Title: A quantum-assisted algorithm for sampling applications in machine learning.

Date: Aug 10, 2016 11:45 AM

URL: http://pirsa.org/16080012

Abstract: An increase in the efficiency of sampling from Boltzmann distributions would have a significant impact in deep learning and other machine learning applications. Recently, quantum annealers have been proposed as a potential candidate to speed up this task, but several limitations still bar these state-of-the-art technologies from being used effectively. One of the main limitations is that, while the device may indeed sample from a Boltzmann-like distribution, quantum dynamical arguments suggests it will do so with an instance-dependent effective temperature, different from the physical temperature of the device. Unless this unknown temperature can be unveiled, it might not be possible to effectively use a quantum annealer for Boltzmann sampling. In this talk, we present a strategy to overcome this challenge with a simple effective-temperature estimation algorithm. We provide a systematic study assessing the impact of the effective temperatures in the learning of a kind of restricted Boltzmann machine embedded on quantum hardware, which can serve as a building block for deep learning architectures. We also provide a comparison to k-step contrastive divergence (CD-k) with k up to 100. Although assuming a suitable fixed effective temperature also allows to outperform one step contrastive divergence (CD-1), only when using an instance-dependent effective temperature we find a performance close to that of CD-100 for the case studied here. We discuss generalizations of the algorithm to other more expressive generative models, beyond restricted Boltzmann machines.

A quantum-assisted algorithm for sampling applications in machine learning

Alejandro Perdomo-Ortiz

Research Scientist, Quantum Artificial Intelligence Laboratory

NASA Ames Research Center, Moffett Field, Calif., USA

Collaborators:

M. Benedetti, J. Realpe-Gomez, and R. Biswas.

Benedetti et al. PRA, 94, 022308 (arXiv:1510.07611).

Funding:

Office of the Director of National Intelligence

IARPA

BE THE FUTURE

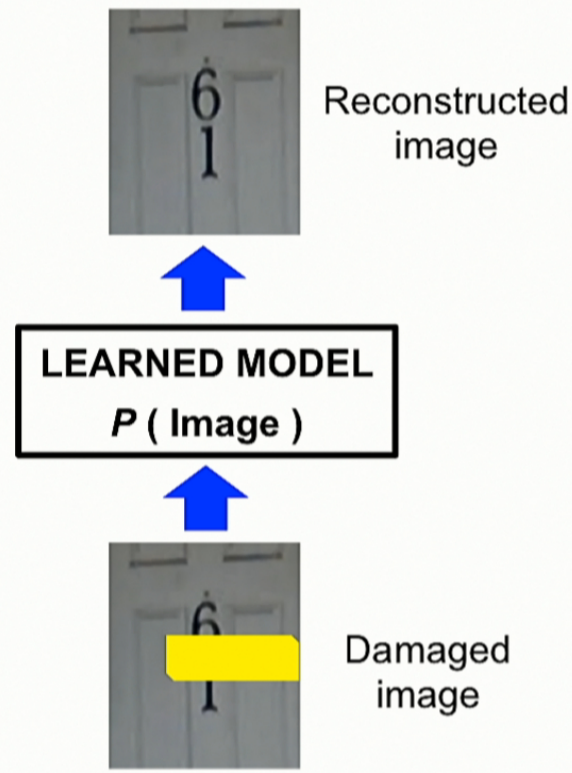
QUANTUM ENHANCED OPTIMIZATION

Quantum Machine Learning @ PI,

Waterloo, Canada, August 10, 2016

Unsupervised learning (generative models)

Example application:
Image reconstruction



2

Unsupervised learning (generative models)

Learn the "best" model distribution that can generate the same kind of data.

MODEL

***P* (Image)**

Learning algorithm

NO LABELS

DATASET

Example application:
Image reconstruction

Reconstructed image

LEARNED MODEL

***P* (Image)**

Damaged image

3

Supervised learning (discriminative models)

Learn the "best" model that can perform a specific task

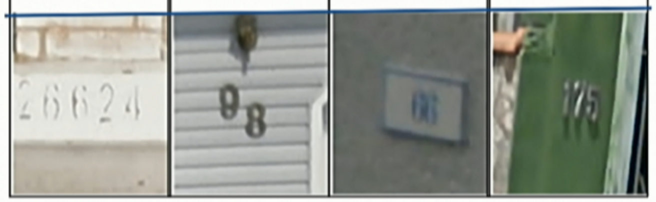
MODEL

***P* (Label | Image)**

Learning algorithm

Labels

26624	98	66	175



DATASET

Example application:
Image recognition

Predicted label

61

LEARNED MODEL

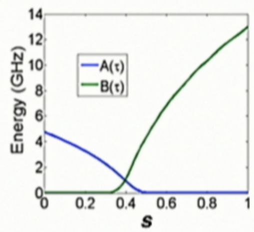
***P* (Label | Image)**

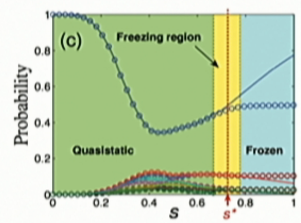
Image to be recognized

4

Outline

- Why is it hard and interesting to sample from a Boltzmann distribution? Why, in principle, is it possible to do classical Gibbs sampling with a quantum annealer?

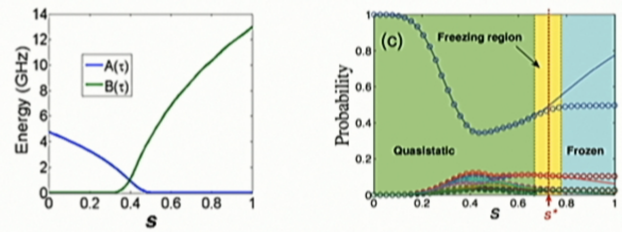




Amin. PRA, 92, 052323 (2015)

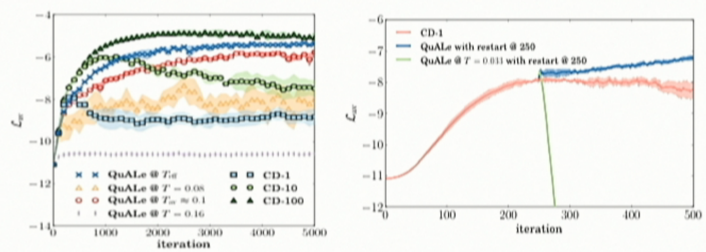
Outline

- Why is it hard and interesting to sample from a Boltzmann distribution? Why, in principle, is it possible to do classical Gibbs sampling with a quantum annealer?



Amin. PRA, 92, 052323 (2015)

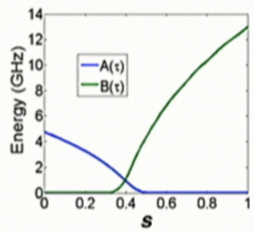
- How to do it experimentally? Results on our quantum-assisted learning (QuALe) algorithm for sampling applications. Feasibility question.

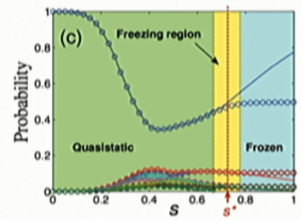


Benedetti et al. PRA, 94, 022308 (arXiv:1510.07611).

Outline

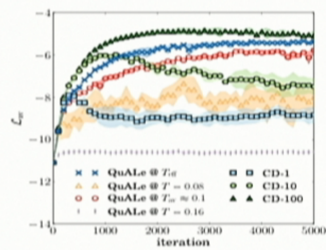
- Why is it hard and interesting to sample from a Boltzmann distribution? Why, in principle, is it possible to do classical Gibbs sampling with a quantum annealer?

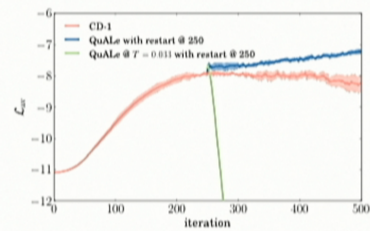




Amin. PRA, 92, 052323 (2015)

- How to do it experimentally? Results on our quantum-assisted learning (QuALe) algorithm for sampling applications. Feasibility question.

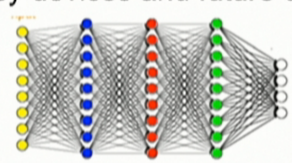




Benedetti et al. PRA, 94, 022308 (arXiv:1510.07611).

- Overcoming the “curse of limited connectivity” in hardware. How to work with general probabilistic graphical models beyond RBM? How to cope with noisy devices and future directions.

General BMs



Deep architectures

Unsupervised learning relies on sampling

"Unsupervised learning [... has] been overshadowed by the successes of purely supervised learning. [... We] expect **unsupervised learning to become far more important in the longer term**. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object."

LeCun, Bengio, Hinton, *Deep Learning*, **Nature 2015**

"In the context of the deep learning approach to undirected modeling, it is rare to use any approach other than Gibbs sampling. **Improved sampling techniques are one possible research frontier.**"

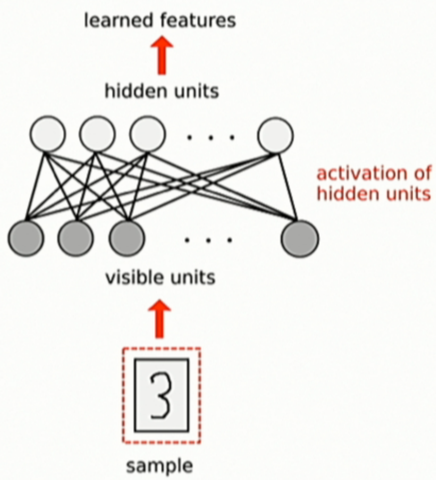
Goodfellow, **Bengio**, Courville, *Deep Learning*, book in preparation for MIT Press, **2016**

"Most of the previous work in **generative models** has focused on variants of **Boltzmann Machines** [...] While these models **are very powerful**, each iteration of **training requires a computationally costly step of MCMC** to approximate derivatives of an intractable partition function (normalization constant), making it **difficult to scale** them **to large datasets.**"

Mansimov, Parisotto, Ba, **Salakhutdinov**, under review for ICLR **2016**

Restricted Boltzmann Machines and Beyond

feature mapping



RBM's:

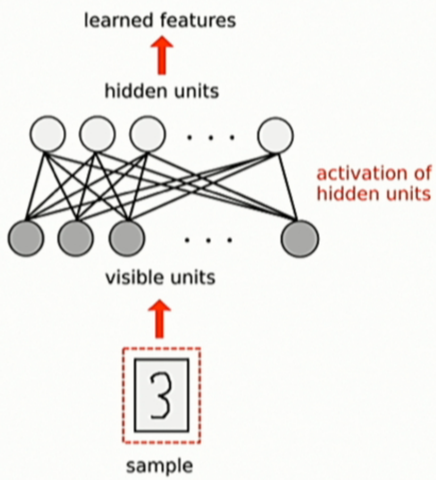
$$E(\boldsymbol{v}, \boldsymbol{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i$$

such that

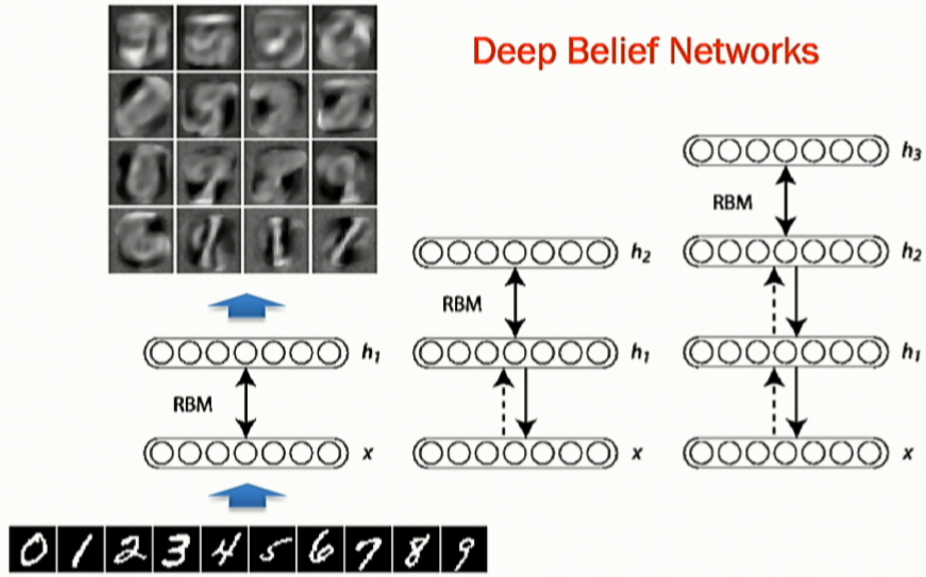
$$p(\boldsymbol{h}|\boldsymbol{v}) = \prod_{i=1}^{n} p(h_i|\boldsymbol{v}) \quad \text{and} \quad p(\boldsymbol{v}|\boldsymbol{h}) = \prod_{j=1}^{m} p(v_j|\boldsymbol{h}).$$

Restricted Boltzmann Machines and Beyond

feature mapping



Deep Belief Networks



RBM's:

$$E(\boldsymbol{v},\boldsymbol{h}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i$$

such that

$$p(\boldsymbol{h}|\boldsymbol{v}) = \prod_{i=1}^{n} p(h_i|\boldsymbol{v}) \quad \text{and} \quad p(\boldsymbol{v}|\boldsymbol{h}) = \prod_{j=1}^{m} p(v_j|\boldsymbol{h}).$$

Model:

$$p(\boldsymbol{v}) = \sum_{\boldsymbol{h}} p(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{Z}\sum_{\boldsymbol{h}} e^{-E(\boldsymbol{v},\boldsymbol{h})},$$

Training Method: Stochastic gradient ascent

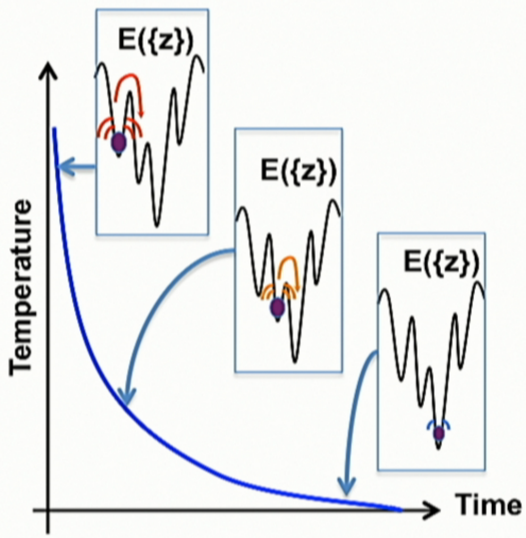
$$\sum_{\boldsymbol{v}\in S} \frac{\partial \ln \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{v})}{\partial w_{ij}} \propto \langle v_i h_j \rangle_{\mathrm{data}} - \langle v_i h_j \rangle_{\mathrm{model}}$$

Foundational Theory of Quantum Annealing

Simulated Annealing

(Kirkpatrick et al., 1983)

- **Algorithm:** Start with a high temperature. Slowly reduce the intensity of these thermal fluctuations aiming for low cost configs..
- Transitions between states are over the barrier and due to thermal fluctuation

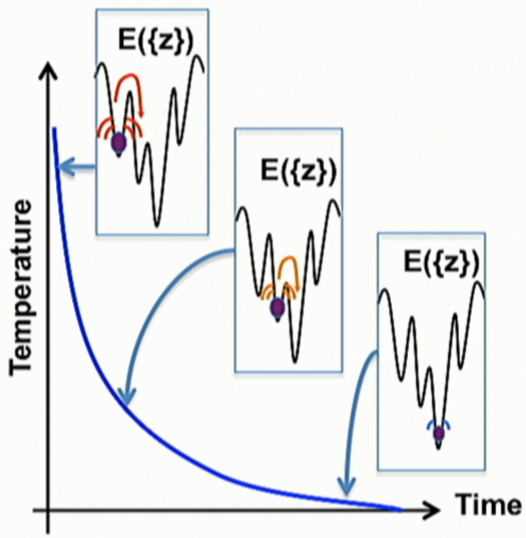


Foundational Theory of Quantum Annealing

Simulated Annealing

(Kirkpatrick et al., 1983)

- **Algorithm:** Start with a high temperature. Slowly reduce the intensity of these thermal fluctuations aiming for low cost configs..
- Transitions between states are over the barrier and due to thermal fluctuation

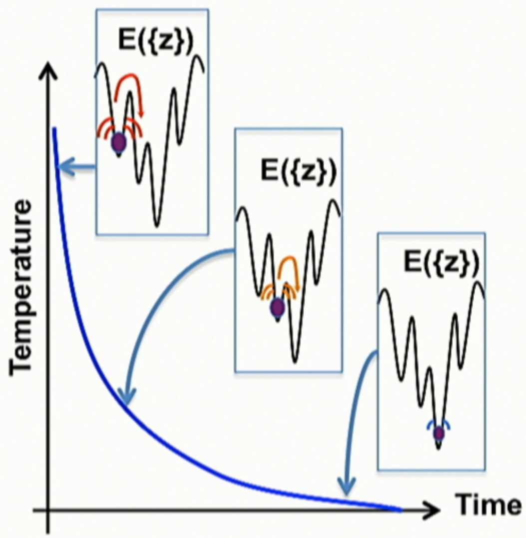


Foundational Theory of Quantum Annealing

Simulated Annealing

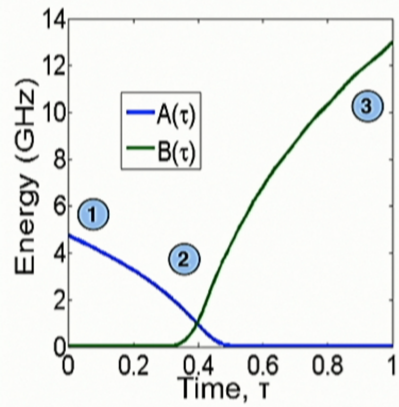
(Kirkpatrick et al., 1983)

- **Algorithm:** Start with a high temperature. Slowly reduce the intensity of these thermal fluctuations aiming for low cost configs..
- Transitions between states are over the barrier and due to thermal fluctuation



Quantum Annealing

(Finnila et al., 1994, Kadawaki and Nishimori, 1998, Farhi et.al., 2001)



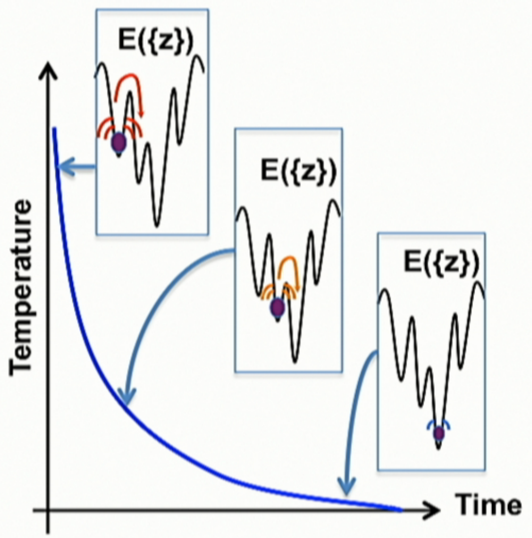
- **Algorithm:** Start with large amplitude $A(\tau)$ responsible for quantum fluctuations. Then, slowly turn it off while turning on the cost function amplitude, $B(\tau)$.
- Transitions between states due to quantum fluctuations (tunneling)

Foundational Theory of Quantum Annealing

Simulated Annealing

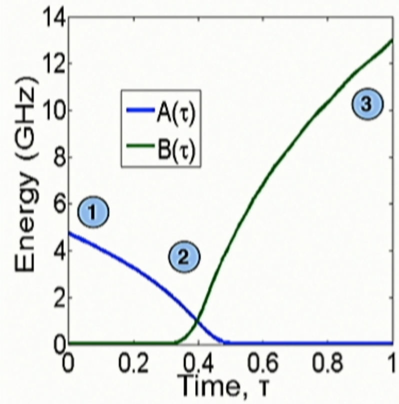
(Kirkpatrick et al., 1983)

- **Algorithm:** Start with a high temperature. Slowly reduce the intensity of these thermal fluctuations aiming for low cost configs..
- Transitions between states are over the barrier and due to thermal fluctuation

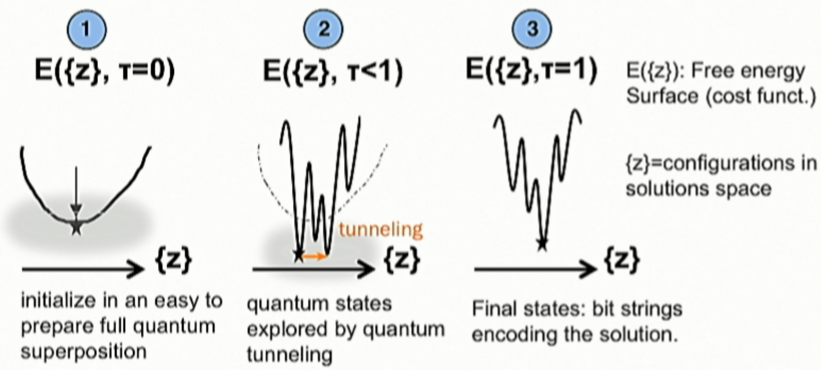


Quantum Annealing

(Finnila et al., 1994, Kadawaki and Nishimori, 1998, Farhi et.al., 2001)



- **Algorithm:** Start with large amplitude A(τ) responsible for quantum fluctuations. Then, slowly turn it off while turning on the cost function amplitude, B(τ).
- Transitions between states due to quantum fluctuations (tunneling)

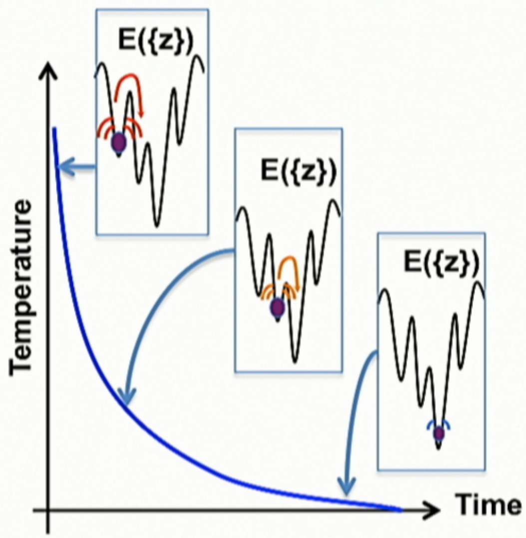


Foundational Theory of Quantum Annealing

Simulated Annealing

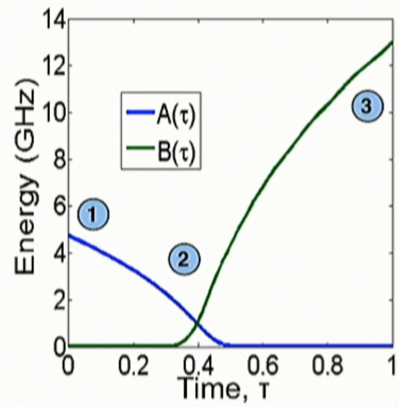
(Kirkpatrick et al., 1983)

- **Algorithm:** Start with a high temperature. Slowly reduce the intensity of these thermal fluctuations aiming for low cost configs..
- Transitions between states are over the barrier and due to thermal fluctuation



Quantum Annealing

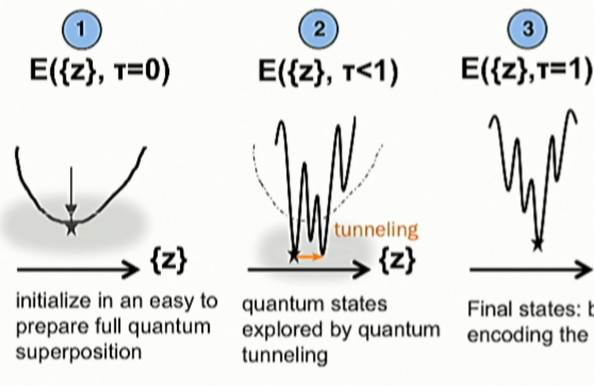
(Finnila et al., 1994, Kadawaki and Nishimori, 1998, Farhi et.al., 2001)



- **Algorithm:** Start with large amplitude A(τ) responsible for quantum fluctuations. Then, slowly turn it off while turning on the cost function amplitude, B(τ).
- Transitions between states due to quantum fluctuations (tunneling)

$$H(\tau) = A(\tau)H_b + B(\tau)H_p$$

$$H_p = \sum_{1 \le i \le N}^{N} h_i \sigma_i^z + \sum_{1 \le i < j \le N}^{N} J_{ij} \sigma_i^z \sigma_j^z$$



E({z}): Free energy Surface (cost funct.)

{z}=configurations in solutions space

D-Wave System Capability

1) As a discrete optimization solver:

Given $\{ h_j , J_{ij} \}$, find $\{ s_k = \pm 1 \}$ that minimizes

NP-hard problem

$$\xi(s_1, \ldots, s_N) = \sum_{j=1}^{N} h_j s_j + \sum_{i,j \in E}^{N} J_{ij} s_i s_j$$

Potential NASA applications: planning, scheduling, fault diagnosis, graph analysis, communication networks, etc.

Also, quantum ML work by Google/DW.

QUBO: Quadratic Unconstrained Binary Optimization (Ising model in physics jargon).

D-Wave System Capability

1) As a discrete optimization solver:

Given $\{ h_j, J_{ij} \}$, find $\{ s_k = \pm 1 \}$ that minimizes

NP-hard problem

$$\xi(s_1, \ldots, s_N) = \sum_{j=1}^{N} h_j s_j + \sum_{i,j \in E}^{N} J_{ij} s_i s_j$$

Potential NASA applications: planning, scheduling, fault diagnosis, graph analysis, communication networks, etc.

Also, quantum ML work by Google/DW.

QUBO: Quadratic Unconstrained Binary Optimization (Ising model in physics jargon).

2) As a physical device to sample from Boltzmann distribution:

$$P_{Boltzmann} \propto \exp[-\xi(s_1, \ldots, s_N) / T_{eff}]$$

Potential NASA applications in machine leaning (e.g., training of deep-learning networks)

Early work:
Bian et al. 2010. The Ising model: teaching an old problem new tricks.

D-Wave System Capability

1) As a discrete optimization solver:

Given $\{ h_j , J_{ij} \}$, find $\{ s_k = \pm 1 \}$ that minimizes

NP-hard problem

$$\xi(s_1, \dots, s_N) = \sum_{j=1}^{N} h_j s_j + \sum_{i,j \in E}^{N} J_{ij} s_i s_j$$

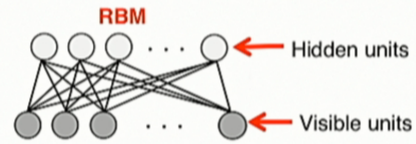
Potential NASA applications: planning, scheduling, fault diagnosis, graph analysis, communication networks, etc.

Also, quantum ML work by Google/DW.

QUBO: Quadratic Unconstrained Binary Optimization (Ising model in physics jargon).

2) As a physical device to sample from Boltzmann distribution:

$$P_{Boltzmann} \propto \exp[-\xi(s_1, \dots, s_N)/T_{eff}]$$



Computationally bottleneck

Widely used in **unsupervised** learning

$$\langle v_i h_j \rangle_{p(\mathbf{h},\mathbf{v})}$$

Potential NASA applications in machine leaning (e.g., training of deep-learning networks)

Early work:
Bian et al. 2010. The Ising model: teaching an old problem new tricks.

Our recent work: Benedetti et al. PRA, 94, 022308 (2015)

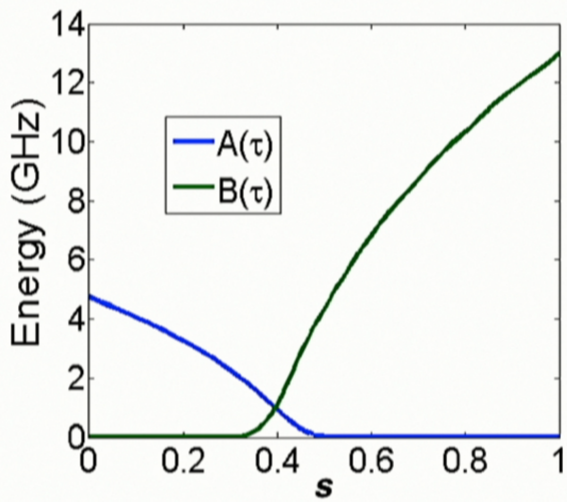
- We provide a robust algorithm to estimate the effective temperature of problem instances in quantum annealers.
- Algorithm uses the same samples that will be used for the estimation of the gradient

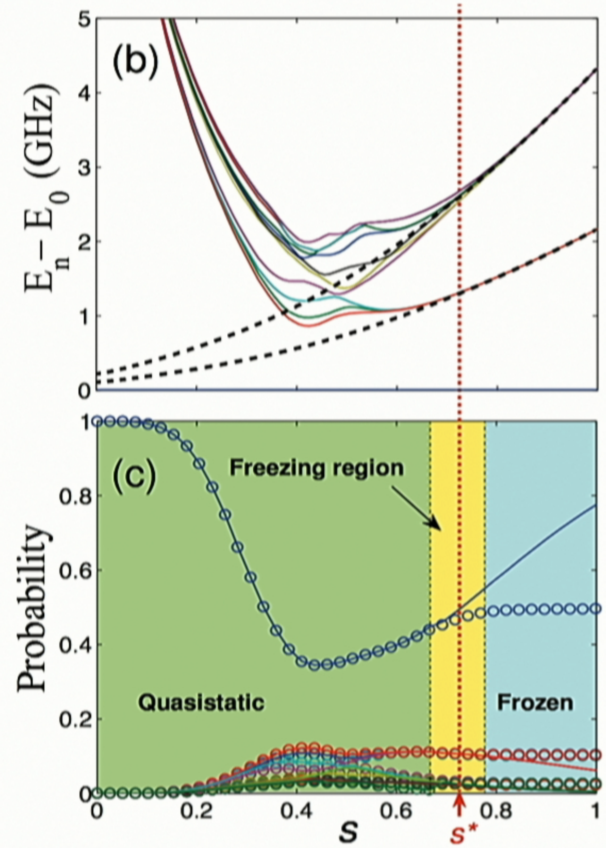
Why sampling from classical Gibbs?

2) As a physical device to sample from Boltzmann distribution:

$$P_{Boltzmann} \propto \exp[-\xi(s_1,\ldots,s_N)/T_{eff}]$$

Potential NASA applications in machine leaning (e.g., training of deep-learning networks)





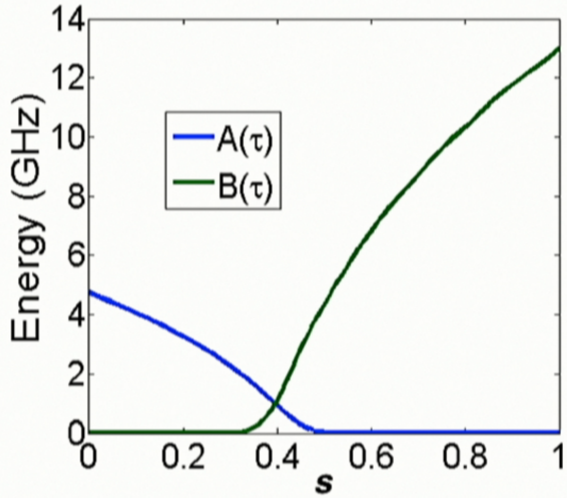
Amin. PRA, 92, 052323 (2015)

Why sampling from classical Gibbs?

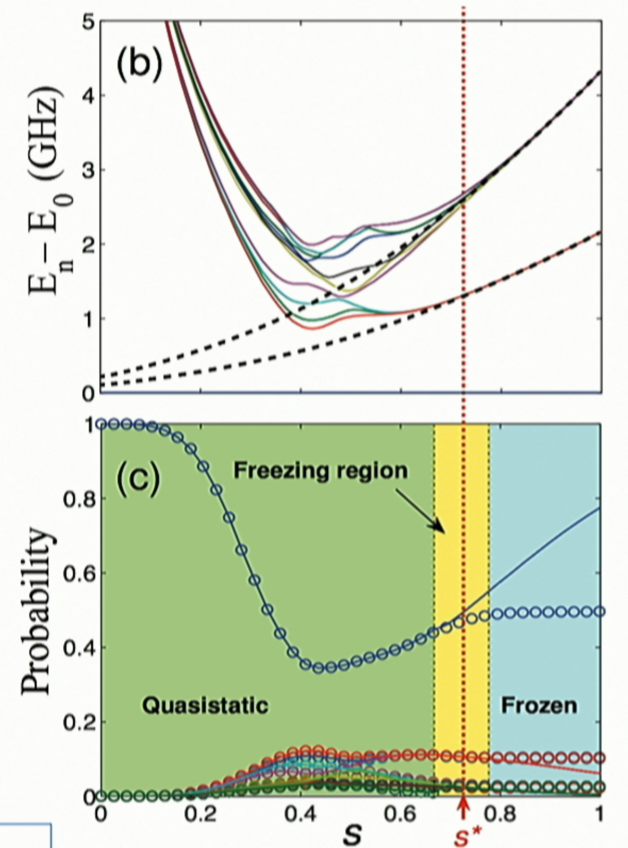
2) As a physical device to sample from Boltzmann distribution:

$$P_{Boltzmann} \propto \exp[-\xi(s_1,\ldots,s_N)/T_{eff}]$$

Potential NASA applications in machine leaning (e.g., training of deep-learning networks)



$T_{eff} > T_{DW2X}$



Amin. PRA, 92, 052323 (2015)

D-Wave System Capability

1) As a discrete optimization solver:

Given $\{ h_j, J_{ij} \}$, find $\{ s_k = \pm 1 \}$ that minimizes

NP-hard problem

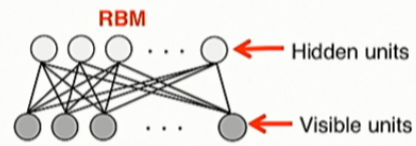
$$\xi(s_1, \ldots, s_N) = \sum_{j=1}^{N} h_j s_j + \sum_{i,j \in E}^{N} J_{ij} s_i s_j$$

Potential NASA applications: planning, scheduling, fault diagnosis, graph analysis, communication networks, etc.

Also, quantum ML work by Google/DW.

QUBO: Quadratic Unconstrained Binary Optimization (Ising model in physics jargon).

2) As a physical device to sample from Boltzmann distribution:



$$P_{Boltzmann} \propto \exp[-\xi(s_1, \ldots, s_N)/T_{eff}]$$

Computationally bottleneck

Widely used in **unsupervised** learning

$$\langle v_i h_j \rangle_{p(\mathbf{h},\mathbf{v})}$$

Potential NASA applications in machine leaning (e.g., training of deep-learning networks)

Early work:
Bian et al. 2010. The Ising model: teaching an old problem new tricks.

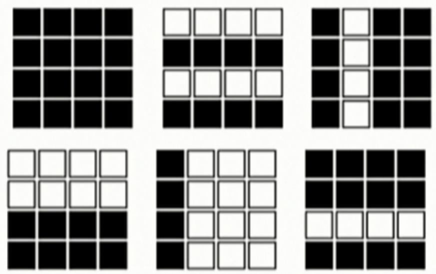
Recent work:
Raymond et al. 2016. Global warming: Temperature estimation in annealers.

Our recent work: Benedetti et al. PRA, 94, 022308 (2015)

- We provide a robust algorithm to estimate the effective temperature of problem instances in quantum annealers.
- Algorithm uses the same samples that will be used for the estimation of the gradient

Quantum-Assisted Learning Vs. Contrastive Divergence

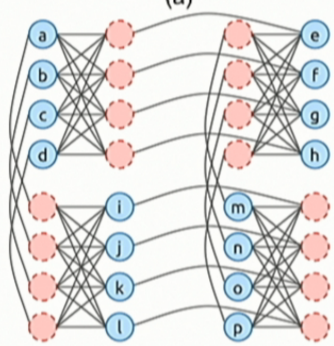
Bars and Stripes dataset



Fisher and Igel. *Pattern Recognition*, 47, 25 (2014)

Embedding on the D-Wave 2X

(a)



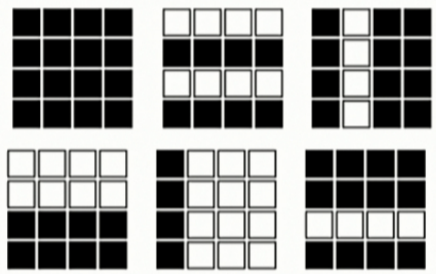
(b)

Pixel blocks embedding

a	b	e	f
c	d	g	h
i	j	m	n
k	l	o	p

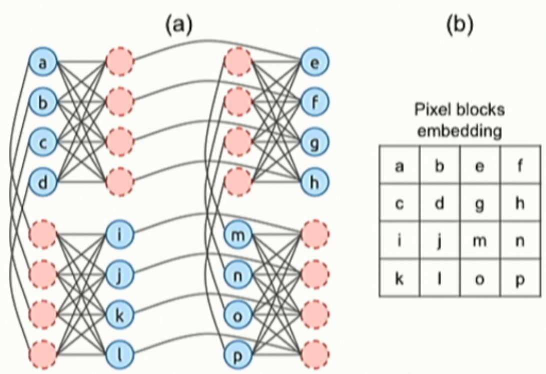
Quantum-Assisted Learning Vs. Contrastive Divergence

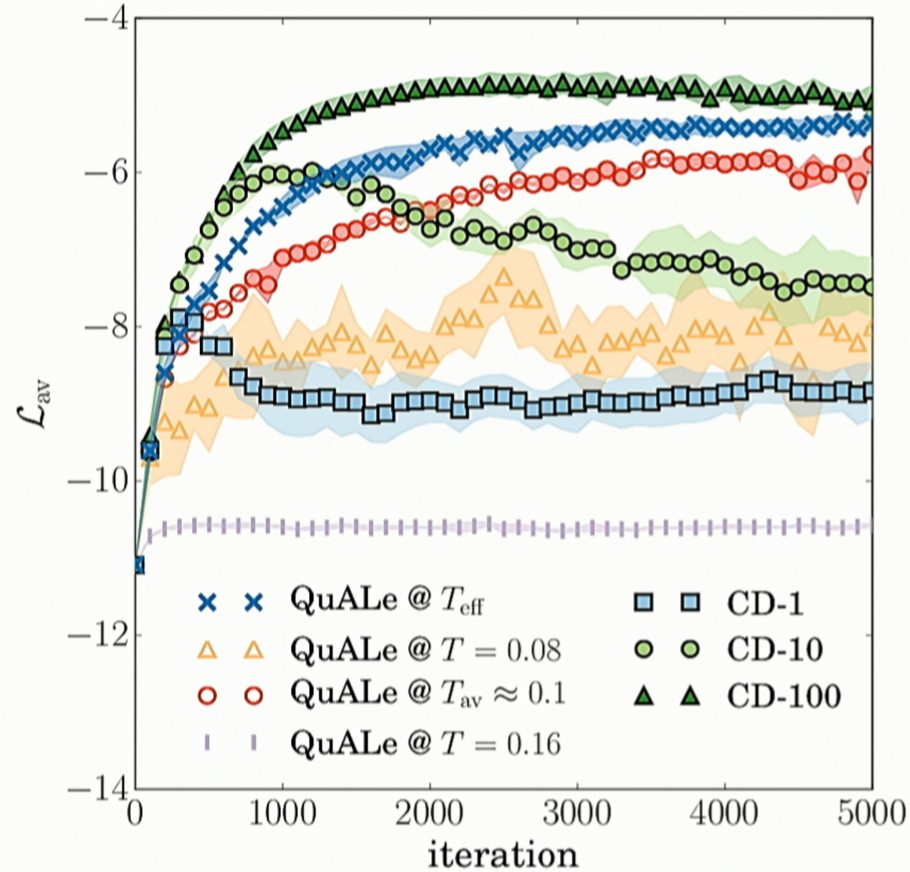
Bars and Stripes dataset



Fisher and Igel. *Pattern Recognition*, 47, 25 (2014)

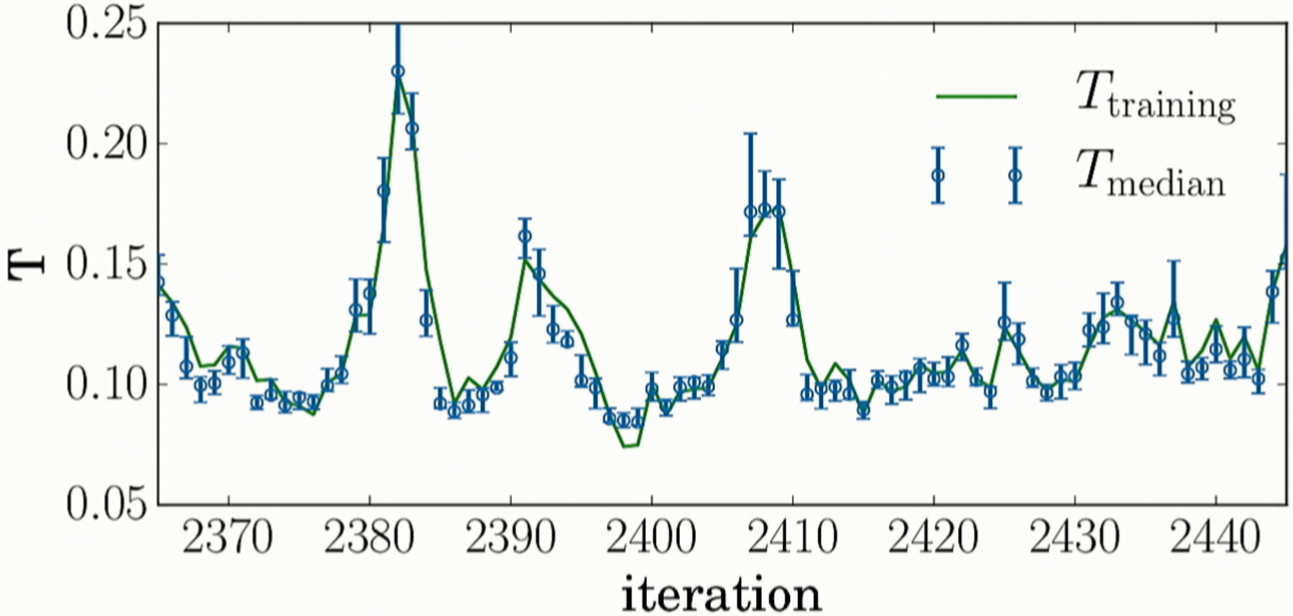
Embedding on the D-Wave 2X





Benedetti et al. PRA, 94, 022308

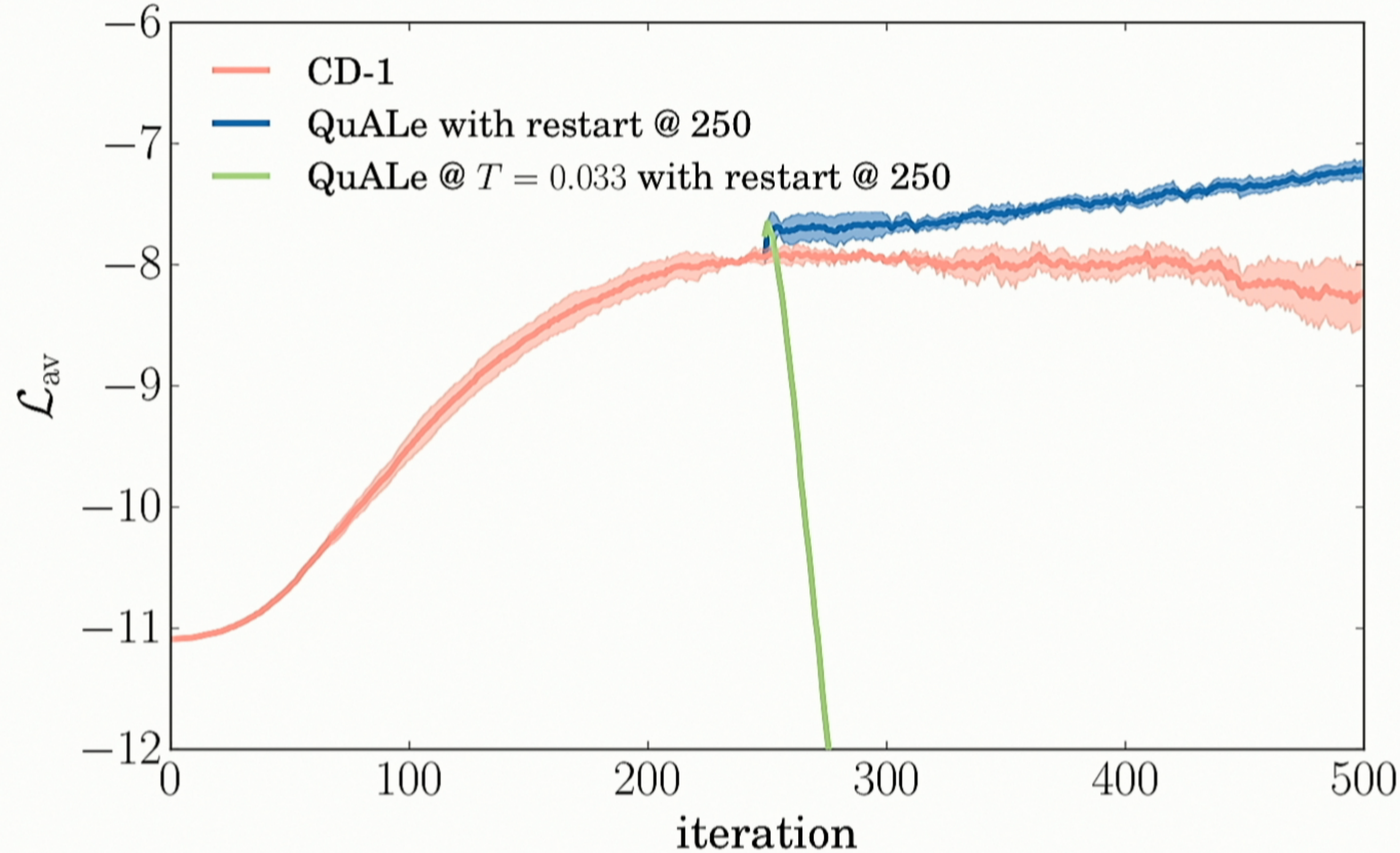
Non-trivial and correlated variations in the temperature



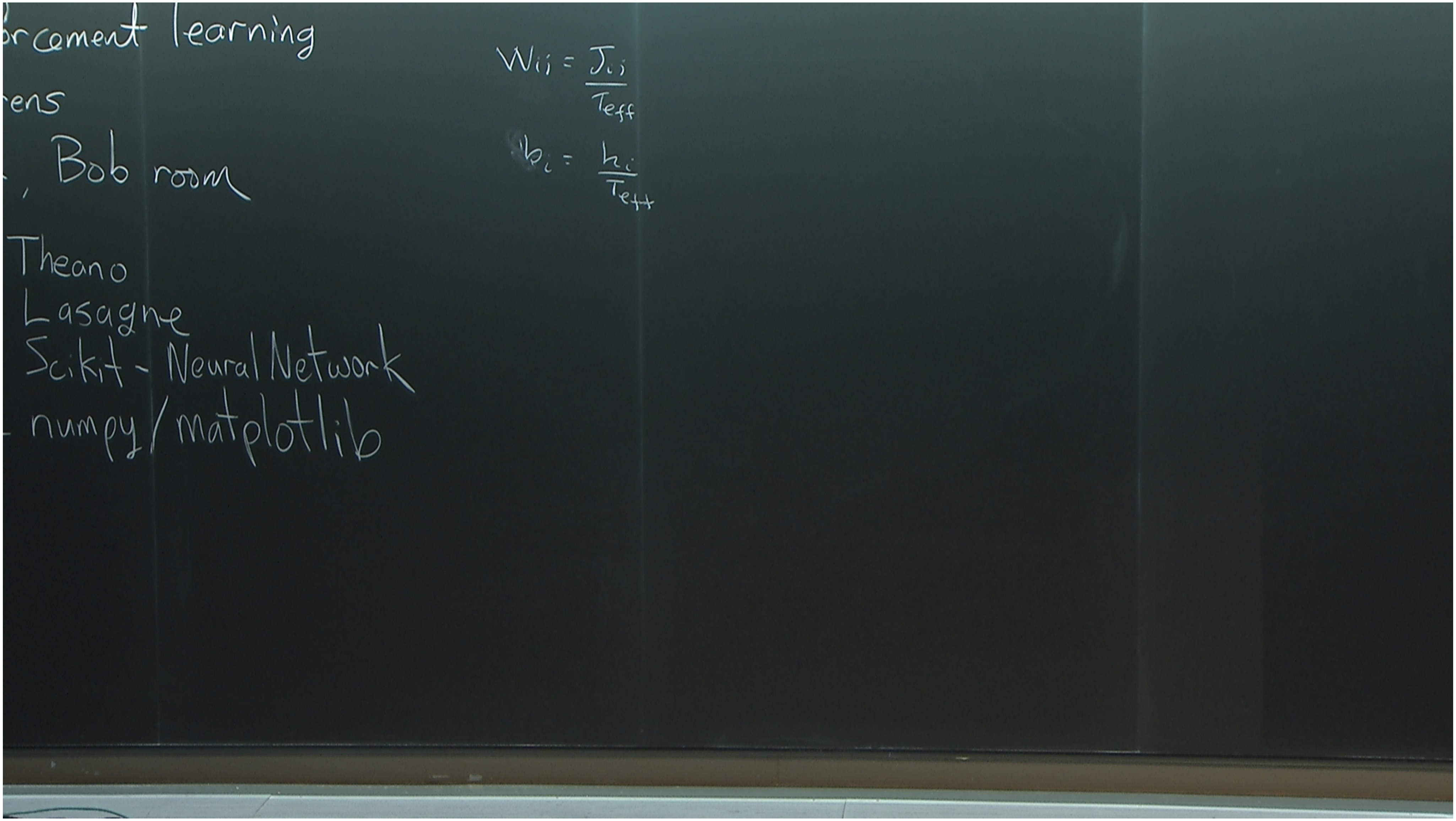
T_{DW2X} = 0.033

Benedetti et al. PRA, 94, 022308

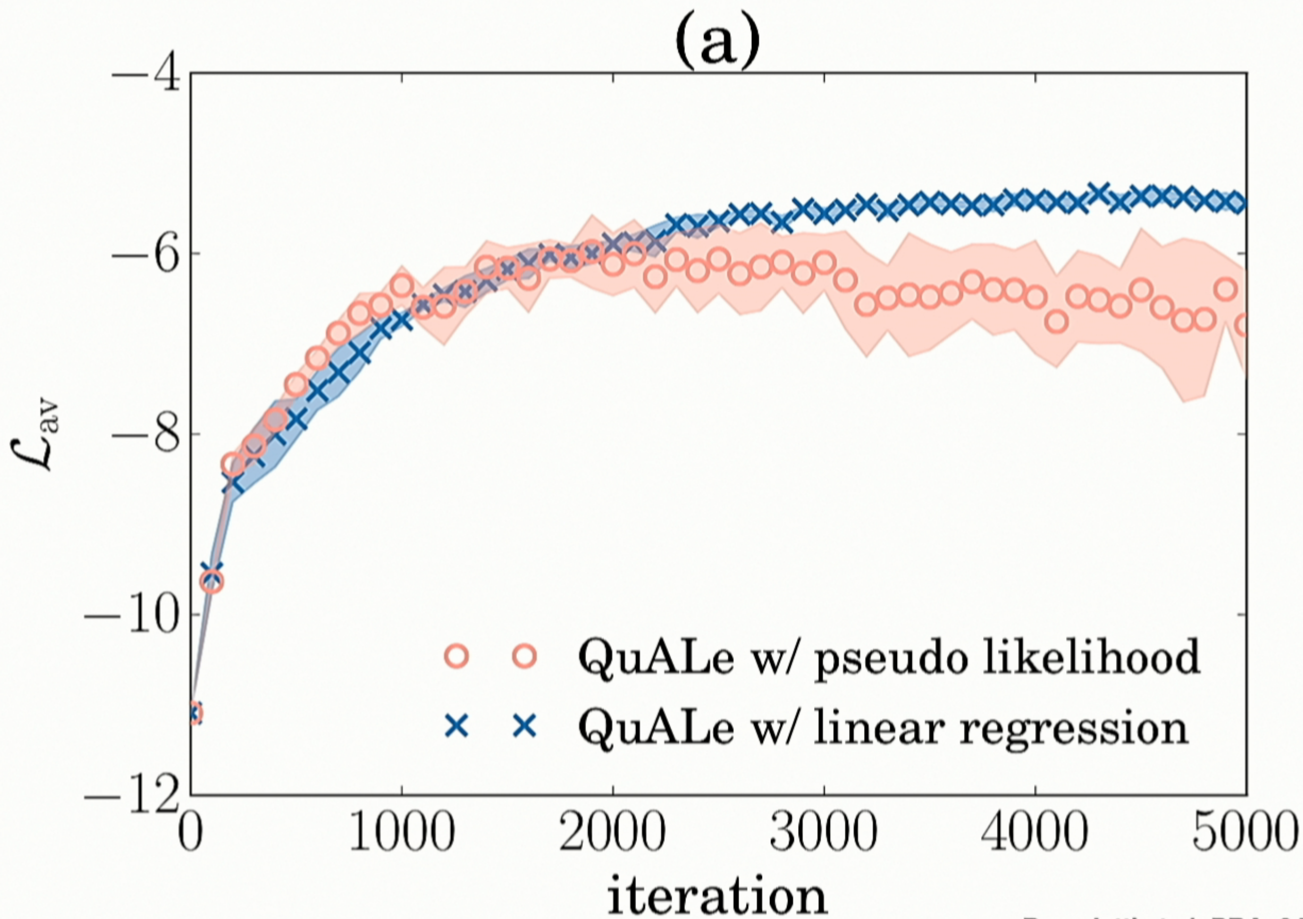
Added features: Restart from CD-k



Benedetti et al. PRA, 94, 022308


rcement learning
ens
, Bob room
Theano
Lasagne
Scikit - Neural Network
numpy/matplotlib
$W_{ij} = \frac{J_{ij}}{T_{eff}}$
$b_i = \frac{h_i}{T_{eff}}$

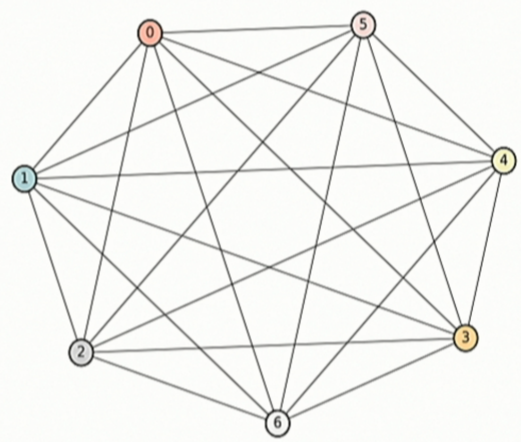
Comparison with pseudo-likelihood



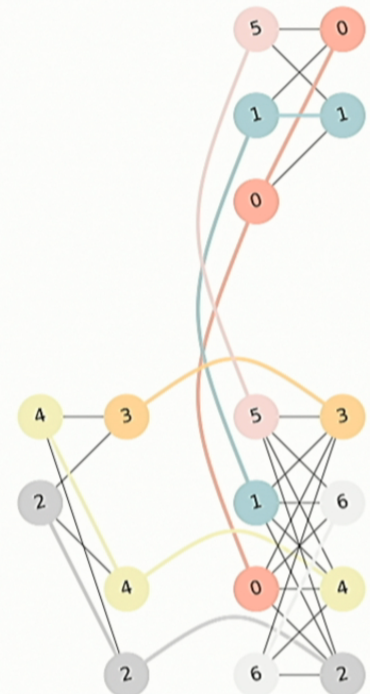
Benedetti et al. PRA, 94, 022308

Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity



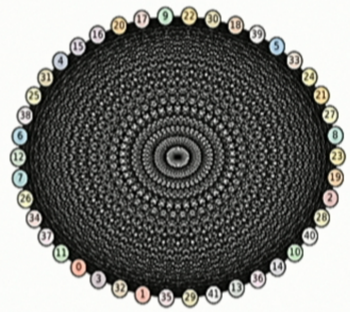
7 logical (visible) variables



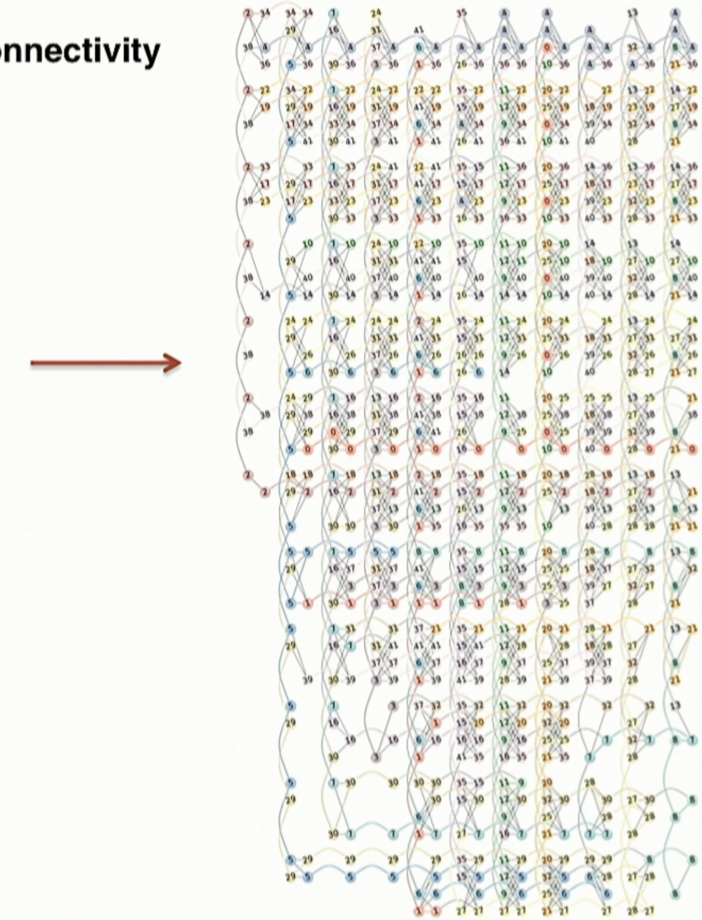
18 physical qubits

Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity in physical devices.



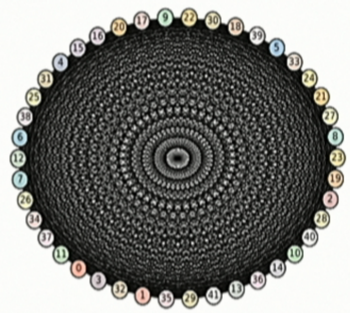
42 fully-connected logical (visible) variables



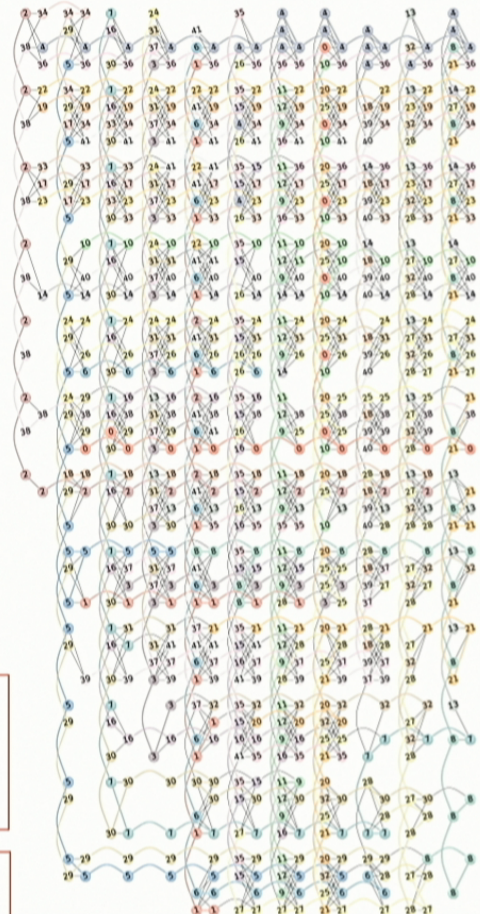
794 physical qubits

Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity in physical devices.



42 fully-connected
logical (visible) variables



794 physical qubits

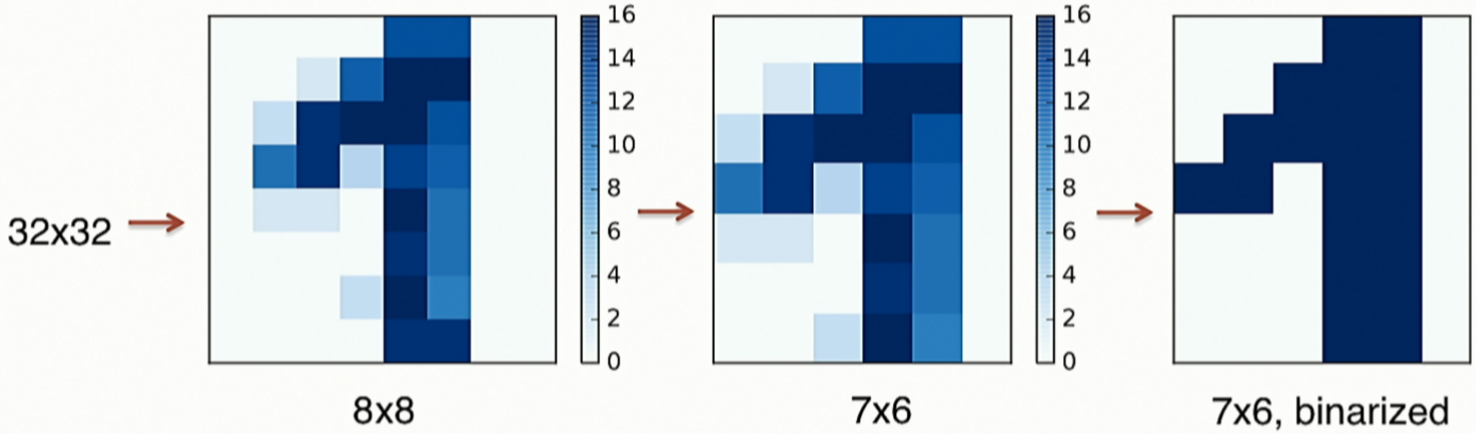
How do we train this 794 qubit problem?
(How do we analyze the (Gibbs) samples from this physical model?

Immediate solution: Keep an eye on a paper coming out with a new gray-model approach for training noisy QA.

Benedetti et al. In preparation.

Quantum-assisted unsupervised learning on digits

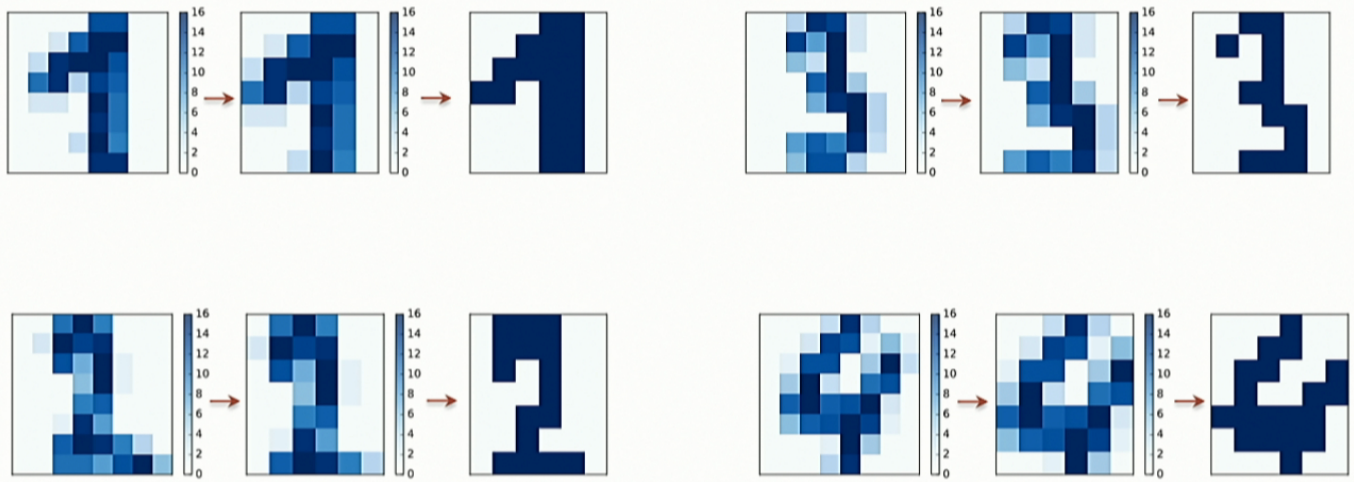
OptDigits Datasets



Dataset: Optical Recognition of Handwritten Digits (OptDigits)

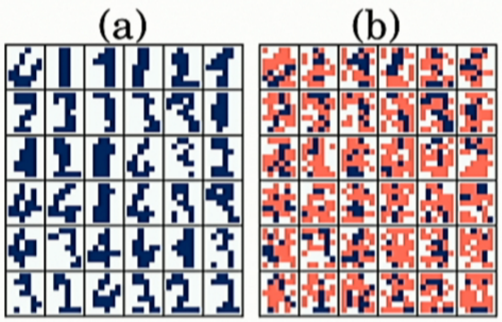
Quantum-assisted unsupervised learning on digits

OptDigits Datasets



Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Quantum-assisted unsupervised learning on digits

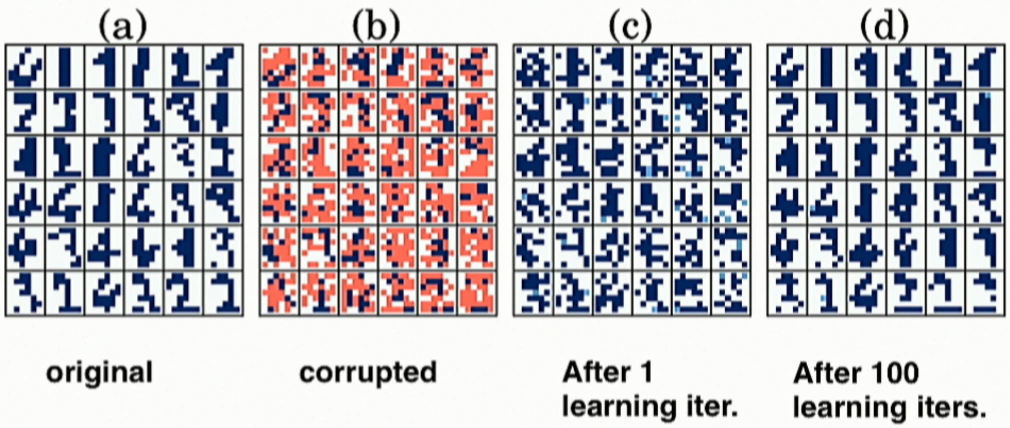


original corrupted

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Benedetti et al. In preparation.

Quantum-assisted unsupervised learning on digits

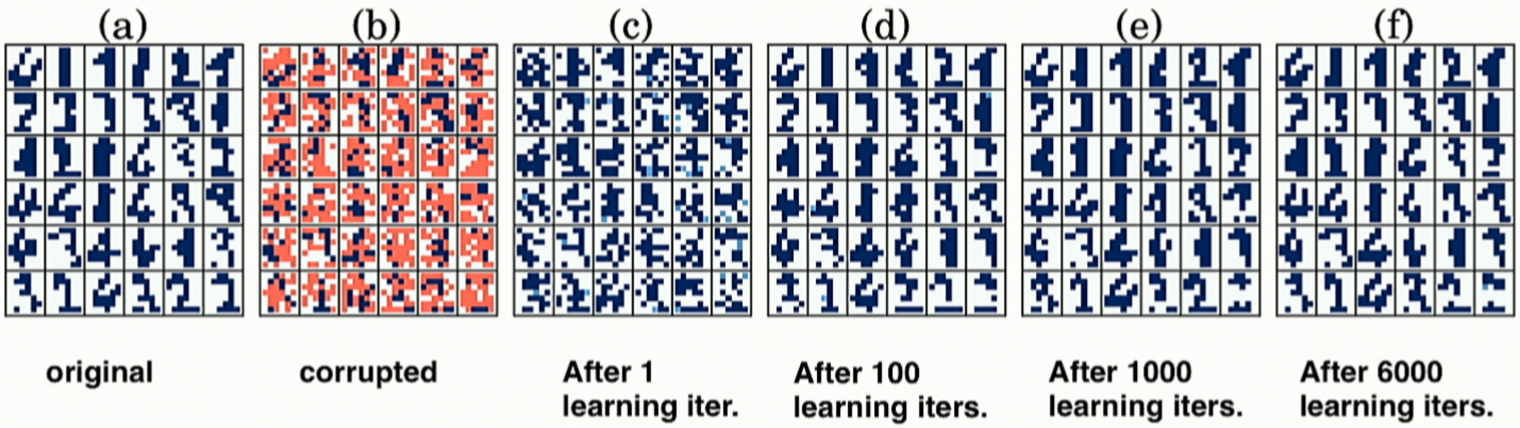


- Experimental realization of quantum-assisted learning algorithm on 794 qubits, for a 42 fully-connected model.
- Fully unsupervised learning and generative model on a digit.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Benedetti et al. In preparation.

Quantum-assisted unsupervised learning on digits



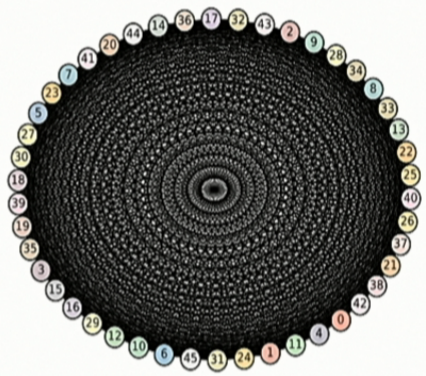
- Experimental realization of quantum-assisted learning algorithm on 794 qubits, for a 42 fully-connected model.
- Fully unsupervised learning and generative model on a digit.

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Benedetti et al. In preparation.

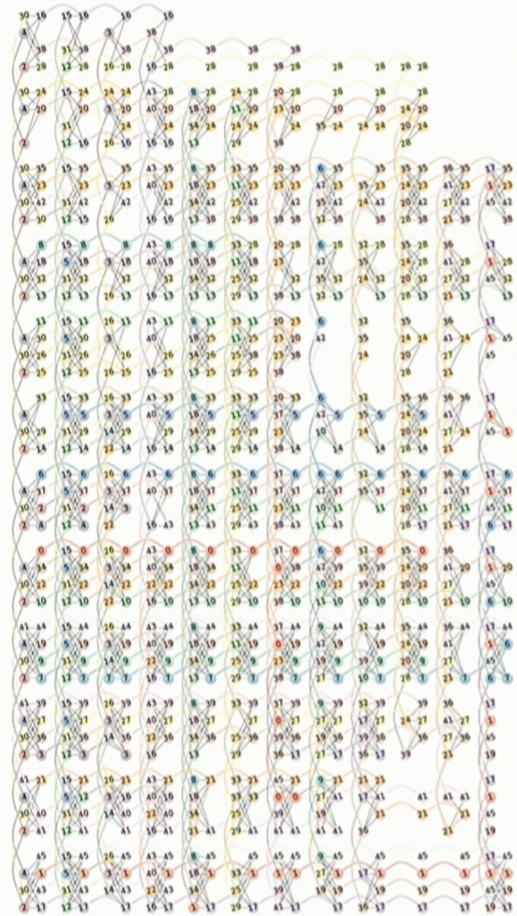
Quantum-assisted unsupervised learning on digits

Overcoming the curse of limited connectivity in hardware.



46 fully-connected
logical (visible) variables

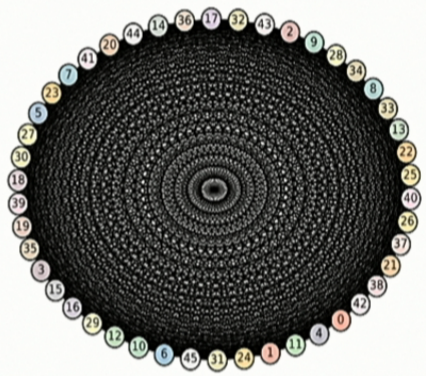
42 for pixels + 4 to one-hot encode the class
(only digits 1-4)



917 physical qubits

Quantum-assisted unsupervised learning on digits

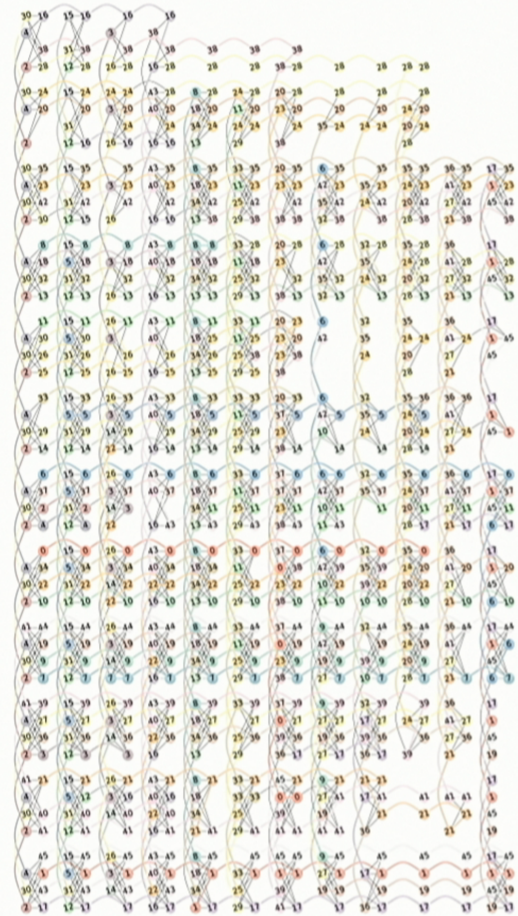
Overcoming the curse of limited connectivity in hardware.



46 fully-connected
logical (visible) variables

42 for pixels + 4 to one-hot encode the class
(only digits 1-4)

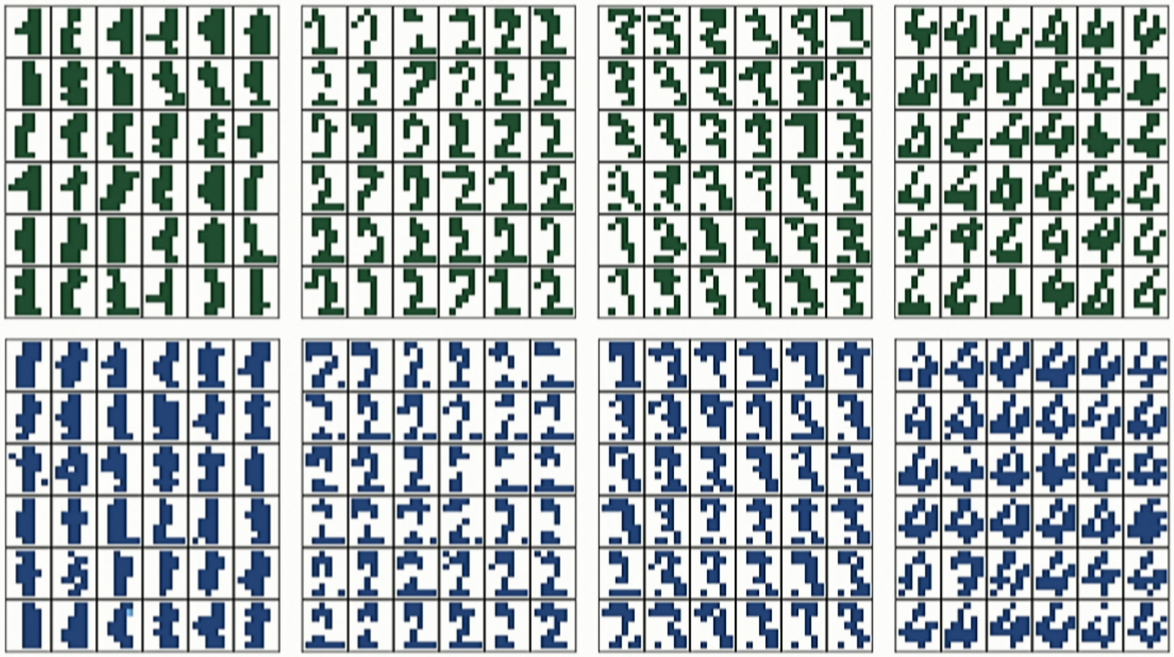
Are the results from this training on 917 qubit experiment meaningful? Is the model capable of generating digits, as expected?



917 physical qubits

Quantum-assisted unsupervised learning on digits

Human or (quantum) machine? (Turing test)

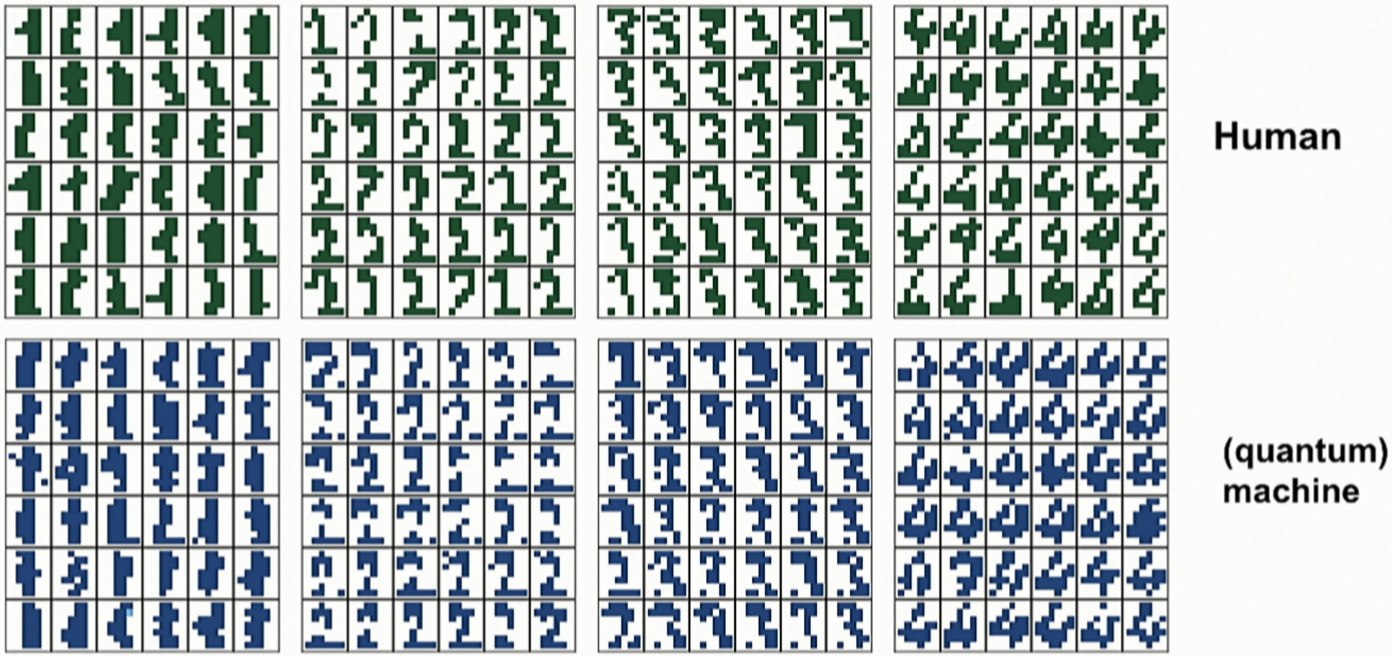


Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Benedetti et al. In preparation.

Quantum-assisted unsupervised learning on digits

Human or (quantum) machine? (Turing test)

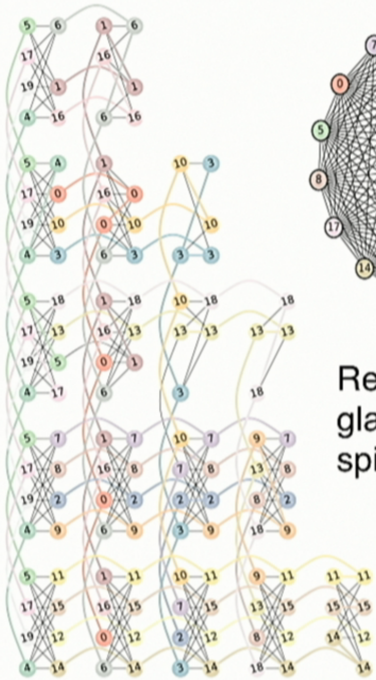


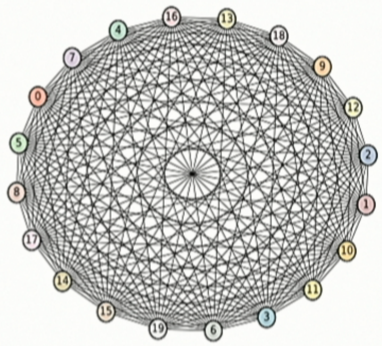
- **Experimental realization of quantum-assisted learning algorithm on 917 qubits, for a 46 fully-connected model.**

Dataset: Optical Recognition of Handwritten Digits (OptDigits)

Benedetti et al. In preparation.

Quantum-assisted unsupervised: artificial model

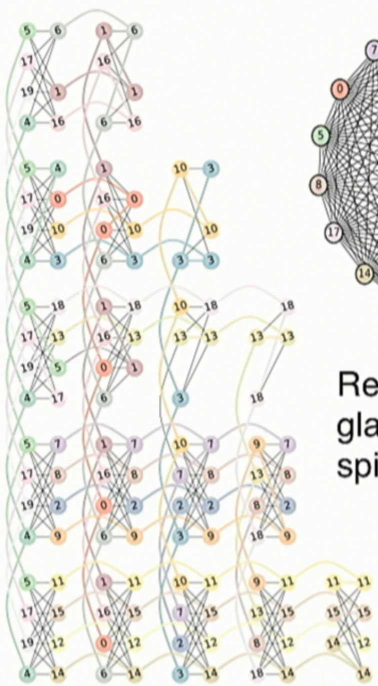


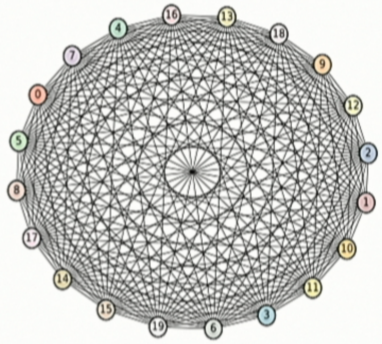


Reference model: Ising spin glass with 20 fully-connected spins (10 instances).

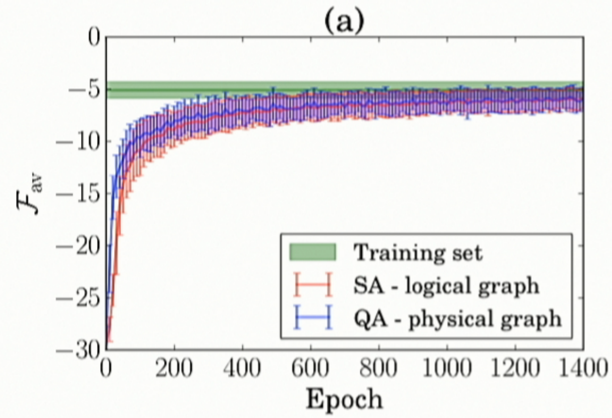
Benedetti et al. In preparation.

Quantum-assisted unsupervised: artificial model



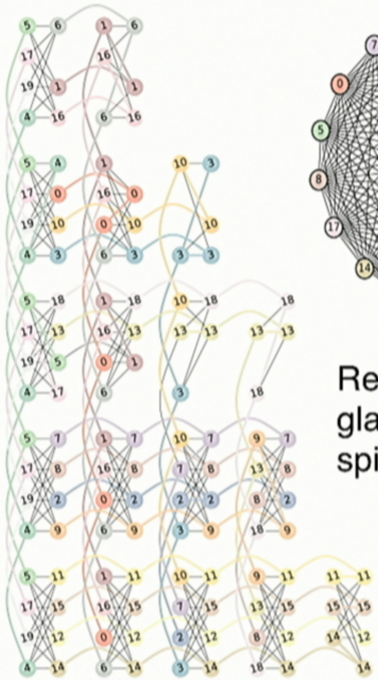


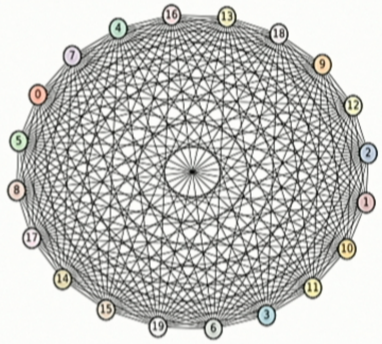
Reference model: Ising spin glass with 20 fully-connected spins (10 instances).



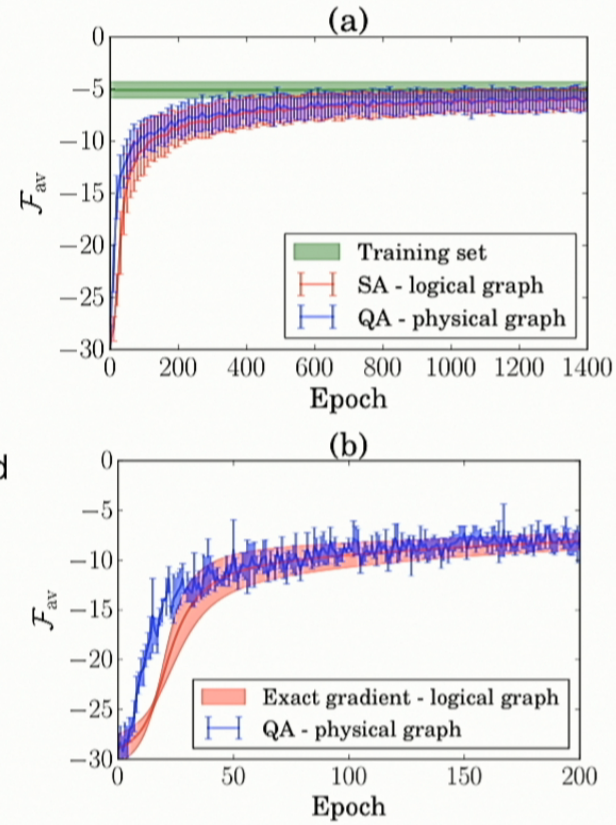
Benedetti et al. In preparation.

Quantum-assisted unsupervised: artificial model





Reference model: Ising spin glass with 20 fully-connected spins (10 instances).



Benedetti et al. In preparation.

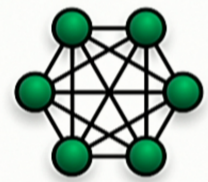
Ongoing research directions

Possible further boosting protocols by considering models to account explicitly for the noise in the quantum device.

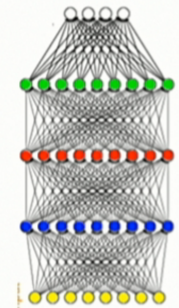
Numerical simulations show that main limitation of current quantum annealers for Boltzmann machines applications is its sparse connectivity.

Extensions to deep learning architectures.

How "Boltzmannian" need the samples to be for QuALE to work.



General BMs

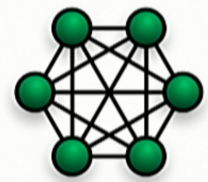


Deep architectures

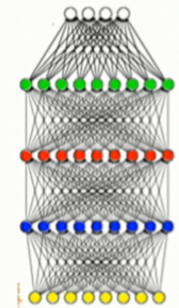
Ongoing research directions

Possible further boosting protocols by considering models to account explicitly for the noise in the quantum device.

Numerical simulations show that main limitation of current quantum annealers for Boltzmann machines applications is its sparse connectivity.



General BMs



Deep architectures

Extensions to deep learning architectures.

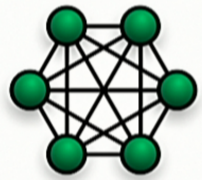
How "Boltzmannian" need the samples to be for QuALE to work.

Inference by using quantum distributions, such as those coming from future generation quantum computing technologies.

Ongoing research directions

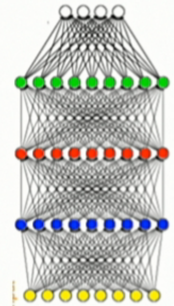
Possible further boosting protocols by considering models to account explicitly for the noise in the quantum device.

Numerical simulations show that main limitation of current quantum annealers for Boltzmann machines applications is its sparse connectivity.



General BMs

Extensions to deep learning architectures.



Deep architectures

How "Boltzmannian" need the samples to be for QuALE to work.

Inference by using quantum distributions, such as those coming from future generation quantum computing technologies.

Is quantum tunneling, or any other quantum computational resource, relevant for machine learning/sampling applications? Can it be any faster than MCMC? Is it possible to achieve quantum supremacy in this domain?