Gibbs Sampling of Periodic Potentials on a Quantum Computer

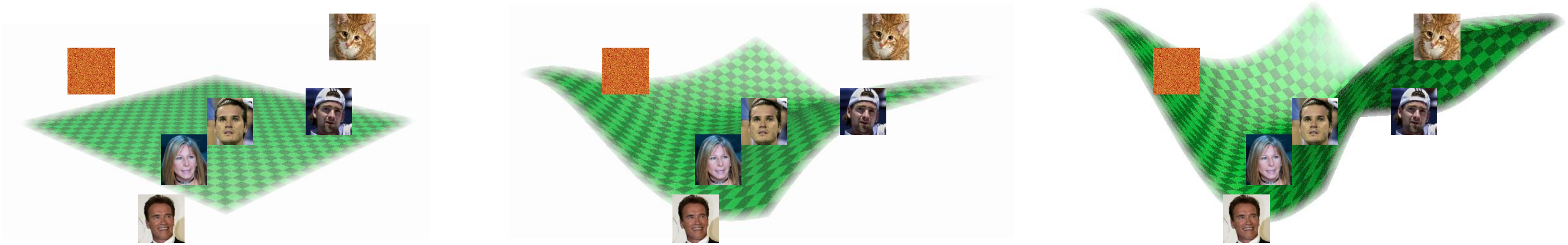
Reference

Gibbs Sampling of Periodic Potentials on a Quantum Computer. arXiv:2210.08104 (2022),
Joint with Arsalan Motamedi.

Pooya Ronagh
Research Assistant Professor | IQC | University of Waterloo | Perimeter Institute
CTO | 1QBit

Energy-based models

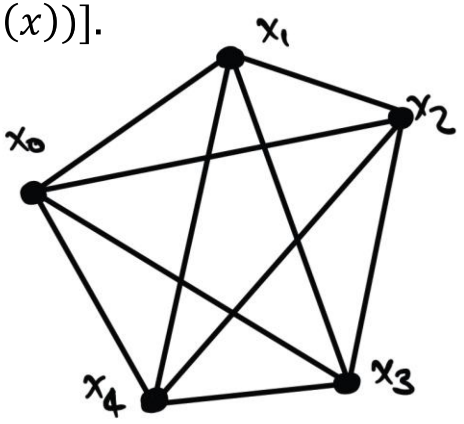
- In an energy-based model the model distribution is a Gibbs distribution $p_\lambda(x) = \frac{1}{Z_{\lambda,\beta}} e^{-\beta E_\lambda(x)}$. The model provides a representation of the **energy potential** $E_\lambda: \Omega \to \mathbb{R}$.



- Training:

$$-\nabla_\lambda \mathcal{L}(\lambda) = \int_x \nabla_\lambda \log p_\lambda(x)\, p(x) dx = \cdots = -\beta \mathbb{E}_p[\nabla_\lambda(E_\lambda(x))] + \beta \mathbb{E}_{p_\lambda}[\nabla_\lambda(E_\lambda(x))].$$

- The first term is called the **positive phase** (easy to approximate from data) and the second term is called the **negative phase** (requires samples from the model/Gibbs distribution).
- Old EBMs (e.g., Boltzmann machines) used the Ising potential $E_\lambda = -\sum_{i,j=1,\ldots,m} \lambda_{i,j} \sigma_i \sigma_j$ which didn't work out so great!

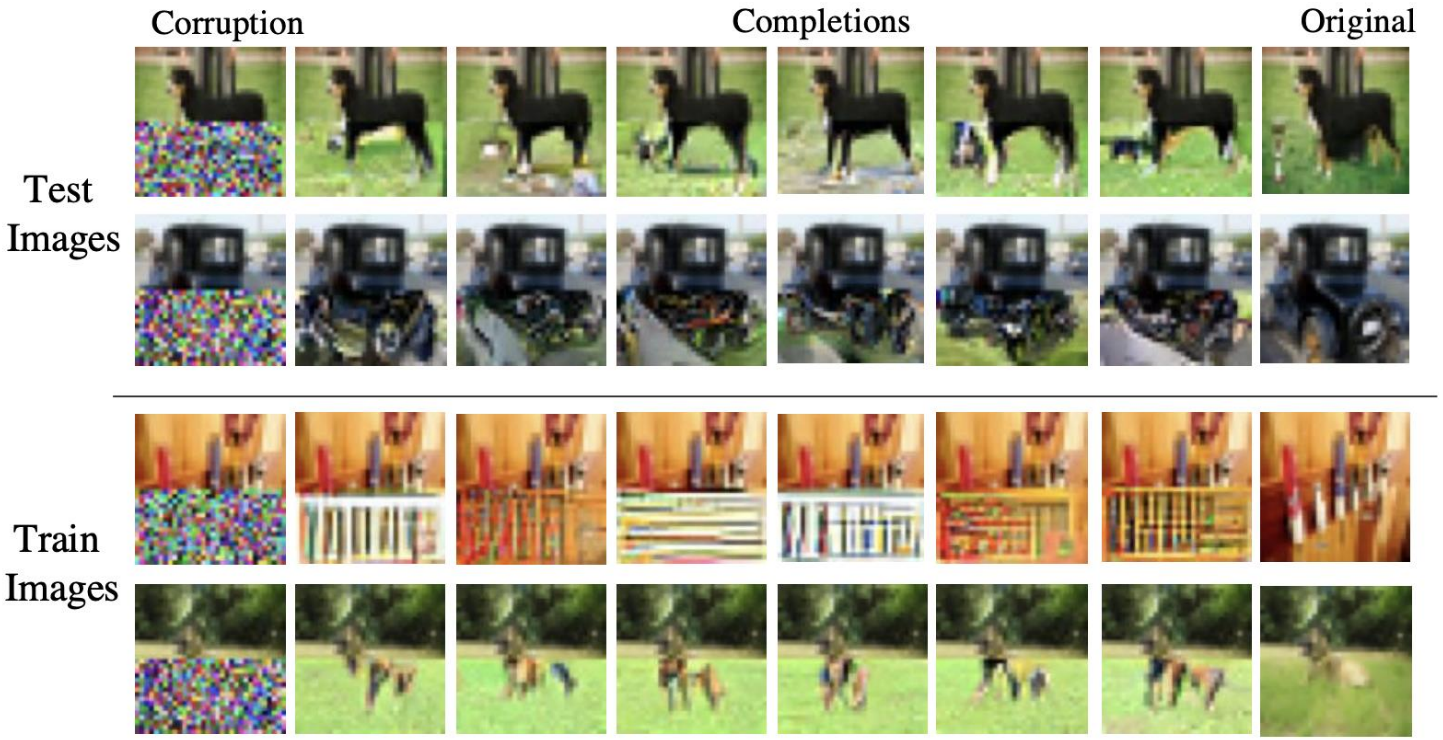


Modern (deep) EBMs

- Du and Mordatch (2019) showed the first large-scale EBM trainings.
- Langevin dynamics:

$$dX_t = -\frac{1}{2}\beta\sigma^2\nabla_X E_\theta(X)\,dt + \sigma\,dW_t.$$

Models	Parameters	Training Time	Sampling Time
EBM	5M	48	3 Hour (Variable)
PixelCNN++	160M	1300	72 Hour
Glow	115M	1300	0.5 Hour
SNGAN	5M	9	0.02 Hour



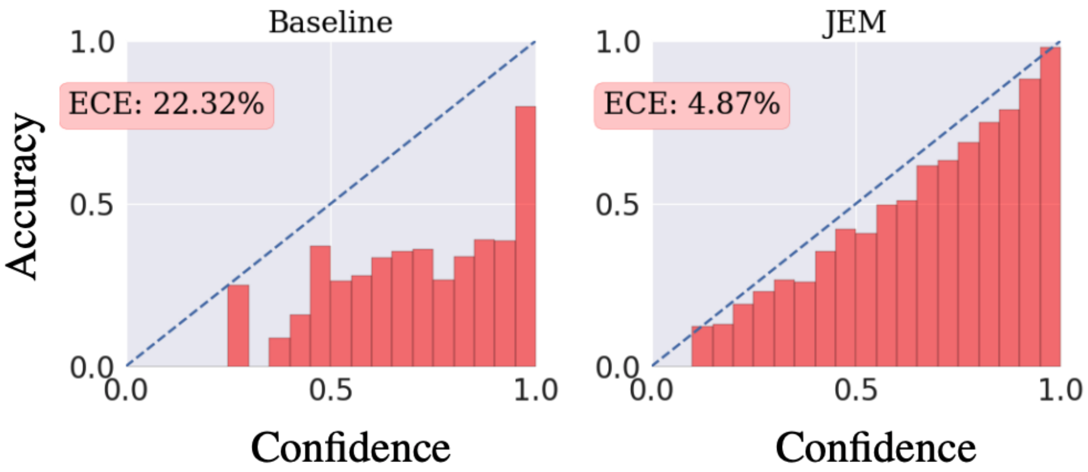
Model	Inception*	FID
CIFAR-10 Unconditional		
PixelCNN [Van Oord et al., 2016]	4.60	65.93
PixelIQN [Ostrovski et al., 2018]	5.29	49.46
EBM (single)	6.02	40.58
DCGAN [Radford et al., 2016]	6.40	37.11
WGAN + GP [Gulrajani et al., 2017]	6.50	36.4
EBM (10 historical ensemble)	6.78	38.2
SNGAN [Miyato et al., 2018]	**8.22**	21.7
CIFAR-10 Conditional		
Improved GAN	8.09	-
EBM (single)	8.30	37.9
Spectral Normalization GAN	**8.59**	25.5
ImageNet 32x32 Conditional		
PixelCNN	8.33	33.27
PixelIQN	10.18	22.99
EBM (single)	**18.22**	**14.31**
ImageNet 128x128 Conditional		
ACGAN [Odena et al., 2017]	28.5	-
EBM* (single)	28.6	43.7
SNGAN	**36.8**	**27.62**

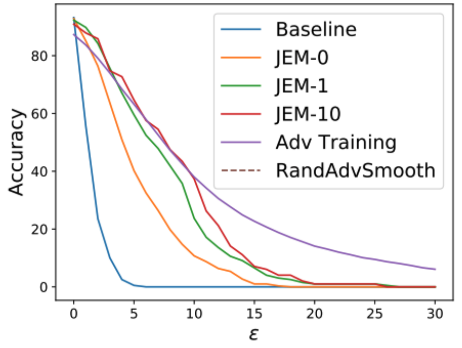
Benefits of EBMs

- Guo et al. (2017): Over the years conventional classifiers have become less and less calibrated.
- Grathwohl et al. (2019): Joint EBM improve the calibration and adversarial robustness of the representation.
- Du and Mordatch (2019): Good at OOD detection by EBMs.

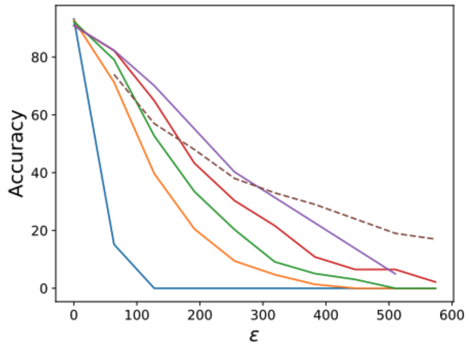
Model	PixelCNN++	Glow	EBM (ours)
SVHN	0.32	0.24	**0.63**
Textures	0.33	0.27	**0.48**
Constant Uniform	0.0	0.0	**0.30**
Uniform	1.0	1.0	**1.0**
CIFAR10 Interpolation	**0.71**	0.59	0.70
Average	0.47	0.42	**0.62**

Figure 10: AUROC scores of out of distribution classification on different datasets. Only our model gets better than chance classification.

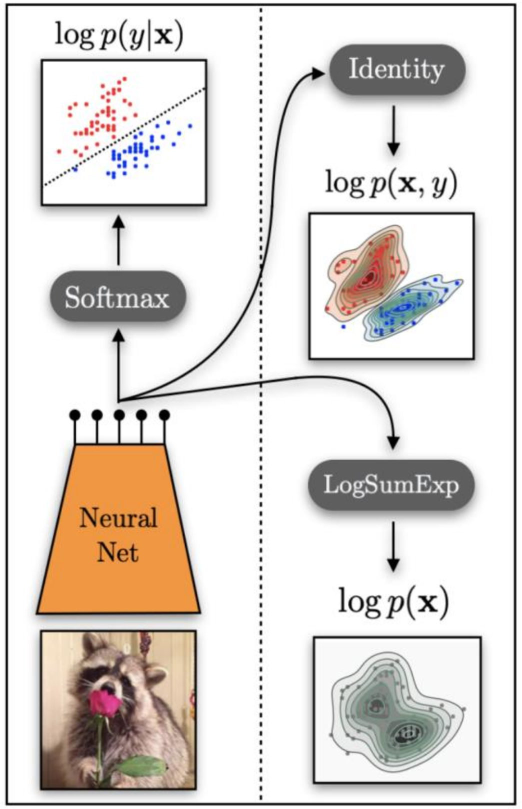




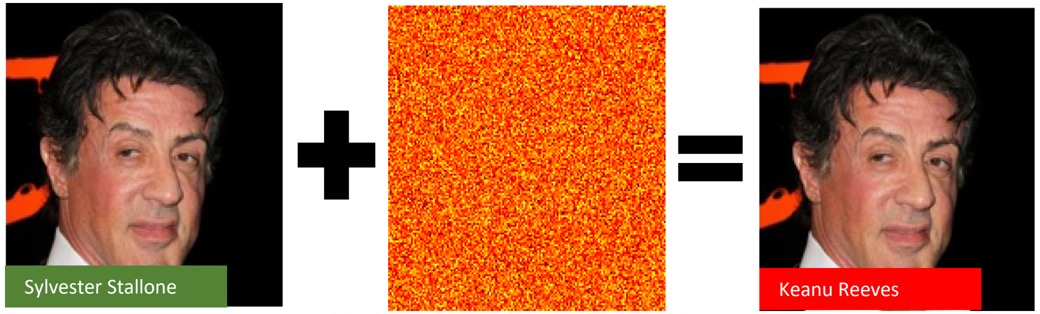
(a) L_∞ Robustness



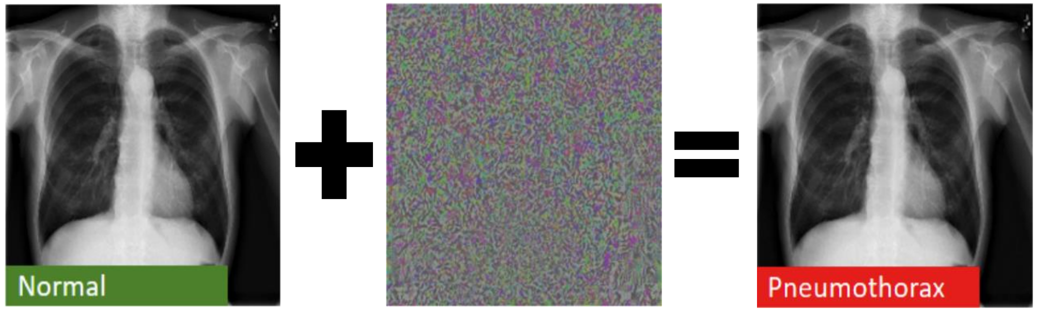
(b) L_2 Robustness



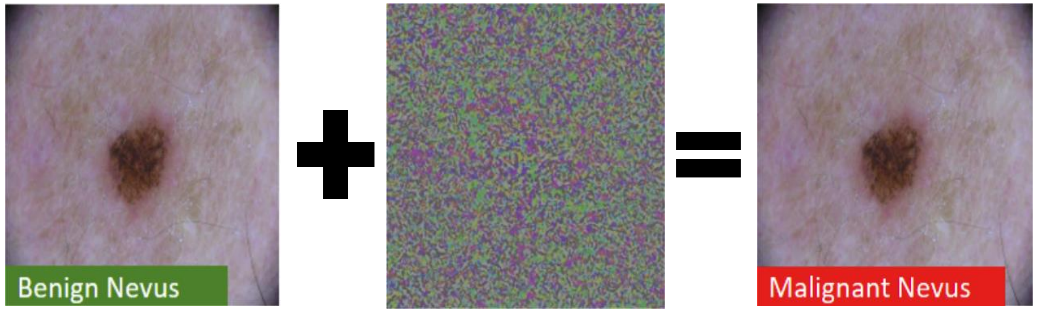
Stochastic security



Mikhailov, Trusov 2017, blog.ycombinator.com

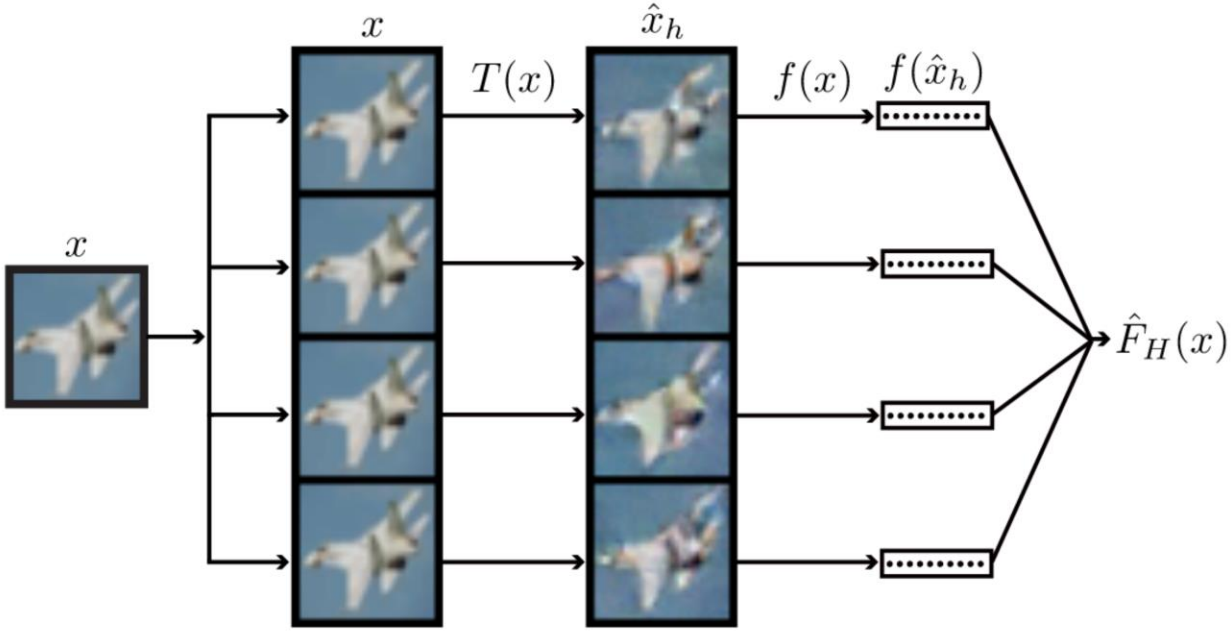


Ma et al. 2020, arXiv:1907.10456v2



Ma et al. 2020, arXiv:1907.10456v2

- Elflein et al., (2021): EBMs outperform normalizing flows in OOD detection.
- Hill-Mitchel-Zho (2021): Convergent EBMs purify an image from adversarial attacks.

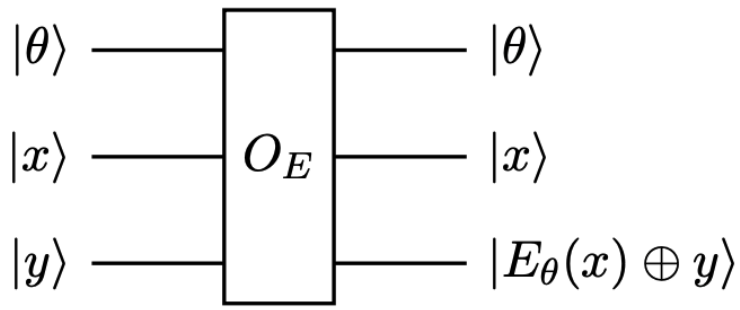


Our agenda

- Instead of Monte-Carlo simulation of Langevin dynamics $dX_t = -\nabla E(X_t)dt + \sqrt{2\beta^{-1}}dW_t$ …
- … solve the **Fokker-Planck equation**

$$\frac{\partial \rho}{\partial t} = \nabla.(\nabla E \rho) + \beta^{-1}\Delta\rho.$$

- A second order PDE; a.k.a. the diffusion equation; a.k.a. Kolmogorov's forward equation.
- Our approach: Linearize the generator $\mathcal{L} = -\nabla E.\nabla + \beta^{-1}\Delta$ and solve using quantum linear system solvers.
- Theorem [Motamedi-R]. A quantum computer can sample from the Gibbs state of an analytic periodic potential using $O(d^3 e^\beta \mathrm{polylog}(1/\epsilon))$ queries to the energy oracle.
- The training scheme does not require QRAM.
- Non convexity is now a non-issue and the complexity of algorithm depends rather on the rate of decay of Fourier coefficients of E.



High precision sampling using Fourier interpolation

Input: Energy function oracle O_E, lattice parameters $N, M \in \mathbb{N}$, solution time $T > 0$

1: Construct an oracle for the discretization $\mathbb{L}$ of the generator of the Fokker–Planck equation (see Fig. 4 in the appendix).

2: Deploy the algorithm of [BCOW17] to prepare a quantum state approximating $|u(T)\rangle$ pertaining to the solution of $\frac{d}{dt}\vec{u} = \mathbb{L}\vec{u}$, with $\vec{u}(0) = \mathbf{1}$, at time $t = T$.

3: Apply the upsampling isometry $F_M^{-1}\iota F_N$ involving quantum Fourier transformations on the prepared state (Theorem 3.1).

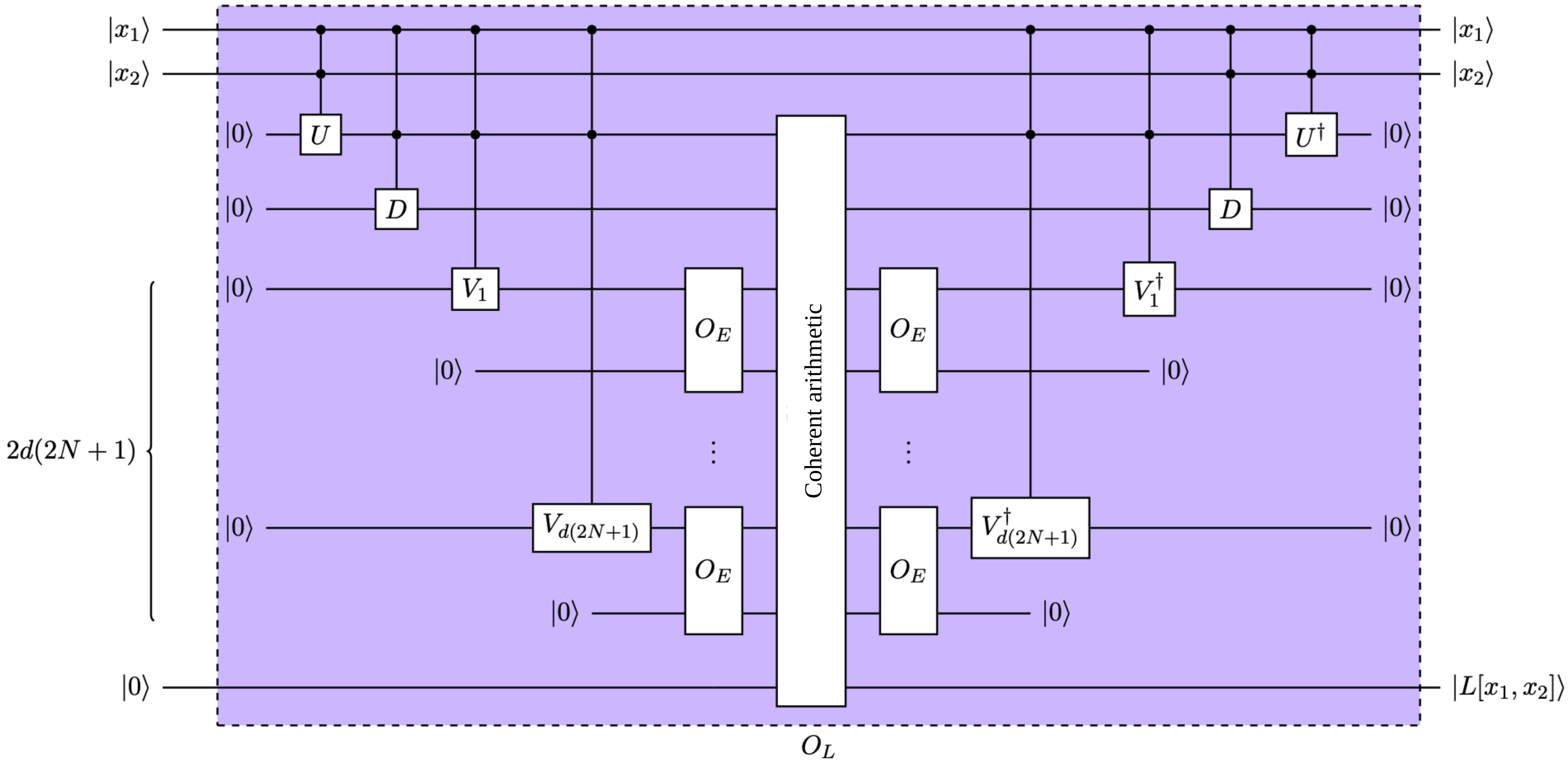
4: Measure the resulting state in the computational basis to obtain a lattice point $x \in [-\ell/2, \ell/2]^d$.

5: Draw a sample $\tilde{x}$ uniformly at random from the box $\prod_{i=1}^{d}\left[x_i - \frac{\ell}{4M+2}, x_i + \frac{\ell}{4M+2}\right]$ around x.

Output: Sample point $\tilde{x}$.

- By writing FPE in the Fourier domain we boost our precision with a technique we call ***upsampling***:
- Given a (C, a)-semi-analytic periodic function u, an integer $N \geq 2ad$, and a quantum state $|\psi_N\rangle \in H_N$ that is ϵ-close to $|u_N\rangle$ we can sample from u^2 with an $O(\epsilon, C)$ precision.

The oracle of the ODE solver



Long mixing time in equilibrium dynamics

- Let ρ_s be the corresponding steady distribution. Further, assume that for a constant $\lambda > 0$ any differentiable function $f \in L^2(\rho_s)$ satisfies

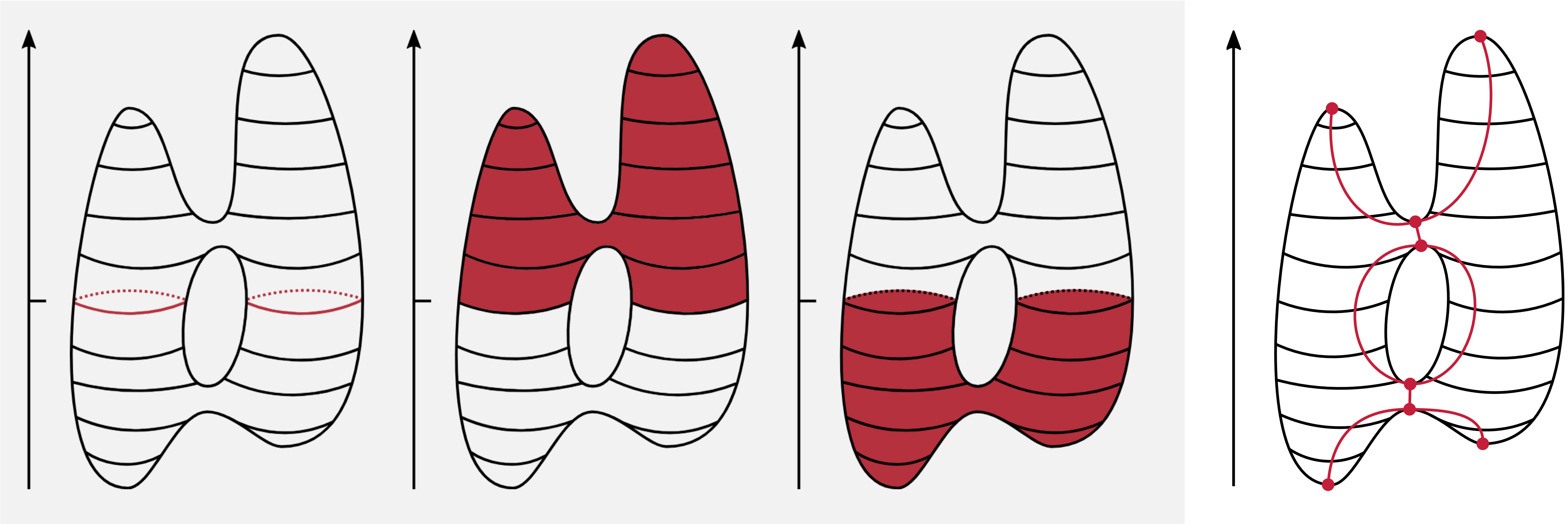
$$Var_{\rho_s}[f] \leq \kappa \mathbb{E}_{\rho_s}\left[|\nabla f|^{\mathbf{2}}\right]$$

- Then

$$\left\|\frac{\rho_t}{\rho_s} - 1\right\|_{L^2(\rho_s)} \leq e^{-\frac{t}{\kappa}} \left\|\frac{\rho_0}{\rho_s} - 1\right\|_{L^2(\rho_s)} .$$

- Note: $TV(\rho_t, \rho_s) \leq \frac{1}{2}\sqrt{Var_{\rho_s}(\rho_t/\rho_s)}$.
- Proposition [Motamedi-R]: Analytic period functions (on a torus) admit a universal Poincare constant $\kappa = O\left(e^{\Delta}\right)$.
- So, no speed up in terms of the mixing time. However, in practice Δ is bounded.

Exponential quantum speedups for Morse potentials



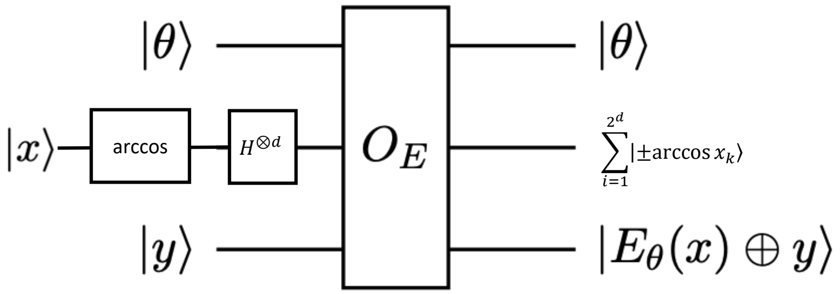
https://bastian.rieck.me/blog/posts/2019/morse_theory/

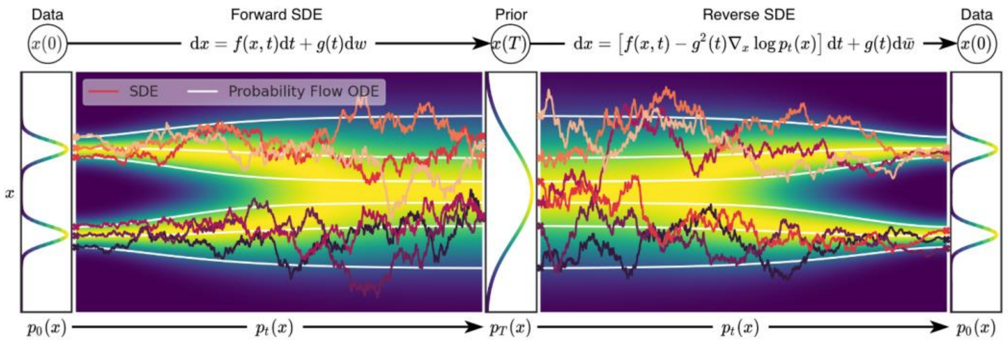
Comparison with other algorithms

Method	Potential type	Query order	Sampling complexity	Norm	Mean estimation complexity
This paper	non-convex periodic	zeroeth	$\widetilde{O}\left(\kappa_{E/2}e^{\Delta/2}d^7\right)$	TV	$\widetilde{O}\left(\kappa_{E/2}e^{\Delta/2}d^7\Delta_f\varepsilon^{-1}\right)$
Rejection sampling	non-convex	zeroeth	$O\left(e^{\Delta}\right)$	TV	$O\left(e^{\Delta}\Delta_f^2\varepsilon^{-2}\right)$
This paper	Morse and periodic	zeroeth	$\widetilde{O}\left(\lambda^{-2}e^{\Delta/2}d^7\right)$	TV	$\widetilde{O}\left(\lambda^{-2}e^{\Delta/2}d^7\Delta_f\varepsilon^{-1}\right)$
Cl. RLA [LE20]	Morse and periodic	first	$\widetilde{O}\left(\lambda^{-4}L^4d^3\varepsilon^{-2}\right)$	TV	$\widetilde{O}\left(\lambda^{-4}L^4d^3\Delta_f^2\varepsilon^{-4}\right)$
Cl. MRW [DCWY18]	convex	first	$\widetilde{O}\left(L^2d^3\varepsilon^{-2}\right)$	TV	$\widetilde{O}\left(L^2d^3\Delta_f^2\varepsilon^{-4}\right)$
Q. ULD [CLL+22]	strongly convex	zeroeth	$\widetilde{O}\left(\mu^{-2}L^2d^{1/2}\varepsilon^{-1}\right)$	W_2	–
Cl. ULD [CCBJ17]	strongly convex	first	$\widetilde{O}\left(\mu^{-2}L^2d^{1/2}\varepsilon^{-1}\right)$	W_2	–
Q. MALA [CLL+22]	strongly convex	first	$\widetilde{O}\left(\mu^{-1/2}L^{1/2}d\right)$	TV	$\widetilde{O}\left(\mu^{-1/2}L^{1/2}d\Delta_f\varepsilon^{-1}\right)$
Cl. MALA [LST20]	strongly convex	first	$\widetilde{O}\left(\mu^{-1}Ld\right)$	TV	$\widetilde{O}\left(\mu^{-1}Ld\Delta_f^2\varepsilon^{-2}\right)$

Outlook

- Outlook 1: Beyond periodic functions
  - Periodicity does not create a constrain in data-driven training.
  - Direct sampling from $[-1,1]^d$ can be achieved by moving to Chebychev spectral methods (WIP).
- Outlook 2:
  - In diffusion models the diffusion equation is solved for a finite time avoiding the long mixing time.
  - This requires solving FPEs with time-dependent drift and diffusion terms (WIP).
- Outlook 3:
  - Digital computation is very expensive for FTQC. Perhaps all this machinery should rather be developed in the bosonic setting (Please talk to me if interested…).





Quantum Physics

[Submitted on 26 Feb 2000 (v1), last revised 7 Jun 2000 (this version, v2)]

Quantum searching with continuous variables

Arun K. Pati, Samuel L. Braunstein, Seth Lloyd

Pooya Ronagh
pooya.ronagh@uwaterloo.ca

1QBit Mitacs

University of Waterloo | IQC Institute for Quantum Computing

NSERC CRSNG | Perimeter Institute

Open positions:
Postdoc 1 (neural quantum states)
Postdoc 2 (quantum algorithms)
PhD (quantum algorithms)