

- Noé, F., Olsson, S., Köhler, J., and Wu, H. Boltzmann generators — sampling equilibrium states of many-body systems with deep learning. *Science*, 365:eaaw1147, 2019.
- Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of chemical physics*, 139(1):07B604_1, 2013.
- Plattner, N., Doerr, S., De Fabritiis, G., and Noé, F. Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature chemistry*, 9(10):1005–1011, 2017.
- Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J. D., Schütte, C., and Noé, F. Markov models of molecular kinetics: Generation and validation. *The Journal of chemical physics*, 134(17):174105, 2011.
- Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, pp. 95–99, 1963.
- Rasley, J., Rajbhandari, S., Ruwase, O., and He, Y. Deep-speed: System optimizations enable training deep learning models with over 100 billion parameters. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD ’20, pp. 3505–3506, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379984. doi: 10.1145/3394486.3406703. URL <https://doi.org/10.1145/3394486.3406703>.
- Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.
- Rezende, D. J., Racanière, S., Higgins, I., and Toth, P. Equivariant Hamiltonian flows. *arXiv preprint arXiv:1909.13739*, 2019.
- Song, J., Zhao, S., and Ermon, S. A-NICE-MC: Adversarial training for MCMC. *Advances in Neural Information Processing Systems*, 30, 2017.
- Swope, W. C., Pitera, J. W., and Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. theory. *The Journal of Physical Chemistry B*, 108(21): 6571–6581, 2004.
- Titsias, M. and Dellaportas, P. Gradient-based adaptive Markov chain Monte Carlo. *Advances in Neural Information Processing Systems*, 32, 2019.
- Tsai, Y.-H. H., Bai, S., Yamada, M., Morency, L.-P., and Salakhutdinov, R. Transformer dissection: An unified understanding for transformer’s attention via the lens of kernel. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 4344–4353, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1443. URL <https://aclanthology.org/D19-1443>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.
- Winkler, C., Worrall, D., Hoozeboom, E., and Welling, M. Learning likelihoods with conditional normalizing flows. *arXiv preprint arXiv:1912.00042*, 2019.
- Wu, H., Mardt, A., Pasquali, L., and Noé, F. Deep generative Markov state models. *Advances in Neural Information Processing Systems*, 31, 2018.
- Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=PzcvxEMzvQC>.