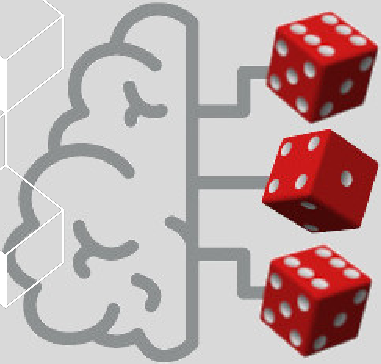
Uncertainty quantification in deep learning for physical simulation



Eiji Kawasaki - SFdS
Uncertainty quantification in machine learning
January 6th 2024

1. Introduction

Physical Simulation & Nuclear Industry

The nuclear industry studies the physics of stochastic processes (multiscale methods) :

- Particle Transport - Quantum Mechanics
- Complex Systems - Statistical Mechanics

Statistical Mechanics describes the behavior of a macroscopic system in terms of the behavior of its microscopic elements through the application of probability.

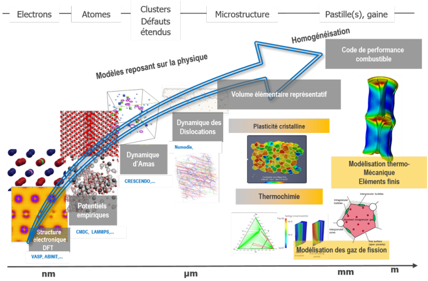


Figure 1 : extracted from M.Karcz's PhD manuscript

Statistical Mechanics 101

We want to compute a physical quantity $\mathcal{O}$. It corresponds to computing an average over all possible physical states x

$$\langle \mathcal{O} \rangle = \int dx\, p(x)\mathcal{O}(x) \simeq \sum_{i=1}^{N} \frac{1}{N}\mathcal{O}(x^{(i)})$$

where $x^{(i)} \sim p(x) \propto e^{-E(x)/T}$ and $E(x)$ is the energy of the state x.

- The computation of the energy $E(x)$ can be expensive.
- The sampling of $p(x)$ can be difficult (local extrema, high dimensions, etc...).

Note : computing $\langle \mathcal{O} \rangle$ using deterministic forces does not solve the challenges : $x_{t+1} = x_t - \nabla E(x_t) + \sqrt{2T}\epsilon$ where ϵ is a Gaussian rv.

ML into the Physical Simulation

The idea is to include ML cheap estimates into the numerical computation.

Either

- Replace the computation of the energy by $E_\theta(x)$.
- Replace the sampling and/or the pdf by $p_\theta(x)$.

θ typically correspond to the parameters of a neural network.

What is the impact of this change in the computation of physical quantities?

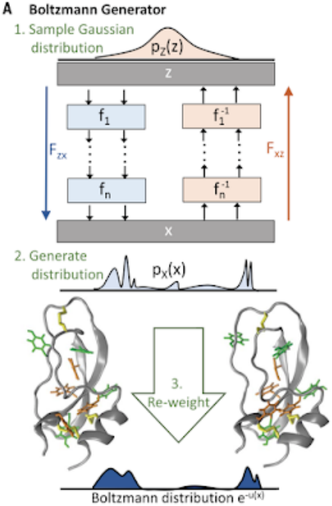


Figure extracted from ….

Uncertainty in AI driven simulation

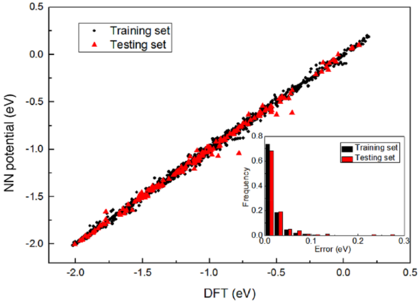
Modeling a physical phenomenon

- Using machine learning, we expect a **difference between the model prediction and the actual value.**
- We can compute what is the likelihood of an outcome given that some aspects of the system are not exactly known. What is the impact of this **uncertainty** in a physical computation ?

Application examples : ML interatomic potentials, Boltzmann generators. Phase diagram - phase transition

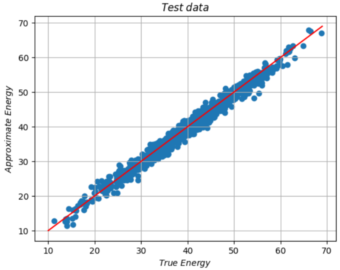
2. Learning the Energy
- D.Tzivrailis (2023-2026)

Uncertainty in the Energy Prediction



The goal here is to accelerate the physical simulation by learning how to estimate the physical energy $E(x)$ that can be expensive to compute. In Watanabe et al., 2021, it can be seen that the tail of the residual distribution is heavy.

Energy Learning Task



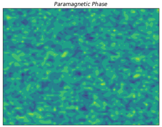
Regression Task

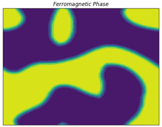
- We design and train a Neural Network that **computes an approximation** of the energy $E_\theta(x)$.
- It is trained by minimizing the loss

$$\sum_i (E_\theta(x_i) - E_i)^2$$

computed using the data set $\mathcal{D} = \{\ldots, (x_i, E_i), \ldots\}$

As a toy model, we consider a physical system containing a phase transition between ordered and disordered x states.





Physical Simulation Failure

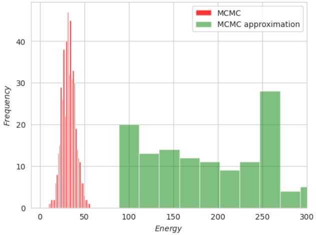
We run a Markov Chain Monte Carlo computation, where the Metropolis-Hastings acceptance writes

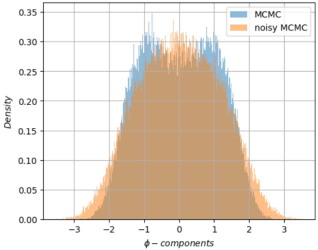
$$a(x \to x') = \min\left(1,\ e^{-\frac{E_\theta(x')}{k_B T}} / e^{-\frac{E_\theta(x)}{k_B T}}\right) \qquad (1)$$

Naively replacing the exact value of $E(x)$ by the Neural Network approximation $E_\theta(x)$, we notice that the physical computation is **not correct** !

This is due to :

- the (epistemic) noise in the prediction
- "out-of-distribution" inputs
- input domain shift

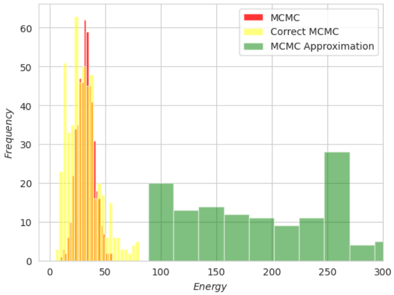




Physical Simulation Fix

We need to compute the prediction uncertainty $p(E|x, \mathcal{D})$ which is difficult for deep learning models (examples of methods include Ensemble, BNN, Last Layer, Deep Kernel Regression). There are many methods to take into account a noisy estimate of the target pdf in a MCMC sampling scheme :

- MH algorithm generalization based on assumption over the noise $\sigma(x)$
- Bayesian inference (Bardenet et al., 2017) : divide-and-conquer, subsampling-based



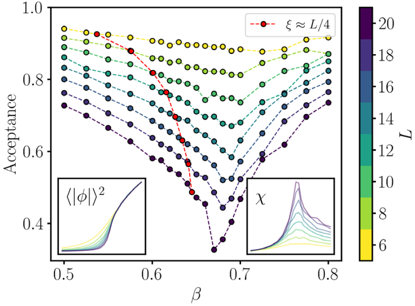
By taking into account the expected difference

$$\sigma(x) = E_\theta(x) - E(x) \tag{2}$$

while modeling the **predictive uncertainty** of the energy we can correct the physical computation.

3. Learning the distribution - M.Karcz (2021-2024)

Physical Simulation Failure

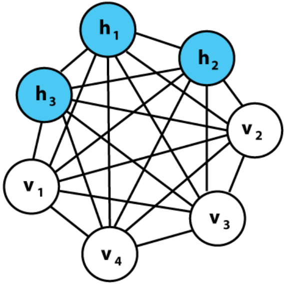


The acceptance rate of configurations proposed by a normalizing flow in a Metropolis-Hastings algorithm is close to 100% for a small system, and drops to less than 40% for a medium-sized system. *Figure extracted from Del Debbio et al., 2021.*

Physical Simulation vs Statistical Learning

Physical Phenomena Modeling

- **Numerical Simulation** : knowledge in the form of simulated / solved laws and equations.
- **Machine Learning** : based on a collection of observations used to optimize the parameters of an effective model.



Machine Learning Errors

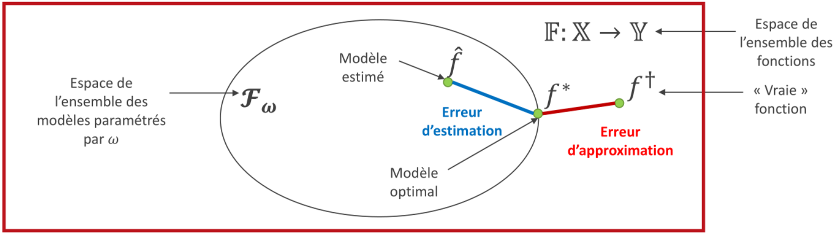


Figure 2 : borrowed from G. Daniel.

Main idea : make an approximation of the target system by an effective variational model where we optimize the parameters.

What about in physics ?

Statistical Physics and Statistical Learning

Variational approaches are ubiquotous in physics. For example, we optimize the variational parameters θ by minimizing the variational free energy $\tilde{F}_{p_\theta}$

$$p(x) = \frac{e^{-\frac{E(x)}{k_B T}}}{Z} \qquad -k_B T \ln Z = F_p \leq \tilde{F}_{p_\theta} \tag{3}$$

This inequality is known as the variational principle of statistical mechanics. Computing the free energy is an intractable problem for all but the simplest models. A powerful approximation method is mean-field theory.

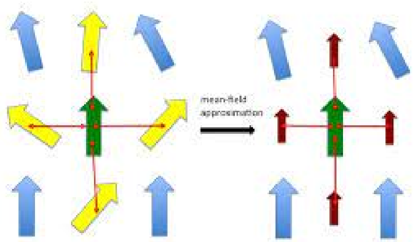
We use a non interacting system,

$$E_\theta^{\mathrm{MF}}(x) = \sum_i E_\theta(x_i) \tag{4}$$

Therefore,

$$p_\theta^{\mathrm{MF}}(x) = \prod_i p_\theta(x_i) \propto \prod_i e^{-E_\theta(x_i)} \tag{5}$$

where we can analytically optimize θ.



Inverted Variational Auto-Encoder - arXiv :2408.14928

We can improve the mean field approximation by taking into account interactions, we need to introduce some auxiliary variables y.

$$p_\theta(x) = \int p(y)p_\theta(x|y)dy \neq \prod_i p_\theta(x_i) \tag{6}$$

where $p(y) = \prod_i p(y_i)$ is some arbitrary tractable distribution. For example,

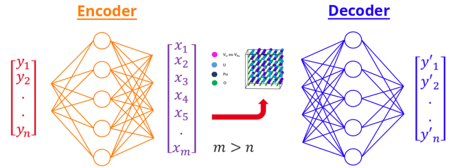
$$p_\theta(x|y) \propto \prod_i e^{-x_i\theta_i(y)} \tag{7}$$

where the function $\theta(z)$ is a neural network. We have,

$$F_p \leq \tilde{F}_{p_\theta} = \mathbb{E}_{\substack{y\sim p(y) \\ x\sim p_\theta(x|y)}} \left[\frac{-E(x)}{k_B T} + \log \frac{q_\phi(y|x)}{p(y)p_\theta(x|y)} \right] \tag{8}$$

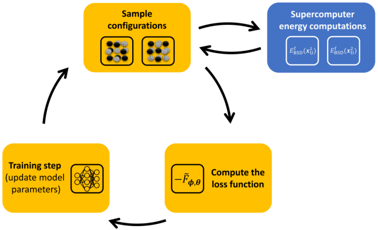
This objective is called the Evidence Lower Bound (ELBO) and it corresponds to the negative variational free energy in physics.

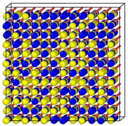
PULSE algorithm – arXiv :2408.14928

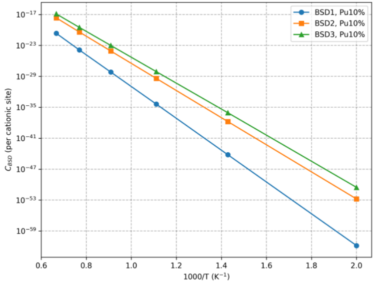


$$C_d(T) = \left\langle e^{-\frac{E^{f}_{d}(x_C)}{k_B T}} \right\rangle \qquad (9)$$

Average defect concentration in uranium-plutonium mixed oxides.







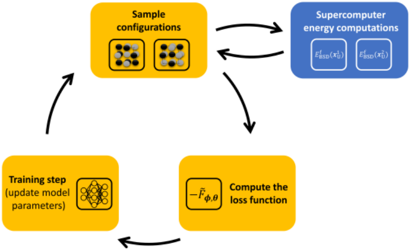
Based on this idea Karcz et al., 2024 : **Partition function Unsupervised Learning Sampling and Evaluation for disordered compounds.**

4. Conclusion

Conclusion

Summary

- **Variational approaches** are a link between physical simulation and machine learning that can be used : **PULSE** (Karcz et al., 2024)
- KARCZ Maciej PhD defence on October 2024 : "Approches d'apprentissage automatique génératif pour la modélisation à l'échelle atomique de composés présentant un désordre chimique"
- **Prediction Uncertainty Quantification** is necessary but can be difficult.



On Going Work & Open for collaborations !

- Link Physical Computation & Machine Learning
- What about **Generative Modeling Uncertainty** ?

Thanks to the collaborators : Emeric BOURASSEAU (CEA), Luca MESSINA (CEA), Alberto ROSSO (CNRS).

References

Bardenet, R., Doucet, A., & Holmes, C. (2017). On Markov chain Monte Carlo methods for tall data. *Journal of Machine Learning Research*, *18*(47), 1‑43. http://jmlr.org/papers/v18/15-205.html

Del Debbio, L., Marsh Rossney, J., & Wilson, M. (2021). Efficient modeling of trivializing maps for lattice ϕ 4 theory using normalizing flows : A first look at scalability. *Physical Review D*, *104*(9), 094507. https://doi.org/10.1103/PhysRevD.104.094507

Karcz, M. J., Messina, L., Kawasaki, E., & Bourasseau, E. (2024). Targeting the partition function of chemically disordered materials with a generative approach based on inverse variational autoencoders [arXiv :2408.14928 [cond-mat]]. http://arxiv.org/abs/2408.14928

Watanabe, S., Li, W., Jeong, W., Lee, D., Shimizu, K., Mimanitani, E., Ando, Y., & Han, S. (2021). High-dimensional neural network atomic potentials for examining energy materials : some recent simulations. *Journal of Physics : Energy*, *3*(1), 012003. https://doi.org/10.1088/2515-7655/abc7f3

Thank you for your attention !

CEA SACLAY- LIST/DIN
Département Instrumentation Numérique
91191 Gif-sur-Yvette Cedex