvectors and the eigenvalues of the \mathbf{Q} tensor. We start by expressing the \mathbf{Q} tensor as

$$\mathbf{Q} = S(\mathbf{n}\mathbf{n}^T - \mathbf{I}/2) = S \begin{pmatrix} \cos(2\theta)/2 & \sin(2\theta)/2 \\ \sin(2\theta)/2 & -\cos(2\theta)/2 \end{pmatrix}, \quad (34.5)$$

where \mathbf{I} is the identity matrix [Define θ]. Now, to look for the eigenvalues λ , we compute the characteristic polynomial

$$\det(\mathbf{Q} - \lambda\mathbf{I}) = \lambda^2 - \frac{S^2}{4}(\cos^2 2\theta + \sin^2 2\theta) = \lambda^2 - \frac{S^2}{4}. \quad (34.6)$$

Thus, $\lambda_{\pm} = \pm S/2$

If we only take the positive eigenvalue, $\lambda = S/2$, we can compute the corresponding eigenvector $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix}$ applying $(\mathbf{Q} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ and find the following relation:

$$b = a \frac{1 - \cos 2\theta}{\sin 2\theta}. \quad (34.7)$$

Choosing $a = \sin 2\theta$:

$$\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sin 2\theta \\ 1 - \cos 2\theta \end{pmatrix}. \quad (34.8)$$

Now, we can readily see that the orientation of the eigenvector gives the orientation of the nematic field:

$$\theta = \text{atan} \left(\frac{b}{a} \right) = \text{atan} \left(\frac{1 - \cos 2\theta}{\sin 2\theta} \right) = \text{atan} \left(\frac{\sin \theta}{\cos \theta} \right). \quad (34.9)$$

Therefore, if we have \mathbf{Q} , we can extract S and \mathbf{n} (or θ) from the diagonalization of \mathbf{Q} .

34.3 Data analysis

Once the local orientation and velocity fields are obtained, the data analysis proceeds according to the particular needs of the research project, e.g., defect density and distribution, geometry of the active nematic, or spectral distribution of active energy. Below, we detail the different methods to obtain these parameters from the basic observables.

34.3.1 Filtering of the nematic director field

It is often convenient to remove spatial noise resulting from the experiments or from the primary data processing detailed in the previous section. To filter the nematic orientational field, we cannot directly apply a conventional spatial filter. This is because of the discontinuities at $\theta = 0$ and $\theta = \pi$. To overcome this problem, we can parameterize the orientation in terms of the \mathbf{Q} tensor, which naturally contains the nematic symmetry and avoids the above discontinuities. Hence, we first compute \mathbf{Q} as $Q_{xx} = \cos(2\theta)/2$ and $Q_{xy} = \sin(2\theta)/2$. Next, we typically apply a disk filter (or mean filter) or a Gaussian filter, usually using convolution.

Then, we can obtain the order parameter S as

$$S = 2\sqrt{Q_{xx}^2 + Q_{xy}^2}, \quad (34.10)$$

and the nematic orientation

$$\theta = \text{atan}2(Q_{xy}, 0.5S + Q_{xx}), \quad (34.11)$$

where the function $\text{atan}2(a, b)$ indicates the angle ϕ between $]-\pi, \pi]$ such that $\tan \phi = a/b$. We could also diagonalize \mathbf{Q} and extract S and θ , as in Subsection 34.2.2, but this method is computationally faster and typi-

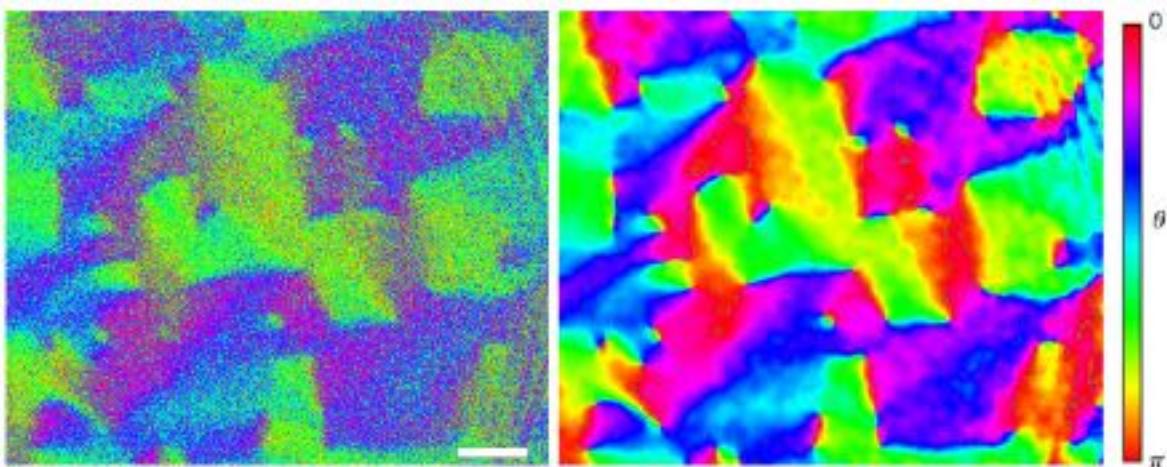


Figure 34.9: **Filtering of the nematic orientation field.** (a) Raw data with a very noisy resolution of the director field. Scale bar is 100 μm . (b) Filtered data as described in the text. [Agnese: add letters](#)

cally good enough. An example is shown in Fig. 34.9, performed with the Matlab script `FilterDirector.m`, available in the repository [@Francesc: can you send me the script to add in the repository?](#).

34.3.2 Defect location and classification. Winding number.

Regions with a low value of S are good candidates to feature topological defects. Hence, we look for areas with S below a threshold value, typically 0.1, and compute the winding number w along a closed loop centered on the point of minimum S . We numerically obtain w as $w = 1/2\pi \oint (\partial\theta/\partial u) du$, where θ is the angle of \mathbf{n} and u is the arc length along the loop. Regions of interest with $w \in \pm[0.49, 0.51]$ are identified as $\pm 1/2$ defects. Note that, because of the nematic symmetry of the director field, we need to include in the code that only jumps in the director smaller than $\pi/2$ between subsequent pixels are allowed. An example of this analysis is shown in Fig. 34.10. A Matlab script to compute the local winding number, `winding.m`, is available in the repository [@Francesc: can you send me the script to add in the repository?](#)

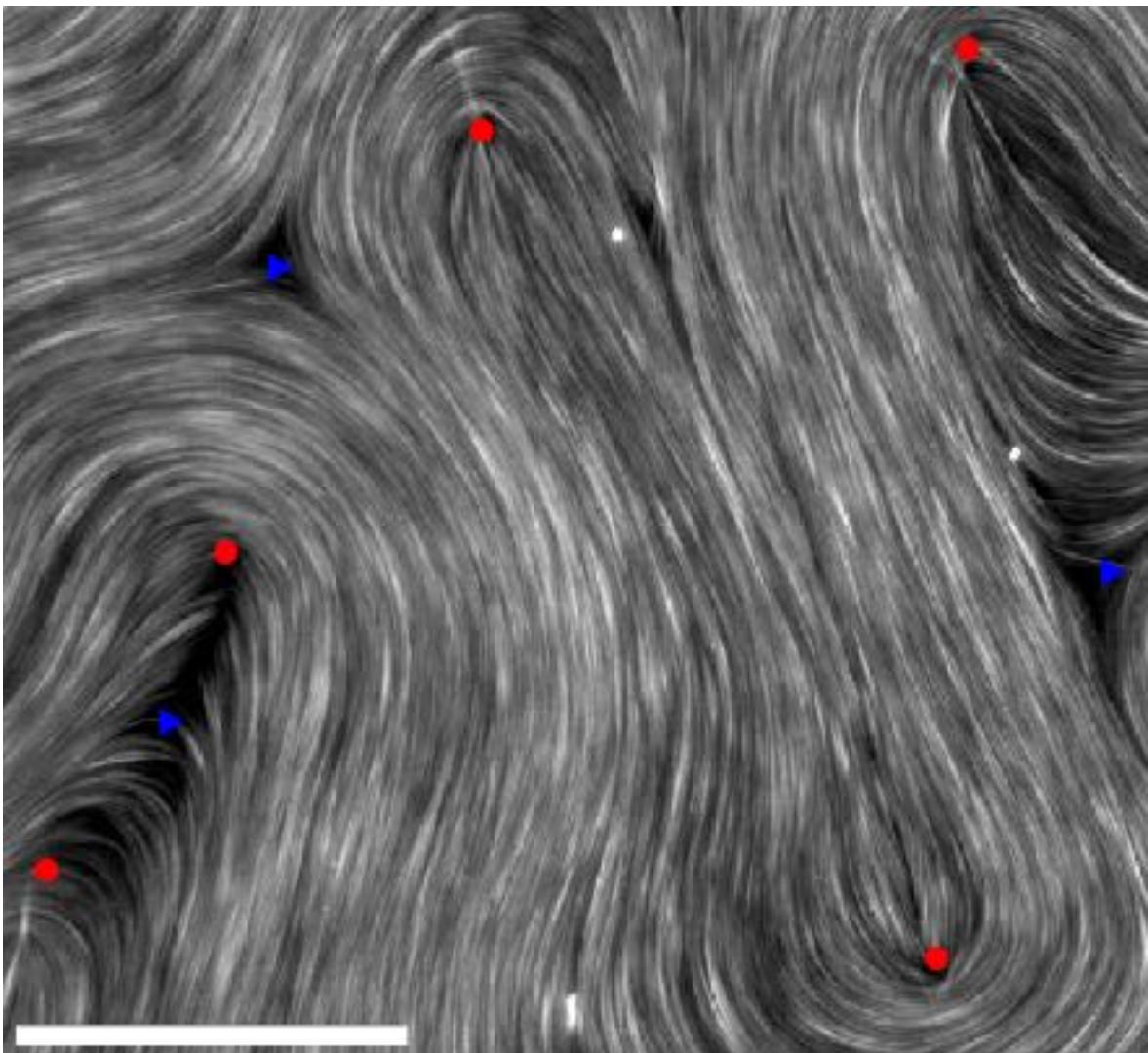


Figure 34.10: **Defect location and classification.** Defects are automatically located and classified in an active nematic. Red circles are $+1/2$ defects, and blue triangles are $-1/2$ defects. Scale bar is $20 \mu\text{m}$.

34.3.3 Analysis of patterns in active nematics

As a first example of detailed analysis, we describe the study of the instability of a radially-aligned active nematic (Fig. 34.11) [16]. Indeed, the material develops an instability triggering the formation of a concentric pattern with a characteristic spacing between dark lanes. To characterize such spacing, we time-average subsequent frames and measure light intensity along a direction orthogonal to the kinks. We finally extract the characteristic wave number q^* (in [16] it appears as k^*) performing a Fast Fourier Transform (FFT) with Matlab. Error bars are calculated as the standard deviation of the mean of 10 measurements along different radial directions in a given experiment.

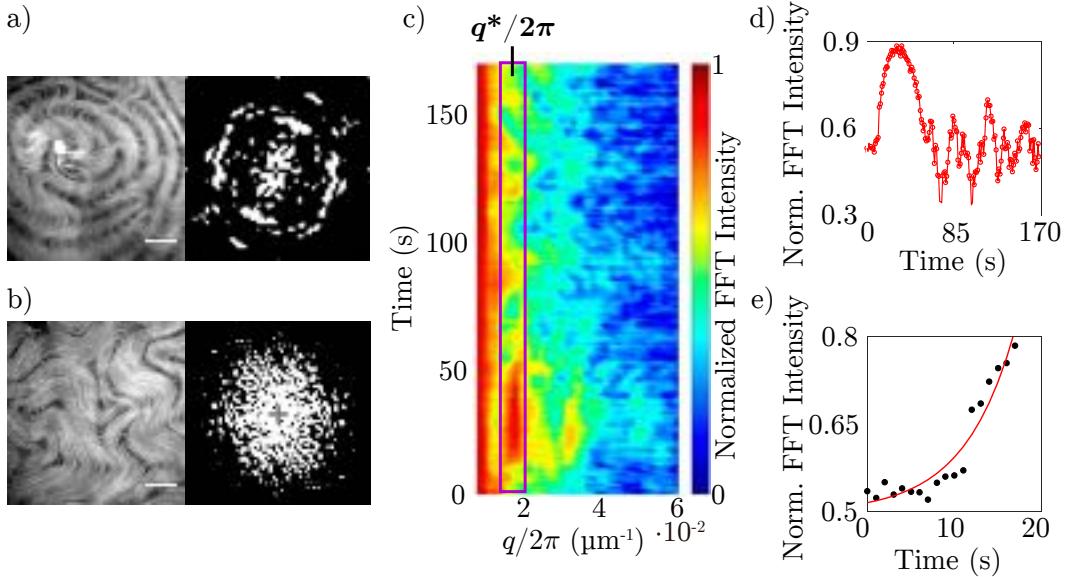


Figure 34.11: Analysis of patterns. (a) Radial pattern and corresponding FFT. (b) Isotropic pattern and corresponding FFT. (c) Kimograph obtained from the FFT of the radial pattern in a. The normalized FFT intensity is plotted as a function of wavenumber and time. (d) Normalized intensity as a function of time for the fastest growing mode. (e) At early times, the fastest growing mode features an exponential growth (line fit through data).

Next, we characterize the time evolution of the instability by monitoring the FFT power spectrum over time with ImageJ. We see that, at short times, the FFT intensity of the characteristic length grows exponentially. Hence, we can fit our data to an exponential trend and extract a characteristic growth rate Ω^* . Error bars are calculated as the uncertainty in the fitted parameter.

34.3.4 Spectral analysis

The power spectrum of an energy density E_{ff} produced by a signal $f(\mathbf{r})$ describes how this energy density distributes in the frequency domain. This could be applied within the context of the active nematic dynamics to analyze how kinetic energy is distributed among different length scales by identifying $f(\mathbf{r}) = \mathbf{v}(\mathbf{r})$, the velocity field.

To compute the power spectrum, we can begin by writing the total energy density of the signal $f(\mathbf{r})$:

$$E_{\text{ff}} = \frac{1}{\mathcal{A}} \sum_{j=1}^{N_x} \sum_{k=1}^{N_y} |f(x_j, y_k)|^2 \Delta x \Delta y = \frac{1}{N_x N_y} \sum_{j=1}^{N_x} \sum_{k=1}^{N_y} |f(x_j, y_k)|^2, \quad (34.12)$$

where, we have assumed that f is discrete and a 2D variable, like all the quantities we can extract from our experiments, and $\mathcal{A} = L_x L_y$ is the area of the system with width L_x and height L_y . We now introduce the Fourier decomposition of $f(x, y) = \frac{1}{M_x M_y} \sum_{j=1}^{M_x} \sum_{k=1}^{M_y} \tilde{f}(q_x^j, q_y^k) e^{i(q_x^j x + q_y^k y)}$, with \tilde{f} the Fourier modes of f and

$(q_x^j, q_y^k) = 2\pi/L(j, k)$, and obtain

$$E_{\text{ff}} = \frac{1}{N_x N_y} \frac{1}{M_x M_y} \sum_{j=1}^{M_x} \sum_{k=1}^{M_y} |\tilde{f}|^2(q_x^j, q_y^k), \quad (34.13)$$

where we have applied the Parseval's theorem for Discrete Fourier Transform. Now, we define the 2D energy density power spectrum $E_{\text{ff}}(q_x, q_y)$ as:

$$E_{\text{ff}} = \sum_{j=1}^{M_x} \sum_{k=1}^{M_y} E_{\text{ff}}(q_x^j, q_y^k) \Delta q_x \Delta q_y = \frac{4\pi^2}{\mathcal{A}} \sum_{j=1}^{M_x} \sum_{k=1}^{M_y} E_{\text{ff}}(q_x^j, q_y^k). \quad (34.14)$$

Comparing Eq. (34.13) with Eq. (34.14), we can obtain the expression for the $E_{\text{ff}}(q_x, q_y)$:

$$E_{\text{ff}}(q_x, q_y) = \frac{\mathcal{A}}{4\pi^2} \frac{1}{N_x N_y} \frac{1}{M_x M_y} |\tilde{f}(q_x, q_y)|^2. \quad (34.15)$$

Furthermore, if we have a time series of the quantity f , we can perform a temporal average, that we denote with $\langle \cdot \rangle$:

$$\langle E_{\text{ff}}(q_x, q_y) \rangle = \frac{\mathcal{A}}{4\pi^2} \frac{1}{N_x N_y} \frac{1}{M_x M_y} \langle |\tilde{f}(q_x, q_y)|^2 \rangle. \quad (34.16)$$

These calculations can be easily performed with the Matlab script `KinEnSpectrum2D.m`, available in the repository [add the script in the repository](#).

The time averaged energy power spectrum can be isotropic. In these cases, one can compute the angle-averaged power spectrum $\langle E_{\text{ff}}(q) \rangle$, where q is the modulus of the wave vector $\mathbf{q} = (q_x^j, q_y^k)$. We define this quantity as:

$$\langle E_{\text{ff}} \rangle = \sum_{l=1}^M \langle E_{\text{ff}}(q_l) \rangle \Delta q. \quad (34.17)$$

Then, taking $\Delta q_x \Delta q_y = q \Delta \varphi \Delta q$ in Eq. (34.14) and introducing Eq. (34.15), we obtain

$$\langle E_{\text{ff}}(q) \rangle = \frac{\mathcal{A}}{4\pi^2} \frac{1}{N_x N_y} \frac{1}{M_x M_y} \sum_{\varphi} \langle |\tilde{f}(\varphi, q)|^2 \rangle \Delta \varphi \quad (34.18)$$

with φ the azimuth. The Matlab script `AngAverage.m`, provided in the repository [add the script in the repository](#), can be used to perform this angular averaging.

34.3.5 Location of vortices

It is also customary to study the distribution of vortices in the active nematic flow. They are detected using the Okubo-Weiss (OW) criterion [17, 18], which parametrizes the stability of two initially close particles immersed in a 2D velocity field at a given time $\mathbf{v}(x, y, t)$ by analyzing the evolution of a small perturbation of the flow field $\delta \mathbf{v}$. We start by writing the evolution of this small perturbation in terms of the Jacobian of \mathbf{v} as:

$$\begin{pmatrix} \delta v_x \\ \delta v_y \end{pmatrix} = \begin{pmatrix} \partial_x v_x & \partial_y v_x \\ \partial_x v_y & \partial_y v_y \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}. \quad (34.19)$$

Now, we assess the stability of the velocity field by looking for the eigenvalues λ_{\pm} of the Jacobian matrix given by:

$$\lambda_{\pm} = \frac{-(\partial_y v_y + \partial_x v_x) \pm \sqrt{(\partial_y v_y + \partial_x v_x)^2 - 4(\partial_x v_x \partial_y v_y - \partial_y v_x \partial_x v_y)}}{2}. \quad (34.20)$$

If we consider the fluid to be incompressible, i.e., $\partial_x v_x + \partial_y v_y = 0$, we can simplify Eq. (34.20) to:

$$\lambda_{\pm} = \frac{\pm \sqrt{-4(-(\partial_x v_x)^2 - \partial_y v_x \partial_x v_y)}}{2} = \sqrt{(\partial_x v_x)^2 + \partial_y v_x \partial_x v_y}. \quad (34.21)$$

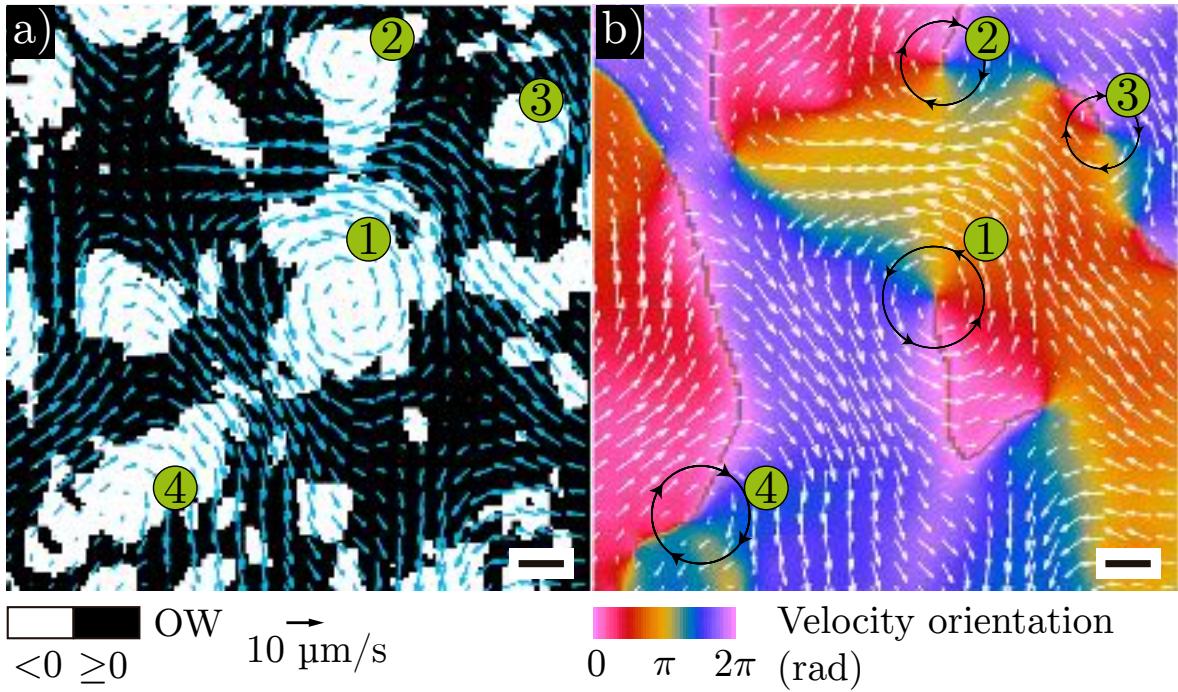


Figure 34.12: **Detection of vortices.** (a) OW parameter computed from the velocity field (blue arrows). (b) The same velocity field is overlaid on its orientation map. Vortices are regions with $\text{OW} < 0$ (white areas in a) and with a 2π rotation of the velocity field around a closed loop (black lines in b). Detected vortices are labeled with numbers in a.

The term inside the square root is the OW parameter ($\text{OW} = (\partial_x v_x)^2 + \partial_y v_x \partial_x v_y)$. In the regions of the fluid with $\text{OW} < 0$, the distance between the two particles embedded in the fluid will not diverge exponentially with time, whereas in the regions with $\text{OW} < 0$, it will diverge. Hence, coherent vortices are expected to be at simply connected regions with $\text{OW} < 0$. Notice here the term *coherent*. Not all regions with $\text{OW} < 0$ feature vortices (Fig. 34.12). Thus, to classify a simply connected region as a vortex, we must check if it contains a singularity with a winding number of +1 of the velocity field (Subsection 34.3.2).

As the velocity fields of our experiments are discrete, it is best to use the five-point method to compute the spatial derivatives of the velocity. Hence, for a function $f(x, y)$ and for the derivative with respect to x with a spacing between points in the grid of Δx , $\frac{\partial f(x, y)}{\partial x}$ is given by:

$$\frac{\partial f(x, y)}{\partial x} \approx \frac{-f(x + 2\Delta x, y) + 8f(x + \Delta x, y) - 8f(x - \Delta x, y) + f(x - 2\Delta x, y)}{12\Delta x}. \quad (34.22)$$

34.3.6 Correlation functions

The final data analysis tool that we will discuss among those that are customary employed during the study of active nematic are the spatial and temporal correlation functions.

Spatial autocorrelation functions

2D spatial autocorrelation of a discrete function $f(x_i, y_j)$, C_{ff} , is defined as:

$$C_{\text{ff}}(\Delta x, \Delta y) = \left\langle \frac{\sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} f(x_i, y_j) f(x_i + \Delta x, y_j + \Delta y)}{\sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} |f(x_i, y_j)|^2} \right\rangle, \quad (34.23)$$

where $\langle \cdot \rangle$ indicates a temporal average.

In practice, the functions we can extract from our experiments are discrete; hence, to compute C_{ff} , we have to multiply f point by point. As a consequence, the calculation of C_{ff} can be computationally very expensive, especially when the system size is large. Therefore, we apply the Wiener-Khinchin theorem, which relates the Fourier transform of f , $\tilde{f}(\mathbf{q})$, with the correlation function as:

$$C_{\text{ff}}(\mathbf{r}) = \left\langle \frac{\mathcal{F}^{-1}[\tilde{f}(\mathbf{q})\tilde{f}(\mathbf{q})^*]}{\mathcal{F}^{-1}[\tilde{f}(\mathbf{q}=0)\tilde{f}(\mathbf{q}=0)^*]} \right\rangle, \quad (34.24)$$

where $*$ indicates the complex conjugate and \mathcal{F}^{-1} the inverse Fourier transform, which we compute with the built-in function `ifft2()` in Matlab. Also, we shift the output with `ifftshift()` in such a way that $r=0$ is at the center of the output matrix.

If the 2D autocorrelation function is isotropic, we angle-average $C_{\text{ff}}(\mathbf{r}) = C(r, \varphi)$ as:

$$C_{\text{ff}}(r) = \frac{1}{2\pi} \sum_{\varphi} C_{\text{ff}}(r, \varphi) \Delta\varphi, \quad (34.25)$$

which features a radial dependence only. This calculation can be easily performed using the Matlab script `autocorrfun.m`, provided in the repository. [add to repository](#)

Time autocorrelation function

We compute the time autocorrelation, C_{tt} as follows:

$$C_{\text{tt}}(x, y, \tau) = \frac{\sum_{i=0}^{N-1} f(x, y, t_i) f(x, y, t_i + \tau)}{\sum_{i=0}^{N-1} |f(x, y, t_i)|^2}. \quad (34.26)$$

Furthermore, to remove the spatial dependence, we also average $C_{\text{tt}}(x, y, \tau)$ over space:

$$C_{\text{tt}}(\tau) = \frac{1}{N_x N_y} \sum_{i=0}^{N_x-1} \sum_{j=1}^{N_y-1} C_{\text{tt}}(x_i, y_j, \tau). \quad (34.27)$$

34.4 Conclusions

In this Chapter, we have reviewed the basic experimental protocols to work with the well-known kinesin–tubulin active nematic, which forms at the water–oil interface between the active gel and the oil of choice. Although project goals may vary, image processing to extract the velocity and orientation field is typically how data processing begins. We have described here basic protocols based on fluorescence microscopy. A homogeneous mixture of fluorescent and non-fluorescent tubulin yields the orientational field with precision. One may also apply the PIV tools to extract the velocity field from the same images, but it is well-known that the lack of discrete spatial features yields a poorly resolved velocity field along aligned filaments. An alternative is to perform a separate experiment where fluorescent and non-fluorescent tubulin is segregated. The velocity field is resolved with higher precision, but the possibility of simultaneously resolving the orientational field is lost. Alternative observation with laser-scanning confocal microscopy that combines fluorescence and reflection modes may take advantage of the latter mode to obtain a more textured pattern, more amenable to PIV analysis. An interesting prospect is to develop high speed polarimetry in a multi-modal microscope that allows to extract the orientational field from the (low) birefringence of the active nematic while sparse fluorescent microtubules would yield, simultaneously, the velocity field.

While the purpose of each project will be different, the analysis tools we have described have a broad applicability, and can be used as the firm starting point for more complex endeavours.

34.5 Problems

[Can you suggest a couple of problems?](#)

Bibliography

- [1] R. D. Vale. The molecular motor toolbox for intracellular transport. *Cell*, 112:467 – 480, 2003.
- [2] R. Subramanian and J. Gelles. Two distinct modes of processive kinesin movement in mixtures of ATP and AMP-PNP. *Journal of General Physiology*, 130:445–455, 2007.
- [3] R. M. Jouybari, A. Sadeghi, B. Khansarinejad, S. S. Abbasian, and H. Abtahi. Production of recombinant streptavidin and optimization of refolding conditions for recovery of biological activity. *Reports of Biochemistry and Molecular Biology*, 6:178–185, 2018.
- [4] A. A. Hyman, S. Salser, D. N. Drechsel, N. Unwin, and T. J. Mitchison. Role of GTP hydrolysis in microtubule dynamics: Information from a slowly hydrolyzable analogue, GMPCPP. *Molecular Biology of the Cell*, 3:1155–1167, 1992.
- [5] C. W. Pratt and K. Cornely. *Essential Biochemistry*. Wiley, United States of America, 4th edition, 2018.
- [6] G. Henkin, S. J. DeCamp, D. T. N. Chen, T. Sanchez, and Z. Dogic. Tunable dynamics of microtubule-based active isotropic gels. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 372:20140142–20140142, 2014.
- [7] S. A. Kuznetsov and V. I. Gelfand. Bovine brain kinesin is a microtubule-activated ATPase. *Proceedings of the National Academy of Sciences*, 83:8530–8534, 1986.
- [8] C. E. Willert and M. Gharib. Digital particle image velocimetry. *Experiments in Fluids*, 10:181–193, 1991.
- [9] W. Thielicke and E. J. Stadhuis. PIVlab – Towards User-friendly, Affordable and Accurate Digital Particle Image Velocimetry in MATLAB. *Journal of Open Research Software*, 2, 2014.
- [10] W. Thielicke. *The flapping flight of birds: Analysis and application*. PhD thesis, University of Groningen, 2014.
- [11] M. Raffel, C. E. Willert, F. Scarano, C. J. Kähler, S. T. Wereley, and J. Kompenhans. *Particle Image Velocimetry. A Practical Guide*. Springer International Publishing AG, Berlin, 3rd edition, 2018.
- [12] T. Olson. *Applied Fourier Analysis*. Birkhäuser, New York, NY, 2017.
- [13] P. W. Ellis, D. J. G. Pearce, Y. W. Chang, G. Goldsztein, L. Giomi, and A. Fernandez-Nieves. Curvature-induced defect unbinding and dynamics in active nematic toroids. *Nature Physics*, 14:85–90, 2018.
- [14] P. W. Ellis, J. Nambisan, and A. Fernandez-Nieves. Coherence-enhanced diffusion filtering applied to partially-ordered fluids. *Molecular Physics*, 118, 2020.
- [15] J. Weickert. Coherence-enhancing diffusion filtering. *International Journal of Computer Vision*, 31:111–127, 1999.
- [16] B. Martínez-Prat, J. Ignés-Mullol, J. Casademunt, and F. Sagués. Selection mechanism at the onset of active turbulence. *Nature Physics*, 15:362–366, 2019.
- [17] L. Giomi. Geometry and topology of turbulence in active nematics. *Physical Review X*, 5:1–11, 2015.
- [18] R. Benzi, S. Patarnello, and P. Santangelo. Self-similar coherent structures in two-dimensional decaying turbulence. *Journal of Physics A: Mathematical and General*, 21:1221–1237, 1988.

Chapter 35

Planktonic Active Matter

ANUPAM SENGUPTA

Phytoplankton are photosynthetic microorganisms, which are at base of nearly all aquatic food webs. As a key player of the ocean and freshwater ecosystems, phytoplankton impact global biogeochemical cycles, produce close to half of the world's oxygen, and are important sources of algal biofuel. The word "phytoplankton" is a portmanteau of the Greek words, *phyton* meaning plant, and *planktos*, signifying a wanderer or passive drifter. For long phytoplankton have been considered to be passive drifters, their spatio-temporal locations determined largely by the environmental fluid flows. However, decades of satellite-, field- and lab-based studies have confirmed that the movement of phytoplankton can occur actively, and they may not be just wandering horizontally or vertically along the water column carried by the fluid flows alone [1, 2, 3]. In this Chapter, we

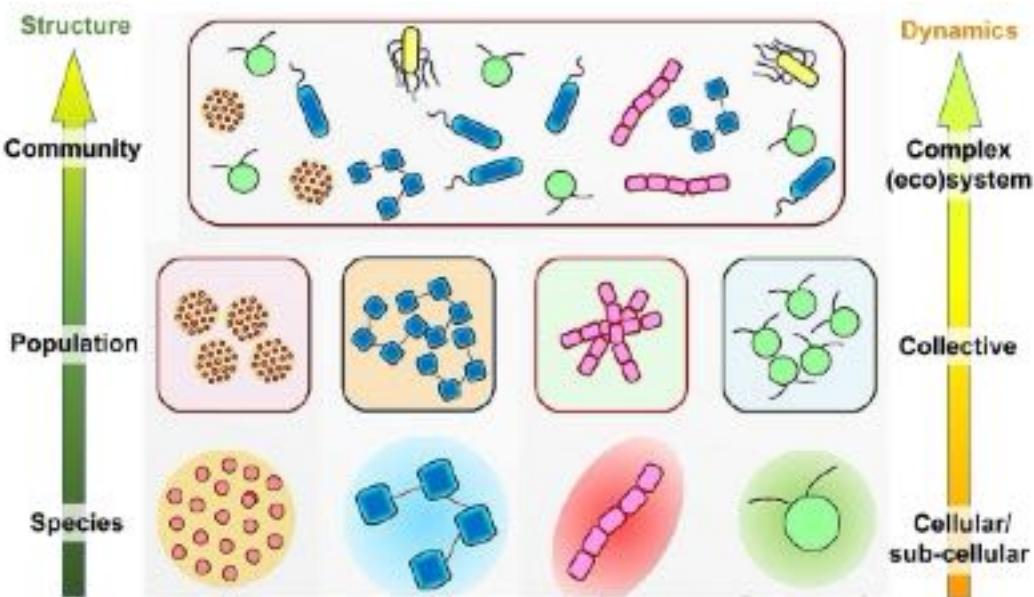


Figure 35.1: Scales and complexity of planktonic active matter. Planktonic active matter presents emergent systems across different scales (species, population, community) and complex dynamics from sub-cellular and cellular scales to collective and ecosystem scales. This cross-scale active matter system responds to both abiotic (temperature, fluid flow, light conditions) and biotic factors (nutrients, pH, secondary metabolites) characteristic of the relevant ecosystems. Active modulation of cell phenotypes, including morphology and motility, enable planktonic microorganisms to interact with other individuals and species, and rapidly adapt to changes in their micro-environment. A multi-scale approach allows us to apply an active matter framework to understand natural ecological systems and their emerging structure and functions due to changing conditions.

give an introduction to planktonic active matter. We will first start with a general overview of phytoplankton in the context of active matter, then focus on their physical ecology, their cell morphology and mass distribution, and the related mathematical modeling, and finally explore gravitaxis, gyrotaxis, and bioconvection across single and collective scales.

35.1 Overview of planktonic active matter

Planktonic active matter represents a highly diverse community of organisms, with hierarchical complexity in their composition, structure, and dynamics ranging from intraspecific (within a given species) to inter-species and species-environment interactions (Fig. 35.1). Planktonic active matter spans orders of magnitude in size (sub-micron to sub-millimeter range), distinct morphologies, ecosystem-dependent photosynthetic activities, and biological functions. Light, nutrients, and turbulence determine the ecology of phytoplankton [4, 5, 6], thus making their position along the vertical water column highly consequential. Phytoplankton occupy the so-called *photic* zone — the light-rich region of the aquatic ecosystems. Light is a key determinant of phytoplankton physiology and fitness, as it mediates vital functions by regulating endogenous circadian cycles of light-harvesting bacteria and algae [7]. Many species of phytoplankton are motile, i.e., they can actively migrate through the water column by leveraging *gravitaxis*, the directional movement in response to gravity [8]. Gravitactic phytoplankton cells migrate upward — against the direction of the gravity force and toward light — during the day, and change their swimming direction downward — toward higher inorganic nutrient concentrations — at night [9, 10, 11, 12]. While the size of individuals undertaking this daily vertical migration is minuscule, the sheer number of cells involved makes the daily vertical migrations some of the largest and most important concerted microbial migrations on Earth. Daily vertical migrations contribute significantly to the sequestration of carbon from the atmosphere to the deep ocean, among other mechanisms through the sinking of particulate organic matter. Groups of motile gravitactic species, including dinoflagellates and raphidophytes, are frequently found to generate harmful plankton blooms, or more commonly, harmful algal blooms. A complex interplay of cell motility and morphology, alongside abiotic and biotic factors like the ambient temperature, fluid flow, nutrient concentrations, pH, and seasonal factors, drives the formation of the harmful algal blooms [13, 14].

Complementing motile species, a second important group of planktonic microorganisms comprises the *diatoms* [15]. Unlike their motile counterparts, diatoms generally lack appendages (e.g., cilia or flagella) required to generate propulsion forces. Their ecology, as a consequence, is tightly coupled to the local fluid flows, which act as conveyors for moving cells from one point to another. Conversely, motile species dominate the calmer regions of the ocean, while diatoms are frequently associated with highly dynamic settings, for instance in the regions of high turbulence. Along the vertical water column, diatoms leverage an array of biophysical mechanisms to either maintain or alter their buoyancy, allowing them optimal access to light and nutrients. Diatoms have evolved different modes to regulate density, including active replacement of heavy ions within vacuoles (triggered by light, nutrients, or osmotic stress [16]), reduction of the starch or carbohydrate inclusions [17, 18], or, over longer timescales, bio-silicification [19]. Remarkably, diatoms show a rapid control of buoyancy under nutrient-limited conditions, thereby potentially increasing the diffusive transport of nutrient molecules to the cell surface [20]. Physiologically, diatoms have been found to access nitrogenous compounds more efficiently (at lower concentrations), thereby exhibiting relatively higher photosynthetic and growth rates compared to the motile dinoflagellates [21].

Planktonic active matter demonstrates exquisite mechanisms to diversify their biophysical traits in response to a range of physico-chemical cues. The ability of phytoplankton to adapt their motility traits — over both short (within a division timescale) and longer timescales (spanning multiple generations) — is underpinned by morphological pliability, intracellular reconfigurability, or modulation of the flagellar beating. Diversification of traits, both physiological and behavioral, enables phytoplankton populations to respond, adjust, and adapt to changes in their environmental conditions, thus maximizing chances of survival [22, 23, 24, 25].

This offers a highly rich testbed for biophysicists to test hypotheses, and drive the field of active and living matter forward. More importantly, understanding how phytoplankton adapt and develop strategies to survive the rapidly evolving nutrient, turbulence, and light conditions of today's oceans remains a crucial challenge. Accounting for the active mechanisms and emergent properties observed frequently across all planktonic systems, could allow accurate predictions of planktonic community compositions, structures, and dynamics across scales and complexities, ultimately advancing the existing models of biogeochemical cycles and biological

pumps for today's aquatic ecosystems.

35.2 Physical ecology of plankton

Ramon Margalef, one of the founding fathers of modern marine biology, was among the first to propose and formalize the dependence of phytoplankton physiology on their environmental settings, specifically, the levels of light, nutrient concentration, and turbulence [4, 26, 27]. In the context of gravitaxis, ambient fluid flows have a direct impact on phytoplankton behavior and physiology, while strong turbulence can be detrimental to motile phytoplankton, potentially leading to flagellar or body-wall damages, triggering enhanced physiological stress and reduced growth [23]. Planktonic species leverage evolutionary coping mechanisms to tackle environmental stressors by devising adaptive strategies based on the intrinsic plasticity of their functional traits [10, 28, 22]. These include minute but rapid adjustments of the cell morphology to adjust the swimming stability [22], the reduction of spine length to promote sinking by the dinoflagellate *Ceratocorys horrida* [29], or the formation of chains by bloom-forming *Alexandrium catenella* to adjust their swimming behavior in response to hydrodynamic shear [30]. However, fluid flows can be beneficial for non-motile diatom species, specifically due to the enhancement of the encounter rates between cells and nutrient molecules [31, 32]. Interestingly, alteration of cell morphology may further enhance the access to nutrients, for instance when single cells transform into chain-like morphology observed in many species [33]. In addition to the cell length, increase of the chain rigidity enhances relative nutrient fluxes, suggesting a critical advantage conferred by the silica frustules often found in diatoms [34].

The different phytoplankton life-forms captured in *Margalef's mandala* (Fig. 35.2a) are based on their ability to adapt and survive in unstable and turbulent environments. Their small size combined with rapid turnover times makes external energy input from turbulence a key determinant of phytoplankton fitness. In its simplest form, fitness can be measured in terms of the growth rates and the carrying capacities of the species (Fig. 35.2b), though, technically, fitness could include multiple associated metrics, including a species' ability of risk- or predator-avoidance, stress amelioration, or maintenance of basic metabolic processes under limited resources. Secondary factors like grazing by predators may be further introduced into the mandala, overlaying

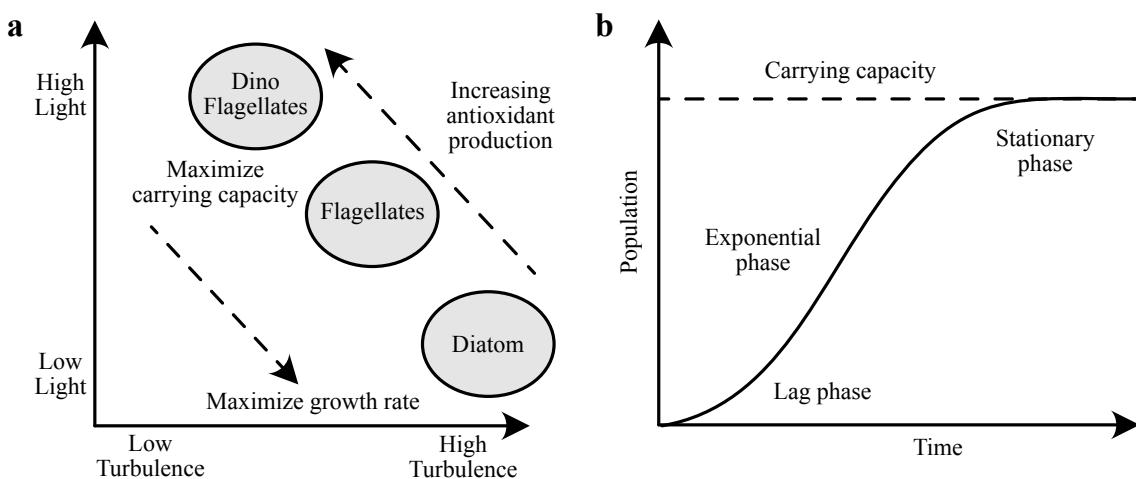


Figure 35.2: Physical ecology of planktonic microorganisms. (a) Depiction of the Margalef's mandala showing the relationship between environmental factors (light and turbulence) and phytoplankton type and antioxidant production in natural habitats. Flagellated species typically occupy low turbulence and high light regions, whereas non-flagellated species, like diatoms, are typically found in regions of high turbulence and low light regions. Adapted from Ref. [26]. (b) The logistic curve captures growth phases of planktonic species growing under laboratory environments: lag phase, exponential phase, and the stationary phase. The specific growth rate of a species is calculated during the exponential phase. The carrying capacity quantifies the average steady state population size under given environmental conditions (representative of a particular habitat), including nutrient availability, light and turbulence conditions, and prey-predator interactions.

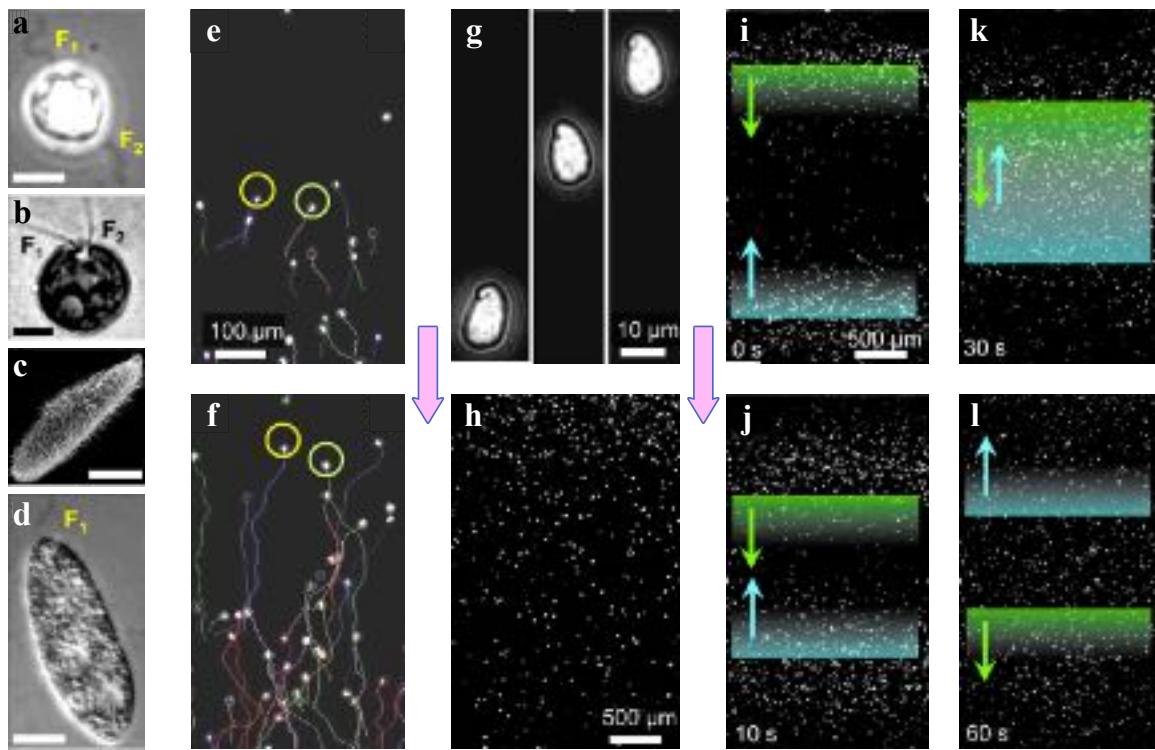


Figure 35.3: Gravitaxis and daily vertical migration. (a-d) Gravitactic species: Phase contrast images of (a) *H. akashiwo* and (b) *Chamydomonas reinhardtii* (images adapted from Ref. [35]); (c) scanning electron micrograph of the ciliate *Paramecium tetraurelia* (image adapted from Ref. [36]); (d) brightfield image of *Euglena gracilis* (image adapted from Ref. [37]). F_1 and F_2 indicate the flagella present in the corresponding species. (e), (f): Population of negatively gravitactic species swimming against the gravity direction (downward arrow). The trajectories of the swimming cells are presented using different hues. (g) Single microplankton cell executing gravitactic motion in a vertical column. Cells typically rotate about their body axis while executing a helical swimming trajectory. (h) Stationary distribution of gravitactic phytoplankton in a vertical chamber. A higher cell concentration is observed around the top of the chamber than at its bottom. (i-l): A sequence of micrographs capturing the motion of co-existing negative (up-swimming cells) and positive (down-swimming cells) gravitactic sub-populations in a vertical chamber.

them with light (closer to the air-water interface) and nutrients (at depths in the water column) as the other two primary axes which govern phytoplankton eco-physiology.

Phytoplankton frequently encounter diverse fluid dynamic environments: the natural environments which they inhabit, engineered confinements such as algal bioreactors, or self-organized emergent flows as in a bio-convecting plume [38, 39, 40]. Large-scale turbulent structures typically cascade to eddy structures at smaller scales, until the turbulent energy is dissipated by viscosity alone [31]. Under typical marine conditions, the mean dissipation rate of turbulent kinetic energy per unit mass varies between $10^{-9} \text{ W kg}^{-1} < \varepsilon < 10^{-5} \text{ W kg}^{-1}$, corresponding to a Kolmogorov scale, $\eta_K = (v^3/\varepsilon)^{1/4} = 0.5$ to 5 mm, with $v \approx 10^{-6} \text{ m}^2 \text{s}^{-1}$ being the kinematic viscosity of sea water [41, 32, 42]. The Kolmogorov scale represents the smallest scale below which the eddy flow breaks down into a simple linear variation in the fluid velocity. The corresponding order of magnitude of the velocity gradient, given by the Kolmogorov shear rate, $S_K = (\eta_K/v)^{1/2}$, thus falls between 0.03 and 3 s^{-1} , with its inverse $\tau_K = 1/S_K$ representing the characteristic time scale of the velocity fluctuations.

With a vast proportion of the phytoplankton species smaller than the characteristic Kolmogorov scale of oceans, cells perceive turbulence as an instantaneous, linearly varying fluid velocity across the cell body [43]. The interplay between phytoplankton motility and ocean turbulence has drawn considerable attention, leading to a deep understanding of both the biophysical and biomechanical aspects, particularly of the vertical migra-

tion. Numerous phenomena have been reported including the formation of thin planktonic layers due to shear flows [44], the emergence of phytoplankton clusters, also referred to as *patchiness* [45, 46, 47], and the enhancement of the vertical migration of chain-forming phytoplankton through turbulence [48]. More recently, it has been shown that major groups of motile phytoplankton (raphidophytes and dinoflagellates) harness active, behavioral changes in response to turbulent cues [22]. The response — manifested as morphological transformations that impact the stability of the swimming cells — is underpinned by the generation of stress markers called *reactive oxygen species* (ROS) [23]. Furthermore, morphological transformations can occur under nutrient-limited settings, which together with the growth and intracellular translocation of energy-storing lipid droplets, can govern migratory strategies of phytoplankton in a species-specific manner [25].

A vast majority of the planktonic microbes perceive gravity and use it to adjust their position in the water column, in combination with other external cues like light and chemical gradients (Fig. 35.3). Directed movement along or against the gravity direction is called *gravitaxis* and can be positive (downward swimming) or negative (upward swimming) [49, 8]. The execution of the negative or positive gravitaxis — observed across raphidophytes, dinoflagellates, and ciliates (Fig. 35.3a-d) — depends on a range of factors, primarily the circadian rhythm [50, 51]. In addition, some ciliates and flagellates are able to perform *gravikinesis* whereby cells modify their swimming behavior by activating a kinetic response: they speed up during the upward swimming phase, and decelerate during downward swimming [52, 53]. This kind of motion, observed typically in larger organisms, depends on the local environmental conditions and enables cells to compensate sedimentation rates, either partly (for instance, in *Paramecium* [54]) or fully (like in *Tetrahymena* [55]). Planktonic microbes have an array of different mechanisms to perceive gravity, including the *statoliths* (heavy bio-mineralized organelles that press onto the cell's gravireceptor), mechanosensitive ion channels (acting as gravireceptors that amplify changes in the gravity, or acceleration, based on which the direction of swimming is altered), and physiological stress markers like the ROS [8, 24].

35.2.1 Morphology and organelles of phytoplankton

The morphology and size of phytoplankton have been found to play a crucial role on the growth, uptake and survival of species. Across different scales and organizational complexity, body size correlates with various traits of species, impacting the composition, structure and dynamics of the phytoplankton networks and food webs, and their stability and resilience to perturbations [56]. In regards to the growth rates, larger motile cells have been reported to outperform motile cells of smaller diameters, possibly due to the relatively higher nutrient uptake rates in larger cells, particularly under turbulent conditions [31, 38, 57]. Under nutrient-limited settings, larger species are often found to outperform smaller ones [58], while cell morphology governs the active hydrodynamic strategies that the cells put in use to navigate their fluidic settings [59, 25]. As turbulent strength increases, shear forces can be detrimental for the motile species, even leading to complete cessation of motility citesengupta2017, carrara2021. Furthermore, in regards to the nutrient uptake, an upper limit is attained, after which size no longer offers a competitive advantage to the swimming cells. For pico- and nanoplankton, the impact of cell morphology on their eco-physiology has been reported to be insignificant. Owing to the dependence of the metabolic constraints with body-size scaling, investigations are currently underway to develop a quantitative framework within which such microscale traits can be incorporated toward description of the structure and functioning of populations and networks at larger scales. The quantitative roles of cell morphology under turbulent cues and nutrient limitation is taken up later in the Chapter, Section 35.3. In addition to cell size and morphology, a growing body of recent literature has demonstrated the role of intracellular organelles in governing the swimming properties (Fig. 35.4). These include the cell nucleus (the heaviest organelle in a cell), energy-storing lipid bodies, carbohydrate and starch reserves, chloroplasts, and gas vacuoles [22, 25, 60]. By altering the physical size, density, and position within the cell body, cellular organelles can alter the swimming speed, orientational stability, velocity correlations (switching from ballistic to diffusive swimming or vice-versa), thereby impacting the overall motility characteristics of the swimmers.

Using bright-field, phase-contrast, and fluorescent-based single cell imaging techniques, precise detection and quantification of the cell morphology and organelles can be made. As shown in Fig. 35.5, raw experimental images can be analyzed by image processing tools to extract the cell and organelle contours and sizes, which can be ultimately fitted to mathematical functions to obtain various feature dimensions [22, 24, 25].

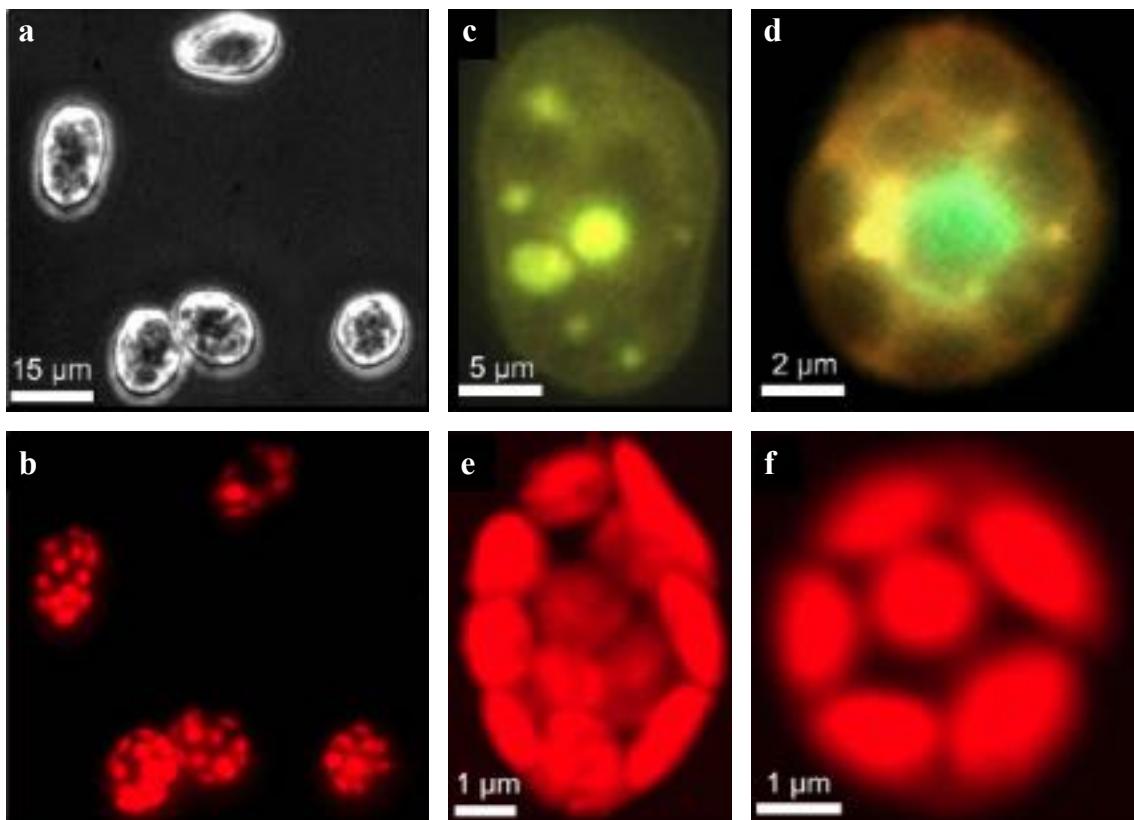


Figure 35.4: Cell morphology and functional organelles. (a) Phase contrast image of raphidophyte *Heterosigma akashiwo*, a motile biflagellate microplankton species. (b) Corresponding autofluorescent micrograph of *H. akashiwo*, illuminated by blue light, revealing the light-harvesting organelles called chloroplasts. (c) Epifluorescence microscopy of intracellular lipid droplets (yellow-green hue) stained using Nile Red dye. (d) Dual-channel epifluorescent imaging captures the cell nucleus (green hue, stained using Syto9 stain) and lipid droplets (yellow-green hue, stained using Nile Red dye). (e), (f) Magnified view of chloroplasts embedded on the outer cell membrane of *H. akashiwo* cells, visualized using autofluorescence imaging.

35.2.2 Encounter rates and kernels

Planktonic organisms experience frequently encounters with the different molecules that make up their micro-environment as well as with one another, either stochastically or while executing prey–predator dynamics. The *encounter rate kernel* [61], a measure of clearance rate (or zone of influence) of a particle or a cell, depends on a number of factors: the mobility of the cell (static, swimming, or sinking), the dynamics of the surrounding fluid (stationary versus flowing conditions), laminar versus turbulent eddies, and the shape (and symmetry) of the cell [31, 62, 63]. Encounters between microorganisms play a crucial ecological role, from predator–prey interactions, to food-web structures and optimal foraging [64, 65, 66], to finding mates in larger plankton like the copepods [67, 68]. Following Ref. [31], the encounter rate E (dimensions: number of encounter per unit volume per unit time, $\text{m}^{-3}\text{s}^{-1}$) between cells in a given population with concentrations C_i and C_j (dimensions: number of particle per unit volume m^{-3}) can be written as:

$$E = \beta C_i C_j = \beta_{\text{enc}} C_1^2, \quad (35.1)$$

where $i = j = 1$ holds for same species and $i \neq j$ for different species (e.g., in a prey–predator interaction); β_{enc} is the encounter rate kernel (dimensions m^3s^{-1}).

For single species, the concentration of cells over time can be written as

$$\frac{dC_1}{dt} = -\beta_{\text{aggr}} C_1^2. \quad (35.2)$$

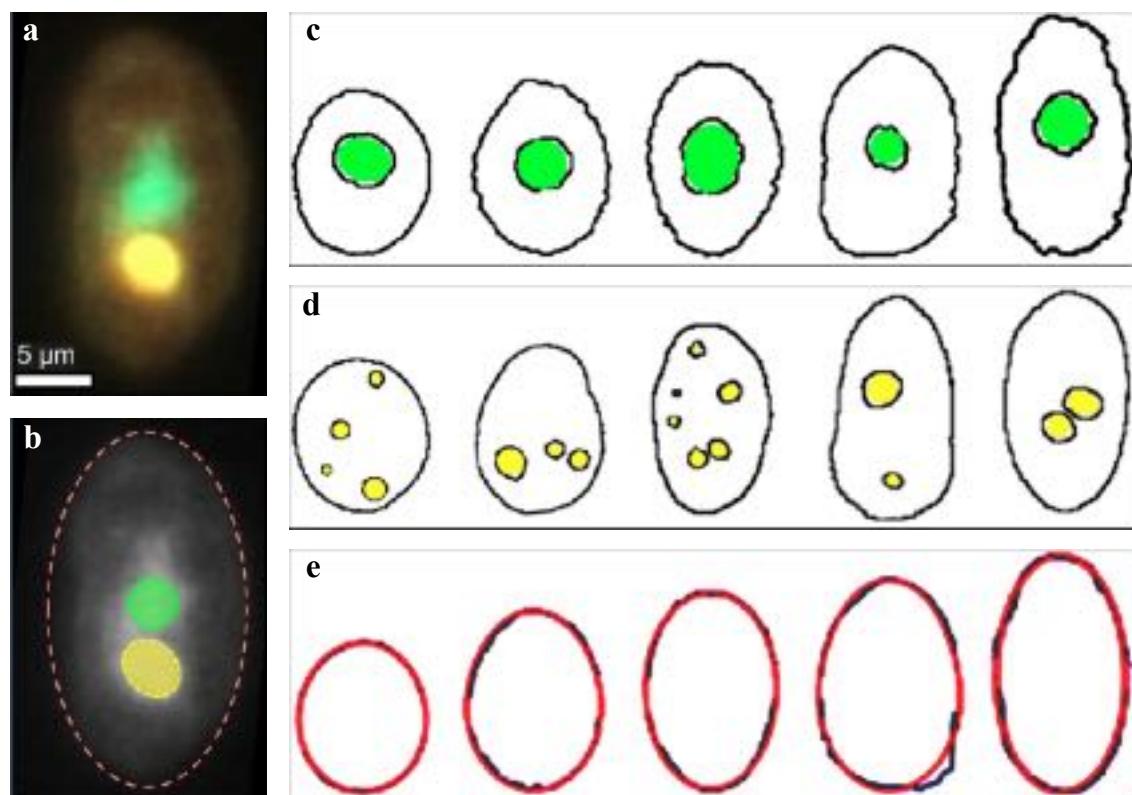


Figure 35.5: Quantifying cell morphology and intracellular organelles. (a) Raw and (b) grey-scale images showing the cell morphology and intracellular organelles within a microplankton cell. Dashed lines denote the contours of the cell body and the organelles. The nucleus (green) and lipid droplet (yellow) were simultaneously visualized using epifluorescence imaging. Image analysis can be used to extract the cell's contour and the positions of the (c) nucleus and (d) lipid droplets. (e) The experimentally extracted cell contours are fitted with a three-parameter curve. Adapted from Refs. [22, 25].

By solving the concentration equation, one can obtain the growth dynamics of cells aggregating, or extend this further for other relevant settings (by appropriately changing the i and j values). The corresponding kernels (also known as coagulation kernels, with dimension of volume rate, i.e., m^3s^{-1}) for different mechanisms (thermal diffusion, settling particles with different speeds, and turbulence driven encounters) are given as

$$\beta_{\text{aggr}} = \begin{cases} 4\pi(D_i + D_j)(a_i + a_j) & : \text{Brownian diffusion} \\ 0.5\pi a_i^2 |u_i - u_j| & : \text{Differential settling for } a_i \leq a_j \\ 1.3\gamma(a_i + a_j)^3 E_{i,j} & : \text{Small-scale turbulent shear} \end{cases}$$

where D_i , a_i , and u_i represent the translational diffusion coefficient, dimension, and speed respectively; γ is the sub-Kolmogorov shear rate; and $E_{i,j}$ represents represents the corresponding numerical factor and is given by:

$$E_{i,j} = \begin{cases} 1 & \text{for } a_i = a_j \\ \frac{7.5(a_i/a_j)^2}{[1 + 2(a_i/a_j)]^2} & \text{for } a_i < a_j \end{cases} \quad (35.3)$$

In the context of prey–predator interactions, one can derive the encounter rate kernels for the various cases, to obtain the following relations [31, 61]:

$$\beta_{\text{enc}} = \begin{cases} 4\pi DR & : \text{Random walk} \\ 1.3\gamma R^3 & : < \text{Kolmogorov scale} \\ 1.37\pi R^2(\varepsilon R)^{1/3} & : > \text{Kolmogorov scale} \\ 4/3\pi R^3 f & : \text{Stop-and-go motion} \\ \pi R^2 u & : \text{Swimming, sinking or feeding current} \end{cases}$$

Here R , f , u , γ , and ε indicate the characteristic dimension, the stop frequency, the velocity difference between the encountering agents, the sub-Kolmogorov shear rate, and the turbulent energy dissipation rate, respectively. At scales smaller than the Kolmogorov length scale, the velocity difference between the encountering agents increases with the separation between the agents; while at scales relatively larger to the Kolmogorov scale, an additional contribution from the turbulent energy dissipation, ε , needs to be accounted for as shown in the expression above. The general framework of encounter rate calculations considers a spherical encounter zone with an effective size [31], which has been recently extended to account for non-spherical morphologies [62]. The initial models, which focused on the encounter rates under diffusive and laminar flow fields, were extended to turbulent eddies by Rothschild and Osborn [69], wherein velocities of both agents (for instance, the prey and the predator) are included. The extended encounter rate model demonstrated that turbulence could increase the contact rates by 50% or more, depending on the size and swimming rate [70]. More recently, Arguedas-Leiva *et al.* have shown that the encounters between neutrally buoyant elongated cells are up to ten-fold higher relative to spherical cells, with further enhancement in encounter rates for those that sink instead of being neutrally buoyant [63].

35.3 Gravitaxis in planktonic active matter

Gravitaxis refers to the movement of organisms in response to the gravity vector. Most phytoplankton species exhibit gravitaxis as a means to execute daily vertical migration. Traditionally, the movement of gravitactic species along the gravity vector was referred to as *geotaxis* [71, 72, 73], however, in light of the generality of this tactic response to gravity forces — not only of the Earth but also to that due to other planetary bodies or artificial accelerations — the term *gravitaxis* has become more widely used [8]. Today, alongside *gravitaxis*, the term *gyrotaxis* is frequently used, particularly to appropriately capture and describe the interplay of gravity with fluid forces that are ubiquitous in the watery environments that the plankton inhabit [74, 75, 1, 76, 77, 78].

The ability of organisms to swim against (negative gravitaxis) or along the gravity vector (positive gravitaxis) depends on the cells' developmental phase, physiological state, time of the day or season, and response or adaptation to exogenous stressors [79, 22, 23, 25]. In the following sections, we will, step-by-step, discuss how planktonic active matter perceive gravity forces, and leverage the gravity–flow interactions to navigate different fluid dynamic and ecological settings.

35.3.1 Gravity-sensing mechanisms

Planktonic microbes sense and respond to the gravitational forces and changes therein using a series of steps — perception, transduction, signal amplification, and response — employing different receptors capable of detecting the gravity signals either directly or indirectly. Organelles for direct sensing include heavy BaSO₄ crystals which function as statoliths, or SrSO₄ crystals which are also used in the statocyst-like organelles of the ciliates *Loxodes* and *Remanella* [79]. Such heavy organelles — found across diverse organisms spanning ciliates and algae — operate by directionally moving (sedimenting) within cells, thereby initiating a mechano-signal transduction chain that ultimately allows cells to distinguish between up versus down, or sense changes in their local accelerations [80, 81]. In some larger planktonic species lacking heavy statoliths, the entire cytoplasmic content of the cell can proxy as a gravity-sensing organelle, exerting pressure on the lower membrane, thereby activating the mechano-(gravi-)sensitive ion channels distributed in the cell membrane [82].

The exact mechanism by which small phytoplankton perceive gravity forces and changes therein (in turn, changes in their orientation relative to the gravity vector) remains unclear, and thus warrants further investigation. Large ($> 75 \mu\text{m}$) unicellular protists like the *Paramecium* and *Tetrahymena* sense gravity by an active physiological mechanism through calcium or potassium mechanosensitive ion-channels [83, 53, 84], which get activated due to the gravitational pressure of the cytoplasm on the lower membrane. In the flagellate *Euglena gracilis*, typically 35 to 50 μm in size, mechanochemical changes in the membrane potential are involved in graviperception [85, 86]. For cells in the size range of the raphidophyte *Heterosigma akashiwo* (10 to 15 μm , Fig. 35.6), the gravitational force on the lower membrane can be approximated by

$$F = (\rho_{\text{cell}} - \rho_{\text{fluid}})Vg, \quad (35.4)$$

where ρ_{cell} is the density of the cell and ρ_{fluid} the density of the surrounding fluid, V is the cell volume, and g represents the acceleration due to gravity (or, any relevant acceleration, in general). For microscale ciliates, this yields a force of the order of tens of pN, while for the larger ciliates, e.g., *Paramecium caudatum*, this results in a force over 100 pN [87].

The magnitude of the force yields work of the order of the thermal noise, assuming that the entire cytoplasmic material functions as a buoy. The work due to the gravitational force on the lower membrane for 1 nm gating distance of the mechanosensitive ion channels can be estimated as $4 \times 10^{-22} \text{ J}$, while the thermal noise, $k_B T / 2$, at room temperature (293 K) is $\approx 2 \times 10^{-21} \text{ J}$, where k_B is the Boltzmann constant [22]. This suggests the possibility of alternative mechanisms for gravity-sensing, particularly for microplankton which lack any other direct sensing mechanism. One such alternative could be the cross-talk between the ion-channels and the production of reactive nitrogen species, wherein the sensing may involve positive feedbacks between these two cellular networks [23, 88].

35.3.2 Biomechanics of gravitaxis

Assuming a body of revolution swimming in a fluid at a low Reynolds number, the translational and rotational equations of motion are decoupled [71, 89, 49, 90], allowing us to write the following equations along the major and the minor axes (Fig. 35.6) under force-free conditions:

$$\left\{ \begin{array}{lcl} F_{\text{prop}} \sin \phi & = & F_{\text{drag}} \sin \theta & \text{horizontal axis} \\ F_{\text{prop}} \cos \phi - F_{\text{drag}} \cos \theta & = & (\rho_{\text{cell}} - \rho_{\text{fluid}})Vg & \text{vertical axis} \end{array} \right. \quad (35.5)$$

where F_{prop} is the magnitude of the propulsion force \mathbf{F}_{prop} originating due to the beating of the flagellum, acting along the major axis of the cell body at an angle ϕ to the gravity vector, F_{drag} is the magnitude of the drag force \mathbf{F}_{drag} acting through the center of the hydrodynamic stress C_H , directed opposite to the cell swimming velocity, at an angle θ relative to the vertical. The volume of the cell and the densities of the cell and the surrounding fluid

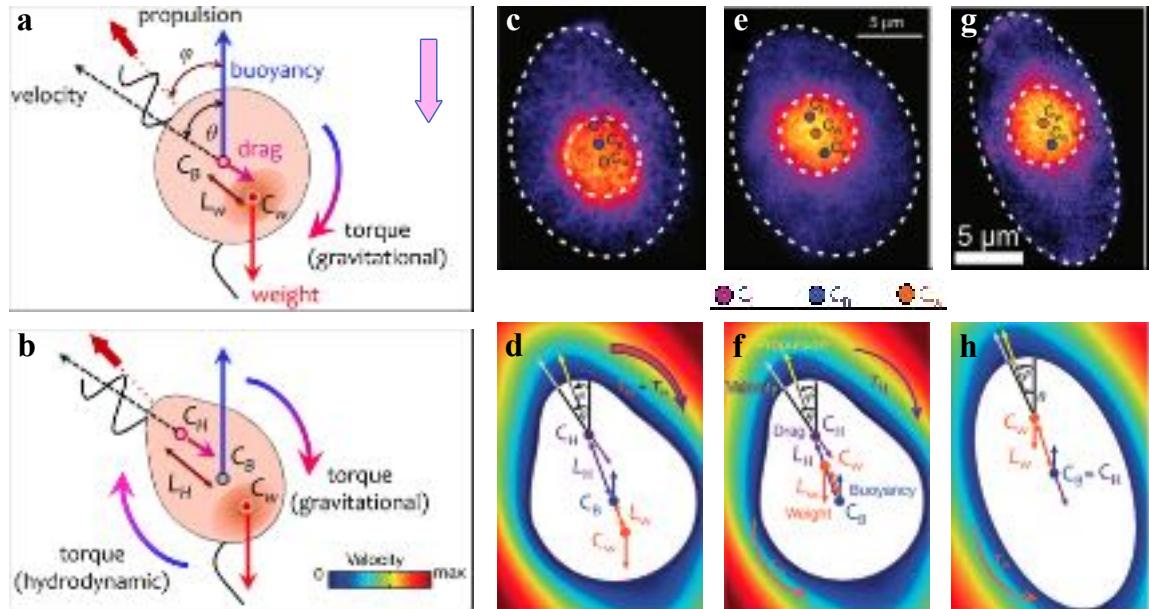


Figure 35.6: Cell mechanics of gravitaxis. (a) Schematic of a symmetrically-shaped bottom-heavy microplankton swimming at low Reynolds number. The cell experiences a propulsion force due to the beating flagellum (or flagella/cilia) and body weight, both acting at the center of mass, C_w . The buoyancy force acts at the geometric center of the cell body (centroid), C_B , which is offset from the center of mass by a distance L_w . The buoyancy force (opposite to the gravity force) and the drag force (opposite to direction of swimming) act through C_B . When the cell is displaced by an angle θ from its equilibrium swimming direction, a stabilizing torque reorients the cell back to the equilibrium orientation (here, in the clockwise direction) due to the bottom heaviness (C_w below C_B). (b) For a cell with fore-aft asymmetry, two separate offset lengths emerge — L_w (as above) and L_h — due to the offset between the centroid and the center of hydrodynamic stress C_h . For a symmetric cell shape, L_w and L_h coincide. The two different offset lengths generate distinct reorientation torques about the centroid of the cell: gravitational (T_w) and hydrodynamic (T_h). The two torques can either reinforce each other (act in the same direction, in the case that the cell is bottom heavy), or counteract if the cell is top-heavy (C_w above C_B). (c) Epifluorescent image of bottom-heavy microplankton cell with fore-aft asymmetry, showing C_w , C_B and C_h , and the corresponding free-body diagram, shown in panel (d). The central orange hue represents the cell nucleus, the heaviest organelle within the cell body. Both T_w and T_h act to orient cells upwards. (e, f) Top-heavy asymmetric cell swimming at an angle θ , such that the rotation rate ω are set by the competition between the gravitational and hydrodynamic torques about C_B , causing cells to orient upwards. (g, h) For the top-heavy symmetric cells, the T_h vanishes, thus causing the cells to orient downwards. Panels (c-h) have been reproduced from Ref. [22].

are V , ρ_{cell} , and ρ_{fluid} , respectively. The drag force \mathbf{F}_{drag} on the moving body in a fluid with dynamic viscosity η at a velocity \mathbf{v} depends on the angle $\alpha = \theta - \phi$ between the body axis and the direction of swimming and can be broken down into two orthogonal components as

$$\mathbf{F}_{\text{drag}} = \mathbf{F}_{\text{drag}}^{\parallel} + \mathbf{F}_{\text{drag}}^{\perp}, \quad (35.6)$$

where $\mathbf{F}_{\text{drag}}^{\parallel}$ and $\mathbf{F}_{\text{drag}}^{\perp}$ are the drag forces corresponding to motion along and perpendicular to the direction of the major axis of the body, respectively. Note that the magnitude of these two components, $F_{\text{drag}}^{\parallel}$ and F_{drag}^{\perp} , are related to F_{drag} via $F_{\text{drag}}^{\parallel} = F_{\text{drag}} \cos \alpha$ and $F_{\text{drag}}^{\perp} = F_{\text{drag}} \sin \alpha$.

Another independent set of equations arises due to the balance of the torques acting at the cell's center of buoyancy (i.e., the geometric center), C_B , giving

$$\mathbf{T}_V = \mathbf{T}_H + \mathbf{T}_W, \quad (35.7)$$

where \mathbf{T}_H is the torque generated by the drag force \mathbf{F}_{drag} and \mathbf{T}_W is the torque generated by the weight of the cell $\mathbf{F}_w = V\rho_{\text{cell}}\mathbf{g}$. On the other hand, the propulsion force \mathbf{F}_{prop} generates no torque about C_B as it passes through it. Note that, referring to Fig. 35.6, each torque vector is perpendicular to the plane formed by the horizontal and vertical axes. The net torque balance equation can be then written as

$$F_{\text{drag}} \sin(\theta - \phi) L_H - F_w [\sin(\phi - \arctan(L_{Nb}/L_{Na}))] L_W = \tilde{R} \eta \omega, \quad (35.8)$$

where $\arctan(L_{Nb}/L_{Na})$ is the contribution to the gravitational torque coming from the offset L_{Nb} of the nucleus within the equatorial plane (L_{Na} and L_{Nb} are the offsets of the nucleus along the longitudinal and equatorial planes, respectively) L_W is the offset of the centroid of the body with respect to the center of mass (see Fig. 35.6), and $\tilde{R} \eta \omega$ is the net viscous torque, \tilde{R} being the coefficient of resistance of the body to rotational motion, and ω the rotation rate (i.e., angular velocity) of the cell. The length-scale L_H is the offset distance between the center of buoyancy C_B and the hydrodynamic stress center C_H (Fig. 35.6c-h). The centre of hydrodynamic stress is the point at which the resultant of all viscous stresses exerted by the fluid on the cell (resulting from the combination of translational motion, reorientation and sedimentation) acts. The center of hydrodynamic stress for bodies with spherical or cylindrical symmetry lies along the axis of symmetry. In case of cells with fore-aft asymmetry, one can obtain the position of C_H by numerically solving the Navier-Stokes equations around the cell body, taking into account the characteristic size and shape determined experimentally by quantitative image analysis. When torque-free condition (sum of all torques on the cell vanishes) are applied, one obtains the coordinates of C_H , determined by minimizing the surface integral of the cross-product between the stress force and the surface of the cell.

For a cell swimming with a speed v , we can numerically solve the system of above equations for the unknowns F_{prop} , ϕ and ω , yielding the angular velocity, $\omega(\theta)$, as a function of the swimming angle, θ , with respect to the direction of the gravity force. The orientational stability of the swimming cell can be extracted from the reorientation timescale τ_{reor} fitting a sinusoid in the plots of $\omega(\theta)$ versus θ for individual cells. At the level of single cells, the reorientation trajectories corresponding to the different fitting parameters represent different reorientation timescales (Fig. 35.7e). At a population-scale, the overall anisotropy of the swimming velocity, represented in Fig. 35.7d, hints at the swimming stability of the individual cells: straight, anisotropic trajectories and consequent distributions are correlated with high stability (Fig. 35.7b), i.e., low reorientation timescales, and ballisticity of swimming (Fig. 35.7e), and vice versa.

When a body of uniform density and arbitrary shape is immersed in a liquid, the upthrust on the body acts through the centre of gravity, and there is no tendency for buoyancy forces to rotate the body. If there is a variation in density, however, a torque is experienced, the magnitude of which depends upon the density differences within the body. The magnitude of the reorientation torque produced by a cell of arbitrary symmetric volume V and density ρ_{cell} is

$$T_W = (\rho_{\text{cell}} - \rho_{\text{fluid}}) V L_W g \sin \theta, \quad (35.9)$$

Taking the opposing viscous torque into account, one can write:

$$\tilde{R} \eta \omega = (\rho_{\text{cell}} - \rho_{\text{fluid}}) V g L_W \sin \theta,$$

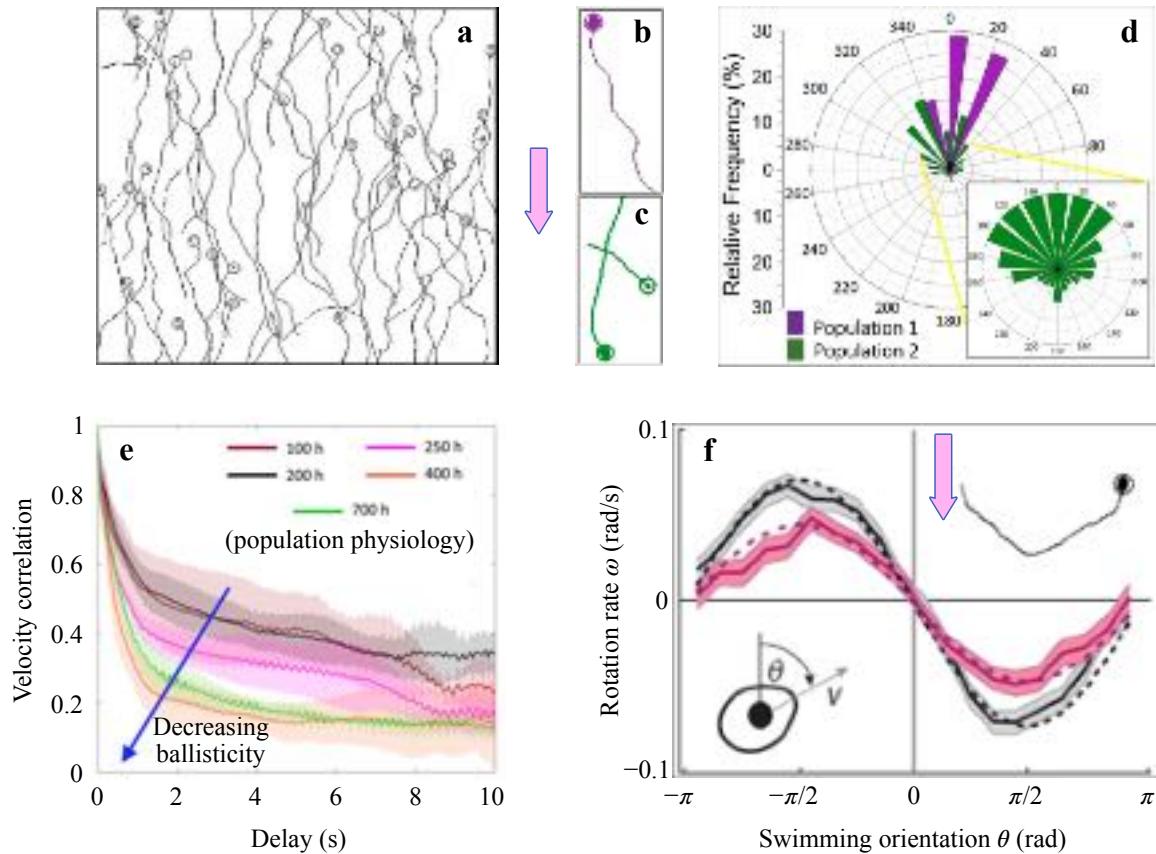


Figure 35.7: Behavior and stability of gravitactic swimming. (a) Trajectories of up-swimming gravitactic microplankton (swimming against the gravity direction). Trajectory patterns (curvature and helicity) can be used to extract the orientational stability of swimming cells. (b) Characteristic helical swimming trajectory of a gravitactic cell. (c) Sample trajectories of down-swimming microplankton (swimming in the gravity direction). (d) Windrose plot presenting the strength and angular spread of the microplankton motility of physiologically distinct swimming populations: strongly ballistic swimming (purple) vs weakly ballistic swimming (green). Higher angular spread of the trajectories of the second population reduces the ballisticity (low swimming anisotropy). The zoomed-in view of the windrose center captures the angular distribution of the second population. Image adapted from [25]. (e) Age-dependent swimming modulation in motile phytoplankton. A shift from ballistic ($t = 100$ h) to diffusive ($t = 700$ h) swimming is captured by plotting the velocity correlations over the delay time [25]. (f) Rotation rate, ω , as a function of the direction, θ , of the instantaneous swimming velocity, v relative to the vertical direction. The rotation rate of the cells, quantified by tracking them in different time intervals (short time shown in grey, while longer time is shown in magenta), averaged over all the cells as a function of θ . The difference between the two curves denotes the presence of cells that reorient more rapidly and others that reorient more slowly. Dashed lines are sinusoidal fits to the experimental data, used to obtain the reorientation timescale τ_{reor} . Solid lines denote the arithmetic mean over all cell trajectories. The reorientation process is shown in the inset to the plot. Image adapted from Ref. [22].

hence,

$$\frac{d\theta}{dt} = \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})VgL_w}{\tilde{R}\eta} \sin \theta, \quad (35.10)$$

where $(\rho_{\text{cell}} - \rho_{\text{fluid}})VgL_w/(\tilde{R}\eta)$ signifies the instantaneous rate of reorientation, and its inverse gives the reorientation timescale τ_{reor} :

$$\tau_{\text{reor}} = \frac{\tilde{R}\eta}{(\rho_{\text{cell}} - \rho_{\text{fluid}})VgL_w}.$$

Changes in both the cell density and cell shape can impact orientational stability of swimming cells [89, 22]. For instance, when a cell expels water or gas molecules by regulating intracellular vacuoles, both the density distribution within the cell and the cell shape undergo considerable change. The distribution of the body mass, for instance top-heavy versus bottom-heavy mass distributions, can significantly impact motility properties, and can even alter swimming behaviors [44, 91, 25]. Given the resemblance of phytoplankton morphologies with prolate spheroids, these shapes have been widely used to model the mechanics of gravitaxis. While the sedimentation rate of symmetrically-shaped prolate spheroids (which are denser than the surrounding fluid) can be estimated analytically [92], the calculations for distorted prolate spheroids with fore-aft asymmetry are more involved. One needs to account for the additional reorientation term, since asymmetric prolate spheroids rotate while sedimenting downwards at instantaneous rates given by the degree of asymmetry [71, 92, 89, 22]. The sedimentation characteristics of an asymmetric cell body, with rotational symmetry can be derived using the *dumb-bell* model described by Happel and Brenner [92], which approximates the asymmetric fore-aft to be constructed equivalently out of two spherical blobs of different radii, connected by a light, rigid inextensible rod. For a single isolated spherical cell of radius R and density ρ_{cell} , settling with a velocity v , the viscous drag $F_{\text{drag}} = 6\pi\eta R v$ (under low Reynolds number conditions) is balanced by the weight of the cell body, $F_w = \frac{4}{3}\pi R^3 (\rho_{\text{cell}} - \rho_{\text{fluid}}) g$, giving us

$$\frac{4}{3}\pi R^3 (\rho_{\text{cell}} - \rho_{\text{fluid}}) g = 6\pi\eta R v.$$

Hence, the speed v is

$$v = \frac{2}{9} \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})g}{\eta} R^2. \quad (35.11)$$

Now, if the dumb-bell shaped cell made of the two spherical components (radii R_1 and R_2 , each with density ρ_{cell}) interconnected by the light inextensible rod is allowed to fall through the medium, one obtains:

$$v_{\parallel} = \frac{2}{9} \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})g}{\eta} \left(\frac{R_1^3 + R_2^3}{R_1 + R_2} \right), \quad (35.12)$$

where v_{\parallel} is the steady-state sedimentation speed the two-sphere dumb-bell with its major axis vertical (i.e., parallel to the gravity vector). Realistically, one needs to consider the hydrodynamic interaction between the interconnected spheres, which is ignored in the analysis of Happel and Brenner [92]. If the total length of the model cell is L , with $e_1 = R_1/L$ and $e_2 = R_2/L$, the equation can be re-written as

$$\begin{aligned} v_{\parallel} &= \frac{2}{9} \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})g}{\eta} L^2 \left(\frac{e_1^3 + e_2^3}{e_1 + e_2} \right) \\ &= \frac{2}{9} \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})g}{\eta} L^2 f_{\parallel} \end{aligned} \quad (35.13)$$

where $f_{\parallel} = \frac{e_1^3 + e_2^3}{e_1 + e_2}$ is the dimensionless shape-factor independent of the physical size of the cell.

By following the above steps, one can derive the corresponding shape-factor when the cell is sedimenting perpendicular to the gravity vector (note that the perpendicular orientation is unstable, with the instantaneous velocity varying from one end to the other). The instantaneous velocity of the dumb-bell is given by the average of the two spheres, as

$$\begin{aligned} v_{\perp} &= \frac{2}{9} \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})g}{\eta} L^2 (e^2 + f^2) \\ &= \frac{2}{9} \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})g}{\eta} L^2 f_{\perp} \end{aligned} \quad (35.14)$$

where $f_{\perp} = (e_1^2 + e_2^2)$ is the dimensionless shape-factor independent of the physical size of the cell.

Extending the above analysis, one can derive the orientational characteristics of the asymmetric dumb-bell, as a function of the angle between the major axis of body and the gravity direction, θ :

$$\frac{d\theta}{dt} = -\frac{1}{\tau_{\text{reor}}} \sin \theta, \quad (35.15)$$

where the maximum reorientation rate (at $\theta = \pi/2$) is given by $\frac{1}{\tau_{\text{reor}}} = \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}})}{\eta} g L f_{\theta}$. The shape-factor f_{θ} can be obtained experimentally.

Happel and Brenner [92] have provided full analytical solutions for the hydrodynamic drag on a prolate spheroid of semi-major and semi-minor axes a and b moving both parallel and perpendicular to the major axis. The directional drag coefficients which an asymmetric prolate cell experiences is given by

$$\begin{aligned} \xi_{\parallel} &= 6\pi\eta R_{\text{eq}} f_{\parallel}, \\ \xi_{\perp} &= 6\pi\eta R_{\text{eq}} f_{\perp}, \end{aligned} \quad (35.16)$$

where $R_{\text{eq}} = (ab^2)^{\frac{1}{3}}$ is the radius of a sphere with volume equal to the prolate spheroid and the directional shape-factors are defined as

$$\begin{aligned} f_{\parallel} &= \frac{1}{48\lambda} \left[\frac{-2\lambda}{\lambda^2 - 1} + \frac{2\lambda^2 - 1}{(\lambda^2 - 1)^{3/2}} \ln \left(\frac{\lambda + \sqrt{\lambda^2 - 1}}{\lambda - \sqrt{\lambda^2 - 1}} \right) \right] \\ f_{\perp} &= \frac{1}{48\lambda} \left[\frac{\lambda}{\lambda^2 - 1} + \frac{2\lambda^2 - 3}{(\lambda^2 - 1)^{3/2}} \ln \left(\lambda + \sqrt{\lambda^2 - 1} \right) \right] \end{aligned} \quad (35.17)$$

where $\lambda = a/b$. The coefficient of resistance of a prolate spheroid to rotational motion is given by [93]

$$R(\lambda) = 8\pi R_{\text{eq}}^3 \frac{2(\lambda^2 + 1)(\lambda^2 - 1)^{3/2}}{3\lambda \left[(2\lambda^2 - 1) \ln \left(\lambda + \sqrt{\lambda^2 - 1} \right) - \lambda \sqrt{\lambda^2 - 1} \right]}. \quad (35.18)$$

35.3.3 Quantifying cell morphology in experiments

Swimming cells imaged in real time can allow us to acquire various projections of the cell morphology, and thereby reconstruct it in three dimensions. The imaged cells can be then analyzed to extract the cell contours (from multiple projections), followed by appropriate curve-fitting of the extracted contours (Fig. 35.8). This technique has been fairly successful in quantifying the cell morphology of microplankton, starting with the pioneering works of Roberts *et al.* [89, 90], and more recently by Sengupta *et al.*, where changes in rotational symmetry of the cell morphology have been experimentally quantified, thus allowing a complete analysis of swimming stability based on the three-dimensional morphological parameters (Fig. 35.9) [25]. A three-parameter equation has been found to reliably capture the projected morphological features of microplankton:

$$S(\varphi, \vartheta) = \frac{ab}{\sqrt{a^2 \sin^2 \vartheta + b^2 \cos^2 \vartheta}} + c \cos \vartheta \quad (35.19)$$

where the first term on the right describes an ellipse with semi-major and semi-minor axes of lengths a and b respectively, and the second term confers a degree of fore-aft asymmetry specified by the length c . The angles ϑ and φ are the polar and azimuthal angles measured from the major axis, and $S(\varphi, \vartheta)$ represents the distance of a point on the surface from the origin with a polar angle ϑ and azimuthal angle φ . Note that the expression of $S(\varphi, \vartheta)$ has no explicit dependence on the azimuthal angle φ because we assume rotational symmetry around the polar axis. Varying the relative ratio of a/b alters the eccentricity of the symmetric ellipsoid of revolution, whereas altering the relative ratio c/a changes the fore-aft asymmetry of the cell morphology. Thus, plotting the values of b/a and c/a gives the family of curves with varying eccentricities and fore-aft asymmetries, as shown in Fig. 35.8a.

It is imperative to note here that the above analyses have been carried out assuming the cell body to be rotationally symmetric, i.e., the cross-sectional plane normal to the major axis of the cell body has a circular shape

(for instance, a pear or an egg). However, many planktonic species lack rotational symmetry, or depending on their physiological status, develop *platelet* shape morphologies with certain degree of flatness [25]. For such cases, one has to account for the lack of rotational symmetry while calculating the gravitactic stability, since the translational and rotational viscous torques are direction dependent. The drag force \mathbf{F}_{drag} on an arbitrary ellipsoid with semi-axes a, b and r (Fig. 35.9), swimming within a fluid with speed u along the major axis (a -direction) is

$$\frac{F_{\text{drag}}}{\pi \eta u} = \frac{16}{\delta + a^2 \chi_a}, \quad (35.20)$$

where δ is given by

$$\delta = \int_0^\infty \frac{1}{\sqrt{(a^2 + x')(b^2 + x')(r^2 + x')}} dx' \quad (35.21)$$

and χ_j , for $j = \{a, b, r\}$, is

$$\chi_j = \int_0^\infty \frac{1}{(j^2 + x') \sqrt{(a^2 + x')(b^2 + x')(r^2 + x')}} dx'. \quad (35.22)$$

The resistive torque \tilde{R} applied by the surrounding fluid due to rotation of the solid body along the major axis with an angular speed ω is given by

$$\tilde{R} = C_0 \frac{b^2 + r^2}{b^2 \chi_b + r^2 \chi_r}, \quad (35.23)$$

where C_0 is a constant pre-factor. The integrals can be calculated numerically, noting that the integral is sensitive to initial discretization of the integration variable x' . Taking the symmetric geometry as a validation case, the discretization can be accurately estimated. Furthermore, the accuracy of the technique can be verified by comparing the resistive viscous torque for the asymmetric case with that of the symmetric case: the two should equate when the semi-minor axes tend to similar values, i.e., when $b \rightarrow r$ (Fig. 35.9e).

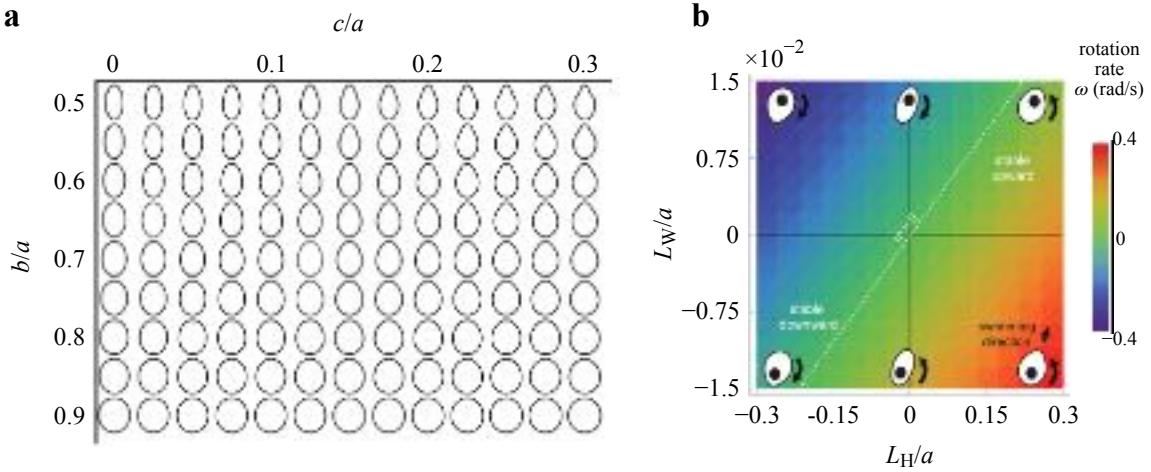


Figure 35.8: Gravitaxis of shape-shifting microplankton. (a) Cell shape variation in terms of the degree of fore-aft asymmetry and minor/major axis ratio. The parameter c is the degree of fore-aft asymmetry, a is the semi-major axis, and b is the semi-minor axis. (b) Two relevant physical features are presented, i.e., the two morphological length scales which determine the cell stability: the asymmetry in shape, quantified by L_H/a , and the mass distribution, quantified by L_W/a , where a is the semi-major axis — L_H is the distance between the centre of buoyancy and the centre of hydrodynamic stress and L_W is the distance between the centre of buoyancy and the centre of mass. The colors denote the cell rotation rate ω following an orientational perturbation. $\omega > 0$ denotes negatively gravitactic cells (stable upward), $\omega < 0$ denotes positively gravitactic cells (stable downward), and $\omega = 0$ (white dashed line) denotes neutrally stable cells. Sample asymmetry configurations corresponding to different locations on the regime diagram are illustrated by the schematics. Adapted from [22].

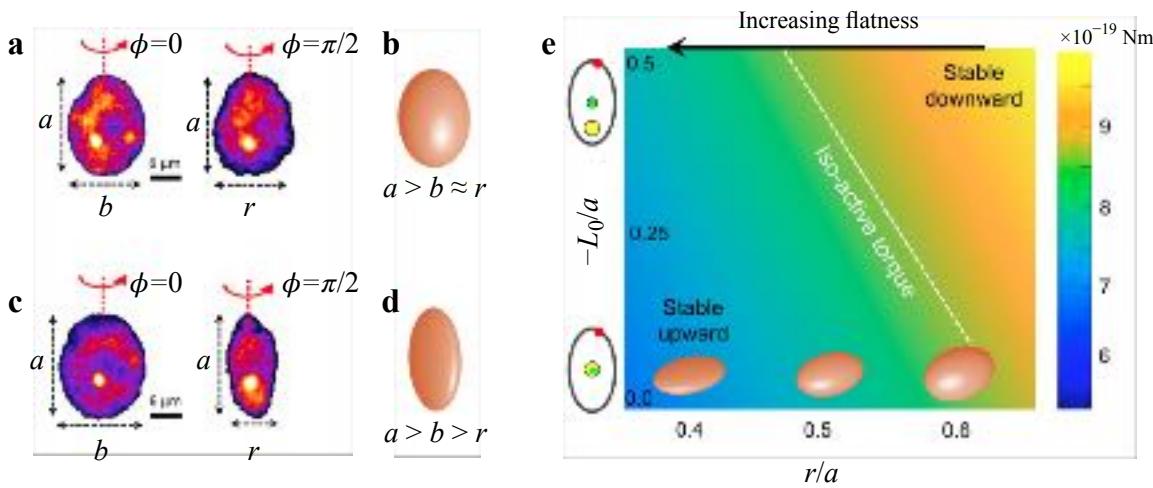


Figure 35.9: **Role of rotational symmetry on gravitaxis.** (a), (b) Axisymmetric morphology results in difference in swimming stability relative to (c), (d) the cells which are flat-shaped (lacking axisymmetry). (e) The active torque (in Nm) required by a cell to reorient itself as an upward swimmer (negative gravitaxis) as a function of the cell flatness r/a and as a function of the effective organelle offset distance L_O (for instance, lipid position from the geometric center) and the semi-major axis a . The dashed white line shows the iso-active torque. Depending on the microplankton species, the active torque requirement decreases as the cells become flatter. Adapted from [25].

35.3.4 Influence of intracellular organelles on gravitactic stability

Planktonic cells contain a range of intracellular organelles within their cell body, including the nucleus (the heaviest organelle within the cell), statoliths (minuscule biomimeticized particles that enable gravity and pressure perception), biomimeticized globules, chloroplasts, gas vacuoles, lipid bodies, and carbohydrate or starch reserves. Depending on the growth conditions and physiological constraints planktonic species experience, intracellular organelles can provide direct biomechanical cues to actively modulate buoyancy, swimming speed and orientational stability of gravitactic species. Furthermore, the orientational stability of the cells could be fine-tuned through dynamic variation of the organelle size and their relative intracellular positioning within the cell cytoplasm. To understand the impact of the organelles on the swimming stability, one can approach the analysis in the framework of an *effective* size, density and position (center of gravity), and how these impact the corresponding cellular parameters. The geometric and the hydrodynamic centers of the cell remain unaffected as long as the overall morphology of the cell remains intact.

Consider a cell with a generic morphology of volume V_C , which, along with its nucleus (volume V_N , density ρ_N), also contains dynamic intracellular organelles (total volume V_O , density ρ_O). In case the effective center of gravity of the organelles, C_O , is located away from the major axis of the cell body, it will generate a rotational moment about the geometric center of the cell (center of buoyancy C_B), yielding the following set of equations:

$$\left\{ \begin{array}{lcl} F_{\text{prop}} \sin \phi & = & F_{\text{drag}} \sin \theta & \text{horizontal axis} \\ F_{\text{prop}} \cos \phi - F_{\text{drag}} \cos \theta & = & (\rho_{\text{cyto}} - \rho_{\text{fluid}}) V_C g + (\rho_N - \rho_{\text{cyto}}) V_N g + (\rho_O - \rho_{\text{cyto}}) V_O g & \text{vertical axis} \end{array} \right. \quad (35.24)$$

where ρ_{cyto} and ρ_{fluid} are the densities of the cell cytoplasm (the intracellular material) and the surrounding fluid, and ϕ and θ are respectively the angles of the cell propulsion (F_{prop}) and the swimming velocity (\mathbf{v}) to the vertical gravity direction (realistically, a cell does not swim exactly in the direction of propulsion). Balancing the torques, one obtains

$$F_{\text{drag}} \sin(\theta - \phi) L_H - F_{W,N} L_N \sin(\phi_N) - F_{W,O} L_O \sin(\phi - \phi_O) = \tilde{R} \eta \omega, \quad (35.25)$$

where L denotes the offset distance from center of buoyancy (C_B), W denotes the weights, the subscripts N, O,

and H refer to the center of gravity due to the nucleus, organelles, and the hydrodynamic center respectively, η and ϕ_O refer to the fluid viscosity and the angle between the cell's long-axis and the line joining C_B and C_O (please refer to Fig 35.6, with the only difference that there is an additional C_O denoting the intracellular organelles.)

The above general system of equations can be extended to include the contributions of other intracellular organelles, for instance, if the cell has n distinct organelles scattered within the cell at $(L_{O,i}, \phi_{O,i})_{i=1,n}$, one arrives at:

$$\left\{ \begin{array}{l} F_{\text{prop}} \sin \phi = F_{\text{drag}} \sin \theta \\ F_{\text{prop}} \cos \phi - F_{\text{drag}} \cos \theta = (\rho_{\text{cyto}} - \rho_{\text{fluid}}) V_C g + (\rho_N - \rho_{\text{cyto}}) V_N g + \sum_i (\rho_{O,i} - \rho_{\text{cyto}}) V_{O,i} g \\ \tilde{R} \eta \omega = F_{\text{drag}} \sin(\theta - \phi) L_H - F_{w,N} L_N \sin(\phi_N) - \sum_i F_{w,O,i} L_{O,i} \sin(\phi - \phi_{O,i}) \end{array} \right. \quad (35.26)$$

35.4 Gravitaxis-related phenomena

Two related phenomena in which gravitaxis plays a pivotal role are gyrotaxis and bioconvection. In gyrotaxis a combination of gravity and fluid velocity gradients in the medium causes motile algal cells to converge actively into the centre of a downward moving stream [74, 94]. Bioconvection occurs in dense cultures when characteristic columns of sinking and rising organisms form spontaneously, driven by the cells themselves [77, 95, 96]. The study of both these effects has not thrown any light on the actual mechanism of gravitaxis because both back-heavy and shape orientation mechanisms can equally well account for the observations.

35.4.1 Gyrotaxis: An interplay of gravitactic and viscous torques

Phytoplankton species exhibiting gravitaxis are often associated with hydrodynamic shear or turbulent flows [43, 44, 91, 45, 46, 97, 48]. Here, we will outline the interaction between gravity and hydrodynamic interactions of motile cells. A gravitactic cell in a flowing liquid experiences viscous torque due to the spatial variation of the fluid velocity. In the simple case of a spherically symmetric (radius R , volume $V = \frac{4\pi}{3}R^3$), bottom-heavy cell (mass m) swimming with constant speed v in the reference frame of the cell, the net torque acting on the cell is given by

$$\mathbf{T}_{\text{gyro}} = 8\pi\mu R^3 \left[\frac{\nabla \times \mathbf{u}}{2} - \boldsymbol{\omega} \right] + (\rho_{\text{cell}} - \rho_{\text{fluid}}) V \mathbf{L}_W \times \mathbf{g}, \quad (35.27)$$

where \mathbf{u} and μ are the velocity and dynamic viscosity of the surrounding fluid, $\nabla \times \mathbf{u}$ is the vorticity, $\boldsymbol{\omega}$ is the angular speed of the spherical cell, \mathbf{L}_W is the position of the center of mass with respect to the geometric center of the cell, and \mathbf{g} is the acceleration. Under low Reynolds number (static torque balance, i.e., $\mathbf{T}_{\text{gyro}} = 0$) and weak vorticity conditions, i.e., $\boldsymbol{\omega} = 0$, one obtains the gyrotactic length scale

$$\lambda_{\text{gyro}} = \frac{4\pi\mu R^3}{(\rho_{\text{cell}} - \rho_{\text{fluid}}) V g L_W} v = \frac{3\mu}{(\rho_{\text{cell}} - \rho_{\text{fluid}}) g L_W} v = \tau_{\text{gyro}} v, \quad (35.28)$$

where

$$\tau_{\text{gyro}} = \frac{3\mu}{(\rho_{\text{cell}} - \rho_{\text{fluid}}) g L_W}$$

is the gyrotactic timescale, while the equilibrium swimming direction θ_{eq} is given by

$$\sin \theta_{\text{eq}} = \frac{3\mu}{(\rho_{\text{cell}} - \rho_{\text{fluid}}) g L_W} \times \frac{du}{dr} = \tau_{\text{gyro}} S, \quad (35.29)$$

where $S = \frac{du}{dr}$ is the gradient of flow velocity around the cell. Gradients in the fluid velocity can disrupt the gravitactic migration of phytoplankton, causing them to accumulate in layers. Whether gravitactic species will form thick layers depends on the relative strengths of the viscous and gravitactic torques. When $S > 1/\tau_{\text{gyro}}$ (i.e., $\sin \theta_{\text{eq}} = S\tau_{\text{gyro}} > 1$), the stabilizing gravitational torque that acts to orient cells upwards is outweighed by the hydrodynamic torque that drives them to reorient, leading to disruption of the upward migration. The cells

under this condition tumble end over end, accumulating where they tumble. For $S < 1/\tau_{\text{gyro}}$, the swimming cells evade the *gyrotactic trap* and can continue swimming following the gravitactic cues. The critical gradient of fluid velocity, S_{crit} , which induces gyrotactic trapping is thus given by

$$S_{\text{crit}} = 1/\tau_{\text{gyro}} = \frac{(\rho_{\text{cell}} - \rho_{\text{fluid}}) g L_W}{3\mu}. \quad (35.30)$$

The gyrotactic motility within a steady vortical flow leads to tightly clustered aggregations of microorganisms. Two dimensionless numbers, characterizing the relative swimming speed and stability against overturning by vorticity, govern the coupling between motility and flow and determine different patchiness regimes. Aggregations can form within a few minutes (multiple overturning timescales), suggesting that vortical flows might be capable of efficiently separating species with different motility characteristics μ . In moving fluids, cells can experience rotation due to gradients in fluid velocity, following the generalized equations:

$$\frac{d\mathbf{p}}{dt} = \frac{1}{2\tau_{\text{reor}}} [\mathbf{k} - (\mathbf{k} \cdot \mathbf{p})\mathbf{p}] + \frac{1}{2}\nabla \times \mathbf{u} \times \mathbf{p} + \alpha \mathbf{p} \cdot \mathbf{E} \cdot [\mathbf{I} - \mathbf{pp}], \quad (35.31)$$

where $\nabla \times \mathbf{u}$ is the fluid vorticity, \mathbf{E} is the rate of strain tensor, \mathbf{I} is the identity matrix, t is time, τ_{reor} the characteristic reorientation timescale (to return to equilibrium orientation \mathbf{k} if $\nabla \times \mathbf{u} = 0$, and $\alpha = (\lambda^2 - 1)/(\lambda^2 + 1)$, λ being the ratio of the major and minor axes of the cell body (assuming prolate ellipsoid geometry; $\lambda = a/b$), and \mathbf{p} is the unit vector along the swimming direction. The above equation applies to organisms much smaller than the scale of ambient velocity gradients, which allows cells to be modeled as point particles. Note that for cells with no preferred swimming direction ($1/\tau_{\text{reor}} = 0$), the Jeffery orbits are recovered [98, 99]. The trajectories of single cells, appearing as periodic rotations due to the orientation-dependent torque, are known as Jeffery orbits. While motility is often isotropic in absence of shear, the presence of shear leads to preferential flow alignment due to the coupling between the shear flow and particle orientations, thus tuning the migratory traits of cells with anisotropic shape. In the framework of the Taylor-Green Vortex flow [100, 101], one can write the non-dimensional equations of motion as under:

$$\begin{aligned} \frac{d\mathbf{p}}{dt} &= \frac{1}{2\Psi} [\mathbf{k} - (\mathbf{k} \cdot \mathbf{p})\mathbf{p}] + \frac{1}{2}\boldsymbol{\omega}(\mathbf{X}) \times \mathbf{p} + \alpha \mathbf{p} \cdot \mathbf{E}(\mathbf{X}) \cdot [\mathbf{I} - \mathbf{pp}] \\ \frac{d\mathbf{X}}{dt} &= \Phi \mathbf{p} + \mathbf{u}(\mathbf{X}) \end{aligned} \quad (35.32)$$

where $\mathbf{X} = [x, y, z]$, $\Psi = \tau_{\text{reor}}\omega_0$, $\Phi = mV_C/\omega_0$, ω_0 is the maximum vorticity in the system, and the time is non-dimensionalized by ω_0^{-1} . The parameters Φ and Ψ respectively measure the swimming speed relative to the flow field and the orientational stability of the swimming cell. The critical vorticity at which a cell is overturned by the vorticity is given by:

$$\omega\Psi > 1. \quad (35.33)$$

Various combinations of the Φ and Ψ values signify the diverse behaviors resulting in the existence of multiple regimes of phytoplankton aggregation in vortical flows, arising due to the complex interplay between motility and flow.

Phytoplankton rotational dynamics in flow, and thus their spatial distribution, can be affected by multiple elements related to cell morphology and motility. First, the presence of flagella alters the effective aspect ratio of the cell compared to the cell body alone (the latter is all that is often visualized when tracking cells). For example, a flagellum protruding along a cell's major axis renders the cell effectively more elongated than its body, as shown for bacteria. Second, the drag associated with flagellar movement also impacts a cell's rotational dynamics in flow. For example, the waveforms exhibited during a beat cycle by the two flagella of the breaststroke swimmer *Chlamydomonas* reduce the organism's effective aspect ratio when compared to the geometrical aspect ratio based on the cell body alone. Third, rotational dynamics can be affected by orientational noise in the swimming direction, which may arise from unsynchronized or asymmetric flagellar beating. Noise typically reduces the effective aspect ratio of cells, making them effectively more spherical by counteracting the aligning effects of fluid flow. Finally, a cell's rotational dynamics may be affected by active changes in behavior as a result of various external stimuli, including light, chemical gradients, or potentially fluid shear. Certain dinoflagellates exhibit a non-uniform orientation distribution in shear that is inconsistent

with predictions from Jeffery orbits (shear dependent orientation of elongated particles), suggesting an active response to flow. *Dunaliella primolecta* was found to swim in the direction of the local vorticity vector when the local shear exceeded 10s^{-1} , an observation difficult to explain based on physics alone. Finally, directional persistence in shear flows has been observed for both *Dunaliella* and *Chlamydomonas*, which contrasts with the continuous rotation expected of cells undergoing Jeffery orbits and may be the result of the cells actively resisting the viscous torques from the shear flow.

35.4.2 Phytoplankton swimming in turbulent eddies

In nature, plankton usually resides in a turbulent environment, experiencing a fluctuating fluid flow along their trajectories. In general, the orientation of gyrotactic swimmers depends on two timescales, which correspond to the effect of gyrotaxis and fluid velocity gradients, denoted by the timescales τ_{reor} and t_η , respectively. In the limit of weak turbulence, $\tau_{\text{reor}} \ll t_\eta$, the effect of turbulence can be regarded as a perturbation from the stable orientation. However, in case of strongly turbulent environment, swimmers have no deterministic orientation, and it is more appropriate to quantify the probability distribution of the orientation of a population of swimmers. According to Lewis [70, 38], the steady distribution of orientation for spherical gyrotactic swimmers in an isotropic turbulence is

$$G(\theta) = \frac{\beta e^{\beta \cos \theta}}{2 \sinh \beta} \quad \text{with} \quad \beta = \frac{\tau_{\text{reor}}^{-1}}{2D_{\text{eff}}} = \frac{D_{\text{eff}}^{-1}}{2\tau_{\text{reor}}} = \frac{\tau_{\text{rot}}}{2\tau_{\text{reor}}}, \quad (35.34)$$

where G is the probability distribution of θ , representing the swimming direction relative to the reference axis (for instance, the direction of the gravity vector), $D_{\text{eff}} = \tau_{\text{rot}}^{-1}$ is the rotational effective diffusivity, and β represents the ratio between the reorientation rate $1/\tau_{\text{reor}}$ and the rotational effective diffusivity. The distribution $G(\theta)$ represents the orientational distribution of gyrotactic swimmers in random flows. The same distribution can be obtained by deriving the Fokker–Planck equation from the constitutive equations in the limit of turbulence being described as Gaussian noise. In terms of timescales, the correlation time of the turbulent velocity gradients experienced along the trajectory of a swimmer has to be much less than the timescale of the change in the orientation for this conclusion to be valid. Effectively, this leads to $\Psi \gg 1$, or $\Psi < 1$ and $\Phi\Psi \gg 1$. The first condition corresponds to extremely weak gyrotaxis, while the second corresponds to the swimmers moving through the smallest flow scale in a short time so that the velocity gradients decorrelate quickly.

For non-spherical cells, the derivation of the distribution is more involved, since swimmers are now subjected to a non-zero strain rate term as well, which depends on the instantaneous orientation. A direct result is that elongated swimmers in turbulence show a preferential orientation with respect to the first and second eigenvectors of the local strain rate tensor, which is similar to non-motile elongated tracers. This preferential alignment with local fluid structures leads to a longer timescale for an elongated swimmer to align with the preferred direction of gyrotaxis. Recently, Borgnino *et al.* [102] have examined the orientation statistics of spheroidal, axisymmetric microswimmers, with shapes ranging from disks to rods, in chaotic turbulence-resembling flows. They showed that alignment is caused by the combined effect of (i) the correlations of fluid velocity and its gradients along the path followed by the swimmer and (ii) the fore–aft symmetry breaking due to both swimming and non-sphericity of the particle shape.

De Lillo *et al.* [46] observed that swimming plankton accumulates in the core of strong vortices because centrifugal acceleration aligns the swimming direction to the center of the vortices, while dead cells remain randomly distributed. Gyrotactic swimmers in turbulence also form patchiness as reported by numerical studies based on direct numerical simulations [46]. This is nontrivial because turbulence often acts as a mixing process. For instance, passive tracers are advected and dispersed randomly in a turbulent flow because they passively follow the fluid streamlines. On the contrary, swimmers can move across the streamlines and form patches if the swimming direction is preferential due to the gravity or shape effects, as discussed below. To analyze the patchiness on the population level of swimmers, we need to quantify the degree of patchiness. The fractal dimension of the patches, referred to as d hereinafter, is one of the most commonly used indicators. It can be calculated from the radial distribution function (i.e., the possibility of finding a pair of swimmers at a distance r):

$$g(r) = \frac{1}{N(N-1)} \frac{dN_r}{dr} \sim r^d, \quad (35.35)$$

where N is the total number of swimmers counted in the calculation of g , and d is defined as the exponent of $g(r)$ at small r . The value of d is equal to the spatial dimension if swimmers are randomly distributed. However, d is smaller than the spatial dimension when swimmers form patches, which is also called fractal clustering. In addition to the fractal dimension, Voronoï tessellation is also commonly used to analyze the patchiness of particles. The value of d is the smallest when $\Psi \sim 1$, and decreases as Ψ becomes larger within this regime of Ψ . This means that swimmers form stronger patches if they swim fast compared to the Kolmogorov velocity scale and their gyrotactic torque is comparable to the effect of vorticity. It is believed that patchiness is a result of preferential sampling in downwelling or upwelling regions. This is supported by the observation that the trend of vertical fluid velocity sampled by gyrotactic swimmers is similar to the trend of d over Φ and Ψ .

Gustavsson *et al.* [103] also found a dependence on cell morphology in stochastic flows: elongated swimming particles can preferentially sample either downwelling or upwelling flow regions, depending on their swimming speed and stability. This is rather intriguing since some gyrotactic cells can change their morphology and biasing direction when exposed to turbulence [22] and in this way control their motility. Moreover, Zhan *et al.* [104] and Gustavsson *et al.* [103] found that turbulent suspensions of elongated gyrotactic cells are generally less clustered than spherical ones, but for the regime in which gyrotaxis is very weak. As expected, other relevant factors including the shape of the swimmer, inertial effects, and the nature of the turbulent environment (bulk turbulence versus surface turbulence) additionally impact the fate of gyrotaxis and gyrotactic clustering of planktonic active matter in natural settings [105, 106].

Significant variability in the response to turbulence exists among species. Turbulence can be beneficial for some species. For example, turbulence has been observed to increase the growth rates of cell cultures. Additionally, theoretical calculations showed that turbulence may enhance mass transport if cells are sufficiently large (equivalent radius greater than 60 to 100 μm). In particular, for typical dissipation rates of turbulent kinetic energy, theoretical analysis indicates that natural levels of turbulence affect uptake by large microorganisms. A cell with an equivalent radius $R = 50 \mu\text{m}$ absorbing small molecules under strong turbulence ($\varepsilon = 10^{-6} \text{ W kg}^{-1}$) gains 18 to 32% in nutrient uptake.

Negative effects can occur based on at least three mechanisms: physiological impairment, physical damage, and behavioral modification. Turbulence can reduce growth rates and, if sustained, lead to mortality — large shear rates can induce cellular disintegration. Turbulence can also negatively affect cells by disrupting various elements of their complex life cycle, such as the cellular clock, mitotic cycle, and the nucleic acid concentrations. It has been hypothesized that the blockage of dinoflagellate division by shaking is caused by physical disturbance of the microtubule assemblage and/or the mechanisms responsible for chromosome separation and genes involved in cell division, including microtubuli synthesis, are down-regulated upon exposure to turbulence. Other negative effects of turbulence include significant (50%) reductions in swimming velocity in *Alexandrium minutum*, loss of flagella and swimming ability in *Gonyaulax polyedra*, and disruption of vertical migration by trapping.

Turbulence is frequently localized within relatively thin layers. In view of the often deleterious effect of strong turbulence, the ability to escape these localized regions of high turbulence will provide a benefit to a phytoplankton population. Specifically, as they swim into a layer of high turbulence, cells will experience progressively increasing levels of turbulence (because the turbulent intensity necessarily tapers off below and above the layer): an evasion response would thus prevent the entire population from entering the high-turbulence core of the layer and potentially being wiped out. In this context, the behavioral response to moderate levels of turbulence found in experiments ($\varepsilon = 3 \times 10^{-8} \text{ W kg}^{-1}$) can provide a direct benefit to a population. This benefit results from the ability of a fraction of the population to avoid the most damaging turbulence (at the cost of temporarily halting growth), while a fraction of the population attempts the crossing of the turbulent region (at the risk of suffering damage), a strategy justified by the lack of information on the actual intensity of turbulence and the benefit of reaching shallower depths optimal for photosynthesis.

Turbulence has long been known to drive phytoplankton fitness and species succession: motile species dominate in calmer environments and non-motile species in turbulent conditions. Recently, a mechanistic understanding of the effect of turbulence on phytoplankton migratory behavior and physiology was provided wherein quantification of stress accumulation and physiology induced by turbulent cues, together with a mathematical model of stress dynamics, revealed that motile phytoplankton use their mechanical stability to sense the intensity of turbulent cues and integrate these cues in time via stress signaling to trigger switches in migratory behavior. The stress-mediated warning strategy we discovered provides a paradigm for how phytoplankton cope with turbulence, thereby potentially governing which species will be successful in a changing ocean.

Finally, we note that changes in the orientation relative to gravity have been neglected as a potential cue for phytoplankton. This is somewhat surprising in view of the fact that the hydrodynamic acceleration of the fluid in Kolmogorov-scale eddies is typically a minute fraction of the acceleration due to gravity ($\sim 0.001 \text{ g}$). We thus predict that, in assessing positive and negative effects of turbulence on phytoplankton, it will be vital to consider the orientation of cells relative to gravity as a cue that is potentially important for the physiology of the cells.

35.4.3 Bioconvection

The term *bioconvection* describes hydrodynamic instabilities and patterns in suspensions of biased swimming microorganisms (Fig. 35.10). Hydrodynamic instabilities arise from coupling between cell swimming behaviors, physical properties of the cells (such as density) and fluid flows. For instance, a combination of viscous and gravitational torques can lead to cells swimming toward downwelling fluid. If the cells are more dense than the fluid, then a gyrotactic instability results. Phototaxis describes the directed response of cells to light, which can also lead to instability. Bioconvection represents a classic system where macroscopic phenomena arise from microscopic cellular behavior in relatively dilute systems. There are ecological consequences for bioconvection and the mechanisms involved as well as potential for industrial exploitation.

Bioconvection (coined by Platt 1961) [107] is a self-organized structure and flow that arises naturally when biased swimming microorganisms are left unstirred in liquid culture and is usually identified by patterns in cell concentration. In an investigation of patterns in *Euglena viridis* published in 1911, Wager [108] described earlier observations of other organisms by Nägeli (1860) and Sachs (1876). Typically, cell concentration patterns and fluid flow develop tens of seconds to minutes after mixing. Characteristic length scales are roughly two or three orders of magnitude larger than the size of individual organisms. The mechanisms are hydrodynamic instabilities mainly due to either the accumulation of non-neutrally-buoyant cells that form unstable density gradients or an unstable swimming orientation field coupled to fluid flow gradients and cell concentration. Figure 35.10 presents typical patterns obtained in the laboratory under controlled conditions for a range of species. Bioconvection has been described as an example of collective vortex behavior, with parallels for larger organisms (social insects, fish, and bats), and of stigmergy, describing mechanisms of coordination through the environment without direct interaction or communication [109]. Complex structures emerge in the absence of planning or control [40, 110, 111].

In most studies of bioconvection, the aim is to describe the macroscale phenomena at a continuum level using information about the microscopic behavior of individual organisms. As discussed earlier, bioconvection sits mainly at the dilute end of the cell concentration spectrum, and so direct cell-cell interactions are mostly neglected. The main constituents of the standard continuum approach are momentum balance, mass balance, cell conservation, and a description of cell orientation moments as a function of the flow. Typically, the individual cell Reynolds number is small (e.g., $\sim 10^{-3}$ for *C. augustae*) relative to the bioconvection Reynolds number (e.g., ~ 10). The general model is based on the incompressible Navier-Stokes equations with an added (negative) buoyancy term [95, 76] (subject to a Boussinesq approximation) that represents the main effect of the difference in density between the cell and fluid, ρ , for cells of concentration $n(x, t)$ and mean volume V_{mean} . Typically, the cells are a little more dense than the fluid (e.g., *C. augustae* are 5% to 10% more dense than the fluid). Additionally, the presence of the swimming cells may impact the bulk stress, $\Sigma(\mathbf{x}, t)$. Hence, we obtain:

$$\begin{cases} \rho \frac{d\mathbf{u}}{dt} = -\nabla p_e + nV_{\text{mean}}\Delta\rho\mathbf{g} + \nabla \cdot \Sigma \\ \nabla \cdot \mathbf{u} = 0 \end{cases} \quad (35.36)$$

where $\mathbf{u}(\mathbf{x}, t)$ represents the velocity of the suspension, $n(\mathbf{x}, t)$ is the cell concentration, $p_e(\mathbf{x}, t)$ is the excess pressure, ρ is the fluid density, and \mathbf{g} is the gravitational acceleration. Pedley and Kessler [76] were the first to consider cell-induced stresses in Σ . They identified three components: swimming-induced stresslets, Batchelor stresses, and stress associated with rotary particle diffusion. It was found that the last two of these could be neglected (with no new qualitative features from the Batchelor stresses and the same tensorial form for the rotary particle diffusion stress as for the swimming stresslets). It is the first term, the leading-order stress for the aggregate swimming stroke of the cells, $\Sigma^{(p)}$, that has been found to play a significant role in concentrated suspensions [113]. Thus, one can write:

$$\nabla \cdot \Sigma = \eta \nabla^2 \mathbf{u} + \nabla \cdot \Sigma^{(p)}, \quad (35.37)$$

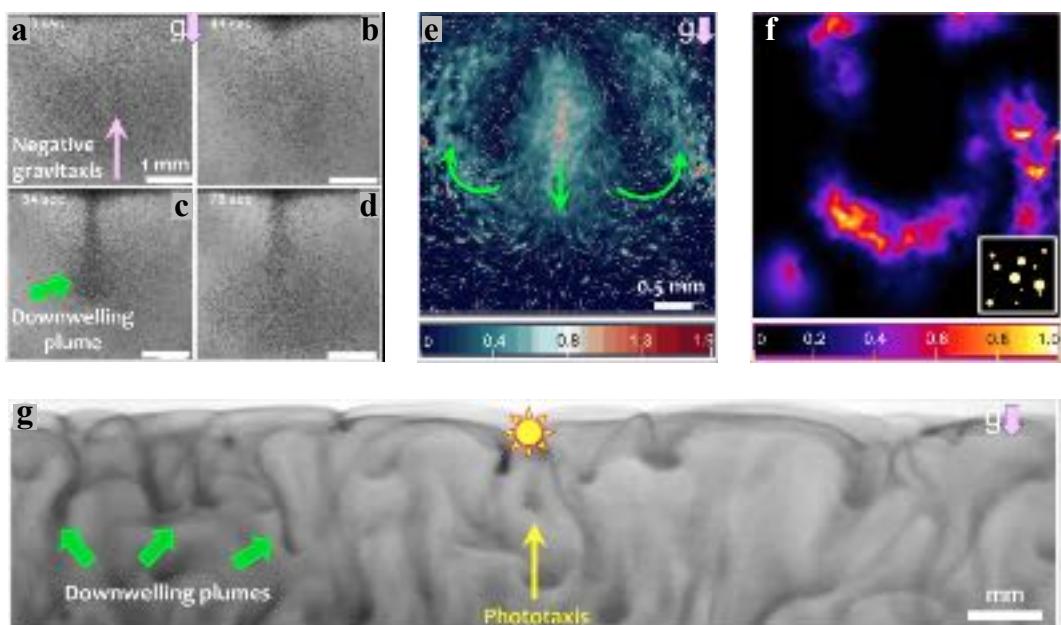


Figure 35.10: Bioconvection of gravitactic microplankton. (a)-(d) Evolution of bioconvective plume due to negatively gravitactic species *Heterosigma akashiwo* confined within a millifluidic chamber. (e) Analyzed image showing regions of high (red-brown) and low (blue-black) collective motion within a bioconvective plume. (f) The plumes generate active motion of the surrounding liquid driving enhanced molecular transport. The simulation snapshot shown here is based on experimental data, using computational fluid dynamics. The inset shows an initial patch of high concentration of molecules. (g) Similar bioconvective patterns can be seen in diverse microorganisms, including phototactic bacteria *Chromatium okenii* (image adapted from Ref. [112]) which swim toward sunlight and accumulates near the surface of lakes, before the onset of bioconvection. Such emergent flows in natural ecosystems have been implicated in distributing critical nutrients required by the microbes in their natural habitats.

where η is the fluid viscosity. The timescale for bioconvection is much shorter than that for cell growth. Therefore, a cell conservation equation of the form can be obtained for n :

$$\frac{\partial n}{\partial t} = -\nabla \cdot [n(\mathbf{u} + \mathbf{v}) - D \cdot \nabla n], \quad (35.38)$$

where the flux terms represent advection by the flow \mathbf{u} , drift relative to the flow with mean swimming velocity $\mathbf{v}(\mathbf{x},t)$, and the swimming diffusion tensor $D(\mathbf{x},t)$ (i.e., a 3×3 tensor, computed using generalized Taylor dispersion theory in terms of spatial moments). Descriptions for individuals are either phenomenological or derived from external torques coupled to low-Reynolds-number fluid mechanics. Appropriate forms of $\mathbf{v}(\mathbf{x},t)$ and $D(\mathbf{x},t)$ for each taxon are described in detail in the Refs. [114, 115, 116]. Initial conditions generally specify a fully mixed suspension or a steady, horizontally uniform, vertical profile. For a detailed treatment, the readers can refer to Ref. [40].

There is considerable potential impact in industry for better understanding the behavior of suspensions of swimming microorganisms and the structures they generate [39]. Applications include extensive and intensive schemes for the production of biodiesel (lipids from stressed algae), synthesis of other products in photobioreactors, such as beta carotene, and gray water treatment. In these systems a significant issue is minimizing energy input [a third of costs is cell dewatering, potentially facilitated by cell focusing; e.g., see Refs. [75, 39]]. Interestingly, behavior is a function of stress, which features strongly in algal hydrogen production [117]. We need to move away from treating bacterial and algal bioreactors as chemical systems and embrace and utilize the biologically motivated biased swimming behavior of the microorganisms.

Sommer *et al.* (2017)[111] provided compelling experimental evidence for bioconvection in a natural water body (the alpine Lake Cadagno in Switzerland) induced by the 10-μm-long, motile, purple bacterium *Chromatium okenii*. These bacteria gain energy from light-driven sulphide oxidation. High cell concentrations were observed in 30 to 120-cm-thick mixed layers of uniform temperature and salinity 12 m below the surface, in deep regions that are otherwise meromictic (monotonically increasing density with depth). The specific gravity of the bacteria were found to be 15% to 27% higher than water, and having swimming speeds of $27 \mu\text{m s}^{-1}$. Simulations of upward-swimming cells exclusively below an imposed oxycline (a sharp gradient in oxygen concentration) led to cell accumulation and instability. The authors found that bioconvection with *C. okenii* was sufficient to explain the observed mixed layer. They indicated that the mean theoretical dissipation rate in the layer was $1.0 \pm 1.5 \times 10^{-10} \text{ W kg}^{-1}$, or 45% of the bacterial energy input rate, matching in situ measurements. The authors also suggested that bioconvection is widespread in regions of lakes where turbulence is relatively weak.

35.5 Outlook

Planktonic active matter represents a rich confluence of physics, biology, and ecology. As described in this Chapter, motility in phytoplankton is strongly regulated by their cell morphology and internal organelles, whose distributions and dynamic adjustments create emergent behaviors such as gravitaxis, gyrotaxis, and bioconvection. These behaviors, in turn, can shape large-scale ecological processes including nutrient cycling, vertical migration patterns, and bloom formation.

Nonetheless, many fundamental questions remain only partially resolved. For instance, we have seen that phytoplankton often rely on morphological and intracellular reconfigurations (e.g., nucleus positioning, lipid storage, and membrane tension) to sense and respond to external cues such as gravity, chemical gradients, and turbulence. Yet, the precise mechanisms by which cells integrate these signals — especially fluid flow or shear cues — are still poorly understood. Recent work hints at *reactive oxygen species* (ROS) functioning as stress biomarkers and providing early warnings (Fig. 35.11), but the spatiotemporal dynamics of ROS signaling and its interplay with mechanosensitive ion channels or cytoskeletal elements remain active areas for exploration.

A deeper understanding of these phenomena demands combined efforts on several fronts. Experimentally, high-speed and high-resolution 3D microscopy under controlled shear or turbulence conditions can capture single-cell responses. Microfluidic platforms, meanwhile, enable fine-grained manipulation of flows, nutrient fields, and light gradients to investigate emergent collective phenomena such as clustering, patchiness, and thin-layer formation. Theoretically, incorporating realistic cell morphologies into continuum or discrete models can shed light on how morphological plasticity and heterogeneous intracellular density distribute the competing torques that govern reorientation and swimming stability.

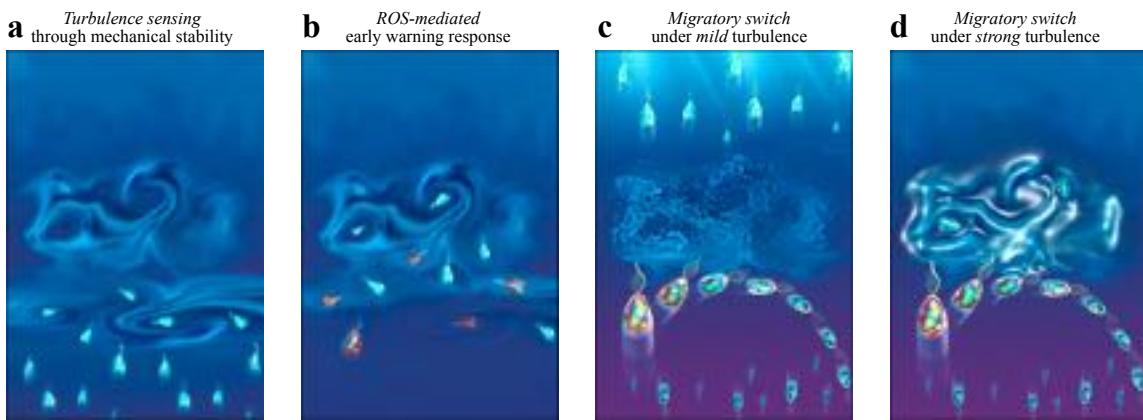


Figure 35.11: Gravitaxis-mediated early warning mechanisms in phytoplankton species. A ROS-mediated early-warning system enables microplankton populations to evade potential biomechanical damages (for instance shearing of flagella or even death) due to turbulence. (a) Turbulent patches can disrupt cells' migratory patterns, specifically when a population migrates vertically against the gravity direction. Cells can sense the intensity of a microscale turbulent eddy through their mechanical stability. Less stable cells are more easily thrown off balance and tumble under the effect of turbulence. (b) Cells achieve the sensing attribute by integrating the ROS signals over multiple tumbling events, within timescales of tens of seconds. (c) A negatively gravitactic population splits into two sub-populations: one that continues to migrate upward and one that switches to downward migration to avoid the turbulence. Cells, shown in red, with higher levels of ROS switch swimming direction from up to down, while others continue swimming upward. (d) Turbulence avoidance by adaptive swimming minimize biomechanical damages to the downward-swimming cells, although they show a reduced growth because of elevated ROS levels. However once the turbulence ceases, the down-swimming population scavenges excessive ROS and thereby recovers fitness.

From an ecological standpoint, quantifying how different phytoplankton species respond to—and even exploit—transient turbulence or nutrient pulses is critical for forecasting changes in community composition under evolving climate scenarios. For example, a better grasp of how gravitaxis couples with intermittent shear may improve models of harmful bloom occurrence and persistence. Including trait-based approaches that link measurable cell properties (size, aspect ratio, intracellular organelles) to fitness-related outcomes (growth rate, nutrient uptake, or survival) will be pivotal to merging active matter physics with classical population and ecosystem models.

Ultimately, a predictive theory of planktonic active matter could inform macroscale processes such as oceanic carbon sequestration, nutrient distribution, and biogeochemical cycling. Ongoing interdisciplinary collaborations—across physics, engineering, oceanography, and microbiology—are poised to unravel how these microscale behaviors scale up to shape global phenomena. For further reading on emerging experimental and theoretical advances, see, for example, Refs. [49, 31, 40] and references therein. The cross-talk between active matter approaches, new imaging technologies, and ecological modeling holds promise for a far deeper understanding of how life organizes in our oceans and freshwater habitats.

35.6 Problems

Problem 35.1: Orientational stability of synthetic microplankton. Consider a spherical *synthetic microplankton* of diameter $2r$, fabricated using two distinct materials in equal proportions, as shown in Fig. 35.12, with the respective densities related as $\rho_1 = 2\rho_2$. The synthetic microplankton is immersed in a liquid of density ρ_L .

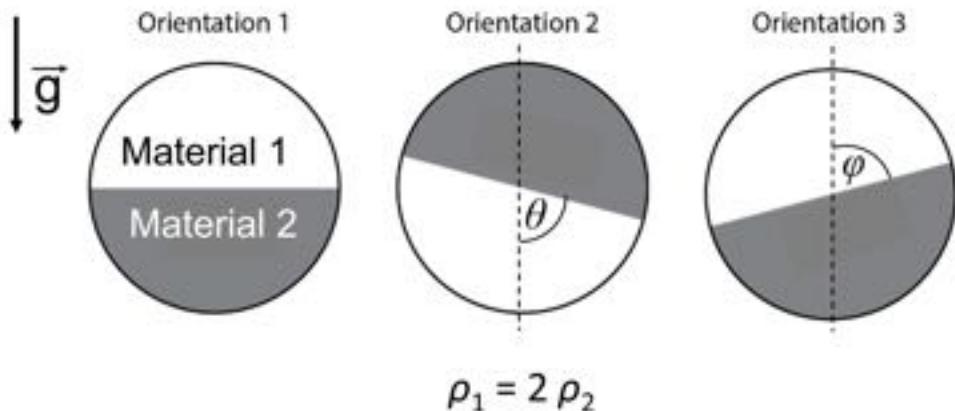


Figure 35.12: Orientational stability of microplankton.

- a. Calculate the gravitational torque acting on the *synthetic microplankton* in each of the three configurations depicted in Fig. 35.12.
- b. Consider that equal masses of the materials (1 and 2) are distributed concentrically within the sphere, such that Material 1 is confined within Material 2. Calculate the orientational torque acting on the sphere if it is displaced by an angle θ relative to the gravity vector.
- c. By what amount will the orientational torque change if the concentric distribution of the materials 1 and 2 is swapped, i.e., Material 2 is confined within Material 1? All other parameters are maintained as in b.
- d. Consider that the synthetic microplankton in a now propels with a speed v along a direction perpendicular to the 1–2 material interface. Draw the free-body diagram, showing all forces acting on the microplankton, and derive the torque equations if the plankton is momentarily titled by an angle θ as in case b.

Problem 35.2: Gravitaxis in micro- and hypergravity conditions. Consider the spherical *synthetic microplankton* in the above question is now placed onboard a parabolic flight simulating a series of micro- and hypergravity conditions. The microgravity ($0.05g$) and the hypergravity ($1.99g$) conditions last 30 seconds each, separated by a 60 seconds time interval of normal gravity (g) conditions.

- a.** Calculate the gravitational torques acting on the *synthetic microplankton* in each of the three configurations depicted in Fig. 35.12 during both micro- and hypergravity conditions.
- b.** Consider that the cell above, swimming at equilibrium orientation against gravity, is tilted by an angle θ . Calculate the time taken by the cell to gain back its equilibrium orientation under microgravity condition.
- c.** How would the time taken to reorient back to the equilibrium orientation change if the gravity conditions are changed to hypergravity?
- d.** Does the strength of the gravity force impact the swimming speed of the cells? Does it impact the sedimentation velocity of the cells?

Problem 35.3: Encounter rates of gravitactic plankton.

- a.** Calculate the encounter rate kernel of a spherical cell of radius r swimming with a speed v for (i) a cell of much larger size than a particle it encounters, and (ii) a cell of size comparable to the particle dimension.
- b.** What is the physical dimension of the encounter rate kernel?
- c.** Consider the situation described in **a**. Consider that the encounter between the cell and the particle now takes place in a liquid which has twice the viscosity. By what magnitude will the encounter rate kernel change?
- d.** Consider the situation described in **a**. However, instead of a swimming cell, the cell is now sinking with a speed v . (i) Calculate the encounter rate kernel. (ii) Does the encounter rate kernel depend on the viscosity and specific gravity of the liquid medium? (iii) How does the encounter rate kernel depend on the magnitude of the gravity force?
- e.** Encounter rate kernel for non-spherical cells: (i) Consider a non-motile cylindrical cell of radius r and height h . Show that the encounter rate kernel with respect to a small moving particle (speed v) is, $Q = 1/2\pi v(h+r)r$. (ii) Derive the encounter rate kernel if the cylindrical cell is motile along its axis with speed u ($\vec{u} \perp \vec{v}$). (iii) How will the encounter rate kernel change if the cell approaches the particle sidewise, moving normally to its axis ($\vec{u} \parallel \vec{v}$)?

Bibliography

- [1] J. O. Kessler. Individual and collective fluid dynamics of swimming cells. *Journal of Fluid Mechanics*, 173:191, 1986.
- [2] D.A. Siegel. Resource competition in a discrete environment: Why are plankton distributions paradoxical? *Limnology and Oceanography*, 43:1133, 1998.
- [3] J. Elgeti, R. G. Winkler, and G. Gompper. Physics of microswimmers—single particle motion and collective behavior: a review. reports on progress in physics. *Reports on Progress in Physics*, 78:056601, 2015.
- [4] R. Margalef. Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanological Acta*, 1:493–509, 1978.
- [5] S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti, and F. M. Cornejo-Castillo. Structure and function of the global ocean microbiome. *Science*, 348:1261359, 2015.
- [6] C. De Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, and M. Carmichael. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348:1261605, 2015.
- [7] R. E. Blankenship. *Molecular mechanisms of photosynthesis*, volume Third Edition. John Wiley & Sons, 2021.
- [8] M. Braun, M. Böhmer, D-P. Häder, R. Hemmersbach, and K. Palme. *Gravitational biology I: Gravity sensing and graviorientation in microorganisms and plants*. Cham: Springer, 2018.
- [9] M. Wada, A. Miyazaki, and T. Fujii. On the mechanisms of diurnal vertical migration behavior of heterosigma akashiwo (raphidophyceae). *Plant Cell Physiol.*, 26:431–436, 1985.

- [10] J. M. Sullivan, E. Swift, P. L. Donaghay, and J. E. B. Rines. Small-scale turbulence affects the division rate and morphology of two red-tide dinoflagellates. *Harmful Algae*, 2:183–199, 2003.
- [11] D. Jin, J. Kotar, E. Silvester, K. C. Leptos, and O. A. Croze. Diurnal variations in the motility of populations of biflagellate microalgae. *Biophysics Journal*, 119:2055–2062, 2020.
- [12] V. B. Tergolina, E. Calzavarini, G. Mompean, and S. Berti. Effects of large-scale advection and small-scale turbulent diffusion on vertical phytoplankton dynamics. *Physical Review E*, 104:65106, 2021.
- [13] D. J. McGillicuddy Jr, L. A. Anderson, N. R. Bates, T. Bibby, Buesseler K. O., C. A. Carlson, C. S. Davis, C. Ewart, P. G. Falkowski, S. A. Goldthwait, and D. A. Hansell. Eddy/wind interactions stimulate extraordinary mid-ocean plankton blooms. *Science*, 316:1021–1026, 2007.
- [14] M. J. Behrenfeld and E. S. Boss. Resurrecting the ecological underpinnings of ocean plankton blooms. *Annual Review of Marine Science*, 6:167–194, 2014.
- [15] J. J. Pierella Karlusich, F. M. Ibarbalz, and C. Bowler. Phytoplankton in the tara ocean. *Annual Review of Marine Science*, 12:233–265, 2020.
- [16] A. Falciatore, d'Alcalà M.R., P. Croot, and C. Bowler. Perception of environmental signals by a marine diatom. *Science*, 288:2363–2366, 2000.
- [17] T. L. Richardson and J. J. Cullen. Changes in buoyancy and chemical composition during growth of a coastal marine diatom: ecological and biogeochemical consequences. *Marine Ecology Progress Series*, 128:77–90, 1995.
- [18] J. K. Moore and T. A. Villareal. Buoyancy and growth characteristics of three positively buoyant marine diatoms. *Marine Ecology Progress Series*, 132:203–213, 1996.
- [19] J. A. Raven and A. M. Waite. The evolution of silicification in diatoms: inescapable sinking and sinking as escape? *New Phytologist*, 162:45–61, 2004.
- [20] B. J. Gemmell, G. Oh, E. J. Buskey, and T. A. Villareal. Dynamic sinking behaviour in marine phytoplankton: rapid changes in buoyancy may aid in nutrient uptake. *Proceedings of the Royal Society B: Biological Sciences*, 283:20161126, 2016.
- [21] S. L. Hinder, G. C. Hays, M. Edwards, E. C. Roberts, A. W. Walne, and M. B. Gravenor. Changes in marine dinoflagellate and diatom abundance under climate change. *Nature Climate Change*, 2:271–275, 2012.
- [22] A. Sengupta, F. Carrara, and R. Stocker. Phytoplankton can actively diversify their migration strategy in response to turbulent cues. *Nature*, 543:555, 2017.
- [23] F. Carrara, A. Sengupta, L. Behrendt, A. Vardi, and R. Stocker. Bistability in oxidative stress response determines the migration behavior of phytoplankton in turbulence. *Proceedings of the National Academy of Sciences (PNAS)*, 118:e2005944118, 2021.
- [24] A. Sengupta. Microbial active matter: a topological framework. *Frontiers in Physics*, 8:184, 2020.
- [25] A. Sengupta, J. Dhar, F. Danza, A. Ghoshal, S. Müller, and N. Kakavand. Active reconfiguration of cytoplasmic lipid droplets governs migration of nutrient-limited phytoplankton. *Science Advances*, 8:ebn6005, 2022.
- [26] J. I. Allen and L. Polimene. Linking physiology to ecology: towards a new generation of plankton models. *Journal of Plankton Research*, 33:989–997, 2011.
- [27] A. E. Kemp and T. A. Villareal. The case of the diatoms and the muddled mandalas: Time to recognize diatom adaptations to stratified waters. *Progress in Oceanography*, 167:138–149, 2018.
- [28] T. J. Smayda. Adaptations and selection of harmful and other dinoflagellate species in upwelling systems. 2. motility and migratory behaviour. *Progress in Oceanography*, 85:71, 2010.

- [29] M. J. Zirbel, F. Veron, and M. I. Latz. The reversible effect of flow on the morphology of ceratocorys horrida. *Journal of Phycology*, 58:46–58, 2000.
- [30] L. Karp-Boss, E. Boss, and P. A. Jumars. Motion of dinoflagellates in a simple shear flow. *Limnology and Oceanography*, 45:1594–1602, 2000.
- [31] T. Kiørboe. *A mechanistic approach to plankton ecology*. Princeton University Press, 2018.
- [32] P. A. Jumars, J. H. Trowbridge, E. Boss, and L. Karp-Boss. Turbulence-plankton interactions: a new cartoon. *Marine Ecology*, 30:133–150, 2009.
- [33] M. M. Musielak, L. Karp-Boss, P. A. Jumars, and L. J. Fauci. Nutrient transport and acquisition by diatom chains in a moving fluid. *Journal of Fluid Mechanics*, 638:401–421, 2009.
- [34] A. M. Young, L. Karp-Boss, P. A. Jumars, and E. N. Landis. Quantifying diatom aspirations: Mechanical properties of chain-forming species. *Limnology and Oceanography*, 57:1789–1801, 2012.
- [35] M. Schröda. Chamydomonas reinhardtii under the light microscope. <https://www.gesundheitsindustrie-bw.de/en/article/news/molecular-chaperons-in-algal-chloroplasts>. Accessed: 2022-12-17.
- [36] M. S. Valentine, A. Rajendran, J. Yano, S. D. Weeraratne, J. Beisson, J. Cohen, F. Koll, and J. Van Houten. Parameciumbbs genes are key to presence of channels in cilia. *Cilia*, 1:1–16, 2012.
- [37] D. P. Häder and R. Hemmersbach. Euglena, a gravitactic flagellate of multiple usages. *Life*, 12:1522, 2022.
- [38] S. S. Guasto, R. Rusconi, and R. Stocker. Fluid mechanics of planktonic microorganisms. *Annual Review of Fluid Mechanics*, 44:373, 2012.
- [39] M. A. Bees and O.A. Croze. Mathematics for streamlined biofuel production from unicellular algae. *Biofuels*, 5:53–65, 2014.
- [40] M.A Bees. Advances in bioconvection. *Annual Review of Fluid Mechanics*, 52:449, 2020.
- [41] M. Estrada and E. Berdalet. Phytoplankton in a turbulent world. *Scientia Marina*, 61:125–140, 1997.
- [42] G. Sutherland, B. Ward, and K. H. Christensen. Wave-turbulence scaling in the ocean mixed layer. *Ocean Science*, 9:597–608, 2013.
- [43] M. T. Barry, R. Rusconi, J. S. Guasto, and R. Stocker. Shear-induced orientational dynamics and spatial heterogeneity in suspensions of motile phytoplankton. *Journal of the Royal Society Interface*, 12:20150791, 2015.
- [44] W. M. Durham, J. O. Kessler, and R. Stocker. Disruption of vertical motility by shear triggers formation of thin phytoplankton layers. *Science*, 323:1067, 2009.
- [45] W. M. Durham, E. Climent, M. Barry, F. De Lillo, G. Boffetta, M. Cencini, and R. Stocker. Turbulence drives microscale patches of motile phytoplankton. *Nat. Commun.*, 4:2148, 2013.
- [46] F. De Lillo, M. Cencini, W. M. Durham, M. Barry, R. Stocker, E. Climent, and G. Boffetta. Turbulent fluid acceleration generates clusters of gyrotactic microorganisms. *Physical Review Letters*, 112:44502, 2014.
- [47] R.E. Breier, C.C. Lalescu, D. Waas, M. Wilczek, and M.G. Mazza. Emergence of phytoplankton patchiness at small scales in mild turbulence. *Proceedings of the National Academy of Sciences (PNAS)*, 115:12112–12117, 2018.
- [48] S. Lovecchio, E. Climent, R. Stocker, and W. M. Durham. Chain formation can enhance the vertical migration of phytoplankton through turbulence. *Sci. Adv.*, 5:7879, 2019.

- [49] A. M. Roberts. The mechanics of gravitaxis in chlamydomonas. *The Biological Bulletin*, 210:78–80, 2006.
- [50] P.L. Lakin-Thomas and S. Brody. Circadian rhythms in microorganisms: new complexities. *Annual review of microbiology*, 58:489, 2004.
- [51] R. Schuech and S. Menden-Deuer. Going ballistic in the plankton: anisotropic swimming behavior of marine protists. *Limnology and Oceanography*, 4:1–16, 2014.
- [52] H. Machemer and R. Bräucker. Gravireception and graviresponses in ciliates. *Acta protozoologica*, 31:185–214, 1992.
- [53] R. Hemmersbach and R. Bräucker. Gravity-related behaviour in ciliates and flagellates. *Advances in space biology and medicine*, 8:59–75, 2002.
- [54] H. Machemer, Machemer-Röhnisch, R. S., Bräucker, and K. Takahashi. Gravikinesis in Paramecium: theory and isolation of a physiological response to the natural gravity vector. *Journal of Comparative Physiology A*, 168:1–12, 1991.
- [55] U. Kowalewski. Responses of tetrahymena pyriformis to the natural gravity vector. *Microgravity Science and Technology*, 6:167–172, 1998.
- [56] G. Woodward, B. Ebenman, M. Emmerson, J. M. Montoya, J. M. Olesen, A. Valido, and P. H. Warren. Body size in ecological networks. *Trends in Ecology & Evolution*, 20:402–409, 2005.
- [57] S. Fraisse, M. Bormans, and Y. Lagadeuc. Turbulence effects on phytoplankton morphofunctional traits selection. *Limnology and Oceanography*, 60:872–884, 2015.
- [58] A. Cózar and F. Echevarría. Size structure of the planktonic community in microcosms with different levels of turbulence. *Scientia Marina*, 69:187–197, 2005.
- [59] J. Padisák, é. Soróczki-Pintér, and Z. Rezner. Sinking properties of some phytoplankton shapes and the relation of form resistance to morphological diversity of plankton—an experimental study. *Hydrobiologia*, 500:243–257, 2003.
- [60] R. Milo and R Phillips. *Cell Biology by the Numbers*. Garland Science, Taylor & Francis Group, 2016.
- [61] A. W. Visser and T. Kiørboe. Plankton motility patterns and encounter rates. *Oecologia*, 148:538–546, 2006.
- [62] A. Andersen and J. Dölger. Planktonic encounter rates with non-spherical encounter zones. *Journal of the Royal Society Interface*, 16:156, 2019.
- [63] J. A. Arguedas-Leiva, J. Słomka, C. C. Lalescu, R. Stocker, and M. Wilczek. Elongation enhances encounter rates between phytoplankton in turbulence. *Proceedings of the National Academy of Sciences (USA)*, 119:e2203191119, 2022.
- [64] S. Sundby. Turbulence and ichthyoplankton: influence on vertical distributions and encounter rates. *Scientia Marina*, 61:159–176, 1997.
- [65] J. Titelman and T. Kiørboe. Motility of copepod nauplii and implications for food encounter. *Marine Ecology Progress Series*, 247:123, 2003.
- [66] J. Dölger, L. T. Nielsen, T. Kiørboe, and A. Andersen. Swimming and feeding of mixotrophic biflagellates. *Scientific Reports*, 7:1–10, 2017.
- [67] J. Gerritsen and J. R. Strickler. Encounter probabilities and community structure in zooplankton: a mathematical model. *Journal of the Fisheries Research Board of Canada*, 34:73–82, 1977.
- [68] D. B. Dusenberry. *Living at Micro Scale*. Harvard University Press, Cambridge, Massachusetts, 2009.

- [69] B. J. Rothschild and T. R. Osborn. Small-scale turbulence and plankton contact rates. *Journal of Plankton Research*, 10:465–474, 1988.
- [70] D. M. Lewis and T. J. Pedley. Planktonic contact rates in homogeneous isotropic turbulence: theoretical predictions and kinematic simulations. *Journal of Theoretical Biology*, 205:377–408, 2000.
- [71] A. M. Roberts. Geotaxis in motile micro-organisms. *Journal of Experimental Biology*, 53:687–699, 1970.
- [72] B. Bean, G. (Eds.) Colombetti, and F. Lenci. *Microbial geotaxis*. In: *Membranes and Sensory Transduction*. Plenum Press, London, 1984.
- [73] T. Fenchel and B.J. Finlay. Geotaxis in the ciliated protozoon *loxodes*. *Journal of Experimental Biology*, 110:17–33, 1984.
- [74] J. O. Kessler. Hydrodynamic focusing of motile algal cells. *Nature*, 313:218, 1985.
- [75] J. O. Kessler. Co-operative and concentrative phenomena of swimming micro-organisms. *Contemporary Physics*, 26:147, 1985.
- [76] T. J. Pedley and J. O. Kessler. A new continuum model for suspensions of gyrotactic micro-organisms. *Journal of Fluid Mechanics*, 212:155, 1990.
- [77] T. J. Pedley and J. O. Kessler. Hydrodynamic phenomena in suspensions of swimming microorganisms. *Annual Review of Fluid Mechanics*, 24:313, 1992.
- [78] M. S. Jones, L. Le Baron, and T. J. Pedley. Biflagellate gyrotaxis in a shear flow. *Journal of Fluid Mechanics*, 281:137–158, 1994.
- [79] M. Braun, M. Böhmer, D-P. Häder, R. Hemmersbach, and K. Palme. *Gravity Sensing, Graviorientation and Microgravity (Chapter 1)*, volume 5Gravitational biology I: Gravity sensing and graviorientation in microorganisms and plants. Cham: Springer, 2018.
- [80] C. Limbach, J. Hauslage, C. Schäfer, and M. Braun. How to activate a plant gravireceptor: Early mechanisms of gravity sensing studied in characean rhizoids during parabolic flights. *Plant Physiology*, 139:1030–1040, 2005.
- [81] A.K. Strohm, K.L. Baldwin, and P. H. Masson. Molecular mechanisms of root gravity sensing and signal transduction. *Wiley Interdiscip. Rev. Dev. Biol.*, 1:276–285, 2012.
- [82] D.P. Häder and R. Hemmersbach. *Gravitaxis in Euglena*. In: Schwartzbach S, Shigeoka S (eds) *Euglena: biochemistry, cell and molecular biology*. Springer, Cham, 2017.
- [83] R. Hemmersbach, D.H. Volkmann, and D.P. Häder. Graviorientation in protists and plants. *Plant Physiology*, 154:1–15, 1999.
- [84] P. R. Richter, M. Schuster, H. Wagner, M. Lebert, and D. P. Häder. Physiological parameters of gravitaxis in the flagellate *euglena gracilis* obtained during a parabolic flight campaign. *Journal of Plant Physiology*, 159:181–190, 2002.
- [85] D. P. Häder and M. Lebert. Graviperception and gravitaxis in algae. *Advances in Space Research*, 27:861–870, 2001.
- [86] P. R. Richter, M. Schuster, M. Lebert, and D. P. Häder. Gravitactic signal transduction elements in *Astasia longa* investigated during parabolic flights. *Microgravity-Science and Technology*, 14:17–24, 2003.
- [87] D.P. Häder, R. Hemmersbach, and M. Lebert. *Gravity and the behavior of unicellular organisms*. Cambridge University Press, 2005.

- [88] A. Besson-Bard, C. Courtois, A. Gauthier, J. Dahan, G. Dobrowolska, S. Jeandroz, A. Pugin, and D. Wendehenne. Nitric oxide in plants: production and cross-talk with Ca^{2+} signaling. *Molecular Plant*, 1:218–228, 2008.
- [89] A. M. Roberts and F. M. Deacon. Gravitaxis in motile micro-organisms: the role of fore–aft body asymmetry. *Journal of Fluid Mechanics*, 452:405–423, 2002.
- [90] A. M. Roberts. The mechanics of gravitaxis in paramecium. *Journal of Experimental Biology*, 213:4158–4162, 2010.
- [91] W. M. Durham, E. Climent, and R. Stocker. Gyrotaxis in a steady vortical flow. *Physical Review Letters*, 106:238102, 2011.
- [92] J. Happel and H. Brenner. *Low Reynolds Number Hydrodynamics: With Special Application to Particulate Media*. In: *Mechanics of Fluids and Transport Processes*, volume 1. Englewood Cliffs, 1983, 1984.
- [93] S. H. Koenig. Brownian motion of an ellipsoid. A correction to Perrin’s results. *Biopolymers*, 14:2421–2423, 1975.
- [94] T. J. Pedley and J. O. Kessler. The orientation of spheroidal microorganisms swimming in a flow field. *Proceedings of the Royal Society of London. Series B, Biological Sciences.*, 231:47, 1987.
- [95] S. Childress, M. Levandowsky, and E. A. Spiegel. Pattern formation in a suspension of swimming microorganisms: equations and stability theory. *Journal of Fluid Mechanics*, 69:591–613, 1975.
- [96] T. J. Pedley, N. A. Hill, and J. O. Kessler. The growth of bioconvection patterns in a uniform suspension of gyrotactic micro-organisms. *Journal of Fluid Mechanics*, 195:223–237, 1988.
- [97] S. Lovecchio, F. Zonta, C. Marchioli, and A. Soldati. Thermal stratification hinders gyrotactic micro-organism rising in free-surface turbulence. *Physics of Fluids*, 29:053302, 2017.
- [98] G. B. Jeffery. The motion of ellipsoidal particles immersed in a viscous fluid. *Proceedings of the Royal Society of London Series A*, 102:161–179, 1922.
- [99] H. C. Berg. *Random walks in biology*. Princeton University Press, Princeton, NJ, 1993.
- [100] G. I. Taylor. LXXV. On the decay of vortices in a viscous fluid. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 46:671–674, 1923.
- [101] S. R. Dungan and H. Brenner. Sedimentation and dispersion of non-neutrally buoyant brownian particles in cellular circulatory flows simulating local fluid agitation. *Physical Review A*, 38:3601–3608, 1988.
- [102] M. Borgnino, G. Boffetta, F. De Lillo, and M. Cencini. Gyrotactic swimmers in turbulence: shape effects and role of the large-scale flow. *Journal of Fluid Mechanics*, 856:R1–11, 2018.
- [103] K. Gustavsson, F. Berglund, P. R. Jonsson, and B. Mehlig. Preferential sampling and small-scale clustering of gyrotactic microswimmers in turbulence. *Physical Review Letters*, 116:108104, 2016.
- [104] C. Zhan, G. Sardina, E. Lushi, and L. Brandt. Accumulation of motile elongated micro-organisms in turbulence. *Journal of Fluid Mechanics*, 739:22–36, 2014.
- [105] J. Qui, N. Mousavi, L. Zhao, and K. Gustavsson. Active gyrotactic stability of microswimmers using hydromechanical signals. *Physical Review Fluids*, 7:014311, 2022.
- [106] J. Qiu, N. Mousavi, K. Gustavsson, C. Xu, B. Mehlig, and L. Zhao. Navigation of micro-swimmers in steady flow: The importance of symmetries. *Journal of Fluid Mechanics*, 932:A10, 2022.
- [107] J. R. Platt. “bioconvection patterns” in cultures of free-swimming organisms. *Science*, 133:1766, 1961.

- [108] H. Wager. On the effect of gravity upon the movements and aggregation of euglena viridis, ehrb., and other micro-organisms. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 201:333–390, 1911.
- [109] J. Delcourt, N. W. F. Bode, and M. Denoël. Collective vortex behaviors: Diversity, proximate, and ultimate causes of circular animal group movements. *The Quarterly Review of Biology*, 91:1–24, 2016.
- [110] M. S. Plesset and H. Winet. Bioconvection patterns in swimming microorganism cultures as an example of Rayleigh–Taylor instability. *Nature*, 248:441, 1974.
- [111] T. Sommer, F. Danza, J. Berg, A. Sengupta, G. Constantinescu, T. Tokyay, H. Bürgmann, Y. Dressler, O. Sepúlveda Steiner, C.J. Schubert, and M. Tonolla. Bacteria-induced mixing in natural waters. *Geophys. Res. Lett.*, 44:9424–9432, 2017.
- [112] N. Pfennig. Chromatium okenii (thiorhodaceae) - bioconvection, aero- and phototactic behaviour. <https://av.tib.eu/media/12292>. Accessed: 2022-12-17.
- [113] T. J. Pedley. Instability of uniform micro-organism suspensions revisited. *Journal of Fluid Mechanics*, 647:335–359, 2010.
- [114] N. A. Hill and M. A. Bees. Taylor dispersion of gyrotactic swimming micro-organisms in a linear flow. *Physics of Fluids*, 14(8):2598–2605, 2002.
- [115] A. Manela and I. Frankel. Generalized Taylor dispersion in suspensions of gyrotactic swimming micro-organisms. *Journal of Fluid Mechanics*, 490:99–127, 2003.
- [116] Y. Hwang and T. J. Pedley. Bioconvection under uniform shear: linear stability analysis. *Journal of Fluid Mechanics*, 738:522–562, 2014.
- [117] C. R. Williams and M.A. Bees. Mechanistic modeling of sulfur-deprived photosynthesis and hydrogen production in suspensions of chlamydomonas reinhardtii. *Biotechnology and Bioengineering*, 111:320–335, 2014.

Chapter 36

Microfluidics

CAROLINE BECK ADIELS, REZA MAHDAVI

Microfluidics, the technology of handling and manipulating fluids on a micrometer scale, has emerged as a crucial interdisciplinary field since the 1980s, bridging material science, engineering, biology, physics, chemistry, and biotechnology to meet diverse application criteria and user demands. Its development is fueled by an increasing number of applications that depend on the technology's custom-made solutions. With the capacity to manipulate small volumes of fluids in channels smaller than a fraction of a millimeter, it offers precise control over microenvironments, an attribute contributing significantly to its increasing popularity in various disciplines. Such control allows for predictable fluid behavior in geometrically constrained sub-millimeter channels or compartments. Microfluidics finds broad applications ranging from the creation of micro total analysis systems (μ -TAS) and physiologically relevant systems like organs-on-chips (OOCs) to high throughput analysis tools for diagnostics in life sciences and environmental fields, demonstrating the technology's vast potential. This Chapter discusses the materials and fabrication processes used in microfluidics as well as its application to single-cell on-chip analysis (Fig. 36.1). This Chapter therefore provides an introduction to microfluidics, exemplifying the technology's wide-ranging applicability in active matter, its potential challenges, and the novel studies it enables.

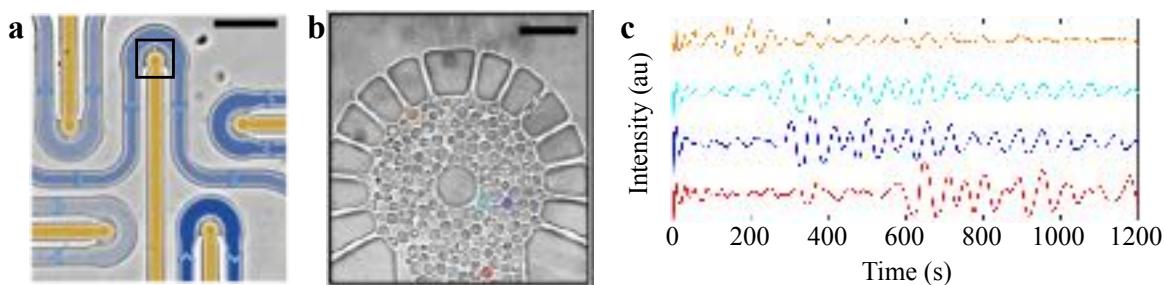


Figure 36.1: **Microfluidic chip for single-cell response data acquisition via fluorescence microscopy.** Yeast cells (*Saccharomyces cerevisiae*) from a single batch are introduced into the chambers (a) through the yellow-shaded inlet channels, forming a monolayer within each chamber, as shown in the zoomed-in view of one chamber (b). When fluid flow is initiated in the blue-shaded channels, the extracellular environment around the cells changes through diffusion via the narrow channels connecting the blue and yellow regions. This environmental change triggers oscillatory behavior in the cells' metabolic networks, which can be observed as intensity fluctuations in an autofluorescent metabolite (c). The figure illustrates the ability to capture longitudinal single-cell responses while modifying the surrounding extracellular environment and perturbing the system using microfluidics. The scale bar represents 200 μ m in (a) and 20 μ m in (b). Reproduced and adapted from Ref. [1].

36.1 Advantages and limitations of microfluidics

Microfluidics, like any experimental approach, presents a balance of advantages and limitations, unique to each application. On the plus side, it offers precise control over small fluid volumes, bringing a host of benefits over conventional sample handling. It allows for the careful handling of precious, sparse, costly, or even hazardous samples and reagents. The speed, efficiency, spatial and temporal control this technology provides outpace traditional methods, even allowing for sample perturbations without disturbing equilibrium — a challenge in standard methodology. The miniaturization of the technology permits the incorporation of multiple fluid inlets for parallel or sequential reactions (high-throughput) without taking up much space.

On the other hand, microfluidics is not without its challenges. One central issue is the introduction of fluids into the device, often leading to complications such as bubbles, leakage, or unstable flow profiles. The technology also grapples with handling the diverse surface properties of various materials used in microfluidic devices, complicating the interaction between the fluid and the material surface. Ensuring adequate concentrations of investigated substrates becomes a daunting task due to the reduced reagent and sample volume. Mixing events in confined micron-sized channels present another hurdle, as controlling flow behavior in such minuscule structures necessitates unique methods. Each of these issues intertwines with the others, primarily revolving around the interaction between the fluids, the surfaces they contact, and the fluid behavior within tight structural confines. These topics will receive a more detailed examination in the subsequent section.

36.2 Fluid Mechanics

Liquids' behavior on the microscale differs from what is experienced in everyday life as different forces dominate. Usually, a large set of parameters must be considered to understand and predict the mechanics and behavior of different fluids. However, the theoretical framework is simplified by assuming all solutions Newtonian, incompressible, and isotropic, i.e., the fluid viscosity is independent of the flow velocity, its density is independent of the pressure, and properties are equal in all directions. This section will give a short overview of fluidic behavior in microfluidics, but the book by Nguyen & Wereley in Ref. [2] is recommended for more details.

36.2.1 Navier-Stokes equations and Reynolds number

The flow behavior within microfluidic chips of typical dimensions is structured and laminar, and its unique traits are beneficial for numerous applications. More specifically, laminar behavior enables predicting how the flow will behave inside the microfluidic design. The fluid flow behavior can be described using the Navier-Stokes equation

$$\rho \left(\frac{\delta \mathbf{u}}{\delta t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) = -\nabla p + \eta \nabla^2 \mathbf{u} + \mathbf{f}, \quad (36.1)$$

where ρ is the fluid density, \mathbf{u} is the flow velocity vector, t is time, p is pressure, η is the fluid's dynamic viscosity, and \mathbf{f} is the body force density. The left-hand side of this equation describes the inertial acceleration of the system as the temporal ($\rho \frac{\delta \mathbf{u}}{\delta t}$) and spatial ($\rho \mathbf{u} \cdot \nabla \mathbf{u}$) variation of the velocity, respectively. The right-hand side describes how pressure, viscosity, and body force densities \mathbf{f} (such as gravity or centripetal forces) act upon the system [3]. Due to the small velocities in microfluidic settings, the dominant terms are the spatial inertial ($\rho \mathbf{u} \nabla \cdot \mathbf{u}$) and the viscous ($\eta \nabla^2 \mathbf{u}$) ones. The inertial term is proportional to $\frac{\rho U^2}{L}$ and the viscous term is proportional to $\frac{\eta U}{L^2}$, where U and L are the microfluidic chamber's typical velocity and length scales. The ratio between the two describes whether a fluid flow is in the laminar or turbulent regime and is termed the Reynolds number (Re):

$$Re = \frac{\frac{\rho U^2}{L}}{\frac{\eta U}{L^2}} = \rho \frac{UL}{\eta}. \quad (36.2)$$

When the viscous forces dominate the inertial forces, a low Reynolds number is attained (typically below 2300 [4]), resulting in laminar flow. For higher Reynolds numbers ($Re > 2300$), inertial forces dominate, and the flow is turbulent and unpredictable, Fig. 36.2. Typically, microfluidic channels exhibit a rectangular cross-section, and the typical length scale can therefore be approximated using the channel's hydraulic diameter. The



Figure 36.2: Flow profiles of a fluid through a channel. A schematic illustration of the laminar (left) and turbulent (right) flow profile of a fluid through a channel. In the latter scenario, the flow pattern is unpredictable in contrast to laminar flows.

hydraulic diameter D_H is defined as four times the open area (A) divided by the wetted perimeter (P),

$$D_H = \frac{4A}{P} = \frac{4ab}{2(a+b)} = \frac{2ab}{(a+b)}, \quad (36.3)$$

where a and b correspond to the channel's width and height, respectively [5].

Typically, in channel dimensions of ten to hundreds of micrometers and flow rates of a few hundred nanoliters per minute, the resulting Reynolds number is below 1. For such low Reynolds numbers, the left-hand side of the Navier-Stokes equation can be neglected. Compared to the viscous forces, usually, also body forces are negligible, and thus, the Navier-Stokes equation can be reduced to the Stokes equation

$$\nabla p = \eta \nabla^2 \mathbf{u}, \quad (36.4)$$

which describes strictly laminar flows. Stokes flow, also named creeping flow or creeping motion, is a type of fluid flow where advective inertial forces are small compared to viscous forces. Compared to the Navier-Stokes equation, this equation lacks time dependency, meaning that the flow behavior described by the Stokes equation is entirely reversible. In contrast to turbulent flows, where continuous mixing occurs due to unstructured streamlines, particle transport between the parallel sheets of fluid in laminar flows relies on diffusion.

36.2.2 Concentration distribution

In the presence of flow, the concentration distribution of solute molecules in solution in a microfluidic channel depends on both advection and diffusion:

$$\frac{\partial c}{\partial t} = D \nabla^2 c - \mathbf{u} \cdot \nabla c, \quad (36.5)$$

where c is the molecule concentration, t is time, D is the diffusion constant, and \mathbf{u} is the velocity of the flow field. In turn, a molecule's diffusion constant (D) depends on its (approximated) radius (r), the viscosity of the surrounding liquid η , and the absolute temperature T ,

$$D = \frac{k_B T}{6\pi\eta r}, \quad (36.6)$$

where k_B is the Boltzmann constant.

In order to distinguish between convection and diffusion as the driving forces for molecule distributions, the dimensionless Péclet number (Pe) can be utilized

$$Pe = \frac{UL}{D}, \quad (36.7)$$

which relates the diffusion constant to the molecule diffusive behavior. As for the Reynolds number, the Péclet number depends on the typical length scale L . In typical microfluidic dimensions, molecule transport is predominantly governed by diffusion due to the decrease in the Péclet number with decreasing length scale.

In the absence of convection, Eq. (36.5) can be simplified to what is referred to as Fick's second law, describing the mass diffusion flux over time:

$$\frac{\partial c}{\partial t} = -D \nabla^2 c. \quad (36.8)$$

However, in a steady state scenario, the direction of molecule movement depends on concentration gradients, and the mass flux of molecules from a higher to a lower concentration (c) can be described by Fick's first law:

$$J = -D\nabla c, \quad (36.9)$$

which relates diffusion (D) of a given solute to the mass flux (J) [3]. The negative sign preceding the equation indicates that the mass flux is positive only when the mass transfer occurs in the direction opposite to the concentration gradient.

36.2.3 Simulations using COMSOL Multiphysics

To prototype the initial model and help the experimental procedures, a computer model can be extremely helpful. In microfluidic devices, simulating the velocity fields and the concentration distribution inside a chamber are immensely important. This task can be accomplished through the use of computational fluid dynamics (CFD) packages. CFD relies on principles derived from fluid mechanics and employs numerical methods and algorithms to solve fluid flow-related problems. Among the available options for CFD analysis, COMSOL Multiphysics stands out as a prime candidate [6]. Although other similar software counterparts like ANSYS Fluent exist, COMSOL has demonstrated versatility and user-friendliness, making it a preferred choice for CFD modeling of microfluidic devices.

COMSOL Multiphysics incorporates a collection of physics interfaces tailored to simulate various phenomena, offering a range of tools to address the specific challenges encountered when modeling micro- and nanoscale flows. The software's module encompasses physics interfaces capable of handling diverse flow regimes, including laminar and turbulent flow, single and multiphase flow, as well as flow in porous media. Moreover, it includes capabilities to treat chemical reactions between dilute species.

Even though COMSOL Multiphysics has a range of physics interfaces to use for fluid flow in a variety of circumstances, typically, the Laminar Flow Interface is a good starting point for many microfluidics problems. This interface obtains the flow field by solving Naiver-Stokes equation and the continuity equation at the same time. It assumes the model boundary and temperature constant within the model geometry. The Laminar Flow Interface best describes problems where Re is less than 1000. In cases where Re is less than 1, the creeping flow regime can be utilized instead. The Creeping Flow Interface simplifies the solution process by neglecting the inertial term in the equations.

By incorporating the Microfluidics module, the flow interface can be coupled with a transport interface to model the movement of chemical species and ions. One commonly used transport interface is the Transport of Diluted Species, which accurately accounts for solutions with diluents constituting approximately 90% or more of the solution in terms of molar concentration. By default, this interface considers species diffusion based on Fick's law and solves the convection-diffusion equation within the fluid bulk. In reality, physical parameters seldom align perfectly with the simplest forms of equations, particularly in biological systems. For example, blood flow in such systems deviates significantly from Newtonian behavior [7], and diffusion constants of solutes can be altered in culture media. Although COMSOL offers various options to accommodate these variations, such as non-Newtonian fluid models, slip boundary conditions, and mass transfer in porous media, acquiring the necessary parameters may not always be feasible. Nevertheless, many studies often utilize the model in its simplest form by, e.g., considering the density independent of the temperature, or substituting water for culture media, as reasonably predictive solutions can be obtained that approximates the real scenario. For further insights into fluid mechanics theory and instructions on how to utilize the software, please refer to the "Fluid Flow, Heat Transfer, and Mass Transport" section on the COMSOL Multiphysics webpage, available at <https://www.comsol.com/multiphysics/fluid-flow-heat-transfer-and-mass-transport>.

36.3 Materials and fabrication methods

Today, numerous microfluidic fabrication methods are available, and the development and evaluation of suitable materials is a constantly ongoing focus area [8]. Therefore, no generic protocols are available, and the methodology guidelines will differ between applications. The different fabrication methods have different advantages and disadvantages and depend on demands of surface roughness, channel ratio aspects (width and length), cost, and fabrication facility or instrument availability. Some standard methodologies are wet or dry

etching procedures, laser ablation, embossing, casting, optical lithography, and, more recently, 3D printing. The latter has become more attractive with the improvement in printing resolution, which has been the major limitation of this methodology for microfluidics. However, the use of stereolithography-based techniques, where a liquid resin is selectively photopolymerized with high resolution (up to $0.6\text{ }\mu\text{m}$ [9]), now places this approach in the same league as, e.g., soft-lithography methodologies.

Regarding the materials suitable for microfluidic devices, there are a few properties that have to be considered. Optimal materials exhibit chemical and thermal stability. The specific range will depend on the application, but, in general, one has to make sure that the studied phenomena are not affected by the material per se. Suppose, for example, that the microfluidic channels are to be used for a biological application. For this purpose, the material needs to be biocompatible, or, at the very least, not leak harmful chemicals that might be toxic to the cells. Since microfluidics often is used in combination with optical or spectroscopic setups, optical transparency for the intended wavelengths is also critical. One of the most commonly used microfluidic materials for bioapplications is the silicon elastomer poly(dimethylsiloxane) (PDMS). However, alternative options are available, both inorganic, such as glass, silicon, and ceramic, and organic, i.e., polymeric and paper-based (see, for excellent summaries, Refs. [8] and [10]). The best material for a desired microfluidic design might not be compatible with the available instruments and setups or will have a too difficult or costly fabrication. Hence, reiterations might be necessary to find the optimal material and fabrication method. The following subsections will focus on PDMS as example material and one of its fabrication processes is described in detail.

36.3.1 PDMS fabrication procedure

This section explains the device fabrication using optical lithography and PDMS molding. The process is outlined in Fig. 36.3 and further detailed protocols are available in Refs. [11] and [12].

The first step of the PDMS microfluidic fabrication process is to create a silicon master mold of ridges corresponding to the microfluidic channel design. The design is either made in house using, e.g., CAD-aided software or purchased from a company that provides off-the-shelf or custom-made masks. For pillars, it is rec-

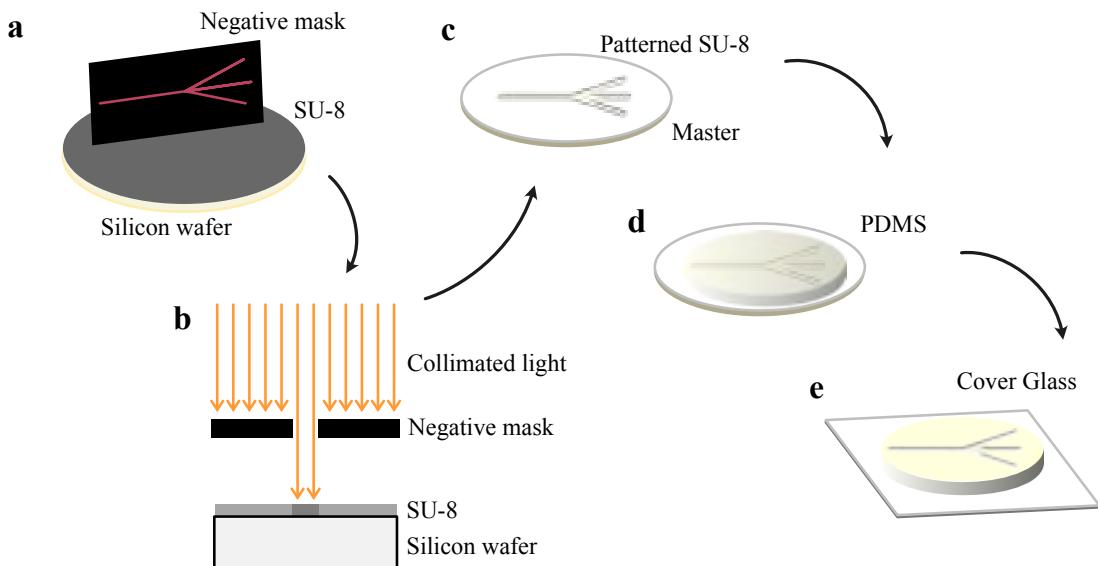


Figure 36.3: Fabrication of PDMS microfluidics using optical lithography. (a) A silicon wafer is spin-coated with (in this case) a negative photoresist (SU-8) to desired thickness and placed in a mask aligner with the negative mask corresponding to the microfluidic channel design on top. (b) The sandwich is exposed to collimated UV light, illuminating the SU-8 only where the mask is transparent. (c) After baking and developing steps, the master is finalized and can be used as a mold. (d) PDMS is poured on top of the master and allowed to cure thermally before being carefully detached. (e) After access holes are punched and cleaned, the PDMS slab and cover glass are plasma treated and allowed to bond for covalent sealing.

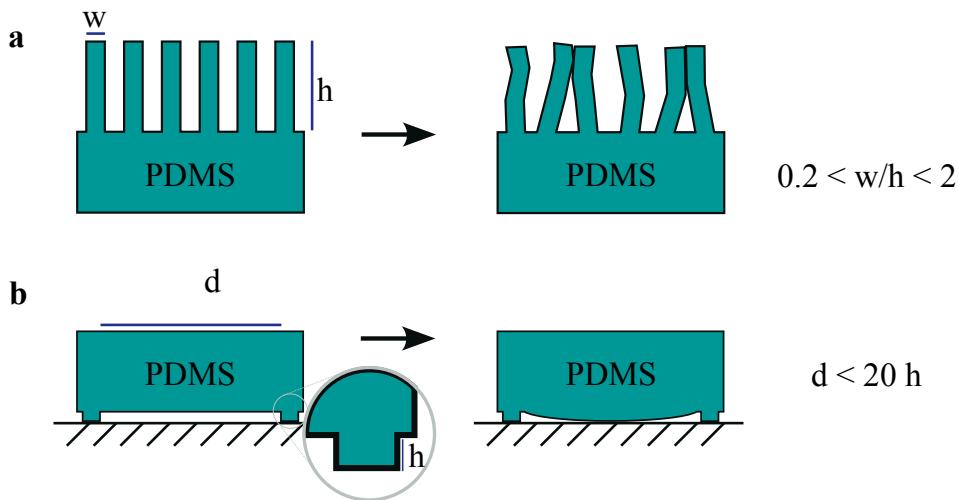


Figure 36.4: An illustration displaying the simple rules to follow for stable PDMS microfluidic structures. (a) Optimal width (w) vs height (h) ratio and (b) optimal distance (d) between walls. If the channels are too deep or shallow with respect to their height, there is a risk of pairing or sagging, respectively.

ommended to keep the width-to-height ($w:h$) ratio between 0.2 and 2 to avoid pairing (Fig. 36.4a). Conversely, to avoid sagging of a channel, the distance (d) between two walls should not exceed the channel height (h) times a factor of 20 (Fig. 36.4b). Following these simple rules can not guarantee intact structures, but the risk of ending up with unintentional channel properties is at least minimized. However, the overall channel stability depends also on other parameters, such as the inherent stability properties of the chosen material.

The mask, in turn, can be a soft, high-resolution e-beam lithography printed version or a solid chromium mask that is used in combination with a positive or negative photoresist substrate. The following steps are exemplified using a negative photoresist substrate (SU-8 series). The photoresist is spin-coated on a silicon wafer. The SU-8 substrate is available in several different viscosities, and adjusting the spin speed and duration will generate coating thicknesses between 2 and 200 μm . The wafer is pre-baked at 65°C and 95°C to densify the coating and evaporate solvents. The design pattern is achieved by first illuminating the wafer and mask with collimated light of 350 to 400 nm using a mask aligner (Fig. 36.3). The photoresist will eventually solidify where the mask is transparent (corresponding to the microfluidic channels). After illumination, a post-baking step, again at 65°C and 95°C, follows before the wafer is submerged in a developer solution and developed under shaking. The baking, illumination, and development durations depend on the coating thickness and recommendations by the manufacturer. Once the master is fully developed, it is rinsed with isopropanol, dried under nitrogen gas, and ready for replica molding.

PDMS is an elastomer consisting of repetitive units of the ene [-Si(CH₃)₂O-]. One of the most common variants, Sylgard 184 [13], consists of a base and a cross-linker. After curing, the resulting flexible material has several beneficial characteristics, such as gas permeability, biocompatibility, and optical transparency. However, the material has a high affinity to solvents, which makes it unsuitable for dry chemistry, such as organic synthesis. An essential trait of PDMS regards its inherent hydrophobicity, a feature discussed in the next paragraph. The elastic polymer PDMS base is mixed in a 10:1 weight ratio with its cross-linker, the curing agent, and subsequently degassed in a desiccator. The mixture is poured onto the master, either in a thin layer (< 0.5 cm) through which inlet and outlet holes can be punched later on, or in a slightly thicker layer (≥ 0.5 cm). In the latter case, metallic posts can be held in place during the molding process to create access holes in the device from the start. If the PDMS layer is very thin, adaptors might be necessary to attach the tubing for fluid introduction (see the following subsection). The PDMS is cured at 90°C degrees for 1 hour, after which it is carefully lifted from the master. The master can be reused for a new replica molding cycle, and the PDMS structure is punched and/or the access holes are cleaned from debris and placed upside down on a cover glass. Exposing the PDMS and a cover glass of choice to oxygen plasma for 30 seconds will, when put together, covalently bond the two surfaces, sealing off the microfluidic channels. The bonds constitute

silanol bridges, Si-O-Si, and form between plasma-treated PDMS, glass, and similar substrates [14]. A final additional one-hour thermal incubation at 90°C may further improve the mechanical stability of the device. The final PDMS device can permeate gases, but it is challenging to predict to which extent accurately. Thin PDMS slabs exhibit exchange between the microfluidic compartment and ambient air [15] driven by gas partial pressure gradients [16]. However, the permeability depends on the curing agent ratio, the device's thickness, storage, and surface treatment [17, 18]. Further treatments or alterations of the final device may be necessary to optimize the environment for the intended sample. It is worth mentioning that PDMS displays unfortunate adsorption and absorption of small hydrophobic molecules [19, 20]. This can be counteracted by inducing durable hydrophilicity via electron beam or gamma irradiation [21].

36.3.2 Handling of fluids

Surface tension and wetting are inherent properties of any given material and are highly relevant in microfluidic work. Wetting is a liquid's ability to keep contact with a solid surface and is a balance between adhesive and cohesive forces. These, in turn, depend on the surface tensions between the solid, the liquid, and the surrounding gas [22]. Their connection is described through Young's equation,

$$\gamma_{LG} \cos \phi_c = \gamma_{SG} - \gamma_{SL}. \quad (36.10)$$

This relation is also displayed in Fig. 36.5. The subscript indices stand for liquid (L), gas (G), and solid (S), and ϕ_c is the droplet's contact angle on the solid surface. As the contact angle increases from low to high values, a liquid can completely wet, partially wet, or not wet a surface. The droplet shape is adjusted to keep the surface free energy to a minimum [22], meaning that, if $\gamma_{SL} < \gamma_{SG}$, it is energetically favorable for a droplet to wet the surface. A hydrophobic surface, such as PDMS, in contact with an aqueous liquid typically displays a contact angle > 90 degrees and shuns the liquid. However, the PDMS surface tension can be altered by modifying the surface roughness via addition of micro- and nanostructures or different surface coatings. In the case of PDMS, the finalized device can be exposed to air- or oxygen plasma [23], or its inner lining treated with an amphiphilic surfactant [24]. Such treatments will benefit the introduction of liquid into the microfluidic device as the solid surface tension decreases with respect to water (γ_{SL}), rendering the surface more hydrophilic. Moreover, a plasma treatment per se results in nanoscale topological alterations, creating a surface roughness that further increases its hydrophilicity [23].

How the fluid is introduced into the microfluidic device can vary. Most commonly, the liquid is pumped into the microfluidic device using syringe-based, pressure-driven, or peristaltic pumps. These pumps work in a broad flow rate regime and can provide pulse-free and consistent perfusion. Relying on gravity-driven perfusion can be an easy solution for smaller volumes, e.g., using a pipette tip filled with the media of interest. However, it will generate an inconsistent and less stable flow rate. When using pumps, the liquids are usually introduced into the microfluidic device via tubing fused directly into inlet holes, or, if necessary, via some adaptors. Commercially available adaptors are glued onto the device, or custom-made punched PDMS slabs can be plasma

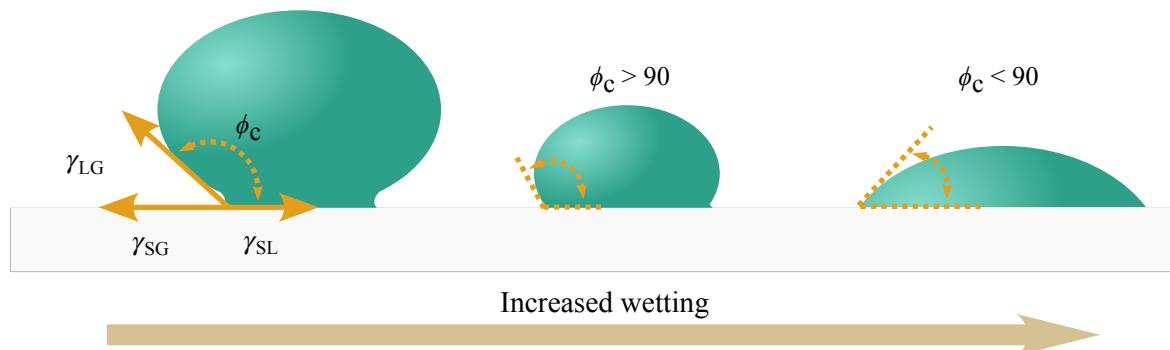


Figure 36.5: **Contact angle and wetting.** The contact angle (ϕ_c) is defined by the surface tensions of three interfaces: γ_{SL} , γ_{SG} , and γ_{LG} , where the indices S, L, and G represent the solid, liquid, and gas phases (leftmost droplet). The higher the wettability, the smaller the contact angle (middle and rightmost droplet).

fused and aligned on top of the inlets to secure the tubing. Commonly, the former is preferred for high flow rates when high resistance is otherwise forcing the tubing out of its placement. As previously mentioned, lateral mixing between different fluids in the laminar regime is diffusion-dominated. When required, a faster mixing can be achieved incorporating active and passive mixers into the microfluidic system (for a recent overview of micromixers, see Ref. [25]). There are also ways of avoiding mixing until a cue is given by activating valves, which can be passive (typically relying on hydrophobic patches) or active. The most renowned active valves are the "Quake" [26] and "Doormat" microvalves [27] and their derivatives. These valves rely on multilayered PDMS channels bonded to a rigid substrate, in which one holds the fluid and the adjacent one constitutes the actuation channel. The Quake microvalve is open until the actuation channel is pressurized, forcing this channel to bulge closing it. The Doormat microvalve instead works the other way around. The flow in the fluid channel is closed in the absence of a vacuum, and only when the actuation channel is underpressurized does the fluid channel open. With today's available miniaturization technology, the incorporation of both sensors and a variety of controls ranging from electric, thermal, and magnetic enables control and manipulation of flow, temperature, and gas distribution. Such progress has made microfluidics a cornerstone for sophisticated, parallelized applications avoiding bulky and expensive instruments by enabling automatized platforms with on-chip analysis and detection possibilities.

36.3.3 Detection

As miniaturization progresses, standardized analysis approaches are exchanged for microfluidic counterparts, and novel technologies are incorporated on-chip. Sensors and detectors that previously required large samples are continuously improved, demanding smaller sample volumes that, in turn, aid their compatibility with microfluidic utilization and parallelization. Today, many analysis approaches are incorporated on-chip [8, 28, 29], but sometimes off-chip methodologies are employed. For instance, mass spectrometry technology requires an off-chip but in-line approach where a microfluidic device handles the cell isolation step and processing steps on-chip, preparing the sample for mass spectroscopy analysis that is run separately [30]. It is also possible to incorporate directly into the microfluidic device biosensors that detect different biological elements. A biosensor is an analytical tool consisting of one part detecting the biological element and a second part that physicochemically transduce the signal into an optical or electrical readout. Many biosensors integrate well with the traditional materials used in microfluidics, but novel materials are on their way with improved signal transduction, such as graphene-based or metal nanomaterials [8].

There is a large variety of means by which a (biological) sample can be monitored on-chip. Since many microfluidic materials are optically transparent (the wavelength interval differs between materials), various optical monitoring technologies are used, including light and fluorescence microscopy, phase or scattering microscopy, as well as UV-Vis, FT-IR illumination, and Raman spectroscopies, to mention a few. However, electrochemical and conductivity measurements are also enabled by incorporating electrical circuits directly into the microfluidic materials (like several other types of sensors) [8, 31]. With these detection possibilities, the research field of single-cell analysis has opened up. Instead of working with millions of cells in bulk, single cells can be analyzed based on their, e.g., behavior, morphology, DNA, or mRNA content. For the latter, the whole pipeline of isolation, extraction, amplification, separation, and detection of the sample can nowadays be executed in a miniaturized manner on-chip [32] rather than by a technician using traditional molecular biology methodologies.

On-chip detection also offers the advantage of enabling long-term imaging, allowing for the continuous monitoring of the same sample over an extended period. This capability enables the observation of cells throughout multiple cell cycles within a controlled environment, such as a bioreactor, while maintaining static, dynamic, or even cyclic conditions [33]. Moreover, with the advancements in AI-aided image analysis, real-time detection and tracking of individual cells are expected to enhance the utility and data extraction from microfluidic-based experiments. In conclusion, microfluidic devices can incorporate various microscale components tailored to specific system requirements, sample types, and available infrastructure. The subsequent section will explore some applications of microfluidics.

36.4 Applications

The controlled manipulation of small sample and reagent volumes, along with the incorporation of custom-made patterns, valves, sensors, and actuators in microfluidic devices, has greatly benefited various disciplines within the field of natural sciences [10, 34]. In the following section, we will explore some notable applications of microfluidics applied to cell biology and active matter. However, for a more comprehensive understanding, interested readers are encouraged to refer to recent reviews in the field, ranging from those focusing on microfluidic-based manipulation and sorting [35], microfluidics for studying soft condensed matter at multiple scales [36], and microfluidic paper-based electroanalytical devices [37].

36.4.1 Single-cell studies

Microfluidics is a powerful tool in the field of biology. One area of particular focus and significance is the ability to investigate individual cells. Microfluidics can be used to manipulate and sort a sample mixture to enrich it with cells with similar characteristics. One method, known as deterministic lateral displacement (DLD) [38] employs microfluidic channels containing custom-made pillars with optimized dimensions and arrangement patterns specific to the sample of interest [38]. These pillars effectively separate cells of varying sizes on the micrometer scale, relying on the parameters of the pillar array structure. However, this sorting technique necessitates downstream cell collection as each cell is encountered only once and subsequently lost for further monitoring.

Another approach, that allows for long-term sample monitoring, combines microfluidics with acoustics [39]. In this method, gentle pressure waves generated inside a channel create an acoustic radiation force that acts upon the introduced sample, separating its content based on their compressibility, density, and size. An emerging field within acoustofluidics is sharp-edge acoustics that focuses on the incorporation of sharp edge structures into microchannel devices [40]. These structures, when subjected to an acoustic wave, initiate oscillations and give rise to highly controllable streamlines in their vicinity. These flows can be utilized for pumping, mixing, cell focusing, trapping, and more. One limitation of this technology is its reliance on external instruments like signal generators and amplifiers, which may not be readily accessible for non-experts.

Apart from sorting and manipulation capabilities, microfluidics provides the advantage of perturbing a sample to repeatedly expose it to specific environments. This enables the investigation of adaptation processes and desensitization of a system. Further, in contrast to large-scale studies, microfluidic devices offer the advantage of rapid addition and removal of desired solutions, eliminating the need for time-consuming washing steps such as centrifugation or sieving. As a result, when integrated with monitoring techniques, microfluidics enable exceptional temporal resolution in data acquisition.

To facilitate imaging possibilities of a sample of interest, the inner lining of the PDMS device can be coated to achieve two objectives simultaneously: rendering the surface hydrophilic and facilitating the anchoring of specific cells. Optical tweezers, in turn, can be utilized to non-invasively trap and attach cells to the coated surface [41]. This approach has been successfully employed in several studies focused on exposing yeast cells to specific extracellular environments, monitoring subcellular localization of fluorescent transcription factors at the single-cell level [42, 43, 44], investigating metabolic reaction networks in response to the sensed environment [45], and exploring the development of coordinated cell behaviors [1]. These experiments are described in a bit more detail below.

Yeast cells are considered a common biological sample and serve as a widely used model organism for studying mammalian cells. Despite being single-celled eukaryotes, they share fundamental reaction networks and evolutionarily conserved proteins with mammalian cells [46]. Unlike eukaryotic cells, yeast cells are protected by an outer cell wall in addition to their membrane, which shields them from external stresses. The cell wall primarily consists of polysaccharides and glycoproteins, which can be leveraged for surface attachment. In microfluidic devices functionalized with the lectin Concanavalin A, binding occurs when the yeast cell wall comes into contact with the inner lining of the microfluidic chambers [41, 44].

In the work by Hamngren Blomqvist *et al.* [44], optical tweezers were combined with a simple four-inlet microfluidic design. This design allowed yeast cells to be trapped while being flown through the device. By raising the microscope stage, the bottom glass surface of the microfluidic device makes contact with the yeast cell trapped in the optical tweezer, allowing the cells to be positioned one by one in a cell array at the intersection of the channels. Two of the inlets served as cell administration channels, introducing different cell

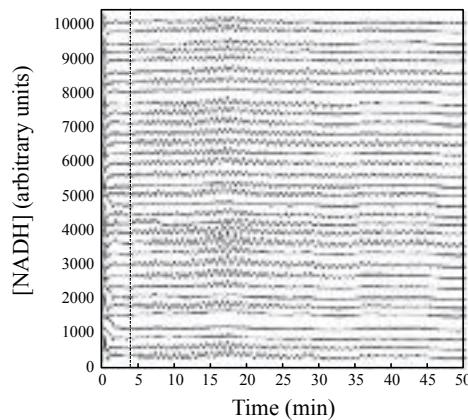


Figure 36.6: Time series for single-cell NADH fluorescence intensities in a microfluidic chamber. Each curve represents a single cell's NADH intensity fluctuations and is separated by 300 arbitrary units in intensity for better visualization. The dotted line indicates the timing of the flow rate infusion alteration. Thus, the environment around the cells is shifted from glucose exposure only to glucose and potassium cyanide, which triggers the glycolytic oscillations. Reproduced from Ref. [45].

strains expressing distinct fusion proteins into two distinct channels, A and B. Individual cells were sequentially trapped from either channel and positioned in predetermined locations on the surface. Once the cell array was established, the extracellular environment was modified by adjusting the flow rates in the different inlets. Simultaneously, fluorescence imaging was initiated to monitor the cellular response during exposure to the new environment. This technique enabled the study of different cell types' responses to the environment within a single experiment by applying the intended solution to the entire cell array.

An alternative scenario is studied in Bendrioua *et al.* [42] where the migration of a fluorescently tagged genetic repressor out of the nuclei upon glucose exposure was monitored. In this study, the same stationary yeast cell population is followed through perturbations of different glucose concentrations, and individual cells' dynamic behaviors are captured with fluorescent microscopy imaging [44, 42, 43].

In a related study [45], a similar approach was employed, but with a focus on monitoring an autofluorescent cytosolic cofactor called nicotinamide adenine dinucleotide (NAD) instead of fusion protein expression. NAD participates in the metabolic reaction network of glycolysis and undergoes cycling between its reduced and oxidized forms. When in its reduced form, NAD exhibits autofluorescence when excited by near-UV light (335 to 350 nm), with an emission peak around 440 to 470 nm [47]. Under specific conditions, the entire glycolysis network begins to exhibit oscillatory behavior, a phenomenon that has been studied since the 1950s [48]. By measuring the fluctuating intensities of NADH through fluorescence imaging, researchers can indirectly observe the oscillating intracellular glycolytic network without the need for external probes. These NADH intensity readouts serve as proxies for the oscillations, which are indicative of the intracellular glycolytic activity. Traditionally, these oscillations were investigated in bulk, where the entire population displayed synchronized oscillations due to the presence of the coupling agent acetaldehyde (ACA) formed during the process [49]. However, attempts to study individual cells at low biomass concentrations were inconclusive, as the cells did not exhibit sustained glycolytic oscillations. In the current study, the authors sought to understand why individual cells demonstrated different behavior compared to the population average. To accomplish this, they employed a microfluidic setup in conjunction with fluorescence microscopy for dynamic live, single-cell imaging. The results clearly demonstrated sustained glycolytic oscillations in individual cells, but with significant heterogeneity, as displayed in Fig. 36.6. The microfluidic setup provided high temporal and spatial resolution data, which were instrumental in developing a detailed kinetic model. This model proved valuable in identifying critical parameters that govern the capacity of individual cells to display oscillatory behavior autonomously, without the need for synchronization, thereby revealing the presence of an inherent oscillatory clock. The utilization of the microfluidic setup and single-cell analysis played a pivotal role in the successful execution of this study. Expanding on these findings, the subsequent objective, discussed next, aimed to uncover the underlying mechanisms that drive the transition from individual cellular behavior to coordinated collective behavior.

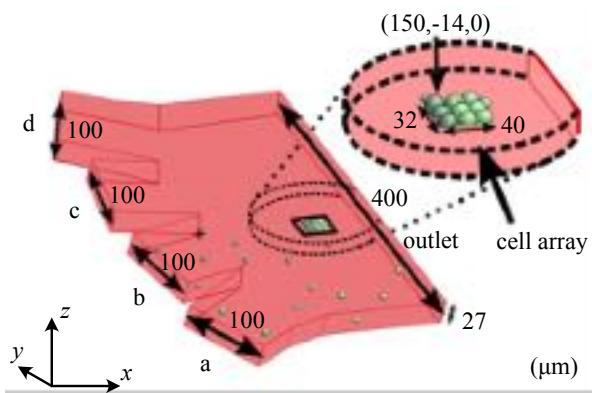


Figure 36.7: Channel junction and the yeast cell array site. The cells are introduced in channels a and b (in this example, two different strains are introduced in these separate inlets), and channels c and d are used for two different exposure solutions. The origin of the microfluidic coordinate system (x, y, z) is at the junction between the two middle inlet channels (marked with +). The cells are not drawn to scale. Reproduced from Ref. [44].

36.4.2 Coordinated cell behavior

Within the human body, organs comprise specialized cells that coordinate their actions both spatially and temporally. The preceding subsection examined glycolytic oscillations in isolated yeast cells using microfluidics and imaging techniques. The subsequent study aimed to investigate how yeast cells transition from individual behavior to influencing one another and establishing synchronized behavior [1]. The glycolytic coupling agent, ACA, generated within yeast cells through metabolic reactions, exits the cells by diffusion across the cell wall. Outside the cells, ACA can impact neighboring cells as long as it is not washed away by the fluid flow (as in the microfluidic device depicted in Fig. 36.7). Inspired by an organ-on-chip (OOC) and diffusion-based perfusion approach [50], the authors employed a new microfluidic design aimed at yeast cells. The cells were introduced into the microfluidic device, and the dimensions of the microfluidic system forced them to form a continuous monolayer. Upon exposure to the substrate solution (potassium cyanide and glucose), various oscillatory behaviors were observed, leading to the rapid emergence of subpopulations of cells that exhibited both spatial differentiation and temporal synchronization (Fig. 36.8). These subpopulations were identified using graph theory-based data analysis (using the software BRAPH, originally developed to study brain connectivity [51]). Adjacency matrices were constructed based on the autofluorescent signals emitted by the cells (Fig. 36.8 and [1] for detailed information). In the final step, the cell identities were overlaid onto the microfluidic cell images for low (Fig. 36.8c), median (Fig. 36.8d), and high (Fig. 36.8e) substrate concentrations. The results demonstrated that cells exposed to the lowest potassium cyanide concentration exhibited well-defined communities. However, at higher concentrations, the boundaries became less distinct (Fig. 36.8d), and cell coupling diminished further at the highest concentration (Fig. 36.8e). The fluctuating NADH intensities eventually led to the propagation of wavefronts through the cell population. By incorporating the architectural structure of the diffusion chamber into the simulation, the observed phenomena were corroborated, leveraging the detailed kinetic model of glycolysis. It was concluded that the generation of subpopulations arises from the interplay between glycolytic oscillations, metabolic cell-cell communication, and the fluid dynamics within the microfluidic device. Despite heterogeneous cell behavior, the spontaneous formation of communities resembles the zoned differentiation observed in multicellular organs such as the mammalian pancreas [52].

36.5 Problems

Problem 36.1: Fluid Dynamics in Microfluidic Systems, Design and Fabrication. Microfluidic systems offer precise control over the microenvironment, which can be a key advantage in experiments.

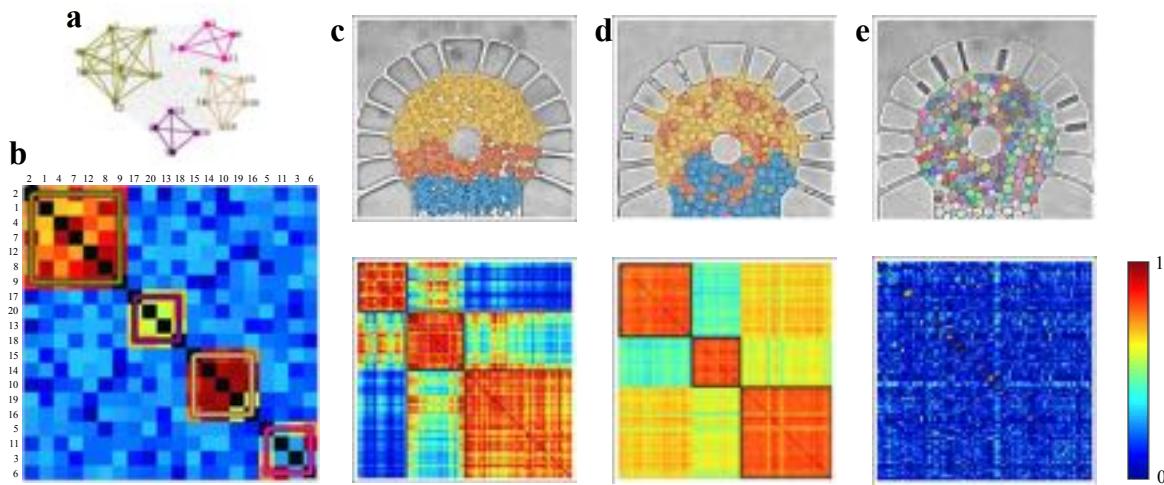


Figure 36.8: Subpopulations of synchronized communities in coordinated cell behavior studies. When analyzing functional graphs (**a**), the correlation of signals between different nodes provides a way to measure the strength of connectivity between them. By examining the resulting graph, we can identify communities of nodes (represented as colored subgraphs) that exhibit strong connections within their own community but weak connections to nodes outside of their community. This community structure is evident in the adjacency matrix (**b**) that represents the graph where the nodes have been rearranged in order to emphasize the community structure more clearly. Similarly, using the NADH autofluorescent signals from the cells (each labeled with a unique number for identification purposes), adjacency matrices (**c-e**) are constructed, and the resulting communities are superimposed onto images of the corresponding cell arrays for three different concentrations of CN^- : (**c**) 12 mM, (**d**) 16 mM, and (**e**) 20 mM. Reproduced from Ref. [1].

- a.** Discuss how the unique fluid behavior within microfluidic channels can be used for perturbation experiments on a population of yeast cells.
- b.** When designing a microfluidic system for studies involving cell communication, what critical design factors must be considered to meet their specific requirements? How would you evaluate the feasibility and reliability of the design?
- c.** What (approximate) dimensional factors should be considered in microfluidic design using PDMS to prevent sagging or unintended interactions between channels?
- d.** What are the key surface and permeability properties of PDMS? How can these properties be altered, if at all? While PDMS offers numerous benefits for working with biological samples, are there any drawbacks to its use? [*Hint: Take into account the role of laminar flow in micro-sized channels and the gas permeability of the material. What tools can be employed to predict fluid dynamics within a given microfluidic design?*]

Problem 36.2: Real-Time Data Acquisition. In microfluidic experiments, maintaining proper flow rates, ensuring component compatibility, and achieving efficient data acquisition are critical for success.

- a.** What challenges arise in maintaining consistent fluid flow and ensuring compatibility between syringes, tubing, and microfluidic chips?
- b.** What are the limitations of current miniaturized sensors and microscopy techniques for real-time data acquisition in microfluidic systems, and how can interdisciplinary research help overcome these obstacles, particularly in studies of cellular communication? [*Hint: Consider how material choice in microfluidic chips can affect imaging and spectroscopic analysis, the influence of flow rate on cell viability and attachment, and how emerging imaging technologies might improve the study of active matter in microfluidic environments.*]

Bibliography

- [1] M. Mojica-Benavides, D. D. van Niekerk, M. Mijalkov, J. L. Snoep, B. Mehlig, G. Volpe, M. Goksör, and C. B. Adiels. Intercellular communication induces glycolytic synchronization waves between individually oscillating cells. *Proceedings of the National Academy of Sciences*, 118:e2010075118, 2021.
- [2] N. T. Nguyen and S. T. Wereley. *Foundamentals and Applications of Microfluidics*. Artech House INC., London, Britain, 2006.
- [3] P. Tabeling. *Introduction to microfluidics*. Oxford University Press on Demand, 2005.
- [4] D. J. Beebe, G. A. Mensing, and G. M. Walker. Physics and applications of microfluidics in biology. *Annual review of biomedical engineering*, 4:261–286, 2002.
- [5] X. F. Peng and G. P. Peterson. Convective heat transfer and flow friction for water flow in microchannel structures. *International journal of heat and mass transfer*, 39:2599–2608, 1996.
- [6] COMSOL Multiphysics®. COMSOL AB, Stockholm, Sweden.
- [7] S. Battat, D. A. Weitz, and G. M. Whitesides. Nonlinear phenomena in microfluidics. *Chemical Reviews*, 122:6921–6937, 2022.
- [8] W. Limbut, K. Promsuwan, S. Kongkaew, P. Thavarungkul, and W. C. Mak. Emerging functional materials for microfluidic biosensors. In *Microfluidic Biosensors*, pages 195–231. Elsevier, 2023.
- [9] Q. Ge, Z. Li, Z. Wang, K. Kowsari, W. Zhang, X. He, J. Zhou, and N. X. Fang. Projection micro stereolithography based 3d printing and its applications. *International Journal of Extreme Manufacturing*, 2:022004, 2020.
- [10] S. Battat, D. A. Weitz, and G. M. Whitesides. An outlook on microfluidics: the promise and the challenge. *Lab on a Chip*, 22:530–536, 2022.
- [11] K. Sott, E. Eriksson, and M. Goksor. Acquisition of single cell data in an optical microscope. *Lab on a Chip Technology: Biomolecular Separation and Analysis*, 2:151–166, 2009.
- [12] A.-K. Gustavsson, A. A. Banaeian, D. D. Van Niekerk, J. L. Snoep, C. B. Adiels, and M. Goksör. Studying glycolytic oscillations in individual yeast cells by combining fluorescence microscopy with microfluidics and optical tweezers. *Current Protocols in Cell Biology*, 82:e70, 2019.
- [13] K. Raj M. and S. Chakraborty. PDMS microfluidics: A mini review. *Journal of Applied Polymer Science*, 137:48958, 2020.
- [14] J. C. McDonald, D. C. Duffy, J. R. Anderson, D. T. Chiu, H. Wu, O. J. A. Schueller, and G. M. Whitesides. Fabrication of microfluidic systems in poly (dimethylsiloxane). *ELECTROPHORESIS: An International Journal*, 21:27–40, 2000.
- [15] C. J. Ochs, J. Kasuya, A. Pavesi, and R. D. Kamm. Oxygen levels in thermoplastic microfluidic devices during cell culture. *Lab on a Chip*, 14:459–462, 2014.
- [16] T. C. Merkel, V. I. Bondar, K. Nagai, B. D. Freeman, and I. Pinna. Gas sorption, diffusion, and permeation in poly (dimethylsiloxane). *Journal of Polymer Science Part B: Polymer Physics*, 38:415–434, 2000.
- [17] D. A. Markov, E. M. Lillie, S. P. Garbett, and L. J. McCawley. Variation in diffusion of gases through PDMS due to plasma surface treatment and storage conditions. *Biomedical microdevices*, 16:91–96, 2014.
- [18] A. Lamberti, S. L. Marasso, and M. J. R. A. Cocuzza. PDMS membranes with tunable gas permeability for microfluidic applications. *RSC Advances*, 4:61415–61419, 2014.
- [19] M. W. Toepke and D. J. Beebe. PDMS absorption of small molecules and consequences in microfluidic applications. *Lab on a Chip*, 6:1484–1486, 2006.

- [20] N. Li, M. Schwartz, and C. Ionescu-Zanetti. PDMS compound adsorption in context. *SLAS Discovery*, 14:194–202, 2000.
- [21] T. G. Oyama, K. Oyama, and M. Taguchi. A simple method for production of hydrophilic, rigid, and sterilized multi-layer 3D integrated polydimethylsiloxane microfluidic chips. *Lab on a Chip*, 20:2354–2363, 2020.
- [22] Y. Yuan and T. R. Lee. Contact angle and wetting properties. In *Surface Science Techniques*, pages 3–34. Springer, 2013.
- [23] D. Bodas and C. Khan-Malek. Hydrophilization and hydrophobic recovery of PDMS by oxygen plasma and chemical treatment – an SEM investigation. *Sensors and Actuators B: Chemical*, 123:368–373, 2007.
- [24] I. Pereiro, A. F. Khartchenko, L. Petrini, and G. V. Kaigala. Nip the bubble in the bud: a guide to avoid gas nucleation in microfluidics. *Lab on a Chip*, 19:2296–2314, 2019.
- [25] S. Acharya, V. K. Mishra, and J. K. Patel. Optimization in fluid mixing in microchannels: A review. *Recent Advances in Thermofluids and Manufacturing Engineering*, pages 73–84, 2023.
- [26] M. A. Unger, H.-P. Chou, T. Thorsen, A. Scherer, and S. R. Quake. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science*, 288:113–116, 2000.
- [27] K. Hosokawa and R. Maeda. A pneumatically-actuated three-way microvalve fabricated with polydimethylsiloxane using the membrane transfer technique. *Journal of Micromechanics and Microengineering*, 10:415, 2000.
- [28] K. Amreen, K. Guha, and S. Goel. An overview of integrated miniaturized/microfluidic electrochemical biosensor platforms for health care applications. *Next Generation Smart Nano-Bio-Devices*, pages 81–101, 2023.
- [29] Y. Liu and X. Zhang. Microfluidic-based plasmonic biosensors. In *Microfluidic Biosensors*, pages 287–312. Elsevier, 2023.
- [30] S. T. Gebreyesus, A. A. Siyal, R. B. Kitata, E. S.-W. Chen, B. Enkhbayar, T. Angata, K.-I. Lin, Y.-J. Chen, and H.-L. Tu. Streamlined single-cell proteomics by an integrated microfluidic chip and data-independent acquisition mass spectrometry. *Nature Communications*, 13:1–13, 2022.
- [31] J. S. del Río, J. Ro, H. Yoon, T.-E. Park, and Y.-K. Cho. Integrated technologies for continuous monitoring of organs-on-chips: Current challenges and potential solutions. *Biosensors and Bioelectronics*, page 115057, 2023.
- [32] J. F. Zhong, Y. Feng, and C. R. Taylor. Microfluidic devices for investigating stem cell gene regulation via single-cell analysis. *Current Medicinal Chemistry*, 15:2897–2900, 2008.
- [33] F. K. Balagaddé, L. You, C. L. Hansen, F. H. Arnold, and S. R. Quake. Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science*, 309:137–140, 2005.
- [34] A. S. Tsagkaris, J. L. D. Nelis, K. Campbell, C. T. Elliott, J. Pulkrabova, and J. Hajslova. Smartphone and microfluidic systems in medical and food analysis. In *Microfluidic Biosensors*, pages 233–257. Elsevier, 2023.
- [35] H. Cha, H. Fallahi, Y. Dai, D. Yuan, H. An, N.-T. Nguyen, and J. Zhang. Multiphysics microfluidics for cell manipulation and separation: A review. *Lab on a Chip*, 22:423–444, 2022.
- [36] N. A. Erkamp, R. Qi, T. J. Welsh, and T. P. J. Knowles. Microfluidics for multiscale studies of biomolecular condensates. *Lab on a Chip*, 23:9–24, 2023.
- [37] J. B. Holman, Z. Shi, A. A. Fadahunsi, C. Li, and W. Ding. Advances on microfluidic paper-based electroanalytical devices. *Biotechnology Advances*, in press, 2023.

- [38] L. R. Huang, E. C. Cox, R. H. Austin, and J. C. Sturm. Continuous particle separation through deterministic lateral displacement. *Science*, 304:987–990, 2004.
- [39] M. Wu, A. Ozcelik, J. Rufo, Z. Wang, R. Fang, and T. Jun Huang. Acoustofluidic separation of cells and particles. *Microsystems & nanoengineering*, 5:32, 2019.
- [40] A. Pavlic, C. L. Harshbarger, L. Rosenthaler, J. G. Snedeker, and J. Dual. Sharp-edge-based acoustofluidic chip capable of programmable pumping, mixing, cell focusing, and trapping. *Physics of Fluids*, 35, 2023.
- [41] E. Eriksson, K. Sott, F. Lundqvist, M. Sveningsson, J. Scrimgeour, D. Hanstorp, M. Goksör, and A. Granéli. A microfluidic device for reversible environmental changes around single cells using optical tweezers for cell selection and positioning. *Lab on a Chip*, 10:617–625, 2010.
- [42] L. Bendrioua, M. Smedh, J. Almquist, M. Cvijovic, M. Jirstrand, M. Goksör, C. B. Adiels, and S. Hohmann. Yeast amp-activated protein kinase monitors glucose concentration changes and absolute glucose levels. *Journal of Biological Chemistry*, 289:12863–12875, 2014.
- [43] R. Babazadeh, C. B. Adiels, M. Smedh, E. Petelenz-Kurdziel, M. Goksör, and S. Hohmann. Osmostress-induced cell volume loss delays yeast hog1 signaling by limiting diffusion processes and by hog1-specific effects. *PLoS One*, 8:e80901, 2013.
- [44] C. Hamngren Blomqvist, P. Dinér, M. Grøtli, M. Goksör, and C. B. Adiels. A single-cell study of a highly effective hog1 inhibitor for in situ yeast cell manipulation. *Micromachines*, 5:81–96, 2014.
- [45] A.-K. Gustavsson, D. D. van Niekerk, C. B. Adiels, F. B. du Preez, M. Goksör, and J. L. Snoep. Sustained glycolytic oscillations in individual isolated yeast cells. *The FEBS journal*, 279:2837–2847, 2012.
- [46] J. Nielsen. Yeast systems biology: model organism and cell factory. *Biotechnology Journal*, 14:1800421, 2019.
- [47] G. H. Patterson, S. M. Knobel, P. Arkhammar, O. Thastrup, and D. W. Piston. Separation of the glucose-stimulated cytoplasmic and mitochondrial nad (p) h responses in pancreatic islet β cells. *Proceedings of the National Academy of Sciences*, 97:5203–5207, 2000.
- [48] P. Richard. The rhythm of yeast. *FEMS Microbiology Reviews*, 27:547–557, 2003.
- [49] P. Richard, B. M. Bakker, B. Teusink, K. Van Dam, and H. V. Westerhoff. Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells. *European Journal of Biochemistry*, 235:238–241, 1996.
- [50] A. A. Banaeian, J. Theobald, J. Paukštyte, S. Wölfel, C. B. Adiels, and M. Goksör. Design and fabrication of a scalable liver-lobule-on-a-chip microphysiological platform. *Biofabrication*, 9:015014, 2017.
- [51] M. Mijalkov, E. Kakaei, J. B. Pereira, E. Westman, G. Volpe, and Alzheimer's Disease Neuroimaging Initiative. Braph: a graph theory software for the analysis of brain connectivity. *PloS One*, 12(8):e0178798, 2017.
- [52] T. Suzuki, T. Kanamori, and S. Inouye. Quantitative visualization of synchronized insulin secretion from 3d-cultured cells. *Biochemical and Biophysical Research Communications*, 486:886–892, 2017.

Chapter 37

3D Printing

STEPHEN EBBENS, PIYUSH KUMAR, DAVID A. GREGORY, PATRICK SMITH, XIUBO ZHAO

Experimentally accessing active matter in an efficient way relies on the ability to control materials shape and composition at small scales. This critical link between active matter shape, composition, and resulting behavior has been established repeatedly. The very first example of artificial active matter exploited a rod-like shape to reduce Brownian rotation via asymmetrical drag, and relied critically on a heterogeneous bi-metallic composition to produce propulsion [2]. Since then, the number of realizations of active matter has grown dramatically [3]. However, compositional asymmetry remains key to the majority of motion-producing mechanisms. The shape and size of active matter determine the competition between Brownian effects, fluid drag, and mechanistic efficiency, and consequently the overall velocity, persistence length, angular velocity, and other trajectory features of the active particles. Using available synthesis and lithographic tools, experimentalists have realized a wide range of active matter shapes and sizes; however, relying on these established methods does impose intrinsic limitations. For example, active matter based on spherical colloids has received far more

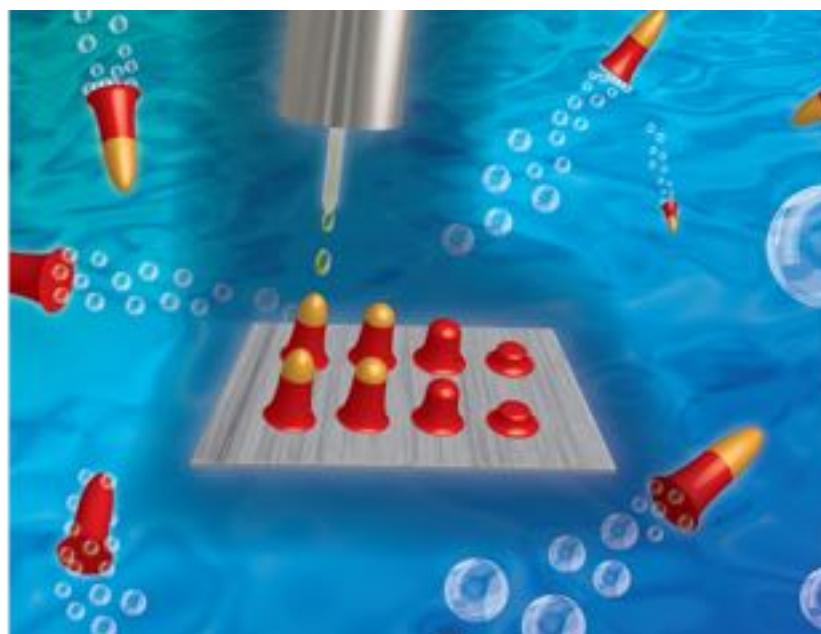


Figure 37.1: **Printing active particles.** Schematic showing the manufacturing process for silk *micro-roockets* using reactive inkjet printing (RIJ) to make biocompatible-enzyme powered devices. Reproduced from Ref. [1].

attention than ellipsoidal or more complex-shaped structures, due to the comparatively straightforward synthesis of materials with spherical symmetry. Additionally, as active matter becomes more complex and targets additional functionality (e.g., cargo capture and transport), the manufacturing methods to achieve the required compositional control become increasingly time-consuming. In this context, the development of the broad category of manufacturing methods termed *3D printing*, and their increasing performance at smaller length scales, has recently been adopted to provide a viable alternative. Indeed, in some cases 3D printing has replaced otherwise time-consuming and expensive multi-stage lithographic processes with a much simpler workflow, where complex active materials and devices are designed digitally, and rapidly produced. The lack of a fundamental limitation on the arrangement of materials and structural complexity that can be achieved via 3D printing also opens up routes to increasingly creative active matter realizations. Moreover, while not yet routine as a scalable manufacturing method, 3D printing at least has the potential, and industrial pull through, to be imagined to progressively fulfill this capacity. This Chapter provides a background to the basics of 3D printing in the active matter context, focusing mainly on inkjet-printing-based methods, which have been deployed to date to make new digitally-defined active materials (Fig. 37.1).

37.1 Introduction to 3D printing

3D printing, “the action or process of making a physical object from a three-dimensional digital model, typically by laying down many thin layers of a material in succession”, was first introduced by Charles Hull in 1983 in the form of *stereolithography* [4]. Since then, 3D printing has been extensively explored to fabricate materials for a variety of applications. With the assistance of *computer-aided design* (CAD), 3D printing can precisely fabricate structures in a bottom-up and layer-by-layer manner. As 3D printing offers high geometric accuracy, it has found extensive applications in the fabrication of 3D structures with complex internal architecture and equally complicated external topography without employing excessive tooling or machining. The diverse applications of 3D printing consequently include a wide range of areas, such as aerospace engineering, automotive industry, soft electronics, tissue engineering, regenerative medicine, food industry, drug delivery and medical devices, and delivery systems [5].

3D printing can be implemented in a variety of ways. This chapter focuses mainly on 3D inkjet printing, due to the demonstrated applications for active matter device manufacture, which will be discussed in detail in the next section. Here, however, we briefly introduce some other methods that can also find utility: stereolithography and extrusion printing via fused deposition modeling.

The basic principle of *stereolithography* is to form solid materials from an initially liquid bath, typically by exploiting photopolymerization [6]. To achieve this, a laser or a digital light projector is used to harden the liquid resin one layer at a time according to a digitally defined template. With augmentation via two-photon based methods, stereolithography can produce structures with 100 nm resolution, albeit with increased cost and reduced speed [7]. However, the intrinsic homogeneous composition of stereolithographically formed materials reduces the overall utility for active matter device manufacture. Despite this, stereolithography has found applications in defining the base structures of complex shaped active devices, which then undergo further functionalization to introduce compositional asymmetry as described in Section 37.2.4. The ability to produce smooth-sided spatially-defined structures has also the potential to be exploited to produce microfluidic channels and other topographic features, which have proven capacity to allow directional control over catalytic Janus colloids [8]. For example, stereolithography has been successfully employed to make channels from PDMS with a width of 50 microns [9]. The ability to digitally define microfluidic channels could allow more efficient testing of the many proposed schemes involving active matter interacting with well-defined topographies, for example to perform separations based on type of motion [10].

Fused deposition modelling, while likely to be considered crude for most active matter research, is the most common and most available 3D printing platform, and responsible for the popularization of the idea of 3D printing amongst hobbyists. It is now routinely used for prototyping in many research labs. For this 3D printing method, molten thermoplastics are extruded from a heated printing nozzle to construct solid materials in layers [11]. This printing method has been used to make fluid containing chambers, however at much larger length scales than stereolithography [12]. The finish of materials produced by this method is also rough, rendering it less appropriate in general for producing topographic features with enough regularity to assist investigations of active matter, compared to stereolithography. Likewise, only larger, crude active device structures could be envisaged to be currently manufactured via extrusion, and the limited material choices would suggest that

further functionalization stages would be required to impart activity. Nonetheless, access to this form of 3D printer can certainly be an enabler for active matter research via the ready ability to fabricate bespoke fluid connectors, sample holders and other enabling components for experiments [13].

37.2 3D inkjet printing

Inkjet printing is a printing method whereby an ink is jetted through the nozzle of a jetting device in the form of droplets in the nano- to picoliter volume range. Inkjet printing can work either using a single-ink system or a multiple-ink system, with both the approaches capable of precisely printing at micrometer resolution with few restrictions in the achievable geometric complexity. Inkjet printing is commonly known through its routine use to 2D print documents, however, with the correct choice of ink, it is possible to use inkjet printing to generate 3D structures in a layer-by-layer fashion. In fact, this form of 3D printing offers key advantages compared to stereolithography and extrusion printing, namely: the much wider range of compositions of materials that can be deposited, and the straightforward deposition of multiple inks with differing compositions. As almost all active matter devices rely on a heterogeneous and well-defined composition and shape, it is clear why inkjet printing is of considerable potential utility for their manufacture. Resultantly, the rest of this Chapter focuses on describing the details of this printing method, and some of the existing applications to make active matter devices. Compared to decorative 2D printing, 3D inks must possess the ability to form solid materials that can be layered on top of each other, while retaining the fluid properties that allow jetting. Numerous approaches to achieve this goal have been developed, including printing curable materials, or exploiting a carrier solvent which is removed via natural or forced drying to form structures from dissolved materials or suspended solids. Much of the challenge in inkjet printing is in formulating an ink that is both printable, but also leads to the desired solid material.

3D inkjet printing is not perfect; some of the limitations of this method include: the need for printable materials to meet certain specific requirements of viscosity and surface tension [14], the mechanical stress and heat involved during the jetting process affecting the activity of sensitive biomaterials, and a range of issues with inks and inkjet nozzles, such as aggregation within ink reservoirs during printing and clogging. Nonetheless, the advantages of inkjet printing overshadow its drawbacks and offer the ability to fabricate multi-material and multi-scale constructs with complex designs at a high printing resolution, as its building block, i.e., the jetted droplet, has a very small volume in the picoliter range.

37.2.1 Actuation methods

Inkjet printing can be classified into two main variants on the basis of the actuation technique or the mode of jetting, namely, thermal and piezoelectric. In *thermal inkjet printing* (Fig. 37.2a), an annular heater surrounds the nozzle which holds the fluid ink. When the CAD signal is received from a computer, the nozzle rapidly heats up ($\approx 300^\circ\text{C }\mu\text{s}^{-1}$), causing sudden vaporization and bubble formation which generates fluid displacement and the jetting of ink droplets. In *piezoelectric inkjet printing* (Fig. 37.2b), an annular layer of piezoelectric material surrounds the fluid cavity of the jetting device. When the CAD signal is received, the piezoelectric material contracts, or squeezes, causing ejection of the fluid out of the nozzle. Other actuation techniques are possible, and these include electrostatic, electrohydrodynamic, solenoid-valve, and acoustic inkjet printing. Some of these methods have more specific fluid criteria, and in some cases reduced resolution. For example, a conductive ink is required in the case of electrostatic inkjet printing. Solenoid-valve inkjet printing produces large droplet sizes of $500\text{ }\mu\text{m}$ or more, causing low resolution printing. Finally, upside-down substrate placement is required in the case of acoustic printing.

37.2.2 Continuous and drop-on-demand inkjet printing

Based on the mode of deposition of the ink, inkjet printing can be classified as either continuous or drop-on-demand. In *continuous inkjet* (CIJ, Fig. 37.3a), a conductive fluid ink is jetted very fast, almost in the form of a continuous stream. After jetting, the ink is electrically charged and passed through an electrical field, which deflects the ink droplets towards the substrate in accordance with the CAD signal received from the computer. The non-deflected droplets are recycled and sent to the reservoir in the printhead so that they can be used for printing again. In *drop-on-demand* (DOD) inkjet printing (Fig. 37.3b), the jetting device actuates and jets

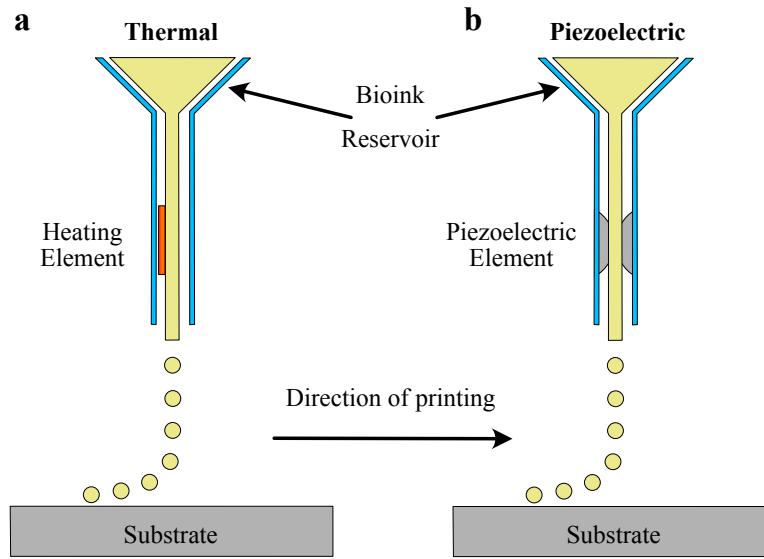


Figure 37.2: Thermal and piezoelectric inkjet printing processes. (a) In thermal inkjet printing, an annular heating element surrounds the nozzle which holds the fluid ink. By rapidly heating up, this element causes the jetting of ink droplets. (b) In piezoelectric inkjet printing, an annular layer of piezoelectric material surrounds the nozzle. By contracting, this element induces the ejection of a fluid droplet out of the nozzle. Reproduced from Ref. [15].

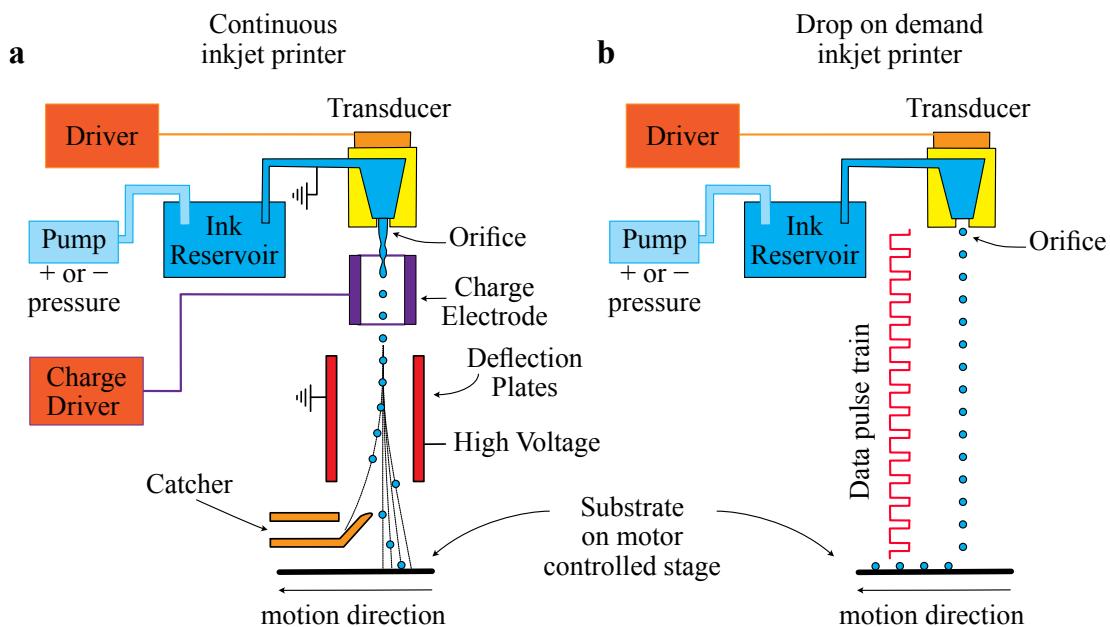


Figure 37.3: Inkjet printing modes. Schematic of (a) continuous and (b) drop-on-demand inkjet printing. Reproduced from MicroFab manual [16].

out the ink droplets only when the electrical signal is received according to the CAD model. For the correct placement of the deposited ink material, either the printhead or the substrate moves in Cartesian coordinates.

From the perspective of active matter and active device manufacture, CIJ has a few disadvantages, principally the obligation to use inks that can be charged, and the relatively lower printing resolution. Therefore, DOD is the most relevant inkjet printing technique for active matter systems. Drop-on-demand inkjet printing allows for drops to be released only when an electric signal is passed into the transducer as illustrated in Fig. 37.3b. In general, the transducer is a piezoelectric material which changes its structure when being electrically charged, as previously described. This structural change of the piezoelectric material alters the pressure and velocity of the fluid and the inner shape of the printhead directs it to the orifice, resulting in a drop of liquid being released. As shown in the illustration in Fig. 37.3b, every pulse programmed into the print driver releases a droplet. This, coupled with a moving sample stage, allows for complex patterns to be printed onto the substrate. The ink from the reservoir needs to be controlled via a back pressure system to make sure that the surface tension of the ink fluid is sufficient to stop it dribbling when there is no signal. This is generally done by means of a very slightly negative pressure but, depending on the ink, it may also be necessary to apply a small positive pressure especially with more viscous or high surface tension fluids.

37.2.3 Reactive 3D ink jet printing

Based on the mode of formation of the final 3D structure, inkjet printing can be categorized as either non-reactive inkjet printing (usually referred to as just inkjet printing), or reactive inkjet printing. In *non-reactive inkjet printing* (Fig. 37.4a), a single ink is simply deposited layer-by-layer on the substrate after which the 3D structure forms via physical changes, such as evaporation of solvent, or through chemical changes. For *reactive inkjet printing* (RIJ, Fig. 37.4b), two or more mutually reactive inks are used which react in-situ after jetting to form the required *printed* product [17]. Apart from the main ink, at least one additional supplementary ink is dispensed layer-by-layer at the same location on the substrate as the main ink, after which it reacts with the main ink and forms rigid 3D structures through irreversible chemical changes, such as cross-linking and polymerization of the precursor fluid in the main ink. A second approach is to simultaneously print additional inks, which then combine through mid-air collision with the main ink, e.g., via tilted jetting devices [18]. RIJ is particularly helpful in situations where the end product is incompatible with the jetting devices, or for fabricating structures composed of materials which are not printable themselves. These include solid end-product materials and materials which are insoluble or barely soluble to form printable liquid inks, but can instead be formed through reaction with different liquid reactant materials. It is also an excellent methodology to avoid having to produce emulsion inks that have a tendency for particles to agglomerate or settle down in the ink reservoir during the printing process and potentially clog the nozzles. The RIJ approach consequently opens a new window for printing a much wider variety of materials and exploring different solidification mechanisms.

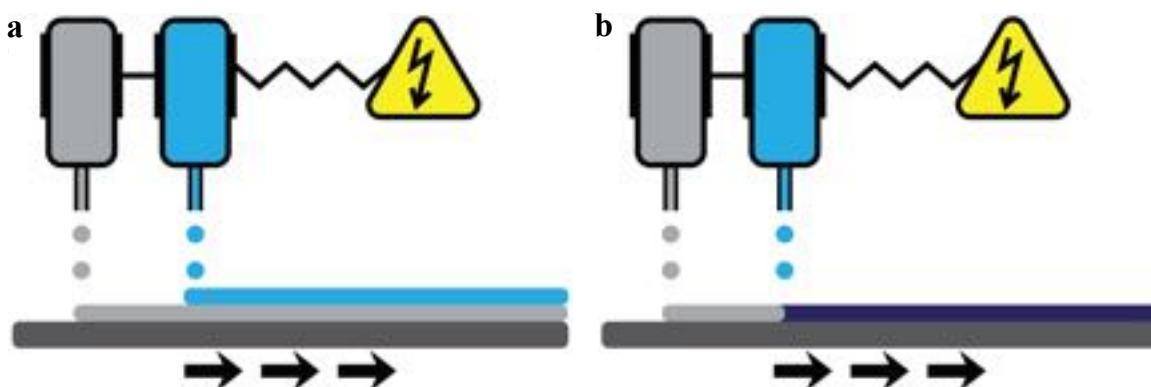


Figure 37.4: **Non-reactive and reactive inkjet printing.** Schematic illustrations of multi-headed (a) non-reactive and (b) reactive inkjet printing processes. In the non-reactive modality, different non-reacting inks form different consecutive layers on the substrate, whereas, in the reactive modality, different reacting inks form a layer of product on the substrate.

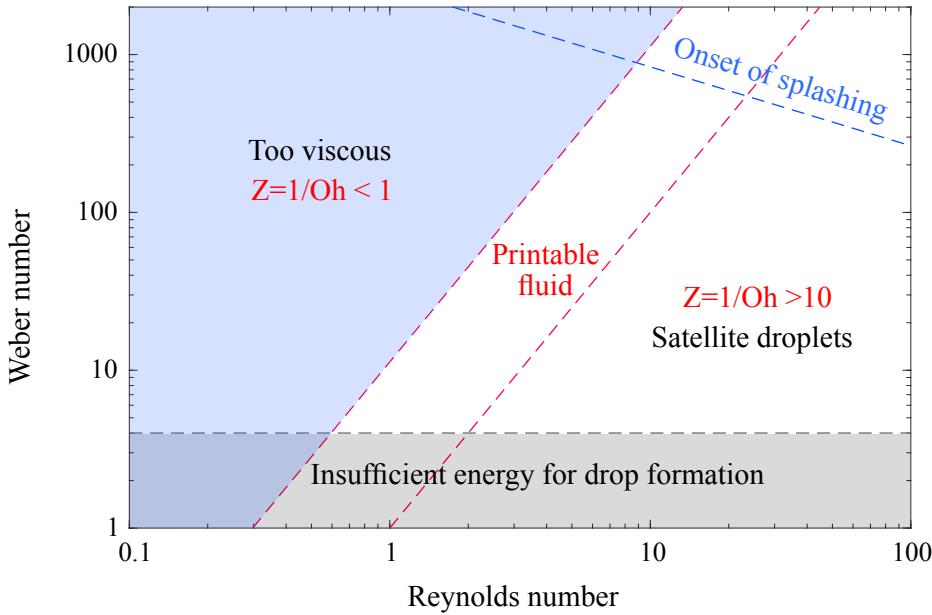


Figure 37.5: **Printability.** Range of Weber number and Reynolds number of a printable fluid. Reproduced from Ref. [19].

37.2.4 Printability

The printability of an ink by inkjet technology depends on the surface tension, viscosity, and inertia of the fluid that forms the bulk of the ink. The printable fluid window is determined by the relative values of the *Weber number* (We , the ratio of a fluids inertia compared to its surface tension) and the *Reynolds number* (Re , the ratio of the fluids inertial and viscous forces), as shown in Fig. 37.5. The requirements can be summarized via a derived quantity, Z , which is the reciprocal of the *Ohnesorge number* ($Oh = We/Re$). If the Z number is too low ($Z < 1$), the viscous forces prevent detachment of a droplet at the nozzle, whereas if the Z number is too high ($Z > 10$), the inertial forces cause formation of several small satellite droplets after the main droplet has already detached. Satellite droplets will significantly reduce the fidelity of the printed materials. Additionally, at low Weber numbers ($We < 4$), there is insufficient energy to form a droplet, and at high Weber numbers the droplets splash too much on contact with the substrate.

37.2.5 Droplet spacing and coverage

Another key parameter for inkjet printing is the droplet spacing, which must be optimized to ensure complete coverage of regions where structures are being deposited. Too small a spacing will result in wastage of materials, and hamper drying or chemical curing stages that must complete before subsequent layers, while too wide spacings will lead to structures that are not continuous (Fig. 37.6a,b). Considering experimental attempts to print clean-edged tracks (Fig. 37.6c), it is apparent that an optimum spacing exists. Indeed, a number of attempts to establish quantitative criteria for smooth-edged track formation have been proposed. A key factor in determining appropriate droplet spacing and material coverage is the ability to predict the diameter of each deposited droplet. As well as ink properties and jetting details, the role of the wettability of the substrate onto which each droplet lands also determines its diameter. Higher contact angle substrates will promote smaller droplets, while completely wetting substrates are clearly not viable as structure forming supports. Pasandideh-Fard *et al.* [21] have proposed a useful equation to determine the spreading factor, written in terms of We and Re , and including the advancing contact angle, θ [22]:

$$\frac{r_{\max}}{r} = \sqrt{\frac{We + 12}{3(1 - \cos \theta) + \frac{4We}{\sqrt{Re}}}}, \quad (37.1)$$

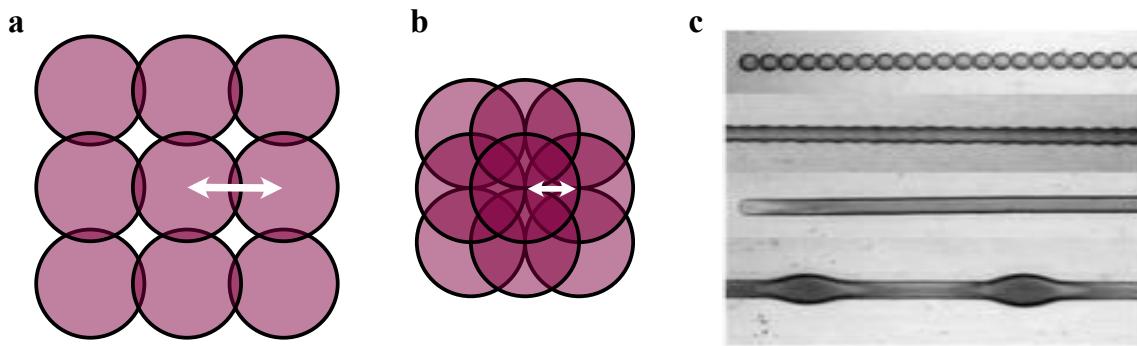


Figure 37.6: Droplet spacing and line printing. Illustration of the effect of droplet spacing on material coverage. (a) A wider droplet spacing is more efficient in use of materials, but can result in gaps in coverage, which may be undesirable. (b) Closer droplet spacing requires more material to be deposited, and can result in additional artefacts in the final material. (c) Effect of decreasing droplet spacing on an inkjet printing line (top to bottom). When spacing is too high, separate drops, or contact pinning effects are seen, when spacing is too low bulges appear [20].

where r is the radius of the expelled droplet and r_{\max} is the maximum radius of the circular area of the substrate that will contact the impacting droplet. However, in 3D printing, the situation is further complicated where droplets will deposit on a range of other materials with varying surface tensions as the structures are formed. Initially droplets contact the substrate, after that they contact the printed structure, so their interactions change during the printing process.

Note that there are analytical methods to calculate jetted droplet size, however these require knowledge of the forces acting on droplet as it is ejected from the nozzle which are hard to determine. Practically, most inkjet devices are equipped with the ability to stroboscopically image the droplets in flight, which can in turn be used to estimate droplet diameter. Contact angles can either be directly measured, or calculated based on knowledge of the materials' surface tension. In practice, optimization of the printing is undertaken via a combined application of the predictive powers of the available equations and tabulated material properties, in combination with a degree of trial-and-error changes based on the usually straightforward ability to image trial structures with optical microscopy or scanning electron microscopy. In addition, real inks for forming 3D structures will often require optimization of other factors as they will usually contain solids or solid forming materials within a volatile carrier fluid. This can indeed introduce further complexity due to drying effects, such as the *coffee ring effect* that can lead to inhomogeneous distribution of materials.

37.3 Printing active rockets and stirrers

This section describes one specific application of ink jet printing from the authors' work: making active enzymatically-driven silk rocket and stirrer devices. Catalase was used as the activity-producing enzyme, due to its ability to decompose hydrogen peroxide fuel and generate oxygen bubbles to locally produce motion as they detach. Silk was chosen as the solid-forming material, due to its biocompatibility and ability to be mixed with catalase and support enzyme activity. Converting the printable silk form ink (silk I) into solid non-soluble stable beta-sheets (silk II) after printing requires a reactive ink jet printing strategy, with an additional print head used to deposit methanol ink as a curing reagent.

We specifically aimed to use printing to investigate how both device shape and enzymatic activity location altered the devices' motion. Our strategy was to form devices from two main inks: *plain* silk, and a catalase/silk mixture. Compared to conventional methods, inkjet printing gave the attractive feature of allowing a wide range of device shapes and compositions to be made simply by changing the digital design.

Example 37.1: Printability of a methanol reactive ink. To build the 3D silk structures for our active

rockets, and stirrers, it was necessary to print methanol. Using the printability criteria established in Section 37.2, and Fig. 37.5, assess the printability of this ink, and suggest the minimum droplet velocity that may be used.

Solution: Firstly, we need to write the relevant dimensionless quantities and ratios given in the text into equations that we can evaluate.

The Weber number is given by:

$$\text{We} = \frac{v^2 \rho l}{\gamma}, \quad (37.2)$$

where v is the fluid velocity at the nozzle, ρ is its density, γ its surface tension and l is the relevant length scale, in this case the nozzle diameter.

The Reynolds number is given by

$$\text{Re} = \frac{v \rho l}{\mu}, \quad (37.3)$$

where μ is the fluid viscosity.

Consequently, we find that Z is given by:

$$Z = \frac{1}{\text{Oh}} = \frac{\text{Re}}{\sqrt{\text{We}}} = \frac{\sqrt{l \rho \gamma}}{\mu}, \quad (37.4)$$

The values of surface tension, density and viscosity for methanol, assuming operation at room temperature and atmospheric pressure, are: $\rho = 792 \text{ kg m}^{-3}$, $\gamma = 0.022 \text{ N m}^{-1}$, $\mu = 0.00056 \text{ Pas}$. A typical inkjet nozzle diameter is $l = 50 \mu\text{m}$, which allows us to obtain $Z = 53$.

Clearly this is outside of the ideal printability range ($Z < 10$). In practice we were able to purchase small diameter, $l = 20 \mu\text{m}$, nozzles to make $Z = 33$, however this was at the expense of increased chances of clogging. Indeed, as predicted by operating in this regime, in our experiments we observed the formation of satellite droplets. However, this was acceptable for deployment as the *curing* reactive ink rather than the main material-forming ink.

We can determine a minimum droplet velocity, using the criterion $\text{We} = 4$, which requires $v = 2.4 \text{ m s}^{-1}$ for a $20 \mu\text{m}$ nozzle. In practice, usual velocity ranges at the nozzle during inkjet printing are of the order of 3 to 10 ms^{-1} .

Example 37.2: Droplet spacing for 3D printing. Assuming that the ejected droplet size is $50 \mu\text{m}$, its velocity is $v = 5 \text{ m s}^{-1}$, the nozzle diameter is $20 \mu\text{m}$, and the droplet contact angle is 60° , estimate the droplet spacing required to produce a *just touching* coverage for the methanol curing ink.

Solution: We can estimate the spreading factor for a methanol droplets using Eq. (37.1). In this case, the Weber number is $\text{We} = 18$ and the Reynolds number is $\text{Re} = 140$. Consequently, the spreading factor is 1.24, indicating that the impacting droplet can generate a maximum diameter of $62 \mu\text{m}$. A droplet spacing of this value would consequently theoretically produce droplets that are touching at their diameters, so to ensure adequate overlap we would usually print with a slightly smaller spacing.

37.3.1 The printing equipment and controllers

While a large range of commercial printers are available, our structures were made using a custom-built printer, shown in Fig. 37.7, giving the advantages of a complete flexibility of geometry, environmental control, and customization. The key features required are a xyz sample translation stage, pressurized fluid supplies, jetting heads (here, we used a jetting device from MicroFab Technologies Inc.), and control software to move the stage and supply the voltages to actuate the jetting heads according to the CAD design. In our case, the software was developed using a LabVIEW platform (National Instruments Corporation, USA). An optional, but very useful addition is the ability to perform stroboscopic imaging of the flight of the droplets to ensure printing conditions are optimized.

The details of droplet expulsion in this piezoelectric, drop-on-demand system are controlled by a voltage pulse supplied to the jetting device (Fig. 37.8). In our case, we were able to adjust the shape of this pulse using our software, which enables both initial optimization and on-the fly adjustments during printing as required, while checking that we were generating single droplets using the stroboscopic imaging capacity (Fig. 37.9).

37.3.2 Computer aided designs of fibroin structures

While sophisticated 3D CAD packages are available, it is also possible to define shapes using spreadsheet software, as the only input that is required for 3D shapes that do not change in the z -direction is a 2D array of coordinates. The geometric shapes of the 3D structures were designed in the form of groups of points in the 2-dimensional xy coordinates (Fig. 37.10). Each complete layer of the device required three printing stages, each using a different print-head charged with the appropriate ink. First, the blue *dots* were printed using the main silk ink. Second, the red *dots* were printed using a second, catalase-doped silk ink. Finally, the outline of the entire shape was overprinted with methanol, to act as the curing stage (Fig. 37.11). This process deposited one layer of each device and was then repeated until the desired thickness was obtained.

37.3.3 Printing details

Using this general framework, additional optimizations were made to lead to our final protocol. Substrate choice is important, as the first layer of material printed is in contact with the supporting surface. We used silicon wafers, which were more hydrophobic than glass slides, resulting in higher fidelity of the 3D shape by reducing spreading of the deposited droplets on the substrate. Another important consideration to make for the choice of substrate for manufacturing active matter is the need to detach the devices easily after fabrication.

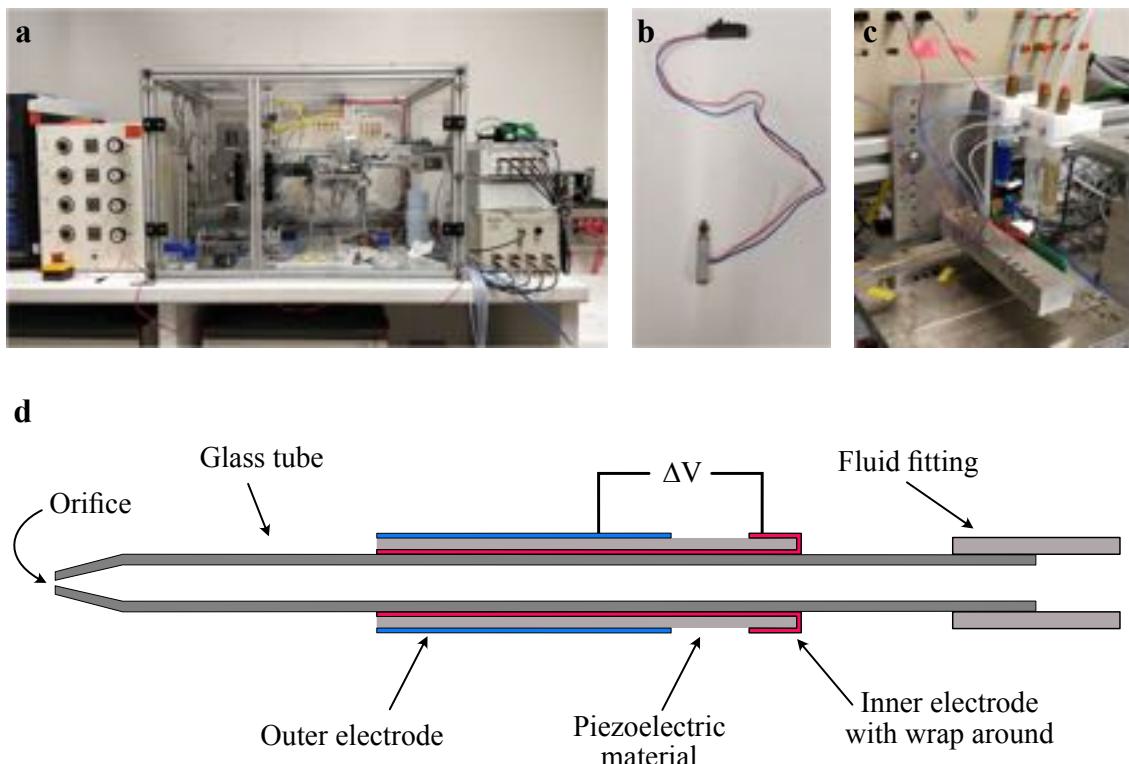


Figure 37.7: **A homebuilt RIJ platform.** (a) Overview of a homebuilt RIJ platform. (b) A printhead and electrical connections and (c) fluid supplies. (d) Schematic cross-section of a printhead. Image (d) reproduced from MicroFab manual [16].

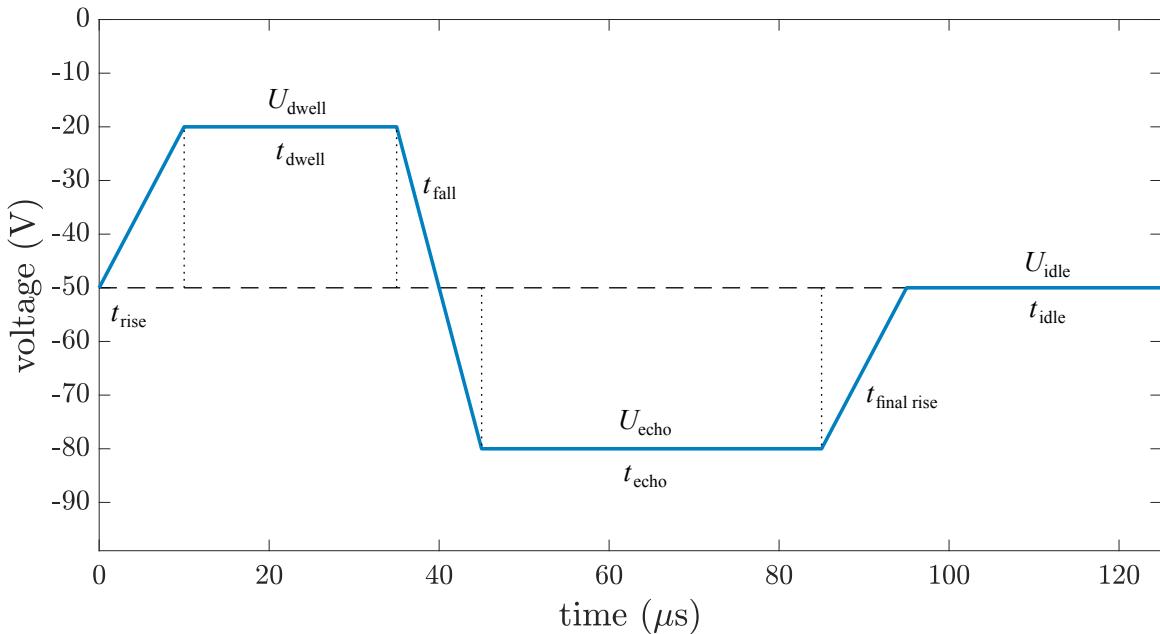


Figure 37.8: **Printing waveform.** A model illustration of the voltage pulse parameters utilised to control the printing. Here, t_{fall} represents the voltage which compresses the piezoelectric material inside the jetting device, resulting in the fall of a droplet. Reproduced from MicroFab manual [16].

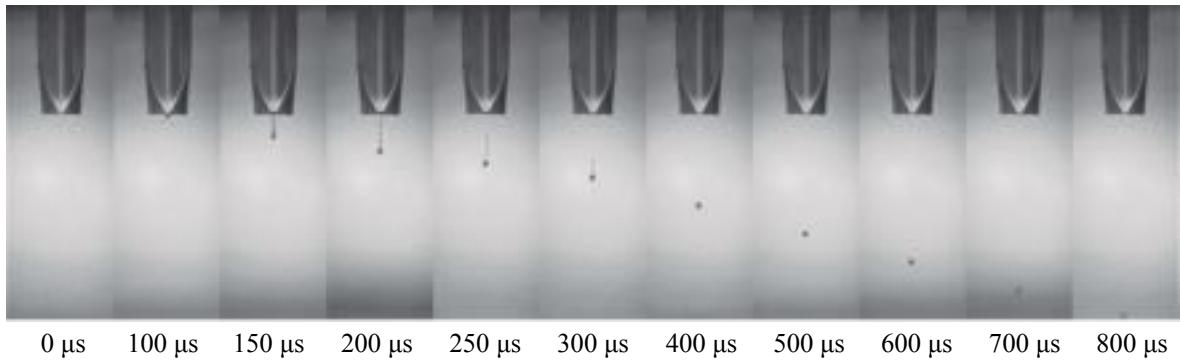


Figure 37.9: **Droplet ejection.** The time-lapse images of the nozzle of a jetting device captured during printing using a macro-lens camera. The ejecting ink droplet is clearly visible. The numbers below the images represent the time elapsed, in microseconds, since the initiation of jetting of the fibroin ink droplet. Reproduced from Ref. [23].

This can entail the need to print an additional sacrificial initial layer to encourage detachment.

Silicon wafers were cleaned with 5% Decon-90 solution followed by 70% ethanol and deionized water, after which they were carefully wiped dry and secured firmly on the printing stage. Inks were loaded into separate reservoirs and then the backpressure to each jetting device was adjusted to ensure inks were not leaking from the jetting devices. Before printing, all the channels were calibrated for accurate positioning on the substrate. For this, a dot array was printed at the same coordinate location with all the printing channels. Any deviations in the droplet positioning were then corrected by shifting the channel position. Printing was done through the alternate layer-by-layer deposition of the primary ink and the curing ink as described previously (Fig. 37.11). For our final *main* structure forming inks, aqueous silk fibroin solution (40 mg mL^{-1}) and the catalase containing ink were mixed with PEG₄₀₀ (14 mg mL^{-1}) to help generate stable droplets. This illustrates the potential need to use additional additives to ensure printability. The distance between the jetting device nozzle and the substrate

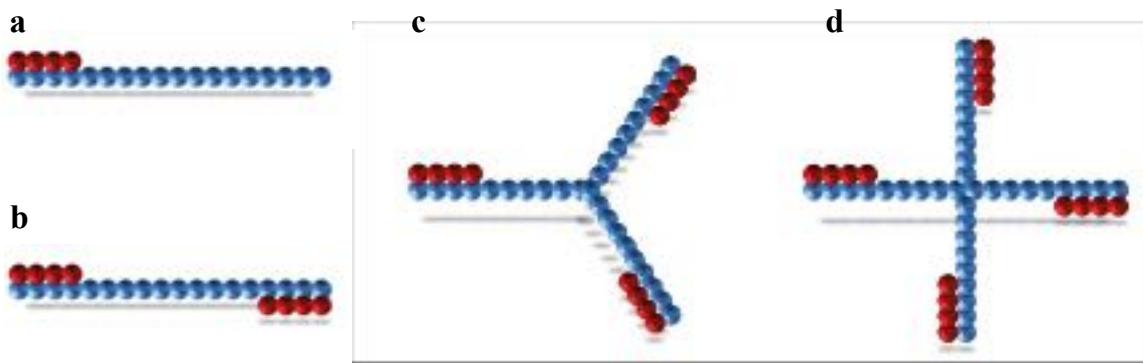


Figure 37.10: Digital shape and composition definition. An example of the 2D CAD of fibroin structures created in a spreadsheet. Blue dots indicate positions where pure silk ink will be printed, red dots indicate where the active silk/catalase ink will be printed. Printing several layers on top of each other in accordance to the 2D CAD gives the final 3D structures. In this case the shapes (**a-d**) were defined to investigate the effect of catalase location and overall device shape on stirring efficiency.

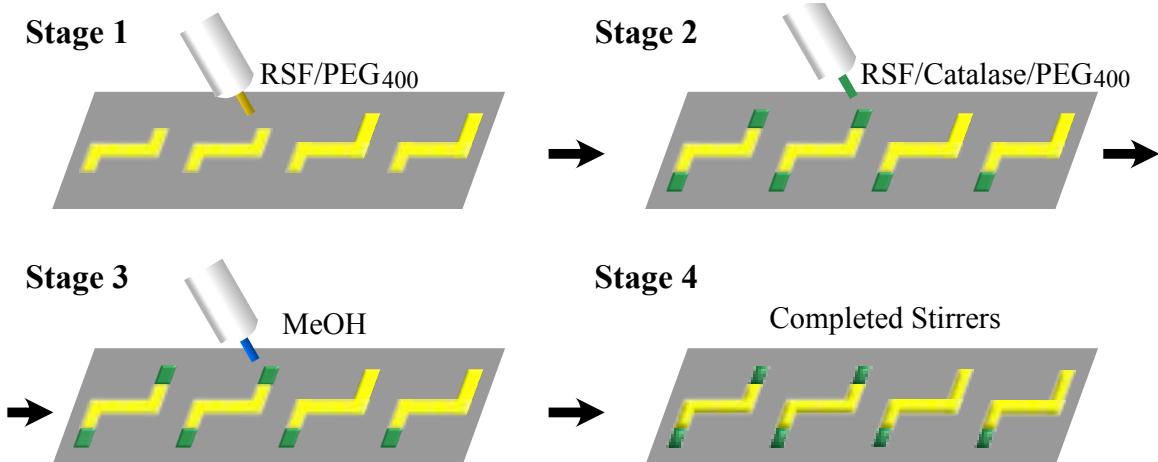


Figure 37.11: Printing process. The sequential stages of printing the active device structures [23]. The plain silk ink is labelled as RSF/PEG₄₀₀. PEG was used as an additional processing aid.

was kept at approximately 10 mm. During the printing process, the room temperature remained in the range of 20°C to 24°C and the relative humidity level in the room remained in the range of 60% to 70%. A good degree of environmental control was vital to our process. The number of layers printed determined the thickness of the finished structure. After printing, the 3D fibroin structures were removed from the substrate by soaking in DI water and gently agitating. Fig. 37.12 shows typical devices after printing, satellite droplets are visible, but these remained on the substrate and so did not hamper the device fidelity unduly.

37.3.4 Device performance

The printed active matter we have investigated so far has been at relatively large length scales (hundreds of microns). At these scales it has been clear that printing-based approaches can efficiently access the heterogeneous compositions needed to drive propulsion. Our earliest printed devices were rocket-shaped units made by building up columnar structures from the substrate [1]. While we digitally defined a constant cross-section for these devices, the layer-by-layer process and repeated methanol applications resulted in a tapered structure instead.



Figure 37.12: **RIJ active stirring devices.** Typical devices obtained after printing. The size of the devices is 8 mm.

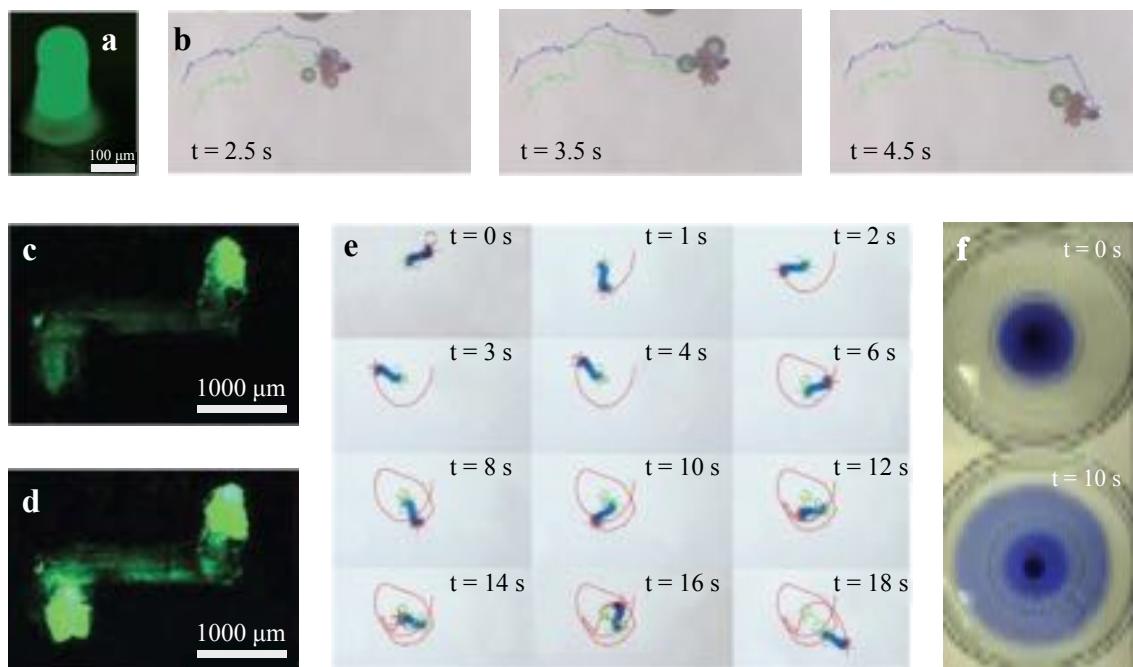


Figure 37.13: **Performance of RIJ active stirring devices.** (a) Rocket-shaped RIJ device with catalase labelled in green. (b) Rocket shaped device with Janus structure translates through fluid powered by bubble propulsion. (c-d) Z-shaped stirring devices with (c) one or (d) two enzymatically active ends (labelled green). (e) Time lapse of rotational trajectories generated by a z-shaped RIJ device with a single active end. (f) RIJ stirring devices show the ability to autonomously enhance fluid mixing. The image sequence shows a droplet of blue ink at the base of a Petri dish being rapidly dispersed by a rotating silk stirrer. Reproduced from Refs. [1, 24].

While not deleterious for our application, this shows the lack of complete fidelity for some inks when comparing 3D design to actual printed objects. The introduction of a Janus structure into these rockets simply by first printing a catalase doped ink, followed by a silk-only ink worked well, demonstrating that straightforward compositional control — a key target for active mater — is achievable. The resulting heterogeneous structures were confirmed by fluorescently labelling the catalase (Fig. 37.13a). As expected, the composition of the structure had a significant effect on device behavior, with the Janus structure producing consistent motion away from the catalase doped end due to bubble release (Fig. 37.13b), whereas a homogeneous distribution of catalase throughout the entire device produced only small translations, due to bubble release occurring isotropically.

From this starting point, we then printed a range of multi-armed structures, according to the designs shown in Fig. 37.11, positioning the catalase domains to encourage rotation rather than translation (Fig. 37.13c) [24]. In this study the utility for RIJ is clear, as we were able to vary both geometry and composition in a straightforward fashion. With one application area for these devices potentially being to stir small volumes of solutions RIJ allows us to simultaneously optimize performance for rotational speed and stirring efficiency, and use tools such as Computational Fluid Dynamics to generate optimum stirrer shapes, which can be easily made and tested. Fig. 37.13e shows example rotary trajectories that we were able to achieve, which we used to enhance mixing within Petri dishes (Fig. 37.13f).

37.4 Additional examples of printed active matter

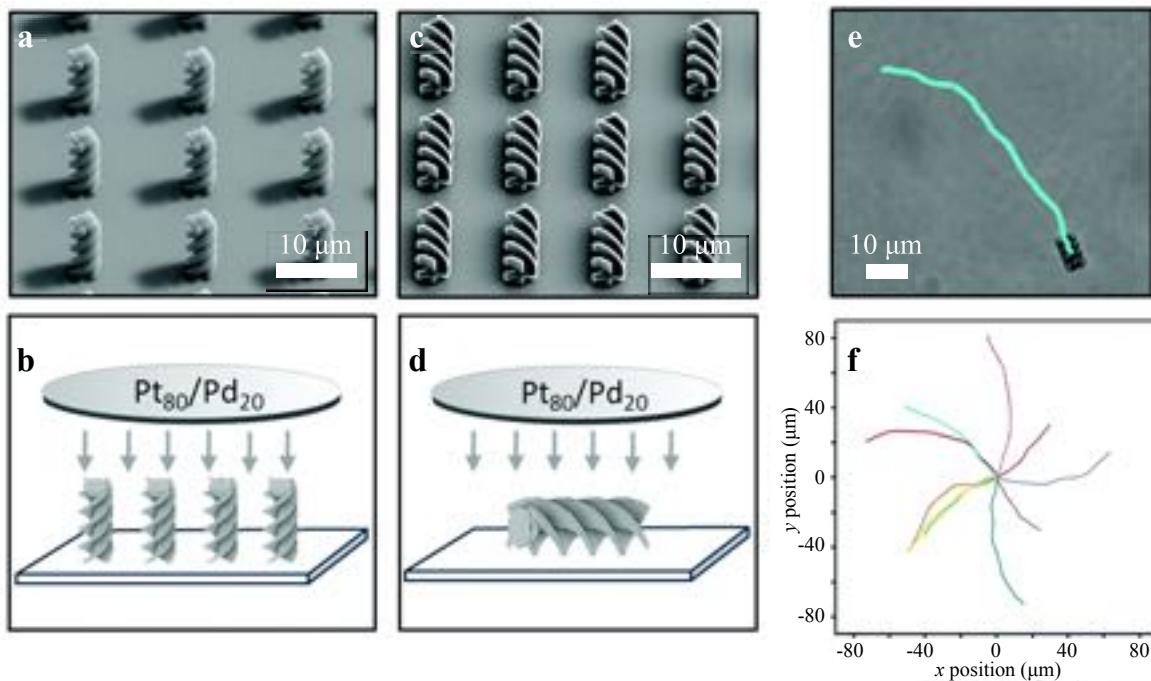


Figure 37.14: Active matter examples. (a,c) 3D printing allowed helical structures to be fabricated in two orientations relative to the substrate. This allowed the effect of the distribution of the active catalyst (Pt/Pd) to be varied during line of site metal evaporation (b,d). (e,f) Examples of the resulting trajectories. Reproduced from Ref. [25].

The Kraft group exploited the exquisite control of device geometry possible via 2-photon polymerization to make a range of active devices, with dimensions of around 10 μm [25]. The printed shapes were subsequently coated with catalytic metals (e.g., Pt) to enable propulsion. Sonication was used to detach the devices to allow investigation of their propulsive motion as a function of their morphology and catalyst distribution. This method allowed access to complex anisotropic geometries such as helices, and controlled the orientation of the devices during the evaporation of the propulsion generating metal coating. This allowed a fine control over the location of the active sites with respect to the device geometry (Fig. 37.14). This can be hard to achieve conventionally, as premanufactured anisotropic shaped materials often lie in a preferred orientation (e.g. flat on the substrate) when dispersed onto a substrate.

There are also some larger scale demonstrations of the potential to print materials to realize active systems. In particular, materials incorporating magnetic components were printed and manipulated using external magnetic fields to obtain complex deformations of the device. In one example, Fused Deposition Modelling was used to make a cm-scale device that *swum* under water in analogue to a Manta ray [26]. The ability to print complex actuated shapes may also provide a new route to access the non-time reversible series of deformations required to produce propulsion at smaller scales where viscous forces dominate [27].

37.5 Future directions and conclusions

It is clear that 3D printing provides a valuable tool-kit for experimental researchers in active matter. As the case study in this Chapter highlights, it is possible to currently use Reactive inkjet Printing to make complete active devices on the hundreds of micrometers scale, with digital control over the overall device shape and distribution of the activity generating materials. This represents a degree of flexibility to trial designs and conduct experimental optimization of performance that could not be achieved via conventional non-digital approaches. Furthermore, it is clear that the ability to accurately define complex shapes at the nanoscale via stereolithography is providing new opportunities for the hybrid manufacture of active devices.

Both RJJ and stereolithographic printing methods currently require medium cost printing equipment. The cheapest most available form of 3D printing, FDM, currently lacks resolution for direct active device manufacture at all but the largest scales. There is significant pull through via the large range of proposed industrial deployments for 3D printing to reasonably expect that instrument specifications are likely to improve with time. It should be pointed out however that the resolution limits for inkjet printing-based methods are to an extent fundamentally limited by the physics of droplet formation and wettability, and so making heterogeneous structures at the micro- and nano-scales that are of chief interest for active matter researchers may remain out of reach. Additional applications beyond direct manufacture of active matter can also be envisaged. Chief amongst these is the ability to digitally define topographic features that can serve to direct active matter and incorporate these structures within microfluidic devices. Indeed, specific 3D printing instrumentation for microfluidics, such as the Dolomite fluidic factory, is already available.

37.6 Problems

Agnese: create some problems based on the examples above. Maybe with different parameters.

Bibliography

- [1] D. A. Gregory, Y. Zhang, P. J. Smith, X. Zhao, and S. J. Ebbens. Reactive inkjet printing of biocompatible enzyme powered silk micro-rockets. *Small*, 12:4048–4055, 2016.
- [2] W. F. Paxton, P. T. Baker, T. R. Kline, Y. Wang, T. E. Mallouk, and A. Sen. Catalytically induced electrokinetics for motors and micropumps. *J. Am. Chem. Soc.*, 128:14881–8, 2006.
- [3] S. J. Ebbens. Active colloids: Progress and challenges towards realising autonomous applications. *Current Opinion in Colloid and Interface Science*, 21:14–23, 2016.
- [4] C. Hull. Apparatus for production of three dimensional objects by stereolithography, U.S. Patent US-6027324-A, 1986.
- [5] P. Kumar, S. Ebbens, and X. Zhao. Inkjet printing of mammalian cells – theory and applications. *Bio-printing*, 23:e00157, 2021.
- [6] J. Huang, Q. Qin, and J. Wang. A review of stereolithography: Processes and systems. *Processes*, 8:1138, 2020.
- [7] K.-S. Lee, R. H. Kim, P. Prabhakaran, D.-Y. Yang, T. W. Lim, and S. H. Park. Two-photon stereolithography. *J. Nonlinear Opt. Phys. Mater.*, 16:59–73, 2007.
- [8] S. Das, A. Garg, A. I. Campbell, J. Howse, A. Sen, D. Velegol, R. Golestanian, and S. J. Ebbens. Boundaries can steer active Janus spheres. *Nature Communications*, 6:8999, 2015.
- [9] G. Comina, A. Suska, and D. Filippini. PDMS lab-on-a-chip fabrication using 3D printed templates. *Lab on a Chip*, 14:424–430, 2014.
- [10] M. Wan, C. Olson Reichhardt, Z. Nussinov, and C. Reichhardt. Rectification of swimming bacteria and self-driven particle systems by arrays of asymmetric barriers. *Physical Review Letters*, 101:1–4, 2008.

- [11] R. B. Kristiawan, F. Imaduddin, D. Ariawan, Ubaidillah, and Z. Arifin. A review on the fused deposition modeling (FDM) 3D printing: Filament processing, materials, and printing parameters. *Open Engineering*, 11:639–649, 2021.
- [12] P. J. Kitson, M. H. Rosnes, V. Sans, V. Dragone, and L. Cronin. Configurable 3D-printed millifluidic and microfluidic ‘lab on a chip’ reactionware devices. *Lab on a Chip*, 12:3267–3271, 2012.
- [13] R. F. Quero, G. Domingos Da Silveira, J. A. Fracassi Da Silva, and D. P. de Jesus. Understanding and improving FDM 3D printing to fabricate high-resolution and optically transparent microfluidic devices. *Lab on a Chip*, 21:3715–3729, 2021.
- [14] D. Jang, D. Kim, and J. Moon. Influence of fluid physical properties on ink-jet printability. *Langmuir*, 25:2629–2635, 2009.
- [15] K. Hözl, S. Lin, L. Tytgat, S. Van Vlierberghe, L. Gu, and A. Ovsianikov. Bioink properties before, during and after 3D bioprinting. *Biofabrication*, 8:032002, 2016.
- [16] Microfab manual, 2012.
- [17] P. J. Smith and A. Morrin. Reactive inkjet printing. *Journal of Materials Chemistry*, 22:10965–10970, 2012.
- [18] M. Y. Teo, L. Stuart, H. Devaraj, C. Y. Liu, K. C. Aw, and J. Stringer. The in situ synthesis of conductive polyaniline patterns using micro-reactive inkjet printing. *Journal of Material Chemistry C*, 7:2219–2224, 2019.
- [19] B. Derby. Inkjet printing of functional and structural materials: Fluid property requirements, feature stability, and resolution. *Annual Review of Materials Research*, 40:395–414, 2010.
- [20] D. Soltman and V Subramanian. Inkjet-printed line morphologies and temperature control of the coffee ring effect. *Langmuir*, 24:2224–2231, 2008.
- [21] M. Pasandideh-Fard, Y. M. Qiao, S. Chandra, and J. Mostaghimi. Capillary effects during droplet impact on a solid surface. *Physics of Fluids*, 8(3):650–659, 1996.
- [22] B. A. Dwiyantoro. Capillary effects during droplet formation on the solid surface. *Advanced Material Research*, 683:889–893, 2013.
- [23] D. A. Gregory, P. Kumar, A. Jimenez-Franco, Y. Zhang, Y. Zhang , S. J. Ebbens, and X. Zhao. Reactive inkjet printing and propulsion analysis of silk-based self-propelled micro-stirrers. *Journal of Visualized Experiments*, 146:e59030, 2019.
- [24] Y. Zhang, D. A. Gregory, Y. Zhang, P. J. Smith, S. J. Ebbens, and X. Zhao. Reactive inkjet printing of functional silk stirrers for enhanced mixing and sensing. *Small*, 15:1–9, 2019.
- [25] R. P. Doherty, T. Varkevisser, M. Teunisse, J. Hoecht, S. Ketsetzi, S. Ouhajji, and D. J. Kraft. Catalytically propelled 3D printed colloidal microswimmers. *Soft Matter*, 16:10463–10469, 2020.
- [26] S. Qi, H. Guo, J. Fu, Y. Xie, M. Zhu, and M. Yu. 3D printed shape-programmable magneto-active soft matter for biomimetic applications. *Composites Science and Technology*, 188:107973, 2020.
- [27] E. M. Purcell. Life at low Reynolds number. *American Journal of Physics*, 45:3–11, 1977.

Chapter 38

Active Mechanical Metamaterials

CORENTIN COULAIS

Active matter consists of distributed energy transducers that pump energy and momentum into local mechanical degrees of freedom. In the vast majority of active matter systems such as active fluids, steric and hydrodynamic

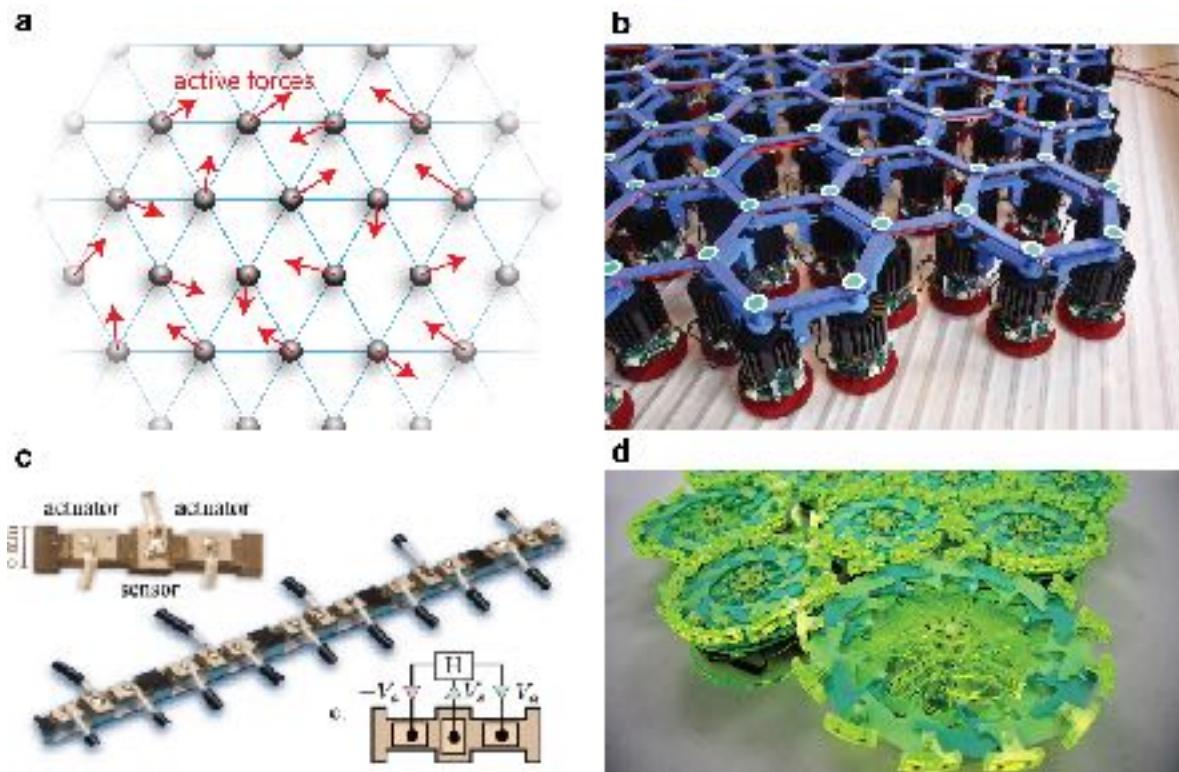


Figure 38.1: **Active mechanical metamaterials.** (a) Active mechanical metamaterials can be minimally represented by elastic networks with active forces. Crucially, these forces can explicitly depend on the time variable t in the case of time-modulated metamaterials or on the configuration of the system $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i, \dots)$ and $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots)$, where \mathbf{u}_i and \mathbf{v}_i are the instantaneous displacement and velocity of node i . (b-d) Possible experimental implementations, where (b) an elastic lattice [1], (c) a flexural beam [2], and (d) a cohesive granular medium [3] are dressed with motors, sensors and microprocessors. These mechatronic components can be used to apply active forces that depends on local displacements, velocities, and time.

interaction dominate so there is no reference state. In stark contrast, in active solids restoring forces are elastic, so there is a well-defined configuration of reference about which a well-defined (linear) response exists. This response is governed by the standard elastic, dissipative, and inertial forces that make up the material, and crucially by the active forces, which may confer novel attributes. These novel properties range from waves in overdamped regimes [4] and synchronized collective actuation [5] to unidirectional waves [6, 7, 8, 2] and limit cycles that do work and locomotion [1]. In this Chapter, we present active metamaterials — active solids that exhibit tunable properties owing to the design of their local energy injection and of their geometry. We discuss basic models, theoretical descriptions, and experimental implementations of such metamaterials and we discuss their distinctive mechanical properties. We cover two types of active metamaterials: (i) *time-modulated metamaterials*, where the property of one local building block is modulated in time such that time-reversal symmetry is explicitly broken; (ii) *odd metamaterials*, where local building blocks are designed to violate reciprocity. We will focus here on the linear regime, which is relatively well established. In contrast, we will leave the much unexplored nonlinear regime out.

38.1 Mathematical description of active mechanical metamaterials

An active mechanical metamaterial can be generically represented by the sketch displayed in Fig. 38.1a: an elastic network, whose masses can undergo dissipative and active forces. These active forces are typically applied by programmable electromagnetic actuators [7, 3, 9, 8, 10, 11, 4, 2, 1, 12, 13], but they can also naturally emerge in dense phases of active colloidal systems [14], of driven suspensions [15, 16, 17], and of biological materials [18] via hydrodynamic interactions. This mechanical system can be described by Newton's equation of motion of particle i :

$$m \frac{\partial^2 \mathbf{u}_i}{\partial t^2} = \mathbf{F}_i^{\text{damping}}(\mathbf{v}) + \mathbf{F}_i^{\text{elastic}}(\mathbf{u}) + \mathbf{F}_i^{\text{active}}(t, \mathbf{u}, \mathbf{v}), \quad (38.1)$$

where all nodes are assumed to have the same mass m , where \mathbf{u}_i and \mathbf{v}_i are the instantaneous displacement and velocity of node i , $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i, \dots)$, $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots)$, and $\mathbf{F}_i^{\text{damping}}$, $\mathbf{F}_i^{\text{elastic}}$, and $\mathbf{F}_i^{\text{active}}$ are the damping, elastic, and active forces.

In this Chapter, we restrict ourselves to the simplest settings where elastic and viscous forces are involved in separable terms. We leave out more intricate viscoelastic settings, which are treated in recent articles [19, 20, 21]. In the limit where all forces are linear in the displacements and velocities, Eq. (38.1) becomes

$$\frac{\partial^2 \mathbf{u}_i}{\partial t^2} = \underbrace{(-\boldsymbol{\Gamma}_{ij}\mathbf{v}_j)}_{\text{damping}} + \underbrace{(-\mathbf{D}_{ij}\mathbf{u}_j)}_{\text{elastic}} + \underbrace{\mathbf{B}_i(t) + \mathbf{D}'_{ij}(t)\mathbf{u}_j + \boldsymbol{\Gamma}'_{ij}(t)\mathbf{v}_j}_{\text{active}}, \quad (38.2)$$

where \mathbf{D}_{ij} is the *dynamical matrix* and $\boldsymbol{\Gamma}_{ij}$ is the *damping matrix*. These are the standard matrices that describe the linear response of passive materials. More specifically, they encode the relation between the displacements and velocities of the nodes and the net forces on the nodes. By virtue of reciprocity, \mathbf{D}_{ij} is necessarily symmetrical [22, 23, 24]. In the following, we will focus on purely elastic settings and therefore consider cases where $\boldsymbol{\Gamma}_{ij}$ is either zero or simply diagonal. In the former case, the dynamic will be purely inertial, while, in the latter case, there is additional viscous drag.

Furthermore, $\mathbf{D}'_{ij}(t)$ is the *active stiffness matrix* and $\boldsymbol{\Gamma}'_{ij}(t)$ is the *active damping matrix*. They encode how the active forces on node i depend on the displacements and velocities of the nodes. Crucially, they need not be symmetrical and, in fact, we will consider below examples where $\mathbf{D}'_{ij}(t)$ is non-symmetrical by design.

Finally, $\mathbf{B}_i(t)$ is an *active force* that may depend explicitly on the time variable t and may be coupled to additional degrees of freedom such as electromagnetic field, chemical potential, temperature, and local polarization. Examples of works treating the case $\mathbf{B}_i(t) \neq 0$ include the case of active Brownian particles [25, 26, 27, 5] or active Ornstein-Uhlenbeck particles [28] on elastic lattices and of certain types of time-modulated metamaterials [3, 12]. We will primarily focus on the cases where such force $\mathbf{B}_i(t)$ is modulated in time and where the active forces depend explicitly on the displacement \mathbf{u}_i or velocity \mathbf{v}_i . In this latter case, there is a well-defined linear response about the equilibrium configuration $\mathbf{u}_i = \mathbf{v}_i = 0$, which allows to study the linear regime.

38.2 Examples of active mechanical metamaterials

In this section, we give a few examples of active metamaterials with designer elastic and/or active forces.

38.2.1 One-dimensional odd chain

Consider the one-dimensional chain of odd springs [8] with lattice spacing $a = 1$ sketched in Fig. 38.2. The elastic forces are given simply by a chain of linear springs, where we assume no damping and we design the active forces so that they bias each spring asymmetrically. In other words, Newton's third law is broken by design. In practice, one breaks this symmetry by injecting linear momentum. This injected linear momentum will do work and thereby the odd chain is active.

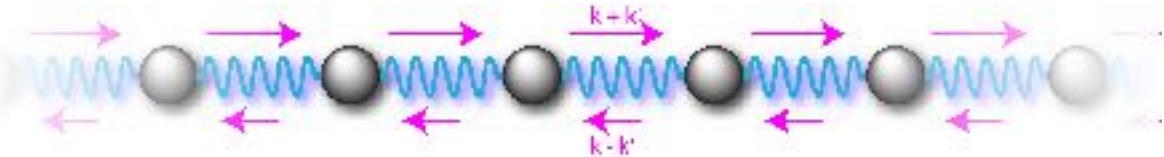


Figure 38.2: **Sketch of an odd chain.** A one-dimensional chain of masses connected by odd springs. The springs have a different stiffness depending on the direction in which they apply the force: If the strain of the spring is ε , they exert a force $(k + k')\varepsilon$ on the right node, but a force $-(k - k')\varepsilon$ on the left node.

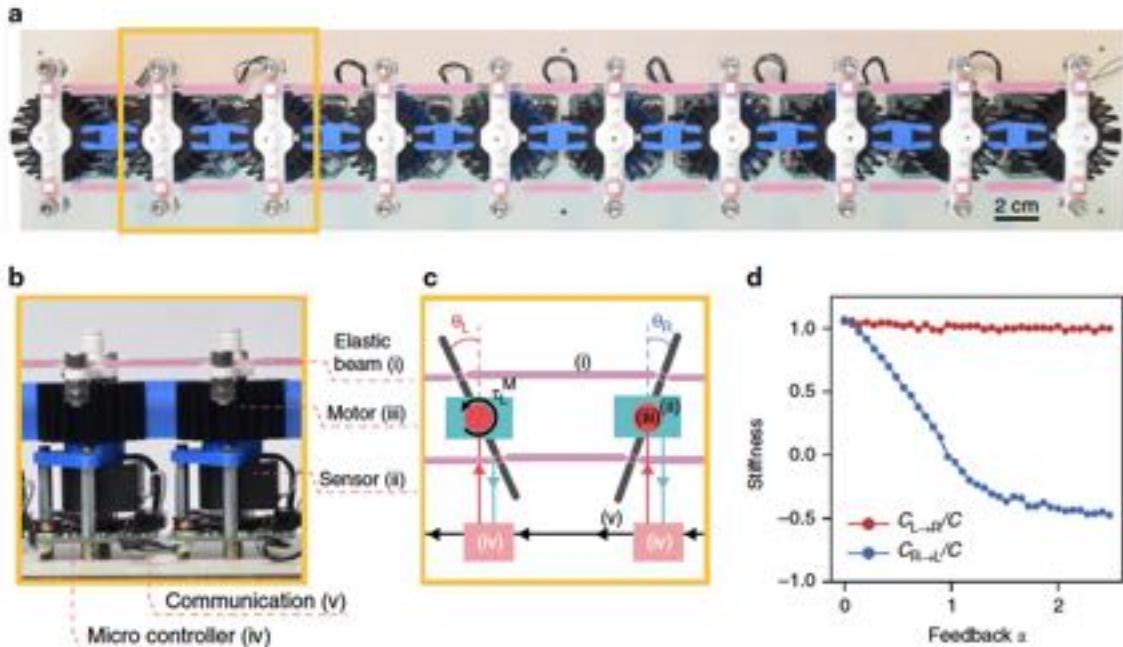


Figure 38.3: **Experimental implementation of an odd chain.** (a) A one-dimensional chain of pendula (white bars with metallic screws) connected by elastic rubber bands (pink). (b-c) Each pendulum is connected to a mechatronic system consisting of a DC coreless motor, an angular encoder, and a microcontroller. The microcontroller collects data from the sensor, exchanges information with its right neighbor, and controls the motor to inject a torque $\tau = \alpha C\varepsilon$, where C is the angular stiffness of the passive elastic links and where ε is the strain between each pendulum and its right neighbor. (d) Measurement of the forward and backward angular stiffnesses $C_{L \rightarrow R}$ and $C_{R \rightarrow L}$. As the non-reciprocal gain is increased, the forward stiffness remains constant, while the backward stiffness decreases. Adapted from Ref. [8].

The odd chain is described by the following equations of motion

$$m \frac{\partial^2 u_i}{\partial t^2} = \underbrace{-k(2u_i - u_{i-1} - u_{i+1})}_{\text{elastic}} \underbrace{-k'(u_{i-1} - u_{i+1})}_{\text{active}}, \quad (38.3)$$

where k is the stiffness of each spring, and k' the active stiffness. The active force renders the springs “odd”: They have a different stiffness seen from the right than from the left. Equivalently, the combination of the elastic and active forces make up an effective dynamical matrix

$$\mathbf{D}_{ij}^{\text{eff}} = \underbrace{\mathbf{D}_{ij}}_{\text{elastic}} + \underbrace{\mathbf{D}'_{ij}}_{\text{active}} = 2k\delta_{ij} - (k+k')\delta_{i,j+1} - (k-k')\delta_{i,j-1},$$

that is not symmetrical, i.e.,

$$\mathbf{D}_{ij}^{\text{eff}} \neq \mathbf{D}_{ji}^{\text{eff}}.$$

Importantly, the active forces inject linear momentum and hence energy into the mechanical degrees of freedom u_i . As we will see below, this odd injection of momentum has dramatic consequence on the wave propagation properties. In practice, such active metamaterial can be realized with a combination of elastic links, motors, sensors and microcontrollers, as shown in Fig. 38.3.

38.2.2 One dimensional chain with time-modulated stiffness

Consider the one-dimensional chain of springs with time-modulated grounding stiffness shown in Fig. 38.4 [7] — see also [29, 6, 9] for related studies. The elastic forces are given simply by a chain of linear springs, where

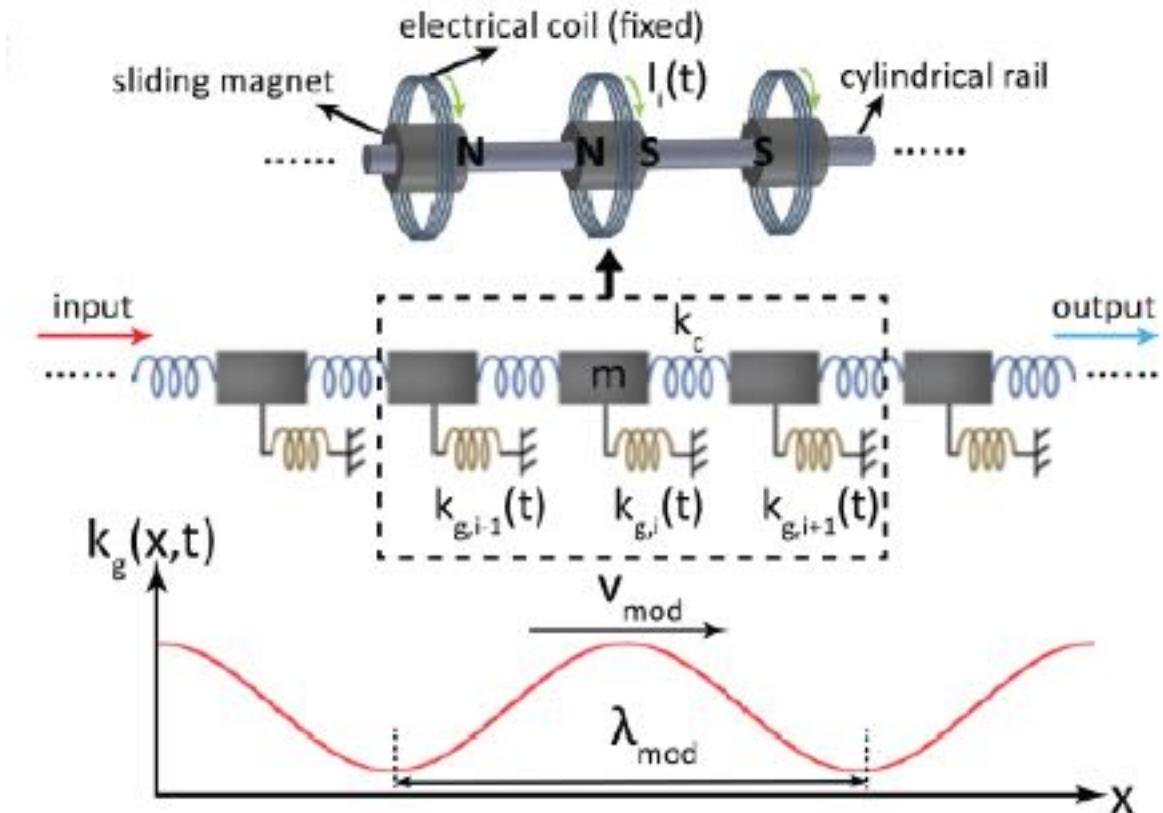


Figure 38.4: **Time modulated metamaterial.** (a) Sketch of the experimental setup. (b) Sketch of the mechanical model. (c) Illustration of the modulation of the stiffness of the grounding spring. Adapted from Ref. [7]. Agnese: Change top, middle, bottom to a, b, c

we assume no damping and we design the active forces so that they apply a spatio-temporal modulation of the springs that connect each mass to the ground. This spatio-temporal modulation travels in a specific direction, e.g., the forward direction, and will interact with the incident elastic waves in the material so that the incident waves whose frequency and wave vector coincide with that of the modulation will resonate and be scattered by the modulation. Since the modulation is unidirectional, the forward and backward incident waves will be scattered differently and leave to non-reciprocal transmission of the incident wave.

Formally, the time-modulated chain is described by the following equations of motion

$$m \frac{\partial^2 u_i}{\partial t^2} = \underbrace{-k(2u_i - u_{i-1} - u_{i+1})}_{\text{elastic}} - k_0 u_i - \underbrace{k' \cos(\omega_{\text{mod}} t + 2\pi i / \lambda) u_i}_{\text{active}}, \quad (38.4)$$

where k is the stiffness of each spring, k_0 is a constant grounding stiffness, and k' the modulated grounding stiffness, Ω is the frequency of the modulation and λ is the wavelength of the modulation. The temporal modulation is spatially patterned thanks to a phase shift of $q_{\text{mod}} = 2\pi/\lambda$, which, we will see later in this Chapter, allows to obtain unidirectional wave propagation.

38.2.3 Metamaterial with time-modulated rest-length that crawls

Consider the one-dimensional chain of springs with time-modulated rest length [12] shown in Fig. 38.5. On its own, each building block cannot move, but when connected to a few neighbors, it can push them off and make them slide, and/or push itself off its neighbors and slide itself. What determines whether it moves and the direction in which it moves is the relative phase of neighboring units.

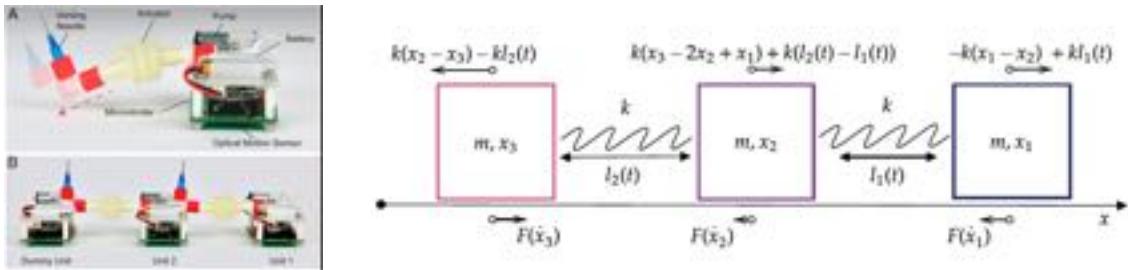


Figure 38.5: **Crawling time-modulated metamaterial.** (a) Pictures of the experimental setup, which consist of active building block in contact with a frictional surface, dressed with pneumatic actuator, whose rest-length can be modulated in time via an embedded pump. (b) Sketch of the mechanical problem. Adapted from Ref. [12]. [Missing panel labels]

Mathematically, the active metamaterial can be described as follows. The elastic forces are given simply by a chain of linear springs, where we assume frictional contact with a surface and we design the active forces so that they apply a temporal modulation of the rest-length of each spring at a instantaneous phase $\phi_i(t)$. Such a chain obeys the following equations of motion

$$m \frac{\partial^2 u_i}{\partial t^2} = -\mu \operatorname{sign}\left(\frac{\partial u_i}{\partial t}\right) \underbrace{-k(2u_i - u_{i-1} - u_{i+1})}_{\text{elastic}} - k\delta\ell(\cos(\phi_{i+1}(t)) - \cos(\phi_i(t))), \quad (38.5)$$

where μ is the friction coefficient, k is the stiffness of each spring, and $\delta\ell$ the stroke of the actuator. In recent articles (such as Refs. [3, 12]), the dynamics is overdamped so one can neglect the left-hand side of the above equation and $\phi_i(t)$ takes the form $\phi_i(t) = \cos \omega_{\text{mod}} t + \psi_i$, where ψ_i is a trainable parameter that is optimized such that the active chain locomotes the fastest. Interestingly, the resulting dynamics are often travelling waves of the form $\psi_i = 2\pi i / \lambda$, where λ is the wavelength, similar to the example above.

38.2.4 Odd elastic lattice

Consider the triangular lattice of chiral odd springs [4] shown in Fig. 38.6. They are odd, but in a different way than the odd springs in the example above. While the odd springs above were directly violating conservation

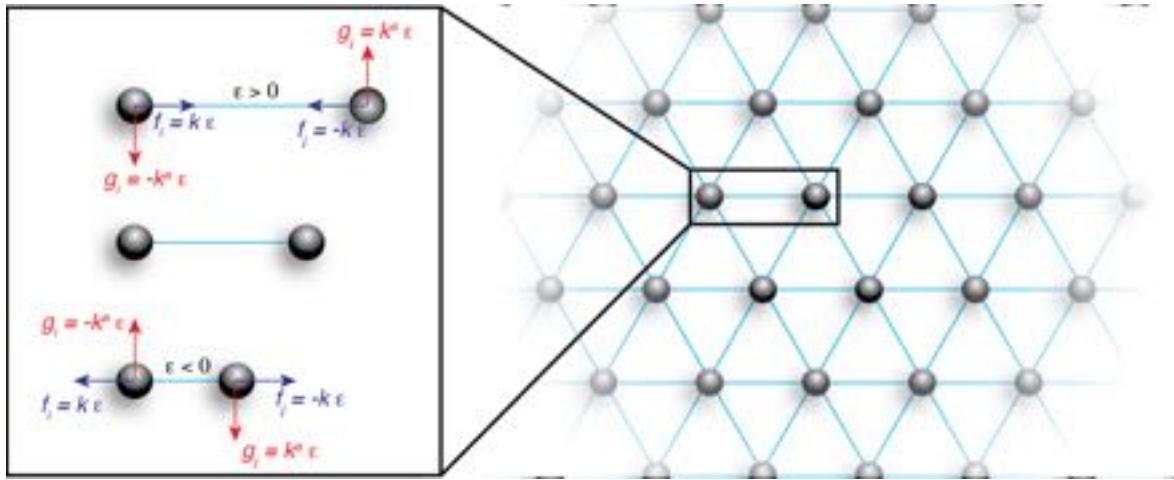


Figure 38.6: **Odd elastic lattice.** Triangular lattice made of chiral odd springs. When stretched, the springs experience a counter-clockwise torque and, when compressed, the chiral odd springs experience a clockwise torque. The net torque is zero if the springs are undeformed.

of linear momentum, the present chiral odd springs conserve *linear* momentum, yet they violate conservation of *angular* momentum. The elastic part of these springs apply a longitudinal force that is proportional to the strain of the spring, equal and opposite on either side of each spring, while the active part applies a transverse force that is proportional to the strain of the spring. Such transverse force applies a net torque.

The inertial dynamics in such a lattice is described by the following equation:

$$m \frac{\partial^2 \mathbf{u}_i}{\partial t^2} = \underbrace{k \sum_{j \in \partial i} (\overbrace{(|\mathbf{r}_i - \mathbf{r}_j| - \ell)}^{\text{spring elongation } \varepsilon} \mathbf{e}_{ij})}_{\text{elastic}} - \underbrace{k^a \sum_{j \in \partial i} (\overbrace{(|\mathbf{r}_i - \mathbf{r}_j| - \ell)}^{\text{spring elongation } \varepsilon} \mathbf{e}_{ij}^*)}_{\text{active}}, \quad (38.6)$$

where k is the stiffness of each spring, k^a is the odd stiffness, \mathbf{r}_i is the position of particle i , \mathbf{e}_{ij} is the unitary bond vector between particle i and j , \mathbf{e}_{ij}^* is the unit vector that is perpendicular to \mathbf{e}_{ij} , and ∂i is the set of particles that are connected to particle i . In the linear regime, when the deformation of the spring remains small, one can approximate the right-hand side as

$$m \frac{\partial^2 \mathbf{u}_i}{\partial t^2} = -\mathbf{D}_{ij}^{\text{eff}} \mathbf{u}_j, \quad (38.7)$$

where $\mathbf{D}_{ij}^{\text{eff}}$ is the effective dynamical matrix that is not symmetrical because it contains the contribution of the active term. While in the 1D case, it was straightforward to write the matrix $\mathbf{D}_{ij}^{\text{eff}}$, in 2D the geometry of the bonds is a bit more complicated. It is in fact an interesting exercise to write down such matrix by hand at least once, see the example below and more generally Ref. [24] for a pedagogical review on how to do this. In practical cases, we use a computer to build such matrix.

Example 38.1: Dynamical matrix of an odd triangular lattice. Consider the particle (i, j) in the lattice described above. It is surrounded by six nearest-neighbors. We denote the horizontal and vertical components of the displacement of particle (i, j) , $u_{i,j}$ and $v_{i,j}$; the horizontal and vertical components of the elastic force on particle (i, j) , $f_{i,j}$ and $g_{i,j}$; and the horizontal and vertical components of the active force on particle (i, j) , $f_{i,j}^a$ and $g_{i,j}^a$. See Fig. 38.7. The vector that quantifies the elongation in each spring connected to particle (i, j) , $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)$, is related to the displacement vector of particle (i, j)

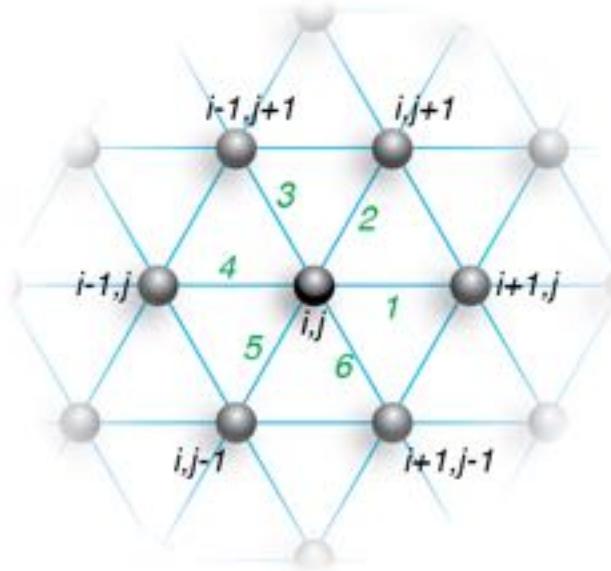


Figure 38.7: **Odd elastic lattice.** Triangular lattice made of chiral odd springs. When stretched, the springs experience a counter-clockwise torque and, when compressed, the chiral odd springs experience a clockwise torque. The net torque is zero if the spring are undeformed. [Agnese: you can add the information in this picture on the right side of Fig. 34.6. Then, please update also the references in the text.](#)

and its six nearest neighbors, i.e.,

$$\mathbf{u} = (u_{i,j}, v_{i,j}, u_{i+1,j}, v_{i+1,j}, u_{i,j+1}, v_{i,j+1}, u_{i-1,j+1}, v_{i-1,j+1}, u_{i-1,j}, v_{i-1,j}, u_{i,j-1}, v_{i,j-1}, u_{i+1,j-1}, v_{i+1,j-1}), \quad (38.8)$$

as $\boldsymbol{\varepsilon} = \mathbf{C}\mathbf{u}$, where

$$\mathbf{C} = \begin{pmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}. \quad (38.9)$$

The vector that quantifies the elastic forces in particle (i, j) and its six nearest neighbors, i.e.,

$$\mathbf{f} = \underbrace{(f_{i,j}, g_{i,j}, f_{i+1,j}, g_{i+1,j}, f_{i,j+1}, g_{i,j+1}, f_{i-1,j+1}, g_{i-1,j+1}, f_{i-1,j}, g_{i-1,j}, f_{i,j-1}, g_{i,j-1}, f_{i+1,j-1}, g_{i+1,j-1})}_{\text{elastic}}, \quad (38.10)$$

is related to the tension in each spring connected to particle (i, j) , $\mathbf{t} = (t_1, t_2, t_3, t_4, t_5, t_6)$, as $\mathbf{f} = -\mathbf{C}^T \mathbf{t}$. The vector that quantifies the active forces in particle (i, j) and its six nearest neighbors, i.e.,

$$\mathbf{f}^a = \underbrace{(f_{i,j}^a, g_{i,j}^a, f_{i+1,j}^a, g_{i+1,j}^a, f_{i,j+1}^a, g_{i,j+1}^a, f_{i-1,j+1}^a, g_{i-1,j+1}^a, f_{i-1,j}^a, g_{i-1,j}^a, f_{i,j-1}^a, g_{i,j-1}^a, f_{i+1,j-1}^a, g_{i+1,j-1}^a)}_{\text{active}}, \quad (38.11)$$

is related to the tension in each spring connected to particle (i, j) , $\mathbf{t} = (t_1, t_2, t_3, t_4, t_5, t_6)$, as $\mathbf{f} = \mathbf{C}' \mathbf{t} k^a / k$,

where $\mathbf{C}' = \mathbf{RC}^T$, where \mathbf{R} is a block diagonal matrix consisting of repeating 2×2 blocks $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ that rotate 2D vectors by $+\pi/2$. The dynamical matrix relates the force vectors to the displacement vectors as $\mathbf{f} = -\mathbf{Du}$. Given that the tension in the springs is readily related to their elongation by the relation $\mathbf{t} = k\mathbf{e}$, where k is the stiffness of each spring, show that

$$\mathbf{D} = \underbrace{k\mathbf{C}^T}_{\text{elastic}} \mathbf{C} - \underbrace{k^a \mathbf{RC}^T}_{\text{active}} \mathbf{C}$$

and specifically that

$$m \frac{\partial^2 u_{i,j}}{\partial t^2} = F_{\text{elastic}}^{(u)} + F_{\text{active}}^{(u)} \quad (38.12)$$

$$m \frac{\partial^2 v_{i,j}}{\partial t^2} = F_{\text{elastic}}^{(v)} + F_{\text{active}}^{(v)} \quad (38.13)$$

with

$$\begin{aligned} F_{\text{elastic}}^{(u)} &= k \left(3u_{i,j} - u_{i+1,j} - u_{i-1,j} - \frac{1}{4}(u_{i,j+1} + u_{i-1,j+1} + u_{i,j-1} + u_{i+1,j-1}) - \frac{\sqrt{3}}{4}(v_{i,j+1} - v_{i-1,j+1} + v_{i,j-1} - v_{i+1,j-1}) \right) \\ F_{\text{active}}^{(u)} &= -k^a \left(\frac{\sqrt{3}}{4}(-u_{i,j+1} + u_{i-1,j+1} - u_{i,j-1} + u_{i+1,j-1}) - 3v_{i,j} + \frac{3}{4}(v_{i,j+1} + v_{i-1,j+1} + v_{i,j-1} + v_{i+1,j-1}) \right) \end{aligned}$$

and

$$\begin{aligned} F_{\text{elastic}}^{(v)} &= k \left(3v_{i,j} - \frac{3}{4}(v_{i,j+1} + v_{i-1,j+1} + v_{i,j-1} + v_{i+1,j-1}) - \frac{\sqrt{3}}{4}(u_{i,j+1} - u_{i-1,j+1} + u_{i,j-1} - u_{i+1,j-1}) \right) \\ F_{\text{active}}^{(v)} &= k^a \left(\frac{\sqrt{3}}{4}(-v_{i,j+1} + v_{i-1,j+1} - v_{i,j-1} + v_{i+1,j-1}) + 3u_{i,j} - u_{i+1,j} - u_{i-1,j} - \frac{1}{4}(u_{i,j+1} + u_{i-1,j+1} + u_{i,j-1} + u_{i+1,j-1}) \right). \end{aligned}$$

Agnese: divide the equations into two lines above so that they don't get out of the box.

38.3 Properties of active mechanical metamaterials

In the previous section, we have surveyed a few examples of active mechanical metamaterials. We now have to discover which properties emerge from the collective interaction of these tunable active elements. To do this, we will focus on two typical approaches (i) *spectral analysis* and (ii) the *continuum limit*. We will survey again the few examples we covered in the previous section and discuss the theoretical tools and the various phenomena that can occur.

38.3.1 Spectral analysis

In the case where active metamaterials are periodic, we can analyze their spectral properties by using the Floquet-Bloch theorem. To this end, we inject the ansatz of plane waves into the equations of motion,

$$\mathbf{u}_i = \mathbf{u}_i^0 \exp i(\omega t - \mathbf{q} \cdot \mathbf{b}), \quad (38.14)$$

where ω is the complex frequency, \mathbf{q} the wave vector, and \mathbf{b} the Bravais lattice vector. Hence, the displacement \mathbf{u}_i is assumed to be a plane wave. The real part of ω denotes the frequency of this wave, whereas its imaginary part denotes the rate of amplification (if $\text{Im } \omega > 0$) or of attenuation (if $\text{Im } \omega < 0$) of the plane wave. Likewise, the real part of $\mathbf{q} \cdot \mathbf{b}$ denotes the mode number of this wave, whereas the imaginary part of $\mathbf{q} \cdot \mathbf{b}$ denotes its spatial amplification (if $\text{Im } \mathbf{q} \cdot \mathbf{b} > 0$) or decay (if $\text{Im } \mathbf{q} \cdot \mathbf{b} < 0$). Generically, one then obtains the equation

$$0 = \sum_i (\omega^2 \mathbf{I} - \tilde{\mathbf{D}}^{\text{eff}}(\mathbf{q})) \mathbf{u}_i^0 \exp i(\omega t - \mathbf{q} \cdot \mathbf{b}), \quad (38.15)$$

where $\tilde{\mathbf{D}}^{\text{eff}}(\mathbf{q})$ is the Fourier transform of the dynamical matrix. The equation above shows that the vibration spectrum can be obtained by diagonalizing the matrix $\tilde{\mathbf{D}}^{\text{eff}}(\mathbf{q})$. In contrast with the dynamical matrix in real space, which was of size $N \times N$, in reciprocal space the size of the dynamical matrix is $D \times D$, where D is the number of degrees of freedom in each unit cell. For example, in simple one dimensional chains, $D = 1$, and in the two-dimensional triangular lattice, $D = 2$.

The non-Hermitian skin effect, a hallmark of odd matter

In the one-dimensional case, the plane wave ansatz becomes $u_i = u_i^0 \exp i(\omega t - qa)$, which once injected into Eq. (38.3) gives the dispersion relation

$$m\omega^2 = \underbrace{2k(1 - \cos qa)}_{\text{elastic}} + \underbrace{2ik' \sin qa}_{\text{active}}. \quad (38.16)$$

Insights can be gained by considering the limit for large wavelengths $q \rightarrow 0$ — or equivalently the continuum limit, which we will treat below. In this case, Eq. (38.16) becomes

$$m\omega^2 = \underbrace{kq^2 a^2}_{\text{elastic}} + \underbrace{2ik' qa}_{\text{active}}. \quad (38.17)$$

We can investigate this dispersion relation in two ways: One can (i) analyze the fate of waves on which we impose a given real frequency ω or (ii) on which we impose a real wave vector q . When we impose the frequency to be real (as if we were to excite the system with a harmonic drive of frequency ω), we find $q = i/(a)(k'/k \pm \sqrt{k'/k - (m/k)\omega^2})$. The imaginary part of q is always of the same sign as k' , regardless of the frequency ω . In other words, for positive k' , all the waves are localized on the right, as shown in Fig. 38.8a for negative k' .

Equivalently, when we impose the wave vectors to be real, we find that

$$\omega(q) = \pm \sqrt{\frac{k}{m} \sqrt{q^2 a^2 + 2ik'/kqa}}, \quad (38.18)$$

which is plotted in Fig. 38.8b. From this dispersion relation, one can readily obtain the group velocity

$$v_g = \frac{\partial \omega}{\partial k} = a \sqrt{\frac{k}{m}} \frac{1}{\sqrt{qa}} \frac{qa + ik'/k}{\sqrt{qa + 2ik'/k}}, \quad (38.19)$$

which is plotted in Fig. 38.8c. Crucially, for $k' > 0$, whenever wave packets propagate forward ($\text{Re } v_g > 0$), the wave amplifies ($\text{Im } \omega > 0$), and whenever wave packets propagate backward ($\text{Re } v_g < 0$), the wave attenuates ($\text{Im } \omega < 0$). The situation is reversed for $k' < 0$.

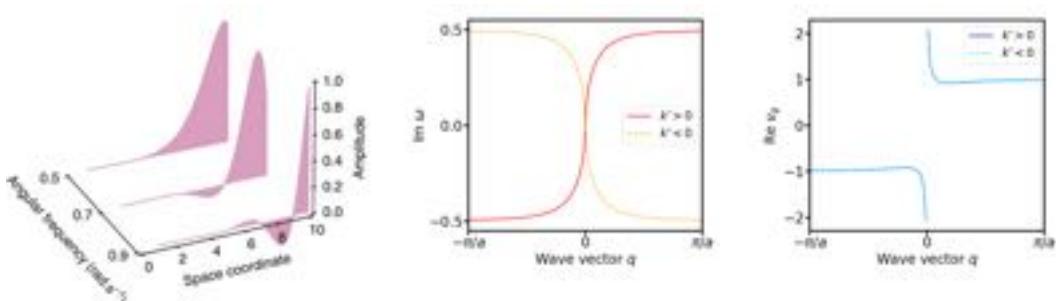


Figure 38.8: **Non-Hermitian skin modes.** (a) Eigenmodes of the form $u_i = u_i^0 \exp i(\omega t - qa)$ with $q = i/(a)(k'/k \pm \sqrt{k'/k - (m/k)\omega^2})$ as a function of frequency ω . Adapted from Ref. [8]. (b) Imaginary part of the frequency $\omega(q)$ vs. wave vector q (relation in Eq. (38.18)). (c) Group velocity $\text{Re } v_g$ vs wave vector q (relation in Eq. (38.19)). Agnese: Add letters to figure. Also, from where are panels b and c taken?

In short, the active forces amplify waves in one direction and attenuate them in the other. This will have the consequence that all vibrational modes are asymmetric and skewed in one direction or, equivalently, that wave packets will be amplified in one direction and attenuated in the other. This is a manifestation of the non-Hermitian skin effect, which was first discovered theoretically in the context of quantum mechanics [30, 31, 32], first observed in active mechanical metamaterials [8, 11, 2] and later in electronics [33], quantum optics [34], and photonics [35].

Dispersion of time-modulated chains

In a time-modulated metamaterials [7], one is interested in waves of the form $u_i = u_i^0 \exp i(\omega t - qai)$, which are governed by the following dispersion relation

$$\omega^2 = \underbrace{k_0 + 2k(1 - \cos qa)}_{\text{elastic}}, \quad (38.20)$$

as shown by the red curve in Fig. 38.9a, where the grounding stiffness k_0 was assumed to be much smaller than k . This relation comes from Eq. (38.4). Those waves will interact with the stiffness time-modulation, which is of the form $\exp i(\omega_{\text{mod}}t - q_{\text{mod}}i)$. Such modulation will create a scattered field $u_i^s = u_i^s \exp i(\omega^s t - q^s i)$, where $\omega^s = \omega \pm \omega_{\text{mod}}$ and $q^s = q \pm q_{\text{mod}}$, as shown by the dotted and dashed curves in Fig. 38.9a. In the limiting case where the amplitude of the modulation is small compared with the stiffness of the springs that couple neighboring nodes, i.e., $k' \ll k$, the scattered field will be negligible, except when it resonates with the incident wave, namely, when the dispersion of the scattered wave crosses that of the incident wave, as shown by the circles in Fig. 38.9a. Since the location of these points is not symmetrical, the resonant scattered modes that will ultimately lead to the wave transmission are not reciprocal and the metamaterials will transmit waves more in one direction than in the other, as shown by the experimental data in Fig. 38.9b. Note that one can also derive a similar result by deriving the coupled modes, which then will predict how the dispersion diagrams are affected [6].

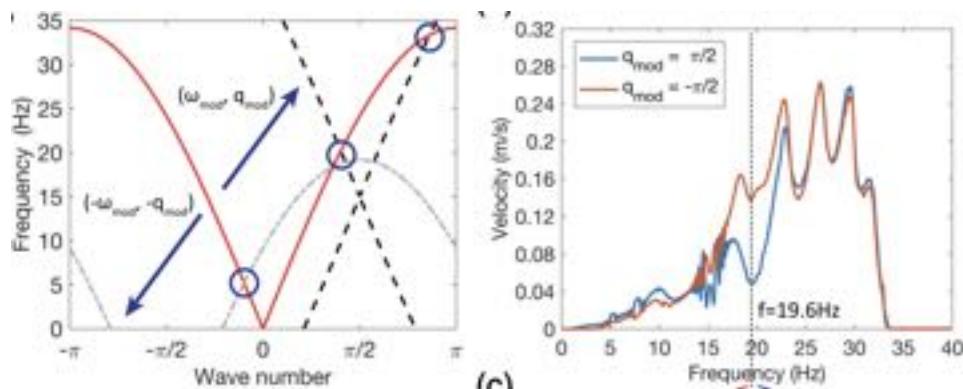


Figure 38.9: **Dispersion and transmission spectrum of a time-modulated metamaterial.** (a) Dispersion relation: The red solid curve describes the original dispersion relation of the unmodulated monatomic lattice. The black dashed and grey dotted-dashed curves describe replicas of the original dispersion relation shifted by the spatio-temporal modulation $\pm(q_{\text{mod}}, \omega_{\text{mod}}) = \pm(\pi/2, 15\text{Hz})$. Crossings between the original and shifted curves are denoted by blue circles and indicate the points where the incident wave is expected to resonate with the modulation. Since these crossings break the $q \rightarrow -q$ symmetry, forward and backward incident waves will scatter differently with the modulation and non-reciprocal wave transmission is expected. (b) Experimental measurement of the transmission spectrum of a metamaterial consisting of 12 nodes. The second node is driven with a sinusoidal force with a varying frequency and the velocity of node 11 is recorded. The spatio-temporal modulation is set to 15Hz with a wave vector $\pm\pi/2$. Adapted from Ref. [7]. Agnese: Reformat this figure. See with author whether also panel c should be included.

Odd elastic lattice

Let us consider again the odd triangular lattice considered in the example above (Example 38.1) and perform a spectral analysis. What we will do here again is to exploit the fact that the lattice is invariant by translation. Therefore, the displacement vectors can be written as $u_{i\pm 1,j} = u_{i,j} \exp i\omega t \mp iq_1$, $u_{i,j\pm 1} = u_{i,j} \exp i\omega t \mp i(\frac{1}{2}q_1 + \frac{\sqrt{3}}{2}q_2)$ and $u_{i\pm 1,j\mp 1} = u_{i,j} \exp i\omega t \mp i(\frac{1}{2}q_1 - \frac{\sqrt{3}}{2}q_2)$. By plugging these expressions into Eqs. (38.12) and (38.13), one finds

$$\mathbf{D}^{\text{eff}}(\mathbf{q}) = k \begin{pmatrix} -\cos(\frac{q_1}{2}) \cos(\frac{\sqrt{3}}{2}q_2) - 2\cos(q_1) + 3 & \sqrt{3}\sin(\frac{q_1}{2}) \sin(\frac{\sqrt{3}}{2}q_2) \\ \sqrt{3}\sin(\frac{q_1}{2}) \sin(\frac{\sqrt{3}}{2}q_2) & 3 - 3\cos(\frac{q_1}{2}) \cos(\frac{\sqrt{3}}{2}q_2) \end{pmatrix} + k^a \begin{pmatrix} -\sqrt{3}\sin(\frac{q_1}{2}) \sin(\frac{\sqrt{3}q_2}{2}) & -3 + 3\cos(\frac{q_1}{2}) \cos(\frac{\sqrt{3}q_2}{2}) \\ 3 - \cos(\frac{q_1}{2}) \cos(\frac{\sqrt{3}q_2}{2}) - 2\cos(q_1) & \sqrt{3}\sin(\frac{q_1}{2}) \sin(\frac{\sqrt{3}q_2}{2}) \end{pmatrix} \quad (38.21)$$

One can readily see that this dynamical matrix is not symmetrical because of the term in front of the odd term k^a . As a result, its spectrum will be complex and therefore some waves will be amplified while others will be attenuated. This spectrum can be obtained by the dispersion relation $0 = \det D^{\text{eff}}(q) - \omega^2 I_2$. **Check this equation. Should we have the symbols in bold?** In Fig. 38.10, we plot $\Re(\omega)$ and $\Im(\omega)$ vs. $\mathbf{q} = (q_1, q_2)$. We see that there is a range of wave vectors where this imaginary part is positive. This means that the wave will be amplified.

One can also consider overdamped dynamics instead of inertial dynamics. To do so, one has to solve $0 = \det D^{\text{eff}}(q) + i\omega I_2$. **Check this equation. Should we have the symbols in bold?** Interestingly, one finds that,

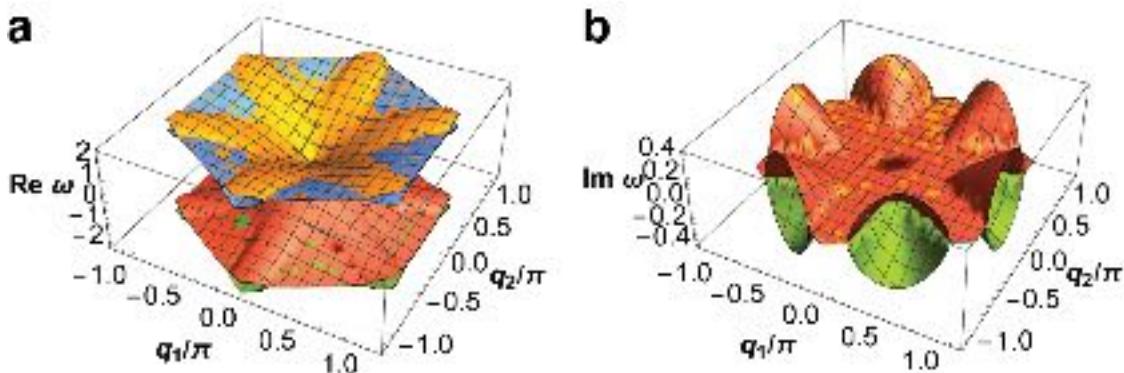


Figure 38.10: **Dispersion of an odd triangular lattice in the inertial case.** (a) Real and (b) imaginary part of the frequency vs. wave vector $\mathbf{q} = (q_1, q_2)$.

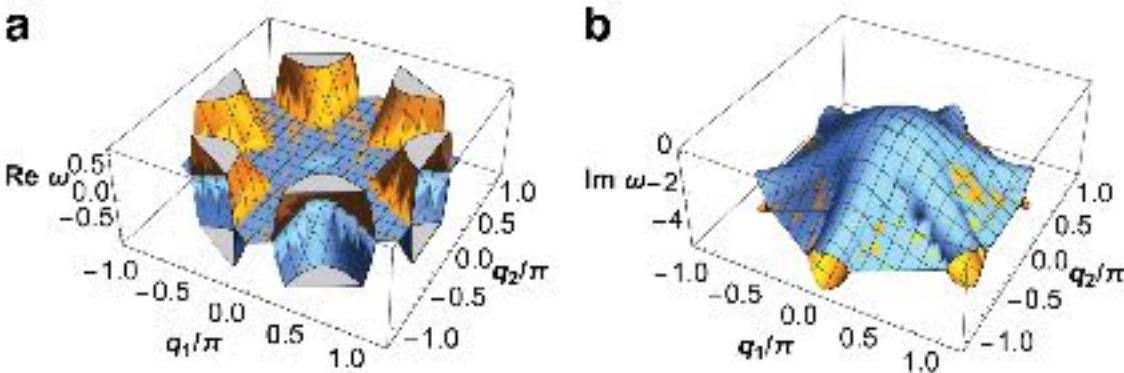


Figure 38.11: **Dispersion of an odd triangular lattice in the overdamped case.** (a) Real and (b) imaginary part of the frequency vs. wave vector $\mathbf{q} = (q_1, q_2)$.

even though the dynamics is in principle overdamped, one can still observe (attenuated) waves, as depicted in the spectrum in Fig. 38.11 — we see that there is a range of wave vectors for which the real part is finite, while the imaginary part is always negative. Physically, these waves exist because of the activity, which couples the two eigenmodes non-reciprocally.

38.3.2 Continuum limit

In the continuum limit, the elastic, viscous, and active forces depend on \mathbf{u}_i , \mathbf{v}_i , and their derivatives. Thus, up to second order in spatial derivative, Eq. (38.24) becomes

$$\frac{\partial^2 \mathbf{u}_i}{\partial t^2} = \underbrace{-a_{ij}\mathbf{u}_j - c_{ijk}\partial_{jk}\mathbf{u}_\ell}_{\text{elastic}} \underbrace{-d_{ij}\mathbf{v}_j - f_{ijk}\partial_{jk}\mathbf{v}_\ell}_{\text{damping}} \underbrace{-a'_{ij}\mathbf{u}_j - b'_{ijk}\partial_j\mathbf{u}_k - c'_{ijk}\partial_{jk}\mathbf{u}_\ell}_{\text{active}} \underbrace{-d'_{ij}\mathbf{v}_j - e'_{ijk}\partial_j\mathbf{v}_k - f'_{ijk}\partial_{jk}\mathbf{v}_\ell}_{\text{active}}. \quad (38.22)$$

Among the passive terms, a_{ij} (d_{ij}) corresponds to elastic (viscous) pinning to a substrate and c_{ijk} (f_{ijk}) corresponds to internal elastic (viscous) forces. The former may inject or take linear momentum away from the medium, while the latter conserves linear momentum. Among the active terms, a'_{ij} (d'_{ij}) and b'_{ijk} (e'_{ijk}) do not convert linear momentum while c'_{ijk} (f'_{ijk}) do. This distinction may appear mundane at first sight, but it has profound consequences on the type of material one has to construct to realize them.

Example 38.2: 1D odd chain. Consider again, the one-dimensional chain of odd springs [8], with lattice spacing $a = 1$. Under the continuum limit approximation, Eq. (38.3) then becomes

$$m \frac{\partial^2 u_i}{\partial t^2} = \underbrace{k \partial_{ii} u_i}_{\text{elastic}} + \underbrace{2k' \partial_i u_i}_{\text{active}}. \quad (38.23)$$

This equation is the standard wave equation, plus an additional term proportional to k' that breaks the mirror symmetry. This term will in fact create waves that are exponentially skewed towards the positive direction for positive odd stiffness $k' > 0$, and towards the negative direction for negative odd stiffness $k' < 0$. This finding is equivalent to that discussed in the wave dispersion analysis above.

Example 38.3: 2D odd elasticity. Consider the two-dimensional triangular lattice made of the chiral odd springs described above. Since it is an isotropic elastic material, in the continuum limit, it is governed by the following equation of motion:

$$\rho \frac{\partial^2 \mathbf{u}_i}{\partial t^2} = \underbrace{-c_{ijkl}\partial_{jk}\mathbf{u}_\ell}_{\text{elastic}} - \underbrace{c'_{ijkl}\partial_{jk}\mathbf{u}_\ell}_{\text{active}}, \quad (38.24)$$

where the effective tensor $c'_{ijkl} = c_{ijkl} + c'_{ijkl}$ governs the elastic response. Crucially, while c_{ijkl} is constrained to obey the symmetries $c_{ijkl} = c_{klji}$, c'_{ijkl} is not because it comprises the active forces as well. It can be shown that the tensor takes the form [4]

$$C_{ijkl} = B\delta_{ij}\delta_{kl} + \mu \left(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl} - \frac{1}{2}\delta_{ij}\delta_{kl} \right) + K^o E_{ijkl} - A\varepsilon_{ij}\delta_{kl}, \quad (38.25)$$

where $E_{ijkl} = \frac{1}{2}[\varepsilon_{il}\delta_{jk} + \varepsilon_{ik}\delta_{jl} + \varepsilon_{jl}\delta_{ik} + \varepsilon_{jk}\delta_{il}]$, and δ_{ij} and ε_{ij} are the Kronecker and Levi-Civita symbols, respectively. Here, B and μ are the well known bulk and shear moduli. The moduli K^o and A are an

example of *odd elastic moduli*. One can write the stress and strain using a pictorial notation,

$$\textcircled{+} = \sigma_{11} + \sigma_{22} \quad \square = u_{11} + u_{22} \quad (38.26)$$

$$\textcircled{\times} = \sigma_{12} - \sigma_{21} \quad \textcircled{\triangle} = u_{12} - u_{21} \quad (38.27)$$

$$\textcircled{+}\textcircled{-} = \sigma_{11} - \sigma_{22} \quad \textcolor{red}{\square} = u_{11} - u_{22} \quad (38.28)$$

$$\textcircled{+}\textcircled{+} = \sigma_{12} + \sigma_{21} \quad \textcolor{red}{\square}\textcolor{red}{\square} = u_{12} + u_{21}. \quad (38.29)$$

Where the stresses $\textcircled{+}$, $\textcircled{\times}$, $\textcircled{+}\textcircled{-}$ and $\textcircled{+}\textcircled{+}$ correspond to pressure ($\textcircled{+}$), torque density ($\textcircled{\times}$), and pure shear stresses ($\textcircled{+}\textcircled{-}$, $\textcircled{+}\textcircled{+}$) respectively and where the strains \square , $\textcircled{\triangle}$, $\textcolor{red}{\square}$ and $\textcolor{red}{\square}\textcolor{red}{\square}$ correspond to dilation (\square), rotation ($\textcircled{\triangle}$), pure shear strains ($\textcolor{red}{\square}$, $\textcolor{red}{\square}\textcolor{red}{\square}$), respectively. In this notation, elasticity Eq. (38.25) then becomes

$$\begin{pmatrix} \textcircled{+} \\ \textcircled{\times} \\ \textcircled{+}\textcircled{-} \\ \textcircled{+}\textcircled{+} \end{pmatrix} = 2 \begin{pmatrix} B & 0 & 0 & 0 \\ A & 0 & 0 & 0 \\ 0 & 0 & \mu & K^o \\ 0 & 0 & -K^o & \mu \end{pmatrix} \begin{pmatrix} \square \\ \textcircled{\triangle} \\ \textcolor{red}{\square} \\ \textcolor{red}{\square}\textcolor{red}{\square} \end{pmatrix}. \quad (38.30)$$

In this notation, one can then readily see the effect of the odd elastic moduli on the elastic response, which then exhibit intriguing non-reciprocal effects. For instance, A converts a dilation into a net torque. However, applying a torque will not induce dilation. Similarly, K^o couples the two shear directions non-reciprocally. Compressing the material in one direction will induce shear, but shearing the material will induce extension instead. Finally, these moduli can generically be derived from the model of the underlying microscopic lattice, using the dynamical matrix in the limit $\mathbf{q} \rightarrow 0$. In the specific example of the triangular lattice, one finds $B = k\sqrt{3}/2$, $\mu = k\sqrt{3}/2$, $A = k^a\sqrt{3}/2$, and $K^o = k^a\sqrt{3}/4$ [4].

38.4 Final Remarks

We conclude this Chapter by giving some perspectives on the field of active mechanical metamaterials.

38.4.1 Difference between active mechanical metamaterials and active matter

An active mechanical metamaterial is active matter. However, there is a difference in perspective: Traditionally, active matter is a testing ground for statistical mechanics, where one investigates the nature of fluctuations, the formation of patterns, the emergence of non-equilibrium steady states, and so on. This is due to the fact that the activity, which typically originates from self-propulsion or active stresses, induces instabilities that are hard to avoid, so one has to study them. In stark contrast, in active mechanical metamaterials, one inherently *designs* the active forces. In the majority of the examples, these active forces are designed such that the metamaterials remains stable and that there is a well-defined linear response close to a well-defined equilibrium. The response then has unconventional properties, such as non-reciprocal response and locomotion.

38.4.2 Blurring active mechanical metamaterials and active matter

We argue here that blurring these two perspectives can in fact be beneficial in multiple ways. First, active mechanical metamaterials can further our conceptual understanding of active matter. For example, it was recently reported that the chiral patterns that emerge in rotating colloids [14] and in rotating star-fish embryos [18] (Fig 38.12a) can actually be interpreted as signatures of odd elasticity [4, 1] (Fig 38.12b). Likewise, recent observations of the emergence of chiral patterns in active solids [5] (Fig 38.12c) share some similarities with recently reported theoretical investigations on phase transition in non-reciprocally coupled agents, such as robots [36] (Fig 38.12d). Second, since active mechanical metamaterials are inherently designed and have a structural integrity with a well-defined reference configuration, one can tailor them to a specific application. Therefore, active mechanical metamaterials are a promising avenue to take active matter closer to actual applications. For example, the concept of non-reciprocal wave transmission could lead to novel concepts for applications in wave-guiding and robotic locomotion. Finally, the impressive array of techniques developed to

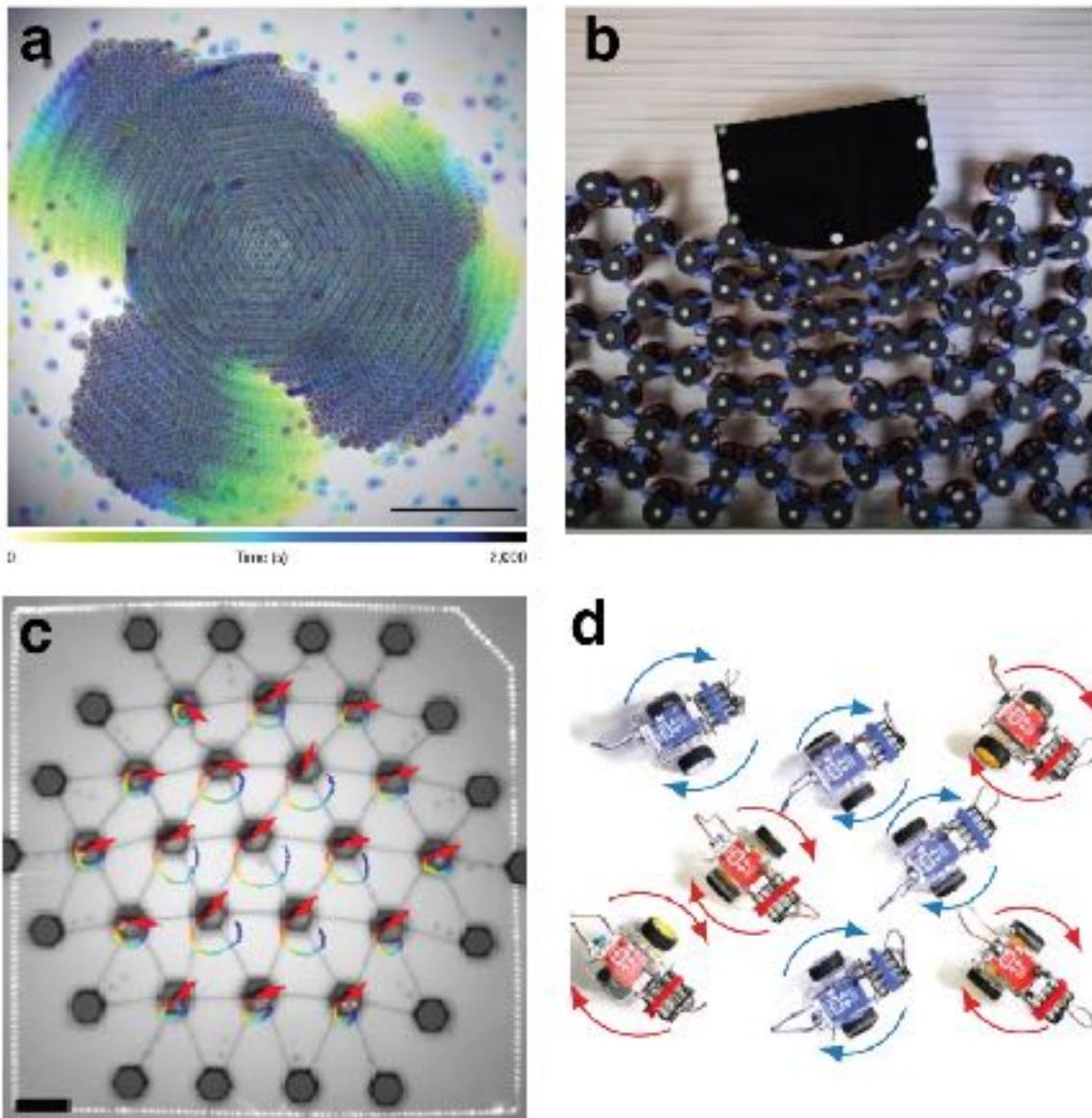


Figure 38.12: Active matter vs. active mechanical metamaterials. (a,c) In active matter, one typically study the emerging dynamics in collections of interacting active units. (b,d) In active mechanical metamaterials, one programs the active force instead to obtain a tunable effective response with unconventional properties. (a) Chiral dynamics in aggregates of star-fish embryos, adapted from Ref. [18]. (b) Odd elastic response of an active mechanical metamaterial, adapted from Ref. [1]. (c) Chiral pattern in an active solid, adapted from Ref. [5]. (d) Chiral pattern in non-reciprocally coupled robots, adapted from Ref. [36].

understand active matter could be in turn used to better understand active mechanical metamaterials beyond the linear response we surveyed here, for instance in the presence of nonlinearities, instabilities and noise, which are widespread in active fluids but which remain completely unexplored in active metamaterials.

Problems

Agnese: Ok figures in problems, but make them into numbered figures also with title and caption

Problem 38.1: Dynamics of a crawling time-modulated metamaterial. Simulate a time-modulated metamaterial with $N = 10$ unit cells using the equation above for a few different values of the wave-length λ .

Problem 38.2: The non-Hermitian skin effect. Consider the case where there are additional next-nearest neighbors odd springs [37]

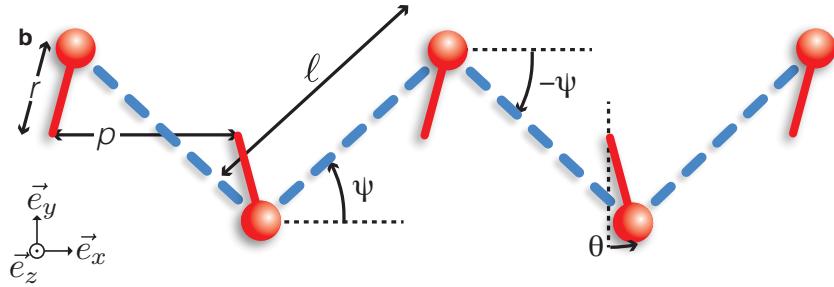
$$m \frac{\partial^2 u_i}{\partial t^2} = \underbrace{-k(2u_i - u_{i-1} - u_{i+1})}_{\text{elastic}} - \underbrace{k'(u_{i-1} - u_{i+1}) - k''(u_{i-2} - u_{i+2})}_{\text{active}}. \quad (38.31)$$

We will focus on the case $m = 1$ and $k = 1$, without loss of generality.

a. Compute the dispersion relation ω vs. q and plot $\operatorname{Re} \omega$ vs. q and $\operatorname{Im} \omega$ vs. q for various values of k''/k' in the case where $k' > 0$.

b. Show that for some value of k''/k' , waves with a positive group velocity can be either amplified or attenuated, depending on the value of the wave vector q .

Problem 38.3: The non-Hermitian skin effect in a topological chain. Consider the active mechanical metamaterial below.



p is the lattice spacing, $\delta\theta_i$ is the angular deflection of rotor i , k the spring constant of the blue springs and ε is the odd gain applied by the motors. Such a system is described by the following equation of motion:

$$J \frac{\partial^2 \delta\theta_i}{\partial t^2} = \underbrace{-kr^2 \cos^2 \theta ((a^2 + b^2) \delta\theta_i - ab \delta\theta_{i-1} - ab \delta\theta_{i+1})}_{\text{elastic}} - \underbrace{\varepsilon kr^2 \cos^2 \theta (b(b \delta\theta_i - a \delta\theta_{i-1}) - a(a \delta\theta_i - b \delta\theta_{i+1}))}_{\text{active}}, \quad (38.32)$$

where the coefficients $a = (p - 2r \sin \theta) / \sqrt{p^2 + 4r^2 \cos^2 \theta}$ and $b = (p + 2r \sin \theta) / \sqrt{p^2 + 4r^2 \cos^2 \theta}$ are geometric parameters.

a. Prove that the dispersion relation is $\omega^2 = kr^2/J(-a + be^{-iq})(-a(1 - \varepsilon) + b(1 + \varepsilon)e^{iq})$.

b. Plot $\Im \omega$ vs. $\Re \omega$ as a parametric plot, parameterized by the wave vector q for $a/b > 1$ and for various values of ε .

c. What happens when $\varepsilon = (a/b \pm 1)/(a/b \mp 1)$?

Problem 38.4: Green's function of an odd chain. Consider a chain made of odd springs, governed in the continuum limit by Eq. (38.23).

a. Fourier transform this equation in the spatial domain and then Laplace transform it in the time domain defined as

$$\hat{u}(q, s) = \int_{-\infty}^{+\infty} dx \int_0^{+\infty} dt u(x, t) \exp(iqx - st), \quad (38.33)$$

and prove that

$$\hat{u}(q, s) = \frac{1}{\frac{s^2}{c^2} + q^2 - 2\varepsilon iq}. \quad (38.34)$$

b. Invert the Fourier transform to prove that

$$\bar{u}(x, s) = \frac{ce^{\varepsilon x}}{2} \frac{e^{-\frac{|x|}{c}\sqrt{s^2 + c^2 \varepsilon^2}}}{\sqrt{s^2 + c^2 \varepsilon^2}}. \quad (38.35)$$

[Hint: use the theorem of residues.]

- c. Invert the Laplace transform and show that

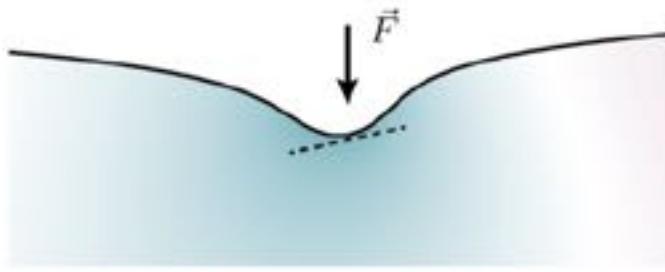
$$u(x, t) = \frac{c e^{\varepsilon x}}{2} \Theta\left(t - \frac{|x|}{c}\right) \left(J_0\left(\varepsilon \sqrt{c^2 t^2 - |x|^2}\right) - \frac{(ct - |x|)^2}{c^2 t^2 - |x|^2} J_2\left(\varepsilon \sqrt{c^2 t^2 - |x|^2}\right) \right), \quad (38.36)$$

where Θ is the Heaviside function and where J_k are Bessel functions of the first kind. [Hint: use the change of variables $s = c \frac{\varepsilon}{W} \left(\xi - \frac{W^2}{4\varepsilon} \right)$ and refer to a book on Bessel's functions such as [38].]

Problem 38.5: Odd elastic square lattice. Consider a square lattice made of chiral odd springs.

- a. Write the dynamical matrix for particle (i, j) in real space.
- b. Write the dynamical matrix in Fourier space and plot the dispersion relation.
- c. Redo the calculations above by adding elastic springs that connect next-nearest neighbors.
- d. Plot the dispersion relation in the case where the dynamics is overdamped and where viscous drag dominate the dynamics. Do you predict the emergence of active waves in such overdamped regime?

Problem 38.6: Point response in an odd elastic material. Consider the active mechanical metamaterial below [1]. It is incompressible, has a shear modulus μ , and an odd elastic modulus K^o . We assume that the odd elastic solid is a 2D half-infinite solid. We apply a point force of magnitude F on this material.



- a. The condition of static mechanical equilibrium $\partial_i \sigma_{ij} = 0$ is equivalent to $\Delta^2 \chi = 0$, where χ is the Airy stress function. Show that the Airy stress function is given by

$$\chi = -\frac{F}{\pi} r \phi \cos \phi. \quad (38.37)$$

- b. The stress relates to the Airy stress function as $\sigma_{ij} = \varepsilon_{il} \varepsilon_{jk} \partial_l \partial_k \chi$. Show that $\sigma_{rr} = -\frac{2F}{\pi r} \cos \phi$ and $\sigma_{r\phi} = \sigma_{\phi\phi} = 0$.

- c. Using the relation between stress and strain of this odd elastic material, show that

$$u_{rr} = \partial_r u_r = -\frac{4F}{\pi E r} \cos \phi, \quad (38.38)$$

$$u_{\phi\phi} = \frac{1}{r} \partial_\phi u_\phi + \frac{u_r}{r} = \nu \frac{4F}{\pi E r} \cos \phi, \quad (38.39)$$

$$2u_{r\phi} = \partial_r u_\phi - \frac{u_\phi}{r} + \frac{1}{r} \partial_\phi u_r = -\nu^o \frac{8F}{\pi E r} \cos \phi. \quad (38.40)$$

- d. Prove that the displacement field, in cylindrical coordinate below the point of application of the point force is

$$u_r = \frac{4F}{\pi E} [-\log r \cos \phi - \nu^o (\sin \phi + \phi \cos \phi)], \quad (38.41)$$

$$u_\phi = \frac{4F}{\pi E} [(1 + \log r) \sin \phi + \nu^o \phi \sin \phi]. \quad (38.42)$$

In the align above there was a “qty” command that gave error and I deleted, because I couldn't understand what it was doing exactly. Not sure if it is needed. The author should check this formula.

Bibliography

- [1] M. Brandenbourger, C. Scheibner, J. Veenstra, V. Vitelli, and C. Coulais. Active impact and locomotion in robotic matter with nonlinear work cycles. *arXiv:2108.08837*, 2021.
- [2] Y. Chen, X. Li, C. Scheibner, V. Vitelli, and G. Huang. Realization of active metamaterials with odd micropolar elasticity. *Nature Communications*, 12(1):5935, 2021.
- [3] S. Li, R. Batra, D. Brown, H. D. Chang, N. Ranganathan, C. Hoberman, D. Rus, and H. Lipson. Particle robotics based on statistical mechanics of loosely coupled components. *Nature*, 567(7748):361–365, 2019.
- [4] C. Scheibner, A. Souslov, D. Banerjee, P. Surówka, W. T. M. Irvine, and V. Vitelli. Odd elasticity. *Nature Physics*, 16(4):475–480, 2020.
- [5] P. Baconnier, D. Shohat, C. Hernandès, C. Coulais, V. Démery, G. Düring, and O. Dauchot. Selective and collective actuation in active solids. *arXiv:2110.0151*, 2021.
- [6] H. Nassar, H. Chen, A. N. Norris, M. R. Haberman, and G. L. Huang. Non-reciprocal wave propagation in modulated elastic metamaterials. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2202):20170188, 2017.
- [7] Y. Wang, B. Yousefzadeh, H. Chen, H. Nassar, G. Huang, and C. Daraio. Observation of nonreciprocal wave propagation in a dynamic phononic lattice. *Physical Review Letters*, 121(19):194301, 2018.
- [8] M. Brandenbourger, X. Locsin, E. Lerner, and C. Coulais. Non-reciprocal robotic metamaterials. *Nature Communications*, 10(1):4608, 2019.
- [9] G. Trainiti, Y. Xia, J. Marconi, G. Cazzulani, A. Erturk, and M. Ruzzene. Time-periodic stiffness modulation in elastic metamaterials for selective wave filtering: Theory and experiment. *Physical Review Letters*, 122(12):124301, 2019.
- [10] H. Nassar, B. Yousefzadeh, R. Fleury, M. Ruzzene, A. Alù, C. Daraio, A. N. Norris, G. Huang, and M. R. Haberman. Nonreciprocity in acoustic and elastic materials. *Nature Reviews Materials*, 5(9):667–685, 2020.
- [11] A. Ghatak, M. Brandenbourger, J. van Wezel, and C. Coulais. Observation of non-hermitian topology and its bulk-edge correspondence in an active mechanical metamaterial. *Proceedings of the National Academy of Sciences of the United States of America*, 117(47):29561–29568, 2020.
- [12] G. Oliveri, L. C. van Laake, C. Carissimo, C. Miette, and J. T. B. Overvelde. Continuous learning of emergent behavior in robotic matter. *Proceedings of the National Academy of Sciences of the United States of America*, 118(21), 2021.
- [13] W. Wang, X. Wang, and G. Ma. Non-hermitian morphing of topological modes. *arXiv:2203.02147*, 2022.
- [14] E. S. Bililign, F. Balboa Usabiaga, Y. A. Ganan, A. Poncet, V. Soni, S. Magkiriadou, M. J. Shelley, D. Bartolo, and W. T. M. Irvine. Motile dislocations knead odd crystals into whorls. *Nature Physics*, 2021.
- [15] T. Beatus, T. Tlusty, and R. Bar-Ziv. Phonons in a one-dimensional microfluidic crystal. *Nature Physics*, 2(11):743–748, 2006.
- [16] T. Beatus, R. Bar-Ziv, and T. Tlusty. Anomalous microfluidic phonons induced by the interplay of hydrodynamic screening and incompressibility. *Physical Review Letters*, 99(12):124502, 2007.
- [17] A. Poncet and D. Bartolo. When soft crystals defy Newton’s third law: Nonreciprocal mechanics and dislocation motility. *Physical Review Letters*, 128(4):048002, 2022.
- [18] T. H. Tan, A. Mietke, J. Li, Y. Chen, H. Higinbotham, P. J. Foster, S. Gokhale, J. Dunkel, and N. Fakhri. Odd dynamics of living chiral crystals. *Nature*, 607(7918):287–293, 2022.

- [19] R. Lier, J. Armas, S. Bo, C. Duclut, F. Jülicher, and P. Surówka. Passive odd viscoelasticity. *arXiv preprint arXiv:2109.06606*, 2021.
- [20] D. Banerjee, V. Vitelli, F. Jülicher, and P. Surówka. Active viscoelasticity of odd materials. *Physical Review Letters*, 126:138001, 2021.
- [21] É. Fodor and A. Souslov. Optimal power and efficiency of odd engines. *Physical Review E*, 104(6):L062602, 2021.
- [22] J. Clerk Maxwell. XLV. On reciprocal figures and diagrams of forces. *Philosophical Magazine Series 4*, 27(182):250–261, 1864.
- [23] Enrico Betti. Teoria della elasticità. *Il Nuovo Cimento (1869-1876)*, 7(1):69–97, 1872.
- [24] T. C. Lubensky, C. L. Kane, X. Mao, A. Souslov, and K. Sun. Phonons and elasticity in critically coordinated lattices. *Reports on Progress in Physics*, 78(7):073901, 2015.
- [25] S. Henkes, Y. Fily, and M. C. Marchetti. Active jamming: self-propelled soft particles at high density. *Physical Review E*, 84(4 Pt 1):040301, 2011.
- [26] E. Ferrante, A. E. Turgut, M. Dorigo, and C. Huepe. Elasticity-based mechanism for the collective motion of self-propelled particles with springlike interactions: a model system for natural and artificial swarms. *Physical Review Letters*, 111(26):268302, 2013.
- [27] E. Zheng, L. Robinet, M. Brandenbourger, P. Schall, E. Lerner, and C. Coulais. Self-oscillation and synchronisation transitions in elasto-active structures. *arXiv:2106.05721*, 2021.
- [28] F. G. Woodhouse, H. Ronellenfitsch, and J. Dunkel. Autonomous actuation of zero modes in mechanical networks far from equilibrium. *Physical Review Letters*, 121(17), 2018.
- [29] G. Trainiti and M. Ruzzene. Non-reciprocal elastic wave propagation in spatiotemporal periodic structures. *New Journal of Physics*, 18(8), 2016.
- [30] N. Hatano and D. R. Nelson. Localization transitions in non-hermitian quantum mechanics. *Physical Review Letters*, 77(3):570–573, 1996.
- [31] V. M. Martinez Alvarez, J. E. Barrios Vargas, and L. E. F. Foa Torres. Non-hermitian robust edge states in one dimension: Anomalous localization and eigenspace condensation at exceptional points. *Physical Review B*, 97(12):121401, 2018.
- [32] S. Yao and Z. Wang. Edge states and topological invariants of non-hermitian systems. *Physical Review Letters*, 121(8):086803, 2018.
- [33] T. Helbig, T. Hofmann, S. Imhof, M. Abdelghany, T. Kiessling, L. W. Molenkamp, C. H. Lee, A. Szameit, M. Greiter, and R. Thomale. Generalized bulk-boundary correspondence in non-hermitian topoelectrical circuits. *Nature Physics*, 16(7):747–750, 2020.
- [34] Lei Xiao, Tianshu Deng, Kunkun Wang, Gaoyan Zhu, Zhong Wang, Wei Yi, and Peng Xue. Non-hermitian bulk-boundary correspondence in quantum dynamics. *Nature Physics*, 16(7):761–766, 2020.
- [35] S. Weidemann, M. Kremer, T. Helbig, T. Hofmann, A. Stegmaier, M. Greiter, R. Thomale, and A. Szameit. Topological funneling of light. *Science*, 368(6488):311–314, 2020.
- [36] M. Fruchart, R. Hanai, P. B. Littlewood, and V. Vitelli. Non-reciprocal phase transitions. *Nature*, 592(7854):363–369, 2021.
- [37] M. I. N. Rosa and M. Ruzzene. Dynamics and topology of non-hermitian elastic lattices with non-local feedback control interactions. *New Journal of Physics*, 22(5), 2020.
- [38] G.N. Watson. *Neumann series and Lommel's functions of two variables*. Press, Cambridge University, 1922.

Chapter 39

Living Robotics

MARIA GUIX, RAFAEL MESTRE

While soft robotic systems have made significant strides in biomimetic designs, emulating the complex actuation mechanisms and functionalities of biological organisms, they have yet to fully replicate the unique properties inherent to living tissues, such as self-healing, energy efficiency, and bio-sensing capabilities. This Chapter ventures into the realm of integrating living muscle cells with robotic frameworks, highlighting their remarkable sensory, self-healing, and adaptive properties. Emphasizing advanced fabrication methodologies, including micro-fabrication and 3D printing, it focuses on achieving a native-like functional muscle structure, ensuring biomaterial compatibility, and maintaining the long-term culture requirements necessary for sustained operation of these systems. Additionally, this Chapter explores various actuation methods, with particular emphasis on electrical pulse stimulation, which enables distinct capabilities such as self-organization and emergent behaviors. These advancements are illustrated in Fig. 39.1, which illustrates the intricate integration of muscle tissue with synthetic materials, emphasizing the need for effective control strategies and methods to prolong the actuator's lifespan. Furthermore, this Chapter delves into the ethical considerations surrounding the design, creation, and application of these living robotic systems, particularly in the biomedical field. It aims to strike a balance between technological innovation and ethical responsibility, navigating the complexities of using living entities in robotics. By covering these multifaceted aspects, this Chapter offers a comprehensive overview of muscle-based living robots, underscoring their potential impact in the advancement of robotic systems and their crucial role in the development of reliable clinical models for personalized medicine.

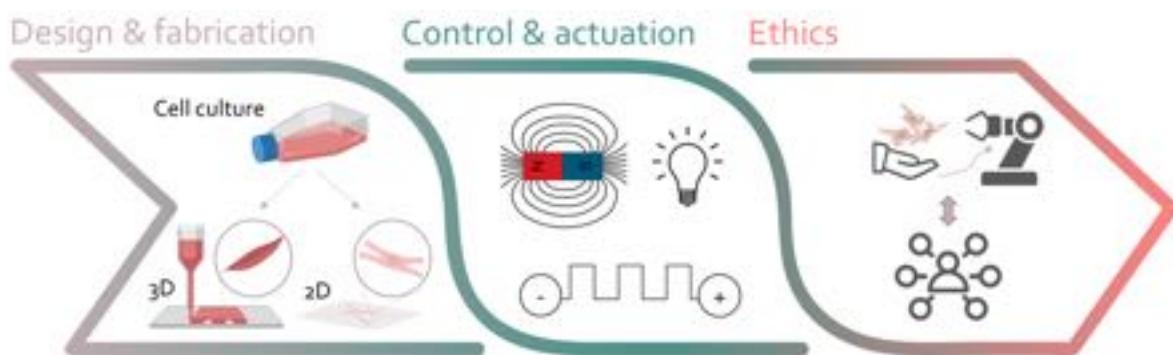


Figure 39.1: **Muscle-based living robots.** A muscle-actuated bio-hybrid robotic system relies on the contractile behavior of the muscle, which is rationally integrated in compliant synthetic materials that provide for mechanical integrity and aid towards the desired actuation (Created with BioRender.com). Effective control strategies and methods to prolong the actuator's lifespan are vital. Equally important is establishing ethical guidelines both for the creation and application of these living robotic platforms.

39.1 Advancements and challenges in bio-hybrid robotic systems

Soft robotic systems often present biomimetic designs that resemble the actuation mechanisms and functionality of certain biological organisms [1]. For example, we find artificial biomimetic designs in soft robots swimming by an antagonistic pair of actuators like fish [2] or swimming like flagellated organisms [3] as well as in other robotic platforms with grasping capabilities resembling a human hand, able to mimic the same complex motion patterns [Maria: can you add a reference?](#). However, there are some unique properties in living organisms that are especially challenging to acquire in fully artificial robotic platforms, even when integrating smart materials or when using advanced control features [4]. In this regard, directly integrating cells in robotic systems provides some of the desired capabilities from such living entities, including self-healing, energy efficiency, high power-to-weight ratio, adaptability, or bio-sensing capabilities [5, 6].

Several bio-hybrid robotics platforms across different scales have been developed [7]. Amongst these, those based on living muscles have attracted increasing attention due to their sensory and self-healing capabilities as well as due to their adaptability when trained under specific protocols, resulting in enhanced force outputs [8]. Besides the muscle's inherent capabilities, it is key to rationally integrate such living entities with compliant materials by exploiting several advanced fabrication technologies, such as microfabrication or 3D printing techniques [9, 10]. When designing such robotic platforms, it is equally crucial to consider the recent advancements in tissue engineering and the contemporary tools available in molecular genetics. Apart from their role as bio-actuators, 3D-printed cell-laden systems also show significant potential in the biomedical domain, particularly for diagnostic studies focusing on *in vitro* 3D tissue development and regeneration, as well as serving as platforms for drug testing.

The rational design and fabrication of bio-hybrid robotic platforms will lead to unique capabilities that will be dynamic over time through training protocols or even with the ability to self-repair. The success of such bio-hybrid platforms will require the interdisciplinary collaboration of biologists, engineers and physicists to properly manipulate cells/tissues, to implement a viable design capable of inducing the desired actuation, and to finally understand the undergoing physical events that cause motion. In addition to the already mentioned key players in the creation of living robots, one has also to take into account ethical considerations on both the design and fabrication of such platforms, ranging from the source and manipulation of the living entities, to the final application of such bio-hybrid systems.

This Chapter will comprehensively address these themes, spanning from the initial conception to the detailed fabrication and thorough characterization of bio-hybrid platforms; it will also delve into the ethical considerations intertwined with such technology, particularly in relation to the utilization of living organisms, as illustrated in Figure 39.1. We will first present the key design features, based on the three main aspects of implementing living robots: (i) achieving a native-like functional muscle structure; (ii) ensuring bio-compatibility of the integrated materials; and (iii) meeting the long-term culture requirements for an extended lifetime. We will then move forward to the description of the main fabrication techniques used to fabricate muscle-based living robots. Subsequently, our discussion will pivot to various actuation methodologies designed to modify the contractile behavior of these systems, thereby achieving effective motion and response. Special attention will be given to electrical pulse stimulation. In addition, we will explore the characterization of force and highlight other distinct features such as self-healing, self-organization, and emergent behaviors. Finally, we will address the ethical considerations pertaining to the use of living organisms and their subsequent applications in the biomedical field; this discussion is vital for fostering the development of a new generation of robotic systems and establishing dependable clinical models for personalized medicine.

39.2 Key design features

The conception of muscle-actuated bio-hybrid robotic systems implies not only the rational platform design for efficient actuation and control, but also a need to account for the compliance/bio-compatibility of the materials used and the proper environment to sustain cell viability (Fig. 39.2a). In this section, we will first explore the various muscle cell sources used in the development of bio-hybrid robotics. Subsequently, we will delve into prominent methods for ensuring optimal cell alignment and organized control over compliant structures, aiming for effective actuation.

Unlike conventional artificial systems, bio-hybrid robot designs are not bound by typical design guidelines. Instead, they heavily rely on the selected living component serving as the actuator, which is precisely stimulated

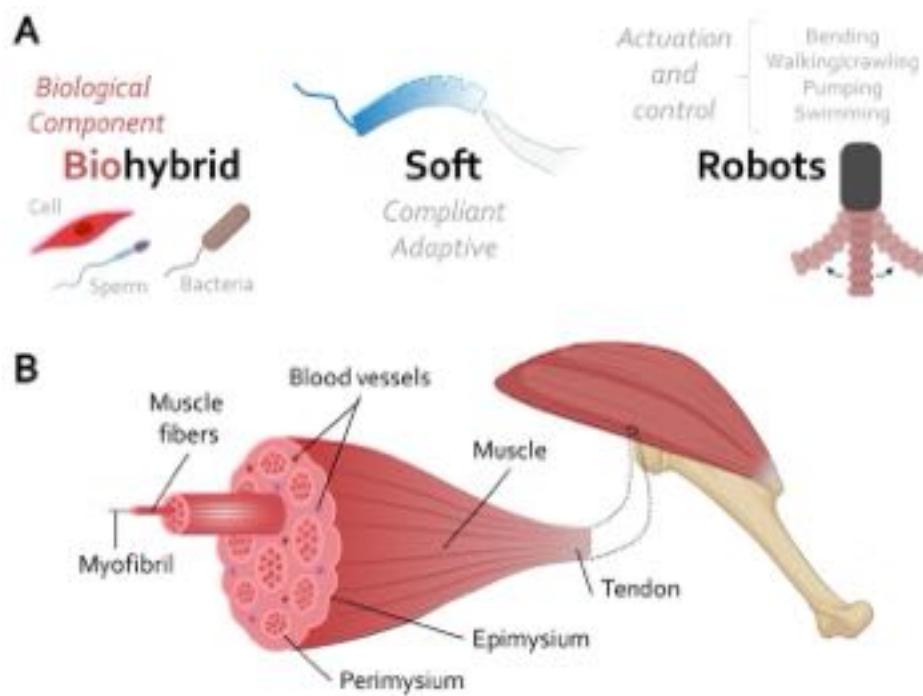


Figure 39.2: Hierarchical organization of muscle tissue. (a) Main characteristics of bio-hybrid soft robots. **(b)** Skeletal muscle tissue structure. Created with BioRender.com.

in a spatiotemporal fashion. The first criterion to define is the size range of the bio-hybrid actuator. At the nano- and microscale, we find examples going from enzyme-based nanoparticle robots [11] to robotic platforms with advanced tactic behavior based on sperm cells or bacteria [12], either as single motile units or as swarms. In contrast, when moving to larger dimensions, the use of entire animals in what is known as cyborg systems is also being explored [13, 14]. There are specific constraints concerning the size and environmental conditions essential for the viability of the biological component. It is imperative to meticulously select the most suitable living materials that align with the needs of the robotic platform and its intended application. Indeed, scaling up these systems presents a distinct challenge [15], which is why the majority of investigated living robots are predominantly small-scale.

In this Chapter, we focus our attention on bio-hybrid systems that fall in the middle range, namely on cell-based systems where muscle tissue is integrated as an active actuator in a robotic platform. Some of the most appealing capabilities of such actuators are their (i) high power-to-weight ratio (for skeletal muscle, ranging from 40 to 225 W kg^{-1} , with a force generation of about 40 N cm^{-2}) and efficiency (10% to 40% during cyclic motion), (ii) good motion control in terms of repeatability and accuracy, (iii) optimal energy storage (self-contained power supply), (iv) robustness and dependability in their native state, and (v) environmental safety (i.e., consuming a renewable fuel source for actuation and not producing harmful waste products) [16]. The most extensively used muscle cell sources are heart muscle cells (or *cardiomyocytes*) and skeletal muscle cells. In both cases, they are generally derived from mammalian sources [8], although there are some exceptions where amphibian cells [17] or insect cells [18] have been used.¹ The organization of such cells will determine the morphology of the resulting bio-hybrid robot, which can be actuated by a single muscle cell, by a muscle cell monolayer, or by some muscle tissue. Additionally, in order to mimic the native neuronal muscular actuation, optogenetic stem-cell-derived neural clusters containing motor neurons can be combined with muscle tissue to increase the complexity and biomimicry of bio-hybrid robots [20]. Therefore, recent advances in molecular biology allows us to exploit the inherent capabilities of certain cell lines as well as to take advantage of genetic editing for improved control capabilities.

When designing muscle-based actuators, the next level of complexity relies on achieving the same hier-

¹In fact, insect cells are especially attractive because of their exceptional durability in a wide temperature range (5° to 40°C), by contrast to mammalian cells, whose optimal viability is only around 37°C [19].

archical organization and cellular alignment present in the native muscle structure. In living organisms, cell maturation and self-organization of fascicles are guided by chemical cues and other bio-mechanical stimuli, leading to emergent properties. Skeletal muscle structure is characterized by 3D fibers organized in a distinct arrangement within the *extracellular matrix* (ECM). This matrix directs the muscle, grouping its fibers into bundles. These bundles then form fascicles that produce contractile force [21]. Notably, when these structures are organized hierarchically, they exhibit enhanced mechanical behavior (Fig. 39.2b) [22]. Although there have been great advances in the field of tissue engineering, even demonstrating the reconstruction of a muscle-based fascicle-like tissue by using anisotropic 3D patterning [23], the translation of such functional tissues in robotic platforms is challenging. For example, cardiac muscle cells are generally cultivated in 2D environments, demonstrating an efficient synchronous beating and efficient actuation. However, recent studies related to cardiac-based spheroids (cardioids) suggest that 3D organoids would be desirable to replicate the complexity of the spreading, differentiation, and migration of cardiac muscle and vascular layers [24]. While this level of complexity has not been documented in the integration of cardiomyocytes into robotic systems, current fabrication approaches for bio-hybrid robots using skeletal muscle cells primarily employ 3D printing techniques. In contrast, for cardiac-based systems, 2D polymeric thin films have yielded impressive outcomes.

Native muscle also presents other elements that can be translated when creating an actuator, such as its ECM and the *muscle-tendon unit* (MTU), which provide physical scaffolding for the cells and mechanical stability, respectively [21]. In the native system, there's an interplay between muscle and bone, ensuring sustained tension during the muscle maturation phase. Such tensile stress will later be translated in the robot's skeleton, ensuring the cell-laden scaffold stability's both during the muscle differentiation/maturation and during the muscle actuation in the robotic system. Such structures have to be compliant to adapt to the tissue contractile behavior and, depending on the size range of the resulting robotic system, different materials and structures can be considered. Micro-electro-mechanical systems (MEMS) were among the initial passive structures to which cells were attached, specifically by seeding rat cardiomyocytes onto the micro-fabricated framework (Fig. 39.3a) [25]. However, one of the most widely utilized artificial components in muscle-based bio-hybrid systems is thin polymeric films; these range from the well-established polydimethylsiloxane (PDMS) to biodegradable or conductive polymers, such as poly(lactic acid) (PLLA) and poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS) [9]. These ultra-thin films offer a substantial surface-area-to-volume ratio and exceptional flexibility; furthermore, they can undergo a broad array of physical and chemical modifications to achieve precise organization of the muscle cells. Cardiomyocytes are the preferred cells from this point of view due to their controlled growth in 2D substrates (such as the soft robotic ray in Fig. 39.3b). In contrast, 3D-printed platforms have been generally used in skeletal muscle-based bio-robots, using bio-compatible polymers such as poly(ethylene glycol) diacrylate (PEGDA) [27] (Fig. 39.3c) or PDMS [28], where different

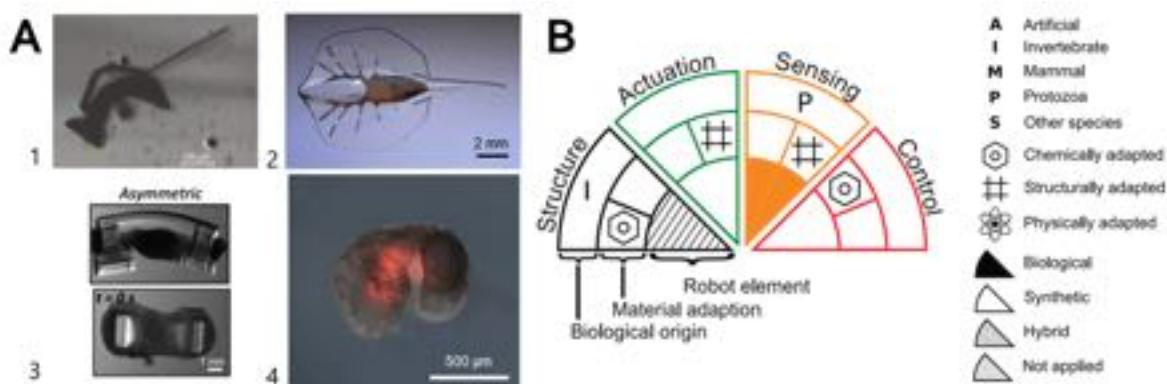


Figure 39.3: Bio-hybrid robot skeletons. (a-d) Different types of passive compliant structures combined with biological muscle-based elements, such as (a) MEMS (reproduced from Ref. [25]), (b) thin polymeric-based film (reproduced from Ref. [26]), (c) 3D printed flexible skeletons (reproduced from Ref. [27]), and (d) non-contractile cell source (reproduced from Ref. [17]). (e) Material Taxonomic Key (MTK) to classify functionality and composition of bio-hybrid robots (reproduced from Ref. [5]). *Agnese: uniform figure to book style. Reorganize panels, for example, putting 1-4 (to become a, b, c, d) in a row above b (to become e).*

geometries where explored for actuation purposes, while ensuring an adequate adjustment of the cell-based actuator without provoking cell damage.

Up to this point, the passive elements ensuring the bio-hybrid system's mechanical stability have been artificial, offering a degree of asymmetry to achieve effective motion. However, a recent example using passive and active (contractile) stem cells from frogs allowed the development of fully biological robots (Fig. 39.3d) [17]. It should be noted that the distribution of active/passive cells was previously simulated *in silico* to predict their behavior by evolutionary algorithms, selecting only the designs that performed efficient displacements. This was the very first time that such biological actuators were simulated prior to their fabrication, which highlights the potential of including prediction methods for the conception of more efficient and novel bio-hybrid systems.

There is a useful taxonomic key proposed by Webster-Wood *et al.* [29], which includes some of the key aspects we briefly discussed in this section, such as (i) the structure of the robotic system (biological origin, material adaption, robotic element), (ii) the actuation system, (iii) the sensing capabilities, and (iv) the control mechanisms. This *robotic taxonomic key* (RTK) is based on a visual schematic where the actuator source (biological, synthetic or bio-hybrid) and the complexity of the robotic system in terms of control are depicted. This classification method was initially envisioned to present a general overview when organic actuators (i.e., bio-compatible polymers, biological components) are combined with other organic/inorganic structural components. However, most of the organic actuators and controllers covered are indeed cell-based systems. In fact, such taxonomy key was later used to classify a wide range of bio-hybrid soft robots based on living materials with a *material taxonomic key* (MTK) shown in (Fig. 39.3e) [5], giving a special emphasis to the functions that such devices perform.

39.2.1 Design versus actuation mode

Actuation in robotic platforms can be reflected in different actions, such as pumping, gripping, walking, crawling or swimming. In artificial soft robotics, these operations typically culminate in the execution of a specific task. This often necessitates the robot to perform the intended physical movement, complemented by various sensory elements to automate the task. Therefore, automation encompasses more than just actuation; it also involves vision-based control techniques that offer feedback about the robot's location and its environment,

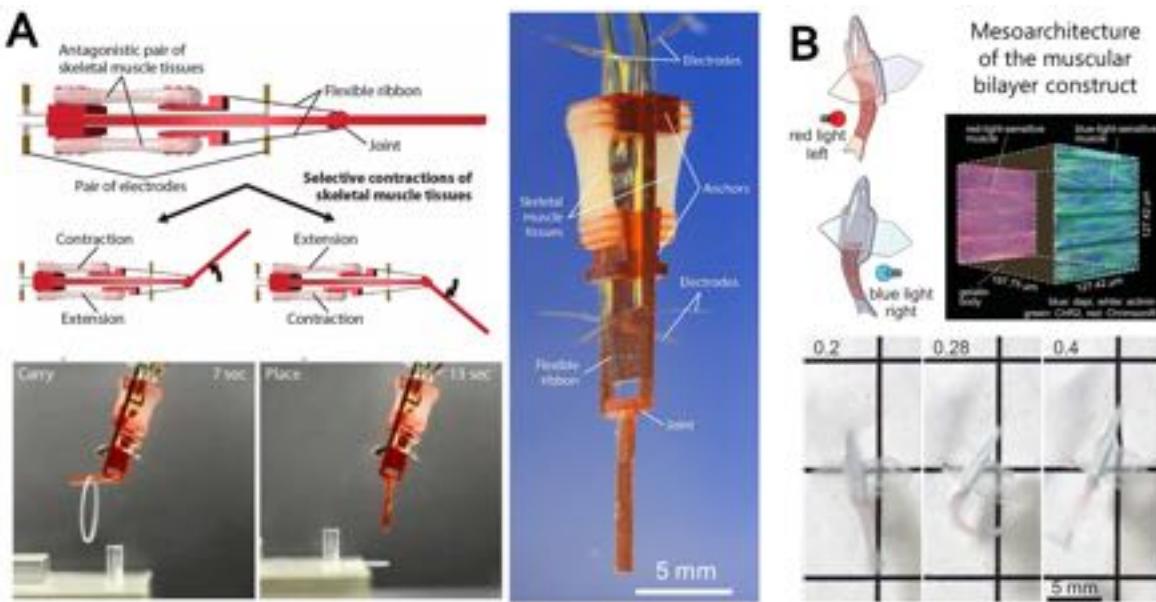


Figure 39.4: **Antagonistic muscle pairs.** Bio-hybrid systems with an antagonistic pair of muscle-based actuators: (a) Using skeletal muscle cells and selectively actuated by electrical stimulation to achieve pick-and-place tasks (reproduced from Ref. [30]); (b) using two different optogenetically modified cardiac muscle cells seeded on each side of a thin film, whose fish-like motion is obtained by alternating blue-light and red-light stimulation (reproduced from Ref. [31]).

guiding its responses to specific tasks, such as manipulating an object.

The mode of actuation in soft robots typically relates to specific robot deformations and is influenced by the intended purpose of the robotic platform. We will first need to take into account the actuator position in the whole robotic structure and, especially in the case of bio-hybrid robots, the surrounding conditions which are crucial to ensure cell viability (e.g., temperature, CO₂ concentration). As an example, we will focus our attention on the implementation of antagonistic pairs of muscles in robots. In fact, the design principles of native antagonistic muscles has extensively been replicated in pneumatic artificial muscles [32, 33], particularly biceps or triceps actuation, where one muscle contracts while the other relaxes, and vice versa.

When designing bio-hybrid robots, such antagonistic pair of muscles have been differently applied depending on the purpose and muscle cell source used, demonstrating the development of a juncture or joint capable of bending and performing pick-and-place tasks [30] (Fig. 39.4a), and a swimming bio-hybrid robot, whose motion is based on reciprocal muscle contraction [31] (Fig. 39.4b). In the first case, two 3D constructs based on three-dimensional skeletal muscle tissue were assembled in a flexible substrate, achieving the desired linear contraction and extension by selectively applying electrical stimulation to each muscle unit. In the second case, a muscular bi-layer construct was fabricated by self-assembling cardiomyocytes that responded to light stimulation in each side of a thin film; therefore, the mechano-electrical properties of cardiac cells were modulated by genetically modifying them to respond to a specific wavelength, which in terms of actuation was translated into an alternating blue-light and red-light stimulation.

As previously indicated, the presence of a skeleton is key to provide proper mechanical stability to the final bio-hybrid robotic platform. Indeed, a proper skeleton design will be a crucial defining player in the resulting performance, as it can provide the desired asymmetry for effective motion, either in terms of friction for walkers or in terms of inertial forces in mesoscale swimmers. The versatility of the design configuration is high when working with 3D-printed technologies. For instance, bio-hybrid walkers can be fabricated by an asymmetric structure composed of two stiff pillars of different lengths joined by a compliant beam, coupled to a skeletal-muscle-laden hydrogel [27], allowing for efficient locomotion (Fig. 39.3c). This same structure was further improved to increase its complexity and enable different muscle tissue configurations, which in turn created different forms of actuation in different directions (Fig. 39.5a) [34]. It is worth noting that 3D scaffolds are preferred when working with skeletal muscle cells, as opposed to cardiomyocytes. On the other hand, non-biomimetic spring-like skeletons can be assembled to skeletal muscle rings to allow not only controlled bending that creates an asymmetry, but also continuous mechanical stimulation of the cells during the

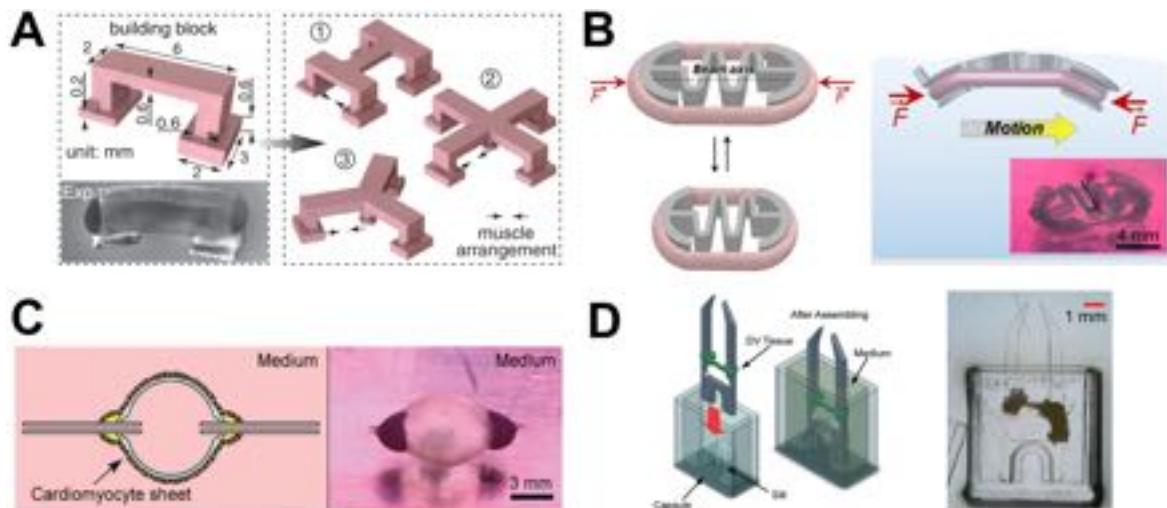


Figure 39.5: Bio-hybrid robots assembled to 3D skeletons. Bio-hybrid platforms with skeletons fabricated by 3D printing technologies with different multi-ring designs for (a) advanced control (reproduced from Ref. [34]) and (b) self-stimulating spring-like skeletons (reproduced from Ref. [28]). (c) Micro-spherical heart-like pump with a cultured cardiomyocyte sheet in cell culture medium (reproduced from Ref. [35]). (d) Atmospheric-operable bioactuator based on the assembly of insect muscle tissue on PDMS-based micropatterns (reproduced from Ref. [36]).

differentiation process, proving the enhanced force output and cell alignment when matured in the presence of a compliant skeleton (Fig. 39.5b) [28]. Another interesting example of a micro-spherical 3D pump actuated by cardiomyocytes was fabricated by promoting the cell attachment using PDMS with cell-adhesive properties (Fig. 39.5c) [35]. In this specific case, no 3D printing techniques were implemented, creating instead a PDMS hollow sphere using a sacrificial sugar ball surface and later including penetrant capillary inlets and outlets, connecting the whole system to a syringe pump to dissolve the sugar ball once the polymerization was complete.

Whereas most of the muscle-based bio-hybrid robotics systems use mammalian cell sources for their interest in biomedical research models, invertebrate tissues have also been explored as they present a more robust performance and longer term culture conditions. The first example was based on the assembly of a micro-mechanical structure (MEMs) with insect muscle bundles (Fig. 39.3a) [25]. More advanced platforms have been developed, exploring for example an interesting bio-hybrid gripper actuator [36], where dorsal vessel insect muscle tissue was assembled on PDMS-based micropatterns operating at atmospheric conditions (both in dry and wet conditions) (Fig. 39.5d).

Regarding cardiomyocytes-based robotic systems, they are generally fabricated by using 2D materials (i.e., flexible thin films), as they prefer to grow as muscle monolayers. One of the main challenges is cell alignment, crucial to obtain an efficient contractile behavior in the resulting bio-hybrid material. One of the most elegant demonstrations of how such alignment could be maximized by mimicking the sparse muscle layout of the jellyfish ephyrae was reported by Parker *et al.* in the medusoid swimming robot [37]. Not only the motion mechanisms resembled in space and time the patterns of a real jellyfish, but they also achieved controlled self-assembly of neonatal rat cardiomyocytes by using micro-patterned ECM cues. It should be noted, however, that simple imitation of the jellyfish muscle architecture did not result in the desired actuation, and the creation of a muscle layout featuring extensive radial components was necessary (Fig. 39.6). This example also highlights the importance of such bio-hybrid robotic systems as models of great interest to reverse engineering, as they allow us to get a better insight on key parameters that rule undergoing biological processes.

The use of whole muscles as an active element has also been demonstrated in, for example, the use of I2 muscle from *Aplysia Californica* in a 3D-printed polymeric framework [29]. Such bio-hybrid robots presented excellent force output, since the muscle was already well-differentiated, as it was directly dissected from the animal. However, associated ethical concerns and the lack of versatility in terms of design, generally makes the use of commercial cell lines for the fabrication of bio-hybrid robots the preferred option for researchers.

Nonetheless, a recent approach by direct cell assembly, with no need of any artificial framework, has opened the door to the development of fully biological living robots. The fabrication parameters discussed so far, namely an active element and a passive one providing support to the contractile cells, as well as attention

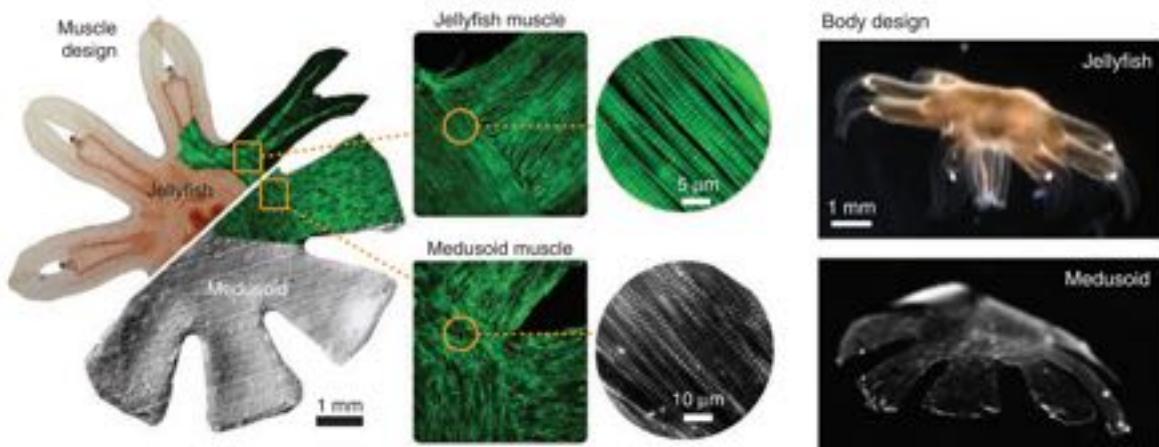


Figure 39.6: **Medusoid swimming robot.** Jellyfish 2D muscle architecture and body design (top) compared to muscle arrangement and body design in a medusoid bio-hybrid swimmers (bottom) (reproduced from Ref. [37]).

to cell alignment and asymmetry, are still valid. Such living robots are the so-called *xenobots*, where the passive and active cells are obtained from pluripotent stem cells of *Xenopus laevis*, a type of frog. By using an evolutionary algorithm, the authors obtained the optimal design of the configurable organism that they later replicated and validated by direct cell assembly (Fig. 39.3d) [17]. Other xenobot configurations have recently been explored, demonstrating self-organization into multiciliated cells (no complex genomic editing required), swimming motion associated to the beating of cilia, and showing self-repair after damage [38] and self-replication [39].

39.2.2 Motion control

One of the big challenges when working with living actuators relies on controlling their actuation, not only by providing on/off motion, but also by being able to guide, and later perform, motion planning. As previously mentioned, flexible skeletons in living robots provide both asymmetry during robotic actuation and mechanical stability of the cell-laden scaffold. However, to achieve a more refined controlled motion, it is important to develop local actuation. The integration of electrodes into the robot's skeleton has opened up new possibilities. In tethered devices, such as the antagonistic pair of skeletal muscle tissue depicted in Fig. 39.4, it is now possible to selectively contract one of the muscle-based actuators [30]. In this regard, micro-fabricated compliant 3D frameworks based on flexible electronics have also been assembled with 3D optogenetically active muscle rings for precise measurement of the spatial and temporal contractile forces [40]. Integrating electrodes in the robot's skeleton has also resulted in improved electrical stimulation, as reported for 3D graphene electrodes coupled to 3D biological tissue systems [41].

The most desirable living robotic platforms are generally untethered, being of great interest for their potential implementation in biomedical applications. For skeletal-muscle-based walking machines, a systematic study over computationally designed skeletons allowed not only for an improved forward motion, but also for a controlled rotational steering by playing with two and four-legged scaffolds, respectively [34]. Recently, a wireless powered controllable bio-hybrid soft robot was reported. This soft robot was constructed by using a stretchable wireless device with 3D-printed accordion-inspired scaffolds where human-induced pluripotent stem cells were grown in an aligned manner. This approach enabled achieving multiscale mechanical properties by playing with different micro-patterns [42]. Controlled cardiomyocyte differentiation was achieved with monophasic pulses, being able to control their speed by using different frequencies. Although frequency-dependent speed modulation had been demonstrated before [26, 28], the local actuation of such wireless platform can provide a more versatile platform to achieve controlled and guided motion in muscle-based actuators. Moreover, the integration of electrical sensors in bio-hybrid systems can also provide additional information over the undergoing biological events taking place during the living robot actuation. Such information is crucial to better understand the factors that can improve the living actuator performance's as well as to provide useful information to reverse-engineer native muscle systems.

In terms of control, light-controlled living robots are of special interest to provide non-invasive stimulation. They also offer easiness of local actuation by providing controlled spatial excitation of optogenetically modified biological elements using light of an appropriate wavelength. The first example of phototactic guidance was demonstrated for a tissue-engineered ray, capable of performing downward and upward contraction depending on the number of muscle layers in each side of a *flexible* thin film [26]. By applying a sequential activation of pectoral muscles, a spanwise bending deformations and chordwise front-to-rear undulatory motion was obtained. Similar antagonistic motion was reported for a fish-like platform by using different blue-and-red light stimulation, a bio-hybrid robot system already discussed in Section 39.2.1. Recently, a new scalable modular bio-fabrication process based on applying static mechanical stretch in the muscle-laden scaffold during the maturation process was presented, improving myotube formation and alignment and resulting in a multi-directional bio-hybrid walker that maximized the active muscle force output [43]. Another interesting approach relies on mimicking the neural circuitry present in the native muscle by using genetically modified neurons that will be later integrated in the 3D muscle actuator, allowing a light-controlled neuromuscular actuation of a bio-hybrid system [20]. However, it should be noted that although genetic editing allows a well-defined spatiotemporal and non-invasive actuation, the biological performance of the genetically modified element gets impacted. For example, in the case of muscle cell lines, several passages are required to introduce such genetic modification, meaning that the cell lines used in that bio-hybrid platform will be from older cultures.

Guiding a robotic platform does not necessary imply having full control over its motion. In fact, the most

basic and elemental control required in a robotic platform is modulating on/off motion. Skeletal-muscle-based tissues present such refined control, being able to stop or activate their actuation on demand, while cardiac cells do not stop beating after being activated (only in the presence of certain compounds, as for example acetylcholine, do heart muscle beats decrease). The subsequent phase involves advanced path planning, achievable only with comprehensive automation incorporated into the system. This necessitates vision-based techniques to gather feedback from the robot's immediate environment, enabling obstacle avoidance and tailoring algorithms to modulate the stimulus source (e.g., light, electrical input) directing the robot towards its intended destination. Such sophisticated control has not been documented in any of the platforms discussed here. While extensively researched in fully artificial robotics, there remains a significant gap in the realm of bio-hybrid robots. Beyond mere actuation or guidance, comprehensive task planning (inclusive of closed-loop feedback control) is paramount. This level of complexity aligns with the advancements seen in artificial soft robots, such as anthropomorphic hands that mimic human dexterity and grip strength [44].

39.3 Actuation by electrical pulse stimulation

Electrical stimulation is the primary method for actuating bio-hybrid robots, given muscle cells' natural responsiveness to it. While light-responsive muscle cells necessitate genetic engineering, which complicates the creation process and potentially affects the performance of the resultant bio-hybrid robotic platform, electrical stimulation remains straightforward and effective. This section will delve into the fundamentals of electrical stimulation and the techniques to measure force in muscle-based actuators. It is vital to remember that as technology advances, exploring alternative stimulation methods will be crucial for broadening the control dynamics of bio-hybrid robots. This mirrors the diverse actuation sources available for soft actuators, which encompass electrical, magnetic, thermal, light, pressure, and even explosive-driven mechanisms [45].

Skeletal muscle contraction is achieved through electrical activation by the nervous system. At the interface between motor neurons and muscle tissue (what is known as the *neuromuscular junction*), a chemical synapse is formed. This synapse transmits electrical signals in the form of neuronal action potentials, which have the ability to stimulate the muscle tissue to achieve contraction. This process is rather complex and involves an interplay of ionic channels and neurotransmitters. In brief, when the voltage of the action potential goes above a certain threshold, voltage-gated calcium channels at the membrane of muscle cells open up, causing an influx of Ca^{2+} into the sarcoplasmic reticulum. Ca^{2+} ions bind to the protein Troponin C, which changes conformation and allows the sarcomere (the smallest unit of skeletal muscle tissue) to contract or relax.

Many bio-hybrid robotic devices that utilize skeletal muscle tissue currently do not incorporate motor neurons with neuromuscular junctions. Even if they do, these neurons lack the capability to autonomously control the muscle. In order to induce contractions and, potentially, achieve some motion, we need to mimic the later stages of this stimulation process, i.e., the induction of a signal that can mimic neuronal action potentials and open up the Ca^{2+} channels to induce the contraction. This can be done through *electrical pulse stimulation* (EPS).

Through EPS, electric pulses of a similar shape to action potentials are applied at a distance to stimulate the tissue. Generally, squared pulses of a small width (from a few milliseconds to tens of milliseconds) are applied. Although the shape of an action potential is not squared, but rather a spike with a faster increase and a lower decrease, this is generally not important, as the aim is to go beyond the stimulation threshold to cause Ca^{2+} release. As EPS is generally performed at a distance, it bears the disadvantage that local stimulation of the tissue is not possible and thus the bio-actuator must be designed with this consideration in mind. Nevertheless, it has the advantage of being non-invasive and very easy to apply. To solve this issue, microelectrodes can be embedded in the bio-robot structure to achieve local stimulation, but this is a rather complex process [41].

The most straightforward way to apply EPS is through two electrodes located at a certain distance in the Petri dish, or culture chamber, where the bio-actuator based on muscle tissue is located. For this type of stimulation, it is important that the tissue is fully immersed in the culture medium, not only to provide all the nutrients to the tissue but also because an electrolyte medium is needed to the electric current to travel to the muscle tissue. In order to use EPS effectively, it is important to carefully control the strength and duration of the electric pulses. The pulses must be strong enough to trigger the opening of the voltage-gated calcium channels, but not so strong that they damage the muscle tissue. Additionally, the pulses must be of the appropriate duration in order to produce the desired level of muscle contraction.

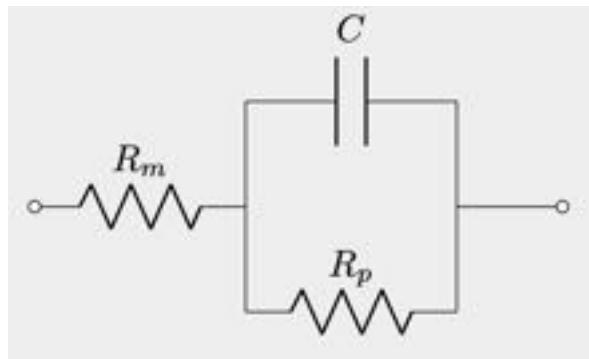


Figure 39.7: Equivalent circuit for the interplay between electrodes and culture medium. R_m is the resistance of the medium, R_p is the polarization resistance and C is the double-layer capacitance formed at the interface of the electrode with the medium. See Example 39.1. [Agnese: Redraw with consistent notation. Ask Maria to add the biological equivalent in the same picture.](#)

There are several factors that need to be considered when designing an EPS system: the material of the electrodes and its interaction with the surrounding medium and the characteristics of the electric pulse (width, amplitude and frequency). The electrolyte medium has a certain resistance to current, which is generally very low, which we can call R_m . At the interface of the electrode with the medium, charge transfer can occur between the electrons in the electrodes and the ions in the physiological medium [46]. There are three main mechanisms for charge transfer here: (i) capacitive/non-Faradaic charge transfer of the double layer formed at the electrodes; (ii) reversible Faradaic reactions; and (iii) irreversible Faradaic reactions [47, 46]. The latter are to be avoided; for instance, electrolysis produces oxidation at the anode and reduction at the cathode, which can produce bubbles, generate byproducts that are toxic for the cells, and chlorine gas, which can be dangerous for the researcher to inhale. Different steps can be taken to minimize these risks. The shape of the signal, as previously mentioned, should be carefully designed to mimic the desired conditions without being too long or producing too high a voltage. The longer the EPS pulses are maintained, the most corroding these electrochemical reactions will be, so it is imperative that the pulses are not maintained for longer than needed. Action potentials for the activation of muscle tissue are usually in the 1 to 5 ms range and it has been proven experimentally that 1 ms pulses are sufficient to generate strong contractions all over the muscle [48]. Square pulses of 2 ms, sufficiently spaced (i.e., at low frequencies) leave time for the majority ($\sim 85\%$) of reactions happening at the interface to be reversed [49]. To further minimize the amount of chemical reactions, however, biphasic pulses can be applied [47, 27], so that the first two types of reactions mentioned above, namely capacitive charge redistribution (non-Faradaic) and reversible Faradaic reactions, can be reversed, decreasing the amount of potential damage to the electrodes. Nevertheless, if the stimulation time is not too long, monophasic pulses of appropriate length are good enough to avoid corrosion and do not produce too many byproducts which are damaging to the cells [47].

The corrosion of the electrodes is heavily dependent on the materials. Carbon (graphite) and platinum are two of the most commonly used materials for cell and tissue stimulation [47, 27, 50, 48, 51, 28]. Both materials are biocompatible and resistant to corrosion. Whereas carbon materials have a lower electrical conductivity than platinum, they have a superior electrochemical stability compared to platinum and are radically cheaper to obtain [52]. Platinum can be purchased as thin but brittle wires, while carbon electrodes can be purchased as carbon rods, rather thick and easy to work with and assemble. Platinum electrodes are still the material of choice for many researchers [27, 51], especially if they are to be deposited in 2D layers [48], but special care should be taken with the duration of the pulse and the electrochemical reactions happening at the interface. Carbon electrodes, however, are generally accepted as the best material for EPS of cells and tissues in stimulation chambers or Petri dishes, due to their resistance to chemical reactions and corrosion, low price, high availability, passive biocompatibility, and superior charge injection characteristics [47].

Example 39.1: Equivalent circuit for EPS. The interplay of reactions between electrodes and culture

medium can be modeled using an equivalent circuit like the one in Fig. 39.7, where R_m is the resistance of the medium, R_p is the polarization resistance and C is the double-layer capacitance formed at the interface of the electrode with the medium. The polarization resistance, R_p , is associated to the Faradaic reactions happening at the electrode interface and, as already explained, some of these are reversible and some are not. The non-Faradaic charge redistribution at the interface is modeled by the capacitor placed in this circuit in parallel. The goal is to minimize the current passing through R_p resistance, as this means that the amount of Faradaic reactions is minimized. This depends on the frequency of the signal. At the beginning of the stimulation (the first few nanoseconds of the pulse), the main mechanism of charge transfer is non-Faradaic charge redistribution, or the formation of the double layer capacitance. In the equivalent circuit shown in Fig. 39.7, this means that all the current passes through the capacitor. If the pulse is long, the capacitor starts to get charged (the double layer is formed) and the current through that non-Faradaic mechanism decreases until reaching zero, when the capacitor is fully charged. At this point, all of the charge transfer occurs through Faradaic reactions and electrode corrosion becomes an important issue. It is therefore crucial to be aware of the equivalent capacitance of the circuit, so that the pulse can be kept small enough for charge redistribution to be the main charge transfer mechanism.

The amplitude of the electric signal is another important parameter, and it is closely correlated to the distance between electrodes. In physiological conditions, we can find ranges of electric field magnitude up to 1 V mm^{-1} , which has been proven to provide enough energy to power skeletal muscle tissue [47, 28, 53, 7]. This means that, if the culture chamber or Petri dish is very wide and the electrodes are separated a longer distance, a higher output voltage is needed so that the electric field magnitude of 1 V mm^{-1} is maintained. If the electrodes are close together, a small voltage will be needed to achieve the same value.

The stimulation frequency is probably the most flexible of the signal characteristics, as it depends on the goals of the experiment. Still, it needs to be chosen with care. Very high frequencies, even if the pulse is short, can lead to faster corrosion of the electrodes and not enough time to reverse some of the reversible reactions happening at the electrode interface. Such a high-frequency stimulation for a long period of time can result in tissue damage due to the release of byproducts but also due to too much activity and rupture of sarcomeres [54]. Low-frequency stimulations are always more desirable, even if they are to be applied for a longer time, but, of course, they will impact the motion or actuation of the bio-hybrid robot. Another important point to consider with high-frequency stimulation is the concept of wave summation. If stimulation happens too quickly before the muscle has had time to relax, the individual contractile peaks become stronger with each stimulation until achieving a saturation point. If the frequency increases to a point where relaxation is not even possible, tetanic contractions take place. This happens when the tension generated is maintained for a long time without relaxing and should be generally avoided for long periods of time, as it can damage the tissue.

39.4 Force measurement

One of the best ways to characterize the performance of the muscle tissue prior to assembly in the bio-robotic structure is to measure its force. The output force of the muscle tissue upon stimulation depends on many parameters, namely the maturation of the tissue, the number of myofibers within it, its width, the nature of the cells, the bio-materials used to encapsulate the cells, and prior stimulation protocols (i.e., if it has been *trained* to increase its force). There is no magic recipe to obtain the strongest muscle tissue, but some guidelines can be provided.

Muscle tissue differentiated for about 4 days can already produce contractions upon EPS, although these tend to be rather weak and it is not recommended to try and achieve actuation or motion at this stage. The most common range of differentiation days in the literature goes from 7 to 14 days [48, 27, 51], although the muscle tissue could be viable for even longer [55]. The width of the tissue construct also impacts its force output. Whereas common reasoning would tell us that the wider the tissue, the more cells inside it and, hence, the stronger the force output, the oxygen and nutrient diffusion limit (of about $200 \mu\text{m}$) needs to be considered [56]. Tissue constructs much wider than this limit will generally show poor cell viability and differentiation at its center [48] and, therefore, the force per unit area will decrease when the width increases. To avoid this effect as much as possible, it is recommendable to keep muscle tissue constructs in the range of 200 to $500 \mu\text{m}$, while using several independent muscle units to achieve larger force outputs.

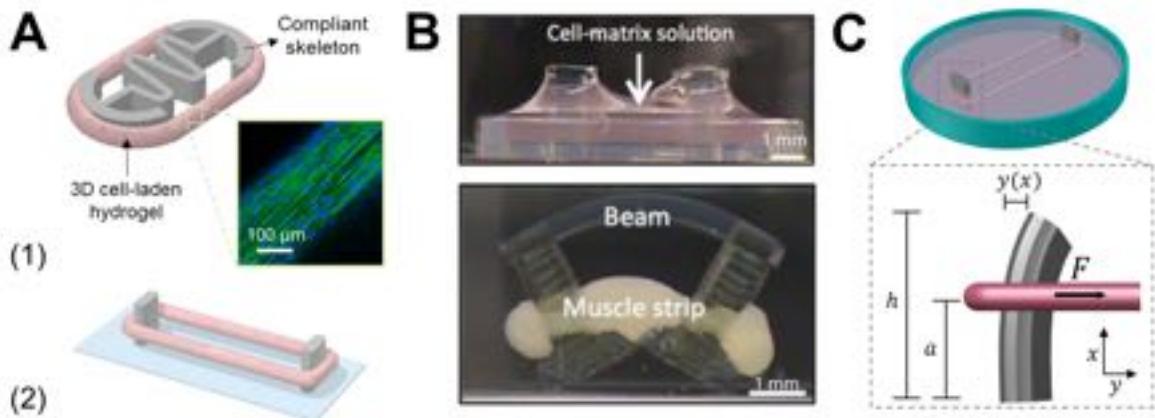


Figure 39.8: Force measurement of bio-hybrid robot actuation. (a) A two-post system to construct a tethered bio-actuator and (b) a tethered bio-robot based on the same principle with a flexible untethered skeleton (reproduced from Ref. [28]). (c) Fabrication process of a bio-hybrid walker based on deflection of a bridge-like structure (reproduced from Ref. [27]). (d) The two-post system in (a) can be used as a force measurement platform by measuring the displacement of the post, $y(x)$, upon electrical stimulation. **Agnese:** uniform to book style. Use letters a b c d (not a1 and a2). **IMPORTANT:** a1 and a2 (now a and b) should be inverted — check with Maria.

The nature of the cells heavily impacts the muscle's performance as well. Immortalized cell lines, like C2C12 myoblasts, are the cells of choice for many researchers due to their availability, price, and ease of use. Primary cell lines extracted from animals like neonatal rats tend to achieve much larger force outputs [30] of about 5 to 10 mN, whereas C2C12-based bio-robots can be in the range of 20 to 1000 mN [57, 27, 48, 28, 51]. However, primary cell lines need to be extracted from live animals, which poses some ethical concerns. Finally, the choice of bio-materials can also influence the final performance. Hydrogels based on gelatin and fibrinogen are common in the literature, as fibrinogen provides a suitable environment with cell attachment sites to ensure cell proliferation and maturation. However, hydrogels based on Matrigel, collagen or decellularized matrices [58] provide a much better environment for these cells. Disadvantages are that they are difficult to mold in 3D shapes and are difficult to 3D-bioprint [59], so the design of the bio-hybrid robot needs to consider that injection molds might be needed for the biological fabrication step.

One of the most straightforward ways to measure the force output of the tissue as a whole is through the deflection of posts, using simple equations to transform the post displacement into a force output. Fig. 39.8a shows an example where a tissue ring made of a bio-ink of gelatin and fibrinogen (with C2C12 cells) has been 3D-bioprinted concurrently with two posts made of PDMS. Once the muscle tissue construct has started differentiation, cell remodeling causes tissue shrinkage, as already discussed in Section 39.2. The two PDMS posts stop this shrinkage and provide two anchor points that generate tension in the tissue ring and assist during maturation [50]. This effect can also help fabricate bio-bots with design features that require a certain tension to create the desired conformation, as can be also seen in Fig. 39.8a and Fig. 39.8b [28, 27]. When applying EPS or other type of stimulation, the tissue will contract, pushing from the posts and creating a small displacement, $y(x)$, which can be measured with conventional bright-field microscopy and converted into a force, F (Fig. 39.8c).

In order to obtain an analytical formula to transform displacement values into force measurements, we can use Euler-Bernoulli's beam bending theory [60] to estimate the tension applied by the tissue to the posts. If we approximate the tension to be a singular force concentrated on a point at a height a from the base of the post depicted in Fig. 39.8d, beam bending equations tell us that, for a post of height h , the displacement has the following dependence:

$$y(x) = \begin{cases} \frac{Fx^2}{6EI}(3a-x) & \text{for } 0 < x < a, \\ \frac{Fa^2}{6EI}(3x-a) & \text{for } a < x < h, \end{cases} \quad (39.1)$$

where F is the force applied along the y -axis, E is the Young's modulus of the material (in this case, PDMS), I is the second moment of area (see Example 39.2, Eq. (39.4) for a definition), a is the height at which the tissue is pulling from, and $y(x)$ is the displacement of the tissue at height x . If we assume that the displacement of the post is measured at the same position where the tissue is pulling, i.e., at $x = a$, then we can simplify the formula to

$$F = \frac{3EIy(a)}{a^3}. \quad (39.2)$$

The second moment of area for this geometry [Which geometry? A pillar? Please, specify] is calculated as

$$I = \frac{w^3 h}{12}, \quad (39.3)$$

where w is the width and h its height. The lateral width along the z -axis does not come into play in this equation, as the displacement happens in the xy plane.

Example 39.2: Second moment of area. The example from Fig. 39.8d is a simple flat slab, or a three-dimensional rectangular shape where w is its width, L is lateral width, and h is its height. The second moment of area, often referred to as the moment of inertia, is a geometric property of an area that indicates how its points are spread out relative to an axis. The second moment of area along the z -axis is defined as [60]:

$$I_z = \int \int y^2 dx dz. \quad (39.4)$$

Example 39.3: Application of Euler-Bernoulli's beam bending theory. Eq. (39.2) gives the deflection of the post, $y(x)$, when a force F is applied on its wider side at a height a (Fig. 39.8c). The Euler-Bernoulli equation [60] can be written as

$$\frac{d^2}{dx^2} \left(EI \frac{d^2 y}{dx^2} \right) = q, \quad (39.5)$$

where q is the distributed load. Since the product of EI is constant, as we assume that the post is homogeneous and the applied forces rather small, this equation can be simplified to

$$EI \frac{d^4 y}{dx^4} = q(x), \quad (39.6)$$

and, subsequently, to:

$$EI \frac{d^2 y}{dx^2} = M(x), \quad (39.7)$$

where $M(x)$ is the bending moment of the post. In our highly symmetric case, the bending moment is simply $M(x) = -Fx$, and the equation finally becomes

$$EI \frac{d^2 y}{dx^2} = -Fx. \quad (39.8)$$

This derivation assumes that the displacements are small and we do not need to consider the viscoelastic beam deformation of the PDMS material. For large deflections, the derivation becomes more complex [61, 62].

39.5 Unique capabilities and main challenges

Although bio-hybrid robots present constraints in the integration of biological elements in living platforms and their later manipulation, the advantages of their use are clear. Compared to their artificial soft-robotic

counterparts, they present exceptional adaptability to the environment, being able to evolve over time, as, for example, by undergoing controlled training protocols [48, 28]. Therefore, when designing a living robot, we are not just restricted to their initial properties, but we can envision improved performance over time by applying some of the reported training, either electrical or mechanical, or even by including certain nanomaterials within the cell-laden hydrogel in order to improve force output [63, 64].

Apart from the capacity to respond to external stimuli, as many other artificial materials in soft robots do to perform a certain task [65], bio-hybrid robots also present other characteristics that make them more advanced. For example, skeletal-muscle-based robots have shown self-healing capabilities: When subjected to mechanical damage during *in vitro* studies, these robots rapidly healed, taking advantage of the natural healing strategies observed in living muscle tissue [66]. Self-healing processes were also shown in xenobots, where robust repair was demonstrated after mechanical damage, as well as self-organization into different living forms, resulting into ciliated epithelium when cultured in certain medium conditions [38]. Also, there are other features that could be exploited in the future, such as the communication ability between biological components, like cell-to-cell communication by using chemical signals [67] or other mechanical cues [68]. Indeed, this topic is of such significance that a whole new community based on artificial cell trials is trying to elucidate how cell communication and signaling work to understand and exploit them [69]. So far, the field is still in its infancy, trying to understand how to create more efficient living robots by using different means, from advanced microscopy techniques to force measurement setups, as those presented in this Chapter.

There is a clear need to work towards developing fabrication methods to obtain bio-hybrid robotic platforms with robust and reproducible performance in order to implement them in future biomedical applications. Inner variability coming from the biological nature of the actuator should also be accounted for. Furthermore, for muscle-based actuators, there are several challenges anticipated in the future related to scaling, whether it is upscaling or downscaling, compared to the mesoscale designs currently in development. From one side, when scaling them up [70], it will be key to account not only for the vascular system of native muscle, but also for nutrient diffusion and efficient oxygenation. The bridging of functional components in such bio-hybrid robots is key, similar to the synaptic connection between motor nerve and muscle (neuromuscular junction) in the native system. Also, it is important to take into consideration the cells immediate environment to account for several multi-scale interactions, such as physical, chemical, and other biological cues that could impact on the bio-hybrid robotic system. On the other hand, when scaling down living systems, it will be important not only to consider the minimal cell size/biological compound, but also if there is a minimum cell number or if several cell types should be present for a proper tissue maturation (i.e., co-culture). Lastly, it's crucial to note that the majority of the living systems developed so far primarily operate in wet conditions, with stringent requirements regarding temperature, CO₂ levels, and nutrient availability. This challenge has been partially addressed in some muscle-based actuators by integrating an outer collagen structure into a skeletal muscle actuator, which extended its lifespan [71]. The longevity was further enhanced by incorporating perfusable tubes within the collagen-based framework, facilitating actuation while supplying the cells with optimal culturing conditions. A cryopreservation method to allow the storage of functional muscle-based actuators for an extended time frame was also developed, demonstrating that they could be frozen and revived on demand to prolong their lifetime [55]. Therefore, the design of future platforms will require the integration of advanced bioreactors, ideally miniaturized, that will not only provide for basic cell maintenance, but also avoid any contamination source from the environment.

39.6 Ethical concerns

The development of bio-hybrid robots, which are made using muscle tissue from animals, spark a number of ethical and societal concerns. Many are shared with classical robotics and artificial intelligence [72], but others are inherent to the unique characteristics of this technology. Some of these concerns include the use of animal sources, the unknown effects on the environment, the potential health effects on humans, the possibility of creating sentient beings, and the impact of this technology on society. In this section, we review some of these ethical and societal implications and call for responsible research and innovation on this technology.

One of the key ethical concerns surrounding the development of bio-hybrid robots is the possibility of creating sentient beings. Because these robots are made using living tissue, they have the potential to develop complex behaviors and abilities that are similar to those of animals and humans. Some people might argue that any being that has consciousness and the ability to experience emotions should be treated with respect and

dignity, regardless of whether it is a human, an animal, or a bio-hybrid robot. This view is based on the idea that all sentient beings have inherent value, and that it is unethical to harm or exploit them. This belief makes us ask ourselves: Where do we draw the line between a group of cells and a sentient being?

Consciousness is a complex and multi-dimensional phenomenon that is not fully understood, and there is no consensus on how to measure or assess it [73], but this does not stop us from reflecting on this issue. In any case, not everyone agrees with the previous view that any entity capable of experiencing consciousness should be treated with the same level of respect than a human being or an animal. Some people argue that only human beings have inherent value, and that it is acceptable to use animals and other non-human beings for scientific research and other purposes. This view is often based on the belief that non-human beings are inferior to humans and do not have the same moral status.

Closely related to this issue is the use of animal sources for the fabrication of bio-hybrid robots that are not aimed at biomedical research. While some people believe that using animals for research is unethical, others argue that it is acceptable if it is for the purpose of curing diseases. In the creation of three-dimensional tissue, researchers can use a range of animal-derived elements, such as cells, serum, and proteins. To address these concerns, some researchers are now exploring the use of xeno-free materials, supplements, or media in tissue engineering research [74]. These components are not derived from animals and do not contain any animal-derived constituents. They are typically synthetic and mimic the properties of animal-derived materials, and are seen as a way to reduce ethical concerns associated with the use of animals in tissue engineering research, as well as a way to reduce other disadvantages, such as lot-to-lot variability or potential microbiological contamination [75]. However, these materials may not have the same biological properties as animal-derived materials, and their development can be more expensive and time-consuming. Despite these challenges, many researchers believe that the use of xeno-free materials is important for more ethical and responsible tissue engineering research. In the case of bio-hybrid robotics research, the use of animal-derived cells is necessary (otherwise they would not be bio-hybrid), but the impact on animal welfare can be minimized in different ways. For instance, researchers using freshly extracted animal muscle tissue could switch to the use of long-established cell lines to reduce their impact on animals, as well as investigate how their biomaterials, serum, and supplements could be replaced by xeno-free alternatives. However, it is known that immortalized cells tend to be weaker than freshly extracted ones, so this poses a trade-off to consider.

At the application level, the use of bio-hybrid robots in certain environments could have unintended consequences. For example, if bio-hybrid robots were released in marine environments to clean oceans, as it has recently been proposed [76], they could potentially compete with wild animals for resources, leading to changes in the local ecosystem. This could have negative impacts on the environment and on the animals that live in our oceans. In order to prevent these negative impacts, it is important for researchers to carefully consider the potential environmental consequences of using bio-hybrid robots in different settings. Furthermore, if large-scale fabrication of these bio-hybrid robots takes place, it is important to consider the negative environmental impacts of the use of materials and energy during their fabrication. In order to minimize these environmental impacts, it is important for researchers to use sustainable materials and production processes in the creation of bio-hybrid robots. Lastly, one of the major concerns about the environmental dangers of bio-hybrid robot research is the potential for the spread of diseases. Because bio-hybrid robots are made using living tissue, there is a risk that they could carry and transmit diseases from the animals from which they are made. This could lead to the spread of diseases to humans, animals, and the environment.

The development of bio-hybrid robots could have significant implications for society. These robots have the potential to blur the line between humans and machines, leading to a shift in societal attitudes towards technology and its role in society. Bio-hybrid robots made with living tissue have been considered for a range of medical applications, such as drug delivery, tissue repair, and microsurgery [77]. For instance, a bio-hybrid arm could replace a human arm, providing mechanical support and movement, as well as some characteristics of living beings like the ability to respond to stimuli and self-repair. This could benefit patients who have lost their arms, but raises ethical concerns like those mentioned above. It could also trigger a range of psychological effects related to having a bio-hybrid arm, ranging from alienation to discomfort, which could impact people's wellbeing. The development of bio-hybrid robots also has implications for *transhumanism*, the philosophical movement exploring the use of technology to enhance human abilities and extend human life. Bio-hybrid robots could be seen as a step towards achieving the goals of transhumanism through the integration of biological and technological systems. Finally, the societal impact of the cost of this technology should also be assessed. If the cost of bio-hybrid robots is too prohibitive, it is likely that they will only be accessible to a limited number

of people. This could create a situation where only certain individuals are able to benefit from the technology, while others are unable to access it due to financial constraints, leading to further societal inequality.

In this section, we have not attempted to provide easy answers to the complex questions raised by bio-hybrid robotics research. Instead, our aim has been to increase awareness of these issues and to provide researchers with information about the potential consequences of their work. Bio-hybrid robotics is a field that has many different potential impacts and, by understanding these impacts, researchers can make informed decisions about their research and its potential consequences. It is important for researchers to adhere to the concept of responsible research and innovation (RRI), which involves conducting research in a way that is ethical, sustainable, and accountable, taking into account the potential impacts of research on society and the environment, and ensuring that research is conducted in a way that is transparent.

To support responsible research and innovation, it is important for researchers to have access to a framework or guideline that can help them conduct their work in an ethical and responsible manner. This framework could provide guidance on issues such as the ethical use of animals in research, the protection of human subjects, and the environmental impacts of research. It can also provide guidance on how to ensure that research is conducted in a transparent and accountable way, and how to engage with stakeholders such as the public and policy makers. This framework should be developed in collaboration with researchers, ethicists, and other stakeholders, and should take into account the ethical and societal implications of this research. It should also be flexible enough to adapt to new developments and changing societal attitudes. All in all, it should allow researchers to make informed decisions about their work, and to ensure that their research has a positive impact on society and the environment.

Finally, it is important that the public is involved in the development of this technology, and that they have a say in how it is used and regulated. This will ensure that the technology is developed in a responsible and ethical manner, and that it benefits society as a whole. Involving the public in the development of this technology could also help ensure that it is used in a way that is safe and beneficial for humans. For instance, the public could be consulted on the potential risks and benefits of using bio-hybrid robots for medical purposes, such as drug delivery or tissue repair. Furthermore, involving the public in the development of bio-hybrid robots could also help ensure that the technology is used in an equitable and inclusive manner. By involving the public in the decision-making processes, it is possible to ensure that the technology is not only accessible to those who can afford it, but also that it is designed to meet the needs of diverse individuals and communities.

39.7 Conclusions

Although the development of living robotics is rather recent, bio-hybrid platforms have already demonstrated their potential by highlighting how including biological components into soft robotic systems can enhance their capabilities. Their unique properties, from their compliance to their adaptability to their capacity to self-regenerate and self-organize, open the door to a new technology of special interest in the biomedical field. Indeed, these bio-hybrid systems hold promise not only for advancing robotic platforms but also for offering valuable insights into underlying biological processes. By helping to reverse engineer these processes, they can contribute to the development of micro-physiological platforms, serving as crucial clinical models for personalized medicine.

However, the use of living components comes with many challenges, mostly related to the robustness and lifetime of the biological actuator in the system. More specifically, for muscle-based living robotics, it is essential to provide a proper cell environment (i.e., the nutrients availability, temperature, CO₂), as well as to use biocompatible hydrogels that provide for attachment sites and aid nutrient diffusion. To these basic conditions, we should add a set of conditions related to the application of the bio-hybrid actuator in wet conditions, which implies the use of bioreactors to ensure cell viability. In fact, it will be also essential to consider further challenges when scaling-up such living robots, such as system vascularization and interfacing with other artificial structures, such as flexible electronics useful both for advanced local actuation and feedback control purposes for their future automation.

Discoveries in future living machines when merging advanced materials to well-known biological entities are still to be unveiled, as emerging behavior and self-organization in biological systems are still under study. However, it is important to account not only for the technical challenges, but also for the legal and ethical considerations when working with living materials. Overall, this emergent field is not only of importance for

the soft robotic community, but it has a direct impact on the development of relevant clinical models of special interest as drug delivery testing platforms.

39.8 Problems

Problem 39.1: Derivation of the deflection of the post equation.

- a. Resolve the second-order differential equation, Eq. (39.8), assuming that the force is applied at the tip of the post, i.e., at a height $h = a$. [Hint: You will need to apply appropriate boundary conditions, so choose your origin of coordinates wisely.]
- b. The equation you obtain is not exactly Eq. (39.1). That equation considers that the force is not being applied at the tip of the post (the tissue would slip out), but at another point $a < h$. Repeat the derivation, but considering that the load is applied at that point.
- c. What would be the displacement of a post with dimensions $L = 2\text{ mm}$, $w = 0.5\text{ mm}$, $h = 3\text{ mm}$, and Young's modulus of $E = 200\text{ kPa}$, when the tissue is pushing with a force of $F = 50\text{ }\mu\text{N}$ at a height of $a = 2\text{ mm}$?
- d. What should the minimum image resolution of our system be in $\mu\text{m p}x^{-1}$ if we want a force sensitivity of $\Delta F = 5\text{ }\mu\text{N}$?

Problem 39.2: Design of efficient EPS protocols. Consider the electric circuit from Example 39.1.

- a. Derive the electrical impedance Z of this circuit, defined as $Z = R + iX$, where R is the resistance and X is the reactance.
- b. The impedance depends on the frequency of the signal you are applying as well as on the values of R_m , R_p , and C . What happens when applying a direct current (DC) or, more precisely, a square pulse with a long width? Which component dominates the impedance, and what does it mean? [Hint: A very long pulse is equivalent to a signal with very low frequencies. You can take the limit of the expression of the electrical impedance Z for the frequency of the signal going to 0.]
- c. What happens when applying very short-lived pulses? Which component dominates and what does it mean? [Hint: a very short pulse is equivalent to very high frequencies. You can take a limit.]
- d. The double layer capacitance is more precisely modeled with what is known as a *constant phase element* (CPE) (see, for instance, Ref. [78]), which represents an imperfect capacitor. Read about the CPE and re-derive the impedance of the equivalent circuit considering a CPE instead of a perfect capacitor.

Problem 39.3: Derivation of the second moment of area.

- a. Calculate the integral of Eq. (39.4) to obtain the result of Eq. (39.3).
- b. The moment of inertia I remains unaffected by the depth L of the post along the z -direction. Given this, would a cylindrical post, with a projection on the xy -plane identical to that of a flat slab, have the same moment of inertia?

Bibliography

- [1] S. Kim, C. Laschi, and B. Trimmer. Soft robotics: a bioinspired evolution in robotics. *Trends in Biotechnology*, 31:287–294, 2013.
- [2] R. Wang, C. Zhang, Y. Zhang, W. Tan, W. Chen, and L. Liu. Soft underwater swimming robots based on artificial muscle. *Advanced Materials Technologies*, page 2200962, 2022.
- [3] B. Behkam and M. Sitti. Design methodology for biomimetic propulsion of miniature swimming robots. *Journal of Dynamic Systems, Measurement, and Control*, 128:36–43, 2006.
- [4] R. Raman and R. Bashir. Biomimicry, biofabrication, and biohybrid systems: The emergence and evolution of biological design. *Advanced Healthcare Materials*, 6:1700496, 2017.
- [5] C. Appiah, C. Arndt, K. Siemsen, A. Heitmann, A. Staubitz, and C. Selhuber-Unkel. Living materials herald a new era in soft robotics. *Advanced Materials*, 31:1807747, 2019.

- [6] V. A. Webster-Wood, M. Guix, N. W. Xu, B. Behkam, H. Sato, D. Sarkar, S. Sanchez, M. Shimizu, and K. K. Parker. Biohybrid robots: recent progress, challenges, and perspectives. *Bioinspiration & Biomimetics*, 18:015001, 2023.
- [7] R. Mestre, T. Patiño, and S. Sánchez. Biohybrid robotics: From the nanoscale to the macroscale. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 13:1–26, 2021.
- [8] Y. Morimoto and S. Takeuchi. Biohybrid robot powered by muscle tissues. In *Mechanically Responsive Materials for Soft Robotics*, pages 395–416. Wiley, 2020.
- [9] L. Ricotti and T. Fujie. Thin polymeric films for building biohybrid microrobots. *Bioinspiration & Biomimetics*, 12:021001, 2017.
- [10] S. Ostrovidov, S. Salehi, M. Costantini, K. Suthiwanich, M. Ebrahimi, R. B. Sadeghian, T. Fujie, X. Shi, S. Cannata, C. Gargioli, A. Tamayol, M. R. Dokmeci, G. Orive, W. Swieszkowski, and A. Khademhosseini. 3D bioprinting in skeletal muscle tissue engineering. *Small*, 15:1805530, 2019.
- [11] X. Arqué, T. Patiño, and S. Sánchez. Correction: Enzyme-powered micro- and nano-motors: key parameters for an application-oriented design. *Chemical Science*, 13:9784–9786, 2022.
- [12] L. Schwarz, M. Medina-Sánchez, and O. G. Schmidt. Sperm-hybrid micromotors: on-board assistance for nature’s bustling swimmers. *Reproduction*, 159:R83–R96, 2020.
- [13] H. D. Nguyen, P. Z. Tan, H. Sato, and T. T. Vo-Doan. Sideways walking control of a cyborg beetle. *IEEE Transactions on Medical Robotics and Bionics*, 2:331–337, 2020.
- [14] N. W. Xu and J. O. Dabiri. Low-power microelectronics embedded in live jellyfish enhance propulsion. *Science Advances*, 6, 2020.
- [15] P. Won, S. H. Ko, C. Majidi, A. W. Feinberg, and V. A. Webster-Wood. Biohybrid actuators for soft robotics: Challenges in scaling up. *Actuators*, 9:96, 2020.
- [16] L. Ricotti and A. Menciassi. Bio-hybrid muscle cell-based actuators. *Biomedical Microdevices*, 14:987–998, 2012.
- [17] S. Kriegman, D. Blackiston, M. Levin, and J. Bongard. A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences*, 117:1853–1859, 2020.
- [18] Y. Yalikun, K. Uesugi, M. Hiroki, Y. Shen, Y. Tanaka, Y. Akiyama, and K. Morishima. Insect muscular tissue-powered swimming robot. *Actuators*, 8:30, 2019.
- [19] K. Uesugi, K. Shimizu, Y. Akiyama, T. Hoshino, K. Iwabuchi, and K. Morishima. Contractile performance and controllability of insect muscle-powered bioactuator with different stimulation strategies for soft robotics. *Soft Robotics*, 3:13–22, 2016.
- [20] O. Aydin, X. Zhang, S. Nuethong, G. J. Pagan-Diaz, R. Bashir, M. Gazzola, and M. T. A. Saif. Neuromuscular actuation of biohybrid motile bots. *Proceedings of the National Academy of Sciences*, 116:19841–19847, 2019.
- [21] A. R. Gillies and R. L. Lieber. Structure and function of the skeletal muscle extracellular matrix. *Muscle & Nerve*, 44:318–331, 2011.
- [22] J. Lamsfuss and S. Bargmann. Skeletal muscle: Modeling the mechanical behavior by taking the hierarchical microstructure into account. *Journal of the Mechanical Behavior of Biomedical Materials*, 122:104670, 2021.
- [23] Y. Jin, E. J. Jeon, S. Jeong, S. Min, Y. S. Choi, S. H. Kim, J. S. Lee, J. Shin, J. H. Yu, D.-H. Ahn, Y.-G. Kim, H. S. Yang, T. J. Kang, S.-R. Cho, N. Choi, and S.-W. Cho. Reconstruction of muscle fascicle-like tissues by anisotropic 3D patterning. *Advanced Functional Materials*, 31:2006227, 2021.

- [24] M. Scalise, F. Marino, L. Salerno, E. Cianflone, C. Molinaro, N. Salerno, A. De Angelis, G. Viglietto, K. Urbanek, and D. Torella. From spheroids to organoids: The next generation of model systems of human cardiac regeneration in a dish. *International Journal of Molecular Sciences*, 22:13180, 2021.
- [25] J. Xi, J. J. Schmidt, and C. D. Montemagno. Self-assembled microdevices driven by muscle. *Nature Materials*, 4:180–184, 2005.
- [26] S.-J. Park, M. Gazzola, K. S. Park, S. Park, V. Di Santo, E. L. Blevins, J. U. Lind, P. H. Campbell, S. Dauth, A. K. Capulli, F. S. Pasqualini, S. Ahn, A. Cho, H. Yuan, B. M. Maoz, R. Vijaykumar, J.-W. Choi, K. Deisseroth, G. V. Lauder, L. Mahadevan, and K. K. Parker. Phototactic guidance of a tissue-engineered soft-robotic ray. *Science*, 353:158–162, 2016.
- [27] C. Cvetkovic, R. Raman, V. Chan, B. J Williams, M. Tolish, P. Bajaj, M. S. Sakar, H. H. Asada, M. T. A. Saif, and R. Bashir. Three-dimensionally printed biological machines powered by skeletal muscle. *Proceedings of the National Academy of Sciences of the United States of America*, 111:10125–30, 2014.
- [28] M. Guix, R. Mestre, T. Patiño, M. De Corato, J. Fuentes, G. Zarpellon, and S. Sánchez. Biohybrid soft robots with self-stimulating skeletons. *Science Robotics*, 6:eabe7577, 2021.
- [29] V. A. Webster-Wood, O. Akkus, U. A. Gurkan, H. J. Chiel, and R. D. Quinn. Organismal engineering: Toward a robotic taxonomic key for devices using organic materials. *Science Robotics*, 2, 2017.
- [30] Y. Morimoto, H. Onoe, and S. Takeuchi. Biohybrid robot powered by an antagonistic pair of skeletal muscle tissues. *Science Robotics*, 3:eaat4440, 2018.
- [31] K. Y. Lee, S.-J. Park, D. G. Matthews, S. L. Kim, C. A. Marquez, J. F. Zimmerman, H. A. M. Ardoña, A. G. Kleber, G. V. Lauder, and K. K. Parker. An autonomously swimming biohybrid fish designed with human cardiac biophysics. *Science*, 375:639–647, 2022.
- [32] K. Althoefer. Antagonistic actuation and stiffness control in soft inflatable robots. *Nature Reviews Materials*, 3:76–77, 2018.
- [33] B. Kalita, A. Leonessa, and S. K. Dwivedy. A review on the development of pneumatic artificial muscle actuators: Force model and application. *Actuators*, 11:288, 2022.
- [34] J. Wang, X. Zhang, J. Park, I. Park, E. Kilicarslan, Y. Kim, Z. Dou, R. Bashir, and M. Gazzola. Computationally assisted design and selection of maneuverable biological walking machines. *Advanced Intelligent Systems*, 3:2000237, 2021.
- [35] Y. Tanaka, K. Sato, T. Shimizu, M. Yamato, T. Okano, and T. Kitamori. A micro-spherical heart pump powered by cultured cardiomyocytes. *Lab on a Chip*, 7:207–212, 2007.
- [36] Y. Akiyama, T. Sakuma, K. Funakoshi, T. Hoshino, K. Iwabuchi, and K. Morishima. Atmospheric-operable bioactuator powered by insect muscle packaged with medium. *Lab on a Chip*, 13:4870–4880, 2013.
- [37] J. C. Nawroth, H. Lee, A. W. Feinberg, C. M. Ripplinger, M. L. McCain, A. Grosberg, J. O. Dabiri, and K. K. Parker. A tissue-engineered jellyfish with biomimetic propulsion. *Nature biotechnology*, 30:792–797, 2012.
- [38] D. Blackiston, E. Lederer, S. Kriegman, S. Garnier, J. Bongard, and M. Levin. A cellular platform for the development of synthetic living machines. *Science Robotics*, 6:eabf1571, 2021.
- [39] S. Kriegman, D. Blackiston, M. Levin, and J. Bongard. Kinematic self-replication in reconfigurable organisms. *Proceedings of the National Academy of Sciences*, 118:e2112672118, 2021.
- [40] H. Zhao, Y. Kim, H. Wang, X. Ning, C. Xu, J. Suh, M. Han, G. J. Pagan-Diaz, W. Lu, H. Li, W. Bai, O. Aydin, Y. Park, J. Wang, Y. Yao, Y. He, M. T. A. Saif, Y. Huang, R. Bashir, and J. A. Rogers. Compliant 3D frameworks instrumented with strain sensors for characterization of millimeter-scale engineered muscle tissues. *Proceedings of the National Academy of Sciences*, 118:e2100077118, 2021.

- [41] Y. Kim, G. Pagan-Diaz, L. Gapinske, Y. Kim, J. Suh, E. Solomon, J. F. Harris, S. W. Nam, and R. Bashir. Integration of graphene electrodes with 3D skeletal muscle tissue models. *Advanced healthcare materials*, 9:1901137, 2020.
- [42] H. Tetsuka, L. Pirrami, T. Wang, D. Demarchi, and S. R. Shin. Wirelessly powered 3D printed hierarchical biohybrid robots with multiscale mechanical properties. *Advanced Functional Materials*, page 2202674, 2022.
- [43] J. Wang, Yueji Wang, Y. Kim, Tianqi Yu, and R. Bashir. Multi-actuator light-controlled biological robots. *APL bioengineering*, 6:036103, 2022.
- [44] U. Kim, D. Jung, H. Jeong, J. Park, H.-M. Jung, J. Cheong, H. R. Choi, H. Do, and C. Park. Integrated linkage-driven dexterous anthropomorphic robotic hand. *Nature communications*, 12:1–13, 2021.
- [45] N. El-Atab, R. B. Mishra, F. Al-Modaf, L. Joharji, A. A. Alsharif, H. Alamoudi, M. Diaz, N. Qaiser, and M. M. Hussain. Soft actuators for soft robotic applications: a review. *Advanced Intelligent Systems*, 2:2000128, 2020.
- [46] D. R. Merrill, M. Bikson, and J. G. R. Jefferys. Electrical stimulation of excitable tissue: design of efficacious and safe protocols. *Journal of neuroscience methods*, 141:171–198, 2005.
- [47] N. Tandon, C. Cannizzaro, P.-H. G. Chao, R. Maidhof, A. Marsano, H. T. H. Au, M. Radisic, and G. Vunjak-Novakovic. Electrical stimulation systems for cardiac tissue engineering. *Nature protocols*, 4:155–173, 2009.
- [48] R. Mestre, T. Patiño, X. Barceló, S. Anand, A. Pérez-Jiménez, and S. Sánchez. Force modulation and adaptability of 3D-bioprinted biological actuators based on skeletal muscle tissue. *Advanced Materials Technologies*, 4:1800631, 2019.
- [49] N. Tandon, A. Marsano, C. Cannizzaro, J. Voldman, and G. Vunjak-Novakovic. Design of electrical stimulation bioreactors for cardiac tissue engineering. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3594–3597. IEEE, 2008.
- [50] R. Mestre, T. Patiño, X. Barceló, and S. Sanchez. 3D bioprinted muscle-based bio-actuators: force adaptability due to training. In *Conference on Biomimetic and Biohybrid Systems*, pages 316–320. Springer, 2018.
- [51] G. J. Pagan-Diaz, X. Zhang, L. Grant, Y. Kim, O. Aydin, C. Cvetkovic, E. Ko, E. Solomon, J. Hollis, H. Kong, T. Saif, M. Gazzola, and R. Bashir. Simulation and fabrication of stronger, larger, and faster walking biohybrid machines. *Advanced Functional Materials*, 28:1801145, 2018.
- [52] M. Devi, M. Vomero, E. Fuhrer, E. Castagnola, C. Gueli, S. Nimbalkar, M. Hirabayashi, S. Kassegne, T. Stieglitz, and S. Sharma. Carbon-based neural electrodes: promises and challenges. *Journal of Neural Engineering*, 2021.
- [53] R. Mestre, T. Patiño, M. Guix, X. Barceló, and S. Sanchez. Design, optimization and characterization of bio-hybrid actuators based on 3D-bioprinted skeletal muscle tissue. In *Conference on Biomimetic and Biohybrid Systems*, pages 205–215. Springer, 2019.
- [54] Z. Orfanos, M. P. O. Gödderz, E. Soroka, T. Gödderz, A. Rumyantseva, P. F. M. van der Ven, T. J. Hawke, and D. O. Fürst. Breaking sarcomeres by in vitro exercise. *Scientific reports*, 6:1–9, 2016.
- [55] L. Grant, R. Raman, C. Cvetkovic, M. C. Ferrall-Fairbanks, G. J. Pagan-Diaz, P. Hadley, E. Ko, M. O. Platt, and R. Bashir. Long-term cryopreservation and revival of tissue-engineered skeletal muscle. *Tissue Engineering Part A*, 25:1023–1036, 2019.
- [56] A. L. Farris, A. N. Rindone, and W. L. Grayson. Oxygen delivering biomaterials for tissue engineering. *Journal of materials chemistry B*, 4:3422–3432, 2016.

- [57] R. Raman, C. Cvetkovic, S. G. M. Uzel, R. J. Platt, P. Sengupta, R. D. Kamm, and R. Bashir. Optogenetic skeletal muscle-powered adaptive biological machines. *Proceedings of the National Academy of Sciences*, 113:3497–3502, 2016.
- [58] A. Urciuolo and P. De Coppi. Decellularized tissue for muscle regeneration. *International journal of molecular sciences*, 19:2392, 2018.
- [59] A. Abaci and M. Guvendiren. Designing decellularized extracellular matrix-based bioinks for 3D bio-printing. *Advanced Healthcare Materials*, 9:2000734, 2020.
- [60] A. Öchsner. *Classical Beam Theories of Structural Mechanics*. Springer Cham, 2021.
- [61] K.-J. Bathe, E. Ramm, and E. L. Wilson. Finite element formulations for large deformation dynamic analysis. *International Journal for Numerical Methods in Engineering*, 9:353–386, 1975.
- [62] K. K. Pradhan and S. Chakraverty. Chapter four - finite element method. In K. K. Pradhan and S. Chakraverty, editors, *Computational Structural Mechanics*, pages 25–28. Academic Press, 2019.
- [63] K. Zhu, S. R. Shin, T. van Kempen, Y.-C. Li, V. Ponraj, A. Nasajpour, S. Mandla, N. Hu, X. Liu, J. Leijten, Y.-D. Lin, M. A. Hussain, Y. S. Zhang, A. Tamayol, and A. Khademhosseini. Gold nanocomposite bioink for printing 3D cardiac constructs. *Advanced functional materials*, 27:1605352, 2017.
- [64] R. Mestre, J. Fuentes, L. Lefaix, J. Wang, M. Guix, G. Murillo, R. Bashir, and S. Sanchez. Improved performance of biohybrid muscle-based bio-bots doped with piezoelectric boron nitride nanotubes. *bioRxiv*, 2022.
- [65] J. Wang, F. Tang, Y. Wang, Q. Lu, S. Liu, and L. Li. Self-healing and highly stretchable gelatin hydrogel for self-powered strain sensor. *ACS applied materials & interfaces*, 12:1558–1566, 2019.
- [66] R. Raman, L. Grant, Y. Seo, C. Cvetkovic, M. Gapinske, A. Palasz, H. Dabbous, H. Kong, P. P. Pinera, and R. Bashir. Damage, healing, and remodeling in optogenetic skeletal muscle bioactuators. *Advanced healthcare materials*, 6:1700030, 2017.
- [67] W. C. De Mello. Exchange of chemical signals between cardiac cells. Fundamental role on cell communication and metabolic cooperation. *Experimental cell research*, 346:130–136, 2016.
- [68] N. V. Burov, K. Haase, and A. E. Pelling. Mechanical cues in cellular signalling and communication. *Cell and tissue research*, 352:77–94, 2013.
- [69] V. Mukwaya, S. Mann, and H. Dou. Chemical communication at the synthetic cell/living cell interface. *Communications Chemistry*, 4:1–12, 2021.
- [70] P. Won, S. H. Ko, C. Majidi, A. W. Feinberg, and V. A. Webster-Wood. Biohybrid actuators for soft robotics: Challenges in scaling up. In *Actuators*, volume 9, page 96. MDPI, 2020.
- [71] Y. Morimoto, H. Onoe, and S. Takeuchi. Biohybrid robot with skeletal muscle tissue covered with a collagen structure for moving in air. *APL bioengineering*, 4:026101, 2020.
- [72] J. Torresen. A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI*, 4:75, 2018.
- [73] A. K. Seth and T. Bayne. Theories of consciousness. *Nature Reviews Neuroscience*, pages 1–14, 2022.
- [74] N. Desai, P. Rambhia, and A. Gishto. Human embryonic stem cell cultivation: historical perspective and evolution of xeno-free culture systems. *Reproductive Biology and Endocrinology*, 13:1–15, 2015.
- [75] M. Cimino, R. M. Gonçalves, C. C. Barrias, and M. C. L. Martins. Xeno-free strategies for safe human mesenchymal stem/stromal cell expansion: supplements and coatings. *Stem cells international*, 2017, 2017.

- [76] N. W. Xu, J. P. Townsend, J. H. Costello, S. P. Colin, B. J. Gemmell, and J. O. Dabiri. Developing biohybrid robotic jellyfish (*aurelia aurita*) for free-swimming tests in the laboratory and in the field. *Bio-protocol*, 11:e3974–e3974, 2021.
- [77] M. Levin, J. Bongard, and J. E. Lunshof. Applications and ethics of computer-designed organisms. *Nature Reviews Molecular Cell Biology*, 21:655–656, 2020.
- [78] Holm S., Holm T., and Martinsen Ø. G. Simple circuit equivalents for the constant phase element. *PLoS One*, 16:e0248786, 2021.

Chapter 40

Locomotor Transitions in Complex Terrain

RATAN OTHAYOTH, QIYUAN FU, SEAN W. GART, QIHAN XUAN, YAQING WANG, CHEN LI

The study of locomotor transitions in complex terrain explores the sophisticated dynamics of movement found in nature, where animals seamlessly switch between various modes like walking, running, and climbing. Such adaptability is crucial for navigating diverse and challenging environments, yet remains only partially understood. In this Chapter, we explore some of the mechanisms behind these transitions, examining the interplay of neural, postural, and ecological factors that influence how animals and, by extension, robots navigate complex terrain. We highlight the existing gaps in our understanding, particularly in modeling and controlling locomotion in three-dimensional landscapes. The Chapter puts forward a physics-based approach, drawing from

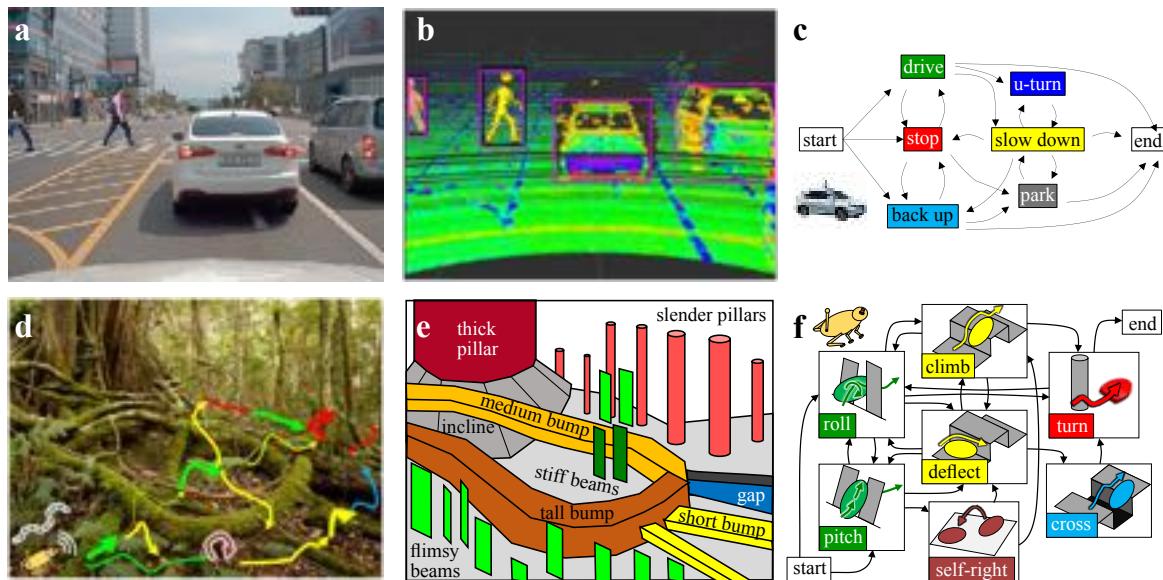


Figure 40.1: **Approaches to multi-pathway transitions to avoid and traverse obstacles.** (a) View from a self-driving car. (b) Geometric map scanned. (c) Multi-pathway driving mode transitions to avoid obstacles. (d) Envisioned capability of robots traversing complex 3D terrain with many obstacles as large as themselves. (e) Abstracted challenges from diverse large obstacles. (f) Envisioned multi-pathway locomotor mode transitions (shown for legged locomotion as an example). Reproduced from Ref. [1].

recent advances in terradynamics and robotics, to develop new theories and models that could revolutionize our comprehension of movement in natural and artificial systems alike. This exploration not only offers insights into animal behavior and evolution but also holds significant implications for the future of mobile robotics in navigating diverse environmental challenges.

40.1 The importance of locomotor transitions

In nature, animals can use many modes of locomotion (e.g., walk, run, crawl, slither, burrow, climb, jump, fly, swim) [2] and often transition between them [3]. Despite this, most mechanistic understanding of terrestrial locomotion has been on how animals generate and stabilize steady-state, limit-cycle-like locomotion using a single mode such as walking, running, and lateral oscillation (for snakes) [4, 5, 6]. Previous studies revealed how terrestrial animals transition across locomotor modes in complex environments. Locomotor transitions, like other animal behavior, emerge from multi-scale interactions of the animal and its environment across the neural, postural, navigational, and ecological levels [7]. At the neural level, terrestrial animals use central pattern generators and sensory information to switch locomotor modes to traverse different media or overcome obstacles. At the ecological level, terrestrial animals foraging across natural landscapes switch locomotor modes to minimize metabolic costs. At the intermediate level, terrestrial animals transition between walking and running to save energy. However, a knowledge gap remains in how locomotor transitions in complex 3D terrain emerge from physical interaction (i.e., terradynamics [8]) of an animal's body and appendages with the environment mediated by the nervous system. We lack theoretical concepts for generating and controlling locomotor transitions in complex terrain on the same level of limit cycles for steady-state, single-mode locomotion. For example, locomotion in irregular terrain with repeated perturbations requires an animal to continually modify its behavior, which limit cycles cannot describe. Similarly, although relatively simple 2D waves can describe rhythmic limbless locomotion on flat ground, such a simplified description is insufficient for understanding how the continuous body moves through complex 3D terrain.

Understanding how to use physical interaction with complex 3D terrain to generate and control locomotor transitions is also critical to advancing mobile robotics. Like personal computers in the 1970s, mobile robots are on the cusp of becoming a significant part of society. Wheeled robots like robotic vacuums and self-driving cars (Fig. 40.1a) already excel at avoiding sparse obstacles to navigate flat homes, streets, and even unpaved roads, by scanning a geometric map of the environment (Fig. 40.1b) and acting upon it to transition between driving modes (Fig. 40.1c). This owes to the fact that wheel–ground interaction is largely understood [9]. Understanding of how to use leg–ground interaction to generate and stabilize steady-state running and walking [5, 4] enabled animal-like legged robot locomotion (such as from Boston Dynamics) on nearly flat surfaces with small obstacles. Snakes' ability to cope with complex 3D environments has inspired the development of snake-like robots for critical tasks like search and rescue in earthquake rubble, building inspection, and extraterrestrial exploration [10]. However, despite progress on robotic design, actuation, and control strategies for multi-modal locomotion [3], robots still struggle to make robust locomotor transitions to traverse obstacles as large as themselves, an ability missing for critical applications such as environmental monitoring in forests (Fig. 40.1d), search and rescue in rubble, and extraterrestrial exploration through rocks. This is mainly due to a poor understanding of physical interaction in complex 3D terrain.

A physics-based approach holds promise for filling this major gap. For aerial and aquatic locomotion of animals and robots, we understand fairly well their fluid–structure interaction thanks to well-established experimental, theoretical, and computational tools, such as wind tunnels and water channels, airfoil and hydrofoil, aero- and hydrodynamic theories, and computational fluid dynamics. By creating new experimental tools and theories, recent studies have elucidated how animals and robots should use physical interaction with granular media to move effectively both on and within sandy terrain (see Ref. [11] for a review), thus beginning to establish a new field of terradynamics in granular media [8]. The general physical principles [11] and predictive physics models [8] not only advanced our understanding of functional morphology, control, and evolution of animals [12, 13, 14], but also led to new design and control strategies [15] that enabled a diversity of robots to traverse granular environments.

Inspired by these successes, our group has been expanding the field of terradynamics of locomotion in complex 3D terrain by integrating biological experiments, robotic physical modeling, and physics modeling (Fig. 40.2). Here, we review our approaches, progress, and opportunities ahead. Although still at an early stage, our work has begun to discover general physical principles, which is remarkable considering that complex 3D

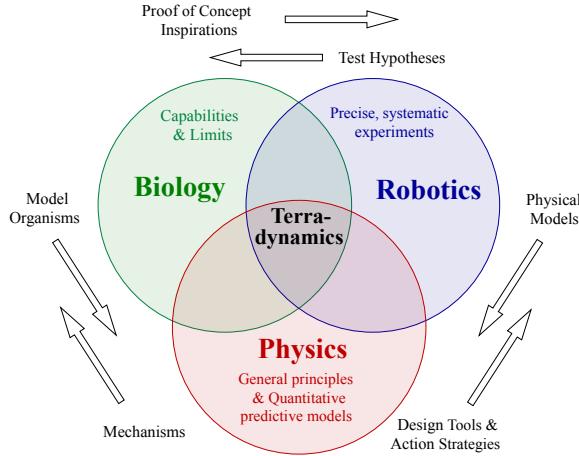


Figure 40.2: Integrative approach of terradynamics. Observations of model organisms inspire robot design and actuation. Simplified robots serve as physical models for testing biological hypotheses or generating new ones [15, 16, 17], while allowing control and variation of parameters to discover general principles. Physical principles and predictive models from this empirical approach provide mechanistic explanations for animal locomotion as well as design tools and action strategies for robots. Reproduced from Ref. [1]. [Agnese: format figure according to book format.](#)

terrain is highly heterogeneous with diverse obstacles.

40.2 Legged locomotion

In this section, we consider the case of legged locomotion in complex 3D terrain using a rainforest-dwelling discoid cockroach as a model organism. The choice of this organism is motivated by its exceptional ability to navigate diverse obstacles, and the focus is on understanding physical interactions during rapid, bandwidth-limited locomotion. The section covers the experimental tools employed, including a terrain testbed and robotic physical models, to systematically investigate locomotor transitions and interactions in challenging terrain. These investigations reveal insights into how animals and robots probabilistically transition between different modes of locomotion based on potential energy landscape modeling, leading to a suite of strategies that modulate these transitions for improved performance. The ultimate goal is to comprehensively understand multi-pathway locomotor transitions in dynamic environments.

40.2.1 Experimental methods

Here, we will give an overview of the employed experimental methods.

Animals We chose the rainforest-dwelling discoid cockroach *Blaberus discoidalis* (Fig. 40.3a) as our model organism because it is exceptional at traversing complex 3D terrain with diverse large obstacles, such as vegetation, foliage, crevices, and rocks. Because we aim first to understand physical interaction, we focused on locomotion in the rapid, bandwidth-limited regime [18] dominated by passive mechanics [2] during escape behavior.

Model terrain To begin understanding physical interaction during locomotion in nature (Fig. 40.1d), we started by abstracting complex 3D terrain as a composition of diverse large obstacles (Fig. 40.1e) that present distinct locomotor challenges. These include compliant beams [24, 25], rigid pillars [22], gaps [21], and bumps [20]. In addition, to enable controlled experiments for each model terrain (analogous to wind tunnels and water channels), we created a test bed that allowed systematic variation of obstacle properties, such as stiffness [24],

size [21, 20], and geometry [22]. Furthermore, because animals and robots often flip over when traversing large obstacles [22, 26, 27], we studied strenuous ground self-righting in which existing appendages must be co-opted [27, 28, 23, 29, 30]. We developed tools (e.g., multi-camera imaging systems, calibration objects) to address the technical challenges in measuring locomotor transitions and locomotor-terrain interaction in complex 3D terrain (Fig. 40.3b-d).

Although studying locomotor transitions to overcome these challenges separately is an amenable first step, in the real world, animals and robots must continually transition over longer spatiotemporal scales (Fig. 40.1f). In addition, besides rapid, bandwidth-limited locomotion during escape, animals can also explore the terrain at their own pace. To study these, we developed a terrain treadmill with modular large obstacles (Fig. 40.3e) to observe locomotion over a long time and distance [19].

Robotic physical models To test biological hypotheses and systematically vary locomotor and terrain parameters beyond the animal’s, we created simplified robotic physical models (Fig. 40.3f-j). Such minimalistic robots offer several advantages as experimental platforms. First, they generate relevant locomotor behavior using minimalist design, actuation, and sensing. In addition, they are more amenable to controlled parameter variation. Moreover, running the robot in an open loop isolates the effects of passive mechanics from that of sensory feedback. Finally, robots as physical models “*can’t violate the laws of physics because robots are enacting, not modeling, the laws of physics*” [31].

We emphasize that our robots were designed and controlled to generate relevant locomotor transitions that we studied, not optimized for maximal performance. However, the physical principles revealed by these tools are generalizable and can predict how to increase performance [15, 16, 26, 27, 21, 20, 22, 24, 23, 29, 30].

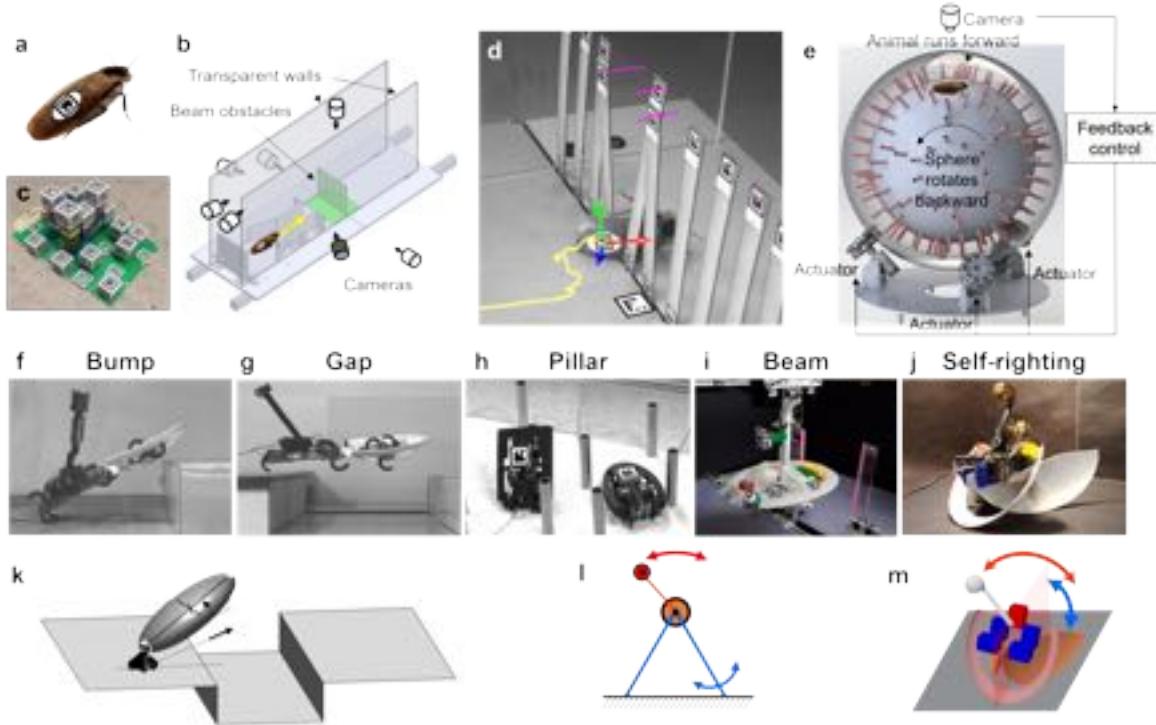


Figure 40.3: **Experimental tools and dynamic models of legged locomotion.** (a) Model organism: the rainforest-dwelling discoid cockroach *Blaberus discoidalis*. (b) Terrain testbed with multi-camera imaging system. (c) Calibration object. (d) Snapshot of obstacle traversal showing automatically tracked trajectories of animal (yellow) and terrain (pink) markers. (e) Terrain treadmill with an untethered animal kept atop by rotating the sphere at the opposite velocity [19]. (f)-(j) Robotic physical models of traversing bump [20], gap [21], pillar [22], beam [23, 24, 25] obstacles and self-righting [26, 27, 28, 29, 30]. (k), (l) Dynamical templates of gap traversal [21] and self-righting [29]. (m) Multibody dynamics simulation of a self-righting robot [30]. Reproduced from Ref. [1]. Agnese: format figure according to book format.

40.2.2 Modelling approaches

Here, we will give an overview of various modelling approaches.

Potential energy landscape modeling Understanding how transitions emerge from locomotor–terrain interaction probabilistically calls for a statistical physics approach. A statistical physics treatment has advanced understanding of complex, stochastic, and macroscopic phenomena in self-propelled living systems, such as animal foraging [32], cooperative transport, and active matter [33, 34]. Here, we use an approach to the potential energy landscape directly inspired by the free energy landscape approach employed in modeling microscopic, multi-pathway protein folding transitions [35, 36, 37]. The near-equilibrium, microscopic proteins statistically transition from higher to lower, thermodynamically more favorable free energy states on the free energy landscape. Thermal fluctuation comparable to free energy barriers induces probabilistic barrier crossings. These physical principles operating on a rugged landscape lead to multi-pathway protein folding transitions. Although our locomotor–terrain systems are macroscopic, self-propelled, and far from equilibrium, their locomotor transitions display similar features, including stochastic, multiple transition pathways, substantial kinetic energy fluctuation (from oscillatory self-propulsion), and preference for some modes over others [24, 21, 20, 22, 23, 29, 30, 25].

Given these similarities, we hypothesized that locomotor transitions are barrier-crossing transitions between basins on potential energy landscapes of locomotor–terrain interaction systems. We tested this hypothesis in each model system (Fig. 40.4). To discover the general principles of locomotor transitions, we systematically varied system parameters and tested how they affect locomotor transitions.

A potential energy landscape approach to modeling locomotor–terrain interaction is plausible also considering the success of potential energy field methods in robotic manipulation. Similar to our systems, robotic part alignment [38] and grasping [39] have continual collisions, multiple pathways to reach the goal [38], and preference of some contact configurations over others [40]. Despite these complexities, quasi-static potential energy field models well explained how system properties like geometry and friction affect part–manipulator interaction and informed planning strategies for desired alignment or manipulation [38, 39]. We emphasize that our potential energy landscapes directly result from physical interaction and are based on first principles, unlike artificially defined potential functions to explain walk-to-run transition [41], and other non-equilibrium biological phase transitions [42], metabolic energy landscapes inferred from oxygen consumption measurements to explain behavioral switching of locomotor modes [43], and artificial potential fields for robot obstacle avoidance [44]. Our potential energy landscapes do not yet capture system dynamics, which is required for quantitative predictions of locomotor transitions. Despite this limitation, they provided substantial insight into the general principles and strategies of obstacle traversal and strenuous ground self-righting.

Dynamic templates and simulations For our model systems, it is often challenging to solve equations of motion analytically due to the hybrid contact [45] and high-dimensional parameter space. As a first step to understanding transition dynamics, we developed dynamical templates for two model systems, namely, large gap traversal [21] (Fig. 40.3k) and strenuous ground self-righting [29] (Fig. 40.3l), whose equations of motion are solvable when two-dimensional dynamics are considered. Templates are the simplest dynamical models that realize the fundamental dynamics of a locomotor behavior using minimal degrees of freedom [46]. For these two systems, our templates enabled quantitative predictions of contact and actuator forces [29], control strategies for traversal or self-righting, and how they depend on system parameters [21, 29].

In addition, for strenuous ground self-righting, we developed multibody dynamics simulations of the robot [30] to study the effect of randomness in wing-leg coordination (Fig. 40.3m). These simulations enabled large-scale variation of relevant parameters for in-depth analysis, revealing nuanced effects difficult to observe in animals and robots. Finally, simulation is faster than experiments [30].

40.2.3 Insights on legged locomotion

Our studies revealed how locomotor transitions depend on system parameters (e.g., gap width, beam stiffness). For each model system, these physical principles are generalizable over the relevant parameter space and improved robot performance. Across all model systems, a potential energy landscape approach helps understand

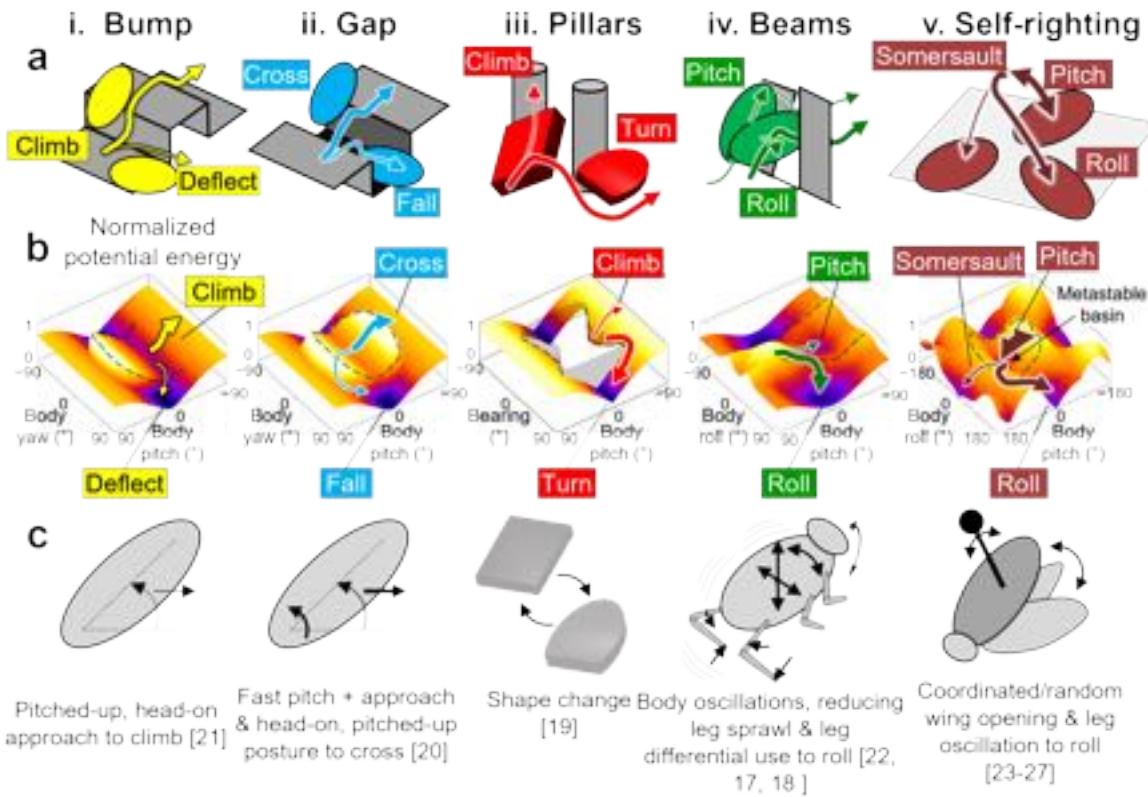


Figure 40.4: **Modulation of legged locomotor transitions on potential energy landscapes via a suite of strategies.** (a) Locomotor modes and transitions of model systems. (b) Each mode is attracted towards distinct basins separated by potential energy barriers, which must be overcome to make transitions. Black dashed curves show potential energy barriers. Arrows in (a) and (b) show representative system state trajectories; thicker arrows show more desirable modes. (c) A suite of strategies can increase the probabilities of desired modes and facilitate transitions to overcome locomotor challenges. We renamed some modes/basins differently from the original papers to better distinguish them across model systems. Reproduced from Ref. [1]. [Agnese needs the source for this figure for reformatting into the book style.](#)

how the probabilistic locomotor transitions of animals and robots are constrained by physical interaction. Comparing across model systems, several general physical principles and new concepts emerge:

- **Animal transitions happen between modes probabilistically and often via multiple pathways.** For all model systems, we found that the animal's transitions between modes occur stochastically, with large trial-to-trial variation [24, 20, 21, 22, 23]. However, statistically, the animal uses some modes or transition pathways more frequently than others, which strongly depends on locomotor and terrain parameters that affect the physical interaction.
- **Locomotor transitions are barrier-crossing transitions on a potential energy landscape.** For all model systems, during each mode, the system's state is strongly attracted to a local minimum basin on the potential energy landscape [24, 21, 20, 22, 29, 30, 23, 25] (Fig. 40.4). This is because self-propulsion induces continual collisions during obstacle interaction and self-righting, which breaks continuous frictional contact between the body and environment and makes the system statically unstable. Because self-propulsion does not always perturb the system sufficiently to escape a basin, the statically unstable system drifts from less (higher potential energy) to more stable (lower potential energy) states within the basin. However, it also does not stay at the minimum due to propulsive forces. Due to this strong attraction to landscape basins, the transition from one locomotor mode to another requires the system to escape from one basin to settle into another. In addition, kinetic energy fluctuation from oscillatory self-

propulsion helps the system stochastically cross potential energy barriers to make transitions [24, 23]. Furthermore, escape from a basin is more likely towards a direction on the landscape along which the potential energy barrier is lower [24, 22, 23]. Finally, when the interaction is dominated by passive mechanics, the system tends to settle into more favorable modes attracted to lower landscape basins when it can cross barriers [24, 22, 23]. Although we focused on understanding the effect of passive mechanics, animals can certainly use active adjustments using sensory feedback. For example, the animal often transitions to rolling during beam traversal when the potential energy barrier becomes comparable but is still a bit higher than body kinetic energy fluctuation [24].

- **There exists a regime of locomotion dominated by the potential energy landscape.** These similar observations across the diverse model systems mean that there is a regime of locomotion dominated by the potential energy landscape [1]. In this regime, along certain directions, large potential energy barriers are comparable to or exceed kinetic energy and/or mechanical work generated by each propulsive cycle or motion. This may happen when propulsive forces are either limited by physiological, morphological, and environmental (e.g., low friction) constraints or are not well directed toward the barriers for a desired transition. These situations are frequent both in complex 3D terrain with many large obstacles and during strenuous ground self-righting, which explains why this approach works well in these scenarios. In this regime, potential energy landscapes provide a useful statistical physics approach for understanding locomotor transitions and allow comparison across systems (different species, robots, terrain, and modes) to discover general physical principles. However, outside of this regime, energy landscape modeling is not useful or unnecessary.
- **Locomotor transitions can be modulated using a suite of strategies to increase performance.** Viewing locomotor transitions as barrier-crossing transitions on the potential energy landscape in our model systems revealed that a suite of strategies can be used to enable, facilitate, suppress, or eliminate transitions (Fig. 40.4a) by steering the system state on the landscape and/or altering landscape barriers and topology (Fig. 40.4b). For each model system, we discovered strategies (Fig. 40.4c) to make desired transitions more probable. These strategies increased robot performance or even enabled new capabilities. Specifically, in bump traversal, approaching with a head-on, pitched-up body posture (Fig. 40.4c, i) better constrains the system state to a desired, albeit unstable, climb basin (Fig. 40.4b, i) and suppresses entrapment in a deflect basin [20]. Similarly, in gap traversal, approaching with large forward velocity and upward pitching velocity and a head-on, pitched-up body posture (Fig. 40.4c, ii) increases kinetic energy useful to overcome the potential energy barrier for transitioning to a desired cross basin (Fig. 40.4b, ii) and steer clear of a fall basin [21]. During pillar interaction, an elliptical body shape (Fig. 40.4c, iii) results in a repulsive turn basin (Fig. 40.4b, iii) that favors escaping from the pillar [22]. By contrast, when climbing up the pillar is desirable, a cuboidal shape eliminates the turn basin and adds an attractive climb basin that favors the body pitching against the pillar [22]. In beam traversal, when beams are stiff and pushing across is undesirable, body oscillation induced by self-propulsion provides kinetic energy fluctuations to overcome the potential energy barrier required to escape from a pitch basin to reach a roll basin (Fig. 40.4b, iv) [24], allowing the body to roll into the gap between beams to traverse. This transition is further facilitated by reduced sprawling and differential use of hind legs (Fig. 40.4c, iv) [25], which presumably destabilizes the system state towards the roll basin. In strenuous ground self-righting, wing opening alters the landscape and reduces the potential energy barrier required to escape from a metastable basin to reach a roll basin (Fig. 40.4b, v). This reduced barrier can then be overcome by small kinetic energy fluctuations from leg oscillation to self-right via body rolling [23]. Self-righting performance is enhanced by proper wing-leg coordination (Fig. 40.4c, v) that best accumulates mechanical energy to overcome energy barriers [29]. Alternatively, it may be improved by adding randomness into the wing-leg to allow the system to find proper coordination [30]. We emphasize that the desirable modes and strategies in the abovementioned obstacle interaction aim toward successful traversal. In different tasks, other modes may be more desirable. Finally, during free exploration of complex 3D terrain over large spatiotemporal scales, the animal likely uses a composition of these strategies (and those yet to be discovered) to modulate its transitions. Experiments using our terrain treadmill (Fig. 40.3e) are beginning to shed light on this [19].

Example 40.1: Diamond body shape and pillar interaction. We read in the paragraphs above that a cuboidal-shaped body is more likely to climb up against the pillar (Fig. 40.5a), whereas an elliptical body is likely to be repelled away from the pillar (Fig. 40.5b). Based on these observations, can you guess which of the sequences shown in the figure are likely to occur for a diamond-shaped body moving toward the pillar?

[Hint: Consider varying length-to-width ratio. When does the contact become similar to that of a cuboidal or elliptical-shaped body? How does the angular impulse direction from the contact change?]

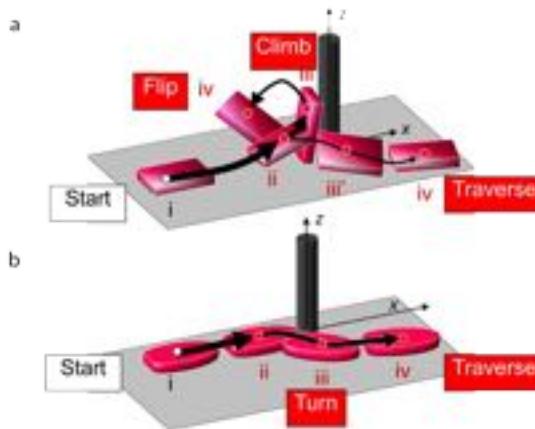


Figure 40.5: **Pillar interaction.** (a) A robot with a cuboidal body shape is more likely to climb up against the pillar than to be repelled away from the pillar. (b) A robot with an ellipsoidal body shape will likely be repelled away from the pillar. Robots can achieve the desired mode by morphing into the corresponding body shape.
Agnese: re-organize the panels in a horizontal fashion.

Example 40.2: Swinging one's tail to pitch one's body up. For the legged robot with an active tail (Fig. 40.3g), how does swinging the tail help the robot to pitch its body up?

[Hint: We can assume that the angular momentum of the robot is roughly conserved, as the active tail swings rapidly and the contributions to angular momentum from the ground reaction forces on the various legs approximately cancel each other out in this process.]

Example 40.3: Rolling mode for stiff beams. For stiff beams, why is the rolling mode more favorable?

[Hint: Consider an extreme case where the beams are infinitely stiff, i.e., unbendable. Now, the system's potential energy only consists of the gravitational potential energy of the animal. What is the minimum height the cockroach needs to reach to traverse along the middle of the two beams in the roll mode compared with the pitch mode?]

Example 40.4: Beam transversal. Consider the schematic for the animal's beam traversal (Fig. 40.6). By approximating the animal as a rigid ellipsoid with uniform density and assuming that it is always in contact with the ground, write the total potential energy of the system as a function of Δz , $\Delta\theta_1$ (left beam angle), $\Delta\theta_2$ (right beam angle), and h . Define total potential energy as zero when the body is horizontal, and beams are vertical.

[Hint: The system's potential energy consists of the gravitational potential energy of the animal and

the beams and the elastic potential energy of the beams.]

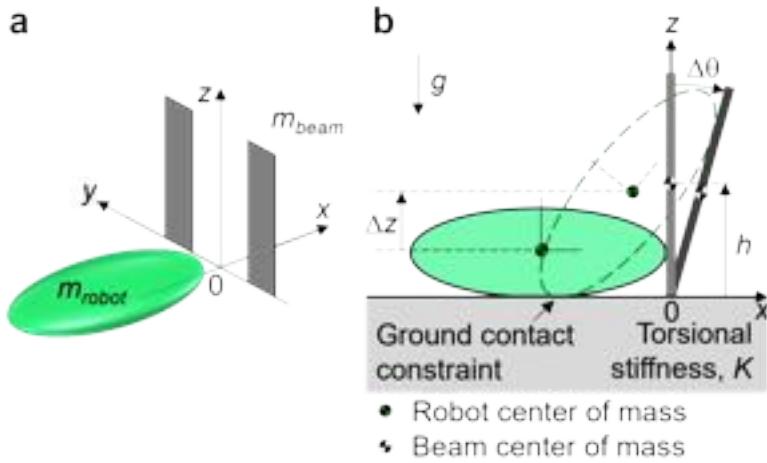


Figure 40.6: **Beam transversal.** (a) Schematic of the reference situation with zero potential energy. The body is horizontal and the beams are vertical. (b) Comparison with the situation when the beam is deflected with an angle $\Delta\theta$ and the body has climbed up against the deflected beam. The ground contact constraint is evidenced in the figure and the position of the center of mass of both the body and the beam. The torsional stiffness of the beam is K . The mass of the robot and each beam is m_{robot} and m_{beam} , respectively. The height of each beam's center of mass is h initially. The robot's center of mass raises Δz as it climbs. [Agnese: Uniform to book style.](#)

Example 40.5: Self-righting for three different species of cockroaches. For the three species of cockroaches shown in Fig. 40.7, assume their body to be rigid ellipsoids of uniform density. Calculate the potential energy barrier to self-right by somersaulting over their head (pitch mode) to that by rolling sideways (roll mode). Use the information in the table below:

	Madagascar	American	Discoid
$2a$ (cm)	6.0	3.34	4.98
$2b$ (cm)	2.24	1.19	2.38
$2c$ (cm)	1.32	0.7	0.96
Mass (g)	7.44	0.66	2.14

40.2.4 Towards multi-pathway locomotor transitions

So far, we have focused on using potential energy landscapes to understand how the physical interaction of relatively simple model systems constrains locomotor modes and simple transitions. We still need energy landscape theories analogous to those of protein folding [35, 37, 47] to understand and predict how multi-pathway transitions (such as those observed in [26]) emerge from physical interaction mediated by the animal's neuromechanical or robot's sensing, control, and planning systems.

40.3 Limbless locomotion

In limbless locomotion, traversing complex 3D terrain poses a significant challenge. Here, we look into the intricacies of limbless locomotion, particularly focusing on the traversing of large steps by snakes and snake-like

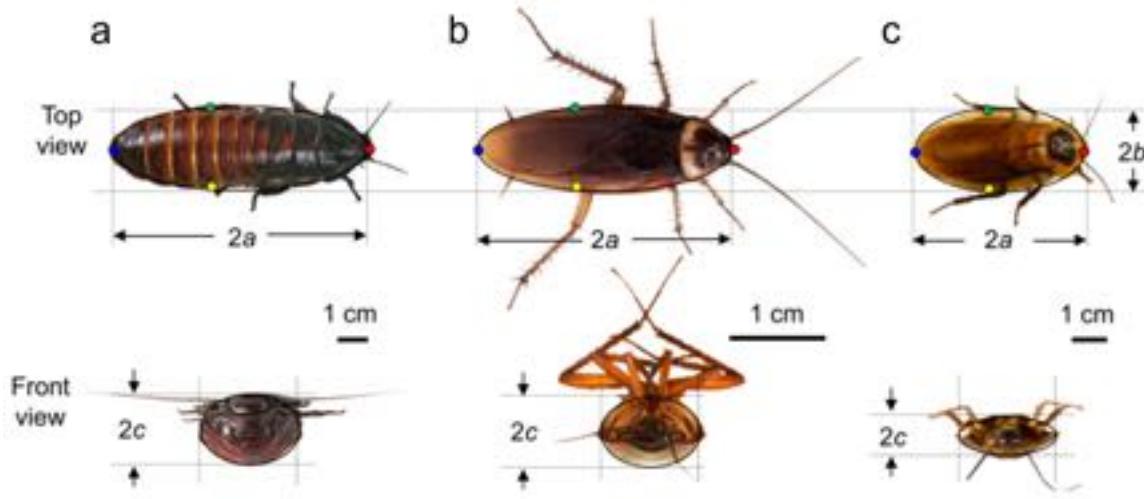


Figure 40.7: Top and front view of three species of cockroaches. (a) Madagascar cockroach, *Gromphadorhina portentosa*. (b) American cockroach, *Periplaneta americana*. (c) Discoid cockroach, *Blaberus discoidalis*. The values of the cockroach length ($2a$), width ($2b$), and height ($2c$) are given in the text of Exercise 40.5. Reproduced from Ref. [28]. Agnese: Uniform to book style.

robots. The stability aspect is explored, emphasizing the difficulties encountered when maintaining stability while bending the body out of the plane. This section also introduces the concept of snake transitions between lateral oscillation and cantilevering to overcome obstacles. Through experimentation with a variable kingsnake and a snake-like robot, insights are gained into the stability advantages of combining lateral oscillation with cantilevering. Furthermore, the importance of body compliance in enhancing traversal performance is examined. This study sheds light on the biomechanical principles governing limbless locomotion and draws parallels and commonalities between legged and limbless locomotion strategies.

40.3.1 Experimental tools

Here, we give an overview of the experimental tools.

Animals We challenged the variable kingsnake (*Lampropeltis mexicana*), a generalist found in diverse rocky environments, to traverse a large step obstacle [48] (Fig. 40.8a).

Robotic physical models We developed a snake-like robot capable of large out-of-plane body bending and used it as a physical model to further study the stability principles [50]. We challenged the robot to traverse a step obstacle with an increasingly large step height (Fig. 40.10c, d).

40.3.2 Insights from snake traversing large steps

For limbless locomotion in complex 3D terrain, a major challenge is to maintain stability. When snakes [6, 51, 52] and snake-like robots [53, 54] use planar gaits to move on flat surfaces, they are inherently stable. However, the more they bend the body out of the plane, the more challenging it is to maintain stability [55, 56]. Arboreal snakes can grip branches and twigs, and desert snakes can brace against depressed sand, thereby using or creating “anchor points” for stability. Similarly, snake-like robots can use or create anchor points in complex environments like ladders, pipes, and desert dunes [56, 57, 58]. In contrast, it is unclear how snakes or snake-like robots can stably traverse large, smooth obstacles lacking anchor points, such as boulders and felled trees.

In the animal experiment, we discovered that the snake traverses large step obstacles by transitioning between two modes (Fig. 40.8b). Both the anterior and posterior body sections oscillate laterally on the horizontal

surfaces above and below the step (Fig. 40.8b, black solid line). Note that lateral oscillation can include both lateral undulation and concertina [59], with the former being smooth and the latter having intermittent start-stop motions. Indeed, we observed lateral undulation-like motion when the step obstacle is low or the surface has high friction, but more concertina-like motion when the step obstacle is high and the surface is slippery [48]. To bridge the large height, the body section in between cantilevers in the air before the head reaches the surface above and is suspended in the air afterward (both referred to as the cantilever mode hereafter Fig. 40.8b, red dashed line). Each body part transitions between these two modes as the snake progresses forward and upward onto the step.

When step height increases, the snake devotes a longer body section to cantilevering and pitches it up more to accommodate the larger height. When surface friction decreases, the snake suffers large lateral and backward slips and moves more intermittently; its cantilevering body section is also closer to the step. However, despite these active adjustments, the overall transition between lateral oscillation and cantilevering is conserved. These observations suggested a simplified control template for limbless traversal of large step obstacles [48] (Fig. 40.8b).

Many previous snake-like robots traversed large, smooth step-like obstacles using a simple follow-the-leader gait [60, 61]. With each body segment following the previous one, these robots often simply bent the body within their sagittal plane to traverse (Fig. 40.9a). To be statically stable during cantilevering before the head reaches the surface above, careful feedback control is necessary to ensure that the center of mass always projects into the narrow base of support formed by the straight body in contact with the surface below (Fig. 40.9a). Otherwise, the robot can be easily tipped over by a lateral perturbation, such as slipping on low-friction steps. As a result, these previous snake-like robots are often slow in traversing large steps.

In contrast, the snake's lateral oscillation during cantilevering may be the key to its success, as it can help achieve a broad base of support to resist lateral perturbations (Fig. 40.9b). Using a newly developed continuous body interpolation method [62], we examined the snake's static stability during large step traversal [49]. Despite its large out-of-plane body bending, the snake always maintains perfect static stability (with its center of mass projection falling into the base of support 100% of the time for all trials) regardless of step height and friction.

In the robot experiment, using the control template of transitioning between lateral oscillation and cantilevering from the snake (Fig. 40.8b), the robot rapidly traversed lower steps with high probability (Fig. 40.10a,

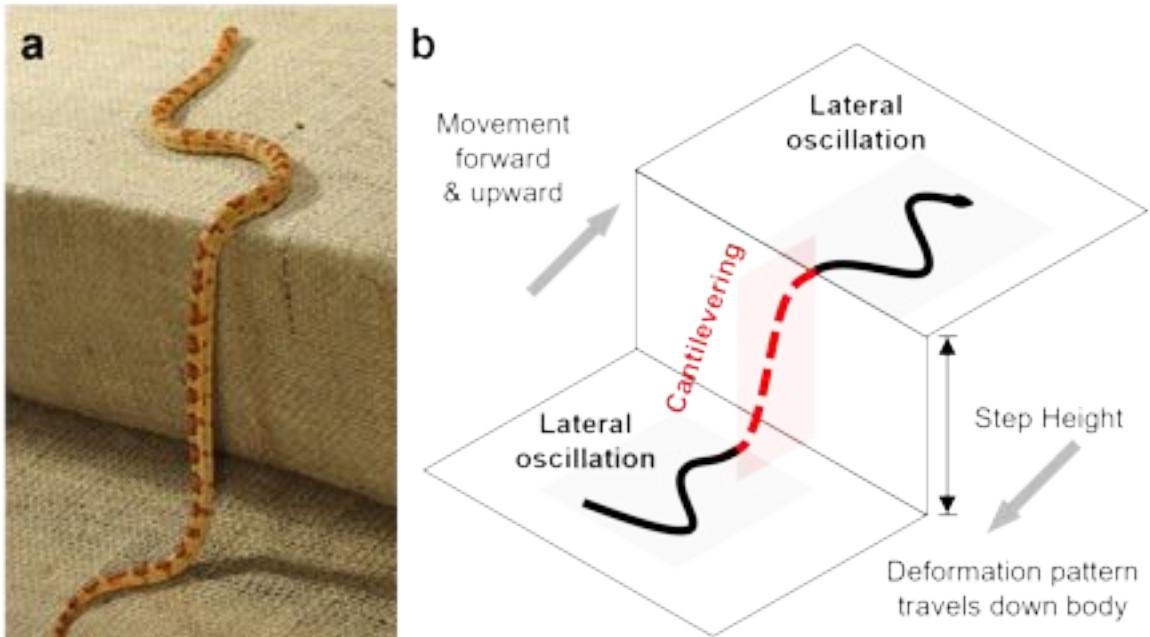


Figure 40.8: Snake traversal of large steps. (a) Snapshot of a kingsnake traversing a large step by transitioning between lateral oscillation and cantilevering. (b) Gait template for step traversal. Reproduced from Ref. [49]. Agnese: format figure according to book format.

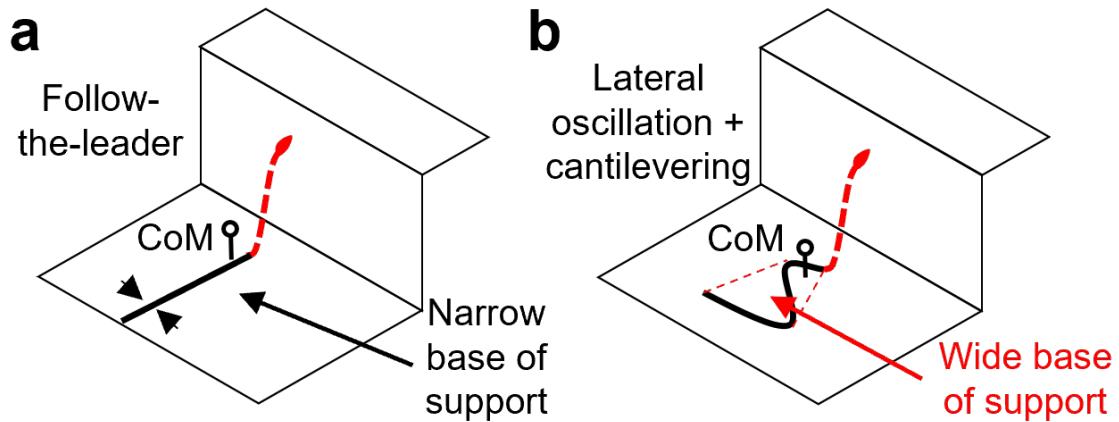


Figure 40.9: Comparison of stability of large step traversal without and with body lateral oscillation. (a) Traditional follow-the-leader gait results in a narrow base of support. **(b)** Snake's locomotor transition between lateral oscillation and cantilevering provides a wide base of support [50]. Reproduced from Ref. [49]. Agnese: format figure according to book format.

black dashed line). An online camera measured the distance between the robot and the step obstacle to determine when to transition between modes. Our robot (both rigid and compliant body) achieved at least 50% higher step traversal speed than most previous robots when normalized to body length while maintaining high traversal probability.

However, the robot's traversal probability diminished quickly as step height increased. The majority of unsuccessful traversal was a direct result of roll failure — the robot rolled so much that it flipped over (Fig. 40.10c). By contrast, even with large slipping on low friction steps, the snake never rolled as far as losing contact with the surface during traversal. The snake's better stability may be due to body compliance, which allows the animal body to deform locally to better engage the surface.

We used our robot as a physical model to test the effects of body compliance by adding a springy suspension system [50]. The compliant robot was more likely to traverse higher steps than the originally rigid body robot (Fig. 40.10a, red solid line vs. black dashed line). This was a direct result of larger body deformation and, thus, reduced roll failure probability (Fig. 40.10b, red solid line vs. black dashed line) compared to the rigid body robot. The added compliance did not significantly reduce the robot's speed during traversal.

The snake still has superior traversal performance without large involuntary body rolling or flipping over. It will be fruitful to study how body continuity, body compliance in multiple directions, and sensory feedback control contribute to the traversal of large, smooth obstacles. Finally, although we focused on a simple large step, the combination of transition between multiple modes and body compliance may be useful for limbless animals and robots moving through more complex 3D terrain [63, 61, 64].

Example 40.6: Snake transversing a gap. For a snake traversing the gap shown in Fig. 40.3g, describe the possible transition between multiple modes.

[Hint: Similar to the snake traversing a large step, the snake needs one mode to bridge the gap, and another before and after traversing the gap.]

Example 40.7: Support polygons of a snake-like robot. The same snake-like robot forms the following two support polygons (blue polygon in Fig. 40.11) on a horizontal surface. With the center of mass projection at the red point, which condition is more stable statically, assuming the height of its center of mass is the same? Why?

[Hint: The snake-like robot will lose stability (i.e., flip over) as the robot rotates and the center of mass

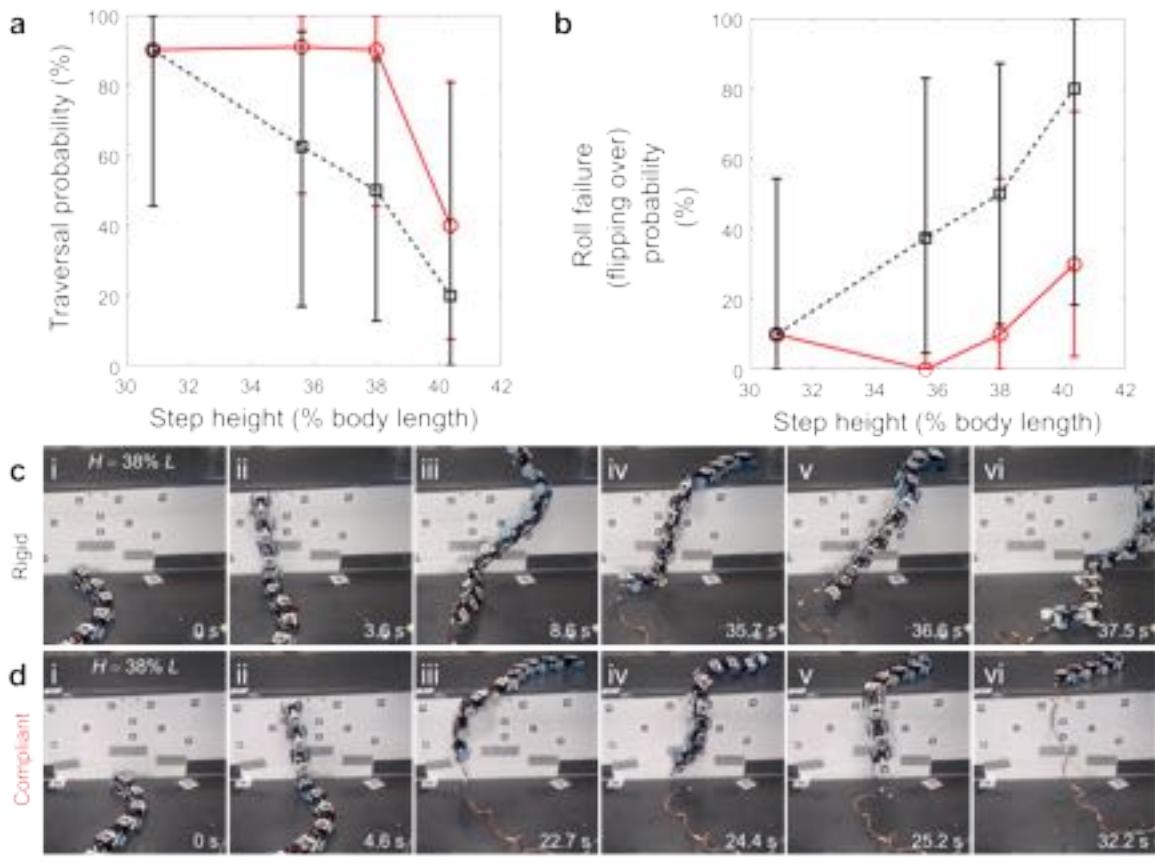


Figure 40.10: **Traversal performance of robot snakes.** (a) Traversal probability as a function of step height. (b) Roll failure (flipping over) probability as a function of step height. (c, d) Snapshots of the robot with a rigid (c) and compliant (d) body traversing a large step [50]. Reproduced from Ref. [49]. Agnese: format figure according to book format.

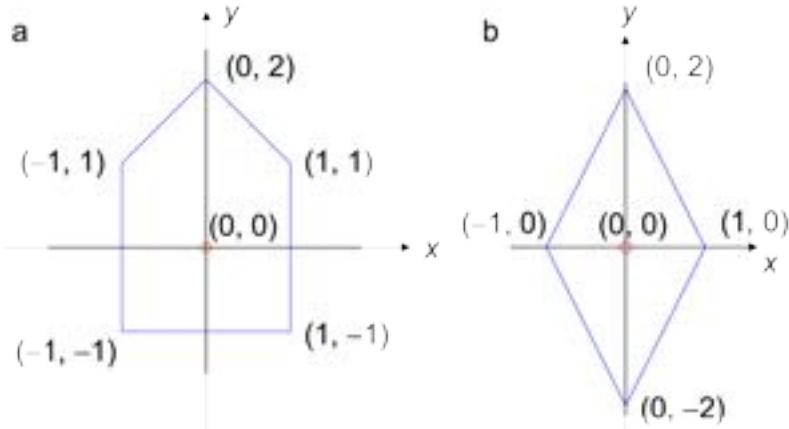


Figure 40.11: **Support polygons of a snake-like robot.** (a) Pentagonal and (b) diamond-shaped support polygon. Agnese: Uniform to book style.

projection falls outside the support polygon. What minimum angle does the robot need to rotate to lose

stability? How much potential energy has to be injected for this to happen?]

40.4 Common features of legged and limbless locomotion

Despite morphological differences between legged and limbless animals, their terrain interaction and resulting locomotor transitions share common features. For both, negotiating obstacles requires the entirety or parts of their body to transition into different locomotor modes. In addition, these transitions and emergent movement in complex 3D terrain are facilitated by body morphology (e.g., rounded, ellipsoidal body shape of insects [26], compliant body of snakes [50]). Furthermore, active adjustments via sensory feedback are crucial to facilitate or even enable desired transitions that may not happen from passive interaction alone [25]. Finally, changes in body movement during locomotor transitions are stereotypical and limited to a low-dimensional space of all the available body degrees of freedom.

Observations of the legged and limbless robots also share similarities. For both, a principled understanding of physical interaction [26, 24, 50, 62, 21, 20, 22, 29, 30, 23] and simplified control templates [21, 29, 50] facilitate locomotor transitions and improve traversal performance. Not surprisingly, with their reduced degrees of freedom, lack of or limited sensing, and well-controlled actuation with reduced noise and randomness [30], our robots' locomotor transitions are neither as robust nor as diverse as the animals.

40.5 Outlook

In the longer term, we envision that models of locomotor-terrain physical interaction from first principles will be pervasive. In addition to self-driving cars that scan streets to create geometric maps and upload them to the cloud for shared use, robots will better use physical interaction to traverse currently unreachable, complex 3D terrain and create physics-based environmental maps and action databases. These maps and databases will be added to the cloud [65] and enable robots to expand our reaches in natural, artificial, and extraterrestrial terrain.

40.6 Problems

Problem 40.1: Traditional approaches and complex, cluttered 3D terrain. Why do traditional geometric mapping and path planning approaches of robotics not work for robots that must dynamically traverse complex, cluttered 3D terrain?

Problem 40.2: Importance of body shape during transversal of cluttered 3D terrain. Why is body shape important during traversal of cluttered 3D terrain, such as grass-like beams, like it is during locomotion in fluids?

Problem 40.3: Analogy between locomotor transitions in complex 3D terrain and protein folding transitions. How are locomotor transitions in complex 3D terrain analogous to protein folding transitions?

Problem 40.4: Step transversal and wheels. Why does the snake-like robot in Fig. 40.10 for step traversal have wheels? Are there any possible drawbacks? Can you provide any possible ways to improve it?

Problem 40.5: Body compliance and step transversal performance. How does body compliance help the snake-like robot in Fig. 40.10 improve its step traversal performance?

Bibliography

- [1] R. Othayoth, Q. Xuan, Y. Wang, and C. Li. Locomotor transitions in the potential energy landscape-dominated regime. *Proceedings of the Royal Society B: Biological Sciences*, 288:20202734, 2021.

- [2] M. H. Dickinson, C. T. Farley, R. J. Full, M. A. R. Koehl, R. Kram, and S. Lehman. How animals move: An integrative view. *Science*, 288:100–106, 2000.
- [3] R. J. Lock, S. C. Burgess, and R. Vaidyanathan. Multi-modal locomotion: from animal to application. *Bioinspiration & Biomimetics*, 9:11001, 2013.
- [4] A. D. Kuo. The six determinants of gait and the inverted pendulum analogy: A dynamic walking perspective. *Human Movement Science*, 26:617–656, 2007.
- [5] R. Blickhan and R. J. Full. Similarity in multilegged locomotion: Bouncing like a monopode. *Journal of Comparative Physiology A*, 173:509–517, 1993.
- [6] J. Gray. The mechanism of locomotion in snakes. *Journal of Experimental Biology*, 23:101–120, 1946.
- [7] G. J. Berman. Measuring behavior across scales. *BMC Biology*, 16:23, 2018.
- [8] C. Li, T. Zhang, and D. I. Goldman. A terradynamics of legged locomotion on granular media. *Science*, 339:1408–1412, 2013.
- [9] H. B. Pacejka. Tyre and vehicle dynamics. *Elsevier*, 2005.
- [10] I. D. Walker, H. Choset, and G. S. Chirikjian. Snake-like and continuum robots. In Bruno Siciliano and Oussama Khatib, editors, *Springer Handbook of Robotics*, volume 4, pages 481–498. Springer, Berlin/Heidelberg, Germany, 2016.
- [11] H. C. Astley, J. R. Mendelson, J. Dai, C. Gong, B. Chong, J. M. Rieser, P. E. Schiebel, S. S. Sharpe, R. L. Hatton, H. Choset, and D. I. Goldman. Surprising simplicities and syntheses in limbless self-propulsion in sand. *Journal of Experimental Biology*, 223, 2020.
- [12] R. D. Maladen, Y. Ding, P. B. Umbanhowar, and D. I. Goldman. Undulatory swimming in sand: Experimental and simulation studies of a robotic sandfish. *International Journal of Robotics Research*, 30:793–805, 2011.
- [13] S. S. Sharpe, Y. Ding, and D. I. Goldman. Environmental interaction influences muscle activation strategy during sand-swimming in the sandfish lizard *scincus scincus*. *Journal of Experimental Biology*, 216:260–274, 2013.
- [14] B. McInroe, H. C. Astley, C. Gong, S. M. Kawano, P. E. Schiebel, J. M. Rieser, H. Choset, R. W. Blob, and D. I. Goldman. Tail use improves performance on soft substrates in models of early vertebrate land locomotors. *Science*, 353:154–158, 2016.
- [15] J. Aguilar, T. Zhang, F. Qian, M. Kingsbury, B. McInroe, N. Mazouchova, C. Li, R. Maladen, C. Gong, M. Travers, R. L. Hatton, H. Choset, P. B. Umbanhowar, and D. I. Goldman. A review on locomotion robophysics: the study of movement at the intersection of robotics, soft matter and dynamical systems. *Reports on Progress in Physics*, 79:110001, 2016.
- [16] Y. O. Aydin, J. M. Rieser, C. M. Hubicki, W. Savoie, and D. I. Goldman. Physics approaches to natural locomotion: Every robot is an experiment. In Shawn M. Walsh and Michael S. Strano, editors, *Robotic Systems and Autonomous Platforms*, Woodhead Publishing in Materials, pages 109–127. Woodhead Publishing, 2019.
- [17] N. Gravish and V. G. Lauder. Robotics-inspired biology. *Journal of Experimental Biology*, 221:138438, 2018.
- [18] S. Sponberg and R. J. Full. Neuromechanical response of musculo-skeletal structures in cockroaches during rapid running on rough terrain. *Journal of Experimental Biology*, 211:433–446, 2008.
- [19] R. Othayoth, B. Strelbel, Y. Han, Francois E., and C. Li. A terrain treadmill to study animal locomotion through large obstacles. *Journal of Experimental Biology*, 225:jeb243558, 2022.

- [20] S. W. Gart and C. Li. Body-terrain interaction affects large bump traversal of insects and legged robots. *Bioinspiration & Biomimetics*, 13:26005, 2018.
- [21] S. W. Gart, C. Yan, R. Othayoth, Z. Ren, and C. Li. Dynamic traversal of large gaps by insects and legged robots reveals a template. *Bioinspiration & Biomimetics*, 13:2, 2018.
- [22] Y. Han, R. Othayoth, Y. Wang, C.-C. Hsu, R. de la Tijera Obert, E. Francois, and C. Li. Shape-modulated obstacle attraction and repulsion during dynamic locomotion. *The International Journal of Robotics Research*, 40:939–955, 2021.
- [23] R. Othayoth and C. Li. Propelling and perturbing appendages together facilitate strenuous ground self-righting. *eLife*, 10:e60233, 2021.
- [24] R. Othayoth, G. Thoms, and C. Li. An energy landscape approach to locomotor transitions in complex 3D terrain. *Proceedings of the National Academy of Sciences*, 117:14987–14995, 2020.
- [25] Y. Wang, R. Othayoth, and C. Li. Cockroaches adjust body and appendages to traverse cluttered large obstacles. *Journal of Experimental Biology*, 225:jeb243605, 2022.
- [26] C. Li, A. O. Pullin, D. W. Haldane, H. K. Lam, R. S. Fearing, and R. J. Full. Terradynamically streamlined shapes in animals and robots enhance traversability through densely cluttered terrain. *Bioinspiration & Biomimetics*, 10:46003, 2015.
- [27] C. Li, C. C. Kessens, R. S. Fearing, and R. J. Full. Mechanical principles of dynamic terrestrial self-righting using wings. *Advanced Robotics*, 31:881–900, 2017.
- [28] C. Li, T. Wöhrl, H. K. Lam, and R. J. Full. Cockroaches use diverse strategies to self-right on the ground. *Journal of Experimental Biology*, 222:186080, 2019.
- [29] Q. Xuan and C. Li. Coordinated appendages accumulate more energy to self-right on the ground. *IEEE Robotics and Automation Letters*, 5:6137–6144, 2020.
- [30] C. Xuan, Q. Li. Randomness in appendage coordination facilitates strenuous ground self-righting. *Bioinspiration & Biomimetics*, 21:1–9, 2020.
- [31] J. Long. *Darwin's Devices: What Evolving Robots Can Teach Us About the History of Life and the Future of Technology*. Basic Books, 2012.
- [32] G. M. Viswanathan, M. G. E. Da Luz, E. P. Raposo, and H. E. Stanley. *The Physics of Foraging*. Cambridge University Press, 2011.
- [33] S. Ramaswamy. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics*, 1:323–345, 2010.
- [34] É. Fodor and M. C. Marchetti. The statistical physics of active matter : From self-catalytic colloids to living cells. *Physica A*, 504:106–120, 2018.
- [35] J. N. Onuchic and P. G. Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14:70–75, 2004.
- [36] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl. The protein folding problem. *Annual Review of Biophysics*, 37:289–316, 2008.
- [37] D. J. Wales. *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*. Cambridge University Press, 2003.
- [38] M. A. Peshkin and A. C. Sanderson. Planning robotic manipulation strategies for workpieces that slide. *IEEE Journal on Robotics and Automation*, 4:524–531, 1988.
- [39] M. T. Mason, A. Rodriguez, S. S. Srinivasa, and A. S Vazquez. Autonomous manipulation with a general-purpose simple hand. *The International Journal of Robotics Research*, 31:688–703, 2012.

- [40] G. Boothroyd and C. Ho. Natural resting aspects of parts for automatic handling. *Journal of Engineering for Industry*, 99:314–317, 1977.
- [41] F. J. Diedrich and W. H. Warren. Why change gaits? dynamics of the walk-run transition. *Journal of Experimental Psychology: Human Perception and Performance*, 21:183–202, 1995.
- [42] J. A. S. Kelso. Multistability and metastability: Understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society B*, 367:906–918, 2012.
- [43] E. L. C. Shepard, R. P. Wilson, W. G. Rees, E. Grundy, S. A. Lambertucci, and S. B. Vosper. Energy landscapes shape animal movement ecology. *American Naturalist*, 182:298–312, 2013.
- [44] E. Rimon and D. E. Koditschek. Exact robot navigation using artificial potential functions. *IEEE Transactions on Robotics and Automation*, 8:501–518, 1992.
- [45] P. Holmes, R. J. Full, D. Koditschek, and J. Guckenheimer. The dynamics of legged locomotion: Models, analyses, and challenges. *SIAM Review*, 48:207–304, 2006.
- [46] R. J. Full, D. E. Koditschek, and R. J. Full. Templates and anchors: neuromechanical hypotheses of legged locomotion on land. *The Journal of Experimental Biology*, 2:3–125, 1999.
- [47] K. A. Dill and H. S. Chan. From Levinthal to pathways to funnels: The “new view” of protein folding kinetics. *Nature Structural Biology*, 4:10, 1997.
- [48] S. W. Gart, T. W. Mitchel, and C. Li. Snakes partition their body to traverse large steps stably. *Journal of Experimental Biology*, 59:jeb–185991, 2019.
- [49] Q. Fu, S. W. Gart, T. W. Mitchel, J. S. Kim, G. S. Chirikjian, and C. Li. Lateral oscillation and body compliance help snakes and snake robots stably traverse large, smooth obstacles. *Integrative and Comparative Biology*, 60:171–179, 2020.
- [50] Q. Fu and C. Li. Robotic modeling of snake traversing large, smooth obstacles reveals stability benefits of body compliance. *Royal Society Open Science*, 7:191192, 2020.
- [51] B. C. Jayne. Kinematics of terrestrial snake locomotion. *Copeia*, 1986:915–927, 1986.
- [52] H. Marvi and David L. Hu. Friction enhancement in concertina locomotion of snakes. *Journal of The Royal Society Interface*, 9:3067–3080, 2012.
- [53] K. J. Dowling. *Limbless locomotion: learning to crawl with a snake robot*. PhD thesis, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1996. Unpublished PhD Thesis.
- [54] S. Hirose and M. Mori. Biologically inspired snake-like robots. In *2004 IEEE International Conference on Robotics and Biomimetics*, pages 1–7, Shenyang, China, 2004. IEEE.
- [55] G. Byrnes and B. C. Jayne. The effects of three-dimensional gap orientation on bridging performance and behavior of brown tree snakes (*boiga irregularis*). *Journal of Experimental Biology*, 215:2611–2620, 2012.
- [56] H. Marvi, C. Gong, N. Gravish, H. Astley, M. Travers, R. L. Hatton, J. R. Mendelson, H. Choset, D. L. Hu, and D. I. Goldman. Sidewinding with minimal slip: Snake and robot ascent of sandy slopes. *Science*, 346:224–229, 2014.
- [57] T. Takemori, M. Tanaka, and F. Matsuno. Ladder climbing with a snake robot. *IEEE International Conference on Intelligent Robots and Systems*, pages 8140–8145, 2018.
- [58] K. Lipkin, I. Brown, A. Peck, H. Choset, J. Rembisz, P. Gianfortoni, and A. Naaktgeboren. Differentiable and piecewise differentiable gaits for snake robots. *IEEE International Conference on Intelligent Robots and Systems*, pages 1864–1869, 2007.

- [59] B. C. Jayne. What defines different modes of snake locomotion? *Integrative and Comparative Biology*, 60:156–170, 2020.
- [60] H. Kurokawa, K. Tomita, A. Kamimura, S. Kokaji, T. Hasuo, and S. Murata. Distributed self-reconfiguration of M-TRAN III modular robotic system. *International Journal of Robotics Research*, 27:373–386, 2008.
- [61] M. Tanaka, M. Nakajima, Y. Suzuki, and K. Tanaka. Development and control of articulated mobile robot for climbing steep stairs. *IEEE/ASME Transactions on Mechatronics*, 23:531–541, 2018.
- [62] Q. Fu, T. W. Mitchel, J. S. Kim, G. S. Chirikjian, and C. Li. Continuous body 3-D reconstruction of limbless animals. *Journal of Experimental Biology*, 2021.
- [63] J. Whitman, N. Zevallos, M. Travers, and H. Choset. Snake robot urban search after the 2017 Mexico City earthquake. *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics, SSRR 2018*, pages 1–6, 2018.
- [64] M. Nakajima, M. Tanaka, K. Tanaka, and F. Matsuno. Motion control of a snake robot moving between two non-parallel planes. *Advanced Robotics*, 32:559–573, 2018.
- [65] B. Kehoe, S. Patil, P. Abbeel, and K. Goldberg. A survey of research on cloud robotics and automation. *IEEE Transactions on Automation Science and Engineering*, 12:398–409, 2015.

Chapter 41

Vibrational Robots in Complex Environments

ALESSANDRO MAGAZZÙ

In this Chapter, we investigate the motion of macroscopic active agents in complex environments using centimeter-scale robotic Hexbugs. We begin by describing the experimental setup, including the arena design and modifications to the active agents. The interactions between active agents and passive obstacles are analyzed to identify emergent behaviors. To characterize motion, we introduce key observables such as mean squared displacement (MSD) and velocity distributions. Next, we explore cooperative and anti-cooperative behaviors that influence collective dynamics. We examine the formation and stability of temporary channels as a function of agent density and obstacle properties. Our results highlight the effects of confinement and interactions on swarm-like behaviors, providing insights into self-organization in active matter. These findings have implications for swarm robotics and autonomous systems.



Figure 41.1: **Toy robots as macroscopic active agents.** Toy robots (Hexbug nano) used as active agents in our experiments. These battery-powered vibrating robots exhibit self-propelled motion, allowing us to study their interactions with passive obstacles and the emergence of collective behaviors. By analyzing their trajectories, we investigate key phenomena such as cooperative and anti-cooperative interactions, temporary channel formation, and the influence of environmental constraints on active motion. This model system provides insight into universal properties of active matter and self-organization [1].

41.1 Macroscopic active agents in complex environments

In Nature, it is common to observe individual agents arranging themselves into biological structures such as swarms of bacteria, flocks of birds, schools of fish, herds of land animals, and even human crowds. The self-organizing ability of such systems does not depend on the complexity of the individual agents but emerges from repeated interactions between the agents and their surrounding environment [2, 3, 4, 5]. In particular, swarming systems exhibit universal and emergent behaviors across different length and time scales, from bacterial colonies to human crowds, independently of the type of agents constituting the swarm [6, 7]. Over the past few decades, various models have been developed to understand swarm behavior at the macroscale.

In 1987, Reynolds developed the Boids model to simulate the aggregate motion of a flock of birds through computer animation. This model represents individual birds as autonomous agents choosing their direction of motion based on local perception, avoiding collisions with both other birds and environmental obstacles. The collective motion emerges from dense interactions among the simple behaviors of these simulated birds [8]. A few years later, in 1995, Vicsek et al. [9] introduced what is now known as the Vicsek model, where a swarm is represented as a collection of self-propelling particles moving at a constant speed while tending to align with the average direction of motion of their local neighbours.

Today, the study of collective motion spans multiple disciplines and scales, from microbiology—examining the movement of bacteria and cells [10]—to synthetic micro-robots [11] and even social transport systems [12]. A deeper understanding of collective behaviors has significant applications, including swarm robotics, autonomous vehicles, and high-accuracy cancer treatment [4, 13, 14]. In particular, models describing swarm self-organization can be leveraged to develop artificial systems that move and coordinate in response to their surrounding environment, such as robotic explorers for hazardous or extraterrestrial terrains [15, 16, 17].

41.2 Setting a complex environment for macro agents

To investigate the motion of macroscopic active agents in controlled conditions, it is essential to design an experimental setup that accurately captures their interactions with obstacles and boundaries. The setup must provide a well-defined and reproducible environment, ensuring precise tracking of agent trajectories while minimizing external influences such as uneven lighting, background noise, and uncontrolled boundary effects.

In this section, we describe the construction of the experimental arena, the choice and modification of active agents, the selection of passive particles that shape the environment, and the role of boundary constraints in influencing agent dynamics. By carefully tuning these parameters, we create a system where emergent behaviors — such as cooperative and anti-cooperative interactions, temporary channel formation, and collective motion — can be systematically analyzed.

41.2.1 Building the arena

To accurately observe and record the motion of our active agents, we designed a controlled experimental setup with proper lighting and a homogeneous background. The goal is to ensure clear tracking of the robotic Hexbugs and their interactions with passive obstacles. Here is how we built it (Fig. 41.2a):

1. **Prepare the arena:** A transparent glass panel measuring 113×98 cm is used as the main surface where the Hexbugs and obstacles will move.
2. **Set up the lighting:** Four spotlights are positioned on the floor at 45° angles, illuminating the glass panel from all sides. Nine additional spotlights are positioned in a 3×3 grid above the arena. These will shine onto a diffusive paper sheet, ensuring even lighting.
3. **Enhance visibility with a diffusive paper sheet:** A diffusive panel sheet is placed a few centimeters above the arena to soften shadows and reduce glare. This creates a uniform background, improving the contrast between bugs and obstacles.
4. **Secure the boundary frame:** A sturdy frame is installed around the arena to prevent the Hexbugs from falling off.

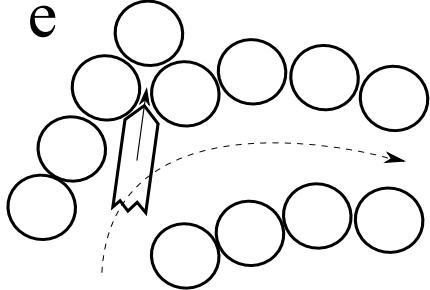
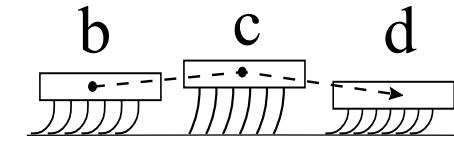
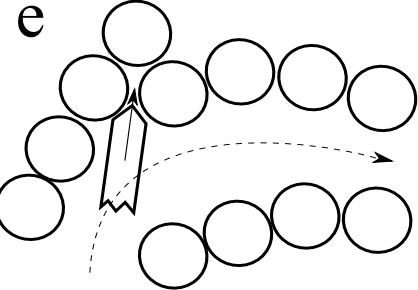
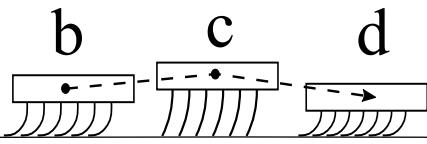
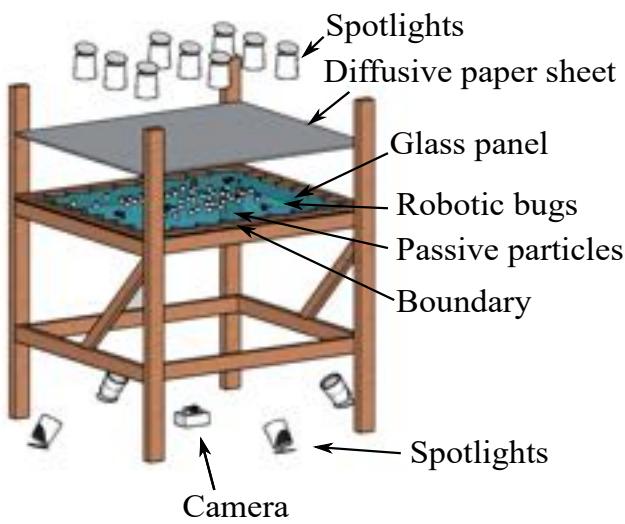
a

Figure 41.2: Arena, Hexbug propelling mechanism, and Hexbug’s interaction with the obstacles. (a) The setup consists of a wooden frame with four table legs, 13 spotlights, a diffusive paper sheet, a transparent glass surface, a boundary frame to prevent particles from falling off the glass, and a camera centered at the bottom of the structure. (b) A toy Hexbug in its initial rest position. (c) As the internal load rotates upward, the Hexbug’s center of mass shifts vertically, reducing the bending of its rubber legs. The legs act like a releasing spring, pushing the Hexbug forward due to the friction between the legs and the surface. (d) As the internal load rotates downward, the Hexbug is pushed down, and its legs act like a compressing spring. Again, due to the specific leg angle and friction with the surface, this compression results in forward propulsion. (e) A robot Hexbug can become stuck between two obstacles (black circles) due to its sharp head, attempting to push them apart and create a new channel instead of following a pre-existing one. The solid black arrow indicates the actual direction of the Hexbug, while the dashed black arrow represents the ideal trajectory.

5. **Position the camera for tracking:** A wide-angle HD camera mounted at the bottom of the setup, pointing upwards is used to record at 30 frames per second (fps) for smooth motion tracking.

With this setup, we create an ideal environment for studying the movement patterns, interactions, and emergent behaviors of our active agents.

41.2.2 Hexbugs as active agents

The active agents used in our experiments are small toy robot Hexbugs (length 4.4 cm, width 1.2 cm, height 1.7 cm, mass 7 g) powered by batteries. These are manufactured by Innovation First, Inc. and commercialized as Hexbug nano (Fig. 41.1). Each Hexbug has six pairs of angled, rubbery legs and an inner electric motor (DC 1.3 V) with an estimated speed of 8000 rpm. Attached to the motor shaft is a semi-cylindrical unbalanced mass. The motor is mounted horizontally (along the length of the Hexbug’s body) so that the angular momentum of the motor is directed along the body axis [18].

When powered, the rotation of the unbalanced mass shifts the center of mass vertically, causing the Hexbug to vibrate along the vertical direction. Due to the curved and flexible legs and to their friction with the bottom surface, this vertical vibration results in forward propulsion, as schematically shown in Fig. 41.2b-d. A single battery can power a Hexbug for up to four hours before needing replacement. It is noteworthy that the Hexbug does not move on surfaces having a very small friction with the legs.

Our active agents exhibit random chirality in their motion, likely due to small asymmetries in the legs, which are inherent to the manufacturing process. Different Hexbugs display varying degrees of chirality, which cannot be predicted in advance. Although this behavior seems significant, it becomes negligible in the presence of passive particles. This is because collisions between the Hexbugs and obstacles reduce their mean free path, making the effects of chirality less noticeable. Over time, as the battery level decreases, the Hexbug’s velocity

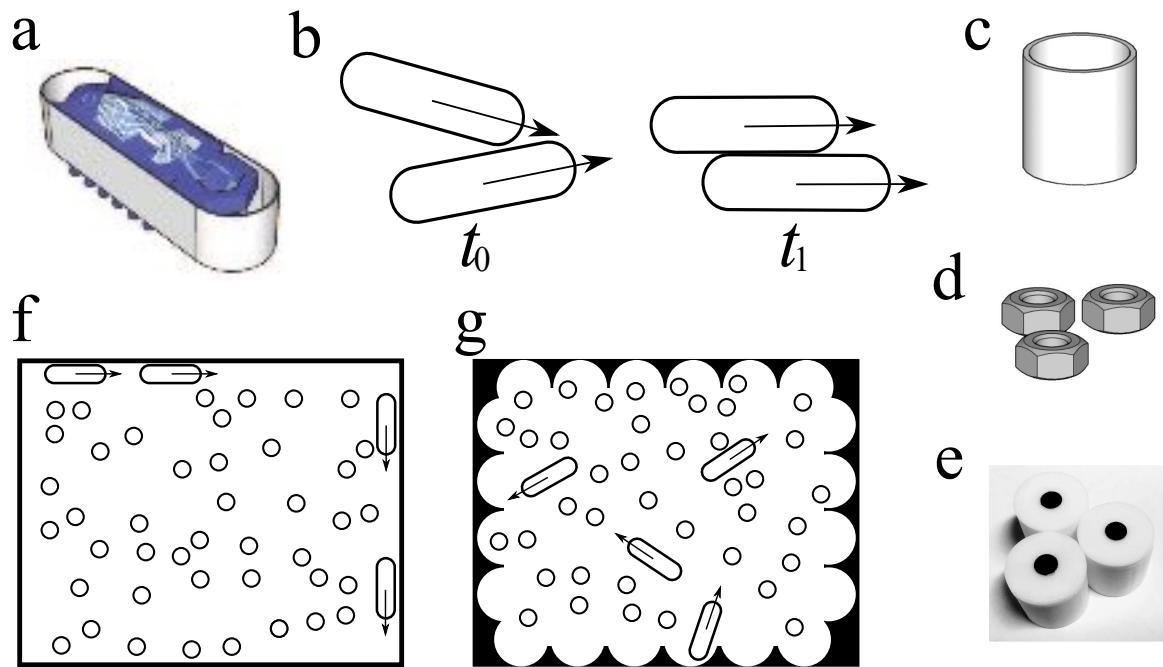


Figure 41.3: Modified Hexbugs and their interaction, obstacles and arena frames. (a) A toy robot Hexbug nano with a strip of paper around its body. The strip of paper acts as a bumper, making the Hexbug's head smoother and preventing them from climbing on each other, flipping, and from getting stuck between obstacles. (b) Sketch of two Hexbugs, having different directions, colliding at t_0 and assuming the same direction after the collision at $t_1 > t_0$. (c) A 3D printed plastic cylindrical cup used as passive particle to create a complex environment for robot Hexbugs. (d) M8 metal nuts used to increase the weight of the passive particles, by placing them inside the cups. (e) Bottom view of the cups, the black dots facilitate the automated counting and tracking of the passive particles. (f) Sketch of rectangular boundary frame, where active agents (black and smooth edge rectangles) tend to move along the boundary walls, avoiding interaction with other agents and obstacles (black circles), black arrows indicate the direction of the active agents. (g) Rounded-shaped boundary frame, where the Hexbugs once reach the edges are re-orientated toward the centre of the arena.

also decreases, making its motion more chiral, often following circular trajectories with progressively smaller radii.

When powered by new batteries, the Hexbugs move vigorously and sometimes climb on top of each other or tip upside down. Another issue arises due to their sharp head, which causes them to get stuck between heavy obstacles, attempting to push them for extended periods instead of following a pre-existing channel, as shown in Fig. 41.2e. To address these issues, we applied a strip of paper around each Hexbug, as shown in Fig. 41.3a. This modification serves as a bumper, making the head less sharp and allowing the Hexbugs to bounce off obstacles instead of getting stuck. Additionally, the paper strip provides lateral stability, preventing tipping and giving the Hexbugs a smooth rod-like shape, which aligns better with theoretical active matter models [19, 20]. Furthermore, this rod-like geometry promotes alignment interactions, facilitating the emergence of collective behavior. As shown in Fig. 41.3b, when two Hexbugs travelling in different directions collide at t_0 , they align during the collision and start moving together in the same direction at t_1 where $t_1 > t_0$. Lastly, although the Hexbugs are available in different colors, we painted their bottom surface black to increase contrast with the white background, improving automatic detection in video recordings.

41.2.3 Setting the passive particles

The passive particles used in our experiments to create a complex environment are 3D-printed plastic cylindrical cups with a diameter of 19.5 mm, a height of 20 mm, and a weight of 2 g (Fig. 41.3c). The weight of these passive particles can be increased by inserting M8 metal nuts inside them (Fig. 41.3d). Each M8 nut weighs

approximately 5 g, and each passive particle can accommodate up to three nuts. Consequently, the total weight of a passive particle can be: 2 g (empty cup), 7 g (cup with one M8 nut), 12 g (cup with two M8 nuts), and 17 g (cup with three M8 nuts).

The passive particles are made of white plastic, allowing them to be easily distinguished from the black Hexbugs. Additionally, each particle is marked with a black dot on its bottom to facilitate automatic tracking and counting during video analysis (Fig. 41.3e).

41.2.4 The right boundary frame

An essential component of the arena is the frame, which prevents the Hexbugs from falling off. However, not all frame designs are suitable for our experiment. A square-shaped arena tends to cause Hexbugs to align with the walls, leading them to spend most of their time moving along the boundaries and limiting their interactions to the frame edges (see Fig. 41.3f). To overcome this issue, a rounded-shaped boundary can be used. This particular design redirects the Hexbugs toward the center of the arena whenever they reach the edges, encouraging more frequent interactions with themselves and with the obstacles (see Fig. 41.3g).

41.3 Hexbugs moving in the arena

In our experiments, we tested different combinations of the number of Hexbugs (N_B), the number of cups (N_C), and the number of M8 metal nuts placed inside the cups (N_W). The specific values we used were:

- $N_B = \{1, 2, 5, 10, 15, 20, 25\}$
- $N_C = \{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300\}$
- $N_W = \{0, 1, 2, 3\}$

Here, N_W refers to the number of nuts placed inside each obstacle. This affects the weight of a passive particle, m_p , which we calculated as

$$m_p = N_W(i) \cdot m_{\text{nut}} + m_{\text{cup}}, \quad (41.1)$$

where each nut weighs $m_{\text{nut}} = 5$ g, and each cup weighs $m_{\text{cup}} = 2$ g. To cover all possible setups, we ran a total of experiments:

$$N_{\text{exp}} = |N_B| \cdot |N_C| \cdot |N_W| + |N_{\text{nc}}| = 370 \quad (41.2)$$

where we also added six additional tests without cups, using different numbers of Hexbugs:

- $N_{\text{nc}} = \{1, 2, 5, 11, 16, 21\}$.

The notation $|A|$ represents the cardinality of the set A , denoting the number of elements it contains.

For each experiment, we first set up the arena by placing the correct number of cups and nuts according to N_C and N_W . Then, we started recording with just one Hexbug and gradually added more every 200 seconds until we reached the target N_B . This method let us capture about 24 minutes of footage per session, covering seven experiments in a single video.

To make sure everything was in order, we used a MatLab script to check the number of cups before every recording. Once the videos were captured, we split them into shorter 3-minute clips, each containing only one specific combination of N_C , N_B , and N_W . During this step, we carefully removed any sections with unwanted disturbances — like shadows or hands appearing when adding Hexbugs.

Next, we used MatLab to correct the wide-angle lens distortion, using a pre-drawn rectangular grid on the arena as a reference. To improve contrast, we converted the footage to black and white, making it easier to distinguish the Hexbugs and obstacles from the background.

Finally, another MatLab script tracked the Hexbugs and obstacles, recording their trajectories and orientations. A typical Hexbug's movement is stored as two data vectors, each with N elements (one for x and one for y positions). Here, N is the total number of frames in the video. The time interval Δt between consecutive frames is:

$$\Delta t = t_{i+1} - t_i = 1/\text{fps} = 0.03 \text{ s}. \quad (41.3)$$

From this, we calculate the time sequence of the Hexbug's positions as:

$$t_i = (i - 1) \cdot \Delta t. \quad (41.4)$$

Since our tracking data is initially in pixel units, we converted it to meters using the known size of the arena. After correcting for lens distortion, we determined that 1 pixel corresponds to 1.4 mm, which we used to convert all measurements.

41.4 What information can we extract from the Hexbug trajectories?

Analyzing the trajectories of the Hexbugs provides critical insights into their motion patterns, interactions, and emergent behaviors within the experimental environment. By quantifying key observables such as mean squared displacement (MSD) and velocity distributions, we can distinguish different motion regimes, including diffusive, subdiffusive, and ballistic transport. These metrics allow us to assess how environmental factors, such as obstacle density and weight, influence the mobility of the active agents. Additionally, tracking individual and collective motion patterns enables the identification of cooperative and anti-cooperative behaviors, temporary channel formation, and other emergent phenomena characteristic of active matter systems. In this section, we outline the computational methods used to extract these quantities and discuss their implications for understanding self-organization in vibrational robots.

41.4.1 Calculating the mean square displacement

The *mean squared displacement* (MSD) quantifies how a particle moves away from its initial position $x(t)$ over a specific time interval τ , providing insights into the diffusion behavior of the particle. For example, in the case of *normal diffusion*, which is typical for a random walk, such as Brownian motion, there is a linear relationship between MSD and τ : $MSD = 4D\tau$, where D is the diffusion constant (and the prefactor 4 is appropriate for a motion in a two-dimensional space). When a random walker or a Brownian particles moves in a crowded environment, such as in presence of obstacles, it typically exhibits *subdiffusion*, characterized by $MSD \propto \tau^\alpha$

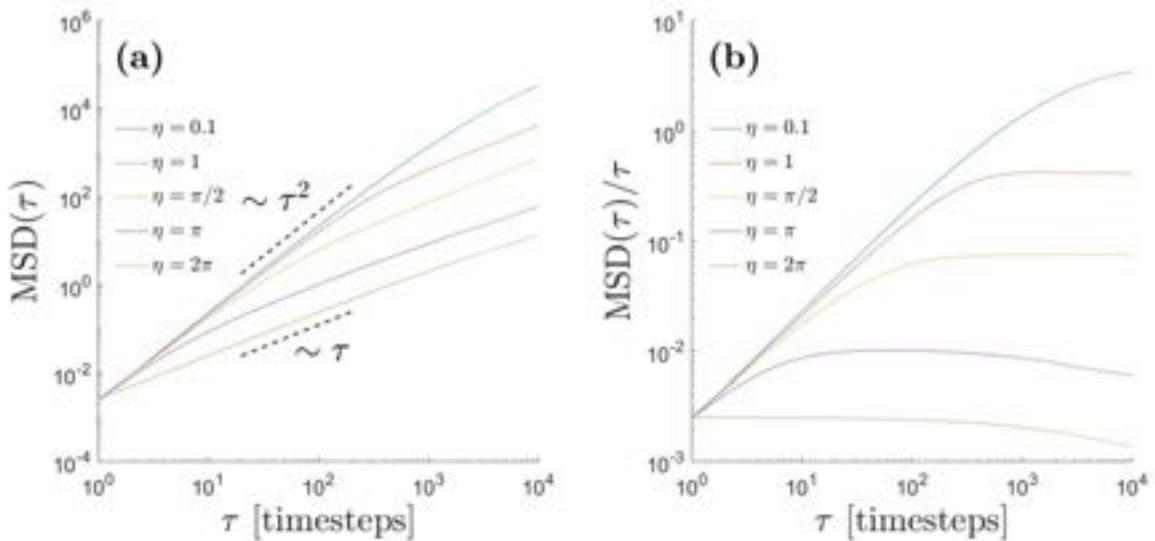


Figure 41.4: **Mean square displacement of active particles in the presence of passive particles.** (a) MSD calculated for active particles with different obstacle noise levels η . As η decreases, particle activity increases. The MSD exhibits a transition from ballistic motion at short times ($MSD \propto \tau^2$ for small τ) to diffusive motion at long times ($MSD \propto \tau$ for large τ). The crossover time between these two regimes increases as the noise level decreases. (b) The ratio MSD/τ decreases in the long-time limit for high noise levels (corresponding to lower particle activity), indicating subdiffusion. Reproduced from Ref. [21].

with $0 < \alpha < 1$, In contrast, active agents exhibit active motion and *superdiffusion*, described by $\text{MSD} \propto \tau^\alpha$ with $1 < \alpha < 2$. Finally, when a particle moves at constant velocity, it undergoes ballistic motion, with $\alpha = 2$ [22, 23].

The MSD can provide valuable information about the motion of the Hexbugs in different scenarios (different combinations of N_B , N_C , and N_W) and may eventually indicate the formation of temporary channels [21]. The MSD along the x direction is defined as

$$\text{MSD}_x(\tau) = \langle |x(t + \tau) - x(t)|^2 \rangle = \frac{1}{N - \tau} \sum_{i=1}^{N-\tau} |x(t + \tau) - x(t)|^2. \quad (41.5)$$

Once we calculate N elements x_i and y_i at different time steps, it becomes more practical to compute the MSD of the Hexbugs as a function of the number n of the time steps t_i as

$$\text{MSD}(n) = \langle |x_{i+n} - x_i|^2 + |y_{i+n} - y_i|^2 \rangle = \frac{1}{N - n} \sum_{i=1}^{N-n} |x_{i+n} - x_i|^2. \quad (41.6)$$

Once we calculate $\text{MSD}(n)$, we can rewrite it as a function of τ , $\text{MSD}(\tau)$, considering that $\tau = n \cdot \Delta t$ with $n = 1, 2, 3, \dots, N - 1$. In 2017, Nilsson and Volpe used the MSD to explore how simulated active particles move when obstacles are present [21]. What they found was that, as the number of obstacles decreased, the active agents naturally formed stable channels — pathways that other active agents could also use. These channels did not only make movement easier; they also increased the characteristic length of active particles when passive particles were around. This effect showed up clearly in the MSD, which revealed a shift from diffusive to ballistic motion. Essentially, the particles moved more freely and the crossover time between these two motion types grew longer (Fig. 41.4a). Furthermore, the MSD also acts as a indicator of subdiffusion in the long-time limit. Looking at Fig. 41.4b, it is possible to see that $\frac{\text{MSD}(\tau)}{\tau}$ decreases for large τ values when enough obstacles are present, showing how the obstacles can slow the active agents down over time.

41.4.2 Calculating the Hexbug velocity

The velocity of the Hexbugs is another important quantity that provides insights into their motion. Given the Hexbug coordinates x_i and y_i the Hexbug velocity $v(t)$ can be computed as

$$v(t_i) = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{\Delta t}, \quad (41.7)$$

where Δt is the time step between two frames.

In Fig. 41.5(a-d), we present the probability distributions of the velocity $P(v)$ for different combination of $N_B = \{1, 5, 15, 25\}$ and $N_C = \{100, 300, 500, 800\}$ while keeping $N_W = 0$. A notable feature in all four cases of N_B is the presence of a peak in $P(v)$ at $v = 0$ m/s. This arises because, even with a low number of obstacles $N_C = 100$, the velocity $v(t)$ can still occasionally reach zero due to Hexbug–Hexbug, Hexbug–obstacle, and Hexbug–boundary collisions.

Additionally, we observe that as N_C increases, $P(v)$ flattens and shifts toward lower velocity values. This trend becomes even more pronounced as the number of Hexbugs N_B increases as seen in Fig. 41.5a-c. The underlying reason for this behavior can be seen in Fig. 41.6a-c, where traffic congestion among the Hexbugs is visible, often persisting for several time intervals Δt . This hindering behavior happens more frequently as the number of Hexbugs increases, and as a consequence, their ensemble mean velocity decreases. As the Hexbug density rises, such congestion events become more frequent, leading to a reduction in the ensemble mean velocity. As the number of obstacles increases, the differences between the velocity distributions $P(v)$ for various Hexbug numbers diminish (Fig. 41.5c-d). This occurs because, at high N_C Hexbug velocity $v(t)$ becomes less influenced by Hexbug–Hexbug interactions and more constrained by obstacle crowding. In extreme cases, such as highly crowded environments with many heavy passive particles ($N_C=800$ and $N_W=3$), $P(v)$ exhibits narrow distributions at lower velocities (Fig. 41.5d). This is due to the overwhelming presence of obstacles, which significantly hinder Hexbug motion.

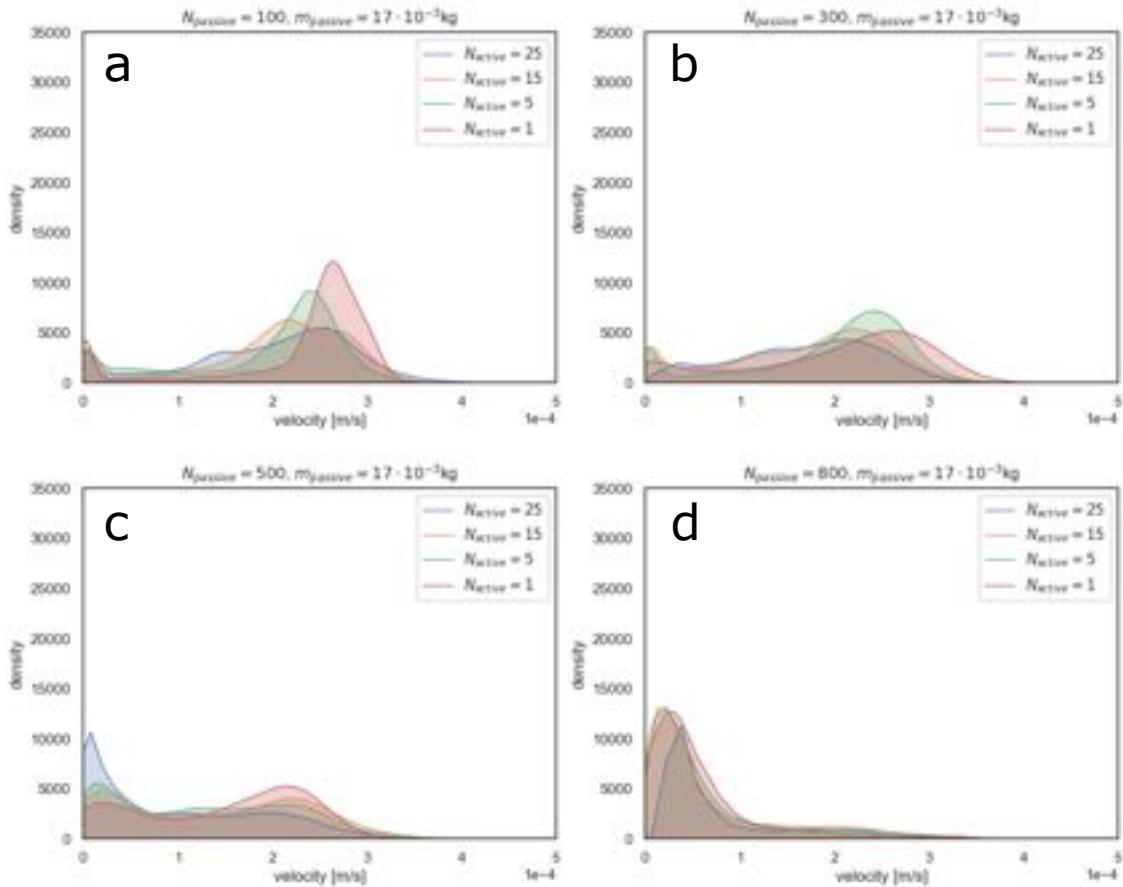


Figure 41.5: Velocity probability distributions. Velocity probability distributions $P(v)$ for different number of Hexbugs $N_B = \{1, 5, 15, 25\}$ (colored lines) calculated for different number of cups N_C = (a) 100, (b) 300, (c) 500 and (d) 800 with $N_W = 3$ corresponding to a passive mass of $m_{\text{passive}} = 17\text{g}$.

41.5 Collective behaviors of the Hexbugs

When multiple Hexbugs move within the experimental arena, their interactions give rise to a variety of collective behaviors that resemble those observed in natural active matter systems, such as bacterial swarms or self-propelled particles. These behaviors emerge from local interactions rather than centralized control, leading to complex group dynamics. Depending on environmental constraints and agent density, the Hexbugs exhibit both cooperative and anti-cooperative behaviors, influencing their ability to navigate the obstacle-laden space. In particular, collective effects play a crucial role in the formation and stability of temporary channels, where multiple agents self-organize into transient pathways. Understanding these behaviors provides valuable insights into swarm dynamics, transport processes in disordered environments, and potential applications in swarm robotics. In this section, we characterize the key collective behaviors observed in our experiments and discuss their implications.

41.5.1 Anti-cooperative behaviors

During our experiments, we observed various collective behaviors, some of which occurred more frequently than others. The most common behavior was Hexbug accumulation at the boundary edges of the arena. This issue was particularly pronounced in rectangular arenas, where Hexbugs tended to get stuck in the corners due to the chirality of their motion. With the round-shaped arena, Hexbugs could only become trapped when approaching the boundary from opposite directions, colliding, and pushing against each other. In such cases,

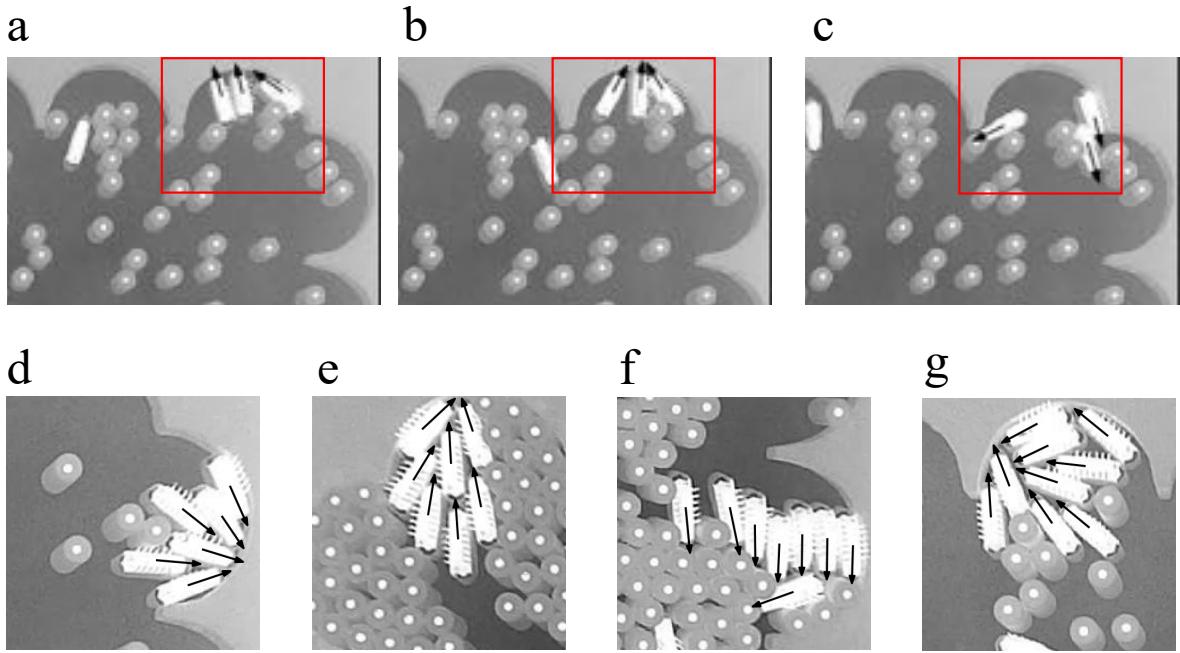


Figure 41.6: Video frames showing Hexbug congestion and anti-cooperative behavior. Red squares and black arrows indicate congestion areas and the Hexbug direction respectively. (a) After approximately $t = 32$ s from the beginning of the experiment, three Hexbugs come into contact at the arena boundary, hindering each other and creating a traffic congestion. (b) One second later, at $t = 33$ s, the Hexbugs continue to obstruct each other. (c) At $t = 34$ s, the three Hexbugs manage to move again. (d) Video frames showing examples of anti-cooperative behavior for different numbers of Hexbugs: $N_B = 6$, (e) $N_B = 7$, (f) $N_B = 8$, and (g) $N_B = 9$.

the boundary constraint caused their resulting force to be zero, preventing the group from moving until the equilibrium was disturbed. For this reason, we define these types of collective behaviors as anti-cooperative behaviors, as they minimize the movement of active particles. This contrasts with cooperative behaviors, where Hexbugs assist each other in maximizing their motion. In Fig. 41.6d-g, we present examples of anti-cooperative behaviors for different numbers of Hexbugs.

41.5.2 Cooperative behaviors

Collective cooperative behaviors occur when Hexbugs actively help each other move more freely. For example, if a Hexbug gets stuck against an obstacle or another Hexbug, an incoming Hexbug can push or nudge it, allowing it to move again. In Fig. 41.7a-c, we see a clear case where a Hexbug is trapped by an obstacle until a second Hexbug arrives and helps it break free by creating a passage.

41.5.3 Formation of temporary channels

Another recurring behavior is when a Hexbug pushes the obstacles forming a temporary ring-shaped channel that it reuses it for several times, as shown in Fig. 41.7d,e. This metastable channel lasts until the Hexbug manages to open a new path by pushing the obstacles further or other Hexbugs come along and disrupt the ring. These temporary channels are more likely to form near the arena boundary, where the fixed wall helps keep part of the channel in place. Because of this, channels near the boundary tend to last longer than those formed in open spaces, where obstacles can move more freely. Every time a Hexbug reuses a channel, it nudges the obstacles slightly, gradually shifting them and after enough use, the channel eventually is destroyed. This happens more quickly when the obstacles are lighter or when there are more Hexbugs in the arena. A great example of this is in Fig. 41.8a, where a single Hexbug creates a long channel stretching across the arena, which was ruined as more Hexbugs move through. If only one Hexbug is present, the channels it creates tend to last much longer, since there's no interference from other Hexbugs. Another interesting behavior happens

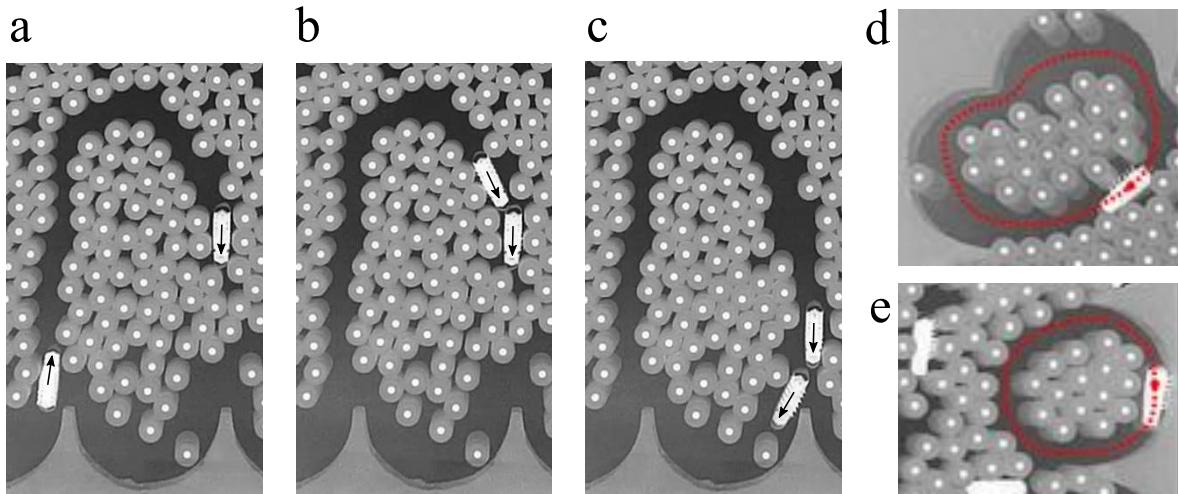


Figure 41.7: **Cooperative behavior and temporary channels.** Video frames showing examples of (a-c) cooperative behavior, where a Hexbug pushes another that is blocked by obstacles. Together they are able to move the obstacles and create a channel. Black arrows indicate the direction of the Hexbugs. This situation occurred for $N_B = 2$, $N_C = 1000$ and $N_W = 1$ at (a) $t = 0s$, (b) $t = 3s$ and (c) $t = 21s$. (d) Hexbug moving in a temporary ring channel for: 7 lapses and (e) 4 lapses. Dashed red lines indicates the Hexbug trajectory and red arrows indicate Hexbug directions.

when multiple Hexbugs start using the same channel. Instead of breaking it apart, they follow along the same path. You can see this in Fig. 41.8b, where different Hexbugs move along a pre-existing channel.

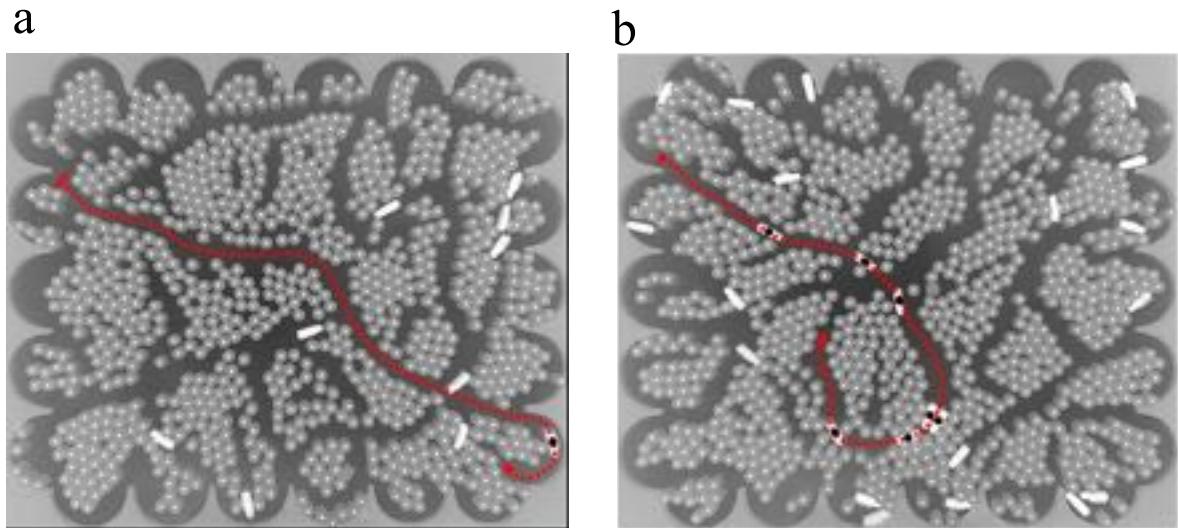


Figure 41.8: **Long temporary channels.** Dashed red lines indicates the Hexbug's trajectory, red symbols \vdash mark the point where the channels starts, and red arrows indicates the direction of the Hexbugs and the end of channels before they are destroyed by other Hexbugs. (a) A video frame showing a Hexbug forming a long temporary channel crossing the entire arena, from the upper left corner to the lower right corner. The Hexbug that created the channel is labelled with a black dot. This scenario occurred for $N_B = 10$, $N_C = 900$ and $N_W = 0$. It is interesting to note the other Hexbugs that are about to cross this channel, opening new paths and ruining it. (b) A video frame showing 7 Hexbugs moving along a pre-existing temporary channels. This behavior occurred for $N_B = 25$, $N_C = 900$ and $N_W = 0$. All the Hexbugs travelling in this channels are labelled with black dots.

As mentioned in the velocity section, a Hexbug's velocity can serve as an indicator of its activity and what it is actually doing in the arena. For example, is it moving freely, or is it pushing obstacles and forming channels? Is it following a pre-existing channel?

These different activities correspond to distinct velocity ranges: a Hexbug moves slowly when pushing obstacles, whereas it moves faster when traveling along a pre-existing channel.

41.6 Problems

During data analysis, it is not always straightforward to deduce the collective behaviors of our active agents solely from the averaged mean squared displacement (MSD) of all individual Hexbugs in an experiment. This is because, in many cases, only a small fraction of the Hexbugs in the arena exhibited collective behavior, and only for a short duration.

You can watch and download the video at the url reported in [24]. Where in the data file name are included all the information about the experimental parameters N_B , N_C and N_W , e.g. a file name: 1W1000C2B indicates that the file or folder is about the experiment with $N_W = 1$, $N_C = 1000$ and $N_B = 2$.

Problem 41.1: Visual analysis of the experiments. Perform a visual analysis of the video experiments. Identify and take note of the various Hexbug behaviors mentioned earlier in this chapter (e.g., performing diffusive or sub-diffusive motion, forming channels, exhibiting cooperative or anti-cooperative behavior, or moving within a channel). Additionally, observe any other behaviors that seem interesting to you.

Problem 41.2: Tracking the agents. Track the positions of the Hexbugs and obstacles over time using any method, software, script, or programming language of your choice [25]. This will allow you to obtain trajectories, the number of Hexbugs, and the configurations and number of obstacles.

Problem 41.3: Mean square displacement and velocity. Calculate the $MSD(\tau)$ and $v(t)$ for the Hexbugs reported in the files 1W300C25B, 1W900C25B, 1W1300C25B as the number of obstacles increases, looking for any transition in the MSD trend indicating a transition of the Hexbug's motion (e.g.: from diffusive to sub-diffusive motion).

Problem 41.4: Channel formation. Analyze channel formation: Compute $MSD(\tau)$ and $v(t)$ for the Hexbug in file 2W800C1B both during the formation of a channel and after the channel is created when the Hexbug moves within it. This analysis should help determine characteristic values of $MSD(\tau)$ and $v(t)$ for different Hexbug behaviors, providing insights into Hexbug activity without relying solely on video analysis. Use the positions of obstacles around the Hexbug to better understand and verify its behavior.

Problem 41.5: Cooperative behavior in channel opening. Examine cooperative behavior in channel opening: Compute $MSD(\tau)$ and $v(t)$ for the Hexbugs in file 1W1000C2B before, during, and after cooperative behavior in opening a channel. Compare their motion characteristics before the channel formation, during the cooperative effort, and after the channel is established. This analysis should help identify characteristic values of $MSD(\tau)$ and $v(t)$ for different behaviors, facilitating behavioral interpretation without direct video observation. Use the positions of obstacles around the Hexbugs to further validate and interpret their activities.

Bibliography

- [1] Hexbug. <http://www.hexbug.com/hexbugnano-single.html#description/>.
- [2] D. J. T. Sumpter. The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1465):5–22, 2006.
- [3] J. Gautrais, F. Ginelli, R. Fournier, S. Blanco, M. Soria, H. Chaté, and G. Theraulaz. Deciphering interactions in moving animal groups. *PLOS ONE*, 2012.
- [4] C. Bechinger, R. Di Leonardo, H. Löwen, C. Reichhardt, G. Volpe, and G. Volpe. Active particles in complex and crowded environments. *Reviews of Modern Physics*, 88:045006, 2016.

- [5] S. Ramaswamy. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics*, 1(1):323–345, 2010.
- [6] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517(3-4):71–140, 2012.
- [7] J. Buhl, D. J. T. Sumpter, I. D. Couzin, J. J. Hale, E. Despland, E. R. Miller, and S. J. Simpson. From disorder to order in marching locusts. *Science*, 312(5778):1402–1406, 2006.
- [8] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, pages 25–34, 1987.
- [9] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6):1226, 1995.
- [10] H. Wioland, E. Lushi, and R. E. Goldstein. Directed collective motion of bacteria under channel confinement. *New Journal of Physics*, 18(7):075002, 2016.
- [11] A. A. Solovev, S. Sanchez, and O. G. Schmidt. Collective behaviour of self-propelled catalytic micromotors. *Nanoscale*, 5(4):1284–1293, 2013.
- [12] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4):1067, 2001.
- [13] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo. Swarm robotics: A review from the swarm engineering perspective. *Swarm Intelligence*, 7(1):1–41, 2013.
- [14] J. Wang and W. Gao. Nano/microscale motors: Biomedical opportunities and challenges. *ACS Nano*, 6(7):5745–5751, 2012.
- [15] M. Mijalkov, A. McDaniel, J. Wehr, and G. Volpe. Engineering sensorial delay to control phototaxis and emergent collective behaviors. *Physical Review X*, 6(1):011008, 2016.
- [16] M. Rubenstein, A. Cornejo, and R. Nagpal. Programmable self-assembly in a thousand-robot swarm. *Science*, 345(6198):795–799, 2014.
- [17] J. Werfel, K. Petersen, and R. Nagpal. Designing collective behavior in a termite-inspired robot construction team. *Science*, 343(6172):754–758, 2014.
- [18] A. Barona Balda, A. Argun, A. Callegari, and G. Volpe. Playing with active matter. *American Journal of Physics*, 92:847–858, 2024.
- [19] F. Peruani. From individual to collective motion of self-propelled particles: The role of particle shape, orientational ordering and clustering. *Physical Review E*, 78(3):031917, 2008.
- [20] W. F. Paxton, K. C. Kistler, C. C. Olmeda, A. Sen, S. K. St. Angelo, Y. Cao, T. E. Mallouk, P. E. Lammert, and V. H. Crespi. Catalytic nanomotors: Autonomous movement of striped nanorods. *Journal of the American Chemical Society*, 126(41):13424–13431, 2004.
- [21] S. Nilsson and G. Volpe. Metastable clusters and channels formed by active particles with aligning interactions. *New Journal of Physics*, 19(11):115008, 2017.
- [22] P. H. Jones, O. M. Maragò, and G. Volpe. *Optical tweezers: Principles and applications*. Cambridge University Press, 2015.
- [23] J. Gieseler, J. R. Gomez-Solano, A. Magazzù, I. P. Castillo, L. P. García, M. Gironella-Torrent, X. Viader-Godoy, F. Ritort, G. Pesce, A. V. Arzola, et al. Optical tweezers: A comprehensive tutorial from calibration to applications. *arXiv preprint arXiv:2004.05246*, 2020.
- [24] A. Magazzu. Bug's videos. <https://1drv.ms/f/c/a72693fe4e25d895/EgjQXIaf-3VDpzI7-XvYO-IBRSqHWhhXH1xWBzoz6bBYAw?e=35pm8c/>.
- [25] B. Midtvedt, S. Helgadottir, A. Argun, J. Pineda, D. Midtvedt, and G. Volpe. Quantitative digital microscopy with deep learning. *Applied Physics Reviews*, 8(1), 2021.

Chapter 42

Experiments with Robot Swarms

VITO TRIANNI

Moving beyond traditional, monolithic robotics, the field of *swarm robotics* proposes an enticing avenue where the collective might of numerous interconnected robots triumphs over the capabilities of a solitary machine, as shown in Fig. 42.1. Inspired by the reliability and adaptability of natural swarm systems, swarm robotics seeks to achieve sophisticated tasks through the self-organized activity of relatively simple autonomous units, each bereft of intricate sensing, substantial computation, or powerful actuation capabilities. The collective behavior of these swarms not only achieves precise sensing, decision-making, and manipulation abilities, but also ensures system robustness, resilience, and adaptability to failure. Drawing parallels with active matter, swarm robotics becomes a paradigmatic example of out-of-equilibrium systems, demonstrating ordered spatio-temporal patterns springing from the actions and interactions of individual units. This Chapter delineates the key principles and experimental studies underpinning swarm robotics, commencing with the exploration of minimal complexity movement to understanding group behaviors such as aggregation, collective navigation, and decision-making. The discourse underscores the promise held by swarm robotics in confronting challenges presented by tasks like search-and-rescue missions or environmental monitoring, paving the way for their



Figure 42.1: **Strength in numbers.** Swarm robotics focuses on the design of robotic systems made of a number of autonomous, independent units that are able to self-organize, cooperate and self-assemble, facing together complex problems with which the single units cannot cope. In this picture, the robots designed within the SWARM-BOTS project [1] self-assemble and move coordinately to support navigation outdoor on uneven terrains. Picture credit: Marco Dorigo.

increasing prominence in future robotics endeavors.

42.1 Introduction to robot swarms

When we think about robots, we often picture an anthropomorphic machine able to work in the same environments where we live, to perceive the world in the same way as we do, and to act with human-like abilities, although somewhat harder, better, faster, stronger. However, humanoid robots represent a relatively small branch of robotics, as most industrial and field robots have a specific morphology tailored to peculiar ways of navigation, manipulation and interaction with their environment. A common feature of most robots is that they are *monolithic* machines, made of sensors, motors and joints interconnected in a single body. This means that, if one of the components breaks, the functionality of the whole robot may get compromised. Hence, a careful engineering design must be made in order to foresee all possible operational conditions and faults from which to possibly recover.

One of the best examples of a well-engineered robot is for sure Perseverance, the NASA rover landed on Mars in 2021 [2]: it is a beautifully designed robot, the dimension of a small van ($3\text{ m} \times 2.7\text{ m} \times 2.2\text{ m}$), which (at the time of writing) has remained operational on a faraway planet for more than four years. Its Mars rover predecessors (Opportunity and Spirit) were not any less. They were operational on Mars for 14 years and 6 years, respectively: some unforeseeable long times. Spirit ceased operation only because it remained trapped in a sandy area from which it could neither escape nor recharge its batteries. If only Opportunity were designed to collaborate and rescue the twin robot, Spirit could have surely remained operational for much longer.

However, designing robots for collaboration and cooperation is not trivial. Multi-robot systems require attentive design so that fruitful interactions are possible and collaborative plans are deployed and brought forth [3, 4]. The larger the number of units in the system, the more complex their control, requiring a complete rethinking of the strategies to engineer the robotic system. This is the challenge undertaken by swarm robotics, a research domain where robotic systems are composed of a multitude of interacting, cooperating autonomous robots capable of dealing with complex tasks that no single unit can carry out individually [5, 6]. This is possible thanks to self-organization, the underlying principle informing swarm robotics research. This also makes it possible to describe robot swarms as active matter, an out-of-equilibrium system capable of displaying an ordered spatio-temporal pattern as the result of the actions and interactions of the individual robots.

Individual robots within a swarm are often relatively simple with respect to the complexity of the task to accomplish. They may lack precise and long-range sensing devices, sufficient computational power or strong-enough actuators. However, when working as a group, they can achieve accurate collective sensing, group decision-making abilities and collective navigation, transport and manipulation abilities [7, 8, 9, 10]. Thanks to the built-in redundancy of swarm robotics systems composed of several inter-changeable units, robustness to failures comes at essentially no cost, as the malfunctioning of one or a few robots does not largely affect the performance of the whole swarm. This makes robot swarms very appealing in comparison with monolithic systems, especially for tasks that can be executed in parallel by multiple robots, and where high uncertainty and lack of infrastructure requires a resilient, autonomous and adaptive approach. Examples of such tasks include search-and-rescue after natural disasters, as well as environmental monitoring for ecological, agricultural, or industrial purposes.

42.2 Exploration of an unknown environment

An unknown environment is one for which a map is not available, or cannot be exploited for some reason. For instance, even if the map of a building is available, rescue operations following some disaster (fires, quakes) cannot assume that the available map is still valid. Also, robots may not have the computational power to handle a map, or to self-localize within a map (e.g., in case of small flying robots with little payload [9]). When standard approaches used in robotics (such as *self-localization and mapping*, SLAM [11]) cannot be exploited, a very minimalistic alternative to exploration is given by *random walks*. Random walks require very little computational complexity to be implemented, still are generally well understood and can lead to somewhat predictable outcomes. Also, they provide stochasticity to the swarm behavior, generating random fluctuations that can be key to trigger self-organization thanks to well-designed feedback loops. For these reasons, random walks are ubiquitous in swarm robotics research, and are considered the basic mode for navigation and

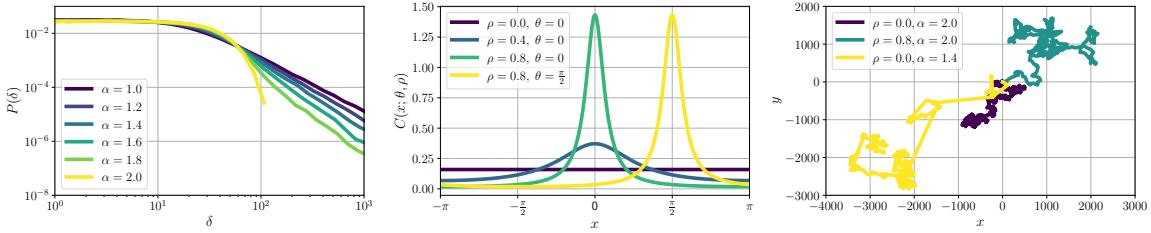


Figure 42.2: **Lévy-modulated correlated random walks (LMCRW).** (a) Power-law distribution of step lengths for different values of α . (b) Wrapped Cauchy distribution of turning angles for different values of ρ and θ . (c) Different types of trajectories obtainable by varying the parameters ρ and α of the LMCRW.

exploration [12].

A random walk can be easily implemented on any robot by alternating straight motion with random rotations on the spot. Different types of random walks can be obtained depending on the distribution of the length of straight steps and the rotation angles. We consider here the case of random walks on a 2D plane, but extensions to 3D space (e.g., in the case of flying or underwater robots) or other topologies are straightforward. A simple and general-purpose approach is the so-called *Lévy-modulated correlated random walk* (LMCRW) [12], which is characterized by a few parameters controlling the distributions of step lengths and turning angles. The step lengths follow a Lévy distribution characterized by a power law $P(\delta) \approx \delta^{-(\alpha+1)}$, with $0 < \alpha \leq 2$ (Fig. 42.2a). For $\alpha = 2$, the distribution becomes Gaussian. For $\alpha < 2$, the distribution has a power-law scaling, meaning that large displacements have non null probability. Finally, for $\alpha \rightarrow 0$, the random walk reduces to ballistic motion. Turning angles are drawn from a wrapped Cauchy distribution having the probability density function described by the following equation:

$$C(x; \theta, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 + 2\rho \cos(x - \theta)}. \quad (42.1)$$

where θ determines the peak of the distribution, while ρ corresponds to the distribution width (Fig. 42.2b). Normally, unbiased rotations require that $\theta = 0$, so that both left and right turns are equally probable. When $\rho = 0$, the distribution becomes uniform, meaning that the choice of the turning angle is isotropic without any correlation between the direction of consecutive movements. As ρ gets larger, the distribution becomes narrower with a peak at θ , meaning that smaller deviation from θ are more likely. In the limit, when $\rho = 1$, a Dirac distribution is obtained, corresponding to continuous rotations of an angle θ , which reduces to ballistic motion when $\theta = 0$. Overall, by playing with the parameters ρ and α , it is possible to obtain many different types of random walk, as shown in Fig. 42.2c: by setting $\rho = 0$ and $\alpha = 2$, we obtain a Brownian motion, which results in a very local search; when $\rho > 0$ and $\alpha = 2$, we obtain a correlated random walk (CRW), with larger diffusion; when $\rho = 0$ and $\alpha < 2$, we obtain a Lévy walk, featuring the characteristics long displacement between periods of local search.

One of the most important values to evaluate the ability of a random walker to diffuse in space is the mean square displacement (MSD) from the starting position, computed as a function of the time elapsed from start. In the case of Brownian motion, for instance, the MSD increases linearly, while for Lévy walks the MSD is superlinear. Computing the MSD can give important information on how the random walk is influenced by different factors, both related to the random walk parameters and external (e.g., collisions, boundaries, interaction with other robots).

42.2.1 Random exploration with a motion bias

Robots usually do not need to explore indefinitely, and often they need to return to a base station or generically a central place where to recharge their batteries or share information about what they have found. A biased random walk guarantees that the movements of the robots are performed in a way that takes into account the preferential direction of motion. To implement a biased random walk, instead of choosing a random turning angle, the robot can choose with a given probability P_b to rotate towards the preferential direction of motion (e.g., the direction of the base station, if available).

A similar approach can be beneficial when robots move in environments where motion may be influenced by some external factors, such as wind for flying robots or water currents for aquatic surface or underwater robots. If the robot is aware of the drag (known currents or perceived force), it can attempt to counteract it by means of a biased random walk. In this case, the bias must not be fixed (e.g., an absolute direction of motion), but can change dynamically over time as the contingencies change. A dynamic bias in the random walk can provide interesting diffusion properties to the robots. A particularly interesting case for robot swarms is when this dynamic directional bias is computed as a function of the interactions with neighboring robots. This can lead to collective exploration strategies that can take advantage of the information retained by the group. One straightforward example is when the motion bias leads the random walker away from its neighbors: when a biased rotation must be made, the angle β is computed with reference to the known neighbors (e.g., by making a weighted sum of repulsion vectors computed after the distance and bearing of other robots within a given range). This can maximize the ability of the swarm to cover large areas while still maintaining a random diffusion at the individual level, especially when the robot has no immediate neighbor. Similar strategies are sometimes referred to as *reinforced random walks* [13, 14].

Example 42.1: Implement a swarm of random walkers. Given the description of a random walk as a sequence of straight movements and random turns, it is possible to develop the algorithm that a robot should follow. Consider that, since robots have a finite velocity, they can cover at most a distance δ in one control step. Similarly, real robots have a finite angular velocity, so that the maximum rotation for each control step is of an angle ω , either clockwise or counterclockwise.

a. Provide a pseudocode to implement the LMCRW controlled by the parameters θ , ρ , and α . Use two variables to store the random displacement t_s and the random rotation t_r , assuming you have functions to obtain correctly distributed values (e.g., `random.levy` and `random.wcd`). Note that the displacement must be non-negative: $t_s \geq 0$. Then, use t_s and t_r as counters, reducing them by δ and ω respectively, until they reach 0. Pay attention to the sign of ω which indicates the rotation direction. When both rotation and straight motion have been performed, reset the counters with new random variables. The following algorithm proposes a suitable implementation.

Algorithm 2 Base control loop for a robot to perform a random walk

```

Require:  $t_s, t_r, \omega, \delta$                                 ▷ Robot state variables
Require: Rotate(), MoveStraight()                      ▷ Robot motion functions

1: if  $t_s = 0$  then
2:    $t_s \leftarrow |\text{random.levy}(\alpha)|$                 ▷ draw a positive random number from a Lévy distribution
3:    $t_r \leftarrow \text{random.wcd}(\theta, \rho)$             ▷ draw a random number from a wrapped Cauchy distribution
4: end if
5: if  $t_r \neq 0$  then
6:   if  $|t_r| > \omega$  then                               ▷  $\omega$  is the maximum rotation angle performed in one control step.
7:     Rotate(sign( $t_r$ ) $\omega$ )
8:      $t_r \leftarrow t_r - \text{sign}(t_r)\omega$ 
9:   else
10:    Rotate( $t_r$ )
11:     $t_r \leftarrow 0$ 
12:  end if
13: else
14:   if  $t_s > \delta$  then                                ▷  $\delta$  is the maximum distance covered in one control step.
15:     MoveStraight( $\delta$ )
16:      $t_s \leftarrow t_s - \delta$ 
17:   else
18:     MoveStraight( $t_s$ )
19:      $t_s \leftarrow 0$ 
20:   end if
21: end if

```

b. Modify the previous algorithm to implement a biased choice of the turning angle, so that the robot moves in a known preferential direction β . The choice of the preferential direction must be performed with a fixed probability P_b every time a new turning angle must be chosen. The following algorithm gives an idea of the modifications to implement.

Algorithm 3 Application of a preferential direction of motion (β) in the choice of the turning angle (see step 3 in Algorithm 2)

Require: β

```

1: if random.uniform(0, 1) <  $P_b$  then          ▷ Choose the preferential direction with probability  $P_b$ 
2:    $t_r \leftarrow \beta$ 
3: else
4:    $t_r \leftarrow \text{random.wcd}(\theta, \rho)$          ▷ draw a random number from a wrapped Cauchy distribution
5: end if

```

42.2.2 Collective exploration strategies

As discussed in the previous section, robots in a swarm can exploit interactions to explore more efficiently. Going beyond the purely random walk approach, robots can explore by creating navigable structures that support both the exploration task and the exploitation of the information available to other robots, for instance to quickly reach the area where something relevant has been found. The simplest such structure is a hexagonal lattice, which can be easily formed by a robot swarm if robots are instructed to maintain a fixed distance from all their neighbors. In other words, two simple rules are sufficient: (i) if the minimum distance from a neighbor exceeds an *aggregation threshold*, the robot moves in the direction of the closest neighbors; (ii) if the minimum distance from all neighbors is smaller than a *repulsion threshold*, the robots moves away from the closest neighbors. It turns out that, if the thresholds are correctly tuned, these simple rules lead to a maximum packing of the swarm. With a similar approach, several different lattice formations can be obtained, with robots only reacting to the presence of neighbors [15]. Once a structure is formed, a reference frame can be autonomously generated by the swarm by assigning IDs to robots as a function of the distance and direction from a relevant point in space (e.g., a base station) [16].

A disadvantage of forming lattice structures is that they are static and do not extend much the area reachable by the swarm. An interesting alternative consists in the formation of chains of robots that maximally extend from a base station [17, 10]. This approach is somewhat inspired by the formation of pheromone trails in ants [18], with the important difference that robots themselves act as pheromones by marking areas that extends away from the base station. The rules for chain formation are easy (Fig. 42.3a), but require that the direction away from the base station can be perceived and/or communicated to neighbors to support navigation along the chain. If this is possible, then robots need just to stop at the farthest place from an already formed chain. When no chain is available or visible, a robot initiates a new chain close to the base station by stopping with a

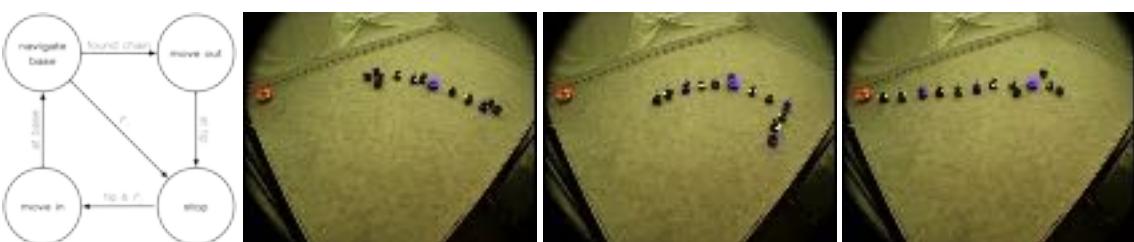


Figure 42.3: **Collective exploration by means of chain structures that extend from a base station.** (a) Finite state machine describing the high-level behavior performed by robots to build and exploit chains. (b)-(d) Three different phases of a chain formation process, which extends from the base station in multiple directions (blue object) and ends up with the robots finding the target (red object). Reproduced from Ref. [10]. [Agnese: add letters to the panels](#)

certain probability P_s . Stopped robots are static, and serve as a beacon to other searching robots. To make the chain formation process dynamic, formed chains should disband with some probability. To this end, a stopped robot at the tip of a chain has a small probability P_L to detach and start searching again. Note that, to maintain a connected structure, robots must always remain in touch with at least one stopped robot or the base station. Figs. 42.3b-d shows different phases of a chain formation process, which ends with the robots finding a target object [10].

42.3 Self-organized aggregation

One of the basic building blocks for swarm robotics behavior is, without doubts, self-organized aggregation [19, 20, 21]. Aggregation is important because it is the pre-requisite of other forms of collaboration, as it allows robots to gather together and interact physically or through communication (e.g., to choose a common direction of motion or to perform a collective decision). The aggregation process can exploit detectable heterogeneities in the environment, resulting in a stimulus that triggers some form of taxis in the robot. This is what happens to many insects that aggregate around light sources (phototaxis), or to bacteria that aggregate following the sugar concentration (chemotaxis). For robot swarms, environmental heterogeneities may not be present, may not provide sufficient stimulus to support aggregation, or may not be relevant for the tasks that the robots have to execute. To remove the dependency from environmental heterogeneities, robots in a swarm can rely on self-organized aggregation mechanisms, whereby the formation of tight groups is performed solely as a result of the interactions among robots. The key to understand self-organized aggregation is noticing that robots themselves should be programmed to produce the heterogeneities that seed the aggregation process. In other words, robots should produce a signal that can be exploited by other robots to aggregate. However, the problem remains on how to design the robot behavior and robot–robot interactions to obtain a reliable aggregation.

Several natural systems are known that perform self-organized aggregation, with different modalities for movement and communication. One very interesting example is the aggregation behavior in the *Dictyostelium discoideum*, whereby a multitude of unicellular amoebae forms a multicellular slug [22]. When the amoebae are starving, they start producing a cyclic adenosine monophosphate (AMP) that diffuses and attracts individuals towards areas of high concentration. As initial aggregates start to form with nearby cells sticking together, the intensity of the chemical signal they produce increases. This leads to a stronger gradient of the chemical signal and, in turns, a stronger attraction for neighboring cells, which further reinforce the gradient. This positive feedback loop ultimately leads to the formation of a large multicellular body (Fig. 42.4a). A similar behavior can be implemented by robots, provided that (i) robots can emit a signal that they are able to track and follow, and (ii) multiple signals emitted by nearby robots add up to create a more intense signal. Figs. 42.4b-d shows different phases of a behavior implemented with robots able to emit a single-frequency sound signal, and able to move in the direction of high sound intensity [19].

An alternative approach to self-organized aggregation that does not make use of signals exploits the ability to detect the presence of neighbors. This approach is inspired by the behavior of cockroaches (*Blattella germanica*) which naturally tend to form aggregates under a shelter [23]. Here, the heterogeneity consists in the presence of other insects, which are recognized by means of direct contact and/or chemical sensing. For

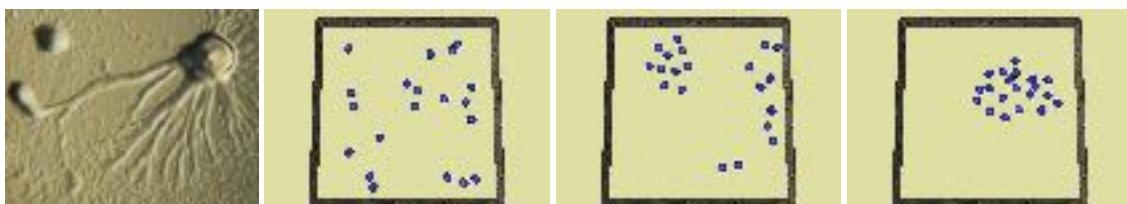


Figure 42.4: Self-organized aggregation. (a) Aggregation in *Dictyostelium discoideum*, where starving amoebae produce cyclic adenosine monophosphate (AMP) that diffuses and attracts others towards areas of high concentration, leading to the formation of a multicellular slug. (b)-(d) Different phases of aggregation for robots emitting single-frequency sound signals and moving towards high sound intensity. Reproduced from Ref. [19]. [Agnese: add letters to the panels](#)

in the swarm robotics implementation, we assume each robot can count the number of neighbors within a limited range (e.g., from visual inspection or short-range communication). The aggregation behavior exploits a non-linear probability of joining or leaving a formed aggregate as a function of the aggregate dimension. If an aggregate is small, the probability to join it is also small (also because it is not easy to find). On the other hand, a large aggregate has much higher chances of being found and also joined. The latter probability can be summarized by the following equation:

$$P_J = \varepsilon + (1 - \varepsilon)(1 - e^{-\gamma n}), \quad (42.2)$$

where n is the number of stopped robots in the neighborhood, γ is a tunable parameter to determine the sensitivity to the aggregate size, and $\varepsilon \in [0, 1]$ is a small probability of stopping spontaneously, forming an aggregate of size one. Conversely, the probability of leaving a small aggregate can be made monotonically decreasing as a function of the aggregate size. This can be computed as follows:

$$P_L = e^{-\lambda n}, \quad (42.3)$$

where λ determines how quickly the probability decays with the number of nearby stopped robots n . Overall, the behavior of a robot can be represented by a simple probabilistic finite state machine, depicted in Fig. 42.5a. Two states are necessary:

Search The robot performs a random walk in search of already-formed aggregates. To this end, it counts the number of stopped neighbors and decides whether or not to stop with probability P_J .

Stop The robot stops and signals its presence to neighbors (if necessary for counting). It continuously checks the number of stopped neighbors, and changes back to state Search with probability P_L .

Following these simple rules, the aggregation process proceeds by formation of multiple small clusters first, some of which quickly disband in favor of others that instead grow in size. The larger the aggregate, the higher the probability that it will grow at the expense of smaller ones. With high probability, a single aggregate remains that embraces most of the robots (Figs. 42.5b-d).

Example 42.2: Implement a self-organized aggregation behavior. Aggregation based on counting the number of neighbors is very easy to implement. Assuming that the count of stopped neighbors is available (through a function called `CountStoppedneighbors()`), determine the rules to implement the probabilistic finite state machine from Fig. 42.5a, which alternates between stopping close to an aggregate, or randomly searching for one (through a function called `RandomWalk()`). Use a state variable `S` to store the current behavior: Stop or Search. Compute the probability of joining and leaving an aggregate according to Eqs. (42.2) and (42.3), respectively, with λ , ε , γ being control parameters for the algorithm. The following algorithm proposes a suitable implementation.

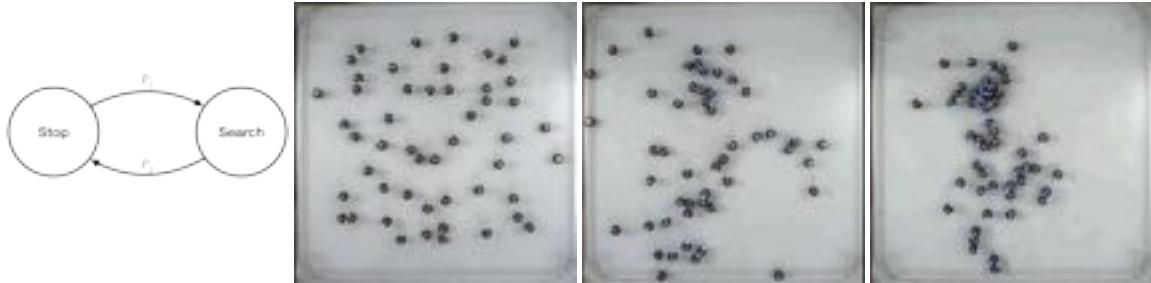


Figure 42.5: **Simple finite state machine for self-organized aggregation.** (a) Simple probabilistic finite state machine for self-organized aggregation. (b)-(c) Different phases of a self-organized aggregation process exploiting the count of neighbors. Reproduced from Ref. [21]. [Agnese: add letters to the panels](#)

Algorithm 4 Base control loop for a robot to perform self-organized aggregation

Require: CountStoppedneighbors(), RandomWalk(), SignalStoppedState()

Require: $\lambda, \varepsilon, \gamma$

```

1: if  $S = \text{Stop}$  then
2:   SignalStoppedState()                                 $\triangleright$  useful to distinguish between moving and stopped robots
3:    $n \leftarrow \text{CountStoppedneighbors}()$ 
4:    $P_L \leftarrow e^{-\lambda n}$ 
5:   if random.uniform(0, 1) <  $P_L$  then
6:      $S \leftarrow \text{Search}$ 
7:   end if
8: else
9:   RandomWalk()                                      $\triangleright$  Search for already-formed aggregates
10:   $n \leftarrow \text{CountStoppedneighbors}()$ 
11:   $P_J \leftarrow \varepsilon + (1 - \varepsilon)(1 - e^{-\gamma n})$ 
12:  if random.uniform(0, 1) <  $P_J$  then
13:     $S \leftarrow \text{Stop}$ 
14:  end if
15: end if

```

42.4 Coordinated motion

Once robots are close to each other, further coordination is necessary for the robots to move together. The main problems to be faced are (i) choosing a common direction of motion, and (ii) proceed in a tight formation while adapting to environmental constraints (e.g., narrow passages). These problems are worsened by the fact that different robots in the swarm may have different, partial information about the goal, and may also receive radically different sensory information about the surrounding environment. For instance, a robot inside the aggregate will have its sensors mostly occluded by its neighbors, while a robot at the periphery can obtain a much better perception of the surrounding environment. It is therefore necessary to deal with highly dynamic availability of information when deciding how robots should move together.

Coordinating the movement of a massive amount of robots is a difficult problem even for a centralized approach. Much work has been devoted to formation movement, where robots must maintain a specific spatial relation with their neighbors [24, 25]. However, maintaining the formation against the presence of obstacles and narrow passages is difficult and sometimes impossible, requiring solutions that can adapt to the environmental contingencies. A different approach takes inspiration from the coordinated movement abilities of birds and fish. Flocks of starlings display mesmerizing movements called murmurations, and have been thoroughly studied thanks to 3D reconstruction and movement tracking [26, 27]. Also fish present interesting collective movements with a variety of collective patterns, from schooling to milling [28, 29]. Interestingly, the individual rules behind these collective movement patterns are strikingly simple [30, 31], and can be exploited for robot swarms [8, 32].

The standard collective movement model requires just three rules: separation, cohesion, and alignment. *Separation* is meant to maintain a safe distance between robots, to avoid collisions. *Cohesion* is meant to maintain contact with other robots in the group, to avoid splitting. *Alignment* is meant to provide robots with a common direction of motion. A simple implementation applies these three rules according to the distance of the closest neighbor. Priority is given to separation, to avoid any collision: in this case, the robot moves away from all neighbors within a safety range D_s , computing a repulsion vector \mathbf{v}_s on the basis of the relative position of the neighbors. If instead the closest neighbor is farther than a maximum cohesion threshold D_c , the robot computes an attraction vector \mathbf{v}_c to remain in contact with all available neighbors. Finally, when neither separation or cohesion are necessary, the robot aligns its motion direction with the one of the neighbors found at distance between D_s and D_c . To this end, the velocity vector of neighbors must be estimated, and the average velocity of the neighbors gives the new direction of motion for the robot. Eventually, all robots end up aligned and moving at constant speed (Fig. 42.6).

While these simple rules are easy to implement, they require knowledge of the position and velocity of all

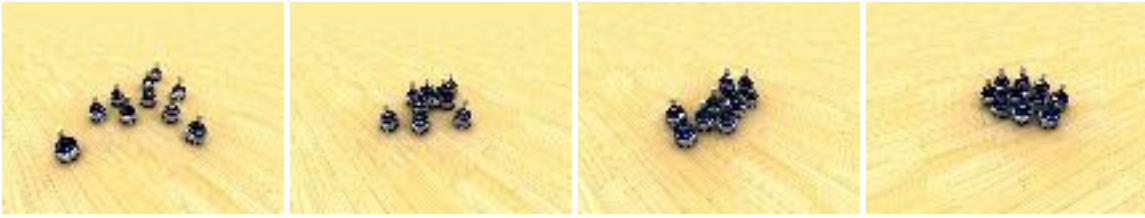


Figure 42.6: **Different phases of coordinated motion with a group of robots.** The robots signal their orientation with red and blue colors on the left and right sides, respectively. This information is sufficient to estimate the direction of motion and leads to coordinated motion. Reproduced from Ref. [34].

neighbors within a given range. Such information is not easy to obtain, especially for the velocity component, which requires the recognition of the relative heading of the robots and possibly also the actual speed. In this cases, methods have been developed that do away with the alignment rule [33]. If however robots have access to their absolute position and velocity (e.g., via some external positioning system like a GPS and an inertial unit), these can be continuously communicated to neighbors, enabling the implementation of the coordinated motion rule (although communication delays can lead to instabilities) [8].

Example 42.3: Implement a flocking behavior for agents in 2D. Coordinated movements can be easily implemented by following the rules of separation, cohesion and alignment. Assume that functions are available to obtain information about neighbors (relative position \mathbf{d} and velocity \mathbf{v}). In particular, the function `GetClosestneighbor()` provides information about the closest robot to implement the separation and cohesion rules. The function `GetAllneighbors(d, D)` returns all neighbors within a distance range $[d, D]$, which can be used to compute the desired motion vector according to the rule to be executed, and then implement the movement (function `Move()`). The following algorithm proposes a suitable implementation.

Algorithm 5 Base control loop for a robot to perform coordinated movements

Require: `GetClosestneighbor()`, `GetAllneighbors()`, `Move()`

```

1:  $R \leftarrow \text{GetClosestneighbor}(D_M)$ 
2: if  $|\mathbf{d}_R| < D_s$  then ▷ Separation rule: move away from neighbors
3:    $\mathcal{R} \leftarrow \text{GetAllneighbors}(0, D_s)$ 
4:    $\mathbf{v}_s \leftarrow \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \mathbf{d}_r$ 
5:    $\text{Move}(\mathbf{v}_s)$ 
6: else if  $|\mathbf{d}_R| > D_c$  then ▷ Cohesion rule: move towards neighbors
7:    $\mathcal{R} \leftarrow \text{GetAllneighbors}(D_c, D_M)$ 
8:    $\mathbf{v}_c \leftarrow \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} -\mathbf{d}_r$ 
9:    $\text{Move}(\mathbf{v}_c)$ 
10: else ▷ Alignment rule: move in the direction of neighbors
11:    $\mathcal{R} \leftarrow \text{GetAllneighbors}(D_s, D_c)$ 
12:    $\mathbf{v}_a \leftarrow \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \mathbf{v}_r$ 
13:    $\text{Move}\left(\frac{\mathbf{v}_a}{|\mathbf{v}_a|}\right)$ 
14: end if

```

42.5 Conclusions

In our discussion of how fundamental behaviors for robot swarms can be obtained by means of very simple individual rules and little interaction among the robots, we have considered exploration, aggregation, and coordinated motion as basic building blocks, along with some possible implementation variants (see also the problems below to further deepen the understanding of these behaviors). Clearly, these basic building blocks

are not the only possible ones [35], and may not be sufficient to implement more complex self-organizing behaviors in robot swarms. Other relevant fundamental abilities studied in robot swarms are collective decision making [36, 7], task-allocation [37, 38], morphogenesis, and self-assembly [39, 16, 40]. The latter in particular is useful to create physical structures with properties that depend on the arrangement of the robots, which should not be pre-determined but functional to the task to be accomplished.

Nevertheless, having a good set of building blocks does not solve the problem of designing swarm robotics systems to accomplish any given task. In fact, it is necessary to have also reliable methods to assemble different building blocks, which is not straightforward. Negative interactions between building blocks may jeopardize their effectiveness or efficiency, strongly departing from the expected properties of the single building block when studied in isolation. Possible solutions may come from the exploitation of design patterns, that is, reusable solutions to tackle commonly occurring problems, which can lead to good results in many cases [41, 36]. Otherwise, one can rely on automatic design techniques to optimize or learn the collective behavior of robot swarms, either offline exploiting simulations or online during task execution [42, 43]. Evolutionary algorithms have been successfully applied to a variety of different tasks involving self-organization in robot swarms, whereby the behavior of the individual robot was encoded by a simple neural network [34]. Similar optimization strategies have been developed for the automatic design of probabilistic finite state machines, exploiting a set of basic behaviors for robots in a swarm [44]. Reinforcement learning is another option that is recently gaining traction thanks to the advancements in machine learning research [45]. Social learning and cultural evolution are also particularly fit for swarm robotics research, and have received much attention in the past years [43, 21].

All the mentioned research studies have strongly contributed to the maturity of swarm robotics as a research field, to the point that we are very close to the deployment of robot swarms in real-world applications [6]. There are already concepts of small robot swarms (both ground and aerial ones) to be deployed in precision agriculture [46, 14]. Logistics, environmental monitoring, surveillance are all tasks that can benefit from collaboration and self-organization of multiple robots. Even space agencies are thinking to use satellite and robot swarms for efficient coverage and planetary exploration [47, 48]. Hence, we can really hope that, in future missions, no planetary rover will ever be left behind.

42.6 Problems

Problem 42.1: Mean square displacement in different realistic conditions. Implement a swarm of random walkers (Algorithm 2) and store the position of the agents in order to compute and visualize the MSD. [*Hint: NetLogo provides a simple multi-agent simulation environment, for this and the following problems.*]

- a. Experiment with different values of the parameters α and ρ to determine how the MSD varies across different conditions.
- b. Evaluate the impact of the robot speed. Differently from an ideal random walkers, robots take time to move and to rotate to a new direction. Vary the robot linear and rotation speed and evaluate the variation of the MSD, both in the case of normal and anomalous diffusion.
- c. Evaluate the impact of the robot embodiment. Collisions with walls, obstacles, and other robots may alter the expected diffusion properties of the random walk. If the robot cannot detect the obstacle, it would keep colliding with the obstacle unless a random rotation leads it away. What is the effect on the MSD? Test it by enclosing the robots within square or circular walls, and by inserting obstacles of different size.
- d. When a robot can detect a collision or the obstacles, it can either perform an avoidance maneuver, or stop the straight motion and draw a random angle away from the obstacles. What is the impact of the diffusion in the latter case? What happens when collisions are possible among robots? Compute the MSD for varying topologies and densities of robots.
- e. Implement an alternative way of introducing a bias in a random walk. Instead of alternating random and biased rotation, apply the bias at every rotation. The turning angle can be obtained by means of a weighted sum between a unit vector in the chosen random direction \mathbf{v}_r , and a unit vector in the direction of the bias \mathbf{v}_β :

$$\mathbf{v} = \mathbf{v}_r + b\mathbf{v}_\beta, \quad (42.4)$$

where b is a weight to give more or less importance to the bias. Implement both the standard and the alternative versions for robots biased to return to a central place, and identify for which values of P_b and b the MSD properties are similar.

Problem 42.2: Variants of self-organized aggregation. Starting from the two presented approaches to self-organized aggregation (one based on the perception of signals emitted by robots, and the other based on counting the stopped neighbors), it is possible to develop alternative solutions, which may also lead to qualitative differences in the resulting aggregate and its properties.

a. Implement aggregation by means of a biased random walk. As discussed above, it is possible to design a random walk behavior that is influenced by the presence of neighboring robots. A directional bias can be computed by taking into account the position of nearby robots, with the goal of creating a robust aggregate. Which parameters of such a biased random walk work better for the creation of a single aggregate? Measure the performance of your behavior by computing the size of the largest aggregate, where two robots are in the same aggregate if their distance is smaller than a given threshold.

b. Aggregation as a decision making process. The aggregation algorithms can be modified to let robots aggregate only in certain areas (e.g., dark areas). When two or more of these areas are present, an aggregate may form in any one, but the collective dynamics can lead the swarm to choose only one of the available areas where to aggregate, as a result of a collective decision. Modify Algorithm 4 to let robots only stop when above specific areas, and study how to tune the control parameters to obtain a reliable collective decision.

c. Design different shapes of the resulting aggregate. The aggregation algorithms discussed above are isotropic, as the robot cluster does not feature any preferential growth direction. However, it is easy to change this property by designing some sort of *preferential attachment* mechanism, which can be embedded in the way in which robots sense the presence of others. In Algorithm 4, change the way in which robots signal the stopping state (e.g., emit a directional signal perceivable only in a given direction), and/or the way in which robots perceive and count neighbors (e.g., limit the angular sector within which signals can be received). What are the effects on the aggregate shape? What are the conditions for the formation of chains?

Problem 42.3: Coordinated motion with informed individuals. When moving together, robots may have some (partial, noisy) knowledge of the goal destination, or the direction to take. To integrate such information, a trade-off must be found between staying with the group and moving in the desired direction. Explore this trade-off, and also deal with the case in which different robots may have different preferred destinations.

a. Implement coordinated motion by removing the partitioning in different zones. Consider the neighbors within a given distance D_M . From the knowledge of their position and velocity vectors compute a cohesion/separation vector (e.g., modulating the distance according to a Lennard–Jones potential) and an orientation vector [33], and average over the available neighbors. Change the way in which neighbors are selected (e.g., vary the distance D_M or choose the first N neighbors, no matter the distance). How does the collective motion behavior change?

b. Include a preferential direction of motion as an additional vector used to compute the robot motion direction. Weigh to the preferential direction with a parameter p , and experiment with different values of the parameter to understand how fast the swarm reorients towards the preferential direction. Then, assign the preferential direction to only a fraction of the swarm, and identify, for different values of p , which is the minimum fraction that leads the whole swarm to move to the preferred destination.

c. Experience with contrasting individual information, by dividing the swarm in three groups: one group of robots with preferential direction \mathbf{V}_1 , another group of robots with preferential direction \mathbf{V}_2 , and a group without preference (referred to as uninformed robots). Vary the fraction of informed/uninformed robots and the angular distance between \mathbf{V}_1 and \mathbf{V}_2 . When does the swarm chooses to follow either direction, and when instead the swarm averages the two moving in the middle?

Bibliography

- [1] F. Mondada, G. C. Pettinari, A. Guignard, I. W. Kwee, D. Floreano, J.-L. Deneubourg, S. Nolfi, L. M. Gambardella, and M. Dorigo. Swarm-bot: A new distributed robotic concept. *Autonomous Robots*, 17(2):193 – 221, 2004.
- [2] N. Jacobstein. NASA’s Perseverance: Robot laboratory on Mars. *Science Robotics*, 6(52):eabh3167, 2021.
- [3] M. Prorok, A. and Malencia, L. Carbone, G. S. Sukhatme, B. M. Sadler, and V. Kumar. Beyond robustness: A taxonomy of approaches towards resilient multi-robot systems. *arXiv*, 2109.12343, 2021.

- [4] Y. Rizk, M. Awad, and E. W. Tunstel. Cooperative heterogeneous multi-robot systems: A survey. *ACM Computing Surveys*, 52(2), 2019.
- [5] M. Dorigo, G. Theraulaz, and V. Trianni. Swarm robotics: Past, present, and future. *Proceedings of the IEEE*, 109(7):1152–1165, 2021.
- [6] M. Dorigo, G. Theraulaz, and V. Trianni. Reflections on the future of swarm robotics. *Science Robotics*, 5(49):eabe4385, 2020.
- [7] G. Valentini, E. Ferrante, and M. Dorigo. The best-of-n problem in robot swarms: Formalization, state of the art, and novel perspectives. *Frontiers in Robotics and AI*, 4:1 – 43, 2017.
- [8] G. Vásárhelyi, C. Virág, G. Somorjai, T. Nepusz, A. E. Eiben, and T. Vicsek. Optimized flocking of autonomous drones in confined environments. *Science Robotics*, 3(20):eaat3536, 2018.
- [9] K. N. McGuire, C. De Wagter, K. Tuyls, H. J. Kappen, and G. C. H. E. de Croon. Minimal navigation solution for a swarm of tiny flying robots to explore an unknown environment. *Science Robotics*, 4(35):eaaw9710, 2019.
- [10] S. Nouyan, R. Gross, M. Bonani, F. Mondada, and M. Dorigo. Teamwork in self-organized robot colonies. *IEEE Transactions on Evolutionary Computation*, 13(4):695 – 711, 2009.
- [11] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. The MIT Press, 2005.
- [12] C. Dimidov, G. Oriolo, and V. Trianni. Random walks in swarm robotics: an experiment with kilobots. In *Proceedings of the 10th International Conference on Swarm Intelligence (ANTS 2016)*, volume 9882 of *Lecture Notes in Computer Sciences*, pages 185–196. Springer Verlag, Berlin, Germany, 2016.
- [13] D. Albani, D. Nardi, and V. Trianni. Field coverage and weed mapping by UAV swarms. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017)*, pages 1–7, July 2017.
- [14] D. Albani, T. Manoni, A. Arik, D. Nardi, and V. Trianni. Field coverage for weed mapping: toward experiments with a uav swarm. In *Proceedings of the 11th EAI International Conference on Bio-inspired Information and Communications Technologies (BICT 2019)*, pages 1–16, Pittsburg, US, 2019. EAI.
- [15] W. M. Spears, D. F. Spears, J. C. Hamann, and R. Heil. Distributed, physics-based control of swarms of vehicles. *Autonomous Robots*, 17(2):137 – 162, 2004.
- [16] M. Rubenstein, A. Cornejo, and R. Nagpal. Programmable self-assembly in a thousand-robot swarm. *Science*, 345(6198):795 – 799, 2014.
- [17] S. Nouyan, A. Campo, and M. Dorigo. Path formation in a robot swarm. *Swarm Intelligence*, 2(1):1 – 23, 2008.
- [18] C. Detrain and Jean-Louis Deneubourg. Self-organized structures in a superorganism: do ants “behave” like molecules? *Physics of Life Reviews*, 3(3):162 – 187, 2006.
- [19] M. Dorigo, V. Trianni, E. Şahin, R Groß, T. H. Labella, G. Baldassarre, S. Nolfi, J.-L. Deneubourg, F. Mondada, D. Floreano, and L. M. Gambardella. Evolving self-organizing behaviors for a swarm-bot. *Autonomous Robots*, 17(2–3):223–245, 2004.
- [20] M. Gauci, J. Chen, W. Li, T. J. Dodd, and R. Gross. Self-organized aggregation without computation. *The International Journal of Robotics Research*, 33(8):1145 – 1161, 2014.
- [21] N. Cambier, D. Albani, V. Frémont, V. Trianni, and E. Ferrante. Cultural evolution of probabilistic aggregation in synthetic swarms. *Applied Soft Computing*, 113:108010, 2021.
- [22] B. N. Vasiev, P. Hogeweg, and A. V. Panfilov. Simulation of *dictyostelium discoideum* aggregation via reaction-diffusion model. *Physical Review Letters*, 73:3173–3176, 1994.

- [23] R. Jeanson, C. Rivault, J.-L. Deneubourg, S. Blanco, R. Fournier, C. Jost, and G. Theraulaz. Self-organized aggregation in cockroaches. *Animal Behaviour*, 69(1):169 – 180, 2005.
- [24] T. Balch and R. C. Arkin. Behavior-based formation control for multirobot teams. *IEEE Transactions on Robotics and Automation*, 14(6):926 – 939, 1998.
- [25] J Fredslund and M.J. Mataric. A general algorithm for robot formations using local sensing and minimal communication. *IEEE Transactions on Robotics and Automation*, 18(5):837 – 846, 2002.
- [26] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences*, 105(4):1232 – 1237, 2008.
- [27] H. Hildenbrandt, C. Carere, and C.K. Hemelrijk. Self-organized aerial displays of thousands of starlings: a model. *Behavioral Ecology*, 21(6):1349–1359, 2010.
- [28] D. S. Calovi, U. Lopez, S. Ngo, C. Sire, H. Chaté, and G. Theraulaz. Swarming, schooling, milling: phase diagram of a data-driven fish school model. *New Journal of Physics*, 16(1):015026, 2014.
- [29] K. Tunstrøm, Y. Katz, C. C. Ioannou, C. Huepe, M. J. Lutz, and I. D. Couzin. Collective States, Multi-stability and Transitional Behavior in Schooling Fish. *PLoS Computational Biology*, 9(2):e1002915 EP –, 2013.
- [30] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *Computer graphics*, 21(4):25 – 34, 1987.
- [31] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517(3–4):71 – 140, 2012.
- [32] A. E. Turgut, H. Çelikkanat, F. Gökc , and E.  ahin. Self-organized flocking in mobile robot swarms. *Swarm Intelligence*, 2(2-4):97 – 120, 2008.
- [33] E. Ferrante, A. E. Turgut, C. Huepe, A. Stranieri, C. Pinciroli, and M. Dorigo. Self-organized flocking with a mobile robot swarm: a novel motion control method. *Adaptive Behavior*, 20(6):460–477, 2012.
- [34] V. Trianni and M. L pez-Ib  ez. Advantages of task-specific multi-objective optimisation in evolutionary robotics. *PLoS ONE*, 10(8):e0136406 – 27, 2015.
- [35] V. Trianni and A. Campo. Fundamental collective behaviors in swarm robotics. In J. Kacprzyk and W. Pedrycz, editors, *Springer Handbook of Computational Intelligence*, pages 1377 – 1394. Springer Berlin Heidelberg, 2015.
- [36] A. Reina, G. Valentini, C. Fern  ndez-Oto, M. Dorigo, and V. Trianni. A design pattern for decentralised decision making. *PLoS ONE*, 10(10):e0140950–18, 2015.
- [37] M. J. B. Krieger, J.-B. Billeter, and L. Keller. Ant-like task allocation and recruitment in cooperative robots. *Nature*, 406(6799):992 – 995, 2000.
- [38] S. Berman, A. Halasz, M. A. Hsieh, and V. Kumar. Optimized stochastic policies for task allocation in swarms of robots. *IEEE Transactions on Robotics*, 25(4):927 – 937, 2009.
- [39] R. Gross and M. Dorigo. Self-Assembly at the Macroscopic Scale. *Proceedings of the IEEE*, 96(9):1490 – 1508, 2008.
- [40] I. Slavkov, D. Carrillo-Zapata, N. Carranza, X. Diego, F. Jansson, J. Kaandorp, S. Hauert, and J. Sharpe. Morphogenesis in robot swarms. *Science Robotics*, 3(25):eaau9178, 2018.
- [41] O. Babaoglu, G. Canright, A. Deutsch, G. A. Di Caro, F. Ducatelle, L. M. Gambardella, N. Ganguly, M. Jelasity, R. Montemanni, A. Montresor, and T. Urnes. Design patterns from biology for distributed computing. *ACM Transactions on Autonomous Adaptive Systems*, 1(1):26 – 66, 2006.

- [42] G. Francesca and M. Birattari. Automatic design of robot swarms: Achievements and challenges. *Frontiers in Robotics and AI*, 3:224 – 9, 2016.
- [43] N. Bredeche and N. Fontbonne. Social learning in swarm robotics. *Philosophical Transactions of the Royal Society B*, 377(1843):20200309, 2022.
- [44] G. Francesca, M. Brambilla, A. Brutschy, V. Trianni, and M. Birattari. AutoMoDe: A novel approach to the automatic design of control software for robot swarms. *Swarm Intelligence*, 8(2):89–112, 2014.
- [45] M. Hüttenrauch, A. Šošić, and G. Neumann. Deep reinforcement learning for swarm systems. *Journal of Machine Learning Research*, 20(54):1—31, 2019.
- [46] T. Blender, T. Buchner, B. Fernandez, B. Pichlmaier, and C. Schlegel. Managing a mobile agricultural robot swarm for a seeding task. In *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 6879–6886, 2016.
- [47] L. A. Nguyen, T. L. Harman, and C. Fairchild. Swarmathon: A swarm robotics experiment for future space exploration. In *2019 IEEE International Symposium on Measurement and Control in Robotics (ISMCR)*, pages B1–3–1–B1–3–4, 2019.
- [48] D. St-Onge, M. Kaufmann, J. Panerati, B. Ramtoula, Y. Cao, E. B. J. Coffey, and G. Beltrame. Planetary exploration with robot teams: Implementing higher autonomy with swarm intelligence. *IEEE Robotics & Automation Magazine*, 27(2):159–168, 2020.

Part IV

Game

Chapter 43

Active Matter Game

LAURA NATALI, JESUS DOMINGUEZ, DAVIDE BREONI, SANDRINE HEIJNEN, DAVID BRONTE CIRIZA, CAROLINA VAN BAALEN, ALIREZA KHOSHZABAN, AGNESE CALLEGARI

This chapter presents the intriguing field of active matter in a playful way through a board game, making it relevant not only to scientists but also to a wide audience. The previous Chapters introduced several aspects about active entities, which can convert energy into self-propulsion. Hopefully, this book has inspired you — and maybe even challenged your way of thinking. However, after trying to absorb all this new information, you might find yourself in a state of cognitive fatigue. You may also want to introduce the field of active matter to students, friends, or family, but find yourself wondering where to start. In both cases, it is useful to have

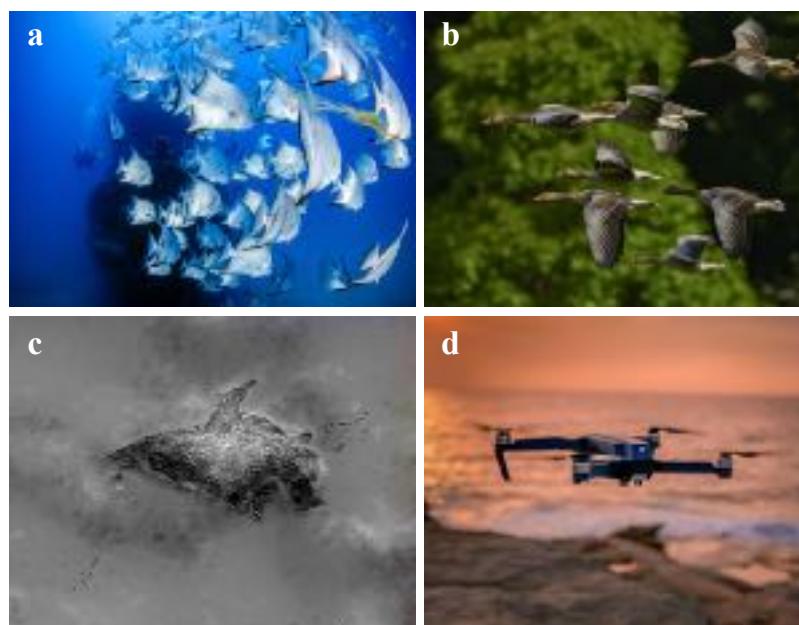


Figure 43.1: **Examples of biological and artificial active matter at different scales.** (a) A school of fish can more often evade predators by moving together than an individual fish can (photo [from Pexels](#)). (b) The V formation in flocks of birds improves flight efficiency, especially over long distances (photo [Pexels](#)). (c) Active matter appears also on the microscopic scale, such as swarming bacteria exploring their environment (image by J. M. Antúnez Domínguez). (d) An example of robot that embeds some of the key features of active systems (photo [from Pexels](#)). All photos on Pexels can be downloaded and used for free ([see Pexels license](#)).

a playful and intuitive way of introducing the topic, and therefore we turned active matter into a board game, presented in this Chapter.

43.1 Background

double-check all references to chapters, both number and title

Active matter encompasses a wide range of self-propelled entities that exhibit complex patterns when interacting with others, a phenomenon known as collective behavior. It includes systems such as swarms of insects, flocks of birds, schools of fish, and groups of synthetic micro- and nanoscale particles. In the *Active Matter Game*, the players' meeples represent bacteria, an important example of biological active matter. Additionally, several physical phenomena are illustrated through the rules of the game.

Active entities have the ability to respond to stimuli through movement, a phenomenon known as *taxis* (see Chapter 27, Chemistry of Micromotors). In simple terms, there are two possible responses: *positive taxis* aiming to approach the source of the stimulus (e.g., animals searching for food or water) and *negative taxis* (e.g., animals escaping from predators). Taxis is at the origin of more complex behaviors (see Chapter 10, Motility-Induced Phase Separation): detecting and escaping predators (Fig. 43.1a), improving flight conditions (Fig. 43.1b), or surviving rough environments (Fig. 43.1c). In the *Active Matter Game*, the bacteria are attracted to the Goal while repelled by the obstacles, which could be translated into positive taxis towards victory and negative taxis to failure.

Active matter can also be programmed in artificial systems both on the microscopic (see Chapter 39, Living Robotics) and macroscopic scales (see Chapter 42, Locomotor Transition in Complex Terrain). These robots can reproduce collective behavior in a direct way, without the challenges of biological systems (see Chapter 42, Robotic Swarms; Fig. 43.1d). To reach the Goal, the bacteria perform a combination of random and directed motion (see Chapter 5, From Passive to Active Brownian Motion). While moving, they encounter several challenges associated to swimming at the microscale. In biology, bacteria can reach most environments thanks to their motility, from tiny cracks in soil to setups in laboratory experiments. However, swimming at the bacterial scale is different from the way we swim, making the field of microfluidics not intuitive for us (see Chapter 7, Low Reynolds Number Hydrodynamics).

Microfluidics (see Chapter 36 Microfluidics) studies the behavior of liquids at the microscopic scale, such as the tiny spaces where our heroes dwell. The players can imagine the game takes place inside a *microfluidic chip* (Fig. 43.2a) in a laboratory experiment. Microfluidic systems are usually characterized by laminar flow, so in absence of turbulence and unpredictable behavior of the fluids. Under such conditions, bacteria can easily swim to maintain their preferred direction. Nevertheless, strong currents in the environment might force them away from their trajectory, or even worse, cause the formation of a bubble around them. While bubbles might not be deadly to bacteria themselves, they certainly are for any experiment with microfluidics. Still, bacteria can be trapped at the interface between liquid and air due to surface tension (see Chapter 30 Active Colloids at Fluid Interfaces), meaning the end of their journey.

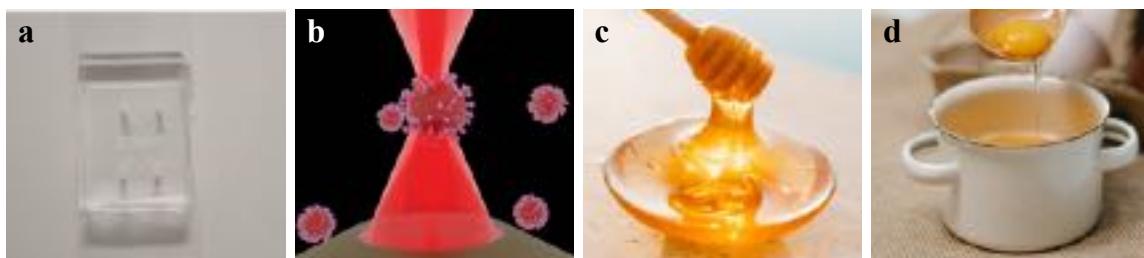


Figure 43.2: Swimming at the microscale. (a) A microfluidic chip, a plastic device for the control of fluids at the microscale (picture by J. M. Antúnez Domínguez). (b) Schematic representation of a virus held in place by an optical trap (illustration by Martin Selin). (c) Honey is an example of a highly viscous fluid, that requires a lot of energy to swim through (photo from Pexels). (d) Egg white is an example of viscoelastic fluid, due to the several different proteins embedded in it (photo from Pexels).

A less tragic form of trapping at the microscale can be found in *optical tweezers* (see Chapter 31, Optical Tweezers; Fig. 43.2b). By applying a focused laser beam, attractive and repulsive forces can be generated on objects. These forces are too small to be relevant at human scales, but they can be used to hold and move microscopic entities around with ease. Optical tweezers do not require contact with the sample, and if used carefully, they do not cause any harm to it. For this reasons, they have been used extensively with particles, droplets, molecules, cells and, of course, bacteria. Therefore, our brave bacteria have to stay away from the scientists trying to trap them for long time far away from their goal.

Other obstacles encountered by our microscopic friends come from the diversity of the environment. Even if the medium is liquid, it is not completely homogeneous, and different regions can have different consistencies (see Chapter 13, Activity in the Cytoskeleton). *Viscosity* is the property that describes how thick or thin a fluid is. For example, honey is a very viscous fluid (Fig. 43.2c), making it difficult to cross. Viscosity is related to how much energy a swimmer dissipates to move through it, and as a result a viscous fluid in the game slows down the bacteria. In other cases not all energy is dissipated, but rather stored in the liquid, similar to what happens when a spring is compressed. This phenomenon is known as *viscoelasticity*. For example, the proteins inside an egg give it viscoelastic properties (Fig. 43.2d). The ability to whip up egg white through fast movements derives from its viscoelasticity. If the liquid absorbs too much energy from a moving swimmer in a short period of time, the energy is projected back into the swimmer, causing it to bounce back. If instead the swimmer moves slower, its energy is dissipated according to its viscosity, allowing it to move forward. In the microscale, bacteria can find changes in viscosity due to different temperatures and substrates. Organic environments tend to have viscoelastic properties, such as the mucus in the digestive track of animals, saliva, and blood (see Chapter 14, From Cells to Tissues).

But you are not playing only against the environment. One of the main interests of active matter systems come from the interaction among its components, by affecting each other trajectories and, in some cases, even communicating (see Chapter 20, Interaction Rules in Models of Active Matter). These collective behaviors lead to the emergence of complex patterns, as in the case of bacteria forming swarms to colonize new environments or prey on other microorganisms. In the game, you can influence the path of neighbors at your reach to cooperate or mess with them!

With all these examples inspired by real physical phenomena, we hope to spark the curiosity of children and adults alike in the field of active matter.

43.2 Playing the game

Welcome to the exciting world of microscale swimming! In the *Active Matter Game*, the bacteria are represented as meeples moving on a discretized space — a playing field consisting of hexagonal tiles (Fig. 43.3). Explore the rules, select your meeple, choose between an existing board or one of your own creation, and get ready for an exhilarating journey into active matter. The objective of this game is to be the first person to reach the *goal*. Although it may initially seem straightforward, proceed with caution. In the microscale world, there are abundant obstacles and traps awaiting you. Stay vigilant when interacting with the other players, for they possess the power to maneuver you, intensifying the game with added layers of strategy and difficulty.

In this section, you will find a step-by-step guide to how to play the Active Matter game, starting from the basics, such as the preparation before a game, the game start, and how to navigate the board, to more complex topics like the interaction with other players and obstacles.

Ready to start?

43.2.1 Preparation

To play a game, you must be at least two (and maximum six) players. You must have a board (Fig. 43.3), a number of meeples at least equal to the number of players, and two classic dice with 6 faces. You can also add some obstacles on the board to make the game more interesting (see Subsection 43.2.5 for how obstacles affect the motion of the meeples).

No scoresheet or point keeping is normally needed. Depending on the board setting (i.e., on the obstacles you interact with) you might have to remember, though, if you or someone else is trapped or hindered in its motion: in such case, some paper and a pencil to annotate the status of each player is recommended. We suggest that, if any simple book keeping is needed during the game, the player that starts the game also is in

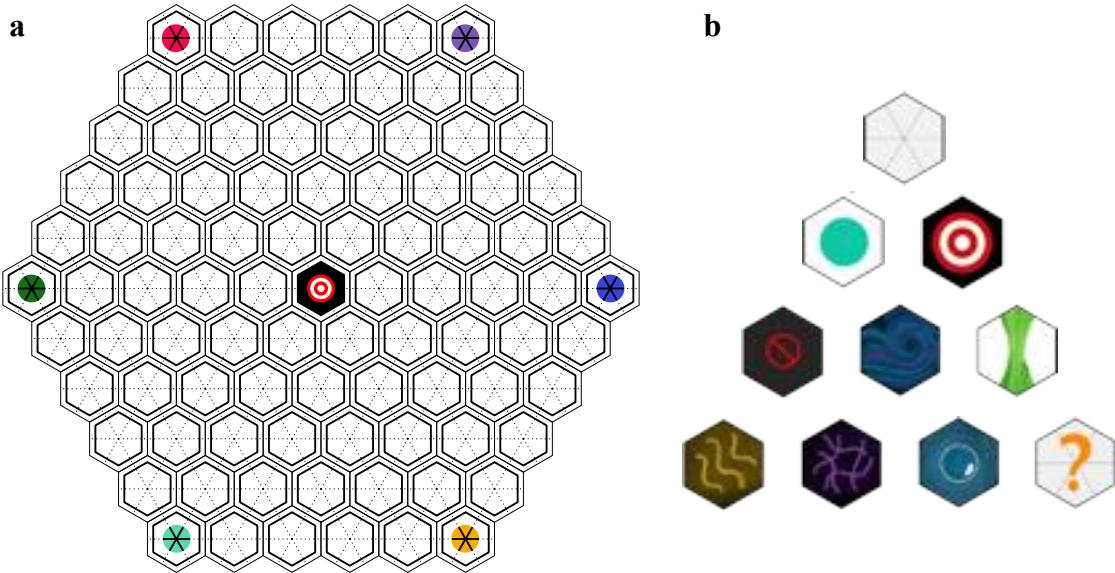


Figure 43.3: The basic board of the Active Matter Game. **a** The board of the *Active Matter Game* represented in its most basic, undecorated form, including only the spots of the starting positions for the meeples, located at the vertices of the hexagonal board, and the *goal*, located at the board center. **b** The various tiles that might appear in the board. The three tiles at the top of the pyramid are the mandatory ones, the remaining tiles are the *obstacles*, also illustrated more clearly in Fig. 43.7, and described in Tab. 43.1 and Subsec. 43.2.5. Starting from the top, left to right: first row – normal tile; second row – starting spot and goal; third row – solid barrier, turbulence tile, optical tweezer; fourth row – viscous tile, viscoelastic tile, bubble tile, random tile.

charge to annotate the general status of each player during a turn. Of course, you should remember to update the status when the trapping or hindering is removed.

When you read through this section for the first time, you can assume there are two or more players, each commanding one meeple, who take turns making their moves during each round. Each player plays for themselves and there is no implied team strategy: each player aims to win and all others are competitors (*competitive setup* of the game). We advise to start with this simple game modality to familiarize with the game and its rules.

Later, when all players are familiar with how meeples move and what are the interaction rules with other players and the obstacles, you can decide to play in a different game modality, accounting for teams and cooperation among players in the same team. To see other possible playing modalities and variations of the game, see Section 43.3.

Once you have set up the board and all players sit around the table, you are ready to start.

43.2.2 Let the game start

At the start of the game, each player throws a die, the one with the highest score starts and decides if the order is clockwise or anticlockwise. In turns, starting from the last player, each one chooses a meeple. The meeples are schematic representations of bacteria with a head and two helical tails (Fig. 43.4a). Each players must position their meeple on the corresponding (same color-coded¹) starting tile (Fig. 43.4b) at the corners of the board, oriented with its head pointing away from the center. Each meeple, at every turn, occupies one of the hexagonal tiles with well-defined orientation and direction. When it is time to move, each swimmer is allowed to *run*, i.e., move forward to the adjacent cell, in the direction from tail to head, and/or *tumble*, i.e., change its orientation while remaining in the same cell. Note that, at the very beginning, all meeples are positioned on a

¹All colors in the board were selected to ensure high contrast, suitable both for black and white printers and for players with color vision deficiency.

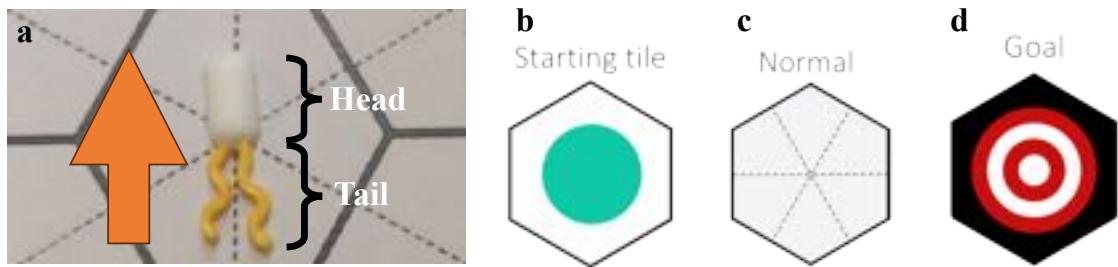


Figure 43.4: Basic elements of the board game. (a) The meeples represent bacteria and consist of two parts. The head points in the direction of motion, while the tails are flagella that allow them to swim. The meeples are initially placed in the starting tiles (b) pointing outwards from the board, move freely across normal tiles (c), and finally reach the goal (d).

corner and oriented such that they look away from the center: if you want your meeple to be able to explore the board, you have to remember to *tumble* to reorient your swimmer in a way where it can find an adjacent cell. Otherwise, your swimmer will not be able to move forward, as it is initially facing a hard wall.

Now that the swimmers are set on the board, you are ready to go!

43.2.3 Navigate the board

As you move across the board, keep in mind that the aim of the game is to be the first player to reach the goal tile (Fig. 43.4d). In each turn, you throw two dice, and the sum of your dice roll determines how many available steps you have. You need to consume all.

Each step can either be a *tumble* of 60 degrees angle or a *run* forward. Therefore, an hexagonal tile can be occupied by the meeple with six different orientations. The orientations are marked in the normal tiles (Fig. 43.4c) as dashed lines, and they apply not only to those, but to all tiles in the board. During your turn, you

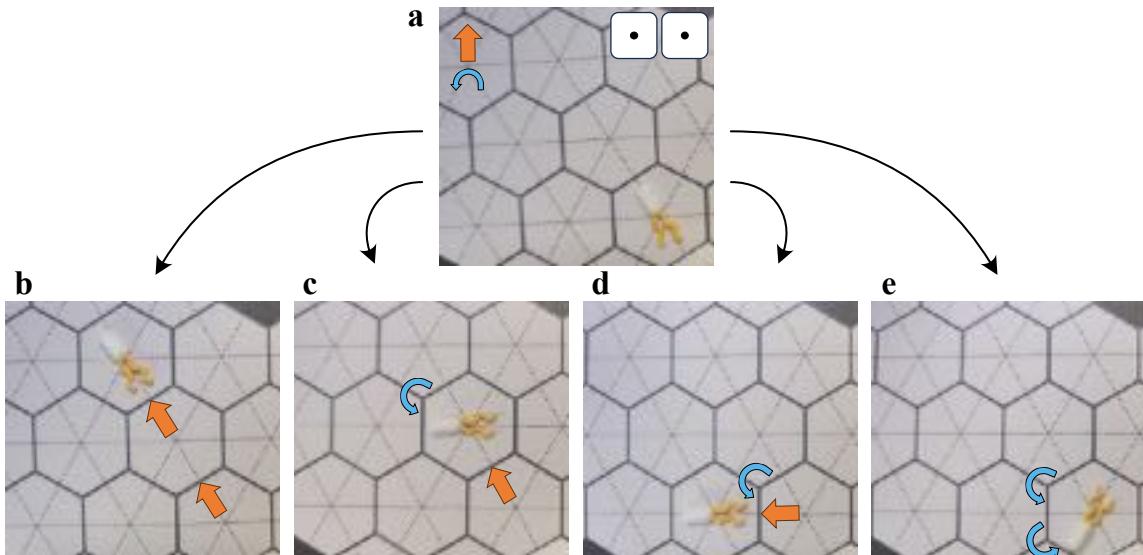


Figure 43.5: Possible moves in a turn. (a) If the sum of the dice values is two, each step can be spent as a run (orange arrow) or a tumble (blue, curved arrow), where each combination leads to a different outcome. (b-e) show the combinations of allowed moves: (b) run + run, (c) run + tumble, (d) tumble + run, (e) tumble + tumble. Keep in mind that tumbling can be done either clockwise or anticlockwise.

can choose to change direction only once, no matter of how many steps. The rotation can be either clockwise or anticlockwise, but not both. You can also choose to tumble in the middle of your turn, or not tumble at all.

In an exemplary turn, after throwing the dice (Fig. 43.5a), there are four possible outcomes: only run (Fig. 43.5b), run first and tumble at the end of the turn (Fig. 43.5c), tumble first and then run (Fig. 43.5d), tumble the whole turn (Fig. 43.4e). Remember that *you can tumble only once* when executing a move. See Example 43.1.

Example 43.1: The tumble-only-once rule. Refer to Fig. 43.5. Assume that, instead of ☒ and ☓, the dice are ☔ and ☕, for a total of 5 steps. Then, the following sequences of steps are admissible:

- Sequence 1:** run (1), **tumble (2)**, run (2)
- Sequence 2:** run (3), **tumble (1)**, run (1)
- Sequence 3:** **tumble (3)**, run (2)
- Sequence 4:** run (3), **tumble (2)**
- Sequence 5:** **tumble (5)**
- Sequence 6:** run (5)

and the following, which contain two or more distinct tumble moments separated by a run, are not admissible:

- Sequence 7 (not allowed):** **tumble (1)**, run (1), **tumble (2)**, run (1)
- Sequence 8 (not allowed):** run (1), **tumble (1)**, run (1), **tumble (2)**
- Sequence 9 (not allowed):** **tumble (1)**, run (1), **tumble (1)**, run (1), **tumble (1)**

From the same tile and dice result, you can end up in very different places. Plan carefully your next move!

43.2.4 Interaction with other swimmers

[The number of points to be subtracted when interacting with another player is confused and not clear.] Bacteria interact inside their colony and your meeples can do the same in the *Active Matter Game*. However, the interaction is limited in distance. During your turn, you can choose to move another player's meeple, instead of your own. This only applies if the meeples are on the same or adjacent tiles. You have to pay double the price though. You consume two of your own steps for each step that the other player has to take. Use this twist to push your rivals into a trap (or help your teammates if you play in teams).

Figure 43.6a shows an exemplary situation where two meeples are in adjacent tiles and, during the yellow player's turn, the dice sum is three. In this case, they can use the steps in two ways: either to move their meeple (Fig. 43.6b) or someone else's (Fig. 43.6c), but they cannot do both. Be aware that you need to decide it *before throwing your dice*, and that an odd even?) number of steps is subtracted one when spent on another player. Note that the other player has no way to counteract your decision. [Not clear. Rephrase it] This interaction is non-mutual (i.e., if you act on another player's adjacent meeple, that player is not compelled to exert their next move on your meeple), but it can be transitive. See Example 43.2.

Example 43.2: Transitivity of the action on an adjacent player. Consider a situation similar to the one depicted in Fig. 43.6a, but with a third meeple (assume it is blue) adjacent to the red meeple, but not adjacent to the yellow one. The yellow player can in principle decide to act upon the blue meeple, which lies at a distance of two steps from its meeple, through the action exerted by the red meeple that is adjacent to both yellow and blue. As in the case of the action upon an adjacent swimmer, the player has to declare their intention at the beginning of their turn, before tossing the dice, and the amount of moves they can act upon the blue meeple has to be calculated in a recursive way.

Let us examine the situation more closely. Suppose you are the yellow player and you declare that you want to act on the blue meeple. [The rule and the examples are not clear. Can you clarify them?]

- a. From tossing the dice, you get ☔ and ☔. Then you would have ☔ to act on the red meeple. The red meeple would be left with only ☐ to act on the blue meeple. Therefore, if you toss a ☔ ☔, you have ☐ to

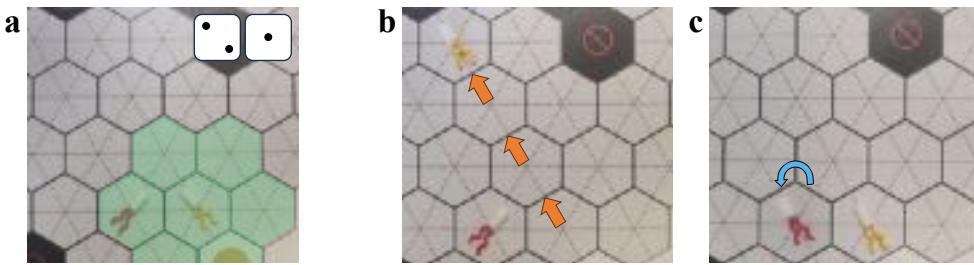


Figure 43.6: Interaction with other players. (a) The yellow player throws the dice and obtains a two and a one, while the red player is inside the interaction area (highlighted in lime green). (b) The yellow player can spend three steps on their meeple. (c) Alternatively, they can force the red player to move one step. [In this case, you are still left with one move. This should be included, otherwise it's not clear]

act upon the blue meeple, that is not in the immediate adjacency.

b. From tossing the dice, you get \square and \square . Then you would have \square to act on the red meeple. The red meeple would be left with only \square to act on the blue meeple. Therefore, if you toss a $\square \square$, you have \square to act upon the blue meeple.

c. From tossing the dice, you get \square and \square . Then you would have \square to act on the red meeple. The red meeple would be left with no move to act on the blue meeple. Therefore, if you toss a $\square \square$, you cannot affect in any way the condition of the blue meeple. Note that you cannot alter your decision *a posteriori* in any case. [Does this mean you can still act on the red meeple or not?]

43.2.5 Interaction with obstacles

Different obstacles are lied out around the board. Look out for them. There are seven types of obstacle tiles, see Table 43.1 and the detailed description in this subsection.

Some of the obstacle tiles require you to throw an auxiliary die, colored in green in the figures, while the dice for the regular rolls are colored in white. If the turn ends on a special tile, its effect applies also on your next turn, which you will start by throwing the auxiliary die. Keep in mind that the direction the head is



Figure 43.7: Obstacle tiles. Be cautious when you see these obstacles when you are on your way to the goal.

pointing at is relevant also for the special tiles, even if the lines are not explicitly drawn. Think carefully about the obstacle tiles, to either avoid them or use them cleverly in your game strategy.

Solid barrier

If you encounter a solid barrier while pointing at it (Fig. 43.8a), you cannot cross. If you are still allowed to tumble, you can freely decide how many steps you spend pushing against the barrier and how many you use to tumble (Fig. 43.8b).

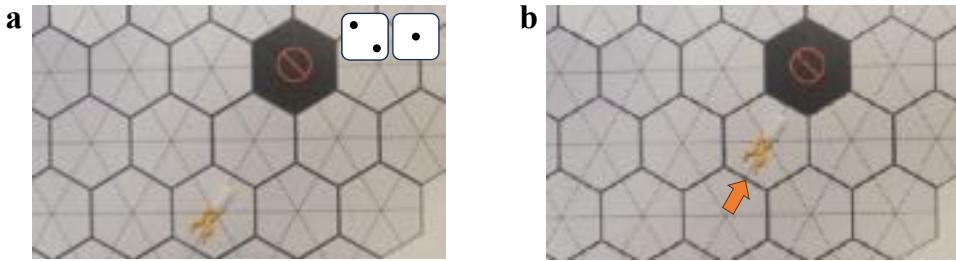


Figure 43.8: **Solid barrier working principle.** (a) You run into a tile next to a solid barrier while pointing at it. (b) You lose the remaining steps for your turn pushing against it. [unless you can tumble, right?]

Flow tile

If you end up on a flow tile (Fig. 43.9a), you can only overcome it by going fast enough. Once you cross one of these tiles, stop there and throw one die. If you get 4 or more, you are fast enough to overcome the flow (Fig. 43.9b) and you can continue to advance like nothing happened. **In this case, you are allowed an additional tumble if you already used it. [This is not consistent with what explained in the caption, can the player do an extra tumble or not?]** If you get less than 4, you get carried away by the current (Fig. 43.9c), so you must reorient in the direction of the flow and continue the rest of your steps accordingly. If there are multiple neighboring flow tiles, follow the arrows all the way, without counting the steps. Continue using your steps from the last flow tile and in the same direction as the arrow.

Optical trap

If you run into an optical trap, you lose the remaining steps of your turn (Fig. 43.10a). Next turn, you start by throwing an auxiliary dice. If the result is lower than 4 (Fig. 43.10b), you remain trapped for another turn. If the dice roll is 4 or higher (Fig. 43.10c), you can throw the dice again and use the steps to escape and move normally through tumble and run. Additionally, if you escape but you invest all steps into a tumble, move normally on the next turn, even if still on the optical trap tile.

Viscous area

Once you enter a viscous area, each movement on the board, for both running and tumbling, costs you two steps obtained with the dice. See Fig. 43.11. The effect lasts for the whole turn, even if the meeple succeeds in transitioning to a non-viscous tile. On the next turn, the meeple can behave normally, if it starts from a non-viscous tile. If it starts from a viscous area instead, it is still affected by the higher viscosity in its movements.

Viscoelastic area

You can interact with a viscoelastic tile (Fig. 43.12a) in two ways. To find out which one applies to you, throw the auxiliary die. If the result is below four, you are slow enough to access the tile and it acts as a viscous area, see Fig. 43.12b. If the result is four or higher, it acts as a boundary and you are forced to run the remaining steps in the opposite direction, see Fig. 43.12c.

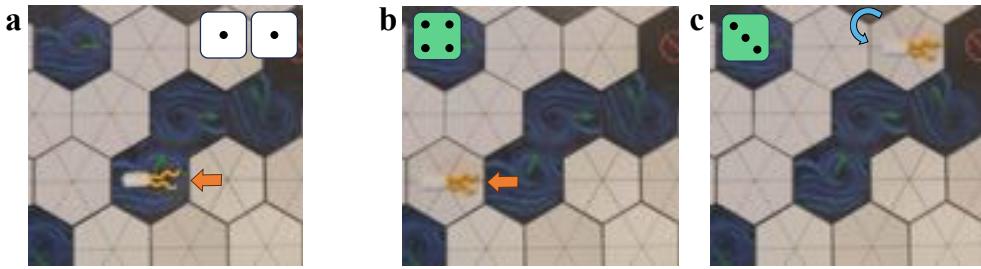


Figure 43.9: Flow working principle. (a) If you run into a flow tile, throw an auxiliary die. (b) If the result is 4 or higher, you can move normally. (c) If the result is lower than 4, then you are carried away by the flow: you are forced to re-orient in the direction indicated by the green arrow on the flow tile (no step is spent in doing this) and you are expelled into the nearby cell (no step is spent in doing this). In the case the next cell is again a flow tile, you do not throw the auxiliary die again, you are forced to re-orient in the direction of the green arrow in the new cell and expelled in the nearby cell. Continue to apply this rule until you end up in a cell that is different from a flow tile. At this point, you continue with your remaining steps from the cell you end up to. In the case represented here, the swimmer, carried away by the flow of the initial flow tile it encounters, is consecutively forced into other two flow tiles. After it is expelled into a normal cell, it still has one step left to use, which it uses for tumbling to the left. Note that, when carried out by a flow and forced to re-orient, no tumbling step is spent: if you have not tumbled in the move yet, you can use your possibility to tumble once at the end when exiting the flow.

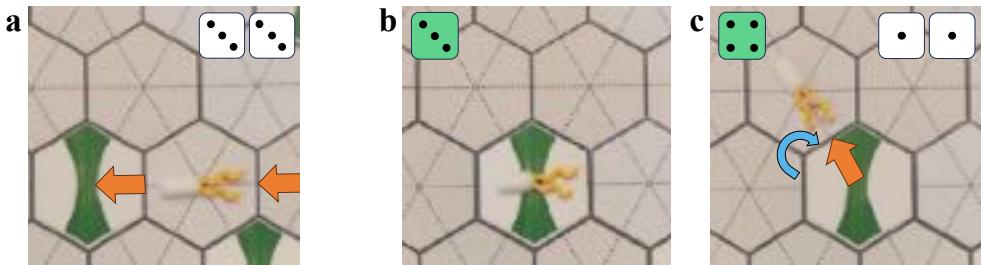


Figure 43.10: Optical trap working principle. (a) If you run into an optical trap, you lose the remaining steps of your turn. Next turn, throw an auxiliary die. (b) If the result is lower than 4, you stay trapped for another turn. (c) If it is 4 or higher, you can throw the dice again and use the resulting steps to move normally.

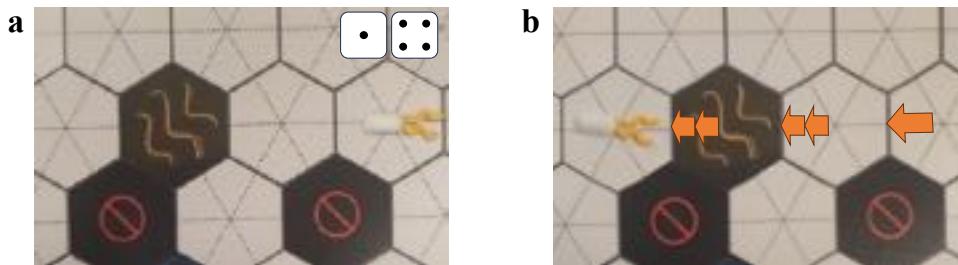


Figure 43.11: Viscous area working principle. If you run into a viscous area (a), every movement will cost more energy. (b) In other words, each run or tumble costs double the steps. If an odd number of steps remain, when entering a viscous tile, you lose one.

Bubble

Once you end up in a bubble, you go back to the start tile and reset both your position and your direction. See Fig. 43.13.

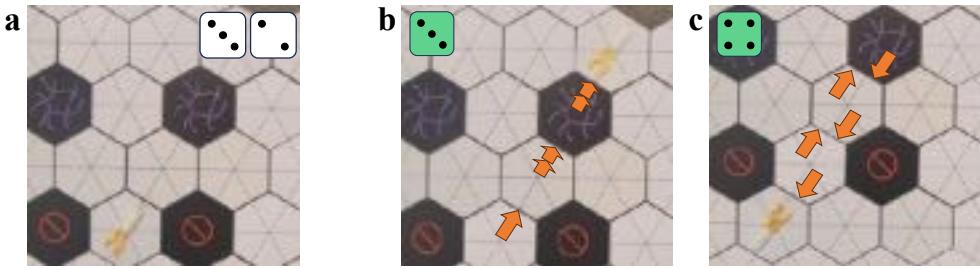


Figure 43.12: **Viscoelastic area working principle.** If you run into a viscoelastic area (a), throw a die. (b) If the result is lower than 4, then the effect is the same as a viscous area. (c) However, if the result is 4 or higher, you spend the remaining steps on a run starting from the viscoelastic area tile and facing the opposite direction of entry.

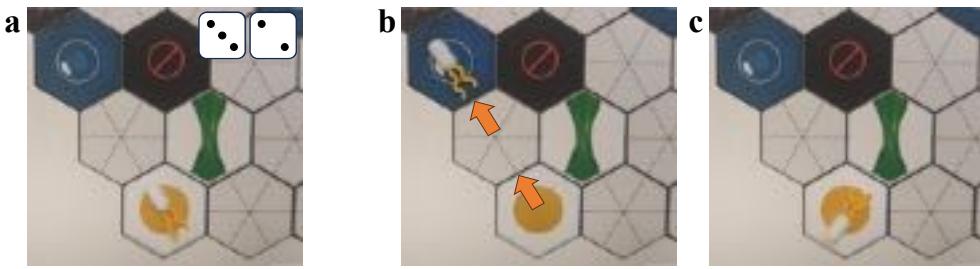


Figure 43.13: **Bubble working principle.** (a) Starting position with results of the dice. (b) If the meeple ends up on a bubble, you lose the remaining steps of the turn and (c) you are forced to return to the starting tile facing outwards.

Random

Random tiles act as a different tile every time you cross them, so throw a die to determine the behavior according to the indication in Tab. 43.2. The behavior applies until you leave that tile. In case the random tile behaves as an optical trap, there is an additional laser hazard sign that can be used to remind yourself and the other players. **However, the effect does not apply to another player controlling your movement. [Explain better]** The random tile appears to other players as a normal tile.

43.2.6 The game end

Once you cross the goal tile, congratulations, you won! If you have extra steps do not worry, you can use them to dance around!

43.3 Variations

In the previous sections, we have explained the basic rules of the game. The simplest way you can play the game is with two players, where one necessarily plays against the other. In the simplest case, each player controls one single meeple only. However, once you are familiar with the rules, you can play in different modalities to make the game more interesting. In the case of two players, you can decide at the beginning if you want to control one, two, or three meeples each. You will move each of your meeples once each round, alternating between the two players, always moving the meeples in the same sequence each round, for the entire game.

43.3.1 Multiple players

You can decide to play in a competitive setup with or without teams, depending on the number of players. The competitive modality (everybody against everybody) is always possible, independently on the number of players. Below (Tab. 43.3) we list the possible combinations we suggest in the case you want to play in teams (more than two players).

When playing in teams, one has to observe certain additional rules:

Alternation of teams' turns in a round. The team members cannot play consecutively: the sequence should be such that there is an alternation of teams. In the case of three teams, before the second player of a team plays, the first player of each team must have played its turn.

Each player moves its own meeple. Each player has a meeple. When the turn comes, the player cannot decide to transfer its moves to its fellow team mate.

Before a team game starts, the players have to decide what is the criterium for a team to win. The player agree if the game is won by the team who first has a meeple to reach the goal, or whether all the meeples of one team must reach the goal for a team to win. In this latter case, the player(s) reaching the goal continue(s) playing and throwing the dice. They can move around and affect other players of the same and of the adversarial team(s). The teams must annotate which players have reached the goal. It does not count if a meeple of a team reached the goal twice: all the players of a team must reach the goal to win the match in this modality.

43.3.2 Mystery mode

Once you mastered playing the game with the standard rules, you can add an extra layer of strategy by trying out the mystery mode! All the rules explained in the previous section apply but this time, you only discover the type of tiles once you step on them.

Required materials (Fig. 43.14):

- 1 blank printable board, where only the starting tiles and Goal are placed.
- 6 copies of each one of the obstacles
- 42 copies of the normal tiles

At the beginning of the game, the obstacle tiles and normal tiles are distributed equally among the players. Before the start of the game, the players fill in the board, except for the goal and starting tiles that are already placed. In turns, each player places one tile of their choice in an empty spot facing down, until the board is completely filled. Place your tiles strategically to pave your way to victory!

The game proceeds according to the regular rules, except the content of the tile (normal or obstacles) remains unknown until the first time a player ends up on it. Afterwards the tile is turned face up and the player is affected by the results. The tile remains uncovered afterwards and it applies to anyone passing through that tile.

43.4 Additional materials

[@authors: links must be added.]

A series of printable materials are provided to easily get started with the game.

Here [add link] are the instructions in a compact and user friendly format.

Here [add link] is the bacteria shaped meeple ready to 3D-print. We colored the tails according to the starting colors. Follow our style, or paint and personalize them as you wish!

We provide four board maps premade and ready to print here [add link]. It is recommended to print it in A3 format, to match the size of the 3D-printed meeples.

We also provide the blank map wit only start and end tiles and the designs for each individual tile here: [add link].

We include a laser hazard card to give to the players as a reminder, while they are stuck in an optical trap.



Figure 43.14: Mystery mode board setup. The board is empty and each of the players has access to a set of independent tiles with both obstacles and normal tiles. At the beginning, They place the tiles face-down in turns, so that the other players do not know what is underneath, leading to a strategic configuration of the map. Be sure the tiles cannot be differentiated when facing down!

43.4.1 Personalized boards

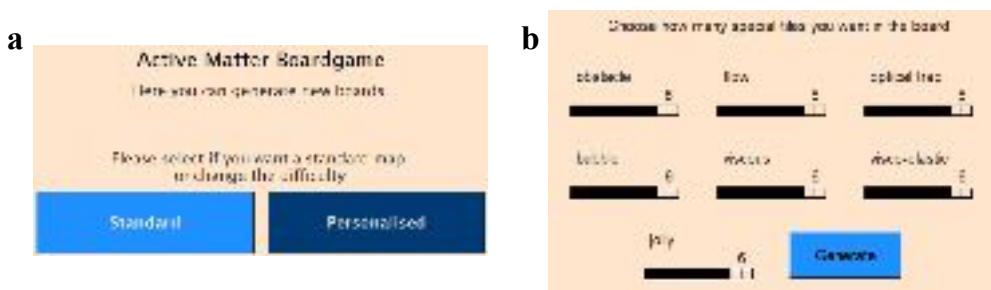


Figure 43.15: Map Generator interface. (a) The main window of the map generator, where you can choose between a standard map and a simplified one. (b) Optional feature window, where you can reduce the number of one or more of the tiles.

[@authors: links must be added.]

After you tried out the board maps provided, you can design your own board, to play endless different matches. You can also tune the level of difficulty as you wish or completely remove some types of obstacles if you do not like them! You can do so by downloading a package to create new boards, with randomized combinations of the obstacle tiles.

The instructions on how to make a map are very simple:

1. download the zip folder here ([link](#)) [[Include link](#)]
2. unzip the compressed file
3. open the folder *main*
4. run the *Map Generator* application

Once you run the program, the window in Fig. 43.15a appears. By selecting the box *standard*, the code generates a pdf file named *board.pdf* with the same number of obstacles but distributed in a new configuration.

In order to reduce the challenge or remove one or more of the obstacle tiles, you can select the box *personalized* instead. The selection guides you to a new window (Fig. 43.15b), where you can choose a number between 0 and 6 for each type of the obstacles described in section 43.2.5



Figure 43.16: **Example configuration.** A close-up of the board around the goal tile, close to the end of the game. Four players A (red), B (green), C (yellow) and D (blue) are one or two tiles away from the goal. Who is going to win?

43.5 Problems

Now you are ready to look at some problems related to the *Active Matter Game*!

Problem 43.1: Run and tumble dynamics of real bacteria. The meeples in the game represent bacteria and they have two types of movements: a run and a tumble.

- a. How do real bacteria move in a directed motion?
- b. How do real bacteria rotate their body?
- c. What is their mechanism to switch between running and tumbling?

Problem 43.2: Interactions among meeples. In the game, the meeples interact with the others by giving or receiving steps, when they are close to each other. How would you include more advanced interactions, such as flocking as a game dynamics?

Problem 43.3: Viscoelastic tiles. One of the most peculiar tiles in the game are the viscoelastic ones.

- a. What physical phenomenon inspires the behavior of viscoelastic tiles in the game?
- b. Why do bacteria crossing viscoelastic tiles manage to get through for small die throws and bounce back for high die throws?
- c. Can you think of more examples of viscoelastic materials?

For the remaining questions we refer to an example situation (Fig. 43.16). Four meeples A,B,C and D are very close to the goal tile. Let us see how they should move in order to be the winner.

Problem 43.4: Let's look at meeple A. The meeple labeled as A is on a normal tile very close to the goal. Only a viscoelastic tile separates them.

- a. What is the lowest number A needs to roll to reach the goal in one turn?
- b. What is the lowest number A needs to roll to reach the goal **safely** in one turn? [Hint: remember A can reach the target through more than one tile!]

Problem 43.5: Let's look at meeple B. The meeple B is the closest to the goal right now.

- a. What are the possible outcomes of its next move?
- b. What is the probability for B to win? [Hint: remember B still has to throw the auxiliary dice!]

Problem 43.6: Let's look at meeple C. The meeple C is also one tile away from the goal, but watch out! Flow tiles are on the way.

- a. What is the best move C can do?
- b. What is the smallest number of turns C needs to win?

Problem 43.7: Let's look at meeple D. Between the meeple D and the goal, there is only one of the mysterious random tiles... What is the probability for D to win in one turn? [Hint: remember the random tile has a probability of 1/6 per each type of behavior.]

Problem 43.8: Win in one turn. Which one of the meeples has the highest probability to win in one turn?

Type	Description
Solid barrier	<p>Sorry, you cannot cross.</p>  <p>If you encounter a solid barrier while pointing at it, you cannot cross it. If you are still allowed to tumble, you can freely decide how many steps you spend pushing against the barrier and how many you use to tumble.</p>
Flow tile	<p>You find yourself in a strong flow.</p>  <p>Once you cross one of these tiles, stop there and throw one die. If you get 4 or more, continue to advance like nothing happened. If you get less than 4, you must reorient in the direction of the flow and continue the rest of your steps accordingly. If there are multiple neighboring flow tiles, follow the arrows [Which arrows? There are no arrows in the flow tile. Which direction do you want the player to follow?] all the way, without counting the steps. Continue using your steps from the last flow tile and in the same direction as the arrow.</p>
Optical trap	<p>Optical forces trapped you on this tile.</p>  <p>If you run into an optical trap, you lose the remaining steps of your turn. Next turn, you start by throwing an auxiliary dice. If the result is lower than 4, you remain trapped for another turn. If the dice roll is 4 or higher, you can throw the dice again and use the steps to escape and move normally through tumble and run. Additionally, if you escape but you invest all steps into a tumble, move normally on the next turn, even if still on the optical trap tile.</p>
Viscous area	<p>Moving is harder in this region.</p>  <p>Once you enter a viscous area, each movement on the board, for both running and tumbling, costs you two steps obtained with the dice. The effect lasts for the whole turn, even if the meeple succeeds in transitioning to a non-viscous tile. On the next turn, the meeple can behave normally, if it starts from a non-viscous tile. If it starts from a viscous area instead, it will still be affected by the higher viscosity in its movements.</p>
Viscoelastic area	<p>It can act as either a viscous area or an elastic barrier.</p>  <p>You can interact with a viscoelastic tile in two ways. To find out which one applies to you, throw a die. If the result is lower than 4, you are slow enough to access the tile and it acts as a viscous area. If the result is 4 or higher, it acts as an elastic boundary: you are forced to invert your direction of 180 degrees and run the remaining steps in this configuration. No tumble allowed.</p>
Bubble	<p>You are tunneled back to your specific start tile.</p>  <p>Once you end up in a bubble, you go back to the start tile and reset both your position and your direction to the starting ones.</p>
Random	<p>Welcome to the mysterious world of the unknown!</p>  <p>Throw a die to determine the behavior according to the indication in Tab. 43.2. The behavior applies until you leave that tile. In case the random tile behaves as an optical trap, you can use the laser hazard sign to remind yourself and the other players. However, the effect does not apply to another player controlling your movement. The random tile appears to other players as a normal tile. [Not clear, explain better]</p>

Table 43.1: **Types of objects and obstacles.** List of the standard types of objects and obstacles that one can find on a decorated board, and a brief description. See Fig. 43.7 for the graphical representation of each object.

Die	Effect
•	optical trap
••	viscous area
•••	viscoelastic area
••••	bubble
•••••	normal, empty tile

Table 43.2: **Random tile working principle.** Consult the table to determine the behavior of the random tile, after tossing a die.

Number of players	Game modality
3	two players with one meeple each versus a single player controlling two meeples
4	two players versus two players, controlling one meeple each
5	two teams, one with three players and one with two players and one extra swimmer
6	two teams of three players or three teams of two players

Table 43.3: **Possible arrangements in teams.** Consult the table to determine the behavior of the random tile, after tossing the dice.

Acknowledgements

This book is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 812780.

Carlo Manzo and **Gorka Muñoz-Gil** would like to thank the co-organizers of the [AnDi challenge](#) for several stimulating discussions about the topics presented in this lecture.

Hartmut Löwen thanks R. Wittkowski, C. Hoell, A. M. Menzel and R. Wittmann for many helpful suggestions.

Francois Nedelec acknowledges that this research was in part supported by the Gatsby Charitable Foundation. We thank Carlos Lugo-Velez, Eashan Saikia and Helen Saville for critically reading the document.

Cristovão S. Dias acknowledges financial support from the Portuguese Foundation for Science and Technology (FCT) under Contracts no. PTDC/FIS-MAC/28146/2017 (LISBOA-01-0145-FEDER-028146), CECIND/00586/2017, UIDB/00618/2020, and UIDP/00618/2020.

Juliane Simmchen acknowledges Priyanka Sharan and Sandra Heckel for proofreading and discussions, as well as Freigeist funding (no 91619) from the Volkswagen foundation.

Félix Ginot and **Chun-Jen Chen** would like to thank Tobias Bäuerle, one of the developer of the experimental system in chapter [28](#). Some of his thesis material has been extensively used during the preparation of chapter [28](#).

Lucio Isa acknowledges that experiments have been carried out and recorded by Dr. Jacopo Vialeto (jacopo.vialeto@mat.ethz.ch) and Dr. Minghan Hu (minghan.hu@mat.ethz.ch). The code for the analysis of the pendant drop measurements has been kindly provided by Prof. Nick Jaensson from TU Eindhoven.

Rafael Mestre would like to thank UK Research and Innovation (UKRI) funding (grant refs ES/X006255/1 and MR/S032711/1).

Maria Guix acknowledges the financial support from the Spanish Ministry of Science through the Ramon y Cajal Grant No. RYC2020-945030119-I and European Union's Horizon Europe Research and Innovation Programme under EVA project (GA No:101047081).