Trainability barriers and opportunities in quantum generative modeling

Manuel S. Rudolph,[1,*] Sacha Lerch,[1,*] Supanut Thanasilp,[1,2,*]
Oriel Kiss,[3,4] Sofia Vallecorsa,[3] Michele Grossi,[3] and Zoë Holmes[1]

[1] *Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2] *Chula Intelligent and Complex Systems, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand, 10330*
[3] *European Organization for Nuclear Research (CERN), Geneva 1211, Switzerland*
[4] *Department of Nuclear and Particle Physics, University of Geneva, Geneva 1211, Switzerland*

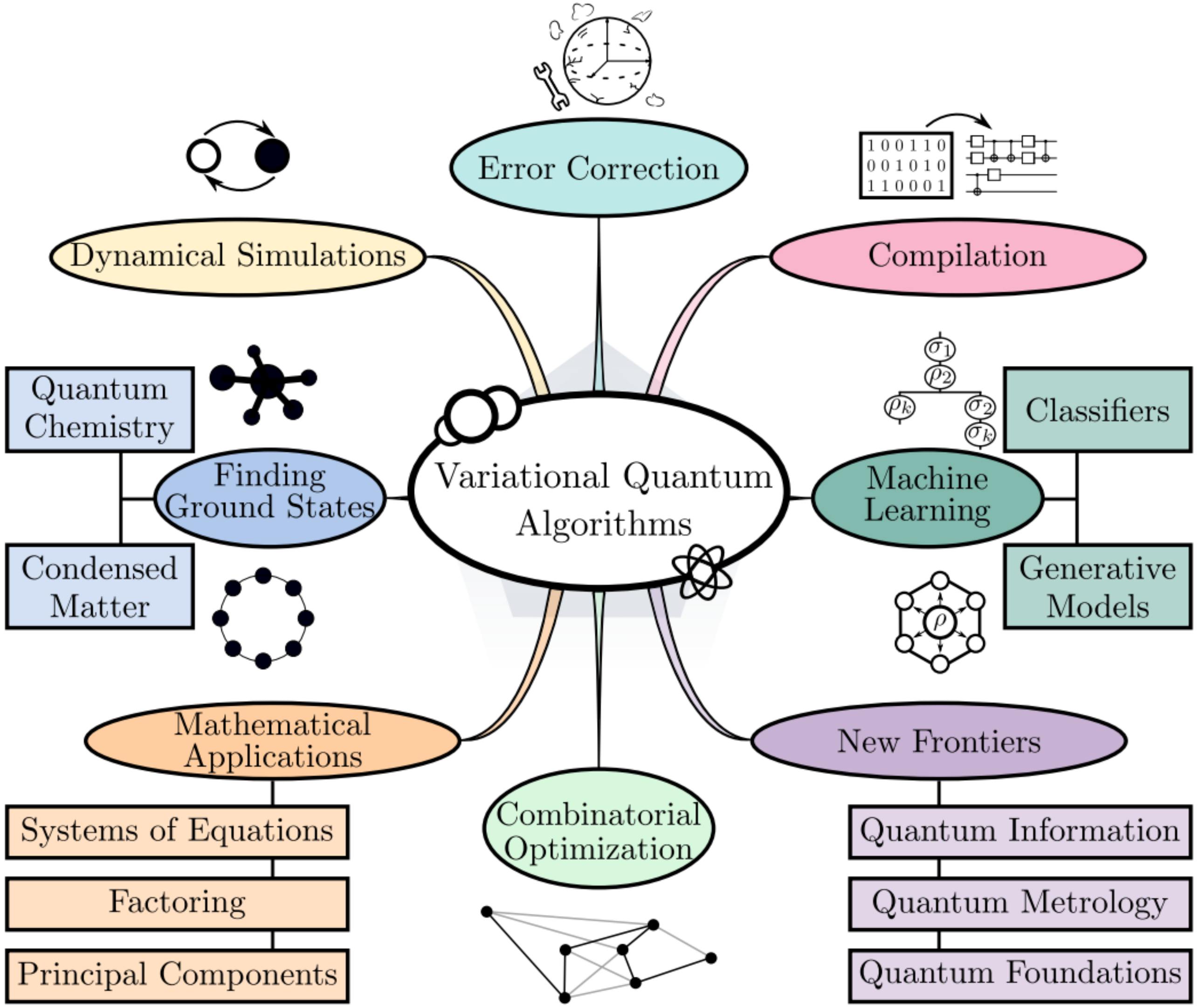
(Dated: May 5, 2023)

arXiv:2305.02881

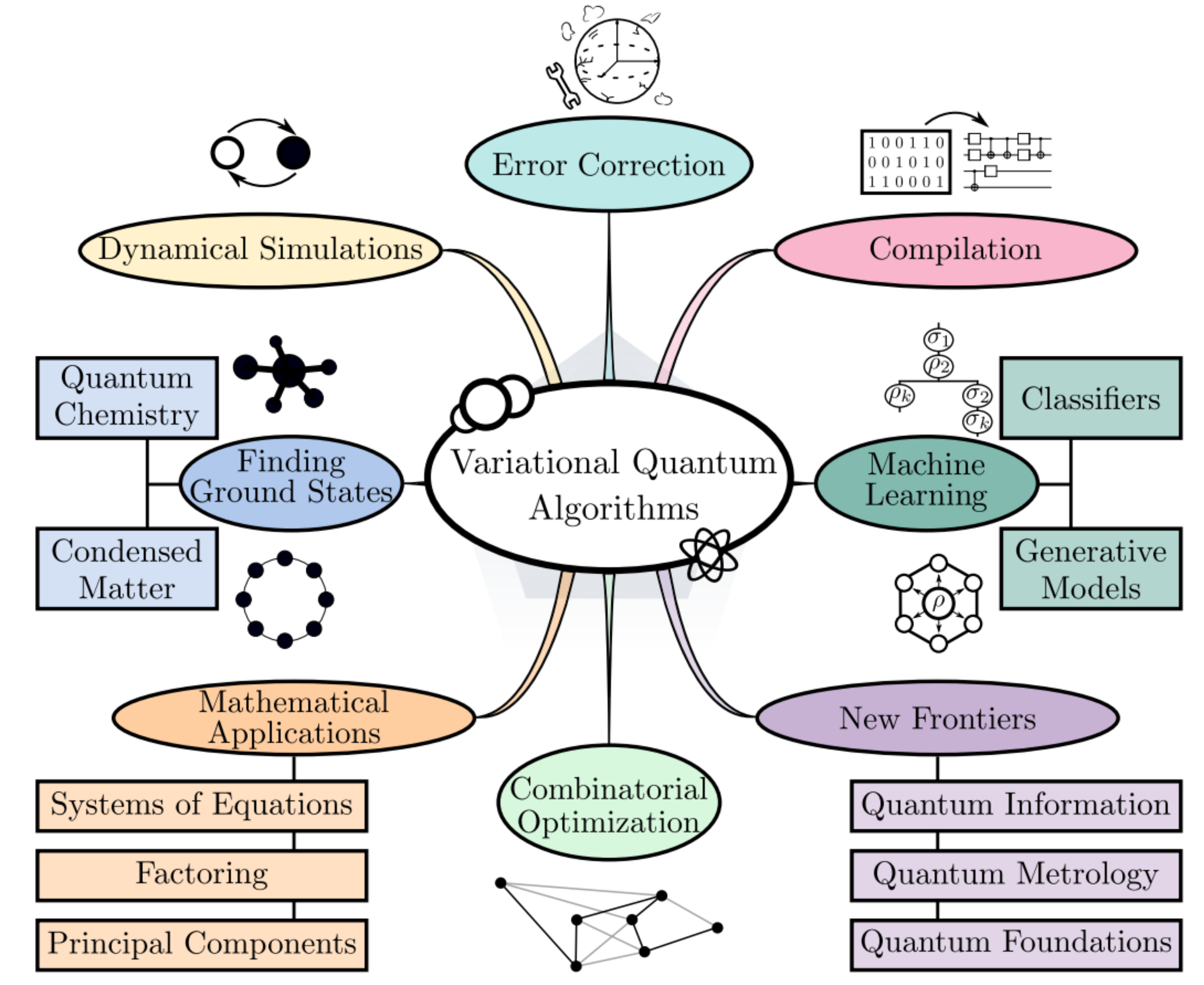
Manuel S. Rudolph

Quantum Information and Computing Lab

Cerezo, Marco, et al. "Variational quantum algorithms." Nature Reviews Physics 3.9 (2021)



Cerezo, Marco, et al. "Variational quantum algorithms." Nature Reviews Physics 3.9 (2021)

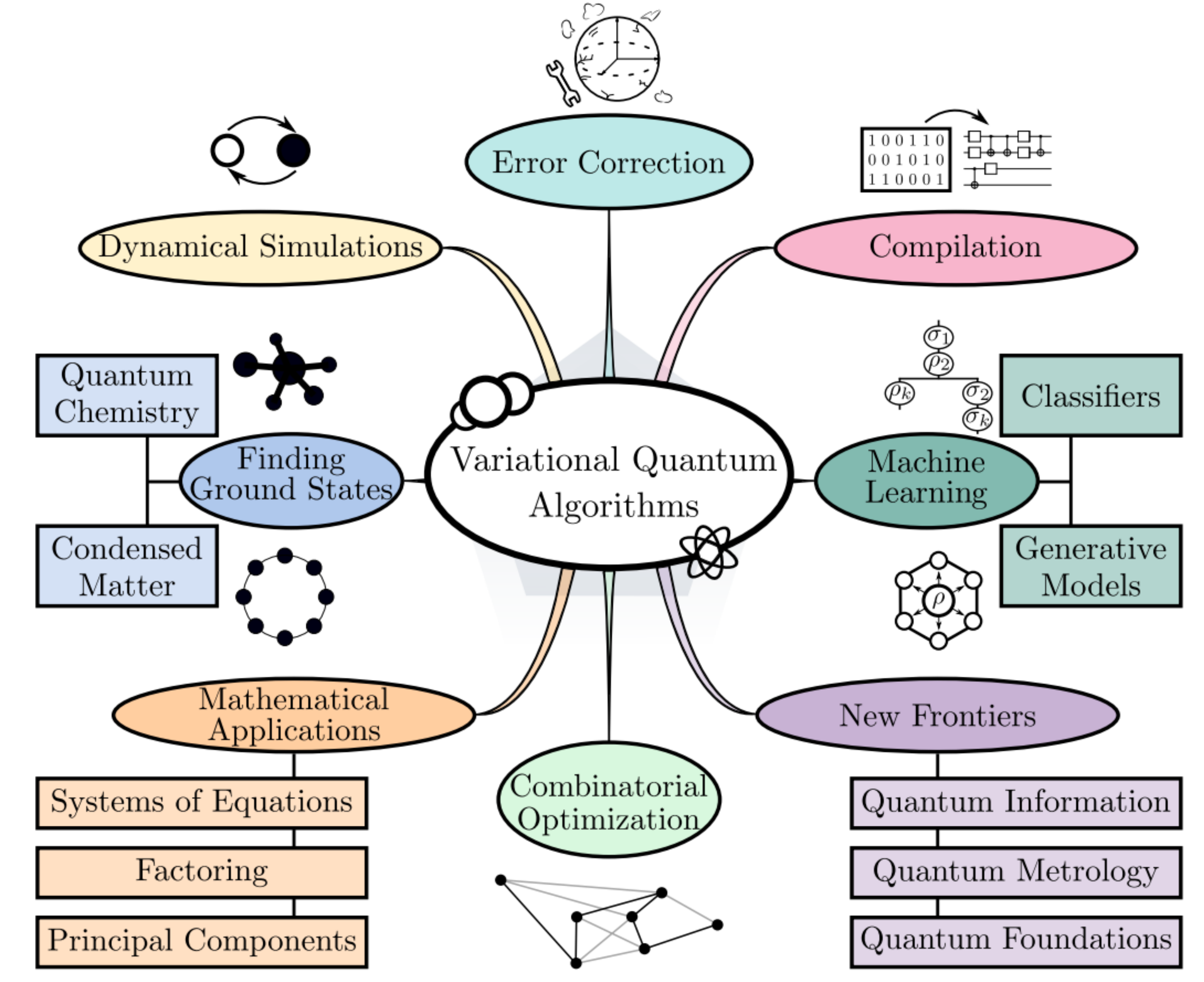


Barren Plateaus

Small gradients or variation

↓

Cerezo, Marco, et al. "Variational quantum algorithms." Nature Reviews Physics 3.9 (2021)



Barren Plateaus

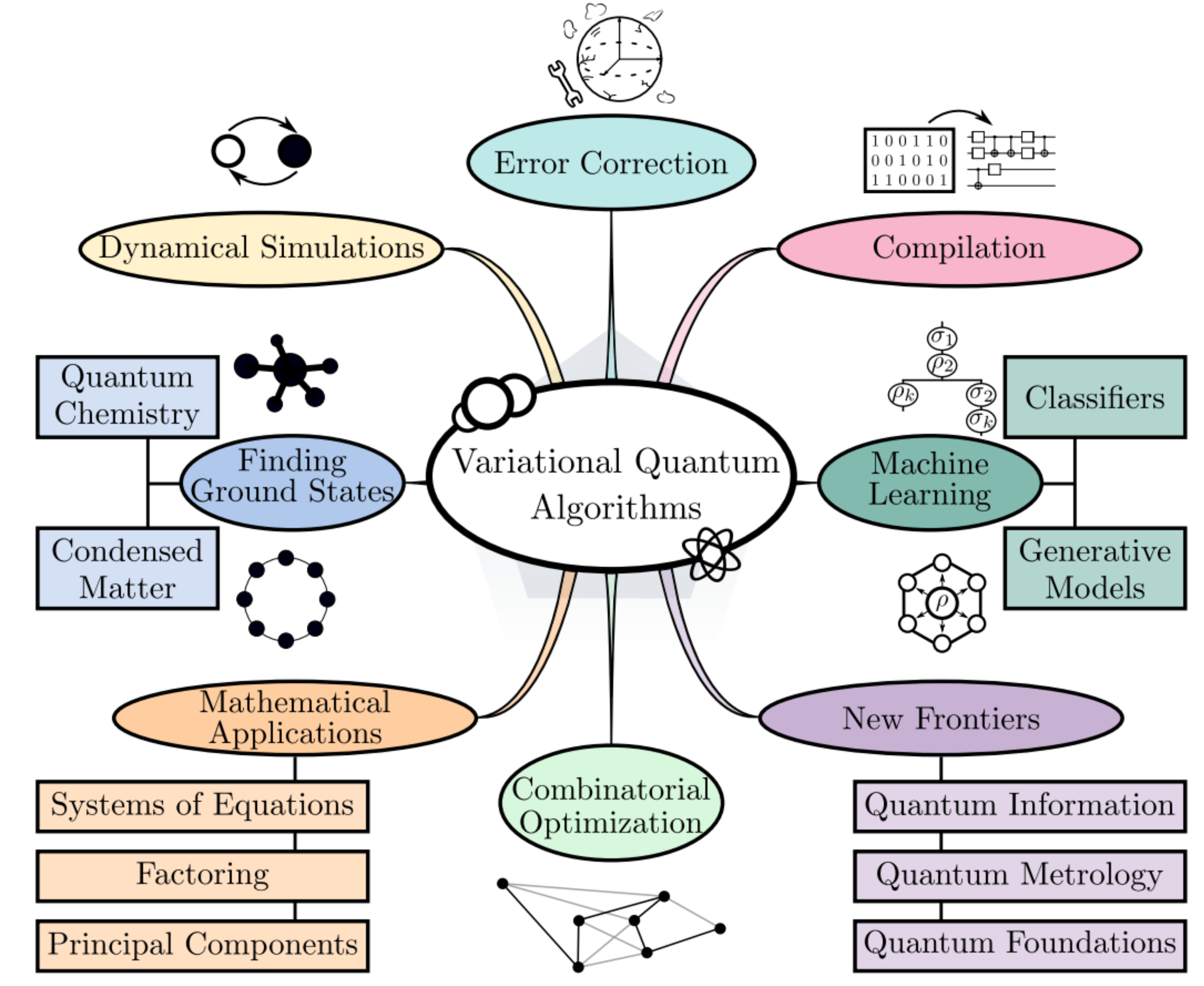
Small gradients or variation

↓

High precision required to find cost-minimizing direction

↓

Cerezo, Marco, et al. "Variational quantum algorithms." *Nature Reviews Physics* 3.9 (2021)



Barren Plateaus

Small gradients or variation

↓

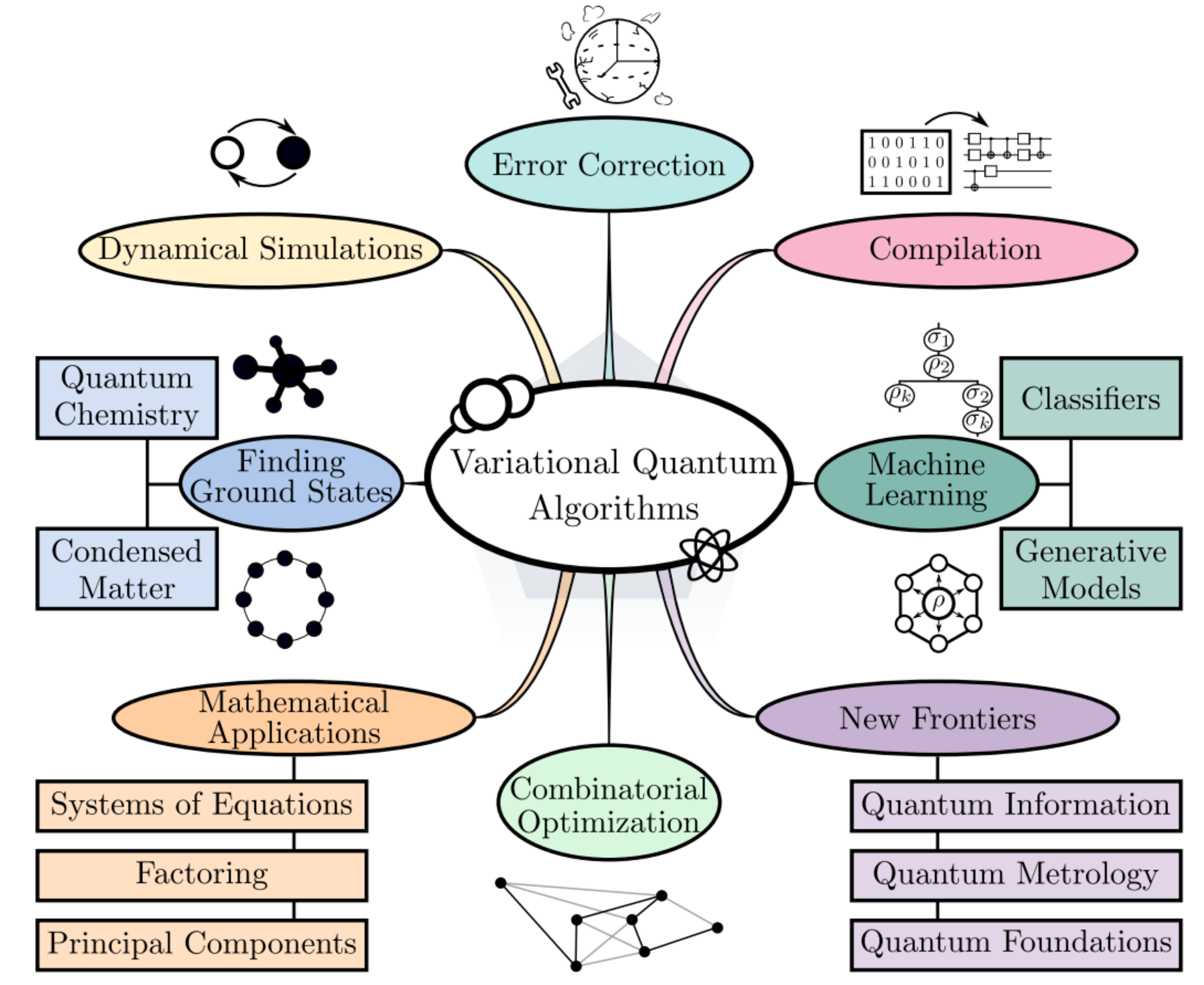
High precision required to find cost-minimizing direction

↓

Large number of measurements

($\frac{1}{\epsilon^2}$ measurements are required estimate a cost to precision ϵ)

Cerezo, Marco, et al. "Variational quantum algorithms." Nature Reviews Physics 3.9 (2021)



WITH HIGH PROBABILITY,
A MEASUREMENT TELLS YOU VERY LITTLE!

Barren Plateaus

Small gradients or variation

↓

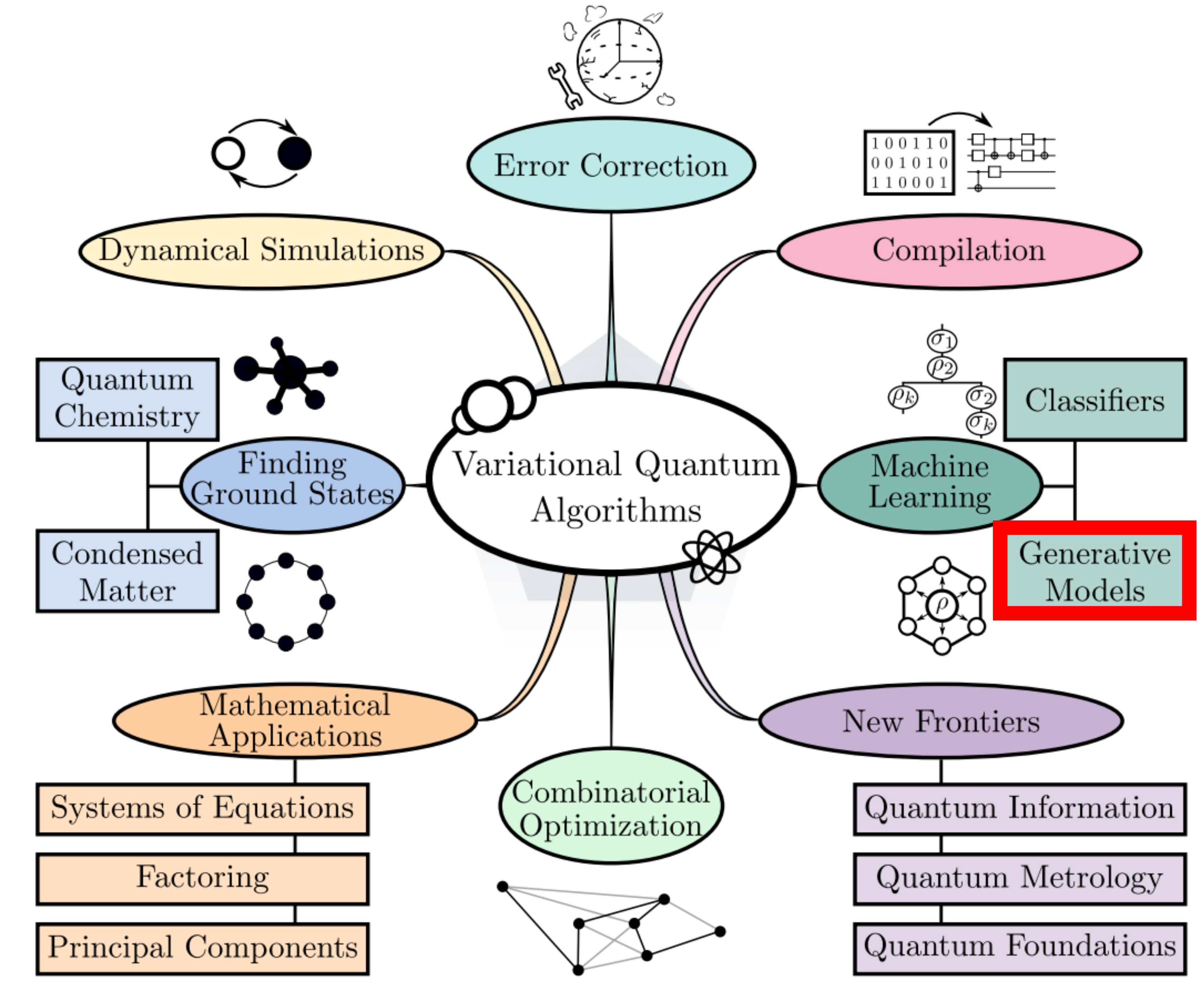
High precision required to find cost-minimizing direction

↓

Large number of measurements

($\frac{1}{\epsilon^2}$ measurements are required estimate a cost to precision ϵ)

Cerezo, Marco, et al. "Variational quantum algorithms." Nature Reviews Physics 3.9 (2021)



WITH HIGH PROBABILITY,
A MEASUREMENT TELLS YOU VERY LITTLE!

Barren Plateaus

Small gradients or variation

↓

High precision required to find cost-minimizing direction

↓

Large number of measurements

($\frac{1}{\epsilon^2}$ measurements are required estimate a cost to precision ϵ)

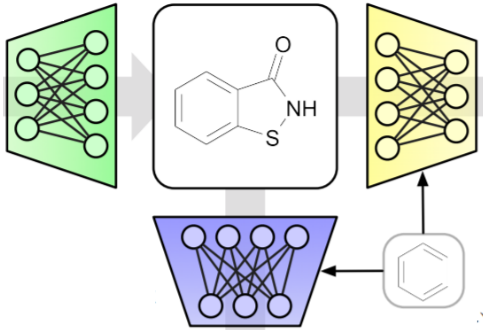
Generative Machine Learning

Generative Machine Learning

Image Generation

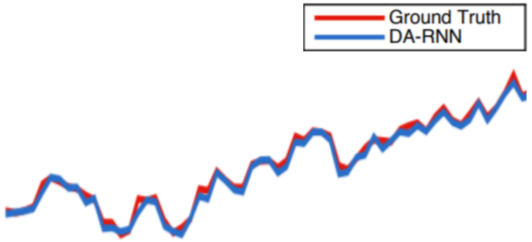
thispersondoesnotexist.com

Molecular Discovery



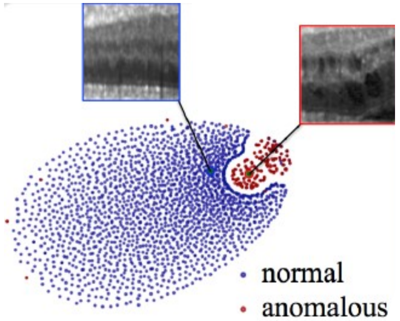
De Cao and Kipf, arXiv:1805.11973

Time-Series Simulation



Qin et al., arXiv:1704.02971

Anomaly Detection



Schlegl et al., IMPI 2017

Natural Language Processing

MA Please make a witty joke about quantum generative models.

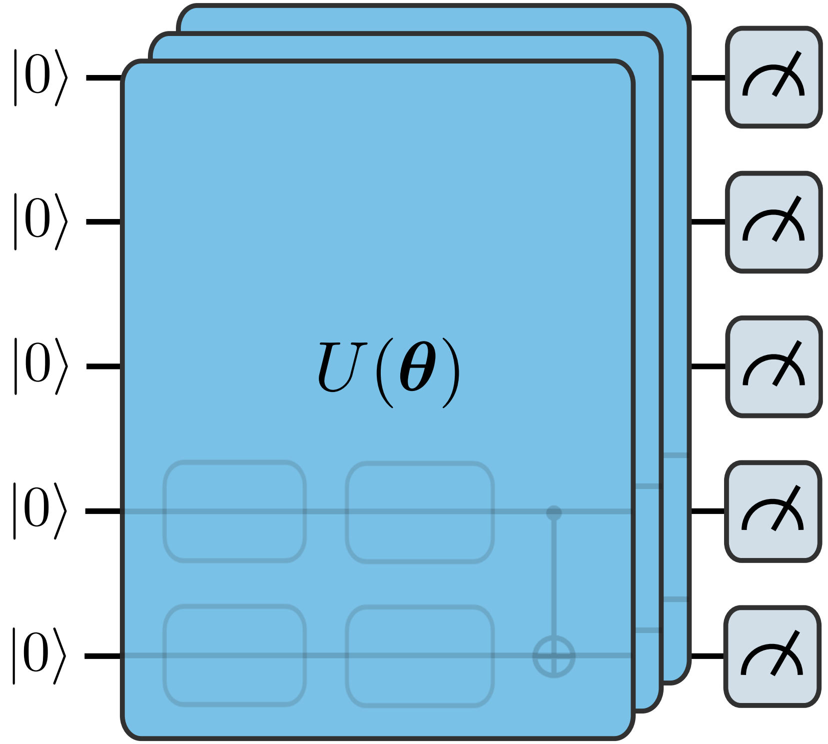
Sure, here's a witty joke about quantum generative models:

Why did the quantum generative model enroll in a comedy class?

Because it wanted to generate laughter in multiple parallel universes simultaneously, leaving no joke unappreciated!

ChatGPT, OpenAI

Quantum Circuit Born Machine (QCBM)

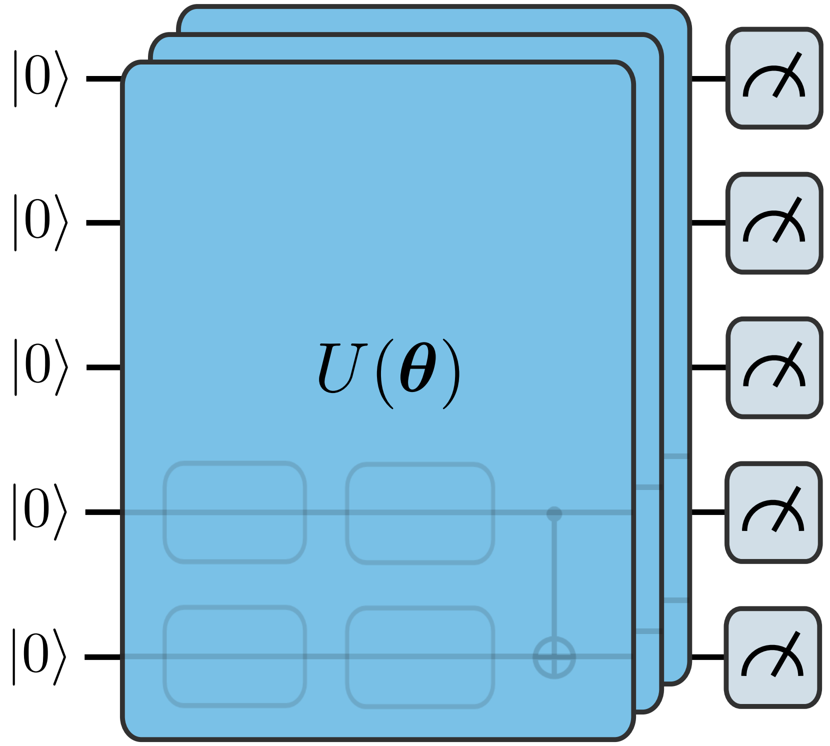


Probability for each sample:

$$q_\theta(x) = \left|\boldsymbol{\alpha}_{x_1 x_2 \ldots x_n}\right|^2 = |\langle x|U(\theta)|0\rangle|^2$$

$$|\psi\rangle = \begin{pmatrix} \alpha_{0\ldots0} \\ \alpha_{0\ldots1} \\ \vdots \\ \alpha_{1\ldots1} \end{pmatrix}$$

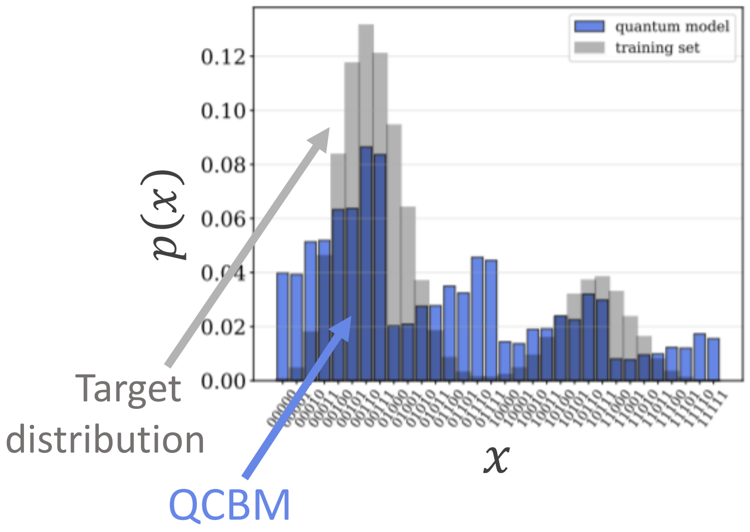
Quantum Circuit Born Machine (QCBM)



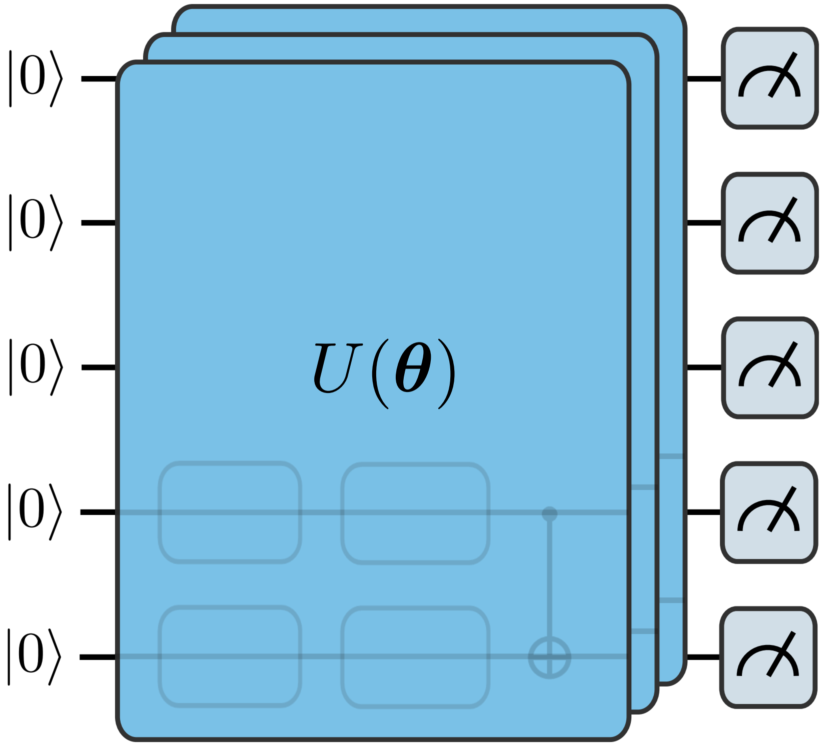
Probability for each sample:

$$q_\theta(x) = \left|\boldsymbol{\alpha}_{x_1 x_2 \ldots x_n}\right|^2 = |\langle x|U(\theta)|0\rangle|^2$$

$$|\psi\rangle = \begin{pmatrix} \alpha_{0\ldots0} \\ \alpha_{0\ldots1} \\ \vdots \\ \alpha_{1\ldots1} \end{pmatrix}$$



Quantum Circuit Born Machine (QCBM)

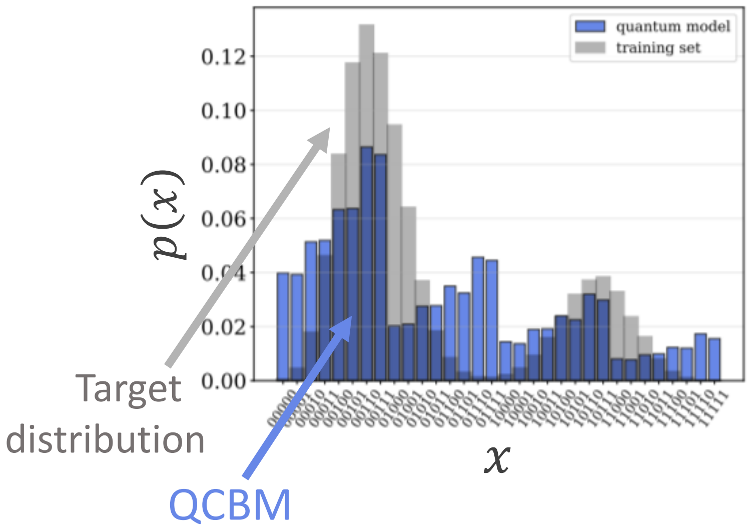


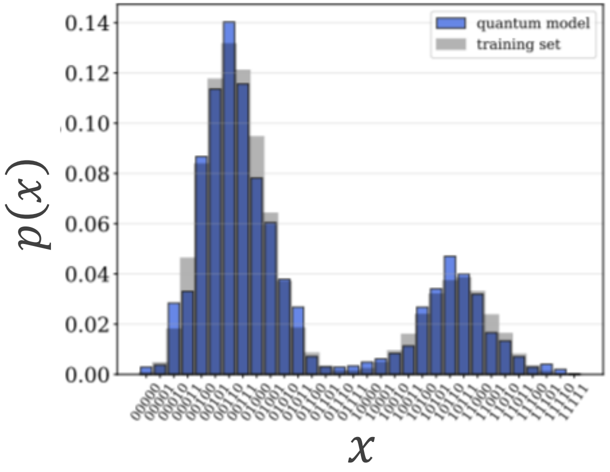
Probability for each sample:

$$q_\theta(x) = \left|\boldsymbol{\alpha}_{x_1 x_2 \ldots x_n}\right|^2 = |\langle x|U(\theta)|0\rangle|^2$$

$$|\psi\rangle = \begin{pmatrix} \alpha_{0\ldots0} \\ \alpha_{0\ldots1} \\ \vdots \\ \alpha_{1\ldots1} \end{pmatrix}$$

A deeper circuit gives more flexibility!





Explicit and Implicit

Explicit Models have access to $q_\theta(x)$ in polynomial time.

Implicit Models

Explicit and Implicit

Explicit Models have access to $\boldsymbol{q_\theta(x)}$ in polynomial time.

Implicit Models only have access to samples from $\boldsymbol{q_\theta(x)}$.

Explicit and Implicit

Explicit Models have access to $q_{\theta}(x)$ in polynomial time.

- *Autoregressive models (RNNs, Transformers)*
- *Tensor Networks without loops (MPS, TTNs)*

Implicit Models only have access to samples from $q_{\theta}(x)$.

Explicit and Implicit

Explicit Models have access to $q_\theta(x)$ in polynomial time.

- *Autoregressive models (RNNs, Transformers)*
- *Tensor Networks without loops (MPS, TTNs)*

Implicit Models only have access to samples from $q_\theta(x)$.

- *Generative Adversarial Networks*
- *Diffusion models*

Explicit and Implicit

Explicit Models have access to $\boldsymbol{q_\theta(x)}$ in polynomial time.

- *Autoregressive models (RNNs, Transformers)*
- *Tensor Networks without loops (MPS, TTNs)*

Implicit Models only have access to samples from $\boldsymbol{q_\theta(x)}$.

- *Generative Adversarial Networks*
- *Diffusion models*
- *Quantum algorithms (QCBMs, QGANs, QBMs)*

Explicit and Implicit

Explicit Models have access to $q_{\theta}(x)$ in polynomial time.

- *Autoregressive models (RNNs, Transformers)*
- *Tensor Networks without loops (MPS, TTNs)*

Explicit Losses are defined via probabilities

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

Implicit Models only have access to samples from $q_{\theta}(x)$.

- *Generative Adversarial Networks*
- *Diffusion models*
- *Quantum algorithms (QCBMs, QGANs, QBMs)*

Implicit Losses are defined via samples

Implicit

$$\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x}, \boldsymbol{y})\right]$$

Explicit Losses

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

Explicit Losses

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

KL Divergence

$$\mathcal{L}^{\mathrm{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)$$

Reverse KL Divergence

$$\mathcal{L}^{\mathrm{rev-KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x})}\right)$$

Jensen-Shannon Divergence

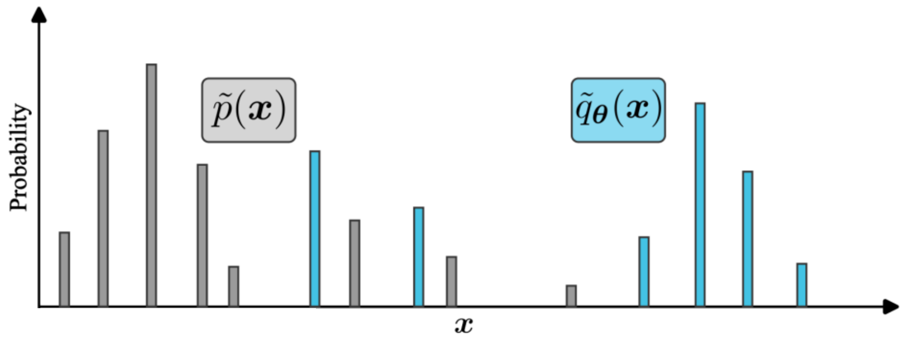
$$\mathcal{L}^{\mathrm{JSD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} \Bigg[p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) + q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)\Bigg]$$

Total Variation Distance

$$\mathcal{L}^{\mathrm{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})|$$

Explicit Losses

Explicit
$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$



KL Divergence
$$\mathcal{L}^{\mathrm{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)$$

Reverse KL Divergence
$$\mathcal{L}^{\mathrm{rev-KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x})}\right)$$

Jensen-Shannon Divergence
$$\mathcal{L}^{\mathrm{JSD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \left[p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) + q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)\right]$$

Total Variation Distance
$$\mathcal{L}^{\mathrm{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})|$$

Explicit Losses

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

KL Divergence

$$\mathcal{L}^{\text{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)$$

Reverse KL Divergence

$$\mathcal{L}^{\text{rev}-\text{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x})}\right)$$

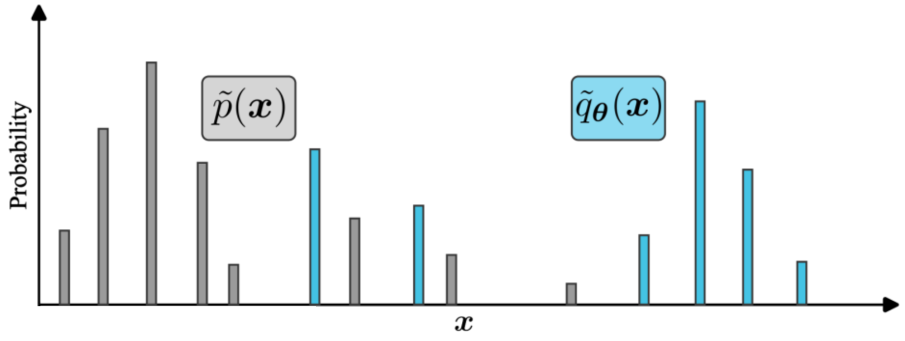
Jensen-Shannon Divergence

$$\mathcal{L}^{\text{JSD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \left[p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) + q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)\right]$$

Total Variation Distance

$$\mathcal{L}^{\text{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})|$$

<u>Question:</u> With what probability do we generate a sample that contributes to the training loss?



Explicit Losses

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

KL Divergence

$$\mathcal{L}^{\mathrm{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)$$

Reverse KL Divergence

$$\mathcal{L}^{\mathrm{rev-KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x})}\right)$$

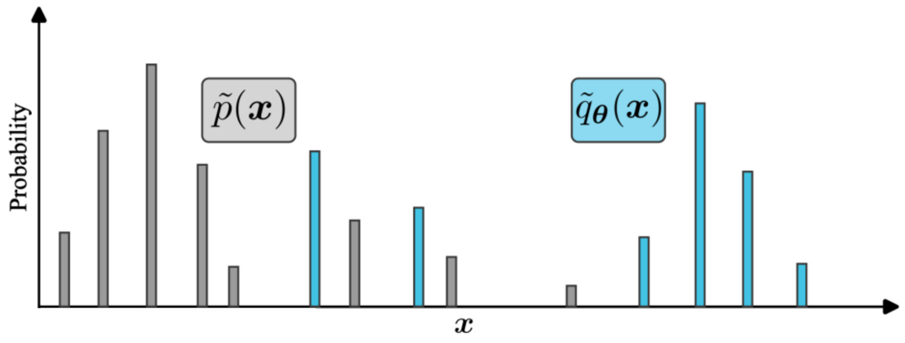
Jensen-Shannon Divergence

$$\mathcal{L}^{\mathrm{JSD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} \left[p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) + q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)\right]$$

Total Variation Distance

$$\mathcal{L}^{\mathrm{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x} \in \mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})|$$

Question: With what probability do we generate a sample that contributes to the training loss?
Answer: Without good initialization or inductive bias, it is **exponentially small**!



Explicit Losses

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

KL Divergence

$$\mathcal{L}^{\mathrm{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)$$

Reverse KL Divergence

$$\mathcal{L}^{\mathrm{rev-KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x})}\right)$$

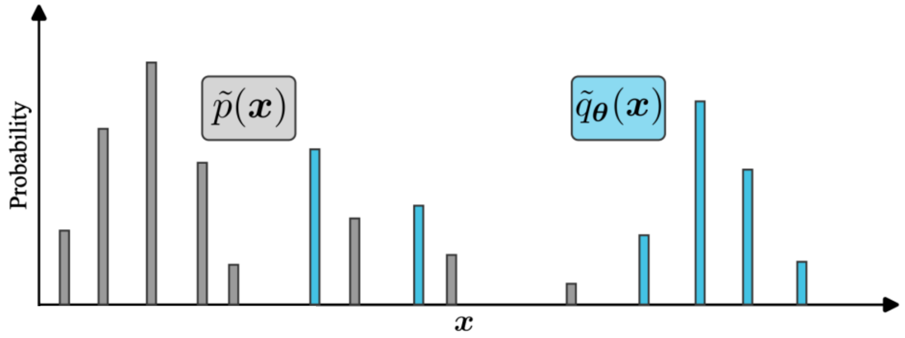
Jensen-Shannon Divergence

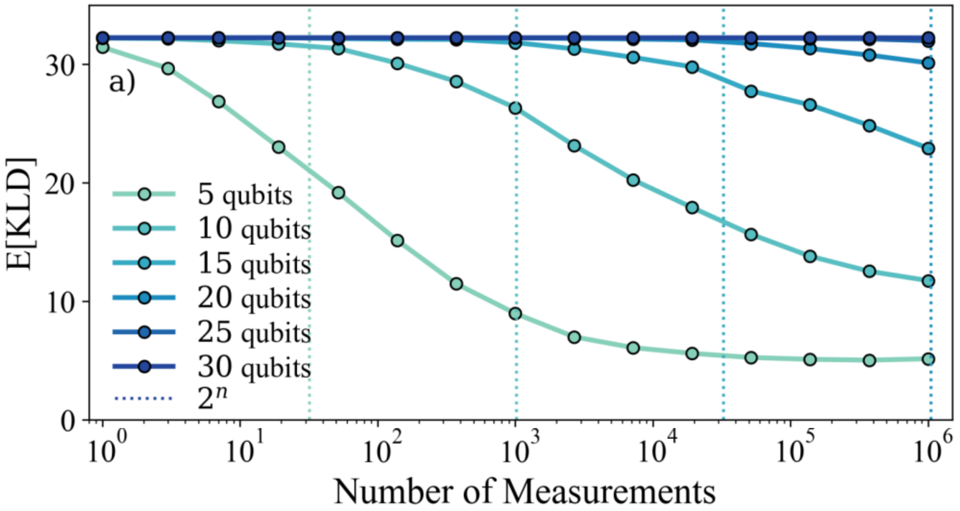
$$\mathcal{L}^{\mathrm{JSD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \left[p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) + q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)\right]$$

Total Variation Distance

$$\mathcal{L}^{\mathrm{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})|$$

Question: With what probability do we generate a sample that contributes to the training loss?
Answer: Without good initialization or inductive bias, it is **exponentially small**!





Explicit Losses

... ARE NOT TRAINABLE WITH GENERIC MODELS

Explicit

$$\sum_{\boldsymbol{x}} f\left(\tilde{p}(\boldsymbol{x}), \tilde{q}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

KL Divergence

$$\mathcal{L}^{\mathrm{KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)$$

Reverse KL Divergence

$$\mathcal{L}^{\mathrm{rev-KLD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x})}\right)$$

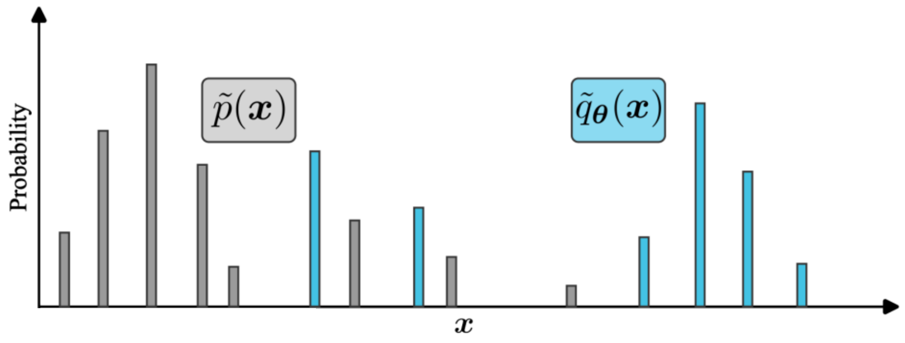
Jensen-Shannon Divergence

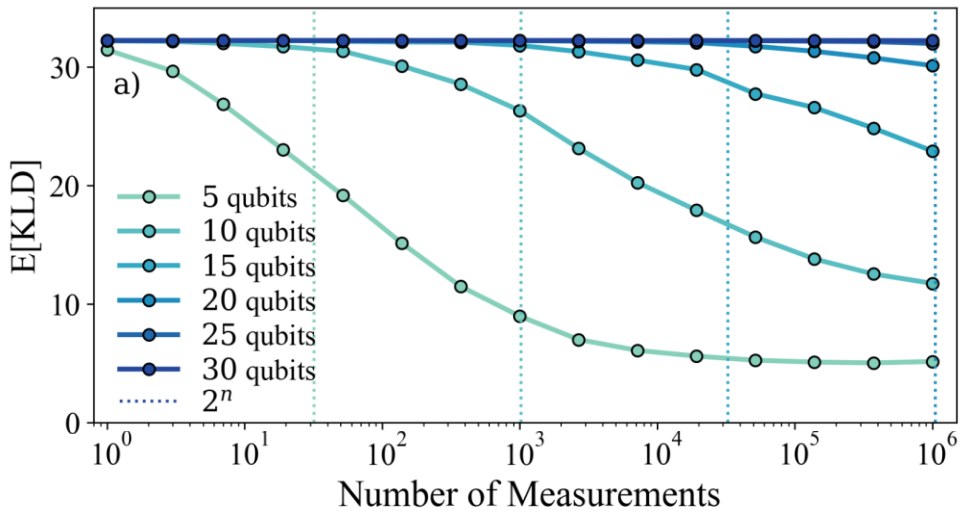
$$\mathcal{L}^{\mathrm{JSD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} \left[p(\boldsymbol{x}) \log\left(\frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right) + q_{\boldsymbol{\theta}}(\boldsymbol{x}) \log\left(\frac{q_{\boldsymbol{\theta}}(\boldsymbol{x})}{p(\boldsymbol{x}) + q_{\boldsymbol{\theta}}(\boldsymbol{x})}\right)\right]$$

Total Variation Distance

$$\mathcal{L}^{\mathrm{TVD}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in\mathcal{X}} |p(\boldsymbol{x}) - q_{\boldsymbol{\theta}}(\boldsymbol{x})|$$

<u>Question:</u> With what probability do we generate a sample that contributes to the training loss?
<u>Answer:</u> Without good initialization or inductive bias, it is **exponentially small**!





Implicit Losses

Implicit

$$\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Implicit Losses

Implicit

$$\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\text{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$

Implicit Losses

Implicit

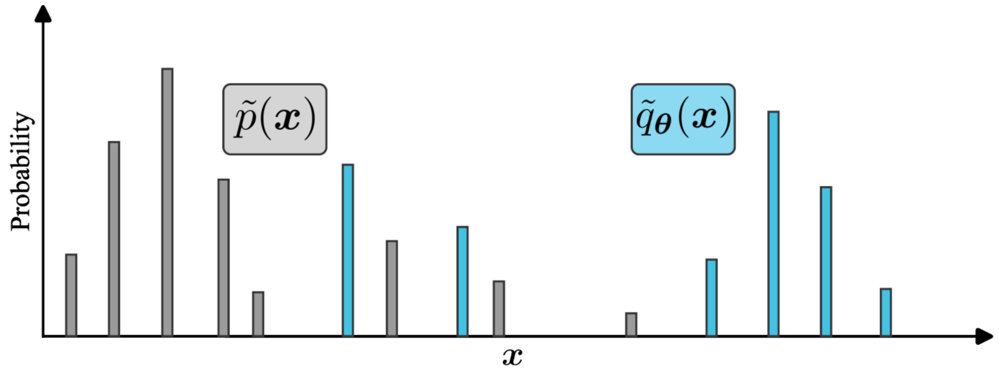
$$\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\mathrm{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$



Implicit Losses

Implicit

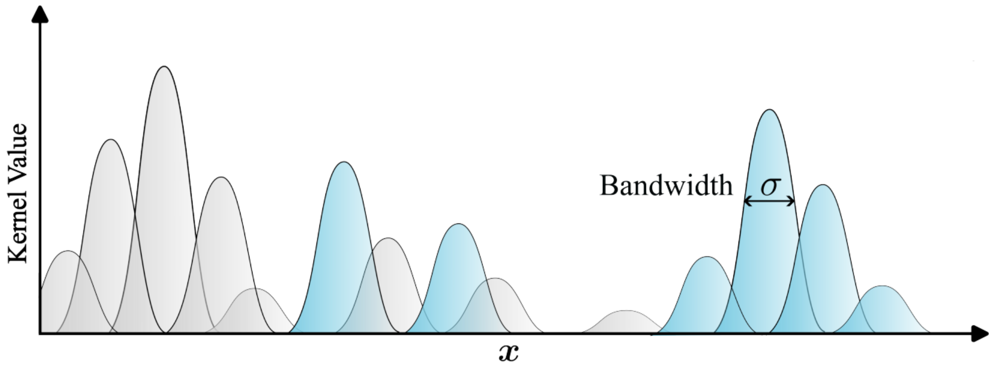
$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\mathrm{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$



Implicit Losses

Implicit

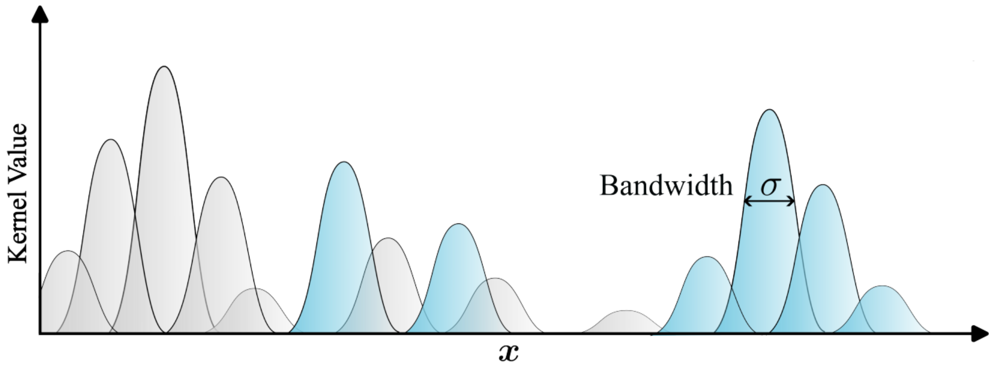
$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\text{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$



As observable

$$\mathcal{M}(\rho,\rho') = \text{Tr}\left[O_{\text{MMD}}^{(\sigma)}(\rho\otimes\rho')\right]$$

$$O_{\text{MMD}}^{(\sigma)} := \sum_{\boldsymbol{x},\boldsymbol{y}} K_\sigma(\boldsymbol{x},\boldsymbol{y})|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\otimes|\boldsymbol{y}\rangle\langle\boldsymbol{y}|\,.$$

Implicit Losses

Implicit

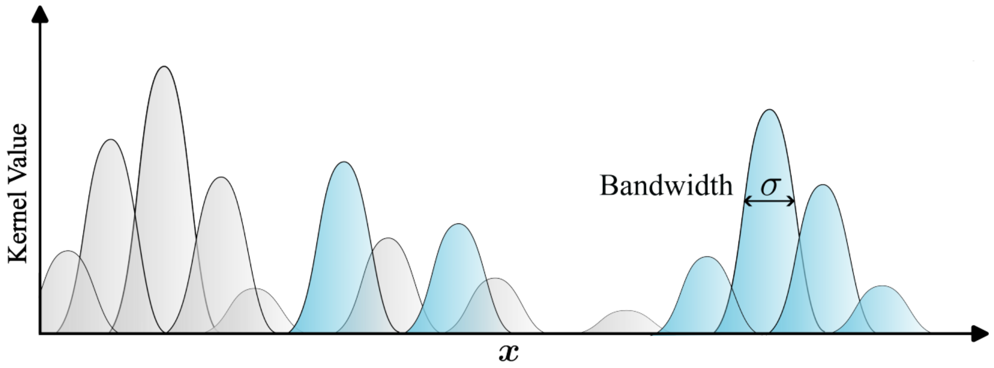
$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\text{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

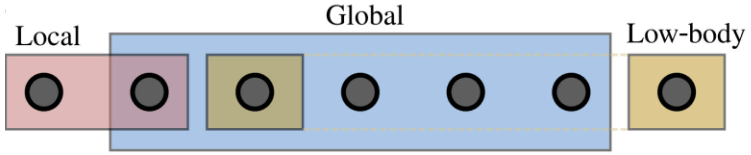
$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$



As observable

$$\mathcal{M}(\rho,\rho') = \mathrm{Tr}\left[O^{(\sigma)}_{\text{MMD}}(\rho\otimes\rho')\right]$$

$$O^{(\sigma)}_{\text{MMD}} := \sum_{\boldsymbol{x},\boldsymbol{y}} K_\sigma(\boldsymbol{x},\boldsymbol{y})|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\otimes|\boldsymbol{y}\rangle\langle\boldsymbol{y}|\,.$$



Local/Low-body: Trainable
Global: Untrainable

Implicit Losses

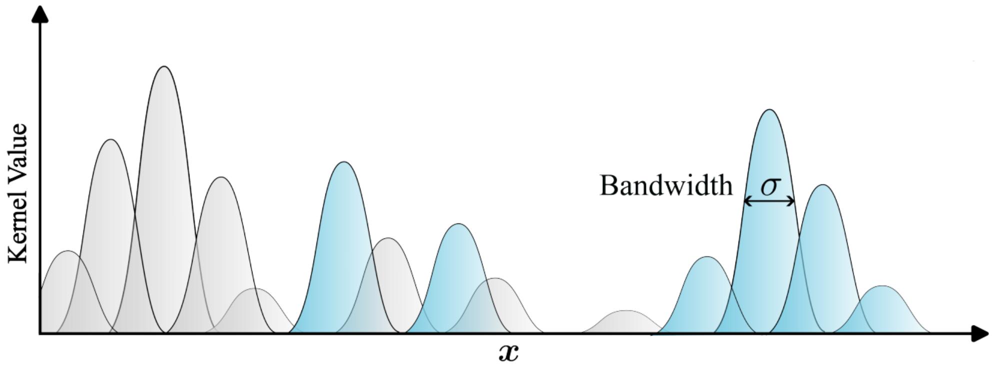
Implicit
$$\underset{\boldsymbol{x},\boldsymbol{y}}{\mathbb{E}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\mathrm{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

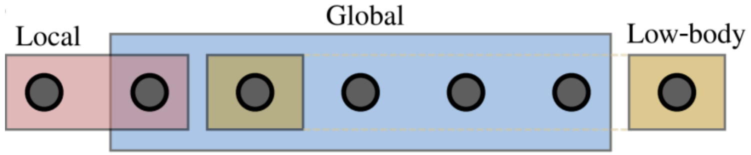
$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$



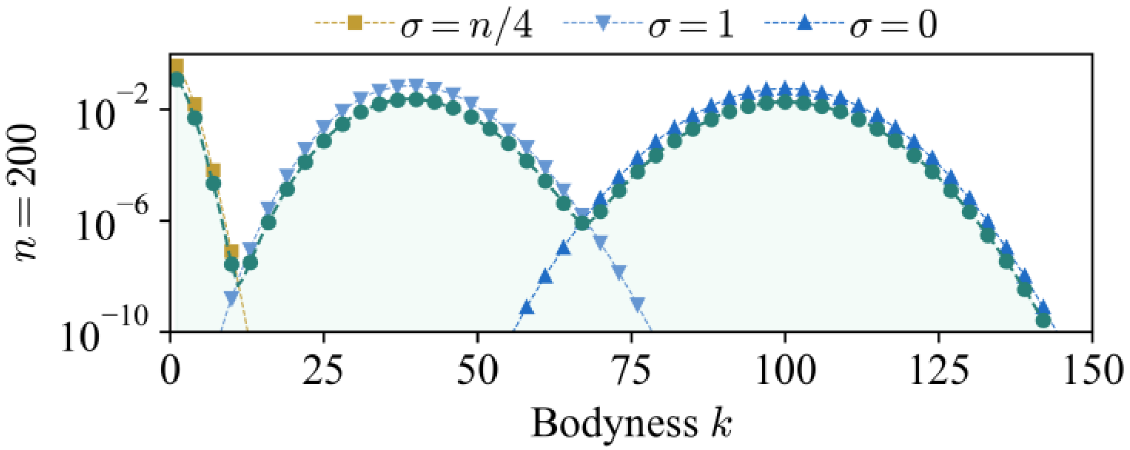
As observable

$$\mathcal{M}(\rho,\rho') = \mathrm{Tr}\left[O_{\mathrm{MMD}}^{(\sigma)}(\rho\otimes\rho')\right]$$

$$O_{\mathrm{MMD}}^{(\sigma)} := \sum_{\boldsymbol{x},\boldsymbol{y}} K_\sigma(\boldsymbol{x},\boldsymbol{y})|\boldsymbol{x}\rangle\langle\boldsymbol{x}| \otimes |\boldsymbol{y}\rangle\langle\boldsymbol{y}|\,.$$



Local/Low-body: Trainable
Global: Untrainable



Implicit Losses

… CAN BE TRAINABLE OR UNTRAINABLE

Implicit

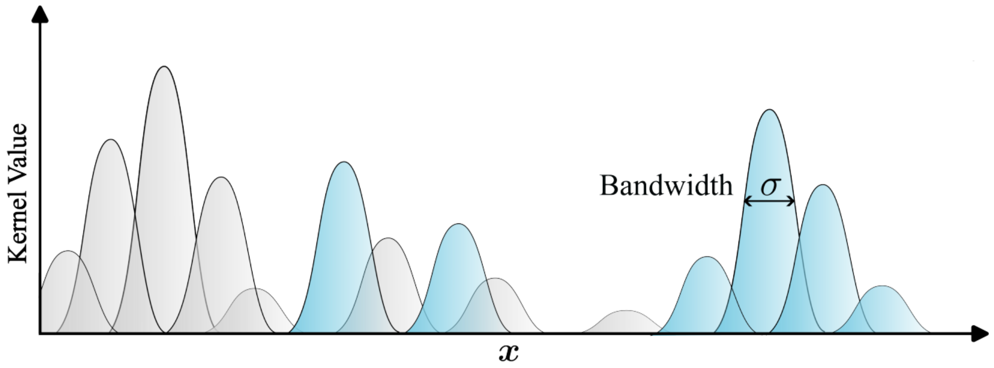
$$\mathop{\mathbb{E}}_{\boldsymbol{x},\boldsymbol{y}}\left[g(\boldsymbol{x},\boldsymbol{y})\right]$$

Maximum Mean Discrepancy

$$\mathcal{L}_{\mathrm{MMD}}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim q_{\boldsymbol{\theta}}}[K(\boldsymbol{x},\boldsymbol{y})] - 2\mathbb{E}_{\boldsymbol{x}\sim q_{\boldsymbol{\theta}},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p}[K(\boldsymbol{x},\boldsymbol{y})]\,,$$

with

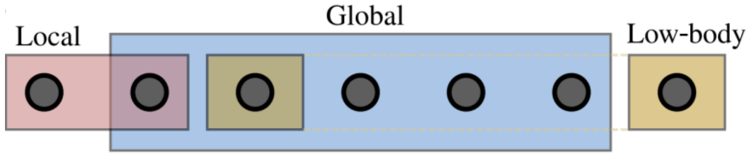
$$K_\sigma(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{2\sigma}} = \prod_{i=1}^{n} e^{-\frac{(x_i-y_i)^2}{2\sigma}}$$



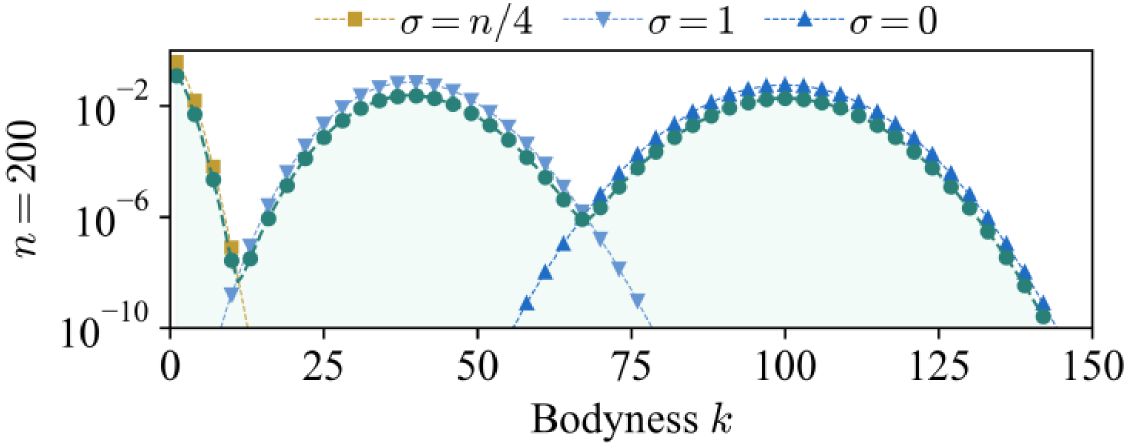
As observable

$$\mathcal{M}(\rho,\rho') = \mathrm{Tr}\left[O_{\mathrm{MMD}}^{(\sigma)}(\rho\otimes\rho')\right]$$

$$O_{\mathrm{MMD}}^{(\sigma)} := \sum_{\boldsymbol{x},\boldsymbol{y}} K_\sigma(\boldsymbol{x},\boldsymbol{y})|\boldsymbol{x}\rangle\langle\boldsymbol{x}|\otimes|\boldsymbol{y}\rangle\langle\boldsymbol{y}|\,.$$

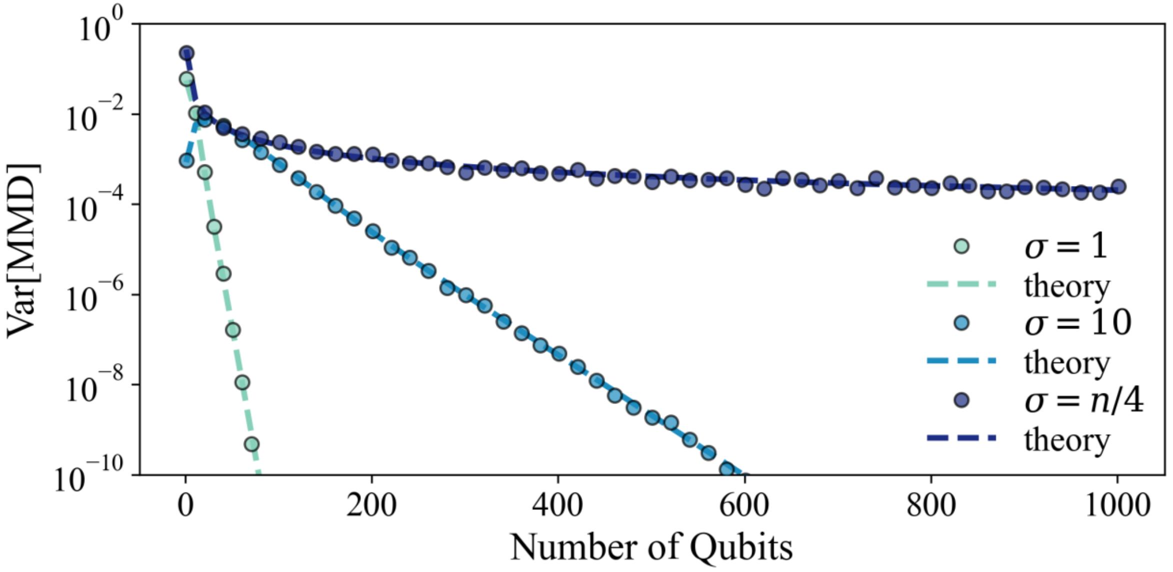


Local/Low-body: Trainable
Global: Untrainable



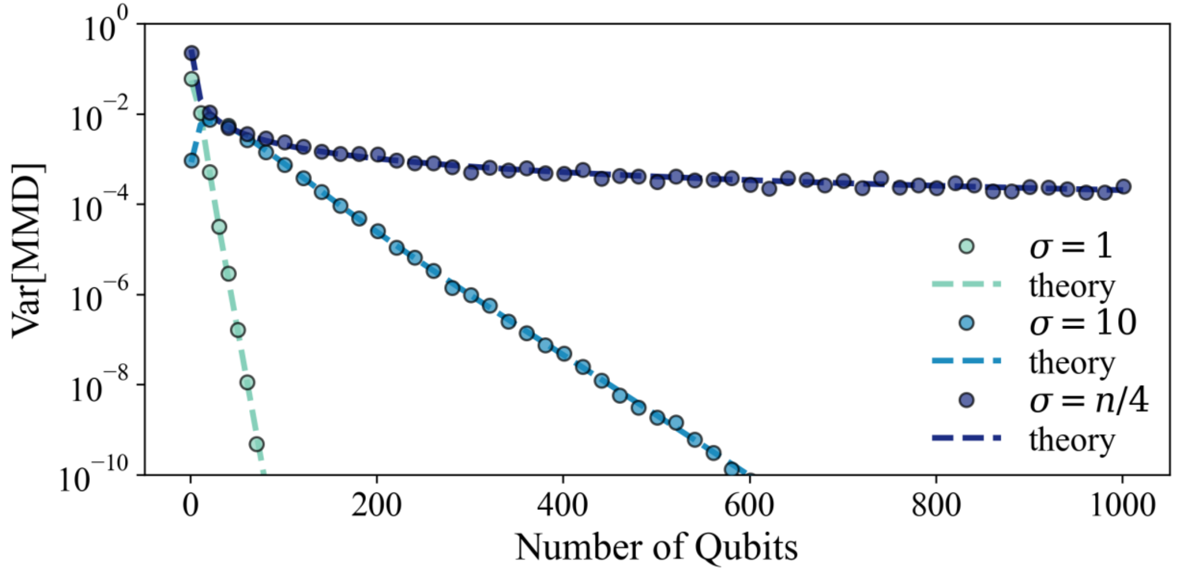
MMD Loss

Exact MMD Loss Variance

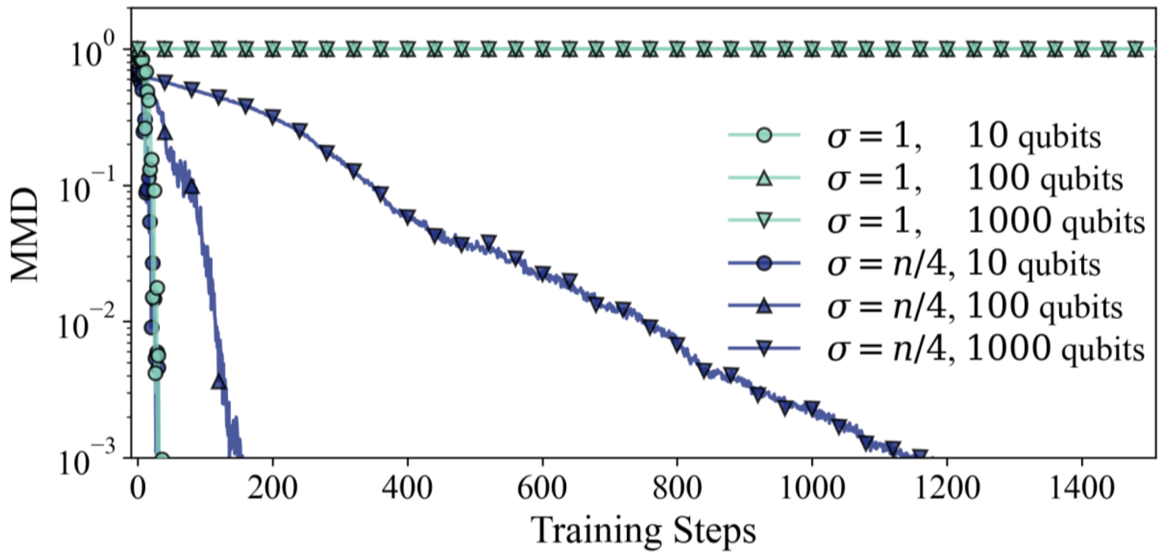


MMD Loss

Exact MMD Loss Variance



MMD Training



Large Gradients are not enough

Large Gradients are not enough

Locality ≡ Probability Marginals

If loss function is **at most k-local,**

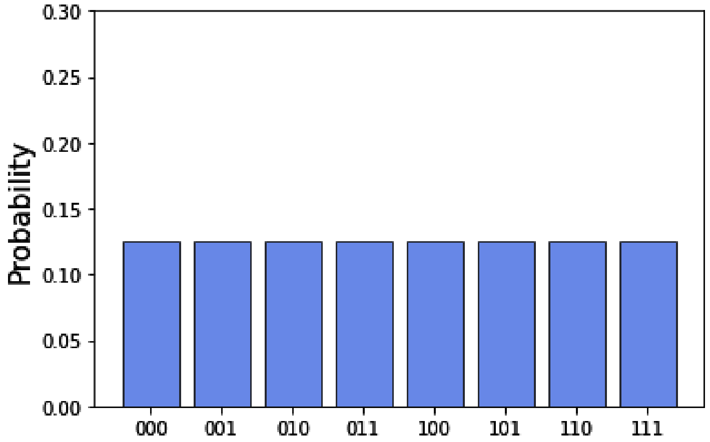
it cannot learn beyond **k-order marginals.**

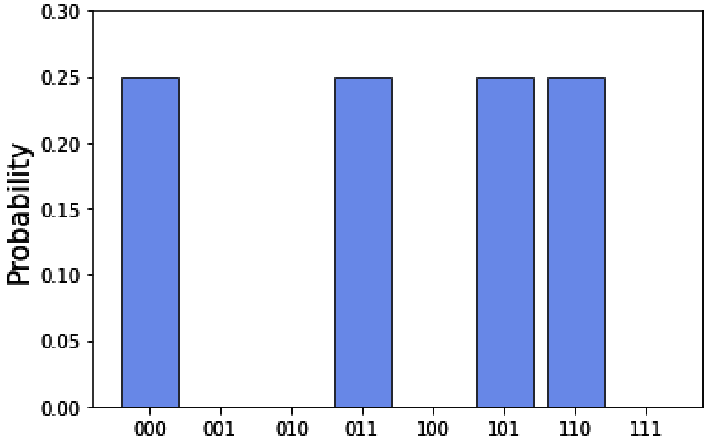
Large Gradients are not enough

Locality ≡ Probability Marginals

If loss function is **at most k-local,**

it cannot learn beyond **k-order marginals.**



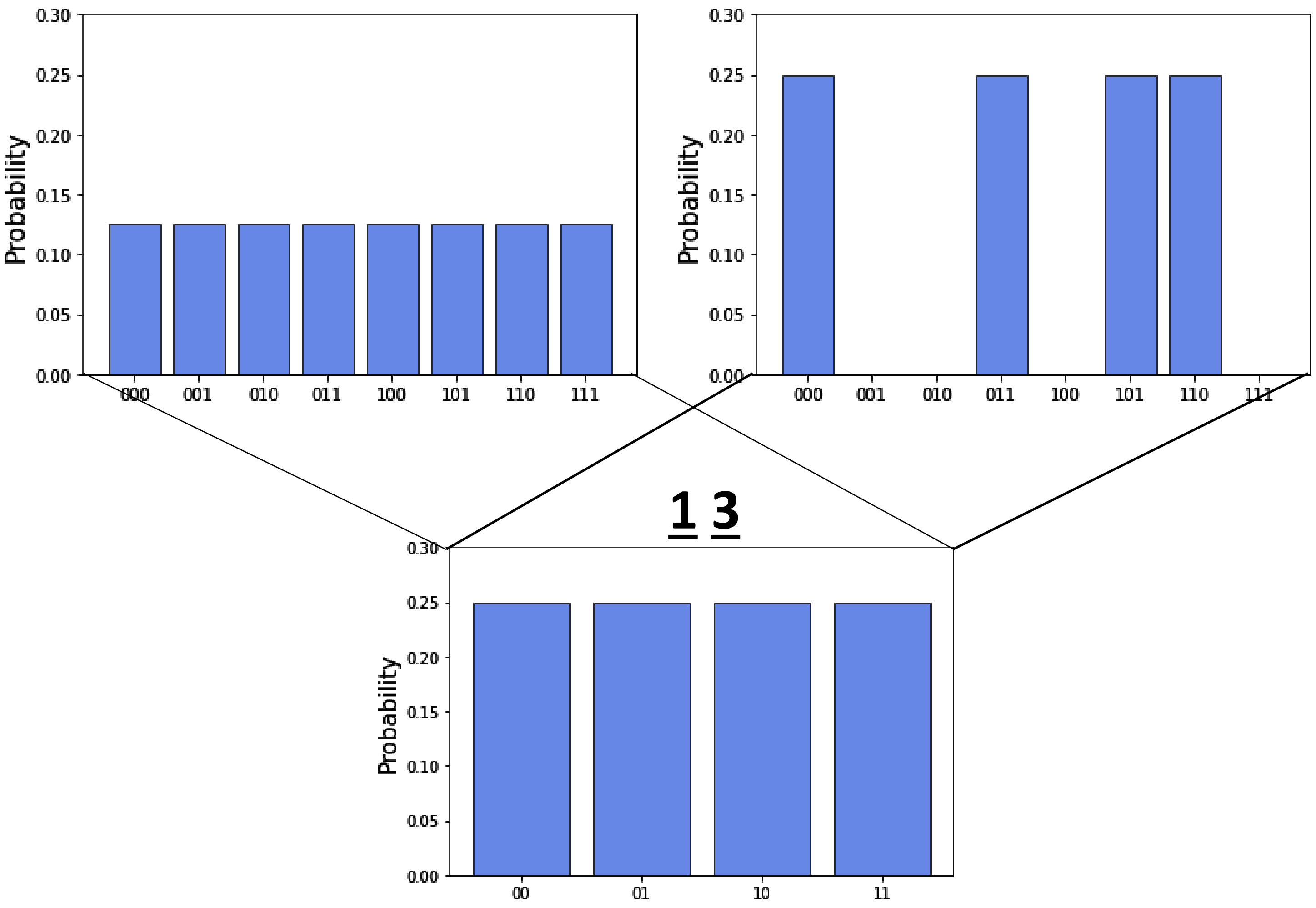


Large Gradients are not enough

Locality ≡ Probability Marginals

If loss function is **at most k-local,**

it cannot learn beyond **k-order marginals.**

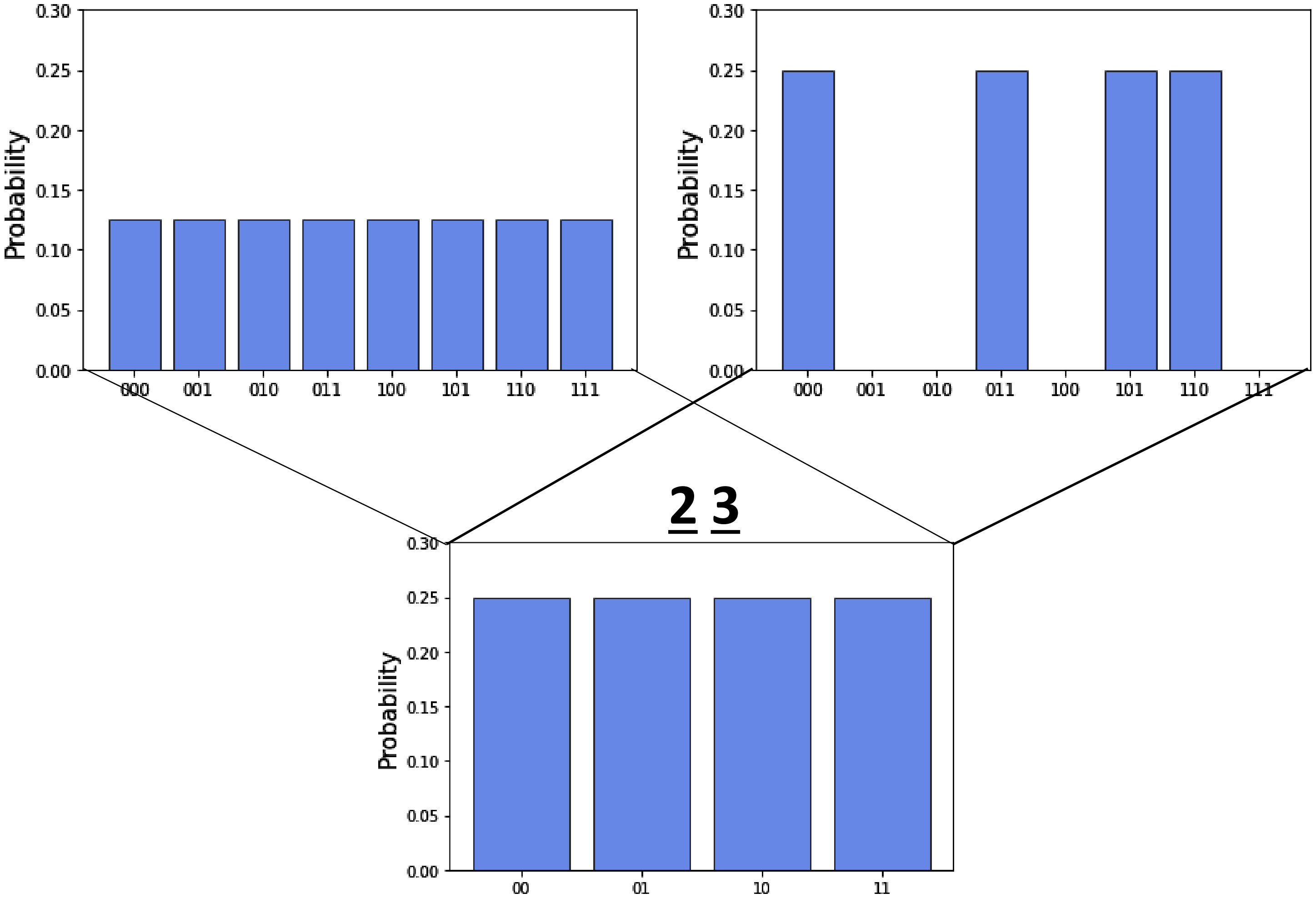


Large Gradients are not enough

Locality ≡ Probability Marginals

If loss function is **at most k-local,**

it cannot learn beyond **k-order marginals.**



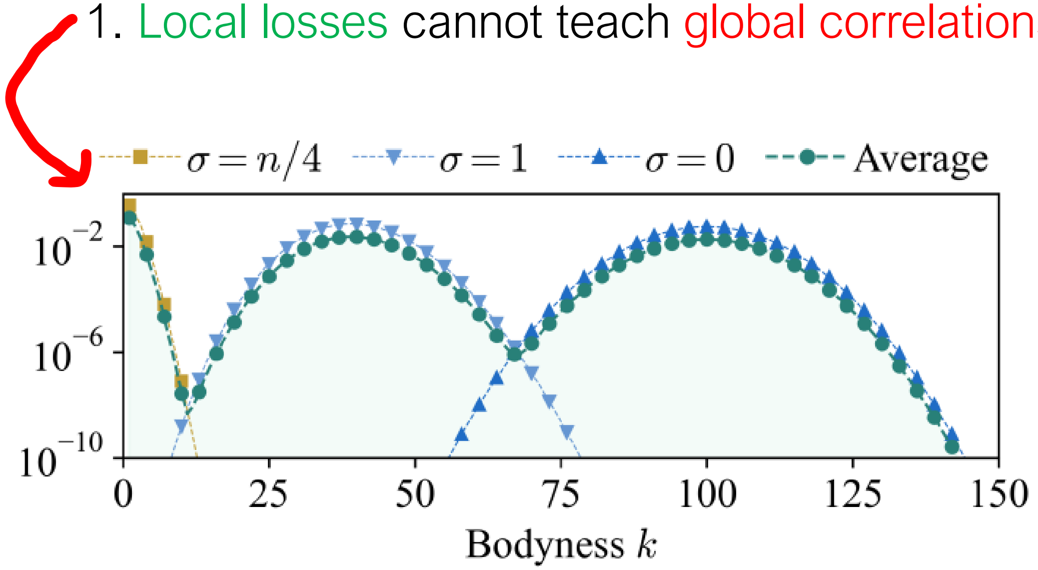
Large Gradients are not enough

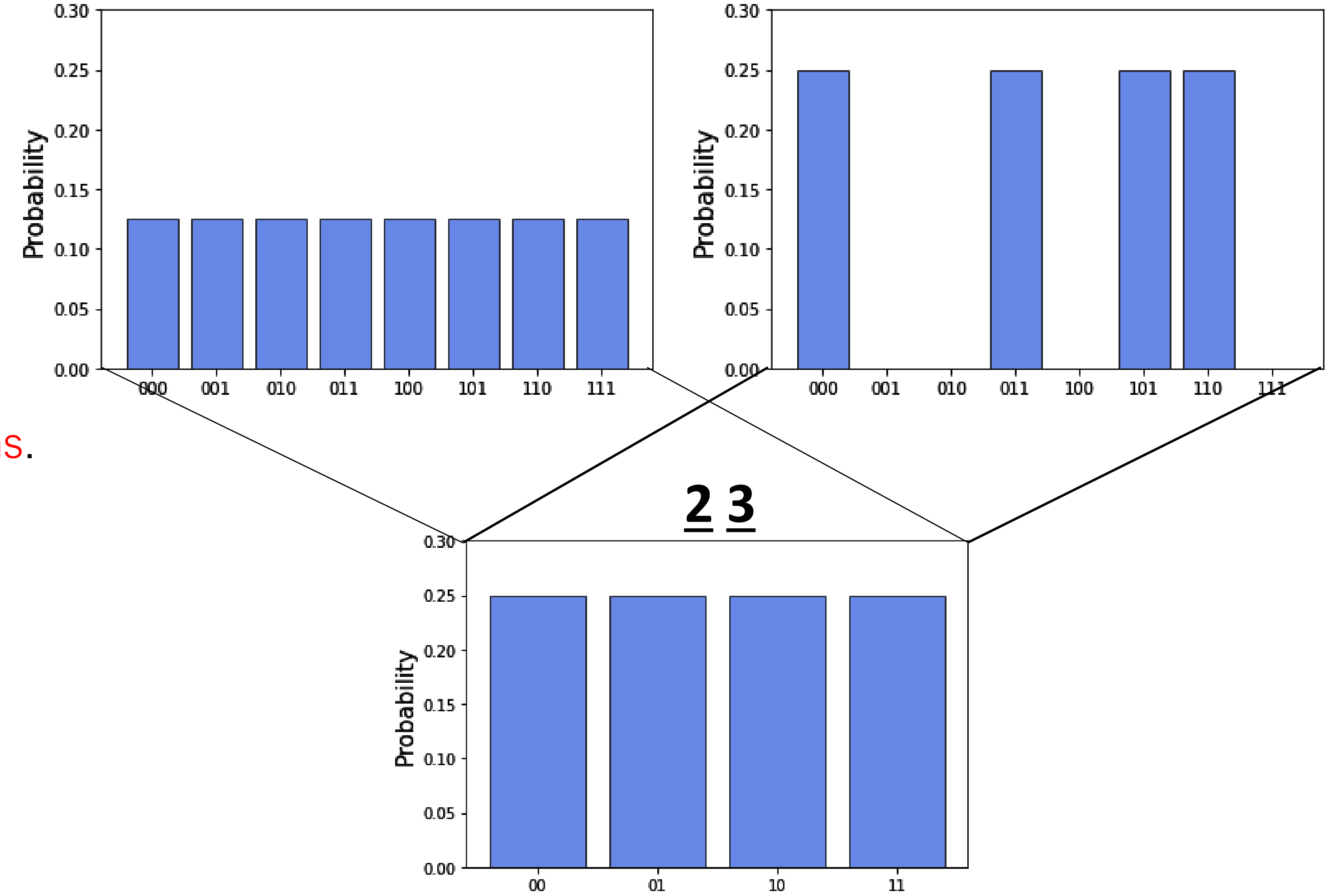
Locality ≡ Probability Marginals

If loss function is **at most k-local,**

it cannot learn beyond **k-order marginals.**

1. Local losses cannot teach global correlations.





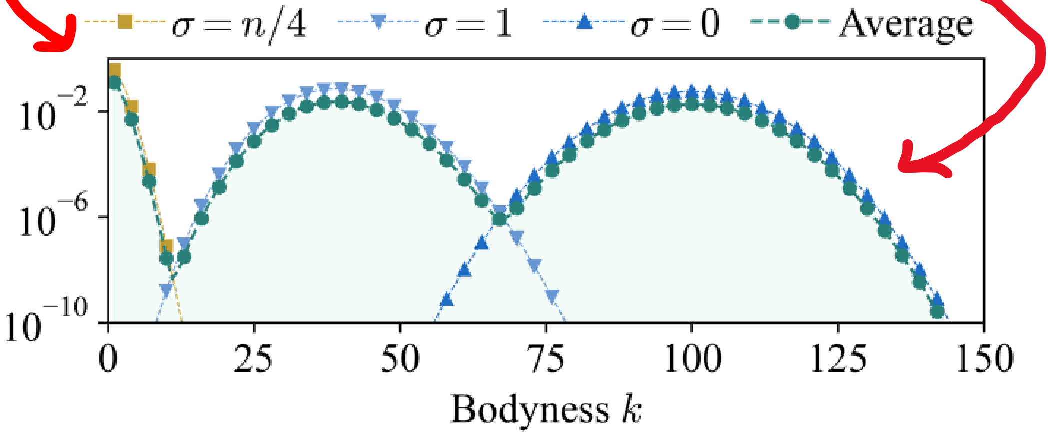
Large Gradients are not enough

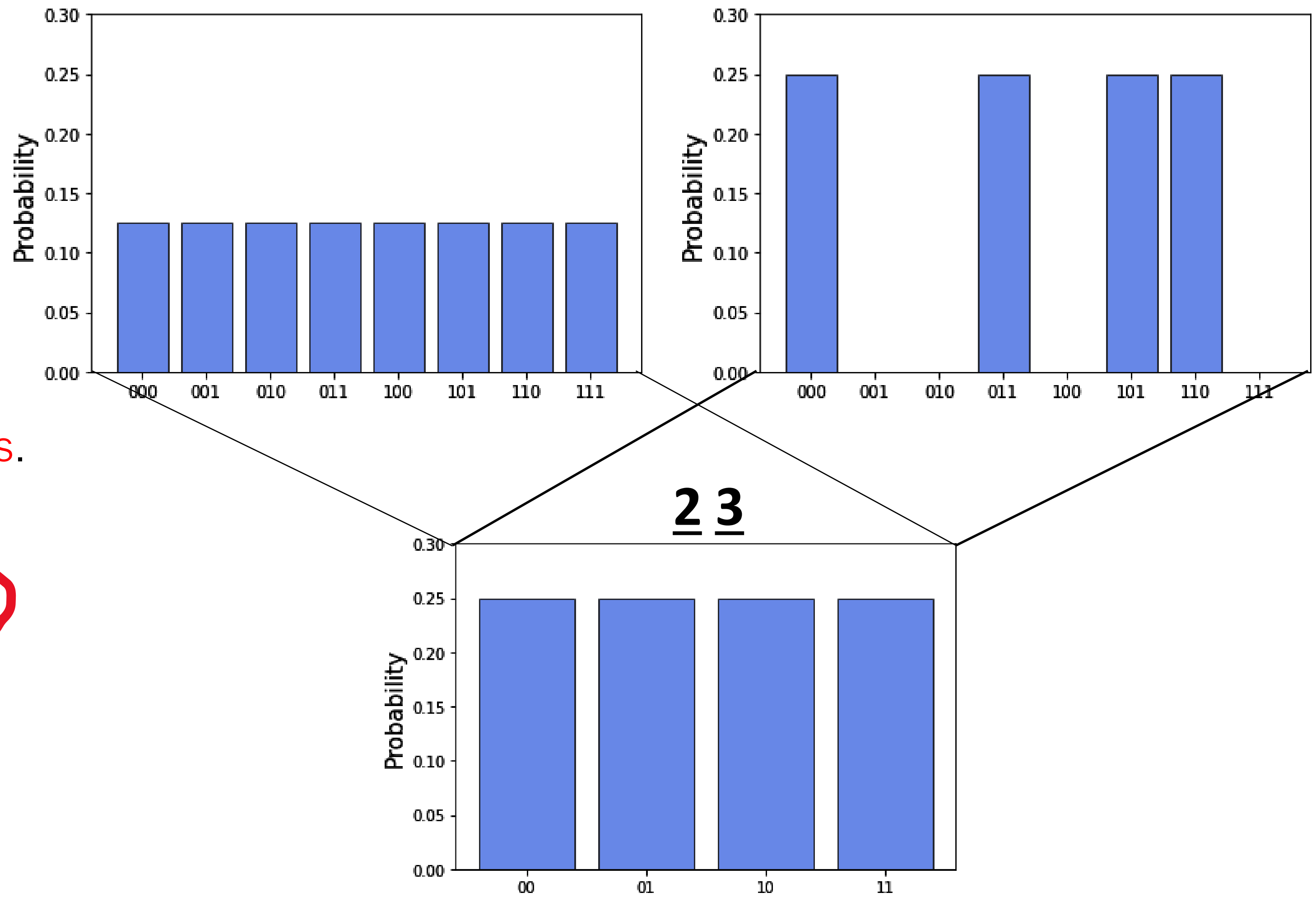
Locality ≡ Probability Marginals

If loss function is **at most k-local,**

it cannot learn beyond **k-order marginals.**

1. Local losses cannot teach global correlations.
2. Global losses are not trainable.



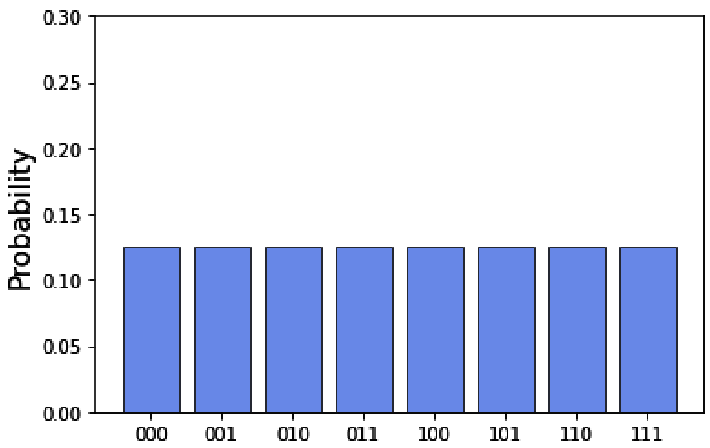


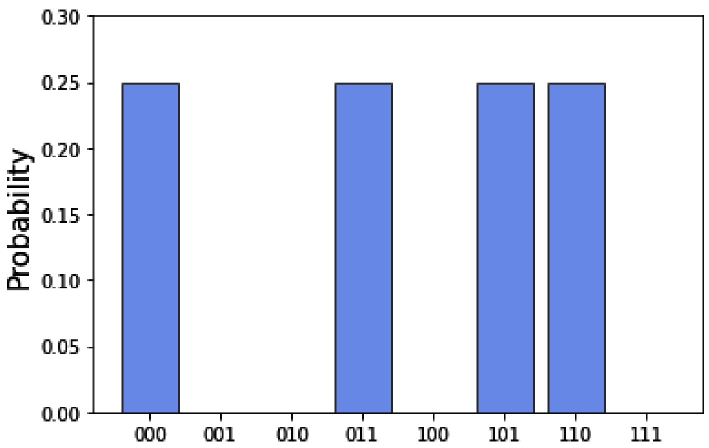
Large Gradients are not enough

Locality ≡ Probability Marginals

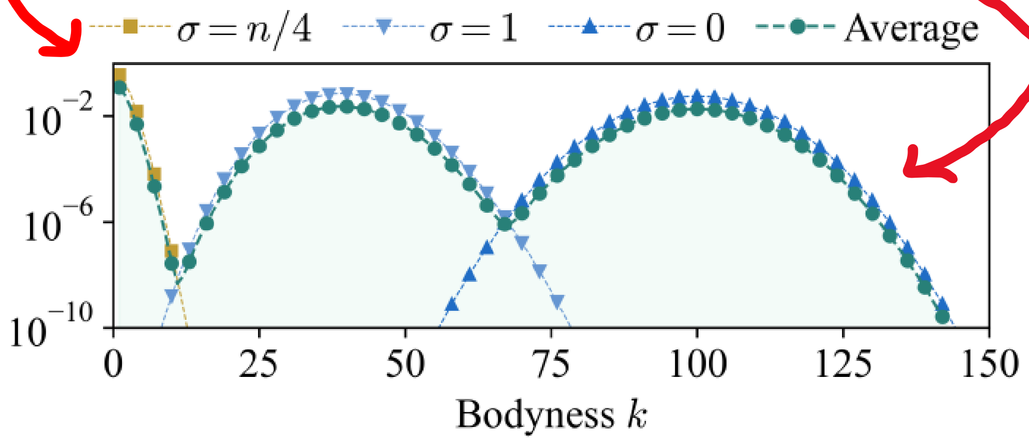
If loss function is **at most k-local,**

it cannot learn beyond **k-order marginals.**





1. Local losses cannot teach global correlations.
2. Global losses are not trainable.



We require
Full-Body Losses
or
Variable Bodyness

e.g., for QCBMs

e.g., for QGANs

Quantum Strategies

Quantum Strategies

Quantum fidelity loss

$$\mathcal{L}_{QF}(\boldsymbol{\theta}) = 1 - |\langle 0|U^\dagger(\boldsymbol{\theta})|\phi\rangle|^2$$

Data target state

$$|\phi\rangle = \sum_{\boldsymbol{x}} \sqrt{\tilde{p}(\boldsymbol{x})}\, |\boldsymbol{x}\rangle$$

Quantum Strategies

Quantum fidelity loss

$$\mathcal{L}_{QF}(\boldsymbol{\theta}) = 1 - |\langle 0|U^{\dagger}(\boldsymbol{\theta})|\phi\rangle|^2$$

Data target state

$$|\phi\rangle = \sum_{\boldsymbol{x}} \sqrt{\tilde{p}(\boldsymbol{x})}\,|\boldsymbol{x}\rangle$$

Local measurement

$$|0\rangle\langle 0| \rightarrow \frac{1}{n} \sum_{i=1\ldots n} |0_i\rangle\langle 0_i| \otimes I_{\bar{i}}$$

Quantum Strategies

Quantum fidelity loss

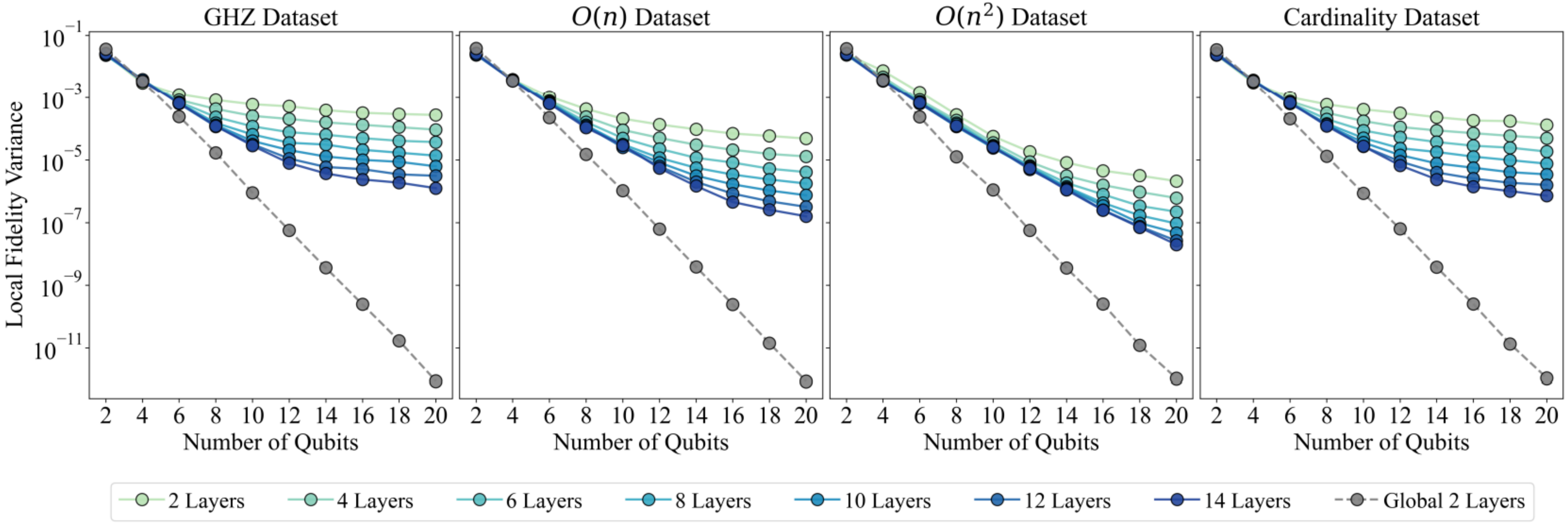
$$\mathcal{L}_{QF}(\boldsymbol{\theta}) = 1 - |\langle 0|U^\dagger(\boldsymbol{\theta})|\phi\rangle|^2$$

Data target state

$$|\phi\rangle = \sum_{\boldsymbol{x}} \sqrt{\tilde{p}(\boldsymbol{x})}\,|\boldsymbol{x}\rangle$$

Local measurement

$$|0\rangle\langle 0| \rightarrow \frac{1}{n}\sum_{i=1\ldots n} |0_i\rangle\langle 0_i| \otimes I_{\bar{i}}$$



Quantum Strategies

… CAN BE BOTH TRAINABLE AND FAITHFUL

Quantum fidelity loss

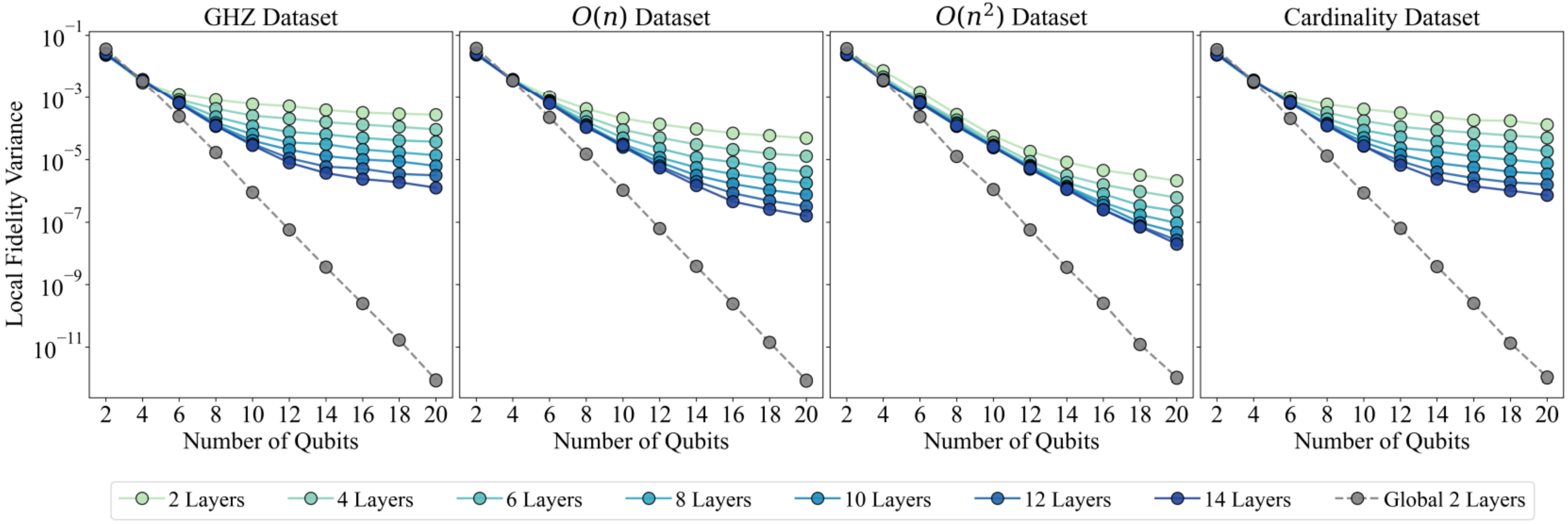
$$\mathcal{L}_{QF}(\boldsymbol{\theta}) = 1 - |\langle 0|U^{\dagger}(\boldsymbol{\theta})|\phi\rangle|^2$$

Data target state

$$|\phi\rangle = \sum_{\boldsymbol{x}} \sqrt{\tilde{p}(\boldsymbol{x})}\,|\boldsymbol{x}\rangle$$

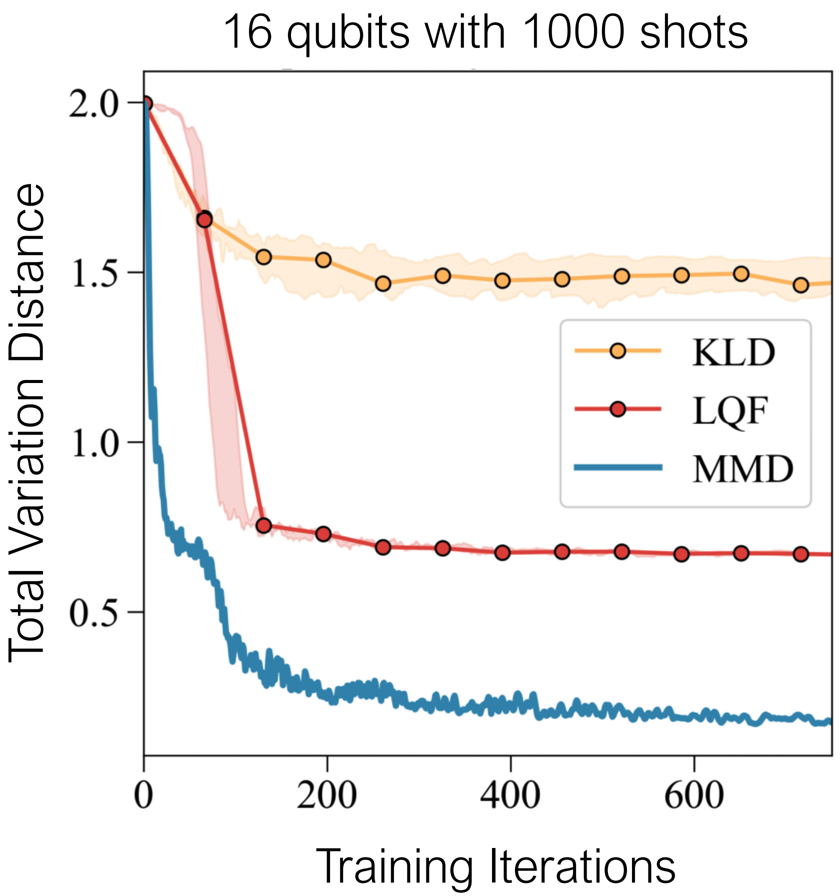
Local measurement

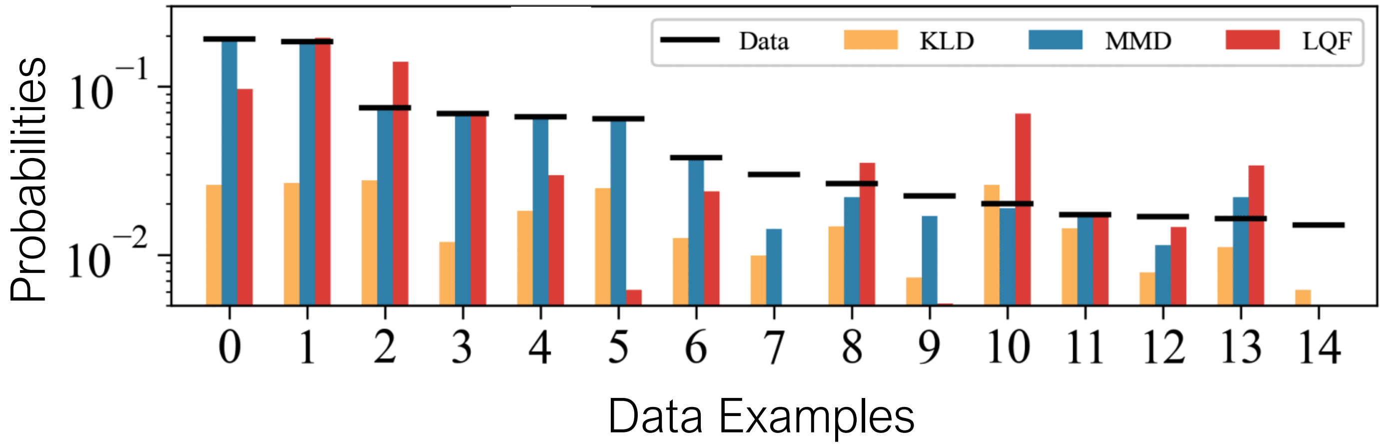
$$|0\rangle\langle 0| \rightarrow \frac{1}{n}\sum_{i=1\ldots n} |0_i\rangle\langle 0_i| \otimes I_{\bar{i}}$$



Final Benchmarks

Dataset: Simulated CERN particle jets





Summary

Circuit depth	Explicit loss (pairwise): Conventional strategy	Explicit loss (pairwise): Quantum strategy	Implicit loss (MMD)
Product	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow			**Yes** ($\sigma \in \Theta(n)$, Conjecture 1)
Deep		**No** [22, 30]	**No** [22, 30]

M. S. Rudolph*, S.Lerch*, S.Thanasilp*,
O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes.
"Trainability barriers and opportunities in quantum generative modeling". *arXiv:2305.02881*.



Summary

Explicit losses are a no-go

Circuit depth	Explicit loss (pairwise): Conventional strategy	Explicit loss (pairwise): Quantum strategy	Implicit loss (MMD)
Product	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow			**Yes** ($\sigma \in \Theta(n)$, Conjec- ture 1)
Deep		**No** [22, 30]	**No** [22, 30]

M. S. Rudolph*, S.Lerch*, S.Thanasilp*,
O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes.
"Trainability barriers and opportunities in quantum generative modeling". *arXiv:2305.02881*.

Summary

Explicit losses are a no-go

Implicit losses can be trainable

Circuit depth	Explicit loss (pairwise)		Implicit loss (MMD)
	Conventional strategy	Quantum strategy	
Product	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow			**Yes** ($\sigma \in \Theta(n)$, Conjecture 1)
Deep		**No** [22, 30]	**No** [22, 30]

M. S. Rudolph*, S.Lerch*, S.Thanasilp*, O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes. “Trainability barriers and opportunities in quantum generative modeling”. *arXiv:2305.02881*.



Summary

M. S. Rudolph*, S.Lerch*, S.Thanasilp*,
O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes.
"Trainability barriers and opportunities in quantum generative modeling". *arXiv:2305.02881*.

Explicit losses are a no-go

Implicit losses can be trainable

Expressivity BP

<table>
<tr><th rowspan="2">Circuit depth</th><th colspan="2">Explicit loss (pairwise)</th><th rowspan="2">Implicit loss (MMD)</th></tr>
<tr><th>Conventional strategy</th><th>Quantum strategy</th></tr>
<tr><td>Product</td><td rowspan="3">No (Corollary 2)</td><td rowspan="2">Yes (Local Quantum Fidelity [31])</td><td>Yes ($\sigma \in \Theta(n)$, Theorem 2)</td></tr>
<tr><td>Shallow</td><td>Yes ($\sigma \in \Theta(n)$, Conjecture 1)</td></tr>
<tr><td>Deep</td><td>No [22, 30]</td><td>No [22, 30]</td></tr>
</table>

Summary

M. S. Rudolph*, S.Lerch*, S.Thanasilp*,
O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes.
"Trainability barriers and opportunities in quantum generative modeling". *arXiv:2305.02881*.

Explicit losses are a no-go

Implicit losses can be trainable

Circuit depth	**Explicit loss (pairwise)** Conventional strategy	**Explicit loss (pairwise)** Quantum strategy	**Implicit loss (MMD)**
Product	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow			**Yes** ($\sigma \in \Theta(n)$, Conjecture 1)
Deep		**No** [22, 30]	**No** [22, 30]

Expressivity BP

Possibly unfaithful

Summary

M. S. Rudolph*, S.Lerch*, S.Thanasilp*,
O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes.
“Trainability barriers and opportunities in quantum generative modeling”. *arXiv:2305.02881*.

Explicit losses are a no-go

Implicit losses can be trainable

Circuit depth	Explicit loss (pairwise)		Implicit loss (MMD)
	Conventional strategy	Quantum strategy	
Product			**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Conjecture 1)
Deep		**No** [22, 30]	**No** [22, 30]

Expressivity BP

Quantum strategies?

Possibly unfaithful

Summary

- Require **inductive bias**!
- Require **good initializations**!
- **Generalization**?

M. S. Rudolph*, S.Lerch*, S.Thanasilp*,
O. Kiss, S. Vallecorsa, M. Grossi, & Z. Holmes.
"Trainability barriers and opportunities in quantum generative modeling". *arXiv:2305.02881*.



Explicit losses are a no-go

Implicit losses can be trainable

Circuit depth	Explicit loss (pairwise)		Implicit loss (MMD)
	Conventional strategy	Quantum strategy	
Product	**No** (Corollary 2)	**Yes** (Local Quantum Fidelity [31])	**Yes** ($\sigma \in \Theta(n)$, Theorem 2)
Shallow			**Yes** ($\sigma \in \Theta(n)$, Conjecture 1)
Deep		**No** [22, 30]	**No** [22, 30]

Expressivity BP

Quantum strategies?

Possibly unfaithful