Machine Learning Trivializing Maps

THE UNIVERSITY of EDINBURGH

Luigi Del Debbio • Joe Marsh Rossney • Michael Wilson | University of Edinburgh, UK

The Task : Estimate expectation values by sampling

$$\langle \mathcal{O} \rangle \approx \bar{\mathcal{O}} = \frac{1}{N_\phi} \sum_{\{\phi\}} \mathcal{O}(\phi) \qquad \phi \sim p(\phi) = \frac{1}{Z} e^{-S(\phi)}$$

The Problem : Efficiency of traditional sampling is enormously diminished as we take continuum limit $\xi \to \infty$

$$\mathrm{SE}_{\bar{\mathcal{O}}}(\{\phi\}) = \sqrt{\frac{2\tau^{\mathrm{int}}_{\mathcal{O}}(\{\phi\})}{N_\phi}}\sigma_{\mathcal{O}} \qquad \tau^{\mathrm{int}}_{\mathcal{O}} \propto \xi^{z_{\mathcal{O}}}$$

Critical Slowing Down

The Idea : Machine-learn an invertible map to a limit where sampling is easy (Albergo, Kanwar, Shanahan [1904.12072])

$$z \overset{\mathcal{F}}{\longmapsto} \phi \qquad z \sim r(z) \overset{\text{e.g.}}{=} \prod_{x \in \Lambda} \mathcal{N}(0,1)$$

$\mathcal{F}^{-1}$ is a Trivializing Map (Lüscher [0907.5491])

The Strategy : Approximate $\mathcal{F}$ using a Normalizing Flow f_θ

1. Train

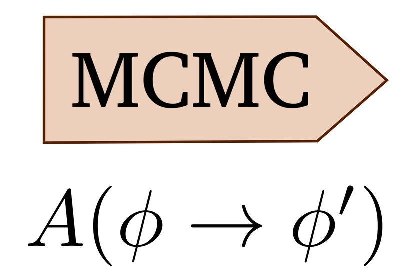
$$f_{\theta^\star} \approx \mathcal{F} \qquad \theta^\star = \underset{\theta}{\operatorname{argmin}} \underbrace{\int \mathcal{D}\phi \, \tilde{p}(\phi) \log \frac{\tilde{p}(\phi)}{p(\phi)}}_{D_{\mathrm{KL}}(\tilde{p}\|p)}$$

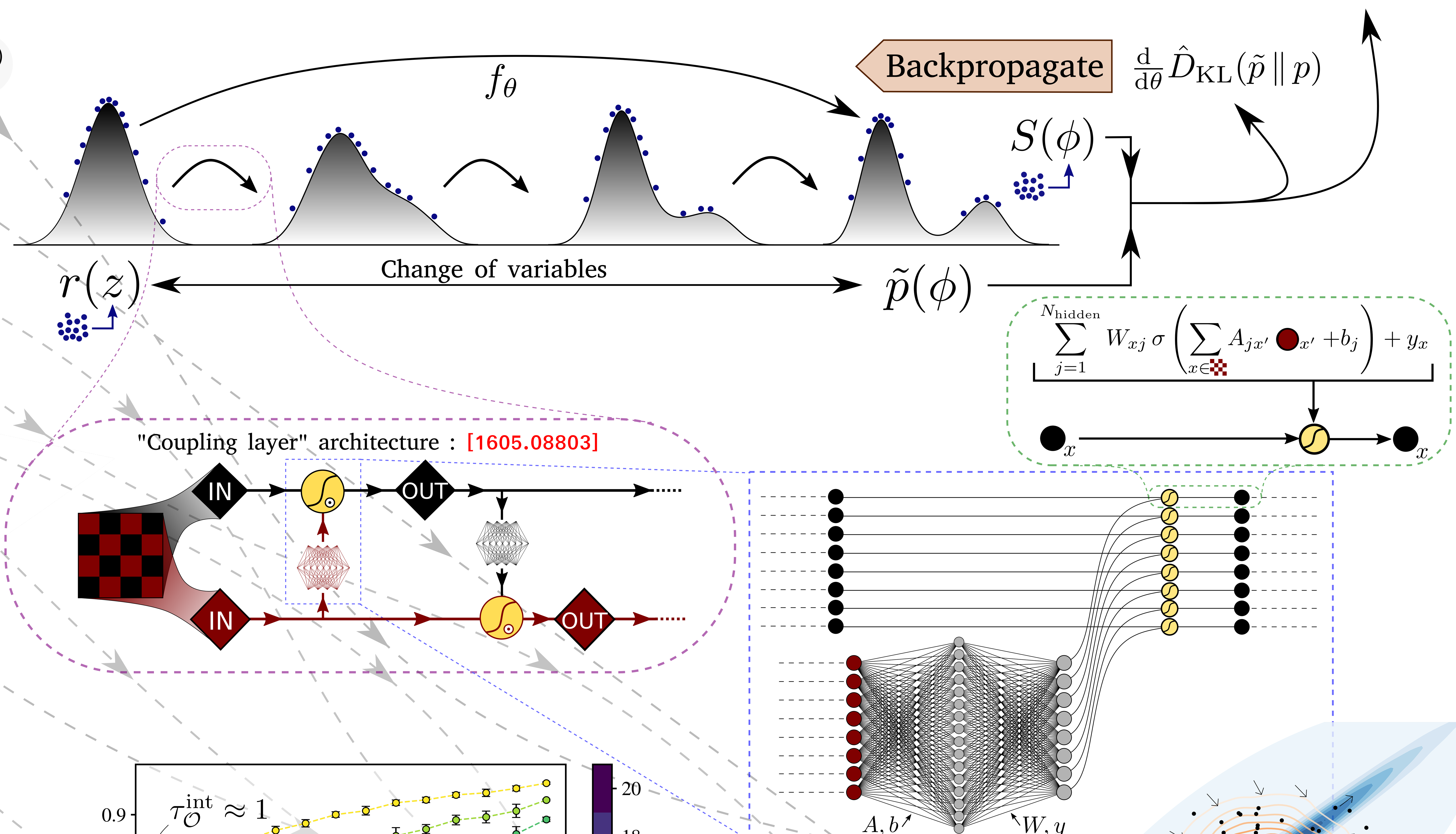
2. Generate

$$z \overset{f_\theta}{\longmapsto} \phi \qquad \log r(z) - \log |\mathbf{J}_{f_\theta}| = \log \tilde{p}(\phi)$$

3. Reweight

$$\tilde{p}(\phi) \overset{\text{MCMC}}{\longrightarrow} p(\phi) \qquad A(\phi \to \phi') = \min\left(1, \frac{\tilde{p}(\phi)}{\tilde{p}(\phi')} \frac{p(\phi')}{p(\phi)}\right)$$





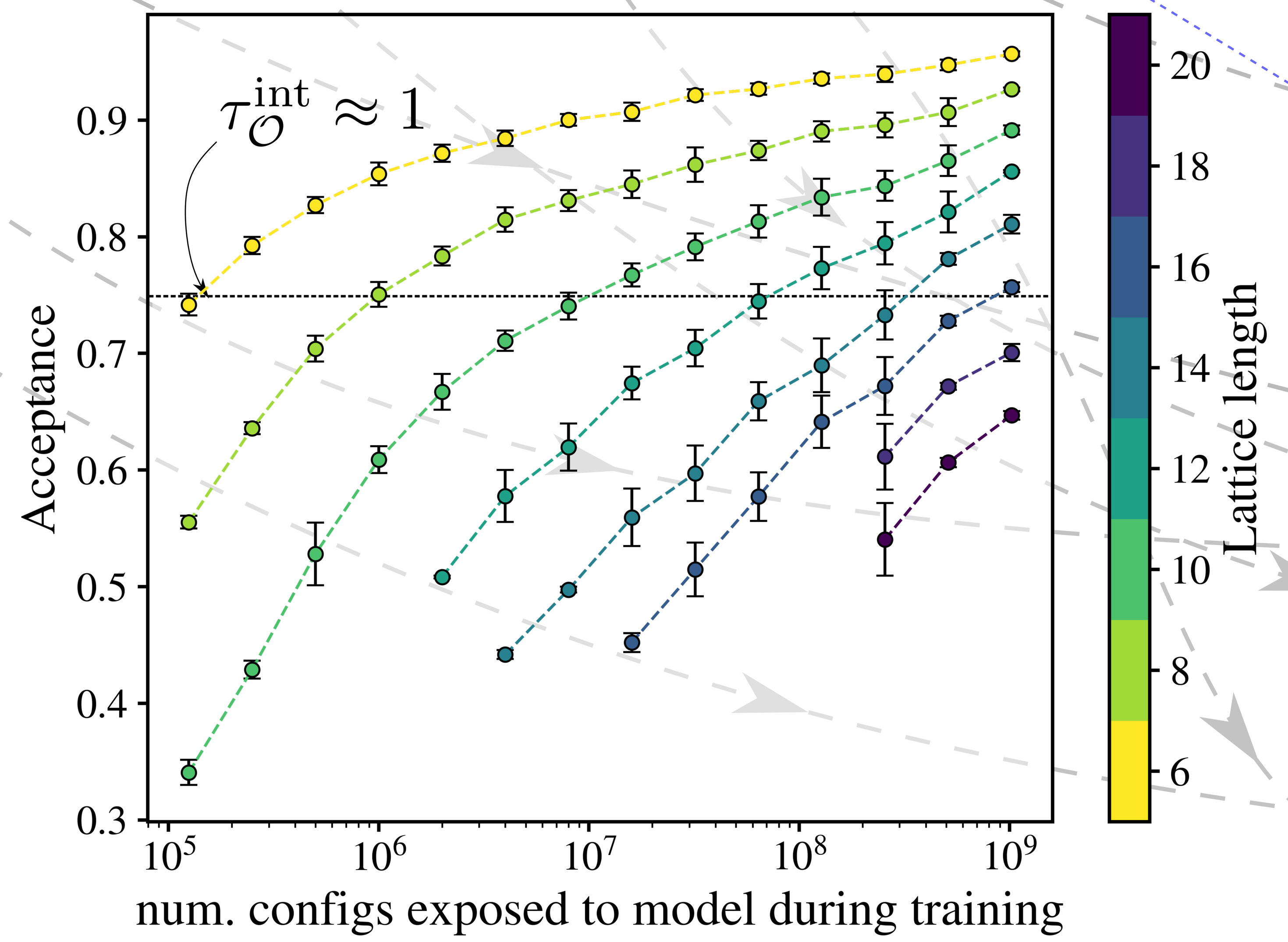
What have we been doing?

We've been experimenting with ϕ^4 [2105.12481] and σ models

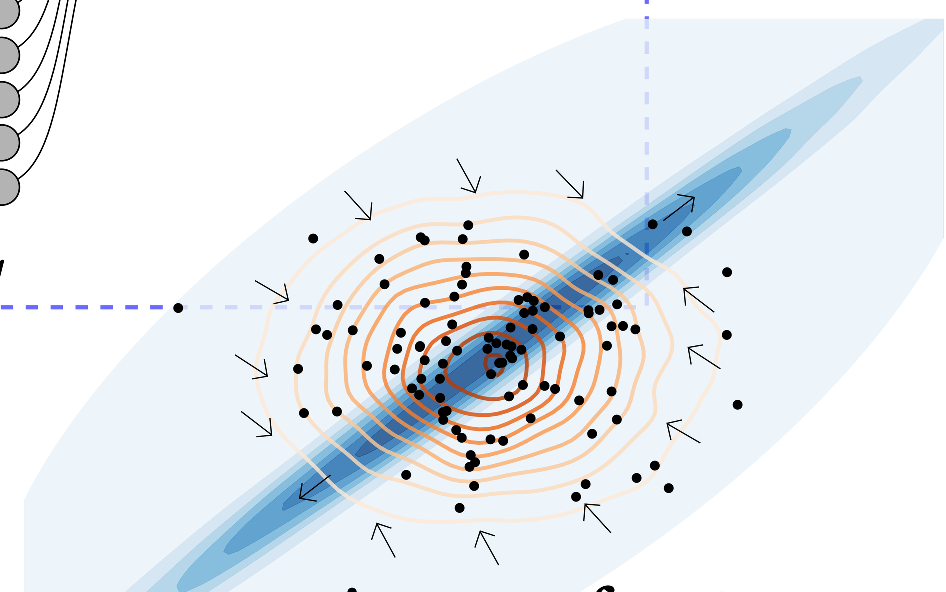
Much improved efficiency in ϕ^4 case using 3 key ingredients:

- Flexible spline transformations
- Shallow neural nets
- Approximate equivariance of f_θ under $\mathbb{Z}_2$

With this recipe, the sampling efficiency is almost entirely dictated by amount of training, not model expressivity...



...but training costs are growing very quickly!



Why are training costs growing so fast?

We're working on it. Ideas are most welcome!

Suggestion (needs validation!): increasingly ill-conditioned covariance of target causes "Critical Slowing Down" in training

Training scheme ('reverse' Kullback-Leibler) known to be slow to fit low-density regions. Can become exponentially (in num. d.o.f) slow for highly structured targets [Huang et al., UAI 2019]

Some alternatives/augmentations explored recently [2107.00734]

A promising novel approach to efficient sampling in Lattice Field Theory...but more work required to understand high training costs!