A Path Towards Quantum Advantage in Training Deep Generative Models with Quantum Annealing

NASA
STINGER GHAFFARIAN TECHNOLOGIES
KBRwyle

How to use quantum annealers to provide quantum advantage on real applications?

- The good:

    1. Quantum annealers are competitive with state-of-the-art classical solvers on natively defined problems

[Mandra, Katzgraber, QST 3; King et al, arXiv:1701.04579; Hen et al. PRA 92; Albash and Lidar, PRX 8]

- The bad:

    2. Relatively small number of available qubits

    3. Quasi two-dimensional (spatially local) connectivities

    4. Control and thermal errors

[Troyer, Katzgraber: AQC2019 talks; Albash et al. QST 4]

- The ugly:

    1. We still don't know how to exploit the good to solve useful problems

Two use cases for quantum annealing

1. Optimization:
   - Quantum annealing developed as a quantum heuristic for optimization
   - Well-established suite of tools for benchmarking: time-to-solution(target) measures, planted solutions…
   - Divide-and-conquer and embedding algorithms face large overheads

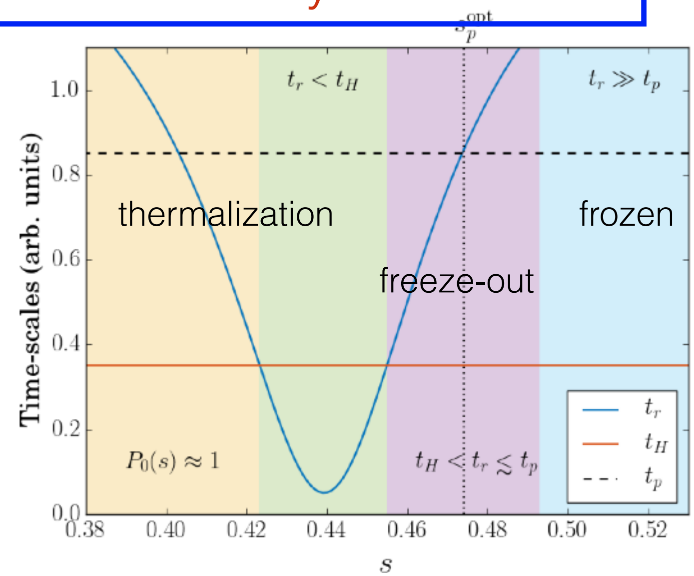
2. Sampling:
   - A more recent application for QA, potential not fully understood
   - Benchmarking is more subtle: KL-divergences are expensive to compute, approximate sampling sufficient for practical applications…
   - Focus on machine learning applications and techniques for 'embeddings'

Quantum annealers are special purpose devices, exploit what they do best:

sampling on native connectivity

- Quantum annealers simulate a transverse field Ising model immersed in a thermal bath

[Amin, PRA 92; Marshall et al., PR Applied 11]



- The required technology is being developed: advanced annealing schedules (pauses, fast quenches, reverse anneals) (D-Wave/QEO)
  - The use of D-Wave quantum annealers as quantum Boltzmann samplers recently demonstrated in material simulations

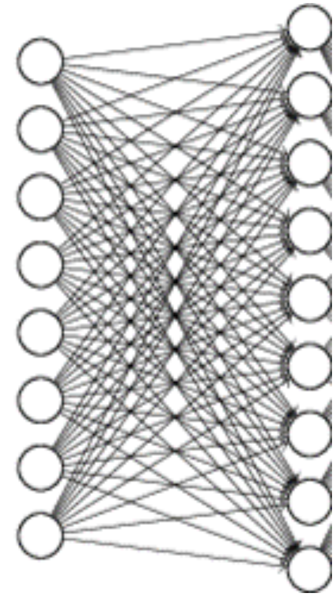
[Harris et al., Science 361; King et al., Nature 560]

- Boltzmann Machines (BM) approximate data distributions as thermal states of classical spin-systems

$$p_{\boldsymbol{\theta}}(\mathbf{z}) \equiv e^{-E_{\boldsymbol{\theta}}(\mathbf{z})}/Z_{\boldsymbol{\theta}}\,, \quad Z_{\boldsymbol{\theta}} \equiv \sum_{\mathbf{z}} e^{-E_{\boldsymbol{\theta}}(\mathbf{z})}$$

$$E_{\boldsymbol{\theta}}(\mathbf{z}) = \sum_{l} z_l h_l + \sum_{l<m} W_{lm} z_l z_m\,, \quad \mathbf{h}, \mathbf{W} \in \{\boldsymbol{\theta}\}$$

$= x$



visible units x

latent units z

  - Boltzmann Machines (BM) and QA: a perfect match?

[Smolensky, '86]

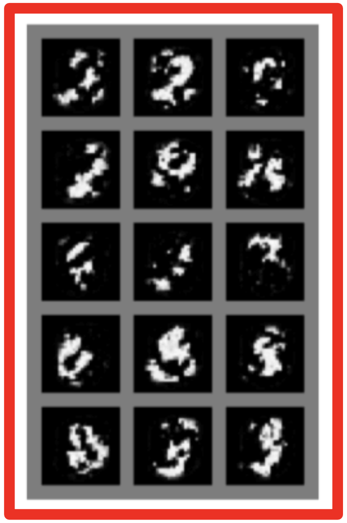
- Training BM requires Boltzmann sampling:
  - State-of-the-art sampling techniques: (Persistent) Contrasting Divergence (P)CD
  - Difficult to scale to large, more powerful BM [Hinton, Science 313]

$\boxed{\text{Employ quantum annealers for faster, more scalable sampling}}$

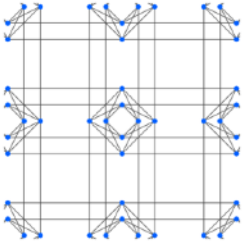
- Connectivity is a very important factor
  - Generative performance of BM on Chimera graph: disappointing

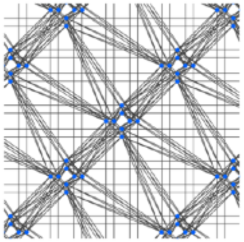
[Dumoulin et al., AAAI Conf, '14]

[MNIST 50k handwritten digits]

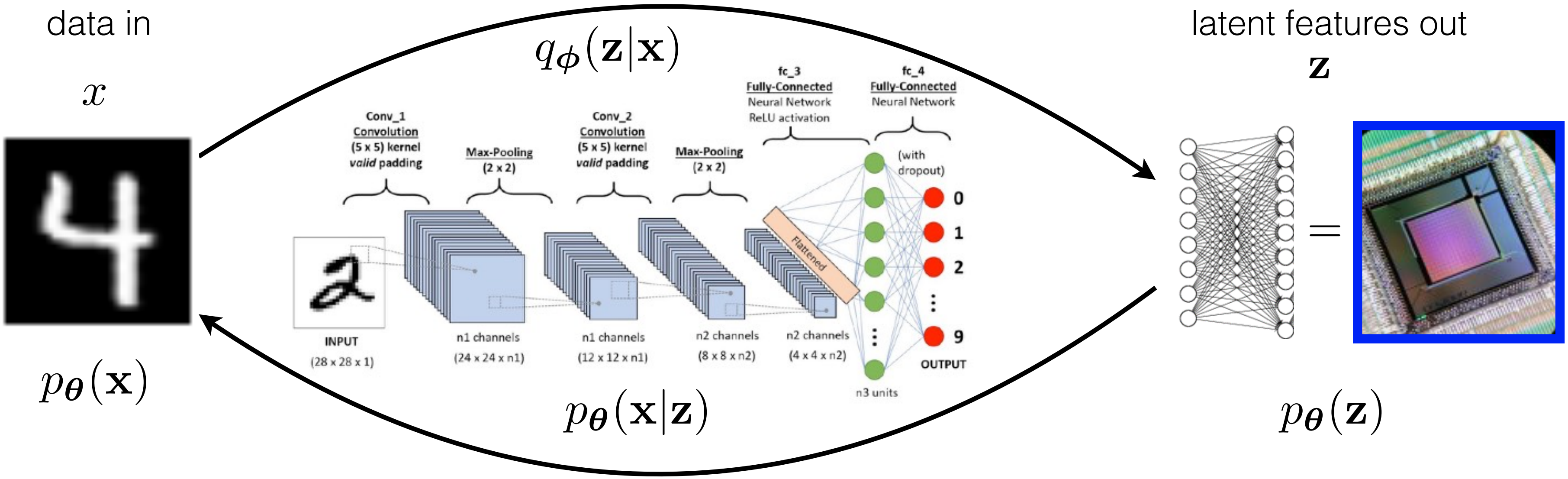


- Rely on technological improvements
  - Chimera (D-Wave) -> Pegasus (D-Wave) -> QEO program





  - Only quasi two-dimensional connectivities available for the foreseeable future
- Rely on common embedding techniques
  - Sampling quality decreases dramatically, likely not scalable



[Benedetti et al., QST 3]

- Quantum/classical joint training
  - Extract the most suitable features for the quantum device
  - Hard-coded specification of the connectivity not required

Generative Adversarial Networks (GAN)

[Wilson et al., arXiv:1904.10573]

Variational Autoencoders (VAE)

[Vinci et al., arXiv:1912.02119]

Invertible Flows

[work in progress]

- Implemented a deep convolutional VAE with discrete latent space
  - 288-dimensional latent space
  - Prior is a Chimera-structured (C6) RBM
- Model trained end-to-end using only samples obtained from the quantum annealer
- Successful training validated by estimating the log-likelihood of the model
  - Showed improvement from a trivial classical baseline (Bernoulli)
  - Match performance of model trained with Population Annealing (PA)

[Khoshaman, Amin, NIPS 2018]

[Trained on MNIST]

[Samples generated with D-Wave 2000Q]

[state-of-the-art: LL~-79.5]

MNIST (dynamic binarization) **LL**		
Sampler	Chimera	Bernoulli
DW2000Q	-82.8 ± 0.2	-83.7 ± 0.2
PA	-82.8 ± 0.1	-84.2 ± 0.05

Reliably sample from large RBM, representing complex multi-modal probability distributions

- Larger RBM
  - Model building to improve use of physical connectivities
  - Develop denser physical connectivities
  - More complex datasets
- Multi-modality
  - Latent-space RBM must develop multi-modal distributions
- Sampling reliability
  - Reduce control errors for more reliable training

Latent-space exploitation with VAEs

$$\mathbf{ELBO} = \mathbb{E}_{\mathbf{x}\sim p_{\mathrm{data}}}\left[\mathbb{E}_{\boldsymbol{\zeta}\sim q_{\boldsymbol{\phi}}(\boldsymbol{\zeta}|\mathbf{x})}\left[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\boldsymbol{\zeta}) - \frac{\log q_{\boldsymbol{\phi}}(\boldsymbol{\zeta}|\mathbf{x})}{\log p_{\boldsymbol{\theta}}(\boldsymbol{\zeta})}\right]\right]$$

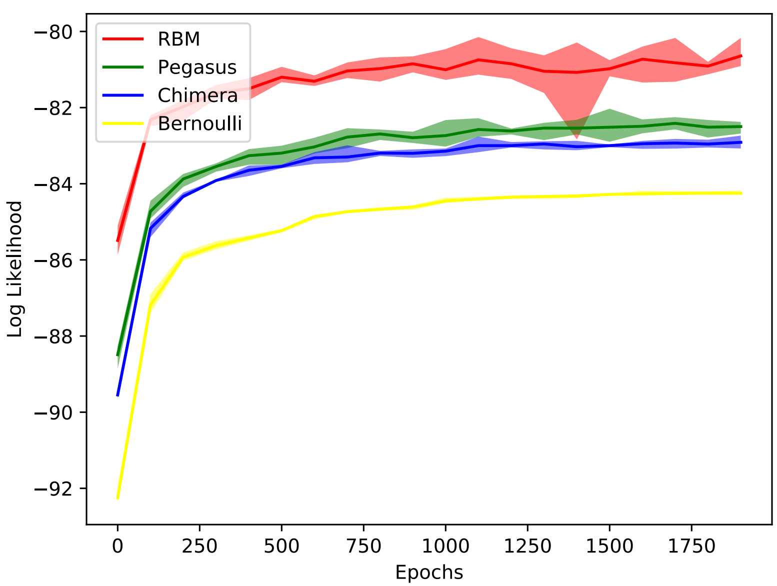
reconstruction

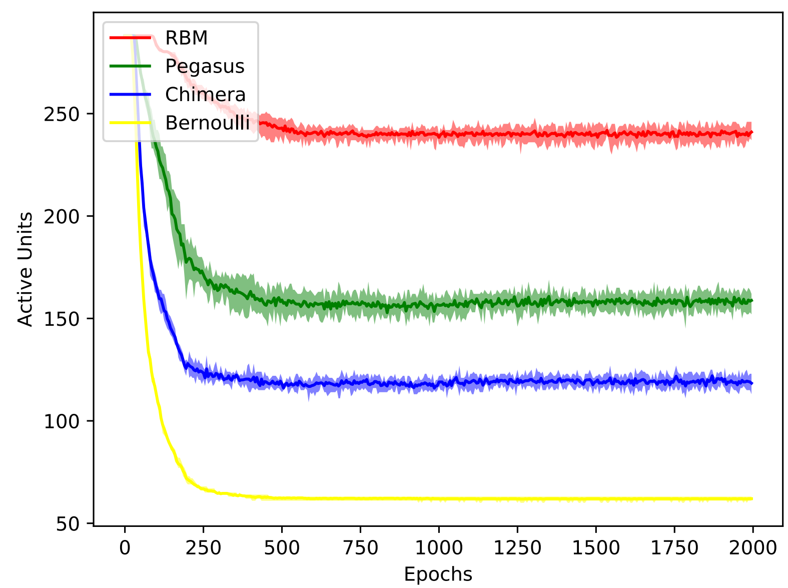
Kl-regularization

Efficient compression: latent units are not used if not necessary

Optimization problem: local minima with sub-optimal number of active units

Denser connectivities = exploit more latent units to achieve better LL





Multimodality in the latent space is not necessary for generative modeling

- RBM can model multimodal distribution, but will they?
  - Block Gibbs sampling from trained RBM:

Bernoulli Chimera Pegasus RBM

Sampling from latent space BM is potentially challenging for classical sampling algorithms

- Summary
  - Demonstrated the use of QA as native samplers in training state-of-the-art deep generative models
  - Provided evidence for the possibility of obtaining quantum advantage within this framework
- Future Directions for QEO
  - Develop meaningful metrics for hybrid generative modeling: understand the limits of classical samplers and performance of quantum annealers as physical samplers.
  - Develop machine learning models for better exploiting quantum annealers
  - Improving sampling and stabilize effective temperatures with advanced anneal controls.
  - Representational power of non-stoquastic Boltzmann machines
  - Ground-state sampling with coherent quantum annealing