Boltzmann priors for Implicit Transfer Operators (BoPITO)

Juan Viguera Diez[1,2], Mathias Schreiner[1], Ola Engkvist[1,2], and Simon Olsson[1].

Department of Computer Science, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden[1]
Molecular AI, Discovery Sciences, R&D, AstraZeneca Gothenburg, Pepparedsleden 1, 431 50 Mölndal, Sweden[2]

Efficient generative models of MD incorporating enhanced sampling data

Accurate prediction of thermodynamic properties is essential in drug discovery and materials science. Molecular dynamics (MD) simulations provide a principled approach to this task, yet they typically rely on prohibitively long sequential simulations. Implicit Transfer Operator (ITO) Learning [1] offers a promising approach to address this limitation by enabling stable simulation with time steps orders of magnitude larger than MD but relies on extensive training data. Here, we introduce Boltzmann Priors for ITO (BoPITO) to enhance ITO learning in two ways. First, BoPITO enables more efficient data generation, and second, it embeds inductive biases for long-term dynamical behavior, simultaneously improving sample efficiency by one order of magnitude and guaranteeing asymptotically unbiased equilibrium statistics. Further, we showcase the use of BoPITO in a new tunable sampling protocol interpolating ITO models trained on off-equilibrium simulation data and an unbiased equilibrium distribution to solve inverse problems in molecular science.

Method

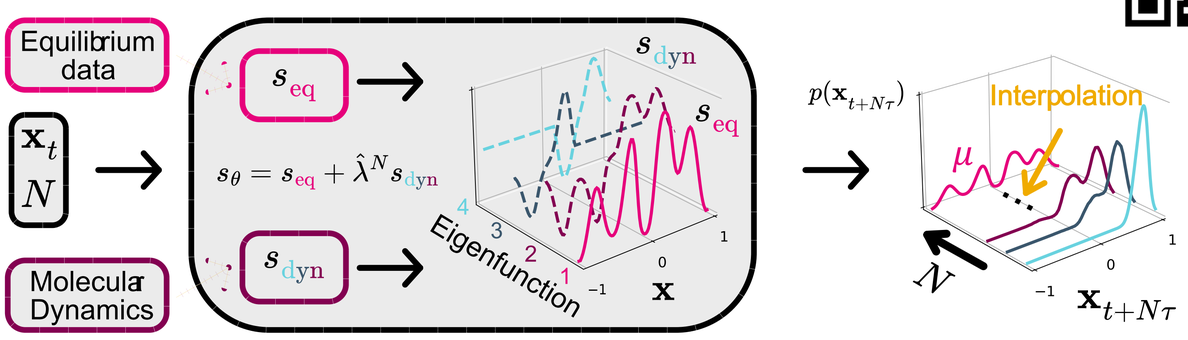
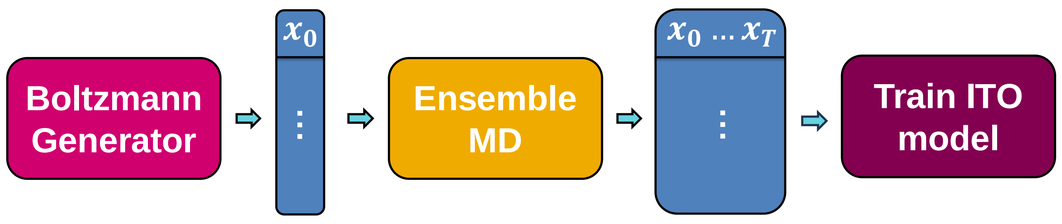


Figure 1: **Boltzmann Priors for Implicit Transfer Operators (BoPITO)** leverage pre-trained Boltzmann Generators [2] to enable data-efficient training of surrogate models of the transition density after N simulation steps. BoPITO is implemented using Score-Based Diffusion Models. The score of the model, s_θ, separates the contributions from the first eigenfunction of the transfer operator (equilibrium density), s_{eq}, from the rest, s_{dyn}. BoPITO embeds inductive biases for long-term dynamical behavior and enables interpolation between off-equilibrium and equilibrium models..

- **Efficient data generation:**



- **Long-term inductive bias diffusion model:**
  - We separate the equilibrium component of the score from the time-dependent ones and introduce a decay:

$$s_\theta\left(x_{t+N\tau}^{t_{diff}}, x_t, N, t_{diff}\right) = s_{eq}\left(x_{t+N\tau}^{t_{diff}}, t_{diff}\right) + \hat{\lambda}^N\, s_{dyn}\left(x_{t+N\tau}^{t_{diff}}, x_t, N, t_{diff}\right)$$

- **Interpolation between off-equilibrium models and the equilibrium distribution:**

$$s_\theta\left(x^{t_{diff}}, x_t, N_{int}, t_{diff}\right) = s_{eq}\left(x^{t_{diff}}, t_{diff}\right) + \hat{\lambda}^{N_{int}}\, s_{dyn}\left(x^{t_{diff}}, x_t, N_{max}, t_{diff}\right)$$

  - We propose to fit the interpolation parameter, N_{int}, using experimental observables and combining interpolation steps with non-interpolation steps for local relaxation.

References

1. Implicit Transfer Operator Learning: Multiple Time-Resolution Surrogates for Molecular Dynamics. Schreiner et al. NeurIPS 2023. DOI: 10.48550/arXiv.2305.18046
2. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. Noé et al. Science **365**, eaaw1147 (2019). DOI:10.1126/science.aaw1147

Results

Efficient data generation :

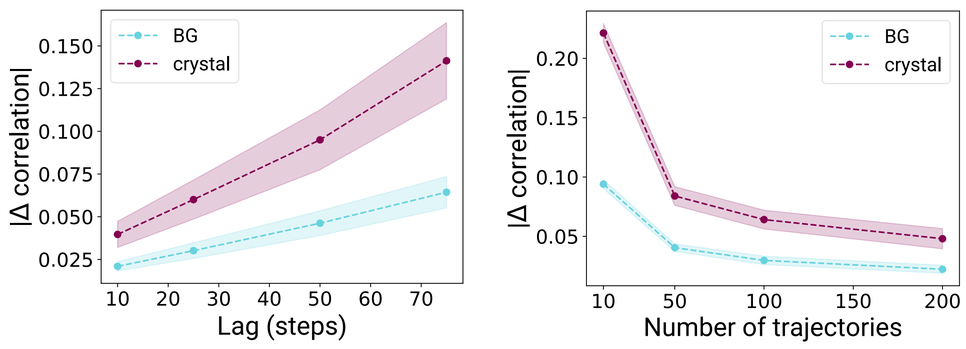


Figure 1: Absolute difference in correlation (lower is better) of models trained on trajectories initialized on samples from a Boltzmann Generator (BG) and a single structure (crystal, x=0.75) for the Prinz Potential under direct sampling. Left plot corresponds to 50 trajectories.

- **Long-term inductive bias diffusion model:**

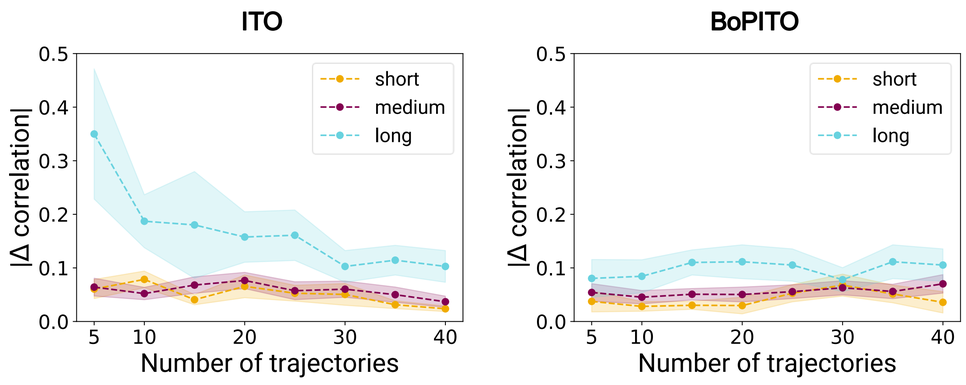


Figure 2: BoPITO can improve sample efficiency. Absolute difference in correlation (lower is better, one-step sampling) between ITO (left) and BoPITO (right) split into short, medium, and long time-scales against the number of training trajectories for Alanine Dipeptide.

- **Interpolation between off-equilibrium models and the equilibrium distribution:**

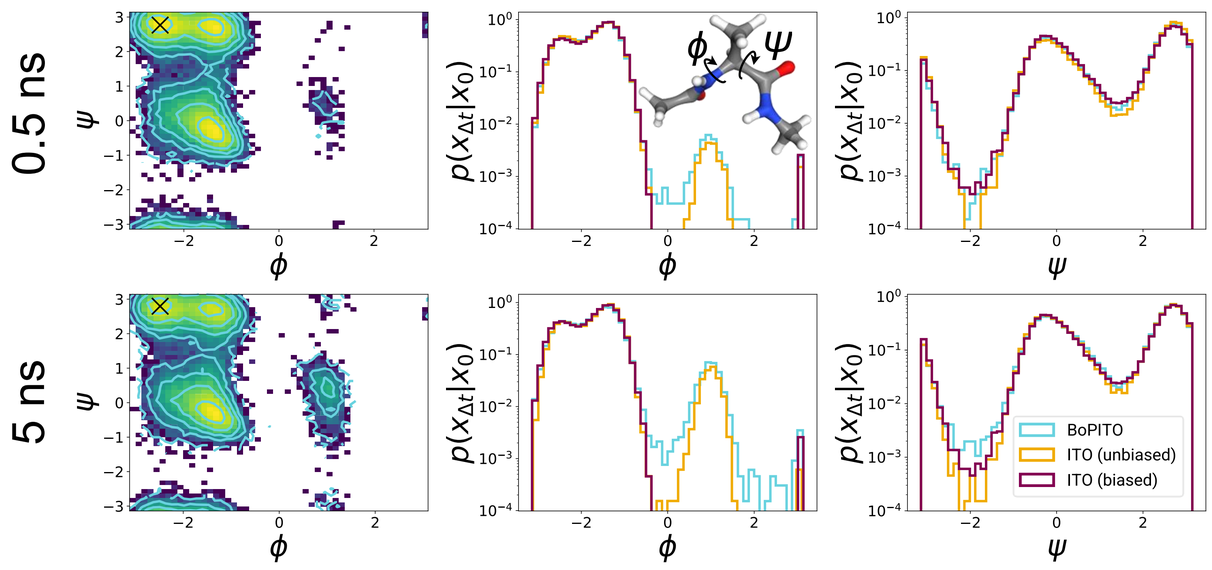


Figure 3: BoPITO can incorporate unbiased dynamic observables to correct a model trained on biased data. Rows of increasing time-lag (from top to bottom). Contour plots correspond to a BoPITO interpolator. The black cross indicates the initial condition.

- **Conclusions and next steps:**
- BoPITO improves sample efficiency and enables interpolation between off equilibrium and equilibrium models.
- We are working on making ITO models transferable across different molecules.