Introduction to Domain Science

By Prof. Roman Voronov
New Jersey Institute of Technology

Outline

- Simulation Scales Overview
- Why use Computational Modeling?
- Molecular Simulations

    Electronic - Density Functional Theory (DFT): Schrödinger Equation

    Atomistic - Molecular Dynamics (MD): F = m * a

    Atomistic - Metropolis Monte Carlo (MC): Random Movement Generation

- Mesoscopic Simulations

    Coarse-Grained: MD & MC

    Lattice Boltzmann Method

- Macroscopic Simulations

    Continuum Mechanics

    Astrophysics

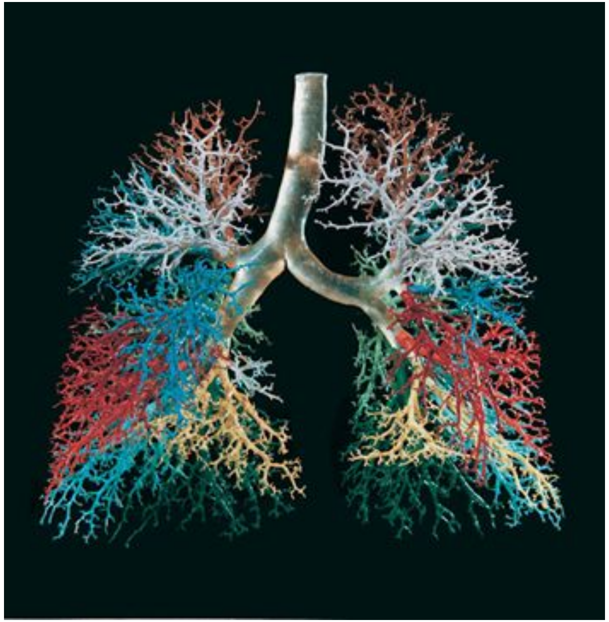
Spatial & Temporal Problem Scales

Length (m)	Phenomena
10^{-18}-10^{-15}	quarks, strings
10^{-15}-10^{-12}	proton, neutron
10^{-12}-10^{-9}	gamma rays, X rays, hydrogen atom
10^{-9}-10^{-6}	DNA, virus, optical light
10^{-6}-10^{-3}	bacteria, fog, human hair
10^{-3}-10^{0}	mosquito, golf ball, football
10^{0}-10^{3}	people, football field, Eiffel tower
10^{3}-10^{6}	Mt. Everest, Panama Canal, asteroid
10^{6}-10^{9}	Moon, Earth, light-second
10^{9}-10^{12}	Sun, light-minute, Earth's orbit
10^{12}-10^{15}	Solar System
10^{15}-10^{18}	light-year, nearest star
10^{18}-10^{21}	galactic arm
10^{21}-10^{24}	Milky Way, distance to Andromeda galaxy
10^{24}-10^{26}	visible universe

Time Scale (s)	Phenomena
10^{-44}	Planck time
10^{-24}	light crosses nucleus
10^{-15}	atomic vibration, visible light
10^{-12}	IBM SiGe transistor
10^{-9}	1 Gz CPU
10^{-6}	protein folding, lightning bolt
10^{-3}	hard disk seek time, blink of an eye
10^{0}	earthquakes
10^{2}	tornadoes
10^{5}	hurricanes
10^{7}	year
10^{9}	human life span
10^{10}	deep ocean mixing time
10^{12}	first homo sapiens
10^{15}	Milky Way rotation period
10^{17}	age of universe

Direct Human Experience

Why Modeling: Real Problems are Complex...



Lung Airways

Brain Vasculature

Why Modeling: Analytical Solutions are Limited!

As an example, let's consider the equations that describe *fluid flow*:

$$\rho g_x - \frac{\partial p}{\partial x} + \mu\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) = \rho \frac{Du}{Dt}$$

$$\rho g_y - \frac{\partial p}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) = \rho \frac{Dv}{Dt}$$

$$\rho g_z - \frac{\partial p}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) = \rho \frac{Dw}{Dt}$$

Navier-Stokes Equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{V}) = 0$$

Continuity

They are **Coupled**, **Nonlinear**, **Partial Differential Equations** ...

Can YOU solve them?

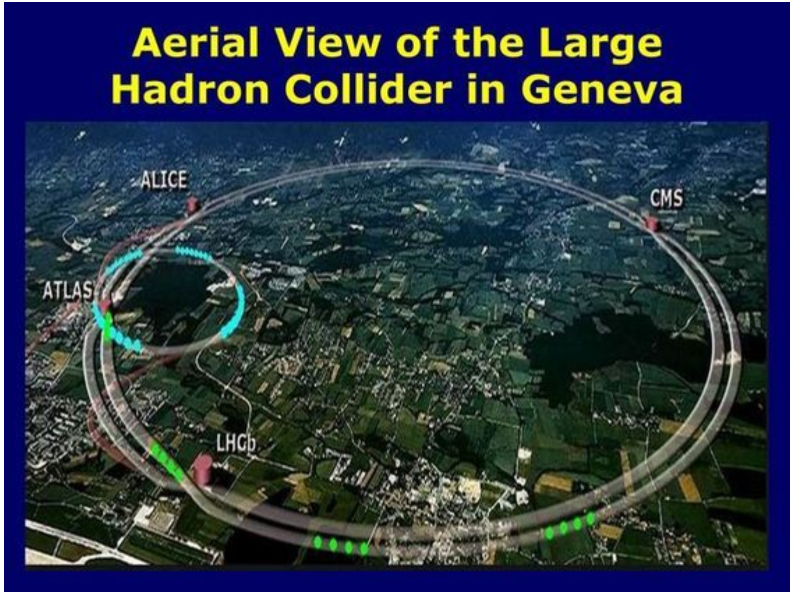
Exact solutions are only available for ***simple*** geometries: like flow in a *Channel*, *Duct*, *Pipe*, *Annulus*; or around *Steps / Cavity*, *Cylinders*, *Spheres*.

It is a "**Millennium Prize**" problem → You get **$1 million** for the ***general*** solution!

https://www.claymath.org/millennium-problems/navier%E2%80%93stokes-equation

Why Modeling: Experiments can be VERY expensive!

James Webb Telescope
(**$10 billion** USD, 2016)



Large Hadron Collider, CERN (**$4.75 billion** +
and **$1billion /yr** operating cost, 2010)

Why Modeling: Political / Ethical / Safety Reasons

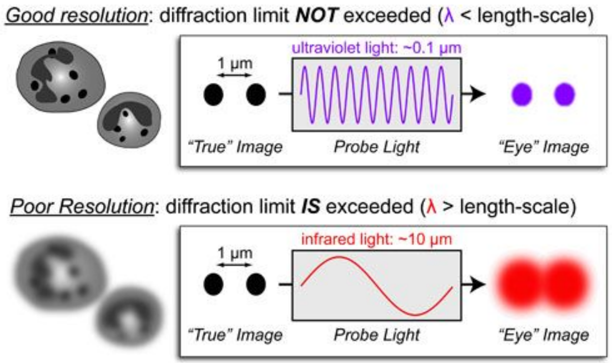
- Human Embryonic Stem Cell research bans (was federal, now by state)
- Nuclear Test-Ban Treaty → Department of Energy built large supercomputers to simulate nuclear weapons:
  Summit Supercomputer was No. 1 in the World
  June 2018 to November 2019

- Chemical Plant Safety Simulations

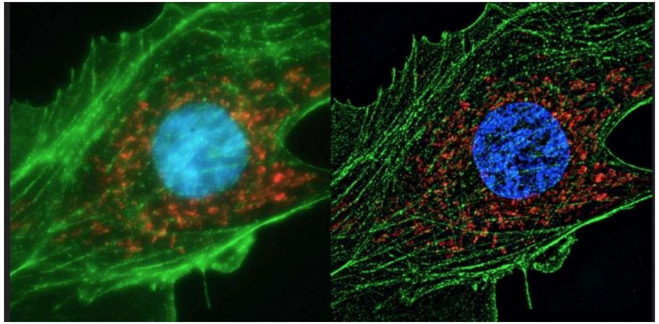
https://www.energy.gov/science-innovation/science-technology/computing

Why Modeling: Inherent Technology Limitations

For example, Optical Microscopy has a limit:

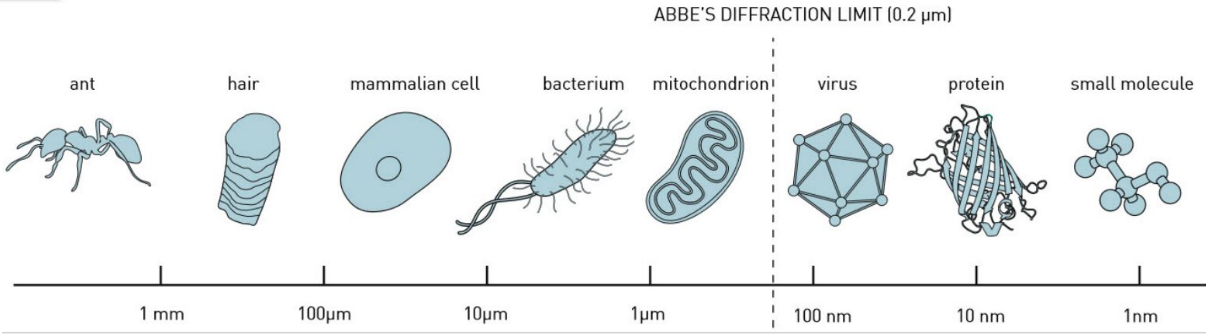


2014 Nobel Prize in Chemistry For Super-Resolution → nm Range



Difference between confocal and super resolution view of a human cell. Credit: Dr. Dr. Kandasamy Biomedical Microscopy Core University of Georgia.

But still limited to **>nm** objects and cannot image deeper than **~100μm**



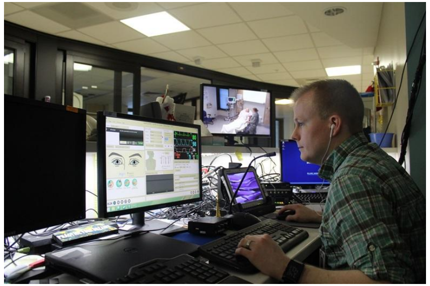
Ultimate Goal: Experiments + Modeling Together

Joke: *"Only the person who coded the simulation believes its results, while everybody but the person who worked in the lab believes the experiment."*

Empirical Validation

Experiment

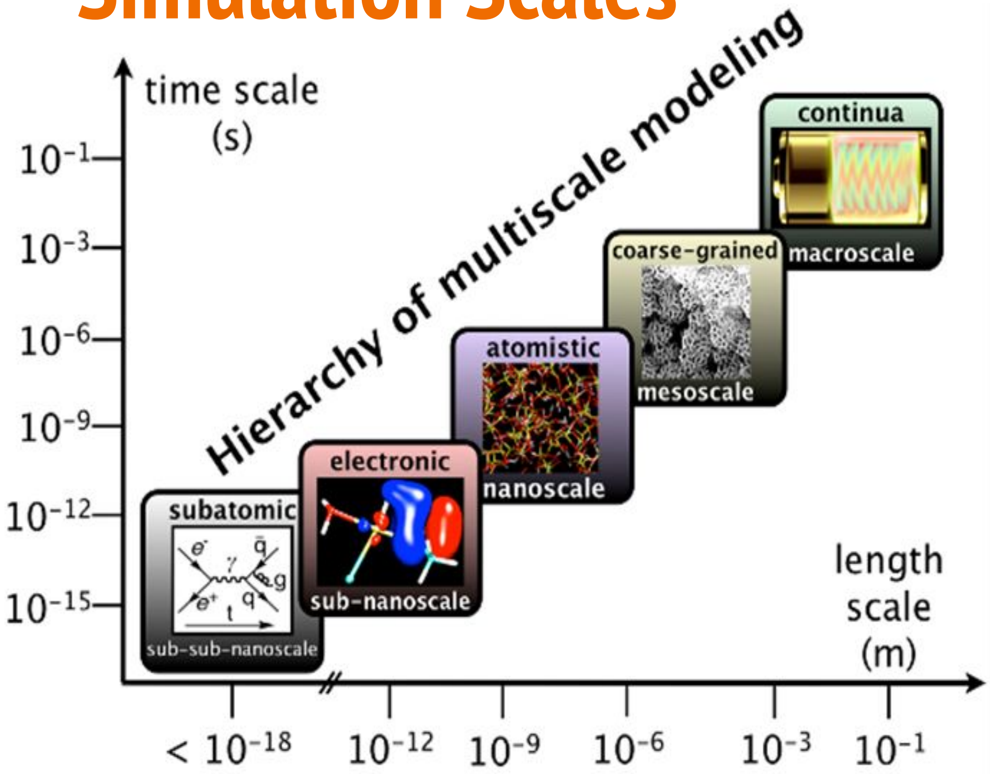
Theory and Computation

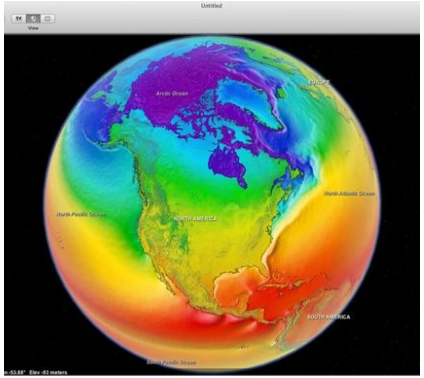


Experimental-Modeling Collaboration

Simulated Prediction

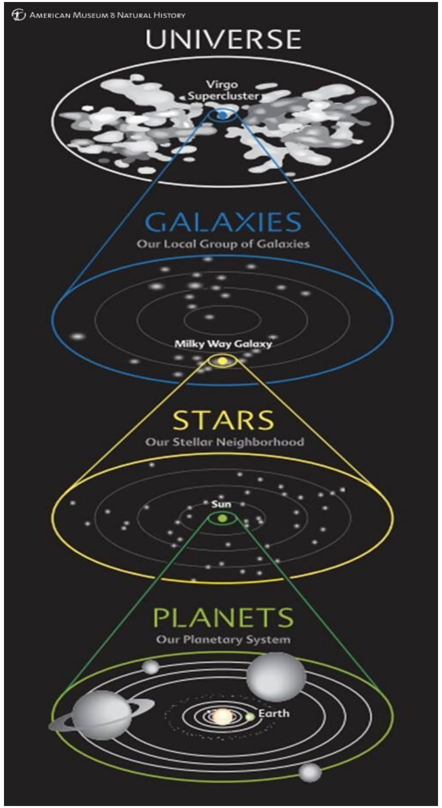
Simulation Scales





Global Weather

Tornado



Astrophysics

MULTI-SCALE MODELING

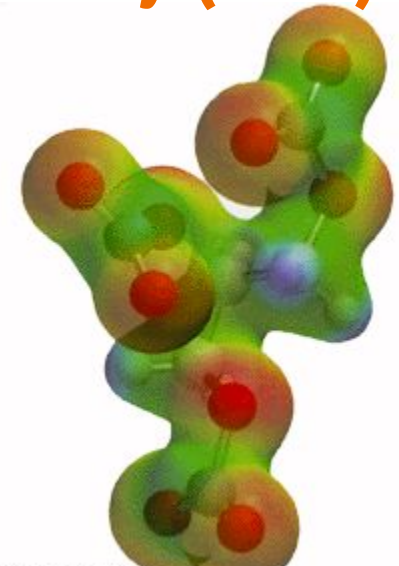
Electronic Structure: Density Functional Theory (DFT)

Nobel Prize in Chemistry 1998

Common Applications: Condensed-matter physics, Computational physics and chemistry (e.g., catalysis)

How it Works: Solves the **Schrödinger** Equation

What it Calculates: Electron Density Around Atoms → Force Fields / Interactions between Atoms



The electron density of nitroglycerine

Popular Codes:

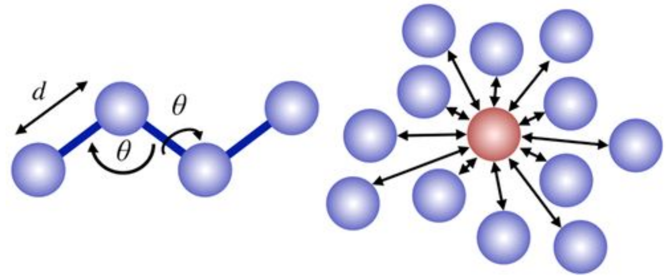
GAMESS | Expanding the limits of computational chemistry | PQS | VASP

https://www.msg.chem.iastate.edu/gamess/ https://gaussian.com http://www.pqs-chem.com https://www.vasp.at

Atomistic Simulations: Molecular Dynamics (MD)

<u>Common Applications</u>: Anything w/ atoms (**high ρ**) - Drug Design, Material Science...

<u>How it Works:</u> Calculates Force Interactions (i.e, solves **F = m*a**) between many atoms. No explicit electrons are modeled, the forces between "atoms" are *assumed:*

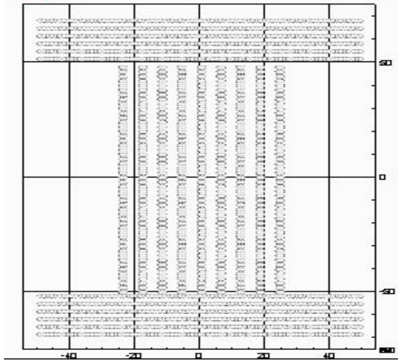


Bonded interactions   Nonbonded interactions

- Van der Waals (i.e., Covalent)
- Coulomb Electrostatic
- Hydrogen Bonds
- Intramolecular Interactions
- Hybrid MD/DFT is possible

(MD = atom motion / DFT = forces)

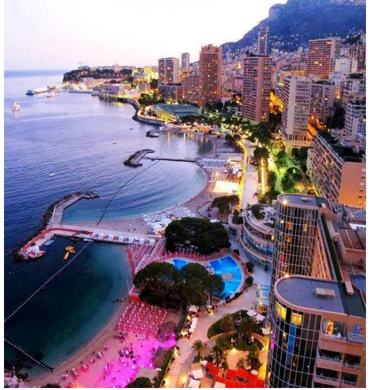
<u>What it Calculates:</u> Atom **Trajectories** → Animations, Statistical Averages (Pressure, Density, etc), Protein Folding Confirmations.



Formation of a Liquid Droplet
Video from Voronov et. al
https://doi.org/10.1021/ie0712941

<u>Popular Codes</u>: LAMMPS https://lammps.sandia.gov/   GROMACS FAST. FLEXIBLE. FREE. http://www.gromacs.org/

Atomistic: Metropolis Monte Carlo (MC)



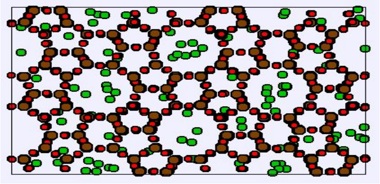
Monte Carlo, Monaco
("Las Vegas" of Europe)

Common Applications: Like MD → Anything with atoms (**Low ρ**)

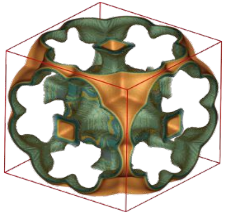
How it Works: Atoms are Moved ***Randomly*** Over Tiny Distances. Each move is then EITHER ***Accepted*** OR ***Discarded*** based on how *likely* it is to occur in nature (Interaction Forces + Boltzmann Distribution)! The forces are either assumed (e.g., MD) or calculated via DFT.

What it Calculates: Statistical Averages (Pressure, Density, etc), Likely Protein Folding Confirmations… but **NOT** Atom **Trajectories!!** → **Time** Sequence is **LOST!**

Popular Codes:



https://www.ccp5.ac.uk/DL_MONTE



https://wiki.iraspa.org/index.php/RASPA

MCCCS

Towhee
http://towhee.sourceforge.net/

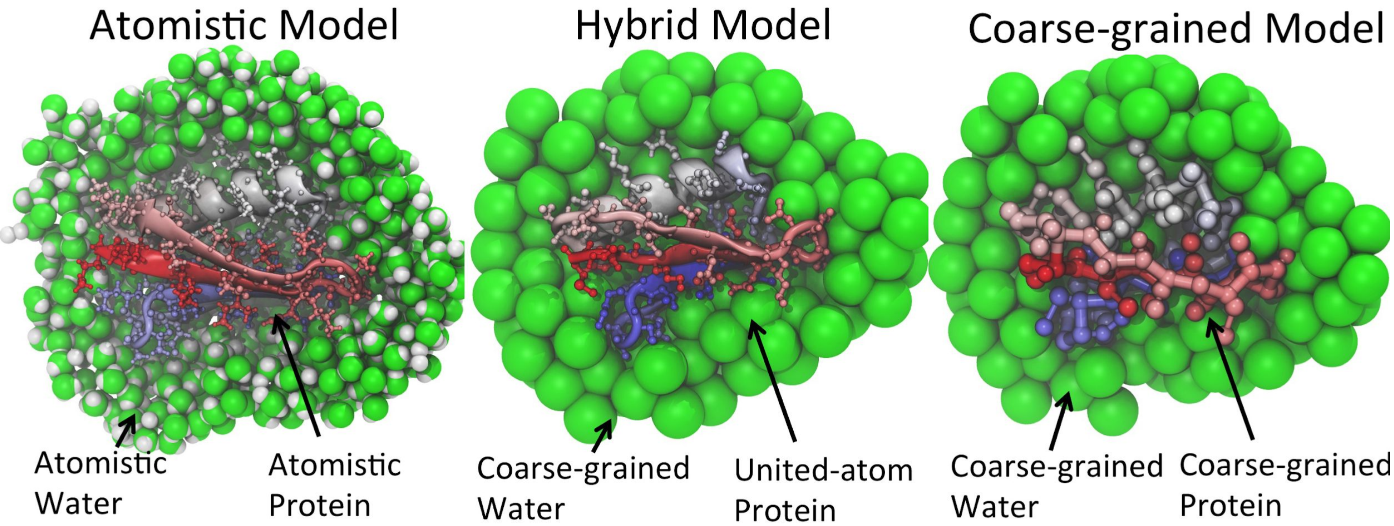
Cassandra
Monte Carlo
SOFTWARE
https://cassandra.nd.edu/

Mesoscopic: Coarse Grained MD/MC Simulations

Common Applications: ***Simple*** Representation of Complex systems (e.g. Proteins).

"United Atom" Model:
A group of atoms (e.g., CH_2, CH_3, etc.) are considered to be a ***single*** unit.

"Coarse Grained" Model:
Several united atoms are taken as one "bead".



Representation of a solvated protein at different levels of details. Protein G (PDB ID: 3GB1) is chosen for the illustration.

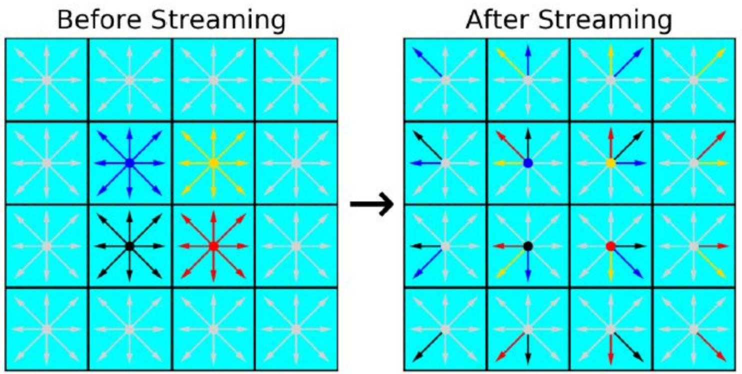
Codes:
Typically Same as MD & MC

Mesoscopic: Lattice Boltzmann / Gas Automata

<u>Common Applications</u>: Very Scalable on Supercomputers → Very **Large** Problems
Also, handles ***Complex Boundaries*** Easily → ***Non-Ideal*** Geometries (e.g., Porous Media)

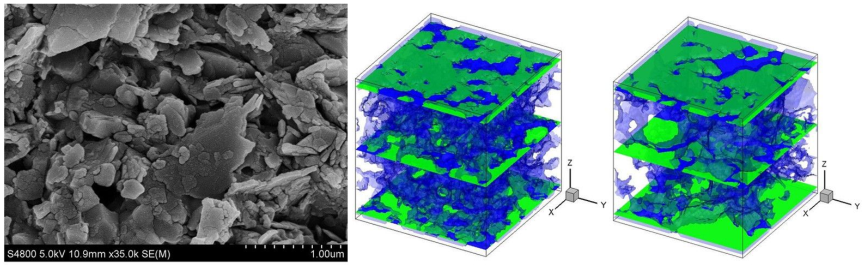
<u>How it Works</u>: Instead of Discrete Particles their ***Densities*** move on a fixed ***Lattice***.

The Molecular Nature of the "Fluid" is ***Hidden*** within the Streaming "Collisions"



The inner four cells are consistently colored to visualize the propagation.

<u>Application Example:</u>



(a) (b) (c)

Natural Gas Transport in Shale Rock

<u>Popular Codes</u>:

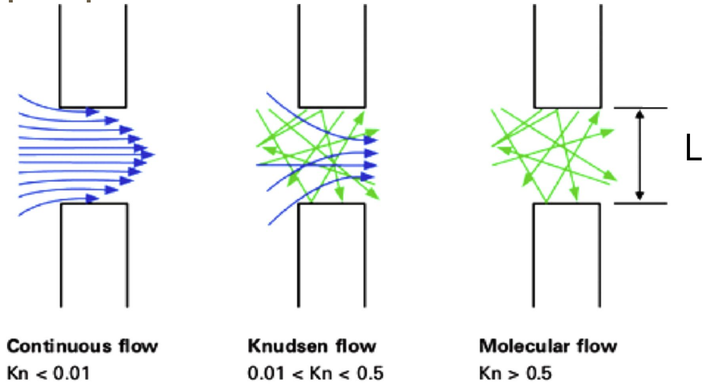
OpenLB
https://www.openlb.net/

PALABOS
https://palabos.unige.ch/

Traditional Continuum Mechanics: Conservation Eqns

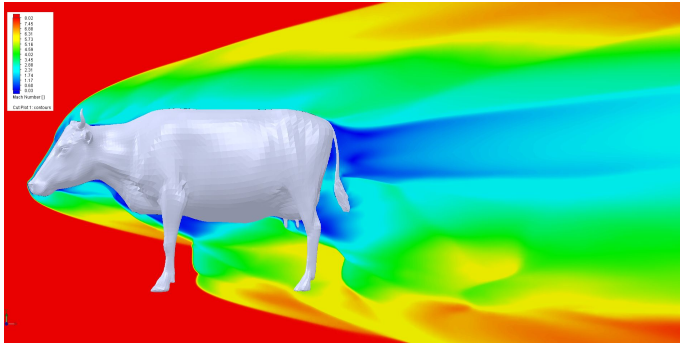
- The **Continuity** Assumption: Matter Has No Discrete Atoms or Molecules
- The Molecular Nature of a Fluid is "***Hidden***" within its Viscosity
- Solid Objects are Accounted for via Boundary Conditions (e.g., No Slip)
- Density, Velocity, Pressure, Temperature are Assumed to Vary ***Continuously*** with Space and Time → Conservation eqns. (i.e., Mass, Force, and Energy balances, like on Slide 5) of macroscopic properties are solved *numerically*.
- The assumption breaks down at **Knudsen (Kn) number > 0.01**

$$Kn = \frac{\text{Mean Travel Distance Before Collision}}{\text{Characteristic Length of System}}$$

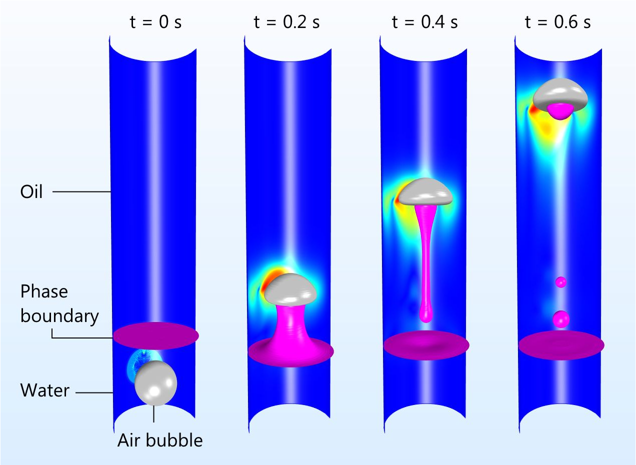


Traditional Continuum Mechanics: Examples & Codes

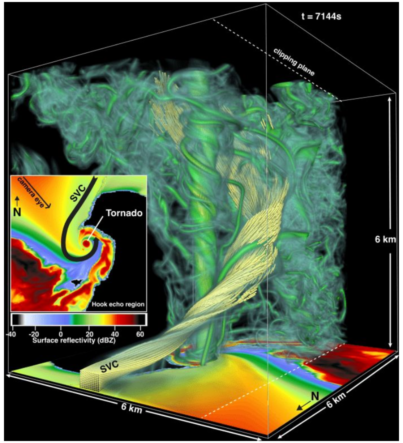
<u>Common Applications</u>: Automotive, Aerospace, Oil & Gas, Pharmaceutical Manufacturing (e.g., Chemical Reactors), Meteorology. Even flow over a cow… :)



Aerodynamics of a Cow



Air Bubble in Oil/Water



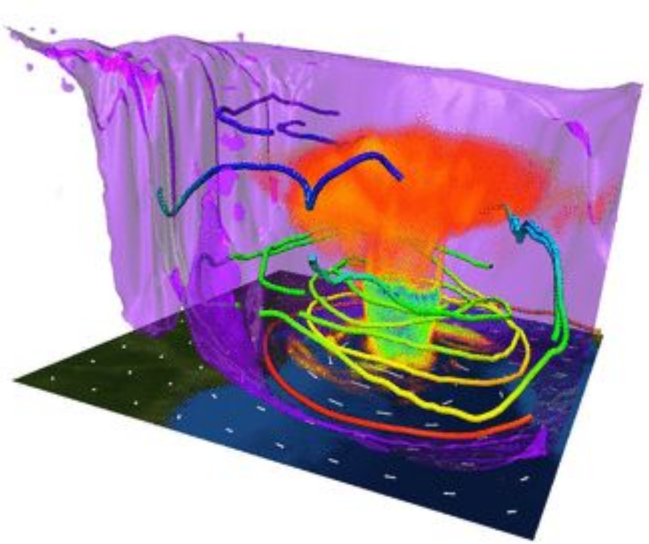
Tornado Structure

OpenFOAM
https://www.openfoam.com/

MFiX Multiphase Flow with Interphase eXchanges
https://mfix.netl.doe.gov/

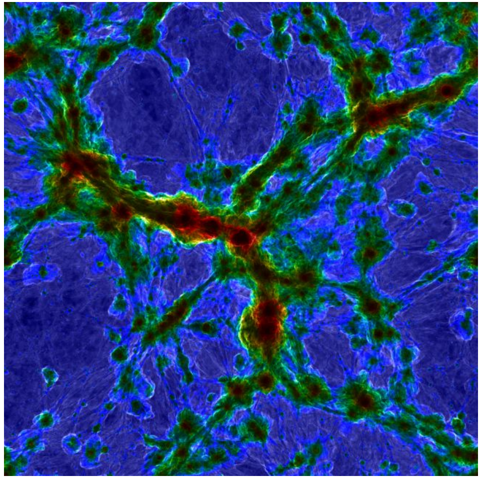
ANSYS Fluent
https://www.ansys.com

Larger Meteorology and Astrophysics Scales



A 3-D animated image of downscaled Global Forecast System (GFS) model data showing Hurricane Katrina making landfall on August 29, 2005. This image was generated with the Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers (VAPOR) tool and ImageMagick.

Large-Scale Simulation of the Known Universe: TNG50



The centers of massive galaxy clusters are super hot (red), while bright structures show diffuse gas from the intergalactic medium shock heating at the boundary between cosmic voids and filaments. (Image: © TNG Collaboration)

- **230 million light-years** wide (TNG 300 is even larger!)
- Contains **tens of thousands** of evolving galaxies
- Tracked more than **20 billion** particles
- Dark matter, gases, stars and supermassive black holes
- **13.8-billion-year** period
- **16,000 cores** on a Hazel Hen Supercomputer (Germany):

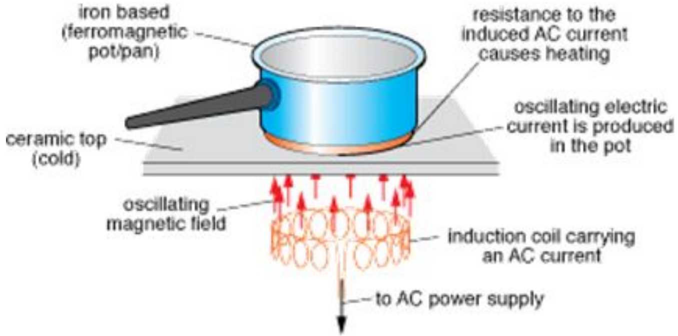
Peak performance	7420 TFlops
Number of compute nodes	7712
Memory/node	128 GB
Power consumption	~3200 KW

- A single processor would take **15,000 yrs** to compute

https://www.space.com/most-detailed-universe-simulation.html
https://www.hlrs.de/systems/cray-xc40-hazel-hen/
https://www.tng-project.org/about/

Other Types of Modeling: Top-Down (Data Science)

The goal is to use "artificial intelligence" to pick up **patterns** in *large* amounts of *data*, instead of relying on fundamental principles (i.e., the code can make predictions *without knowing how the system actually works*).



Physics-Based

← VERSUS →

By Making Associations (i.e., Machine Learning)

Prominent Example:

IBM Watson

https://en.wikipedia.org/wiki/Watson_(computer)