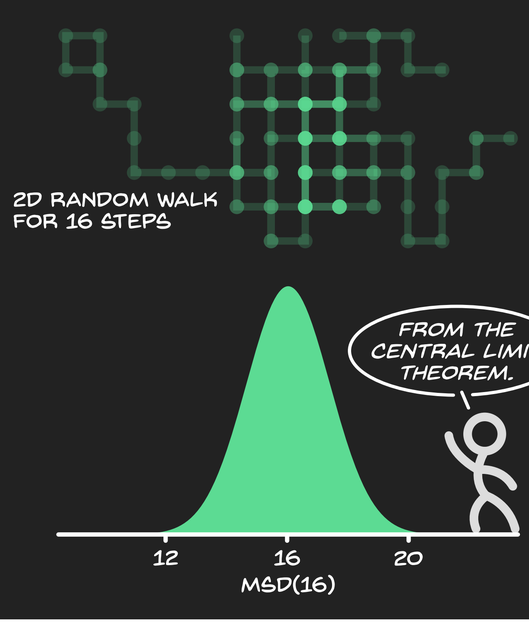
ACCURATE ESTIMATION OF DIFFUSION COEFFICIENTS AND THEIR UNCERTAINTIES FROM COMPUTER SIMULATION

ANDREW R. MCCLUSKEY[1,2], ALEXANDER G. SQUIRES[3], JOSH DUNN[1], SAMUEL W. COLES[4,5], BENJAMIN J. MORGAN[4,5]

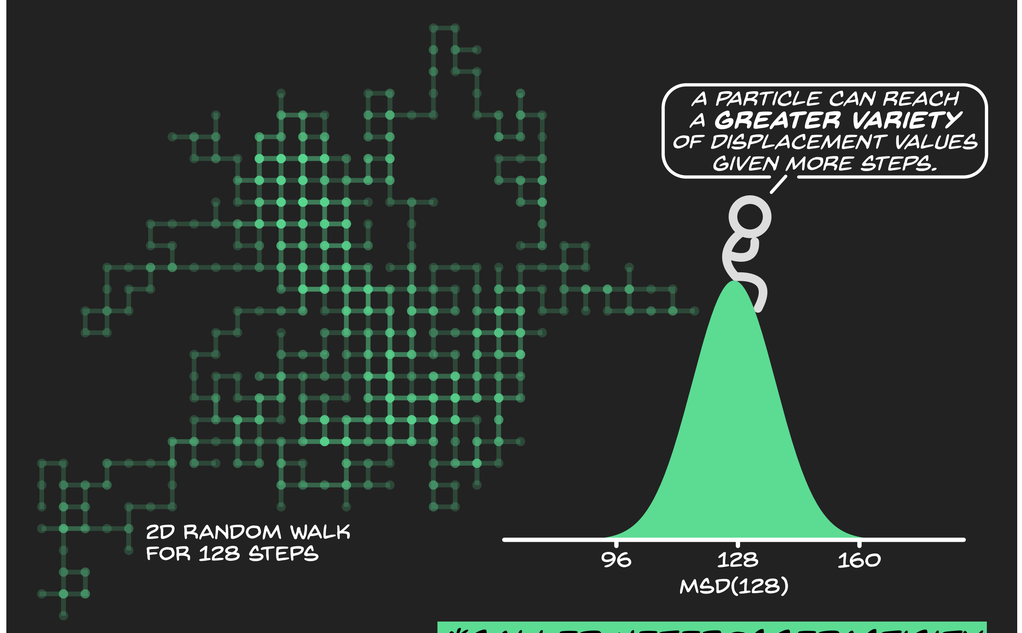
1. SCHOOL OF CHEMISTRY, UNIVERSITY OF BRISTOL, CANTOCK'S CLOSE, BRISTOL BS8 1TS, UK. 2. EUROPEAN SPALLATION SOURCE ERIC, OLE MAALØES VEJ 3, 2200 KØBENHAVN N, DK. 3. SCHOOL OF CHEMISTRY, UNIVERSITY OF BIRMINGHAM, EDGBASTON, BIRMINGHAM, B15 2TT, UK. 4. DEPARTMENT OF CHEMISTRY, UNIVERSITY OF BATH, CLAVERTON DOWN, BATH, BA2 7AY, UK. 5. THE FARADAY INSTITUTION, QUAD ONE, HARWELL SCIENCE AND INNOVATION CAMPUS, DIDCOT, OX11 0RA, UK. EMAIL: ANDREW.MCCLUSKEY@BRISTOL.AC.UK

*THE OPTIMAL SCHEME TO ESTIMATE THE DIFFUSION COEFFICIENT REQUIRES **BOTH** VARIANCE AND COVARIANCE TO BE CONSIDERED.*

THE MEAN-SQUARED DISPLACEMENT OVER A GIVEN NUMBER OF STEPS IS NORMALLY DISTRIBUTED...

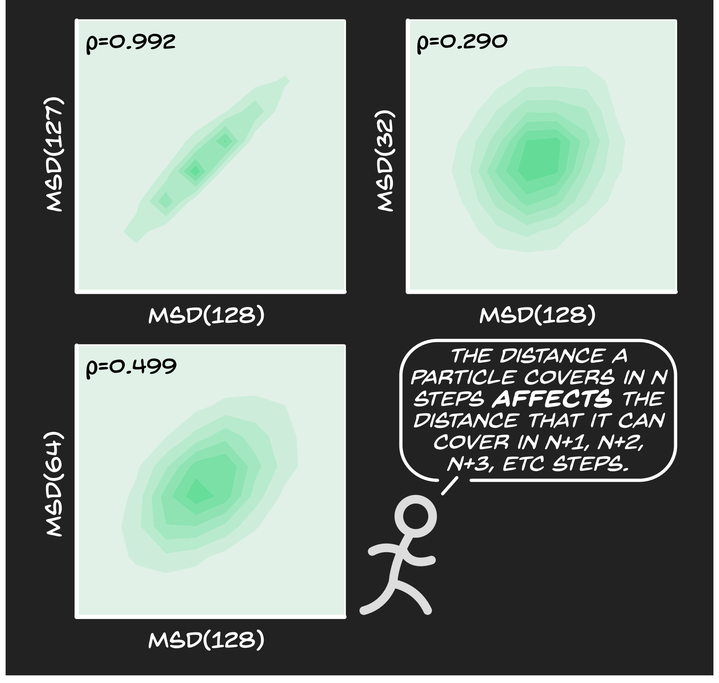


...**HOWEVER**, AS THE NUMBER OF STEPS INCREASES SO DOES THE VARIANCE*...



*CALLED HETEROSCEDASTICITY

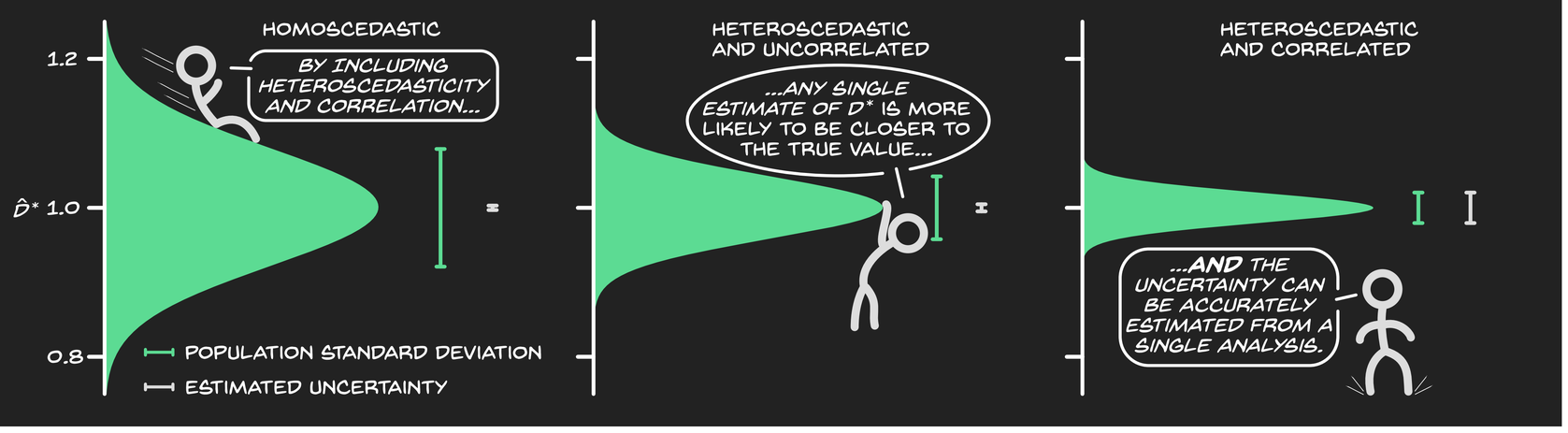
...**AND** THE DISPLACEMENTS ARE CORRELATED.



*A **COVARIANT MULTI-DIMENSIONAL NORMAL DISTRIBUTION** CAN BE USED AS A GENERATIVE MODEL FOR MEAN-SQUARED DISPLACEMENT...*

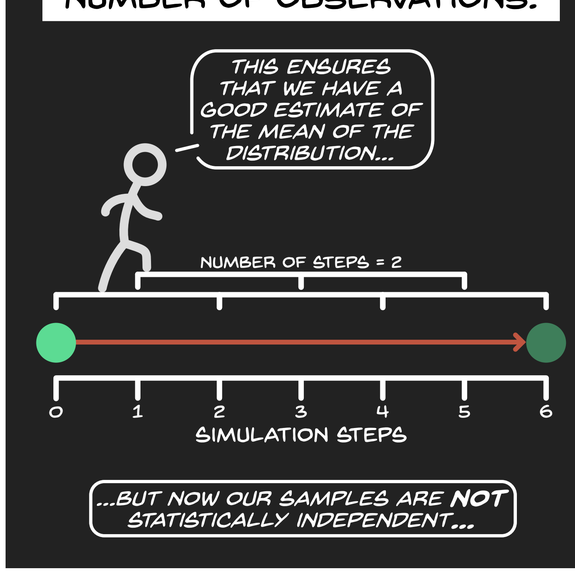
...WHERE EACH DIMENSION IS A DIFFERENT NUMBER OF STEPS OR TIME-INTERVAL.

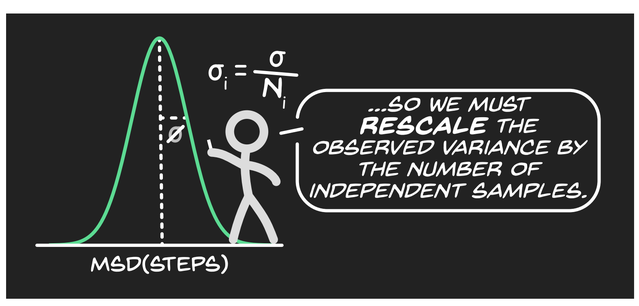
HETEROSCEDASTICITY AND CORRELATION MUST BE ACCOUNTED FOR TO EFFICIENTLY ESTIMATE THE DIFFUSION COEFFICIENT AND ACCURATELY ESTIMATE THE UNCERTAINTY.



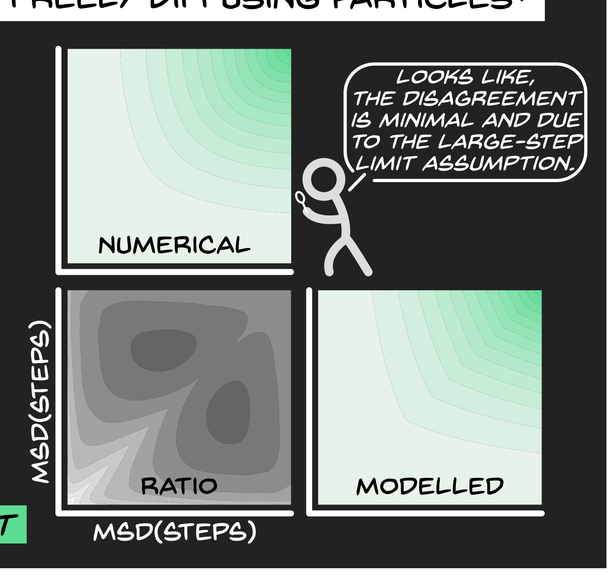
*DATA FROM **MOLECULAR DYNAMICS SIMULATION** CAN BE USED TO PARAMETERISE THE MODEL.*

OVERLAPPING SAMPLES ARE DRAWN TO MAXIMISE THE NUMBER OF OBSERVATIONS.





THE FULL COVARIANCE MATRIX CAN BE MODELLED BASED ON FREELY DIFFUSING PARTICLES*



AVAILABLE NOW! OPEN-SOURCE!

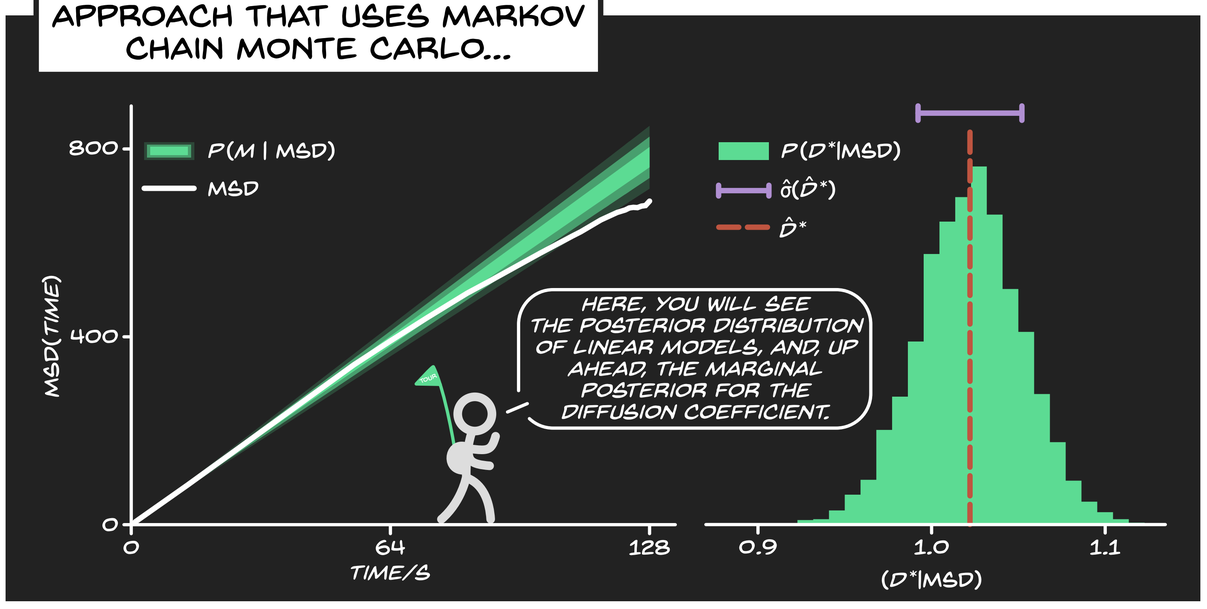
WITH SUPPORT FOR PYMATGEN, ASE, AND MDANALYSIS OBJECTS

kinisi

pip install kinisi

*WE CAN SAMPLE THE PARAMETERISED DISTRIBUTION LIKELIHOOD WITH A HEAVISIDE PRIOR TO GET AN ESTIMATE OF THE DIFFUSION COEFFICIENT DISTRIBUTION, WHICH WOULD TYPICALLY REQUIRE **MANY 1000s OF SIMULATIONS**.*

THIS IS A BAYESIAN SAMPLING APPROACH THAT USES MARKOV CHAIN MONTE CARLO...



...**AND** USING THIS METHOD, ON 4096 INDIVIDUAL SIMULATIONS, GIVES AN UNBIASED, EFFICIENT ESTIMATE OF D^* WITH GOOD VARIANCE ESTIMATE.