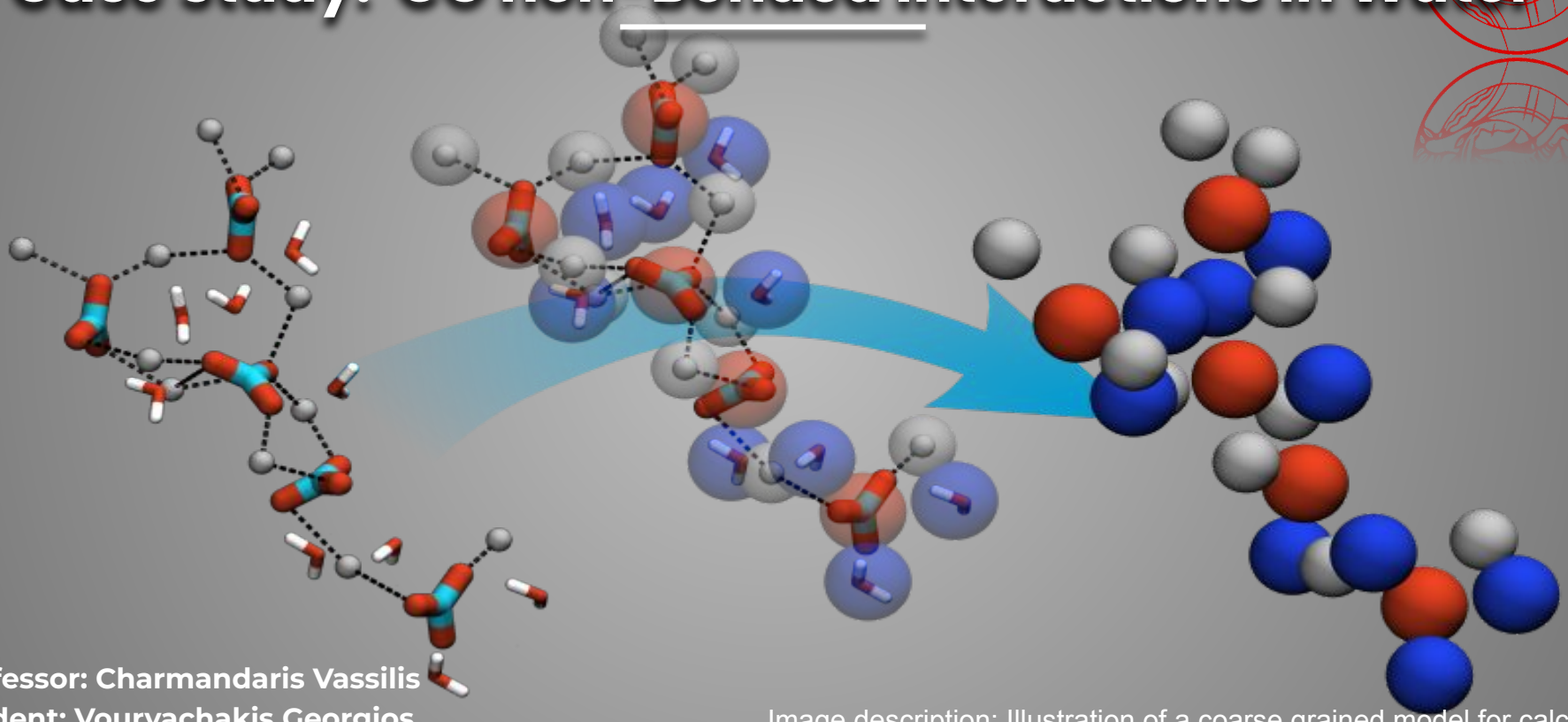
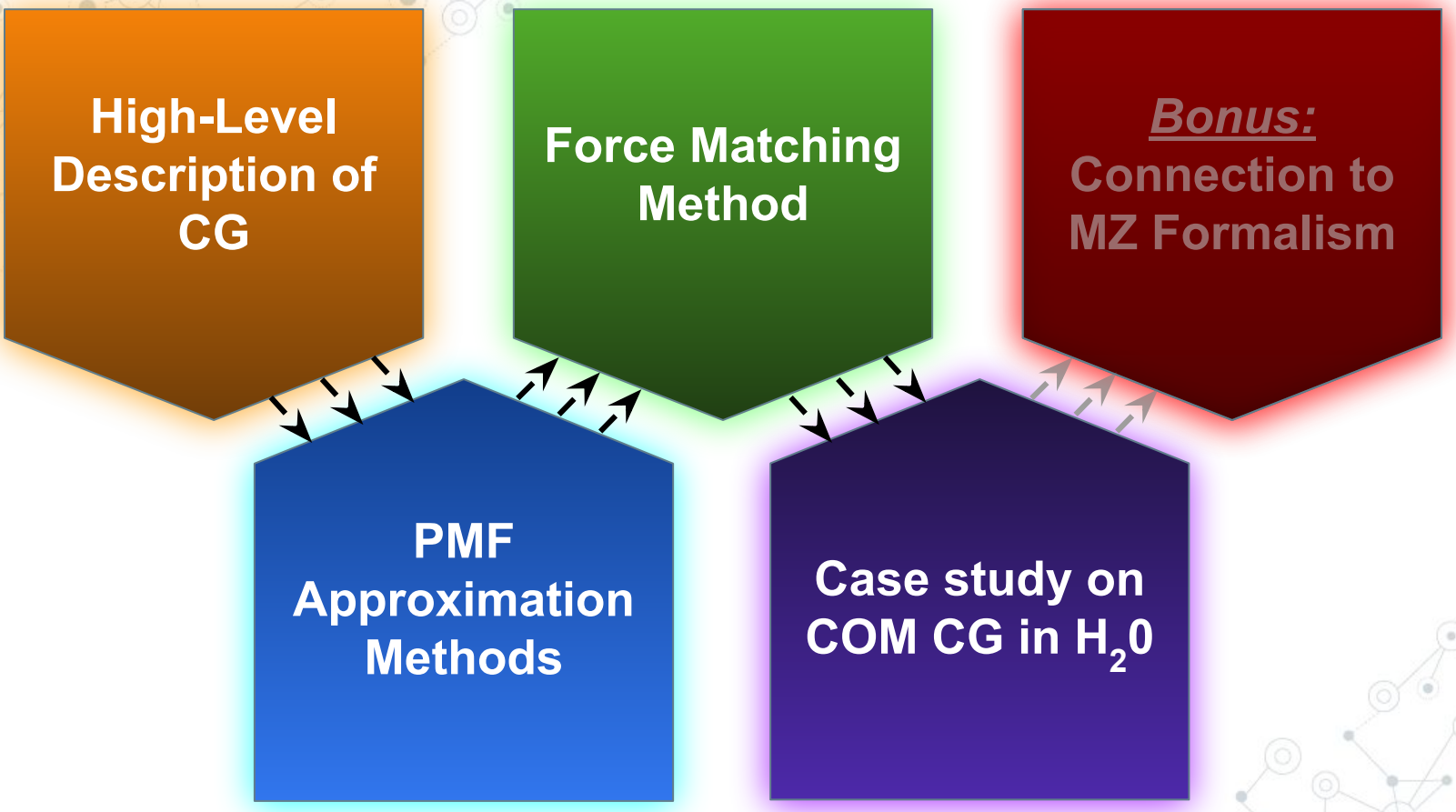
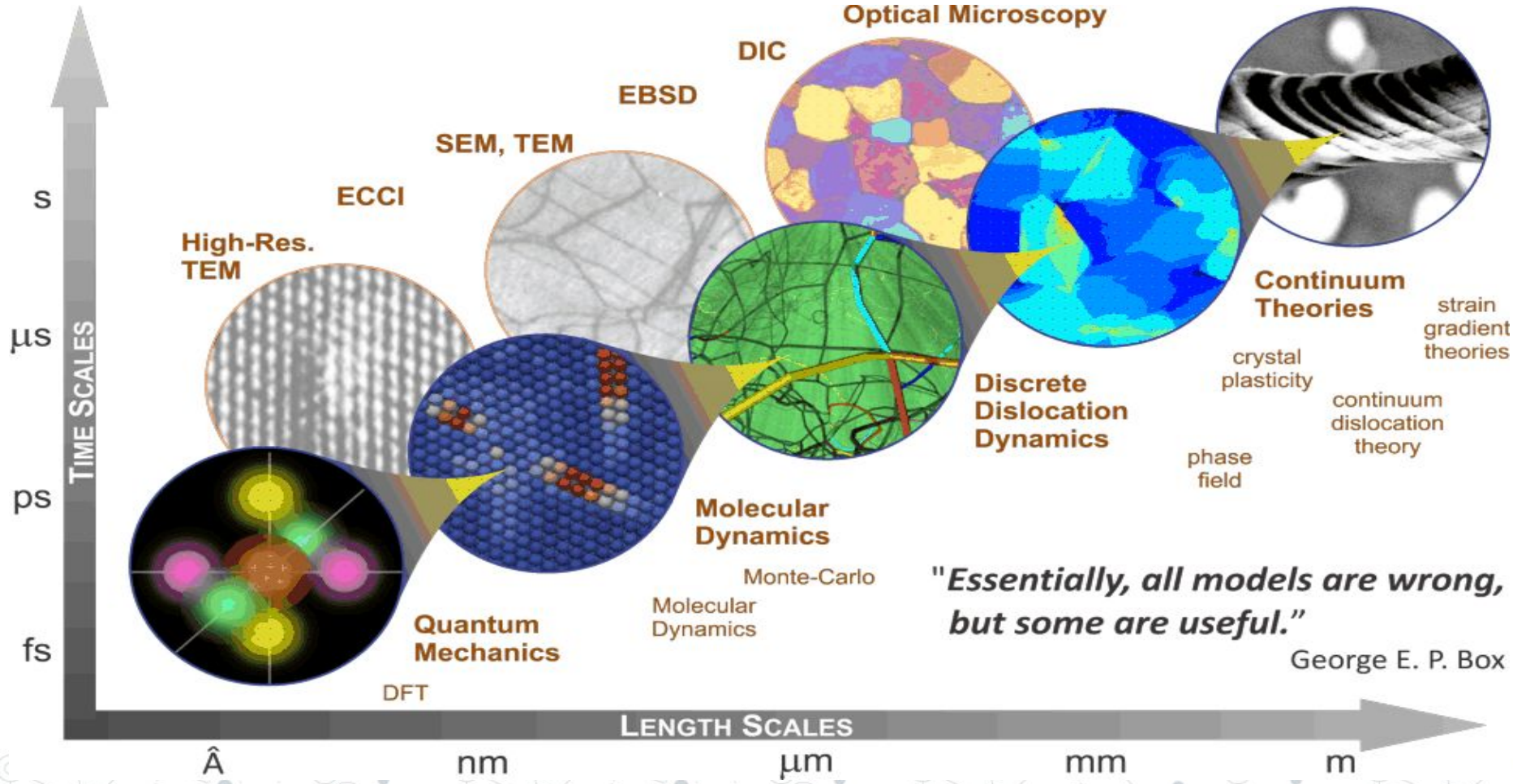
Coarse-Grained Molecular Dynamics
Case Study: CG non-Bonded Interactions in Water



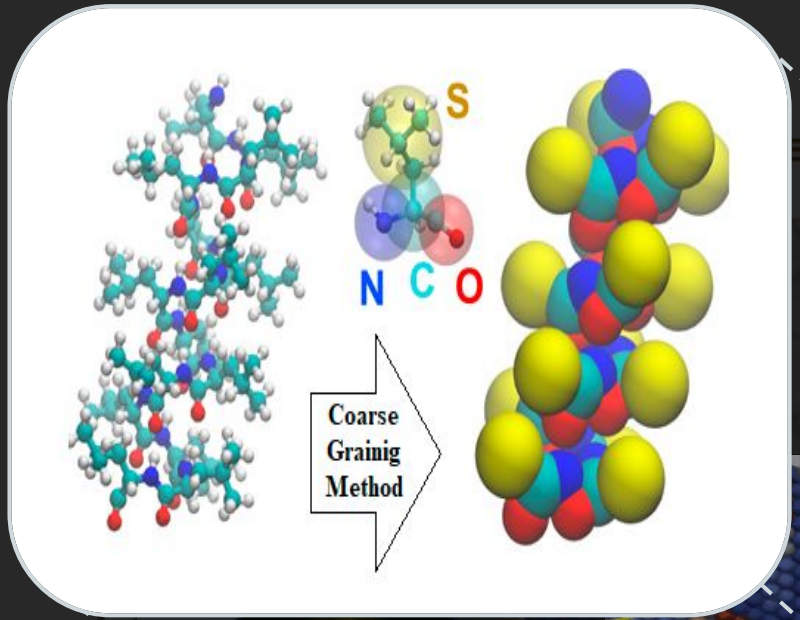
Professor: Charmandaris Vassilis
Student: Vourvachakis Georgios
Date: 27/05/2025

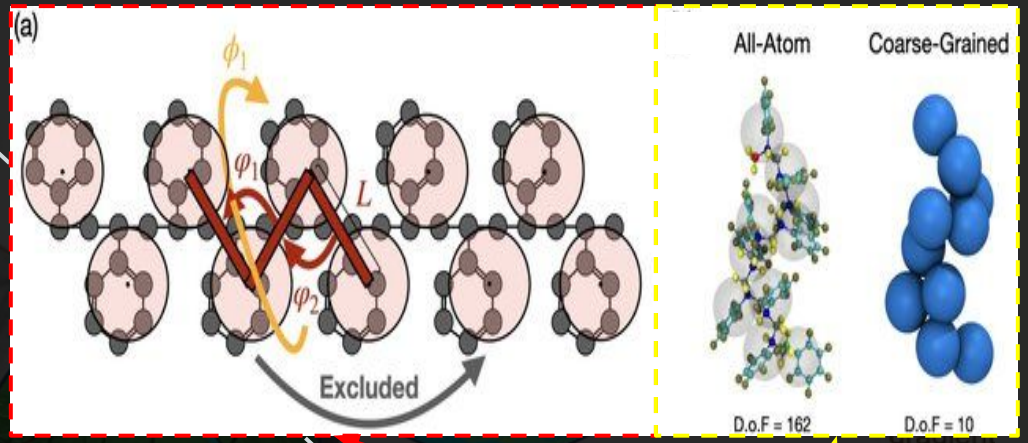
Image description: Illustration of a coarse grained model for calcium minerals [University of Konstanz, VDM package]


High-Level Description of CG
Force Matching Method
Bonus: Connection to MZ Formalism
PMF Approximation Methods
Case study on COM CG in H_20



[ETH Zurich, Mechanics and Materials Laboratory]





Bonded interactions separation

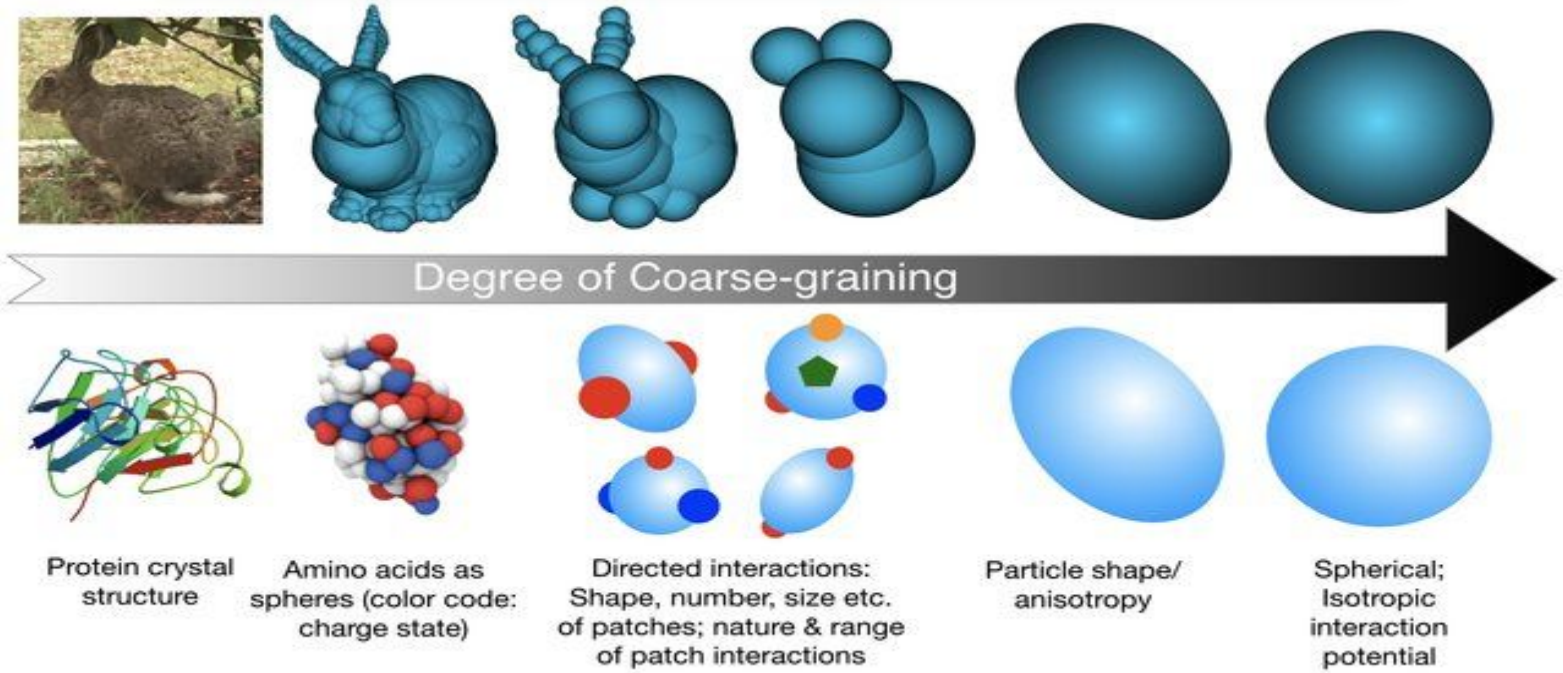
CG mapping between AA and CG model of a 10-monomer PS chain

Extend the **time** and **length scale** of AA and MD simulations to cover **biological scales**

[LAMMPS tube, Coarse Grained Modeling]

[Zhenghao Wu et al., JoCTC **20**(4) (2024)]

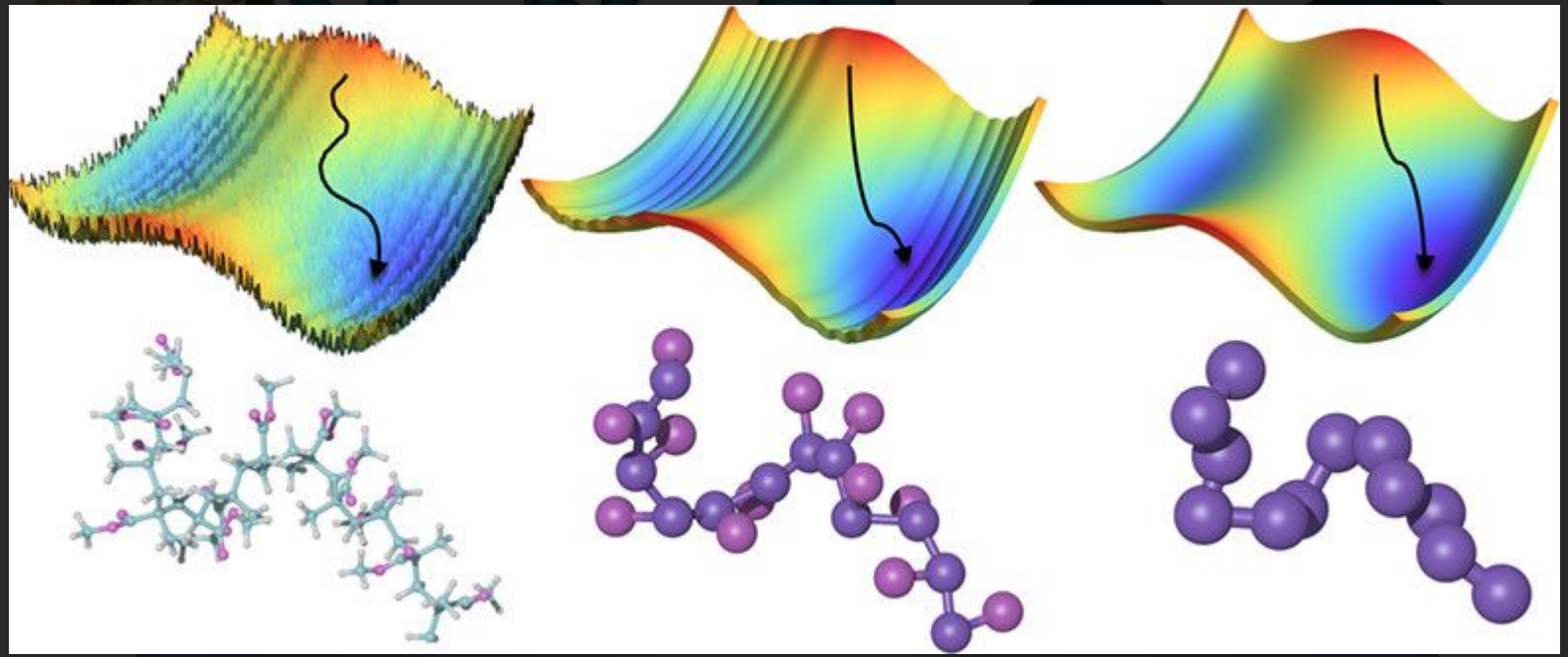
Optimal Coarse-Graining Strategy



CG approaches must be developed to simultaneously **reduce computational complexity** and **retain necessary chemical details** for large-scale systems, encompassing more than ~10^7 atoms and larger than ~100 nm in space

[Anna Stradner et al., Soft Matter **16**(2), 307-323 (2019)]

Optimal Coarse-Graining Strategy



CG approaches must be developed to simultaneously **reduce computational**

The dynamics of the system is ***artificially accelerated*** since the PES is smoothed

[Anna Stradner

[Rui Shi et al., WIREs CMS **13**(6) (2023)]

Paradigm of Coarse-Graining

- The governing equations of AA MD → CG:

$$\mathbf{Q}_i = \frac{\sum_{j \in \mathrm{CG}_i} m_j \mathbf{q}_j}{\sum_{j \in \mathrm{CG}_i} m_j (\equiv M_i)}$$

$$\dot{\mathbf{q}}_i = \mathbf{v}_i + \mathbf{f}_{\mathrm{baro}}$$
$$m_i \dot{\mathbf{v}}_i = -\nabla_{\mathbf{q}_i} U(\mathbf{q}) + \mathbf{f}_{\mathrm{thermo}}$$

$$\dot{\mathbf{Q}}_i = \mathbf{V}_i + \mathbf{C}$$
$$M_i \dot{\mathbf{V}}_i = -\nabla_{\mathbf{Q}_i} U_{\mathrm{CG}}(\mathbf{Q}) + \mathbf{D}$$

conservative CG force field

Henderson's theorem

Unique static structure of the system

- Under CG, auxiliary terms **C** and **D** become *stochastic* due to the absence of high-frequency dof (e.g., solvents << CG sites in dilute polymer solutions).

The purpose of coarse-graining is to find an **appropriate set of parameters** that describe U (≈U_{PMF}), **C**, and **D** from atomistic (AA) models, reproducing the same structures and dynamics generated from the fine-grained (DEM in our case study) EoM.

PMF Estimator(s) (discussed here): Force Matching (or MSCG)

resolved (CG)

unresolved (orthogonal)

EoM (C and D): Mori- Zwanzig projection operator formalism

PMF Description and Approximation

General objective of PMF approximators (FM, RE, structural-based): Propose a family of interaction potential functions $\bar{U}_{\text{eff}}(Q; \theta \in \Theta)$, and seek for the optimal $\bar{U}_{\text{eff}}(Q; \theta^*)$ that "best approximates" the PMF (canonical ensemble),

$$\bar{U}^{\text{PMF}}(\mathbf{Q}) = -\frac{1}{\beta} \log \int_{\Omega(\mathbf{Q})} e^{-\beta U(\mathbf{q})} d\mathbf{q}.$$

Equilibrium marginal of the full atomistic distribution

$Q \in \mathbb{R}^{3M}$ denotes the configuration of generalized coordinates in CG space with M(<N) particles and $\Omega(\mathbf{Q}) := \{\mathbf{q} \in \mathbb{R}^{3N} : \Pi(\mathbf{q}) = \mathbf{Q}\}$ntities on the CG space are annotated with the "-" symbol).

$\bar{Z} = \int e^{-\beta \bar{U}_{\text{eff}}(\mathbf{Q})} d\mathbf{Q}$ is the corresponding *partition function*.

- The many body PMF can be described as being composed by two-body, three-body, e.t.c., interactions (pairwise distance $R_{ij} = \|Q_i - Q_j\|$, $i, j = 1, \ldots, M$):
  (a two-body effective pair potential usually suffices, $\bar{U}_{\text{eff}}(\mathbf{Q}) = \sum_{i,j} u(R_{ij}) \approx \bar{U}^{\text{PMF}}(\mathbf{Q})$.)

$$\bar{U}^{\text{PMF}}(\mathbf{Q}) = \sum_{i,j} u_2(R_{ij}) + \sum_{i,j,k} u_3(R_{ij}, R_{ik}, R_{jk}) + \ldots$$

[E. Kalligiannaki et al., Eur. Phys. J. Special Topics **225**, 1347–1372 (2016)]

PMF Description and Approximation

- **General objective of PMF approximators (FM, RE, structural-based)**: Propose a family of interaction potential functions $\bar{U}_{eff}(Q; \theta \in \Theta)$, and seek for the optimal $\bar{U}_{eff}(Q; \theta^*)$ that "best approximates" the PMF (canonical ensemble), $\bar{U}^{\mathrm{PMF}}(\mathbf{Q}) = -\frac{1}{\beta}\log\int_{\Omega(\mathbf{Q})} e^{-\beta U(\mathbf{q})}d\mathbf{q}.$

  $Q \in \mathbb{R}^{3M}$ denotes the configuration of generalized coordinates in CG space with M(<N) particles (quantities on the CG space are annotated with the "-" symbol).

- We denote the *mean force* $\bar{F}^{\mathrm{PMF}} : \mathbb{R}^{3M} \to \mathbb{R}^{3M}$ corresponding to the PMF assuming it exists, by $\bar{F}_i^{\mathrm{PMF}}(\mathbf{Q}) = -\nabla_{Q_i}\bar{U}^{\mathrm{PMF}}(\mathbf{Q}),\ i = 1, \ldots, M.$

- We denote by $\bar{\mu}(d\mathbf{Q}) = \bar{Z}^{-1}\exp\{-\beta\bar{U}_{\mathrm{eff}}(\mathbf{Q})\}d\mathbf{Q}$ the *equilibrium probability measure* at the CG configurational space for the given CG potential function $\bar{U}_{eff}(Q; \theta)$, where $\bar{Z} = \int e^{-\beta\bar{U}_{\mathrm{eff}}(\mathbf{Q})}d\mathbf{Q}$ is the corresponding *partition function*.

- The many body PMF can be described as being composed by two-body, three-body, e.t.c., interactions (pairwise distance $R_{ij} = \|Q_i - Q_j\|$, i, j = 1, . . . , M):
  (a two-body effective pair potential usually suffices, $\bar{U}_{\mathrm{eff}}(\mathbf{Q}) = \sum_{i,j} u(R_{ij}) \approx \bar{U}^{\mathrm{PMF}}(\mathbf{Q}).$)

$$\bar{U}^{\mathrm{PMF}}(\mathbf{Q}) = \sum_{i,j} u_2(R_{ij}) + \sum_{i,j,k} u_3(R_{ij}, R_{ik}, R_{jk}) + \ldots$$

[E. Kalligiannaki et al., EPJST. **225**, 1347–1372 (2016)]

PMF Description and Approximation

- **General objective of PMF approximators (FM, RE, structural-based)**: Propose a family of interaction potential functions $\bar{U}_{eff}(Q; \theta \in \Theta)$, and seek for the optimal $\bar{U}_{eff}(Q; \theta^*)$ that "best approximates" the PMF (canonical ensemble), $\bar{U}^{\mathrm{PMF}}(\mathbf{Q}) = -\frac{1}{\beta} \log \int_{\Omega(\mathbf{Q})} e^{-\beta U(\mathbf{q})} d\mathbf{q}.$

  $Q \in \mathbb{R}^{3M}$ denotes the configuration of generalized coordinates in CG space with M(<N) particles (quantities on the CG space are annotated with the "-" symbol).

- We denote the *mean force* $\bar{F}^{\mathrm{PMF}} : \mathbb{R}^{3M} \to \mathbb{R}^{3M}$ corresponding to the PMF assuming it exists, by $\bar{F}_i^{\mathrm{PMF}}(\mathbf{Q}) = -\nabla_{Q_i} \bar{U}^{\mathrm{PMF}}(\mathbf{Q}),\ i = 1, \ldots, M.$

- We denote by $\bar{\mu}(d\mathbf{Q}) = \bar{Z}^{-1} \exp\{-\beta \bar{U}_{\mathrm{eff}}(\mathbf{Q})\} d\mathbf{Q}$ the *equilibrium probability measure* at the CG configurational space for the given CG potential function $\bar{U}_{eff}(Q; \theta)$, where $\bar{Z} = \int e^{-\beta \bar{U}_{\mathrm{eff}}(\mathbf{Q})} d\mathbf{Q}$ is the corresponding *partition function*.

- The many body PMF can be described as being composed by two-body, three-body, e.t.c., interactions (pairwise distance $R_{ij} = \|Q_i - Q_j\|$, i, j = 1, . . . , M):
  (a two-body effective pair potential usually suffices, $\bar{U}_{\mathrm{eff}}(\mathbf{Q}) = \sum_{i,j} u(R_{ij}) \approx \bar{U}^{\mathrm{PMF}}(\mathbf{Q}).$)

$$\bar{U}^{\mathrm{PMF}}(\mathbf{Q}) = \sum_{i,j} u_2(R_{ij}) + \sum_{i,j,k} u_3(R_{ij}, R_{ik}, R_{jk}) + \ldots$$

[E. Kalligiannaki et al., EPJST. **225**, 1347–1372 (2016)]

PMF Description and Approximation

- **General objective of PMF approximators (FM, RE, structural-based)**: Propose a family of interaction potential functions $\bar{U}_{\text{eff}}(Q;\theta \in \Theta)$, and seek for the optimal $\bar{U}_{\text{eff}}(Q;\theta^*)$ that "best approximates" the PMF (canonical ensemble), $\bar{U}^{\text{PMF}}(\mathbf{Q}) = -\frac{1}{\beta}\log\int_{\Omega(\mathbf{Q})} e^{-\beta U(\mathbf{q})}d\mathbf{q}.$

  $Q \in \mathbb{R}^{3M}$ denotes the configuration of generalized coordinates in CG space with M(<N) particles (quantities on the CG space are annotated with the "-" symbol).

- We denote the *mean force* $\bar{F}^{\text{PMF}} : \mathbb{R}^{3M} \to \mathbb{R}^{3M}$ corresponding to the PMF assuming it exists, by
$$\bar{F}_i^{\text{PMF}}(\mathbf{Q}) = -\nabla_{Q_i}\bar{U}^{\text{PMF}}(\mathbf{Q}),\ i = 1,\dots,M.$$

- We denote by $\bar{\mu}(d\mathbf{Q}) = \bar{Z}^{-1}\exp\{-\beta\bar{U}_{\text{eff}}(\mathbf{Q})\}d\mathbf{Q}$ the *equilibrium probability measure* at the CG configurational space for the given CG potential function $\bar{U}_{\text{eff}}(Q;\theta)$, where $\bar{Z} = \int e^{-\beta\bar{U}_{\text{eff}}(\mathbf{Q})}d\mathbf{Q}$ is the corresponding *partition function*.

- The many body PMF can be described as being composed by two-body, three-body, e.t.c., interactions (pairwise distance $R_{ij} = \|Q_i - Q_j\|$, i, j = 1, . . . , M):
  (a two-body effective pair potential usually suffices, $\bar{U}_{\text{eff}}(\mathbf{Q}) = \sum_{i,j} u(R_{ij}) \approx \bar{U}^{\text{PMF}}(\mathbf{Q}).$)

$$\bar{U}^{\text{PMF}}(\mathbf{Q}) = \sum_{i,j} u_2(R_{ij}) + \sum_{i,j,k} u_3(R_{ij}, R_{ik}, R_{jk}) + \dots$$

[E. Kalligiannaki et al., EPJST. **225**, 1347–1372 (2016)]

Force Matching: A Least Squares Problem

❑ Since the PMF bridges the gap between chemical-specific CG and AA models, evaluating the mean force directly provides a natural CG parametrization approach (Ercolessi and Adams (1994)). We construct the following mean *least-square minimization*:

$$\min_{\theta\in\Theta} \mathbb{E}_{\mu}\left[\|h(\mathbf{q}) - \bar{F}(\mathbf{\Pi}(\mathbf{q});\theta)\|^2\right]$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^{3M}$ and $\mathbb{E}_{\mu}[\cdot]$ averages with respect to the *Gibbs canonical (probability) measure* (*ensemble average*) $\mu(d\mathbf{q}) = Z^{-1}\exp\{-\beta U(\mathbf{q})\}d\mathbf{q}$, given a CG (contraction) mapping $\mathbf{\Pi}: \mathbb{R}^{3N} \mapsto \mathbb{R}^{3M}$ s.t. $\mathbf{q} \mapsto \mathbf{\Pi}(\mathbf{q}) \in \mathbb{R}^{3M}$ on the microscopic state space.

➢ The reference field $h(\mathbf{q}) \in \mathbb{R}^{3M}$ is the local mean force where each component, $h_i(\mathbf{q})$, is the force exerted at the i^{th} CG site that is a function of the microscopic forces.

● One can interpret FM as a geometrical projection of the atomic forces (Voth et al. (2008)). Namely, the $\bar{F}^{\mathrm{PMF}}(\mathbf{Q})$ is a projection of a local mean force $h(\mathbf{q})$ onto the space of square integrable functions w.r.t $\mu(d\mathbf{q})$:

$$L^2(\mu;\mathbf{\Pi}) = \{F \in L^2(\mu) \mid \exists\, \bar{F}: \mathbb{R}^{3M} \mapsto \mathbb{R}^{3M} \ s.t.\ F(\mathbf{q}) = \bar{F}(\mathbf{\Pi}(\mathbf{q}))\}$$

Force Matching: A Least Squares Problem

- Since the PMF bridges the gap between chemical-specific CG and AA models, evaluating the mean force directly provides a natural CG parametrization approach (Ercolessi and Adams (1994)). We construct the following mean *least-square minimization*:

$$\min_{\theta\in\Theta} \mathbb{E}_\mu \left[\|h(\mathbf{q}) - \bar{F}(\mathbf{\Pi}(\mathbf{q});\theta)\|^2\right]$$

[V. Harmandaris et al., *J. Chem. Phys.* **143**(8), 084105 (2015)]

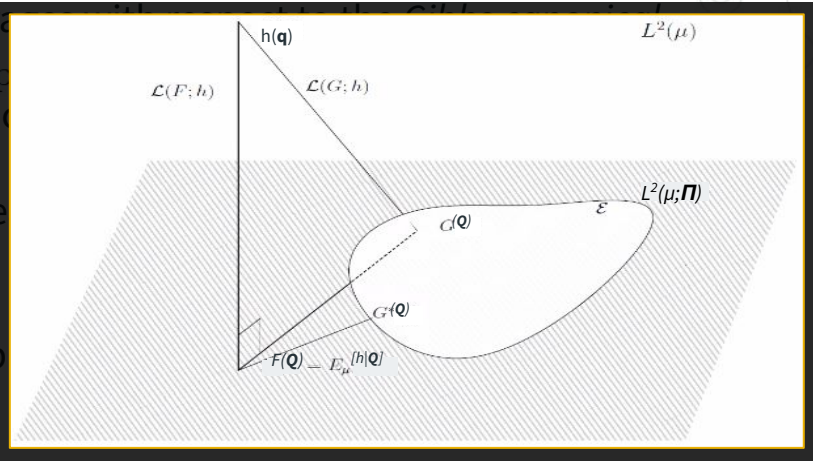
For a given $h \in L^2(\mu)$ the minimization problem

$$\inf_G \mathcal{L}(G;h) = \inf_G \mathbb{E}_\mu \left[\|h - G\circ\mathbf{\Pi}\|^2\right],$$

where inf is taken over all $G \in L^2(\mu;\mathbf{\Pi})$ has the unique solution

$$F^*(\mathbf{Q}) = \mathbb{E}_\mu[\mathbf{h}|\mathbf{Q}], \quad \mathbf{Q}\in\mathbb{R}^{3M}.$$

$F^{PMF}(\mathbf{Q}) = \mathbb{E}_\mu[\mathbf{h}|\mathbf{Q}]$, for the chosen h



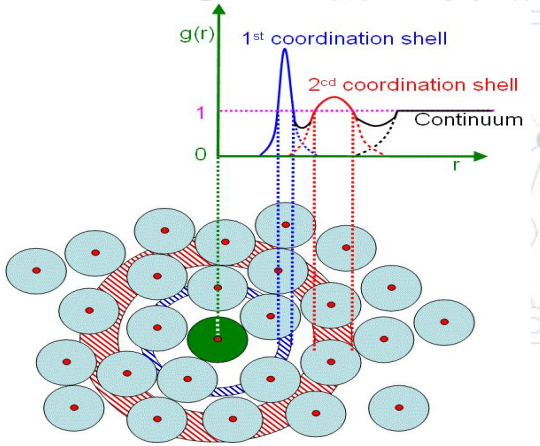
- One can interpret FM as a geometrical projection of the atomic forces (Voth et al. (2008)). Namely, the $\bar{F}^{\mathrm{PMF}}(\mathbf{Q})$ is a projection of a local mean force h($\mathbf{q}$) onto the space of square integrable functions w.r.t $\mu(d\mathbf{q})$: $L^2(\mu;\mathbf{\Pi}) = \{F \in L^2(\mu) \mid \exists\, \bar{F} : \mathbb{R}^{3M} \mapsto \mathbb{R}^{3M} \ s.t.\ F(\mathbf{q}) = \bar{F}(\mathbf{\Pi}(\mathbf{q}))\}$

Case Study on Water Molecules: CG procedure

- After identifying each molecule's atom indices {j∈I} (CG_i=I), the CG position $\mathbf{Q}_I$ is computed as the mass-weighted CoM:

$$\mathbf{Q}_I = \frac{\sum_{j\in I} m_j \mathbf{q}_j}{M_I} \text{ where } m_O = 15.9994,\ m_H = 1.008 \text{ u}$$

This mapping eliminates ***internal vibrational dof***.



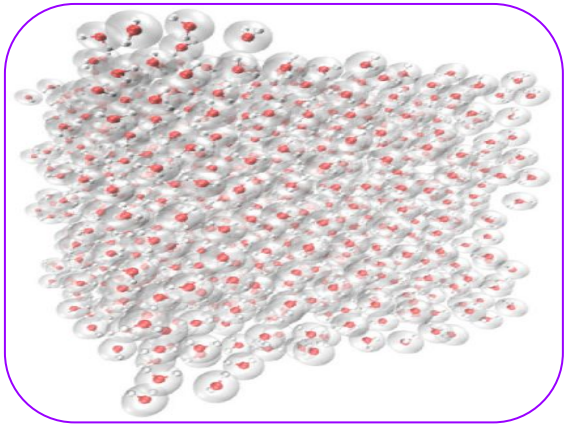
- To ensure consistent dynamics, atomic forces $\mathbf{f}_j$ are projected to CG sites by the same mass weights:

$$\mathbf{F}_I = \frac{\sum_{j\in I} m_j \mathbf{f}_j}{M_I}$$

- Pair correlation or Radial Distribution Function (RDF) (CG_i=I):

$$g(r) = (4\pi r^2 \Delta r \rho M)^{-1} \sum_{I<J} \delta^{(3)}(r - r_{IJ})$$

- Complexity and KD-Tree Acceleration:
  - **Brute-Force Cost:** Naive pairwise check costs $O(N_H N_O)$.
  - **KD-Tree:** Tiling positions to handle PBC and indexing with a cKDTree reduces the neighbor-search cost to $O(N_O \log N_H)$ (Friedman et al. (1977)).



[E. Kalligiannaki et al., EPJST. **225**, 1347–1372 (2016)]

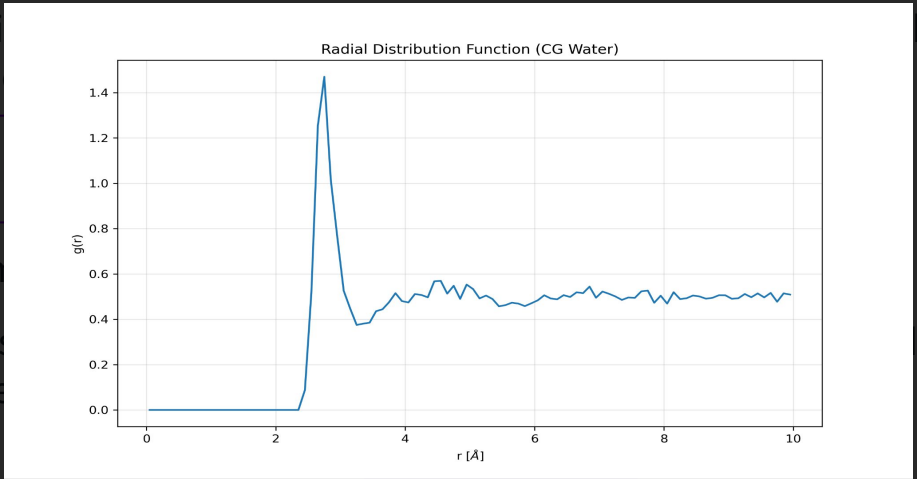
Case Study on Water Molecules: CG procedure

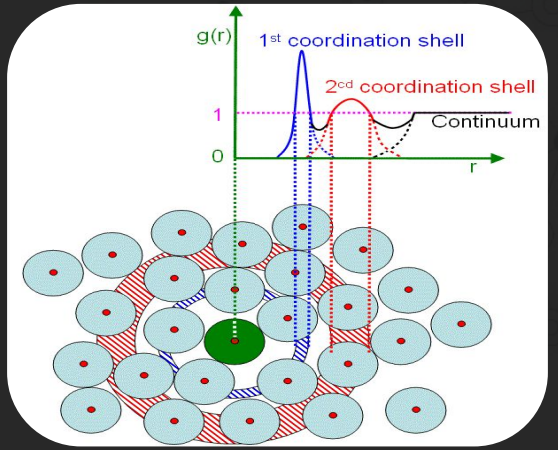
- After i … he CG positi …

$$Q_I = \dots$$

This n …

- To ens … to CG sites by the …





- Pair correlation or Radial Distribution Function (RDF) (CG_i=I):

$$g(r) = (4\pi r^2 \Delta r \rho M)^{-1} \sum_{I<J} \delta^{(3)}(r - r_{IJ})$$

- Complexity and KD-Tree Acceleration:
  - **Brute-Force Cost:** Naive pairwise checks costs $O(N_H N_O)$.
  - **KD-Tree:** Tiling positions to handle PBC and indexing with a cKDTree reduces the neighbor-search cost to $O(N_O \log N_H)$ (Friedman et al. (1977)).

[E. Kalligiannaki et al., EPJST. **225**, 1347–1372, (2016)]

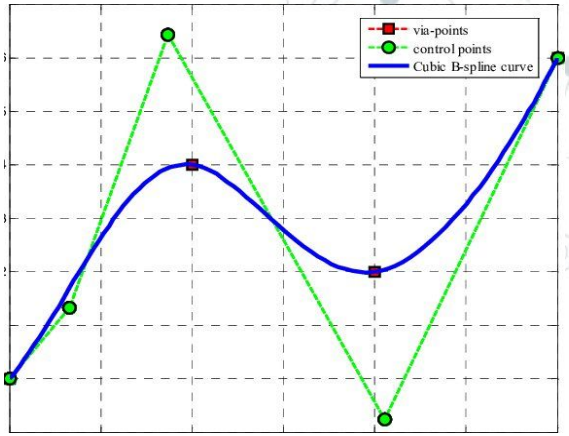
Case Study on Water Molecules: Cubic Spline Representation

- Each CG pair interaction $U_{IJ}(R)$ is modeled as a cubic spline through K knot points $\{(R_k,\theta_k)\}$, such that

$$U_{IJ}(R) := \sum_{k=1}^{K} \theta_k B_k(R)$$

where $B_k(R)$ are the **cubic B-spline basis functions** with

local support between adjacent knots.

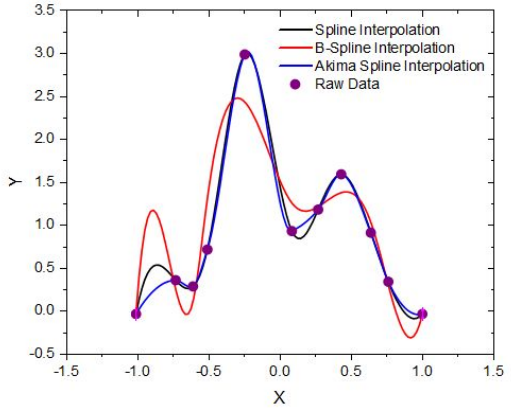


- The force between sites I and J follows from the negative derivative:

$$\bar{F}_{IJ}(R) = -U'_{IJ}(R) = -\sum_{k=1}^{K} \theta_k B'_k(R)$$

and the vector force on site I is

$$\bar{\mathbf{F}}_I \approx -\nabla_I U_{eff}(\mathbf{Q};\theta) = \sum_{J\neq I} \bar{F}_{IJ}(R_{IJ})\hat{\mathbf{R}}_{IJ}\ ,\ \hat{\mathbf{R}}_{IJ} = \frac{\mathbf{Q_I} - \mathbf{Q}_J}{\|\mathbf{Q_I} - \mathbf{Q}_J\|(\equiv R_{IJ})}$$



[C. Shuai et al., IEEE ICIA, 634-639 (2014)]

Case Study on Water Molecules: FM Optimization

- Minimizing the discrepancy between CG forces $\bar{\mathbf{F}}_I$ derived from a parametric potential and reference atomistic forces $\mathbf{h}_I$ mapped onto the same CG sites. The *mean-squared error* objective

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{h}) = \frac{1}{n_{\text{conf}}} \sum_{k=1}^{n_{\text{conf}}} \sum_{I=1}^{M} \| h(\mathbf{q}_k \in I) - \bar{\mathbf{F}}_I(\mathbf{Q}_k; \boldsymbol{\theta}) \|^2$$

- Optimization employs deterministic (L-BFGS-B) or stochastic (Adam) minimizers, using either full batches or random subsets of configurations to compute gradients via finite differences.

- For large datasets ($n_{conf} \gg 1000$), the algorithm performs **mini-batch** updates:
  1. Randomly sample subset of configurations,
  2. Compute batch MSE and approximate gradient $\nabla_{\boldsymbol{\theta}} \mathcal{L}$ via central differences (δ=1e-6),
  3. Update parameters **θ** by Adam's updating rule:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_{\boldsymbol{\theta}}\mathcal{L}, \; v_t = \beta_2 v_{t-1} + (1-\beta_2)(\nabla_{\boldsymbol{\theta}}\mathcal{L})^2,$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \; \hat{v}_t = \frac{v_t}{1-\beta_2^t}, \; \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha \frac{\hat{m}_t}{\sqrt{v_t} + \epsilon}$$

Tested Hyperparameters: **α=1e-2, 1e-3**, decay=0.98, β_1=0.9, β_2=0.999, ε=1e-8, K=20, frames-per-step = 100

Case Study on Water Molecules: Initial Guess Scheme

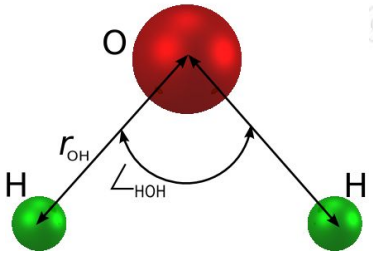
1. **Lennard-Jones–Like Potential:** Naive LJ–like formula with a *constant repulsive core* for R<σ:

$$U(R_i) = \begin{cases} \epsilon & R_i < \sigma \\ \epsilon\left[\left(\frac{\sigma}{R_i}\right)^{12} - 2\left(\frac{\sigma}{R_i}\right)^{6}\right] & R_i \geq \sigma \end{cases}$$

where ε=0.1 kcal/mol is the energy scale, σ=3.0 Å is the location of the potential minimum, $R_i \in \{R_1, R_2, \ldots, R_K\}$ are the spline knot positions. This gives the initial valu $\theta_i^{(0)} = U(R_i)\ , \forall i \in [K]$

2. **SPC/E-Inspired Initialization:** The potential is defined piecewise as

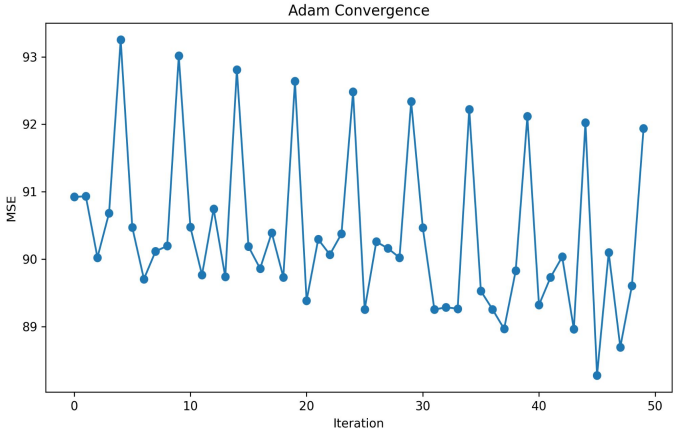
$$U(R_i) = \begin{cases} 100\left(\frac{\sigma}{R_i}\right)^{12} & R_i < R_{\text{cut,hard}} \\ 4\epsilon\left[\left(\frac{\sigma}{R_i}\right)^{12} - \left(\frac{\sigma}{R_i}\right)^{6}\right] & R_{\text{cut,hard}} \leq R_i \leq R_{\text{cut}} \end{cases}$$

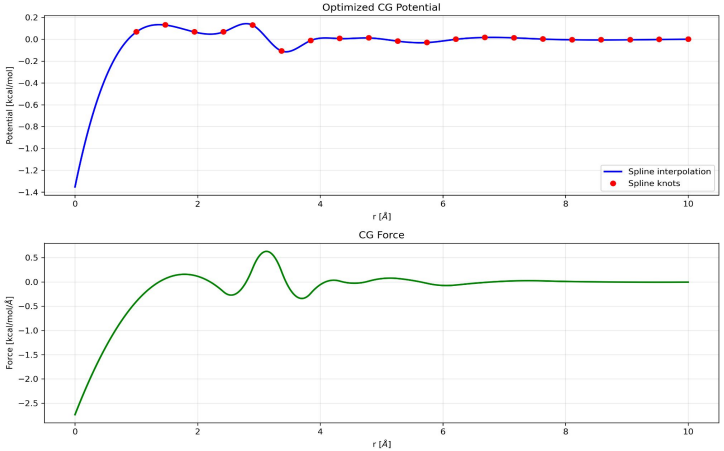


where (parameters from the SPC/E water model) ε = 0.65 kcal/mol, σ = 3.166 Å, $R_{cut,hard}$ = 2.5 Å (transition from hard core to LJ form), Rcut = 10 Å. Then *Savitzky-Golay smoothed* to remove oscillations and shifted s.t. $\theta_K^{(0)} = 0$.

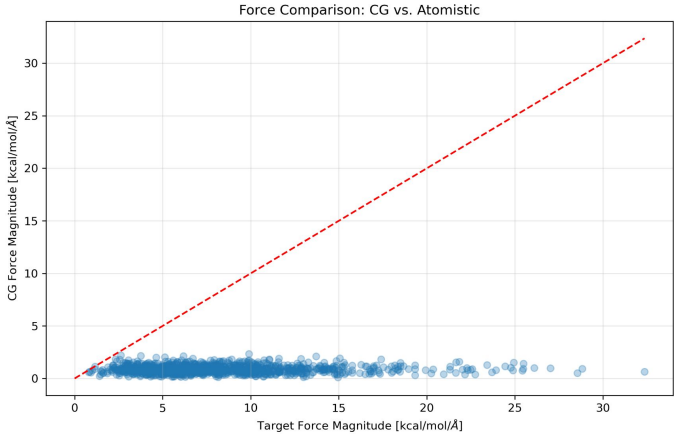
Initializing spline parameters: a good “starting potential” can accelerate convergence and avoid unphysical wells or barriers during optimization.

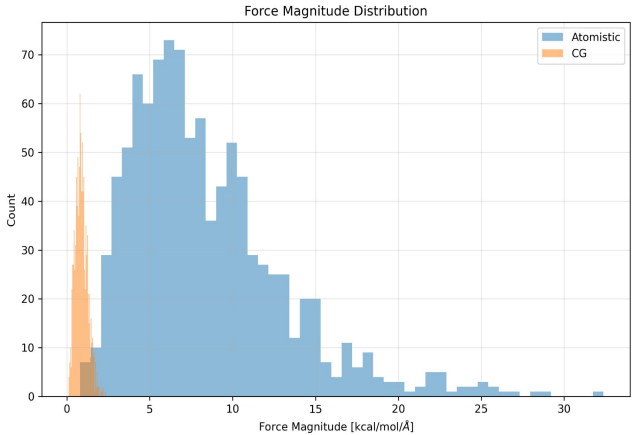
Results on Naive Initialization with α=1e-3





A single-site CG bead per water molecule cannot capture directional hydrogen-bond forces or intramolecular stiffness, hence the low variance in CG forces.





The inherent limitations of a *purely pairwise*, *single-site* model manifest as significant <u>underestimation of force variability</u> and a high MSE

MSE: 89.13 kcal/mol/Å
Best Val Loss: 91.94

Results on Naive Initialization with α=1e-3

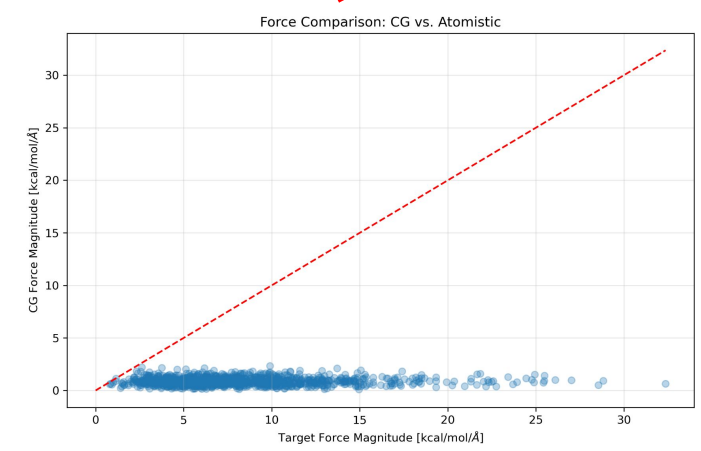
Upshot: A **pairwise spline** basis cannot represent these many-body contributions, which manifest as large local forces when bonds form or break

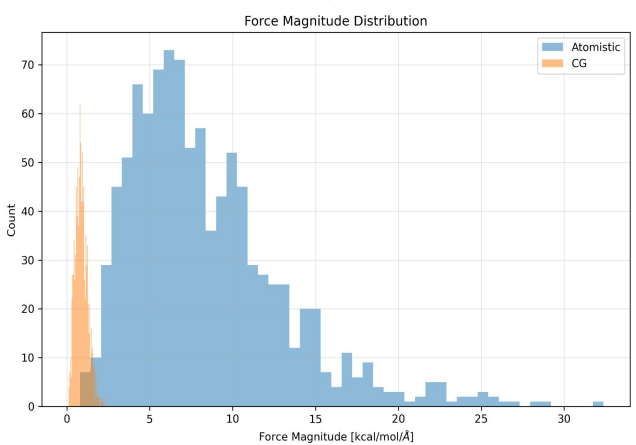
FM fits only the Markovian projected force; it omits the fluctuating term, causing underestimation of force variance

A single-site CG bead per water molecule cannot capture directional hydrogen-bond forces or intramolecular stiffness, hence the low variance in CG forces.

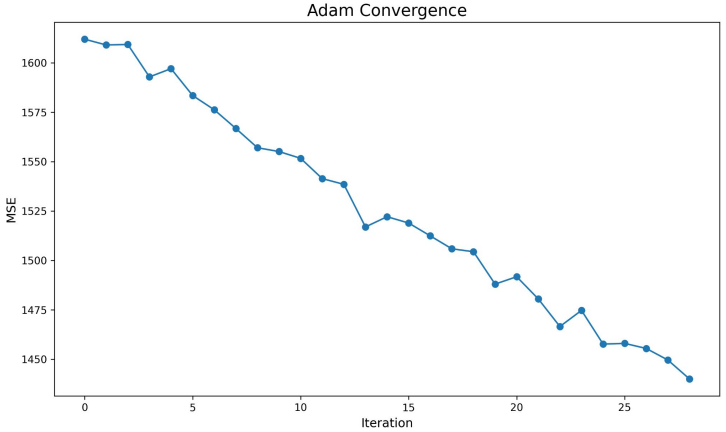
This is a direct consequence of a *purely pairwise*, *single-site* model manifest as significant underestimation of force variability and a high RMSE

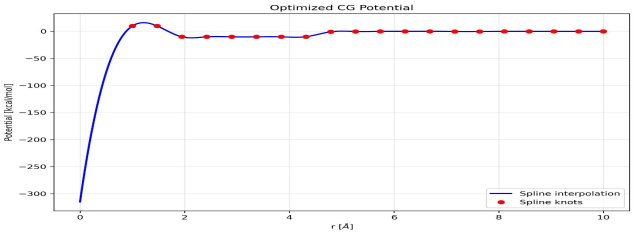
MSE: 89.13 kcal/mol/Å
Best Val Loss: 91.94

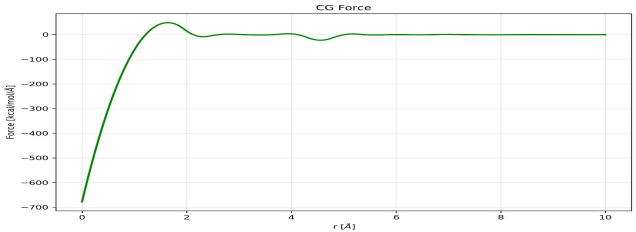


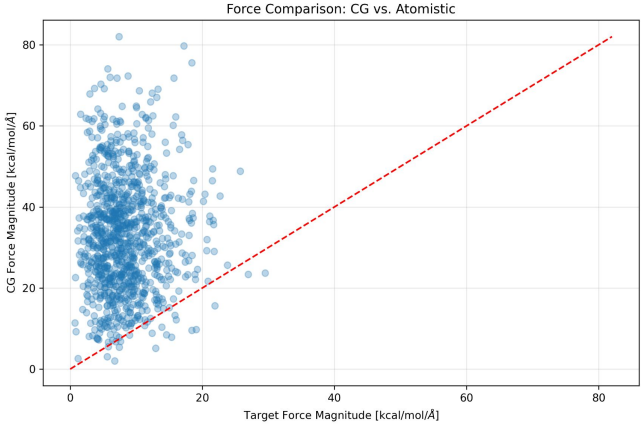


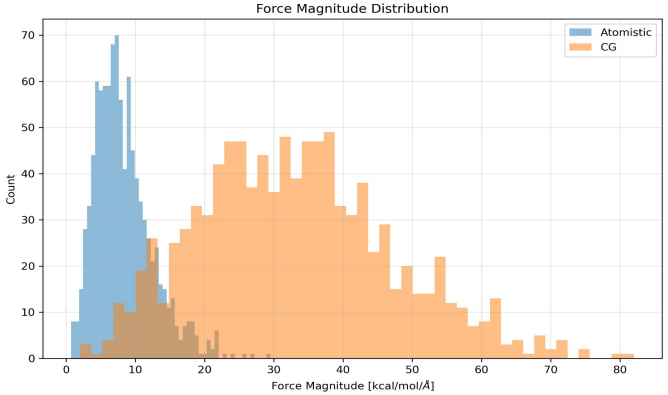
Results on SPC/E Initialization with α=1e-2









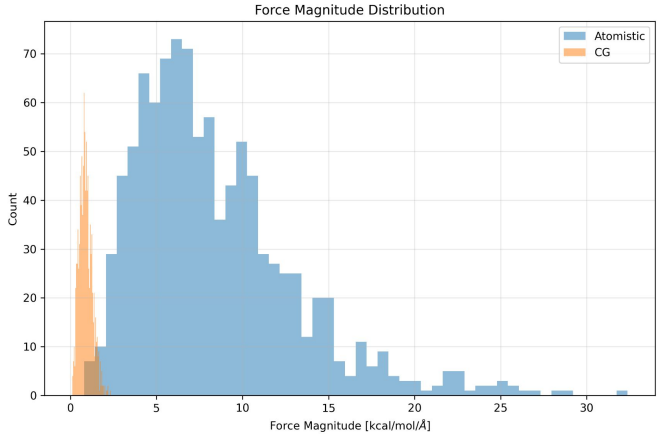


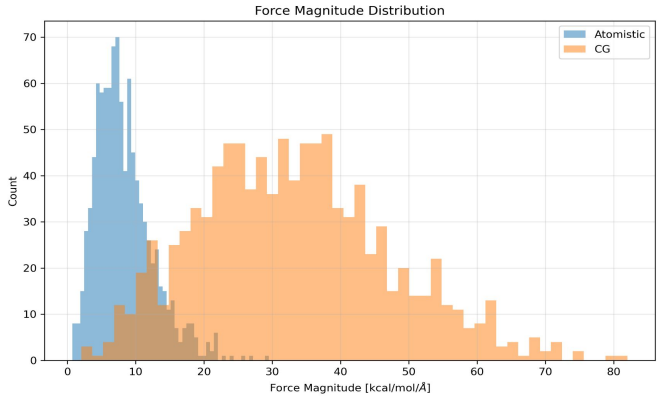
Swings to the other extreme, overestimating the spread, and in fact pays for it in a worse cross-validated MSE

SPC/E behaves like an overly “flexible” model (high variance, possibly overfitting)

MSE: 1439.47 kcal/mol/Å
Best Val Loss: 1424.55

Remarks on Initialization Strategies





Adaptive weighting during training

Regularize deviation in standard deviation from the reference force set

Central tendency ⊕ spread

$$\mathcal{L} = \mathrm{MSE}(\mathbf{F}_{\mathrm{CG}}; \mathbf{h}) + \lambda \left| \mathrm{std}(\mathbf{F}_{\mathrm{CG}}) - \mathrm{std}(\mathbf{h}) \right|$$

“Tight” LJ ⊕ “broad” SPC/E baseline

Convex combination of LJ and SPC/E coefficients (**hybrid initialization**)

$$\mathrm{MSE}_{(\mathrm{batch})}(\mathbf{F}_{\mathrm{CG}}; \mathbf{h}) = \frac{1}{N_{\mathrm{batch}}} \sum_{k=1}^{N_{\mathrm{batch}}} \sum_{I=1}^{M} \left\| \mathbf{h}_I(\mathbf{q}_k) - \mathbf{F}_{\mathrm{CG},I}(\mathbf{Q}_k; \boldsymbol{\theta}) \right\|^2$$

Thank you for your attention!

I am here because

I enjoy creating and presenting ideas.

Contacts:

gv@materials.uoc.gr / gv@csd.uoc.gr

Georgios Vourvachakis | LinkedIn

Dynamics of the CG Model: The Mori-Zwanzig Formalism

- The MZ formalism originates from the **microscopic Hamiltonian** and utilizes a projection operator to decompose the dynamics into slower macroscopic dynamics.
- Heisenberg time-evolution of an observable operator A(t): $\dot{A} = iLA$ with solution $A(t)=e^{iLt}A_0$ and iL=[H, .], the Liouville operator.
- Let $P=(A_0,A_0)^{-1}(\,.\,,A_0)A_0$, the projection operator onto the initial observable, and Q=1-P the one onto P's orthogonal subspace. Then, applying the operator identity (*Duhamel identity* or *Dyson expansion):* $e^{iLt} = e^{iQLt} + \int_0^t ds e^{iL(t-s)} PiLe^{iQLs}$ into the equation $\dot{A}(t) = iLe^{iLt}A_0$ and defining the *fluctuation term* $F(t)=\exp(iQLt)iQLA_0$, the *frequency term* $\Omega=(iLA_0,A_0)(A_0,A_0)^{-1}$ s.t. $\Omega A_0 = iPLA_0$, and the *memory kernel* $K(t)=(iLF(t),A_0)(A_0,A_0)^{-1}=(F(t),F(0))(A_0,A_0)^{-1}$ (2^{nd} FD Theorem) we obtain the **GLE form** of the MZ formalism (after invoking the *Mori identity*):

$$\dot{A} = \Omega A + \int_0^t \mathrm{d}sK(s)A(t-s) + F(t)$$

- The time evolution is decomposed into its **slow counterpart**, represented by A (i.e., ΩA), and the term F(t) corresponds to the **orthogonal residue dynamics**. These two parts are connected through the memory kernel K, which accumulates all the unresolved components from the initial stages.

Dynamics of the CG Model: The Mori-Zwanzig Formalism

Markovian approximation
→ rapid decay of ζ(t)

ζ(t)~δ(t) →
Langevin dynamics

- ❑ …the **microscopic Hamiltonian** and utilizes …macroscopic dynamics.
- ❑ Heisenberg time evolution of an observable operator A(t): $\dot{A} = iLA$ with solution $A(t)=e^{iLt}A_0$ and iL=[H, .], the Liouville operator.
- Let P=$(A_0,A_0)^{-1}(\ldots)$ … Q=1-P the one onto P's orthogonal sub… *Dyson expansion):* $e^{iLt} = e^{iQLt} + \int_0^t ds e^{iL(t-s)} PiLe^{iQLs}$ into the equation $\dot{A}(t) = iLe^{iLt}A_0$ and defining the *fluctuation term* F(t)=exp(iQLt)iQLA$_0$, the *frequency term* Ω=(iLA$_0$,A$_0$)(A$_0$,A$_0$)$^{-1}$ s.t. ΩA$_0$ ≡ iPLA$_0$, and the *memory kernel* K(t)… theorem … the MZ formalism (after invoking the *Mori identity*):

C(t) (memory) D(t) (noise)

$$\dot{\mathbf{P}}(t) = -\nabla_{\mathbf{Q}_i}\bar{U}^{\mathrm{PMF}}(\mathbf{Q}) - \int_0^t \bar{\zeta}(t-\tau)\mathbf{P}(\tau)d\tau + \mathbf{F}_R(t)\ ,\ t>0$$

[pairwise decomposition] $A \equiv \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_M\}$

$$\dot{A}(t) = \underbrace{\Omega A} - \int_0^t \zeta(t-\tau)A(\tau)d\tau + F(t)$$

Instantaneous (Markovian) Drift

$$\langle \mathbf{F}_R(t)\mathbf{F}_R(t')\rangle = 2\beta^{-1}\bar{\zeta}(t-t')\ ,\ t>t'$$

Fluctuation-Dissipation Theorem

- ❑ The time evolution is decomposed into its **slow counterpart**, represented by A (i.e., ΩA), and the term F(t) corresponds to the **orthogonal residue dynamics**. These two parts are connected through the … kernel, which … the … of the initial states.

[Baxevani et al., SIAM **21**(4) (2023)] [Sergei Izvekov, J. Chem. Phys. **138**(13), 134106 (2013)]

Fun Fact: if the kernel follows ζ~t^{-s} with s∈(0,1), the "random" force exhibits **pink noise characteristics**.

Dynamics of the CG Model: The Mori-Zwanzig Formalism

Elimination ⇒ Memory & Noise: By **eliminating** fast variables, one loses their explicit EOM but cannot erase their effects.

- Heisenberg time-evolution of an observable operator A(t): $\dot{A} = iLA$ with solution $A(t)=e^{iLt}A_0$ and iL=[H, .], the Liouville operator.
- Let $P=(A_0,A_0)^{-1}(\ldots$ … $Q=1-P$ the one onto P's orthogonal sub… … *Dyson expansion):* $e^{iLt} = e^{iQLt} + \int_0^t ds\, e^{iL(t-s)} PiLe^{iQLs}$ into the equation $\dot{A}(t) = iLe^{iLt}A_0$ and defining the *fluctuation term* $F(t)=\exp(iQLt)iQLA_0$, the *frequency term* $\Omega=(iLA_0,A_0)(A_0,A_0)^{-1}$ s.t. $\Omega A_0 = iPLA_0$, and the *memory kernel* K(t) … the MZ formalism (after invoking the *Mori identity*):

C(t) (memory) D(t) (noise)

$$\dot{\mathbf{P}}(t) = -\nabla_{\mathbf{Q}_i} \bar{U}^{\mathrm{PMF}}(\mathbf{Q}) - \int_0^t \bar{\zeta}(t-\tau)\mathbf{P}(\tau)d\tau + \mathbf{F}_R(t)\ ,\ t>0$$

[pairwise decomposition] $A \equiv \{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_M\}$

$$\langle \mathbf{F}_R(t)\mathbf{F}_R(t')\rangle = 2\beta^{-1}\bar{\zeta}(t-t')\ ,\ t>t'$$

Fluctuation-Dissipation Theorem

$$\dot{A}(t) = \Omega A - \int_0^t \zeta(t-\tau)A(\tau)d\tau + F(t)$$

Instantaneous (Markovian) Drift

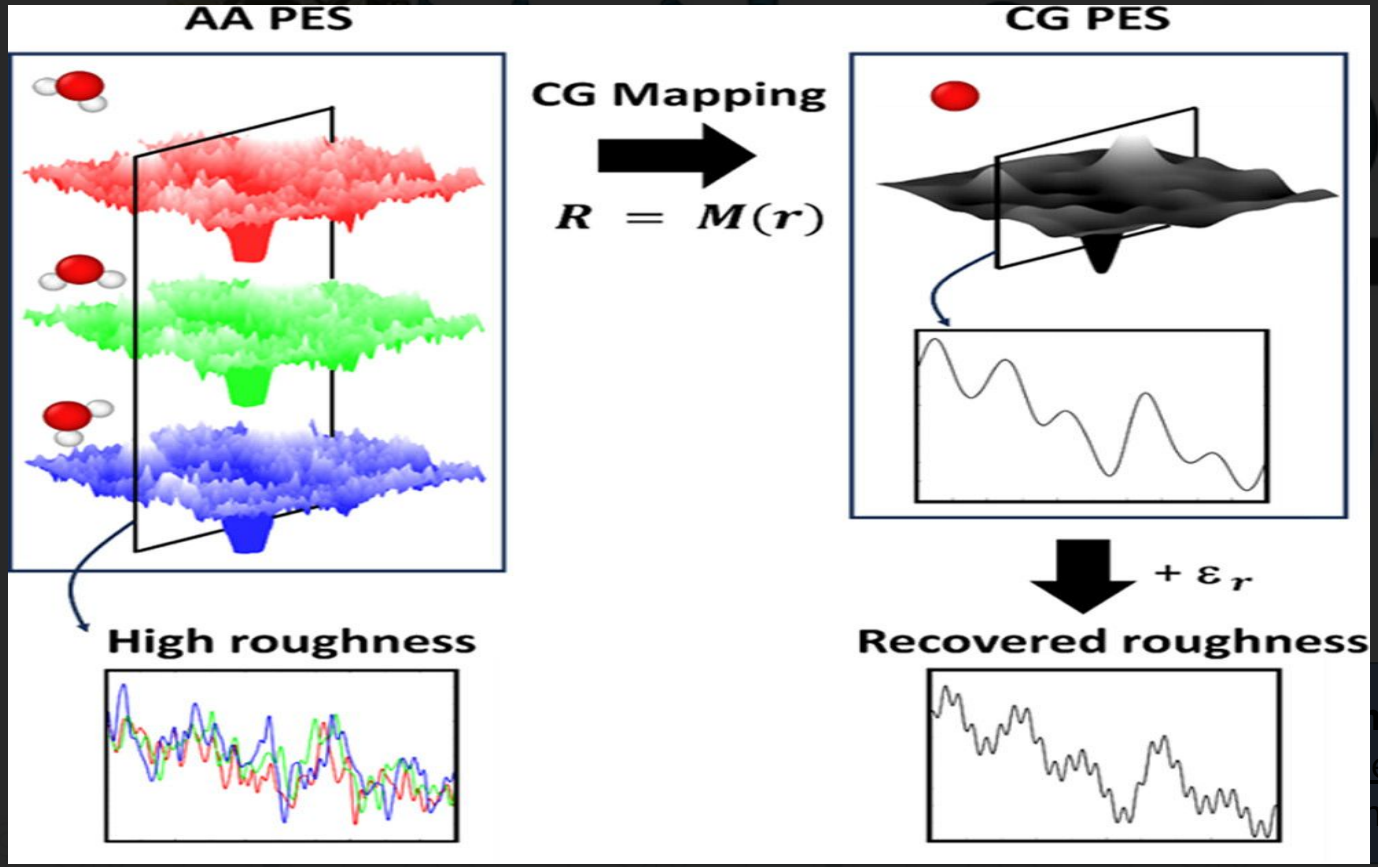
- The time evolution is decomposed into its **slow counterpart**, represented by A (i.e., ΩA), and the term F(t) corresponds to the **orthogonal residue dynamics**. These two parts are connected through the memory kernel K, which accumulates all the unresolved components from the initial stages.

[Baxevani et al., SIAM **21**(4) (2023)]

[Sergei Izvekov, J. Chem. Phys. **138**(13), 134106 (2013)]

Fun Fact: if the kernel follows $\zeta \sim t^{-s}$ with $s \in (0,1)$, the "random" force exhibits **pink noise characteristics**.

Optimal Coarse-Graining Strategy



Visualization of an idealized schematic of a rugged **all-atom landscape** (top-left) and a smoothened **CG landscape** (top-right). The sliced potential energy surface at the bottom provides a qualitative representation of **high-frequency "rough" features**, which can be emulated by **introducing perturbation to the CG potential**.

[Ishan Nadkarni et al., The JoPC A **128**(46) (2024)]