Coarse-Grained Parameterizations of Biomolecular Systems

Peter R. Kramer
Department of Mathematical Sciences
Rensselaer Polytechnic Institute. Troy, NY
Supported by NSF CAREER Grant DMS-0449717
Overview

- **Homogenization Theory** for Effective Transport Properties of Brownian Motors

- **Stochastic Immersed Boundary Method** for Microbiological Simulation

- **Stochastic Drift-Diffusion Parameterization of Water Dynamics near Solute**
Biological Disclaimer

(www.moleculararium.com, S. Garde et al)
Molecular Motors

Biological engines capable of transforming thermal energy, through chemically activated processes, into mechanical work.

• Molecular pumps.
• Walking motors: Kinesin, Myosin.
• Flagellar motors.
• Polymer Growth.
Molecular Motors

Biological engines capable of transforming thermal energy, through chemically activated processes, into mechanical work.

- Molecular pumps.
- Walking motors: Kinesin, Myosin.
- Flagellar motors.
- Polymer Growth.

http://multimedia.mcb.harvard.edu
Brownian Motors as Mathematical Model

\[ dX = U \, dt, \]
\[ mdU = (-\gamma U - \nabla V(X(t), F(t)) + Y(t)) \, dt + \sqrt{2\gamma k_B T} \, dW(t), \]
Brownian Motors as Mathematical Model

\[ dX = U \, dt, \]
\[ m \, dU = (-\gamma U - \nabla V(X(t), F(t)) + Y(t)) \, dt + \sqrt{2\gamma k_B T} \, dW(t), \]

- System represented by low-dimensional state variable \( X(t) \)
- Spatially periodic potential \( V \) modeling repetitive structure or cycle
- External driving (chemical activation) represented through stochastic processes \( Y(t) \) and \( F(t) \)
- Interaction with environment: friction coefficient \( \gamma \) and thermal fluctuations
Brownian Motors as Mathematical Model

\[ dX = U \, dt, \]

\[ m \, dU = (-\gamma U - \nabla V(X(t), F(t)) + Y(t)) \, dt + \sqrt{2\gamma k_B T} \, dW(t), \]

Potential with spatial period \( L: V(\mathbf{x} + L \hat{e}_i, \cdot) = V(\mathbf{x}, \cdot), \) for coordinate vectors \( \hat{e}_i. \)

\( dW(t) \) represents Gaussian white noise forcing with

\[ \langle dW(t) \rangle = 0, \]

\[ \langle dW_i(t) dW_j(s) \rangle = \delta_{ij} \delta(t - s) dt \, ds. \]
Brownian Motors as Mathematical Model

\[ \text{d}X = U \, \text{d}t, \]
\[ m \text{d}U = (-\gamma U - \nabla V(X(t), F(t)) + Y(t)) \, \text{d}t + \sqrt{2\gamma k_B T} \, \text{d}W(t), \]

Often overdamped limit adequate:

\[ \text{d}X(t) = \gamma^{-1} (-\nabla V(X(t), F(t)) + Y(t)) \, \text{d}t + \sqrt{2D} \, \text{d}W(t). \]

with \( D = k_B T / \gamma \). Remove \( \gamma \) by appropriate rescaling of parameters.
Gaussian Long-Time Statistics from Central Limit Argument

Effective transport properties characterized entirely by drift vector and diffusion matrix:

\[ U_{\text{eff}} = \lim_{t \to \infty} \frac{\langle X(t) \rangle}{t}, \]

\[ D_{\text{eff}} = \lim_{t \to \infty} \frac{\langle (X(t) - \langle X(t) \rangle) \otimes (X(t) - \langle X(t) \rangle) \rangle}{2t}. \]
Homogenization Theory Provides Means for Calculating Drift and Diffusion Coefficients

Potential advantages over existing theoretical calculations

- Does not rely on one-dimensionality or special structure
- Provides means of calculating diffusion coefficient
- Sets up further calculations in various asymptotic parameter limits


- Originally one-dimensional with discrete states; two-dimensional generalization (2007)
Our Homogenization Studies So Far

With Juan Latorre and Grigorios Pavliotis (Imperial)

- Connection between Homogenization and WPE equations.
- Systematic application of homogenization theory on simple flashing ratchet example
  - Previous work with homogenization (Pavliotis 2005) and WPE (2003) considered steady potentials with constant mean force $F$
  - First passage time formulation (Reimann et al, 2002)
- Numerical methods to apply Homogenization Theory to more complex models
Homogenization Theory Derivation (Pavliotis, 2005)

The process \( \{X(t), F(t)\} \) is characterized by the Backward-Kolmogorov equation:

\[
\frac{\partial u(x, f, t)}{\partial t} = \mathcal{L}u(\cdot) = \left( -\nabla V(x, f) \cdot \nabla x + D \Delta x + \mathcal{L}_f \right) u(\cdot).
\]

\( \mathcal{L}_f \) is the infinitesimal generator associated to the stochastic process \( F(t) \); can depend on \( X \). We have omitted the process \( Y(t) \) for now.
Homogenization Theory Derivation (Pavliotis, 2005)

The process \( \{X(t), F(t)\} \) is characterized by the Backward-Kolmogorov equation:

\[
\frac{\partial u(x, f, t)}{\partial t} = \mathcal{L}u(\cdot) \\
\equiv \left( -\nabla V(x, f) \cdot \nabla x + D\Delta x + \mathcal{L}_f \right)u(\cdot).
\]

\( \mathcal{L}_f \) is the infinitesimal generator associated to the stochastic process \( F(t) \); can depend on \( X \). We have omitted the process \( Y(t) \) for now.

**Long-Time limit analysis.** Rescale time and space with respect to a parameter \( \epsilon \ll 1 \).
Effective Drift Computation

**Advective** large-scale, long-time rescaling:

\[ \tilde{x} = \epsilon x, \]
\[ \tilde{t} = \epsilon t. \]

The rescaled Backward-Kolmogorov equation is:

\[
\frac{\partial u^\epsilon(\cdot)}{\partial \tilde{t}} = \frac{1}{\epsilon} \left( - \nabla V\left(\frac{\tilde{x}}{\epsilon}, f\right) \cdot \nabla \tilde{x} + D \Delta \tilde{x} + \mathcal{L}_f \right) u^\epsilon(\cdot),
\]
Effective Drift Computation

Look for a two-scale solution using singular perturbation theory:

\[ u^\varepsilon(\tilde{x}, f, \tilde{t}) = u_0(\tilde{x}, \frac{\tilde{x}}{\varepsilon}, f, \tilde{t}) + \varepsilon u_1(\tilde{x}, \frac{\tilde{x}}{\varepsilon}, f, \tilde{t}) + \ldots \]
Effective Drift Computation

Look for a two-scale solution using singular perturbation theory:

\[ u^\epsilon(\tilde{x}, f, \tilde{t}) = u_0(\tilde{x}, \frac{\tilde{x}}{\epsilon}, f, \tilde{t}) + \epsilon u_1(\tilde{x}, \frac{\tilde{x}}{\epsilon}, f, \tilde{t}) + \ldots \]

The solvability condition at first order gives a transport equation for \( u_0 \):

\[ \frac{\partial u_0(\tilde{x}, \tilde{t})}{\partial \tilde{t}} = U_{\text{eff}} \cdot \nabla u_0(\tilde{x}, \tilde{t}), \]

with drift velocity

\[ U_{\text{eff}} = \int_Q \int_{E_f} -\nabla V(z, f) \rho(z, f) \, dz \, df, \]

\[ \mathcal{L}_0^* \rho(z, f) = -\nabla_z ( -\nabla V(z, f) ) + D \Delta_z \rho + \mathcal{L}_f^* \rho = 0, \]

with periodic boundary conditions in \( z \) and \( \rho \in L^1 \).
**Effective Diffusivity Computation**

**Diffusive** large-scale, long-time rescaling:

\[ \tilde{x} = \epsilon (x - U_{\text{eff}} \frac{t}{\epsilon^2}), \]
\[ \tilde{t} = \epsilon^2 t. \]
Effective Diffusivity Computation

Look for a two-scale solution

\[ u^\varepsilon(\tilde{x}, f, \tilde{t}) = u_0(\tilde{x}, \frac{\tilde{x}}{\varepsilon}, \tilde{t}) + \varepsilon u_1(\tilde{x}, \frac{\tilde{x}}{\varepsilon}, f, \tilde{t}) + \ldots \]
Effective Diffusivity Computation

Look for a two-scale solution

\[ u^\varepsilon(\tilde{x}, f, \tilde{t}) = u_0(\tilde{x}, \frac{\tilde{x}}{\varepsilon}, \tilde{t}) + \varepsilon u_1(\tilde{x}, \frac{\tilde{x}}{\varepsilon}, f, \tilde{t}) + \ldots \]

The solvability condition at second order gives a diffusion equation for \( u_0 \):

\[ \frac{\partial u_0(\tilde{x}, \tilde{t})}{\partial \tilde{t}} = D_{\text{eff}} : \nabla \nabla u_0(\tilde{x}, \tilde{t}), \]

with diffusivity matrix:

\[ D_{\text{eff},ij} = D \delta_{ij} + \langle [\nabla V(z, f)]_i \chi_j(z, f, y) \rangle_\rho + 2D \langle \frac{\partial \chi_i}{\partial z_j} \rangle_\rho. \]

The auxiliary field \( \chi(z, f) \) is the solution to the first order equation and satisfies:

\[ -\mathcal{L}_0 \chi = -\nabla V(z, f) - U_{\text{eff}}. \]
Wang-Peskin-Elston Approach

• Discretize Brownian motor dynamics as Markov chain, preserving detailed balance structure.
Wang-Peskin-Elston Approach

- **Discretize** Brownian motor dynamics as Markov chain, preserving detailed balance structure.
- **Matrix equation for effective drift** is clear discretization of homogenization formula.
Wang-Peskin-Elston Approach

- Discretize Brownian motor dynamics as Markov chain, preserving detailed balance structure.
- Matrix equation for effective drift is clear discretization of homogenization formula.
- Matrix equation for effective diffusivity is not direct discretization of homogenization formula, but can be obtained by working with adjoint problem.
**Flashing Ratchet Models**

**Amplitude** of potential fluctuates between two random values \( \{f_1, f_2\} \).

One spatial dimension:

\[
dX(t) = -V'(X(t))F(t)dt + \sqrt{2D}dW(t),
\]

Two-state example: \( F(t) \in \{f_1, f_2\} \) with transition rates \( k_{ij} \) from state \( i \to j \). Infinitesimal generator \( \mathcal{L}_f \) is matrix composed of transition rates.
**Flashing Ratchet Homogenization Equations**

Stationary **Fokker-Planck Equation** is PDE system \((f\) dependence in subscript)

\[
\begin{align*}
\partial_x (V'(x) f_1 \rho_1(x)) + D \partial_{xx} \rho_1(x) - k_{12} \rho_1(x) + k_{21} \rho_2(x) &= 0 \\
\partial_x (V'(x) f_2 \rho_2(x)) + D \partial_{xx} \rho_2(x) + k_{12} \rho_1(x) - k_{21} \rho_2(x) &= 0.
\end{align*}
\]

with

\[
\begin{align*}
\rho_i(x + L) &= \rho_i(x) \\
V'(x) f_i \rho_i(x + L) + D \partial_x \rho_i(x + L) &= V'(x) f_i \rho_i(x) + D \partial_x \rho_i(x), \\
i &= 1, 2
\end{align*}
\]

\[
\int_0^L (\rho_1(x) + \rho_2(x)) \, dx = 1.
\]

**Cell problem for \(\chi\) has similar structure.**
Flashing Ratchet with a Piecewise Linear Potential
Amplitude switches randomly between $A_1 = A f_1$ in state 1 (in blue) and $A_2 = A f_2$ when in state 2 (in red).
Flashing Ratchet with a Piecewise Linear Potential

Homogenization equations take form of ODE system with piecewise constant coefficients; solvable through numerical linear algebra.
On-Off Flashing Ratchet

\[ dX(t) = -V'(X(t))F(t)dt + \sqrt{2D}dW(t), \]

with \( F(t) \in \{1, 0\} \).
On-Off Flashing Ratchet

\[ dX(t) = -V'(X(t))F(t)dt + \sqrt{2D}dW(t), \]

with \( F(t) \in \{1, 0\} \).

Sample trajectory
On-Off Flashing Ratchet

\[ \frac{dX(t)}{dt} = -V'(X(t))F(t)dt + \sqrt{2D}dW(t), \]

with \( F(t) \in \{1, 0\} \).

Explore how design parameters influence:

- drift \( U_{\text{eff}} \)
- Peclét number: measure of coherence of transport

\[ Pe = \frac{LU_{\text{eff}}}{2D_{\text{eff}}}. \]
On-Off Flashing Ratchet

\[ \text{d}X(t) = -V'(X(t))F(t)\text{d}t + \sqrt{2D}\text{d}W(t), \]

with \( F(t) \in \{1, 0\} \).

Nondimensionalize/rescale: \( L = 1, A = 1 \).

Remaining parameters: \( D, k_{ij}, \alpha \).
Effective Drift As Function of Transition Rates

\[ D = 10^{-1.5}, \alpha = 0.75 \]
Peclét Number as Function of Transition Rates

$D = 10^{-1.5}, \alpha = 0.75$
Observations from Numerics and Analysis

- Drift-optimizing transition rates $k_{12}$ and $k_{21}$ have different scaling with respect to $D$ than those optimizing $Pe$. 
Observations from Numerics and Analysis

• Drift-optimizing transition rates $k_{12}$ and $k_{21}$ have different scaling with respect to $D$ than those optimizing $Pe$.

• Intermediate asymptotic regime where $Pe \approx 1$: well-approximated by continuous-time random walk (CTRW) with strong bias.
Observations from Numerics and Analysis

- Drift-optimizing transition rates $k_{12}$ and $k_{21}$ have different scaling with respect to $D$ than those optimizing $Pe$.
- Intermediate asymptotic regime where $Pe \approx 1$: well-approximated by continuous-time random walk (CTRW) with strong bias.
- Coarse-grained CTRW model inconsistent over parameter subdomain (where $Pe > 1$).
Observations from Numerics and Analysis

- Drift-optimizing transition rates $k_{12}$ and $k_{21}$ have different scaling with respect to $D$ than those optimizing $Pe$.
- Intermediate asymptotic regime where $Pe \approx 1$: well-approximated by continuous-time random walk (CTRW) with strong bias.
- Coarse-grained CTRW model inconsistent over parameter subdomain (where $Pe > 1$).
- Development of explicit formulas for CTRW rates in various asymptotic regimes.
Flashing ratchet, multiplicative Gaussian noise

\[\begin{align*}
\frac{dX}{dt} &= -V'(X(t))Y(t) + \sqrt{2D}dW(t), \\
\frac{dY}{dt} &= -\frac{1}{\tau}Y(t) + \sqrt{\frac{2\sigma^2}{\tau}}dW_2(t).
\end{align*}\]

Amplitude \(Y(t)\) is Ornstein-Uhlenbeck process with variance \(\sigma^2\) and correlation time \(\tau\).
Flashing ratchet, multiplicative Gaussian noise

\[ \begin{align*}
\mathrm{d}X &= -V'(X(t))Y(t) + \sqrt{2D}\mathrm{d}W(t), \\
\mathrm{d}Y &= -\frac{1}{\tau}Y(t) + \sqrt{\frac{2\sigma^2}{\tau}}\mathrm{d}W_2(t).
\end{align*} \]

Amplitude \( Y(t) \) is Ornstein-Uhlenbeck process with variance \( \sigma^2 \) and correlation time \( \tau \).

First step toward incorporating continuous variability in chemical “states” (i.e., interference with nearby structures)
Flashing ratchet, multiplicative Gaussian noise

\[ \text{d}X = -V'(X(t))Y(t) + \sqrt{2D}\text{d}W(t), \]
\[ \text{d}Y = -\frac{1}{\tau}Y(t) + \sqrt{\frac{2\sigma^2}{\tau}}\text{d}W_2(t). \]

Amplitude \( Y(t) \) is Ornstein-Uhlenbeck process with variance \( \sigma^2 \) and correlation time \( \tau \).

Stationary Fokker-Planck equation for \( \rho \in L^1([0, L] \times \mathbb{R}) \), periodic in \( x \):

\[ \partial_x (V'(x)y\rho(x, y)) + D\partial_{xx}\rho(x, y) \]
\[ + \frac{1}{\tau} (\partial_y (y\rho(x, y)) + \sigma\partial_{yy}\rho(x, y)) = 0, \]
Flashing ratchet, multiplicative Gaussian noise

\[ dX = -V'(X(t))Y(t) + \sqrt{2D}dW(t), \]
\[ dY = -\frac{1}{\tau}Y(t) + \sqrt{\frac{2\sigma^2}{\tau}}dW_2(t). \]

Amplitude \( Y(t) \) is Ornstein-Uhlenbeck process with variance \( \sigma^2 \) and correlation time \( \tau \).

Solve numerically through spectral decomposition with respect to orthonormal Hermite polynomials \( H_n(y) \) and Fourier expansion in \( x \) (similar to continued-fraction approach in Risken, *Fokker-Planck Equation*).
Numerical Results. $V'(x) = \sin(2\pi x)$

Effective Diffusivity, $D=0.1, \sigma^2 = \tau^{-1}$

- Spectral Algorithm
- Monte Carlo
- Discrete OU, NS=21
- Dichotomous noise
Ongoing and Future Work

- Extension to multiple degrees of freedom
  - motor and cargo
  - particle sorters
- Applications to more realistic models
- “Quenched” random modulations of potential
- Numerical efficiency studies for computational schemes (homogenization vs. WPE)
Stochastic Immersed Boundary Method

Charles Peskin developed Immersed Boundary Method as efficient simulation of fluid systems with immersed structures.

- philosophy is to treat heterogenous system as a single fluid
- material properties of immersed structures reflected in forces applied to Navier-Stokes equations

Originally developed for macrobiology (blood flow through heart), more recently extended (with Paul Atzberger (UCSB) and Peskin (NYU)) with a view toward microbiology applications with important thermal and fluid components

- molecular motors
- bacterial locomotion including osmotic gel swelling effects
- exocytosis of vesicles from cells
Stochastic Immersed Equations I

\[ \rho \left( \frac{\partial u(x, t)}{\partial t} + u(x, t) \cdot \nabla u(x, t) \right) = \mu \Delta u(x, t) - \nabla p(x, t) + f(x, t) + f_T(x, t), \]
\[ \nabla \cdot u(x, t) = 0, \]

Navier-Stokes equations:

- \( u(x, t) \) is fluid velocity,
- \( \rho \) is fluid density (often assumed constant),
- \( \mu \) is dynamic viscosity,
- \( f(x, t) \) is structural force density acting on fluid.
- \( f_T(x, t) \) is thermal force density.
Stochastic Immersed Boundary Equations II

Immersed particles induce force on fluid:

\[ f(\mathbf{x}, t) = -\sum_{j=1}^{N} \nabla_j \Phi(\{X^{[j']}(t)\}) \delta_a(\mathbf{x} - X^{[j]}(t)) \]

where:

- \( N \) is the number of particles,
- \( \Phi \) is the total potential on the particles due to interaction and/or external influences.
- \( X^{[j]}(t) \) is location of particle \( j \) at time \( t \).
- \( \delta_a(\cdot) \) is function which spreads force over region with linear dimension \( a \) (representing effective particle size).

Immersed membranes and fibers handled similarly as...

\[ \text{Coarse-Grained Parameterizations of Biomolecular Systems – p. 25/4} \]
Stochastic Immersed Boundary Equations III

Thermal force density from fluctuation-dissipation theorem:

\[
f_T(\mathbf{x}, t) = \sum_k f_{T,k}(t)e^{2\pi i k \cdot \mathbf{x}/L},
\]

\[
f_{T,k}(t) = \sqrt{\frac{4\pi^2 k^2 \mu k_B T}{L^5}} \frac{d}{dt} \tilde{\mathbf{W}}_k(t)
\]

- \(L\) is length scale of domain.
- \(T\) is absolute temperature; \(k_B\) is Boltzmann’s constant.
- independent complex Brownian processes \(\{\tilde{\mathbf{W}}_k(t)\}_k\):

\[
\langle d\tilde{\mathbf{W}}_k(t) \otimes d\tilde{\mathbf{W}}_k(t') \rangle = 0,
\]

\[
\langle d\tilde{\mathbf{W}}_k(t) \otimes d\tilde{\mathbf{W}}_k^*(t') \rangle = 2\mathcal{I} \delta(t - t') \, dt \, dt',
\]
Stochastic Immersed Boundary Equations IV

The immersed particles move at a velocity obtained by interpolating the local fluid velocity over length scale $a$:

$$\frac{dX[j](t)}{dt} = u_a(X[j](t), t),$$

$$u_a(\mathbf{x}, t) = \int_{\Omega} u(\mathbf{x}', t) \delta_a(\mathbf{x} - \mathbf{x}') \, d\mathbf{x'},$$

where $\Omega$ is the spatial domain and the interpolation function $\delta_a$ is the same function as was used to spread force. (Important for energy conservation properties).
Discretization

Velocity and pressure defined on periodic fluid lattice $h\mathbb{Z}_K^3$, with

- $\mathbb{Z}_K^3 \equiv [1, 2, \ldots, K]^3$, $K = L/h$.

- Atzberger is developing adaptive version.

Elementary particle positions $X^{[j]}$ anywhere in continuous period domain.

Time discretization allows underresolution of fast fluid modes.
Motor Velocity-Force Relation
Numerical simulations

Latest version of code due to Atzberger

• Pre-release version available on request from http://www.cims.nyu.edu/~paulatz/computational_projects/projects.html

• Polymer knot and kinesin model simulations

• developing simulations for osmotic pumps, membranes, Golgi apparatus

Undergraduate research by Sam Hughes on microswimmer models.
Simplified statistical description of water dynamics as possible basis for implicit solvent method to accelerate molecular dynamics simulations for proteins, etc. (with Adnan Khan and Shekhar Garde (Biochemical Engineering))

As a first step, we explore stochastic parameterization of water near $C_{60}$ buckyball molecule.

- isotropic, chemically simple
Molecular Dynamics Snapshot of Buckyball Surrounded by Water
Statistical dynamics encoded in biophysical literature in terms of a diffusion coefficient:

(Makarov et al, 1998; Lounnas et al, 1994, . . .)

\[
D_B(r) \equiv \left\langle \left| \frac{X(t + 2\tau) - X(t)}{6\tau} \right|^2 \right| \begin{array}{c}
X(t) = r \\
- \left\langle \left| \frac{X(t + \tau) - X(t)}{6\tau} \right|^2 \right| X(t) = r \end{array} \right\rangle
\]

But this seems to mix together inhomogeneities in mean and random motion.
We explore capacity of models of the form

\[ dX = U(X(t)) \, dt + D(X(t)) \, dW(t), \]

for water molecule center-of-mass position \( X(t) \).

- drift vector coefficient \( U(r) \)
- diffusion tensor coefficient \( D(r) \)

For isometric solute (buckyball):

- \( U(r) = U_{\parallel}(|r|) \hat{r} \)
- \( D(r) = D_{\parallel}(|r|) \hat{r} \otimes \hat{r} + D_{\perp}(|r|)(I - \hat{r} \otimes \hat{r}) \)

for position \( r = |r| \hat{r} \) relative to center of symmetry.
Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

\[ U(\mathbf{r}) = -\gamma^{-1} \nabla \phi(\mathbf{r}), \]

\[ D(\mathbf{r}) = D_0 1. \]
Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

\[ U(r) = -\gamma^{-1} \nabla \phi(r), \]
\[ D(r) = D_0 l. \]

- Potential of mean force obtained from measuring concentration \( c(r) \) and Boltzmann distribution \( c(r) \propto \exp(-\phi(r)/k_BT) \).
- Diffusivity unchanged from bulk value.
- Friction coefficient from Einstein relation \( \gamma = k_BT/D_0 \).
Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

\[ U(r) = -\gamma^{-1} \nabla \phi(r), \]
\[ D(r) = D_0 l. \]
Physically Inspired (DD-I) Model

In analogy to Brownian dynamics simulations, take

\[ U(r) = -\gamma^{-1} \nabla \phi(r), \]
\[ D(r) = D_0 l. \]
Systematic, Data-Driven Parameterization (DD-II) Model

Parametrize drift and diffusion functions from mathematical definitions:

\[ U_{\parallel}(|r|) = \lim_{\tau \downarrow 0} \left\langle \frac{X(t+\tau) - X(t)}{\tau} \cdot \hat{r} \bigg| X(t) = r \right\rangle, \]

\[ D_{\parallel}(|r|) = \lim_{\tau \downarrow 0} \left\langle \frac{|X(t+\tau) - X(t) - U_{\parallel}(r)\hat{r}\tau|^2}{2\tau} \bigg| X(t) = r \right\rangle. \]
Systematic, Data-Driven Parameterization (DD-II) Model

Parametrize drift and diffusion functions from mathematical definitions:

$$D_{\perp}(|r|) = \lim_{\tau \downarrow 0} \left( \frac{1}{4\tau} \langle |X(t + \tau) - X(t)|^2 \rangle \right)^2 \left| X(t) = r \right\rangle.$$  

Obtain statistical data from MD simulations.
Time Difference $\tau$ must be chosen carefully

Taking $\tau = \Delta t$ (time step of MD simulation) may not be appropriate

- Limit $\tau \downarrow 0$ implicitly refers to times large enough for drift-diffusion approximation to be valid.

Must choose $T_v \ll \tau \ll T_x$, where:
- $T_v$ is time scale of momentum.
- $T_x$ is time scale of position.

See also Pavliotis and Stuart (2007) about need to undersample.
How choose $\tau$ in practice?
Explore simple Ornstein-Uhlenbeck (OU) model

\[ dX = V \, dt, \]

\[ m\, dV = -\gamma V \, dt - \alpha X \, dt + \sqrt{2k_B T \gamma} \, dW(t) \]
Explore simple Ornstein-Uhlenbeck (OU) model

\[
d\mathbf{X} = \mathbf{V} \, dt,
\]

\[
md\mathbf{V} = -\gamma \mathbf{V} \, dt - \alpha \mathbf{X} \, dt + \sqrt{2k_B T \gamma} \, d\mathbf{W}(t)
\]

Forces: friction, potential, and thermal.
Explore simple Ornstein-Uhlenbeck (OU) model

\[
d\mathbf{X} = \mathbf{V} \, dt, \\
md\mathbf{V} = -\gamma \mathbf{V} \, dt - \alpha \mathbf{X} \, dt + \sqrt{2k_B T \gamma} \, d\mathbf{W}(t)
\]

Nondimensionalize:

\[
d\mathbf{X} = \mathbf{V} \, dt, \\
d\mathbf{V} = -a \mathbf{V} \, dt - a \mathbf{X} \, dt + a \, d\mathbf{W}(t)
\]
Explore simple Ornstein-Uhlenbeck (OU) model

\[ dX = V \, dt, \]
\[ m\, dV = -\gamma V \, dt - \alpha X \, dt + \sqrt{2k_B T \gamma} \, dW(t) \]

Nondimensionalize:

\[ dX = V \, dt, \]
\[ dV = -a V \, dt - a X \, dt + a \, dW(t) \]

where \( a = \gamma^2 / (m\alpha) \) is ratio of position to momentum time scale.
Explore simple Ornstein-Uhlenbeck (OU) model

\[ d\mathbf{X} = \mathbf{V} \, dt, \]
\[ m d\mathbf{V} = -\gamma \mathbf{V} \, dt - \alpha \mathbf{X} \, dt + \sqrt{2k_B T \gamma} \, d\mathbf{W}(t) \]

Nondimensionalize:

\[ d\mathbf{X} = \mathbf{V} \, dt, \]
\[ d\mathbf{V} = -a \mathbf{V} \, dt - a \mathbf{X} \, dt + a \, d\mathbf{W}(t) \]

Exact drift-diffusion coarse-graining when \( a \gg 1 \).
Explore simple Ornstein-Uhlenbeck (OU) model

\[ dX = V \, dt, \]

\[ m dV = -\gamma V \, dt - \alpha X \, dt + \sqrt{2k_B T \gamma} \, dW(t) \]

Nondimensionalize:

\[ dX = V \, dt, \]

\[ dV = -a V \, dt - a X \, dt + a \, dW(t) \]

What if we try to obtain this from analysis of trajectories with finite but large \( a \)?
Drift and diffusion coefficients of exact OU solution sampled with finite time difference $\tau$.

Longitudinal Drift for OU, $a=132$

Blue : Exact
Red : Asymptotic
Drift and diffusion coefficients of exact OU solution sampled with finite time difference $\tau$. 

Longitudinal Diffusivity for V–OU, $a=132$

Blue : Exact  
Red  : Asymptotic
Inferences from OU model

- Good choice of $\tau$ may be the one which maximizes drift magnitude and diffusivity.
- Beginning estimate obtained from OU model with same $\alpha$ value.
To obtain time scales, approximate main well in potential of mean force by quadratic.
OU Model Insights → MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of $\tau$. 
OU Model Insights → MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of $\tau$.

![Longitudinal Drift Diagram]

- Blue: $\tau=20$ fs
- Green: $\tau=40$ fs
- Red: $\tau=60$ fs
- Magenta: $\tau=80$ fs
- Cyan: $\tau=100$ fs
OU Model Insights → MD Data Parameterization in DD-II Model

Examine drift and diffusivity computed from various choices of $\tau$. 

![Longitudinal Diffusivity Graph]

<table>
<thead>
<tr>
<th>Color</th>
<th>$\tau$ (fs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>60</td>
</tr>
<tr>
<td>Green</td>
<td>80</td>
</tr>
<tr>
<td>Magenta</td>
<td>120</td>
</tr>
<tr>
<td>Cyan</td>
<td>160</td>
</tr>
<tr>
<td>Black</td>
<td>200</td>
</tr>
<tr>
<td>Yellow</td>
<td>240</td>
</tr>
<tr>
<td>Red</td>
<td>280</td>
</tr>
</tbody>
</table>
OU Model Insights → MD Data Parameterization in DD-II Model

Examine **drift** and **diffusivity** computed from various choices of \( \tau \). Both desiderata about drift and diffusivity behavior not simultaneously satisfiable.

- Correct bulk diffusivity behavior more important

We choose \( \tau = 0.2 \, \text{ps} = 200 \, \text{fs} \).
Parameterization Used in DD-II Model

Longitudinal Drift

Blue: DD-II output
Red: DD-II input from MD data

U \parallel (\text{nm/ps})

r (\text{nm})

Coarse-Grained Parameterizations of Biomolecular Systems – p. 42/44
Parameterization Used in DD-II Model

Longitudinal Diffusivity

Blue: DD-II output
Red: DD-II input from MD Data

$D_{||}$ (nm$^2$/ps)

$r$ (nm)

$10^{-3}$
Parameterization Used in DD-II Model

![Graph showing lateral diffusivity vs. distance (r) in nanometers. The graph compares DD-II output (blue) with DD-II input (red) from MD data. The x-axis represents r (nm) ranging from 0.6 to 1.8, and the y-axis represents D⊥ (nm²/ps) ranging from 0 to 8 x 10⁻³.](image)
Compare Predictions of Biophysical Diffusivity Formula

\[ D_B(\mathbf{r}) \equiv \left\langle \left| \frac{\mathbf{X}(t + 2\tau) - \mathbf{X}(t)}{6\tau} \right|^2 \right| \mathbf{X}(t) = \mathbf{r} \right\rangle \\
- \left\langle \left| \frac{\mathbf{X}(t + \tau) - \mathbf{X}(t)}{6\tau} \right|^2 \right| \mathbf{X}(t) = \mathbf{r} \right\rangle \]
Compare Predictions of Biophysical Diffusivity Formula

\[
D_B(r) \text{ for DD–I Model}
\]

\[
D_B(r) \text{ for DD–II Model}
\]

Coarse-Grained Parameterizations of Biomolecular Systems – p. 43/44
Future Work

Next steps

• anisotropies

• chemical heterogeneity