ECEA 2019
5th International Electronic Conference on Entropy and Its Applications
18–30 November 2019
Chaired by Prof. Geert Verdoolaege
Sponsored by: entropy MDPI

Systematic Coarse-Grained Models for Molecular Systems Using Entropy

Vagelis Harmandaris [1,2]*, Evangelia Kalligiannaki [2]*, and Markos Katsoulakis [3]

[1]Department of Mathematics and Applied Mathematics, University of Crete;
[2]Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas;
[3]Department of Mathematics and Statistics, University of Massachusetts, Amherst, USA.

* Corresponding author: harman@uoc.gr (V.H.), evangelia.kalligiannaki@iacm.forth.gr (E.K.)

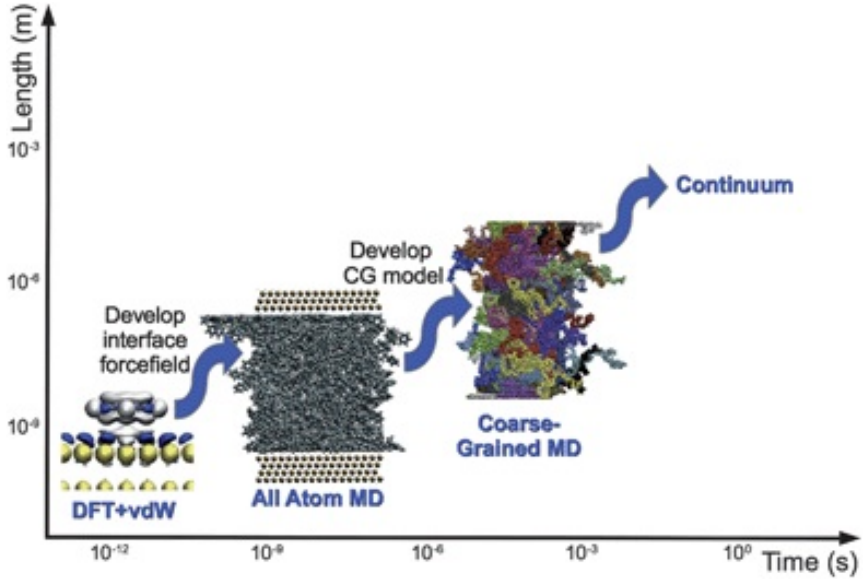
Abstract: The development of systematic coarse-grained mesoscopic models for complex molecular systems is an intense research area. Here we first give an overview of different methods for obtaining optimal parametrized coarse-grained models, starting from detailed atomistic representation for high dimensional molecular systems. We focus on methods based on information theory, such as relative entropy, showing that they provide parameterizations of coarse-grained models at equilibrium by minimizing a fitting functional over a parameter space. We also connect them with structural-based (inverse Boltzmann) and force matching methods. All the methods mentioned in principle are employed to approximate a many-body potential, the (n-body) potential of mean force, describing the equilibrium distribution of coarse-grained sites observed in simulations of atomically detailed models. We also present in a mathematically consistent way the entropy and force matching methods and their equivalence, which we derive for general nonlinear coarse-graining maps. Finally, we apply, and compare, the above-described methodologies in several molecular systems: gas and fluid methane, water, and a polymer.

Keywords: coarse-graining; data-driven; relative entropy; path-space; uncertainty quantification

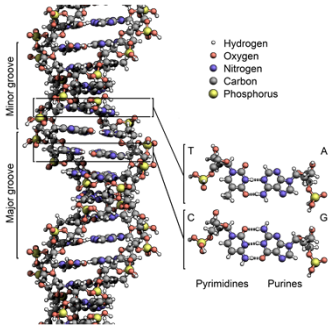
Motivation

Simulating complex molecular systems: Enormous range of Length-Time scales

- Need to reduce complexity and system size.
- Dimensionality reduction: Coarse-graining
- Statistical equilibrium: Structural properties.
- Non-equilibrium: Dynamical properties.



❑ K. Johnson, V. Harmandaris, Soft Matter, 2013, 9, 6696-6710



Coarse-graining and Potential of Mean Force

❑ Equilibrium Statistical Mechanics

A system of N >> 1 particles;

U(x), x ∈ R^{3N} interaction potential, Gibbs configurational probability density

$$\mu(q) = Z^{-1} e^{-\beta U(q)}, \; q \in \mathbb{R}^{3N}$$

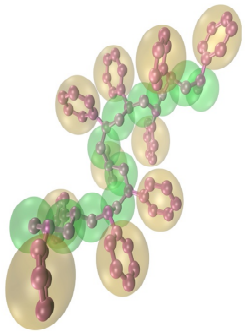
❖ Coarse-graining (CG) as transformation operator

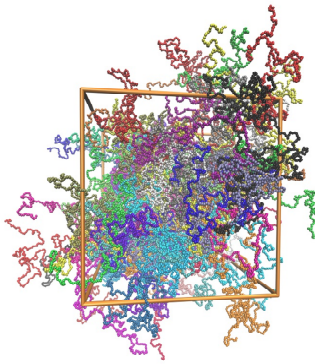
Π : $R^{3N} \rightarrow R^{3M}$, M < N

Examples:

✓ Linear map: Center of mass of groups

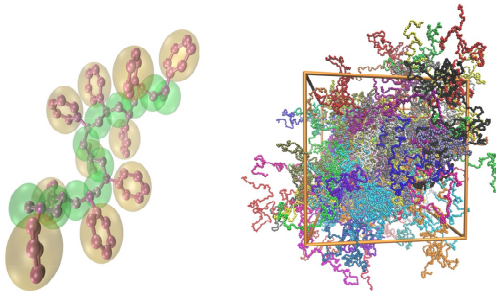
✓ Non-linear map: Bond angle, end-to-end vector





Coarse-graining and Potential of Mean Force

- Exact CG model at equilibrium Q $\in R^{3M}$



Potential of Mean Force (PMF)

$$U^{pmf}(Q) = -\frac{1}{\beta}\ln\int_{A(Q)} e^{-\beta U(q)}dq,$$

- CG probability density

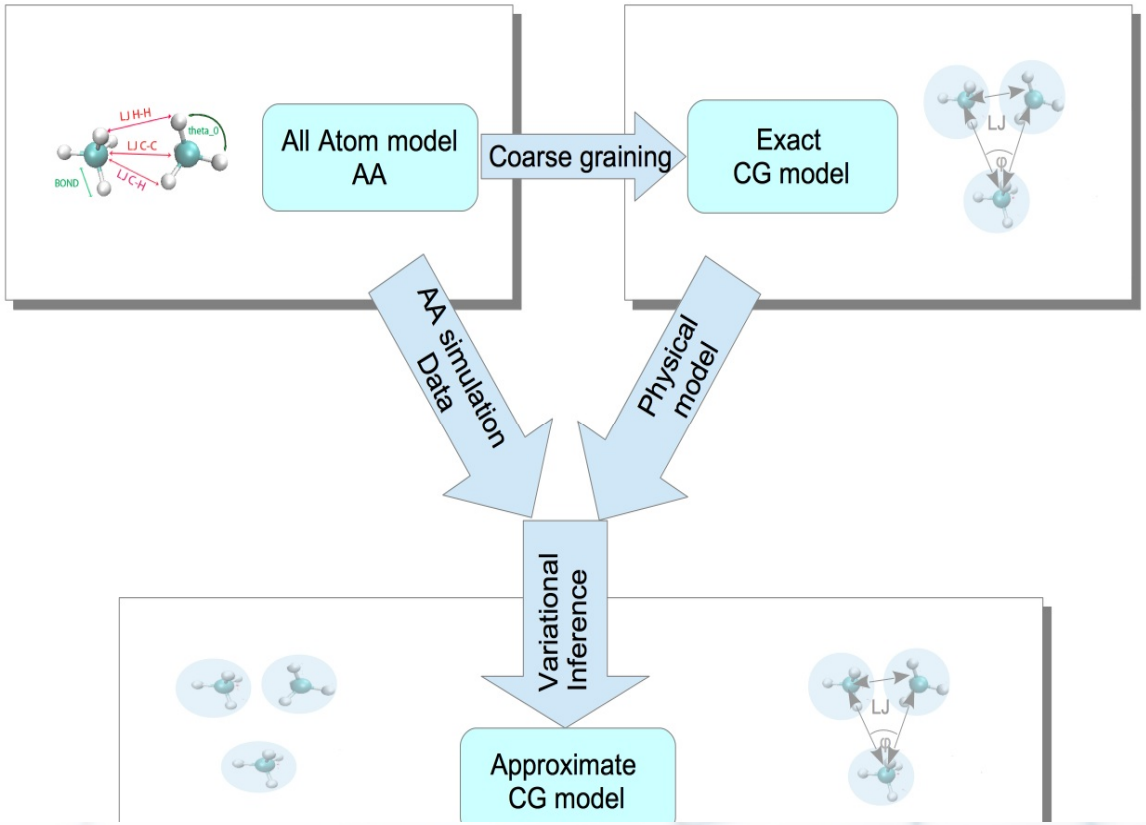
$$\mu^{pmf}(Q) \propto e^{-\beta U^{pmf}(Q)} \qquad A(Q) = \{q \in \mathbb{R}^{3N} : \Pi(q) = Q\}$$

✓ Exact! But still High-dimensional!

Approximate – Effective Coarse Models

- **Main goal:** Derive effective CG models consistent with structure and dynamic properties of the microscopic system.

"Digital Twin"



Use atomistic information to find effective mesoscopic model

$$\overline{U}(Q;\theta) \sim U^{pmf}(Q)$$

Parameter set

$$\theta \in \Theta$$

Methods for optimal parametrization of mesoscopic models

- How to find optimal $\theta \in \Theta$ $\quad \overline{U}(Q;\theta) \sim U^{pmf}(Q)$
  s.t.

- In what sense optimal? Various Methodologies

Objective

- Force matching

  Forces

  Voth et.al. J. Phys. Chem. B 2005, Noid et.a.l. 2008,
  J.F. Rudzinski and W.G. Noid (2012) ,
  Kalligiannaki et. al. J. Chem. Phys. 2015

- Variational inference: Relative entropy minimization Gibbs measures

  M.S. Shell J. Chem. Phys. 2008, A. Chaimovich, M.S. Shell 2009

- Inverse Boltzmann, Inverse Monte Carlo

  Pair correlation

  A.K. Soper Chem. Phys. 1996, F. Muller-Plathe Chem.Phys.Chem. 2002,
  A. P. Lyubartsev and A. Laaksonen, N. Meth. Soft Matter Sim., 2004

Variational inference: Relative Entropy minimization

- Relative Entropy minimization (Information theory).

$$\min_\theta \mathcal{R}\left(\mu \,||\, \bar{\mu}_\theta\right)$$

$$\bar{\mu}_\theta \propto exp\{-\beta \bar{U}(\cdot;\theta)\}$$

$$\mathcal{R}\left(\mu \,||\, \bar{\mu}_\theta\right) = \int \mu(x) \log \frac{\mu(x)}{\bar{\mu}_\theta(x)} dx$$

- Relative Entropy measures the Information loss when using probability ν instead of μ

$$\mathcal{R}(\mu|\nu) \geq 0\,, \qquad \mathcal{R}(\mu|\nu) = 0 \text{ iff } \mu = \nu$$

$$\mathcal{R}(\mu|\nu) \neq \mathcal{R}(\nu|\mu)$$

Thus, the optimal parameter set θ^* is the solution of the optimization problem:

$$\min_{\theta\in\Theta} \left\{ \beta \mathbb{E}_\mu \left[\bar{U}(\mathbf{\Pi}(\mathbf{q});\theta) - U(\mathbf{q})\right] - \left[\log Z^\theta - \log Z\right] \right\}$$

Dynamics: Path-space Relative Entropy minimization

❑ **Dynamics is described by stochastic processes:**

Microscopic (atomistic): $\mathrm{X}_t = (q_t, p_t), \ \in \mathbb{R}^{6N}, \ t > 0$

Mesoscopic (coarse-grained): $\overline{X}_t = (Q_t, P_t), \ \in \mathbb{R}^{6M}, \ t > 0$

Back-mapped coarse-grained: $\widetilde{X}_t = (\widetilde{q}_t, \widetilde{p}_t), \ \in \mathbb{R}^{6N}, \ t > 0$

✓ With path-space probabilities

$$P_X^{[0,T]} \qquad P_{\overline{X}}^{[0,T]} \qquad P_{\widetilde{X}}^{[0,T]}$$

Dynamics: Path-space Relative Entropy minimization

- **Dynamics is described by stochastic processes**

- Path-space Relative Entropy

$$\mathcal{R}(P_X^{[0,T]}|P_{\tilde{X}}^{[0,T]}) = \mathbb{E}_{P_X^{[0,T]}}\left[\log \frac{dP_X^{[0,T]}}{dP_{\tilde{X}}^{[0,T]}}\right]$$

- Relative Entropy Rate

$$\mathcal{H}(P_X|P_{\tilde{X}}) = \lim_{T\to\infty} \frac{1}{T}\mathcal{R}(P_X^{[0,T]}|P_{\tilde{X}}^{[0,T]})$$

Coarse-graining Langevin dynamics

Microscopic (Atomistic) representation

position $q \in \mathbb{R}^{3N}$

momentum $p \in \mathbb{R}^{3N}$

Mesoscopic CG mapping

CG map: $\mathbf{\Pi} : \mathbb{R}^{6N} \to \mathbb{R}^{6M}$, $M < N$

Approximate dynamics model

Atomistic

$$\begin{cases} dq_t = M^{-1} p_t dt \,, \\ dp_t = F(q_t)dt - \gamma M^{-1} p_t dt + \sigma dB_t \,, \end{cases}$$

Coarse-grained

$$\begin{cases} d\bar{q}_t = M^{-1} \bar{p}_t dt \,, \\ d\bar{p}_t = \bar{F}(\bar{q}_t; \theta)dt - \gamma M^{-1} \bar{p}_t dt + \sigma d\bar{B}_t \,. \end{cases}$$

Atomistic Force

$$F(q) = -\nabla_q U(q)$$

CG Force/Potential

$$\bar{F}(\cdot; \theta) = -\nabla \bar{U}(\mathbf{Q}; \theta)$$

Coarse-graining Langevin dynamics

- Derivation of CG model: CG force/potential need to be parametrized

Path-space Relative Entropy minimization reduces to 'path-space force-matching'

$$\begin{aligned}\theta^* &= \text{argmin}_\theta \mathcal{R}(P_X^{[0,T]} | P_{\tilde{X}}^{[0,T]}) \\ &= \text{argmin}_\theta \mathbb{E}_{P^{[0,T]}} \left[\frac{1}{2\sigma^2} \int_0^T \left\| \mathbf{\Pi} F(q_s) - \bar{F}(\mathbf{\Pi} q_s; \theta) \right\|^2 ds \right]\end{aligned}$$

$$\mathbf{\Pi} F = \sum_{\{i \in \text{ CG particle } j\}} F_i(\mathbf{q})$$

V. Harmandaris, E. Kalligiannaki, M. Katsoulakis, P. Plechac, J. Comp. Phys. 2016,
M. Katsoulakis, P. Plechac, 2013,

Relative Entropy Rate Minimization and Force Matching

In stationary (equilibrium) dynamics, path-space relative entropy minimization reduces to Relative Entropy Rate minimization which in turn reduces to the Force Matching method

$$\theta^* = \text{argmin}_\theta \mathbb{E}_\mu \left[\frac{1}{2\sigma^2} \left\| \mathbf{\Pi} F(q) - \bar{F}(\mathbf{\Pi} q; \theta) \right\|^2 \right]$$

For discrete time path observations $\mathcal{D}_{n_s,n_t} = \{\mathbf{X}_1^k, \ldots, \mathbf{X}_{n_t}^k\}_{k=1}^{n_s}$

Path-space Relative Entropy

$$\hat{\theta} = \underset{\theta}{\text{argmax}} \sum_{k=1}^{n_s} \log \frac{\bar{P}(\mathbf{\Pi} X_1^k, \mathbf{\Pi} X_2^k, ..., \mathbf{\Pi} X_{n_t}^k)}{\bar{Q}^\theta(\mathbf{\Pi} X_1^k, \mathbf{\Pi} X_2^k, ..., \mathbf{\Pi} X_{n_t}^k)}$$

Relative Entropy Rate

$$\hat{\theta} = \text{argmax}_\theta \sum_{i=1}^{n_t - 1} \log q^\theta(\mathbf{\Pi} X_i, \mathbf{\Pi} X_{i+1})$$

$q^\theta(\bar{\mathbf{X}}, \bar{\mathbf{X}}')$ Parametric transition probability density

Quantifying Uncertainty in Coarse-grained models

Goal: Provide confidence sets for the derived optimal CG model

- Large number of observations $n_s >> 1$

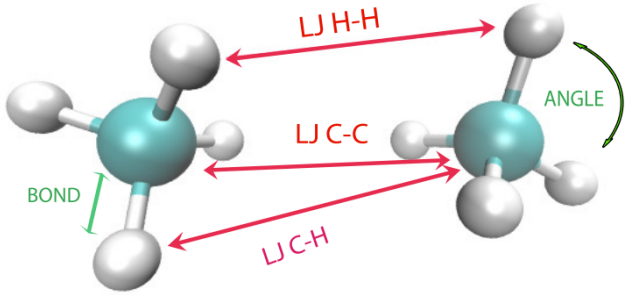
  Point-estimates: Asymptotic standard error

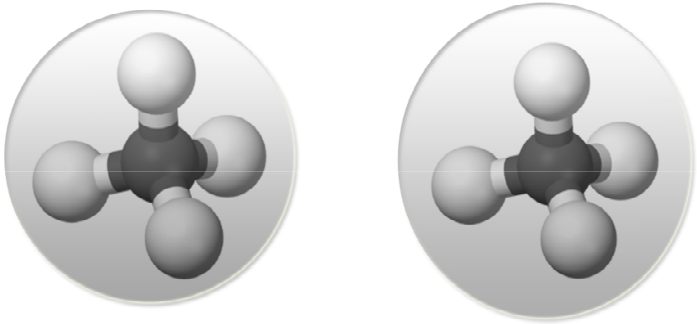
- Small number of observations n_s

  Frequentist statistics tools: Bootstrap, jackknife

  Bayesian statistics tools

Example: Simple fluid - Bulk methane CH_4





❑ Coarse-graining map: $\Pi : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3M}$

Center of mass $Q_i = \sum_{k\in \text{ mol } i} m_k q_k / \sum_{k\in \text{ mol } i} m_k$

❖ CG parametrized interaction (two-body, pair) potential: $\bar{U}(\mathbf{Q};\theta) = \sum_{i,j} u(R_{ij};\theta)$

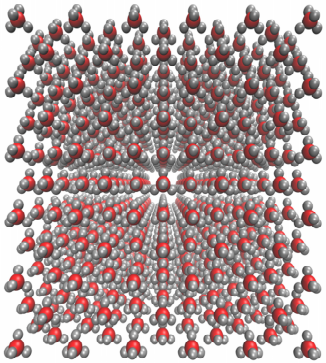
✓ B-splines $u(R;\theta) = \sum_{k=1}^{n} \theta_k \psi_k(R)$

✓ Lennard-Jones $u(R;\theta) = \theta_1/R^{12} - \theta_2/R^6$

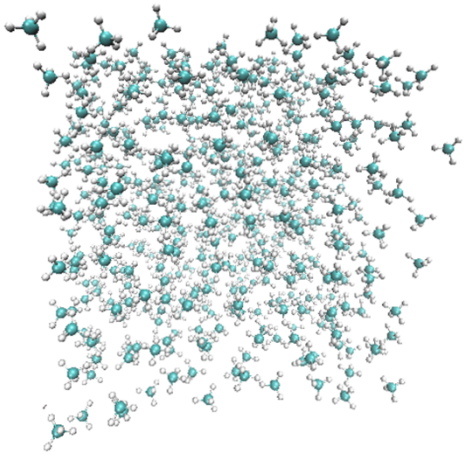
Study at Equilibrium and Transient Time Regimes

- Data are generated from Molecular dynamics simulations for M = 666 methane molecules at temperature 100K (density is 0.3825 gr/cm^3), and initial positions of molecules are at a FCC (Face Centered Cubic) crystal structure.

✓ Initial configuration: FCC crystal



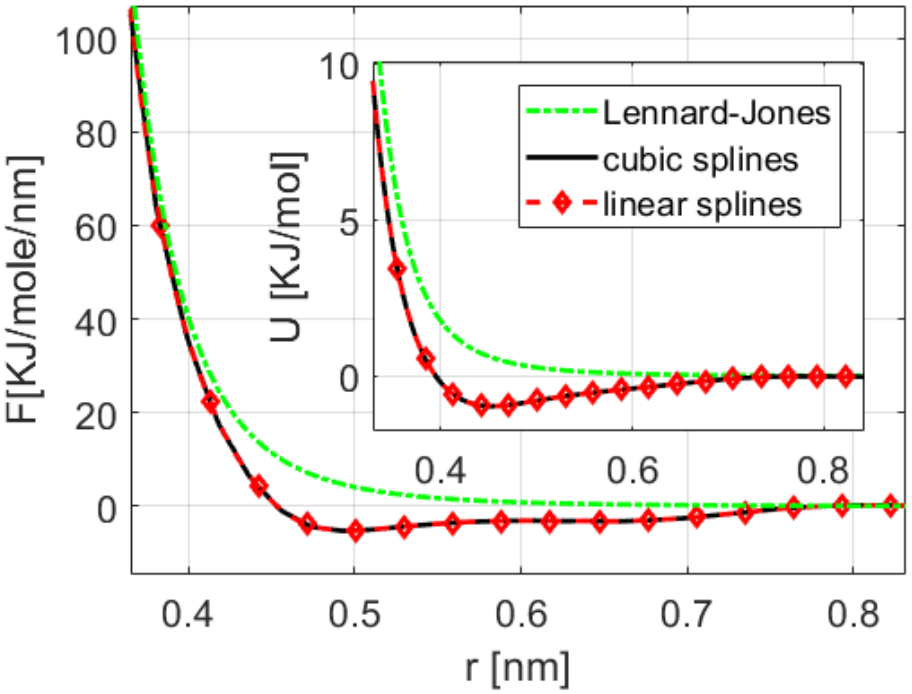
✓ Equilibrated configuration:



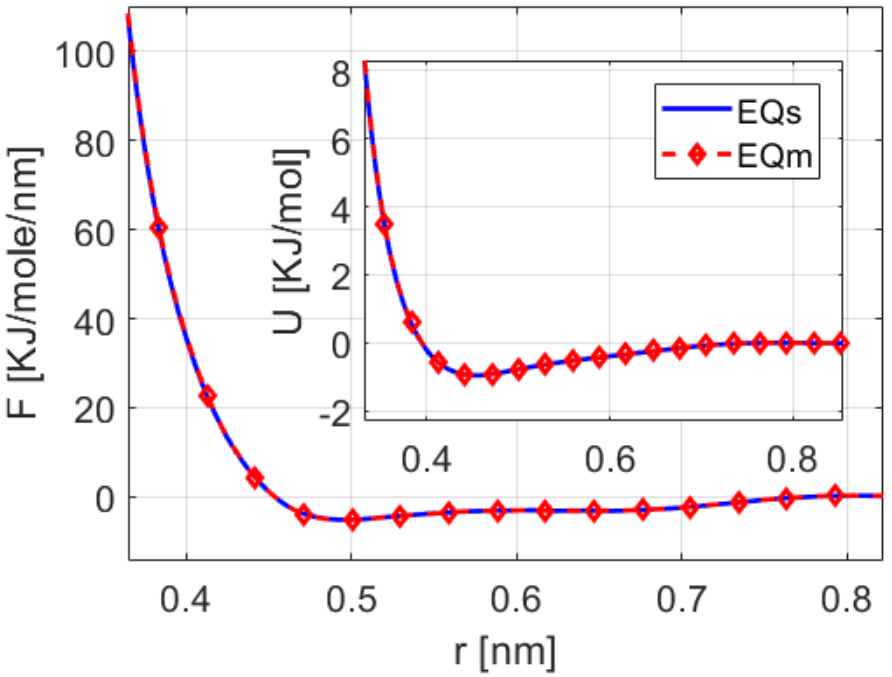
System	Paths (n_p)	time step dt (ps)	Time interval (ps)	Configurations (n_t) per path
tFCC	10	0.1	[0,100]	1000
EQm	10	0.1	[900,1000]	1000
EQs	1	1	[1000,10000]	9000

Results at Equilibrium Regime

Path-space Force Matching



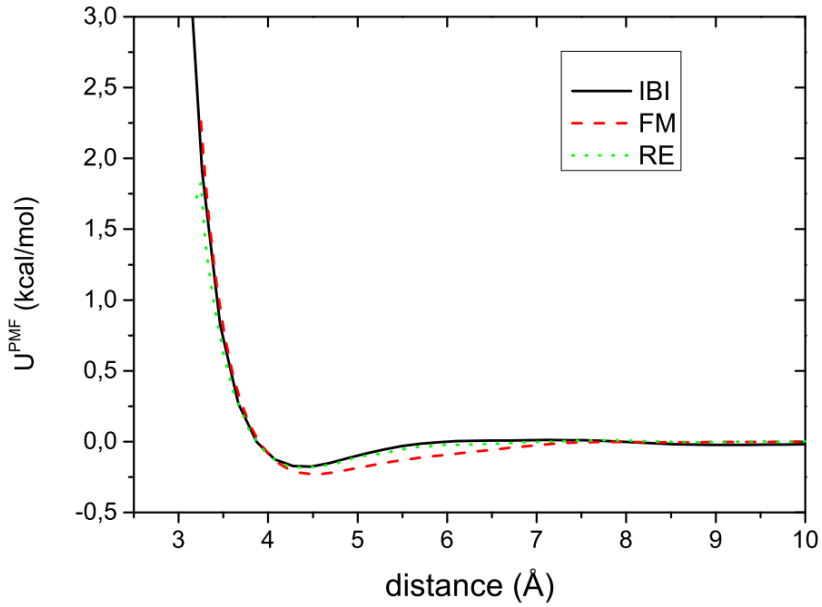
Force Matching &
Path-space Force Matching



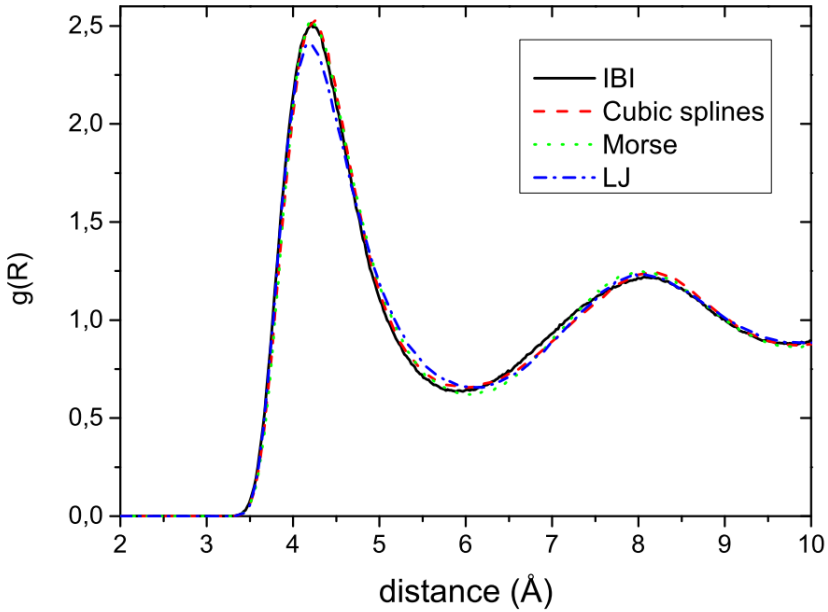
- ✓ Forces of the equilibrium data set (Eqm), obtained through the PSFM method.
- ✓ In the inset is the derived CG pair effective potential.

Comparison of Equilibrium Methods

Effective pair interaction potential u(R;θ)

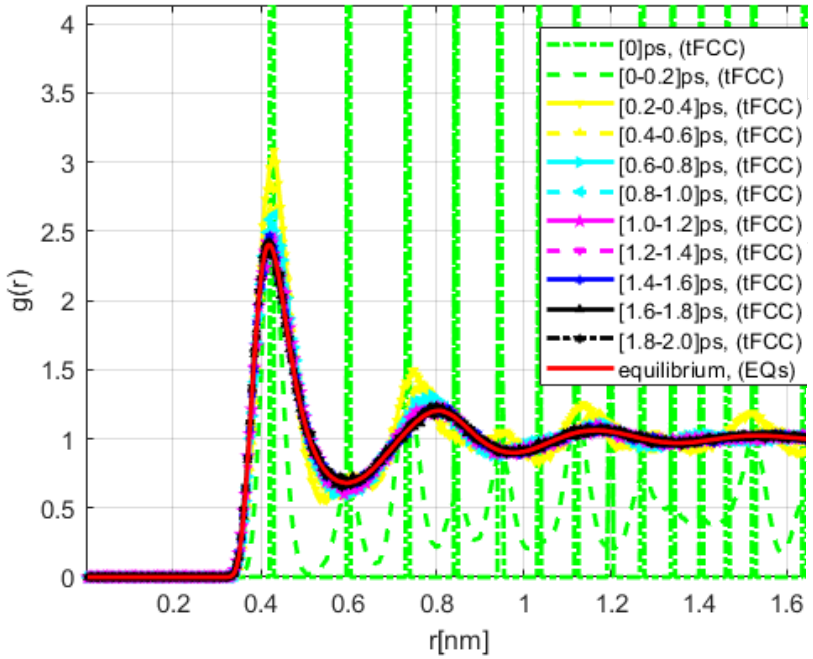


Pair correlation function

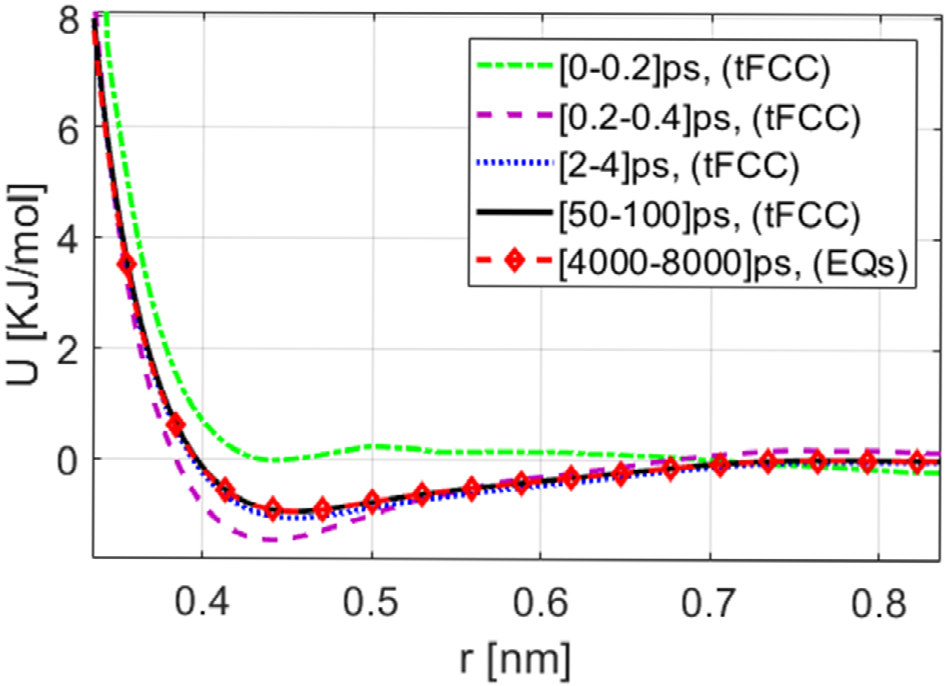


✓ E. Kalligiannaki., A. Chazirakis, A. Tsourtis, M. Katsoulakis, P. Plechac, V. Harmandaris, EPJ ST, 225, 1347–1372, 2016

Results at Transient Time Regime



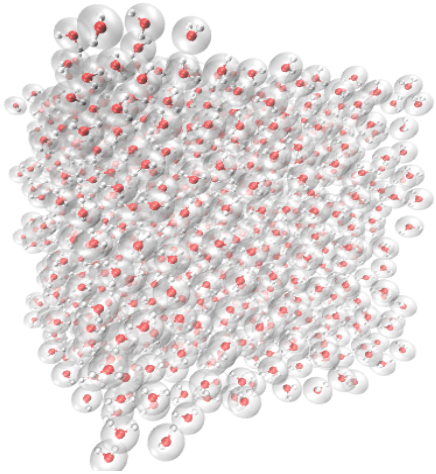
✓ Evolution of the RDF g(r) of the data set tFCC with initial FCC crystal structure, for different time sub-intervals from the all-atom simulation



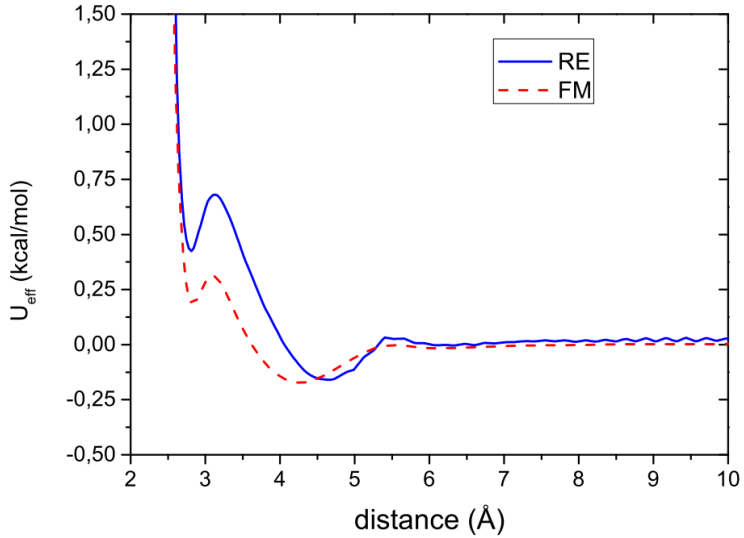
✓ Evolution of the effective CG potential with cubic splines, for different time sub-intervals of the all-atom simulation.

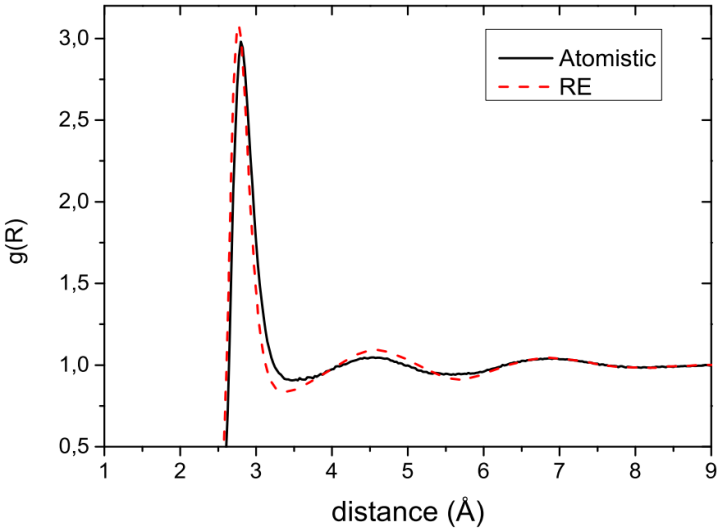
Water, one-site CG representation

- Simulated all-atom water, using the SPC/E force field. The model system consists of 1192 molecules at ambient conditions , T = 300 K, P = 1 atm. All-atom configurations were recorded every 10 ps.



- CG effective interactions for CG water molecules by analyzing the all-atom data, using force matching and relative entropy techniques





Polymer Bulk System: Polyethylene

- Atomistic simulated system: 96 PE chains with 99 monomers each, N = 96 × 99, temperature = 450K

- CG map 3:1, i.e., CG system size M = 96×33

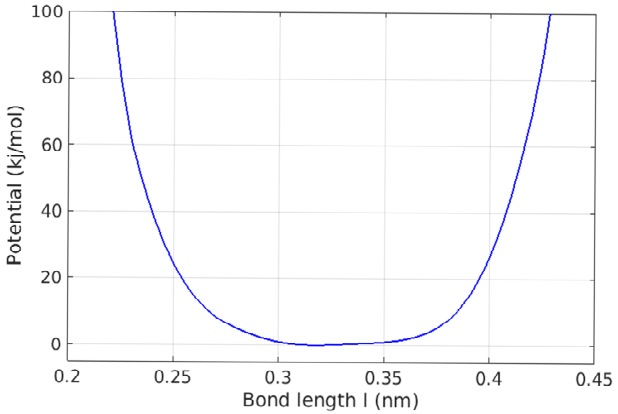
$$\bar{U}(Q;\theta) = \bar{U}_b(Q;\theta) + \bar{U}_{nb}(Q;\theta)$$

- ✓ CG Bonded (Bonds; Angles; Dihedrals) interactions: Estimated with the Iterative Inverse Boltzmann method in tabulated form.

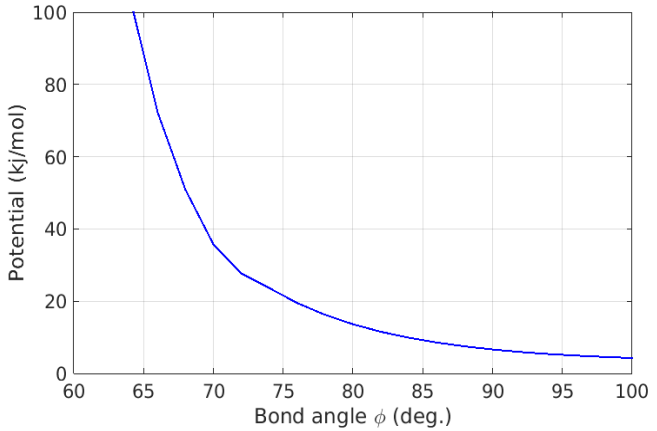
- ✓ CG Non-bonded interaction potential: Two-body pair potential: u(R ; θ) cubic B-splines. Estimated with the Force-Matching method.

CG Polyethylene: “Bonded” Interactions

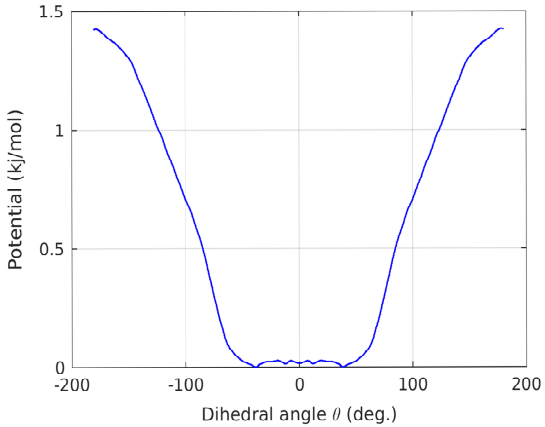
Effective interaction between CG “Bonds”



Effective interaction between CG “Angles”



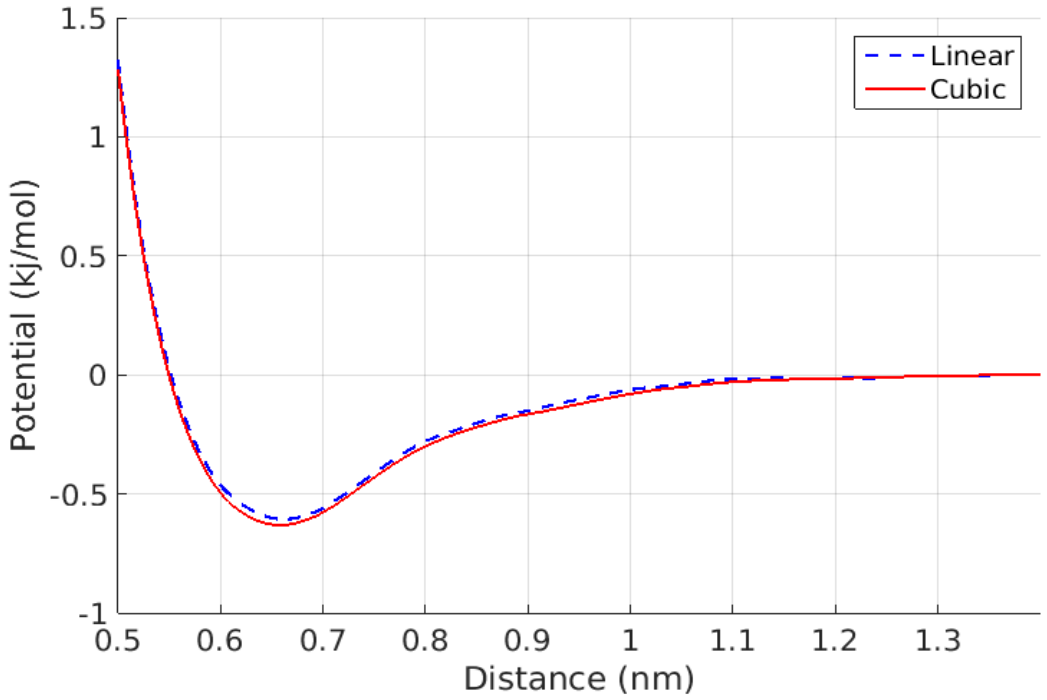
Effective interaction between CG “Dihedrals”



Non-bonded potentials: Large & small number of observations

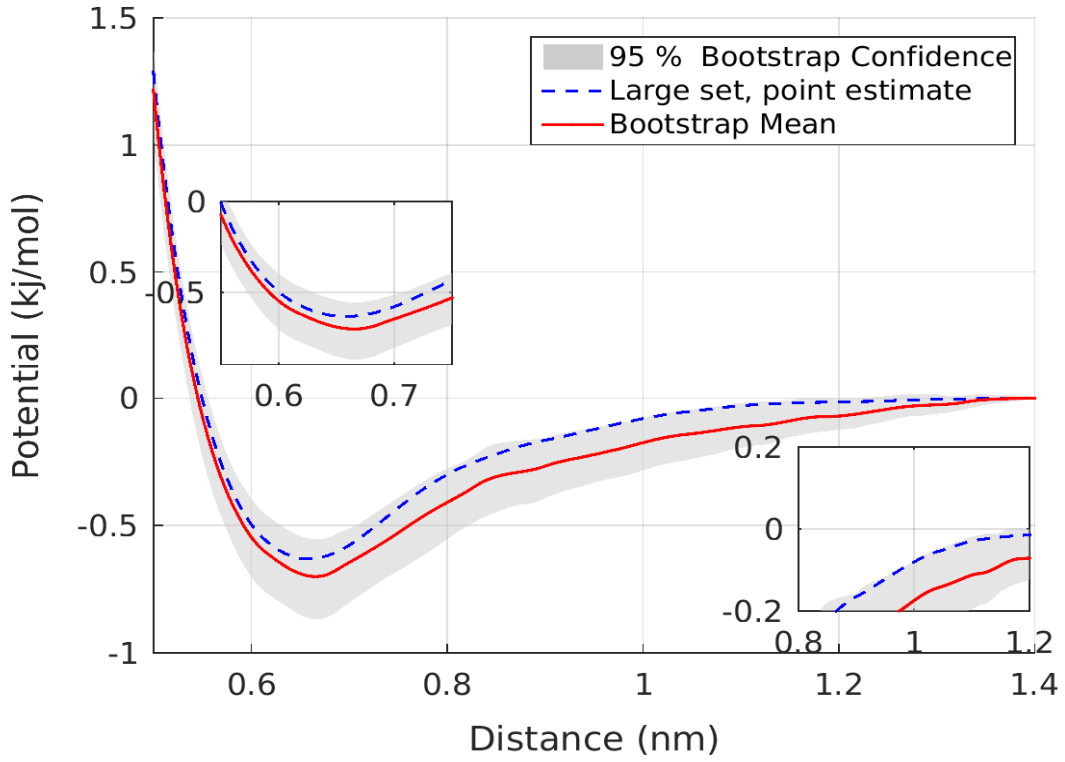
- ❑ Accuracy of the CG non-bonded effective interaction depends on the dataset.

- ✓ Point estimates for a large data et (2000 configurations): Linear versus cubic B-splines, pair potential representation



Non-bonded potentials: Large & small number of observations

- ✓ Estimates for a small dataset (200 configurations): Bootstrap results for the cubic B-spline representation



Conclusion and Discussion

- Variational inference for mesoscopic models; continuous and discrete time observations;
- Hybrid data driven physics-based coarse-graining approach;
- Systems out of equilibrium;
- Quantify uncertainties;
- Transferability;
- Challenging: dependencies and correlations in space/time and between model elements (molecules, parameters, and mechanisms), regions of sparse data

Related References

- G. Baxevani, E. Kalligiannaki, V. Harmandaris. Study of the transient dynamics of coarse-grained molecular systems with the path-space force-matching method. Procedia Computer Science, 156, 59- 68, 2019.

- V. Harmandaris, E. Kalligiannaki, M. A. Katsoulakis Computational Design of Complex Materials Using Information Theory: from Physics-to Data-driven Multi-scale Molecular Models, ERCIM News 115, 19-20, 2018.

- E. Kalligiannaki, A. Chazirakis, A. Tsourtis, M. Katsoulakis P. Plechac and V. Harmandaris, Parametrizing coarse grained models for molecular systems at equilibrium, EPJ ST, 225(8), 1347-1372, 2016, DOI:10.1140/epjst/e2016-60145-x.

- V. Harmandaris, E. Kalligiannaki, M. Katsoulakis and P. Plechac, Path-space variational inference for non-equilibrium coarse-grained systems, J. Comp. Phys., 314(1), 355–383, 2016, DOI:10.1016/j.jcp.2016.03.021.

- E. Kalligiannaki, V. Harmandaris, M. Katsoulakis and P. Plechac, The geometry of force matching in coarse graining and related information metrics, J. Chem. Phys., 143, 084105, 2015, DOI:10.1063/1.4928857.

Supplementary Material

❖ Links:

MACOMMS: Mathematical and Computational Modeling of Complex Molecular Systems.

SISDECS: Statistical Inference with Stochastic Differential Equations and applications in Complex Stochastic systems.

M. Katsoulakis, Department of Mathematics and Statistics , UMass, Amherst

Mathematics & Applied Mathematics, University of Crete

Institute of Applied and Computational Mathematics, Foundation for Research and Technology-Hellas

Acknowledgments

E.K. acknowledges this project has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No [52], SISDECS.