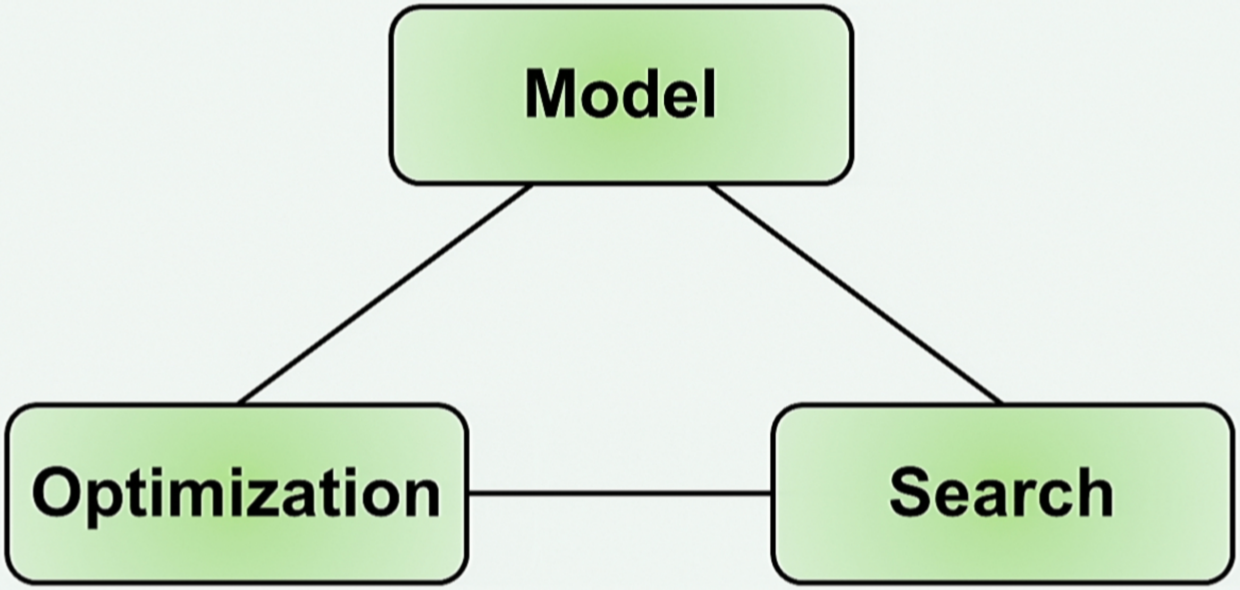
Title: Quantum Boltzmann Machine using a Quantum Annealer

Date: Aug 10, 2016 11:00 AM

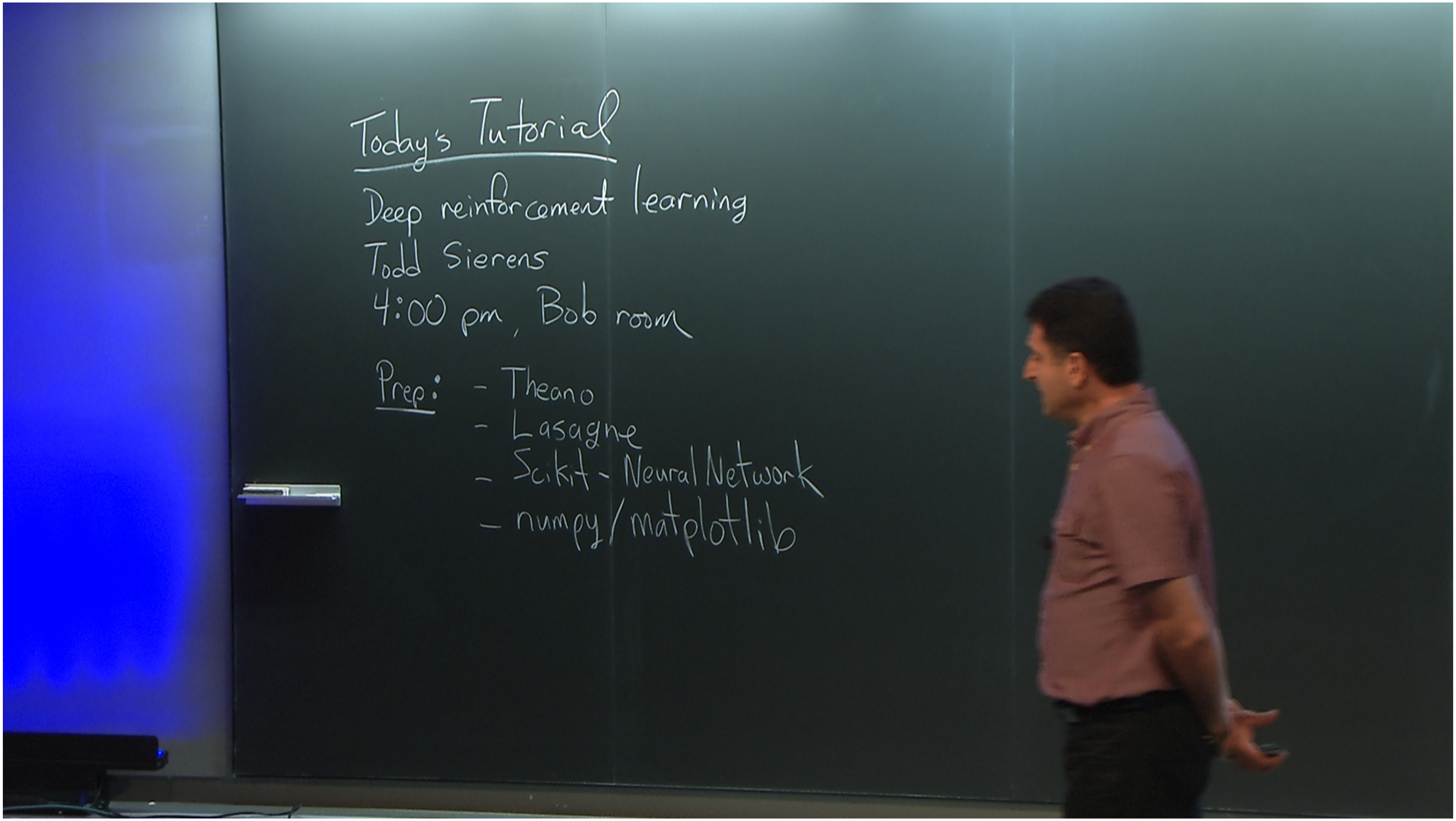
URL: http://pirsa.org/16080009

Abstract: Machine learning is a rapidly growing field in computer science with applications in computer vision, voice recognition, medical diagnosis, spam filtering, search engines, etc. In this presentation, I will introduce a new machine learning approach based on quantum Boltzmann distribution of a transverse-field Ising Model. Due to the non-commutative nature of quantum mechanics, the training process of the Quantum Boltzmann Machine (QBM) can become nontrivial.Â I will show how to circumvent this problem by introducing bounds on the quantum probabilities. This allows training the QBM efficiently by sampling. I will then show examples of QBM training with and without the bound, using exact diagonalization, and compare the results with classical Boltzmann training. Finally, after a brief introduction to D-Wave quantum annealing processors, I will discuss the possibility of using such processors for QBM training and application.

Machine Learning Techniques



Thanks to Ali Ghodsi (U of Waterloo)



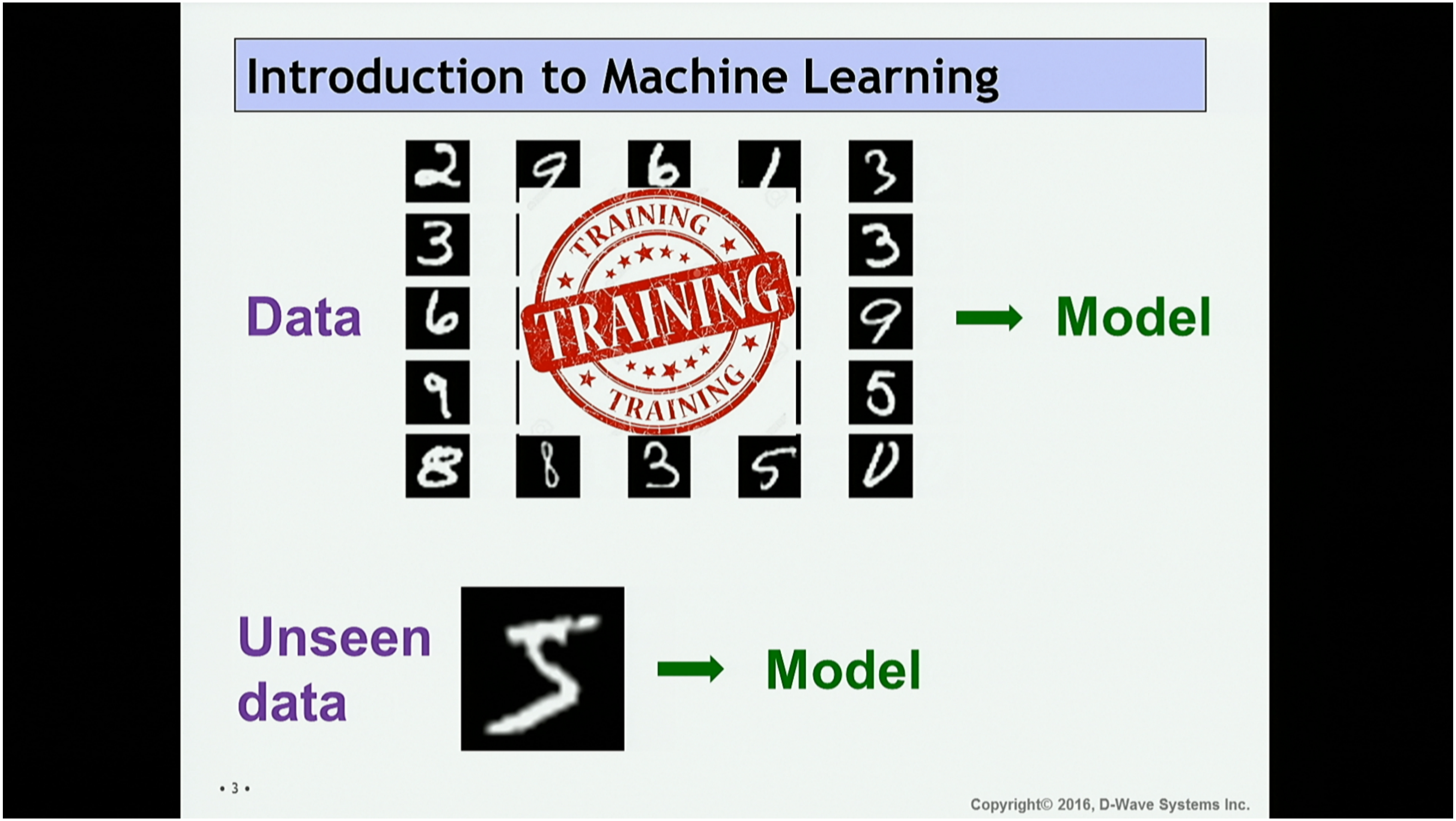
Today's Tutorial

Deep reinforcement learning

Todd Sierens

4:00 pm, Bob room

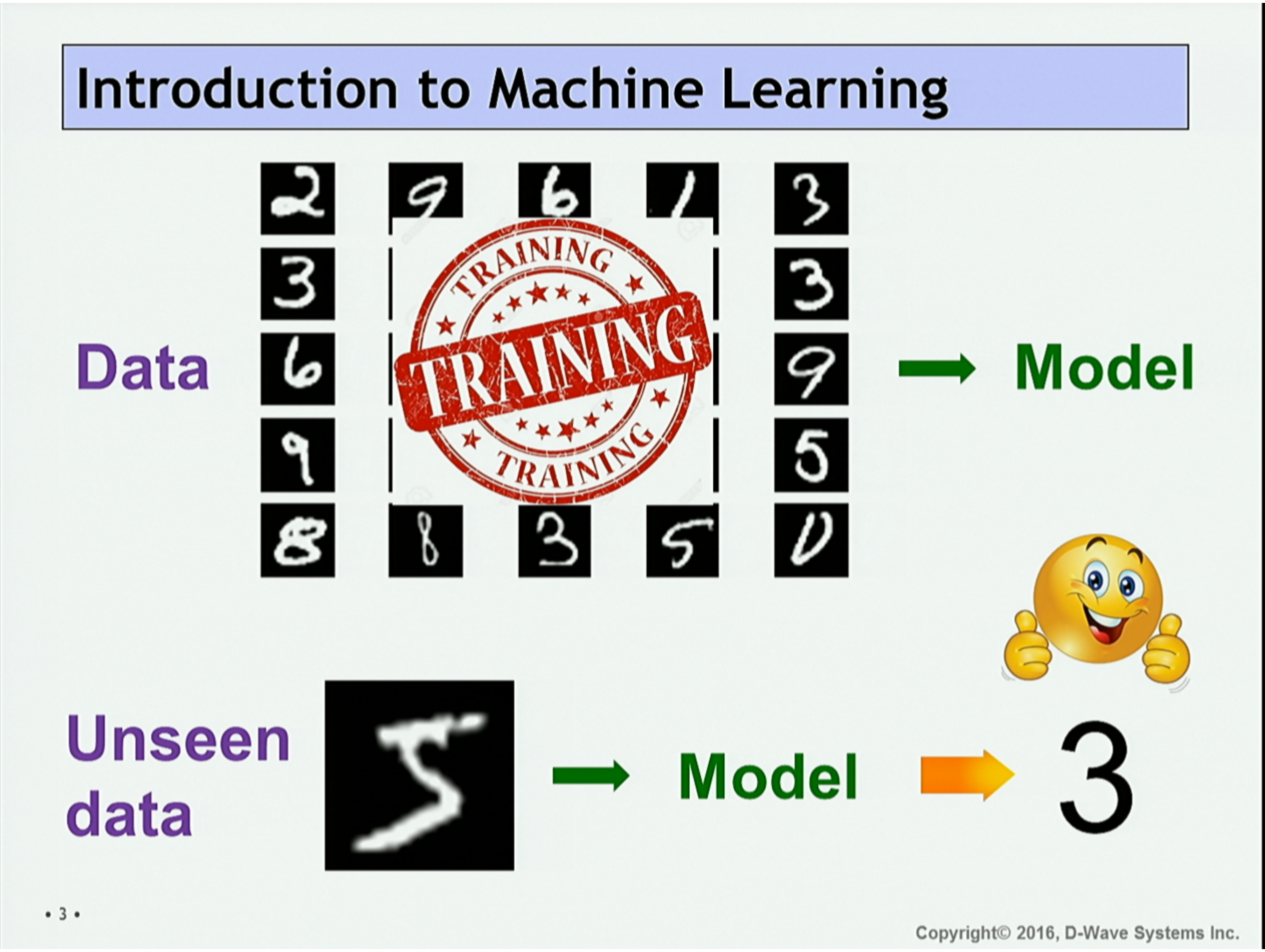
Prep:
- Theano
- Lasagne
- Scikit - Neural Network
- numpy/matplotlib



Introduction to Machine Learning

Data → **Model**

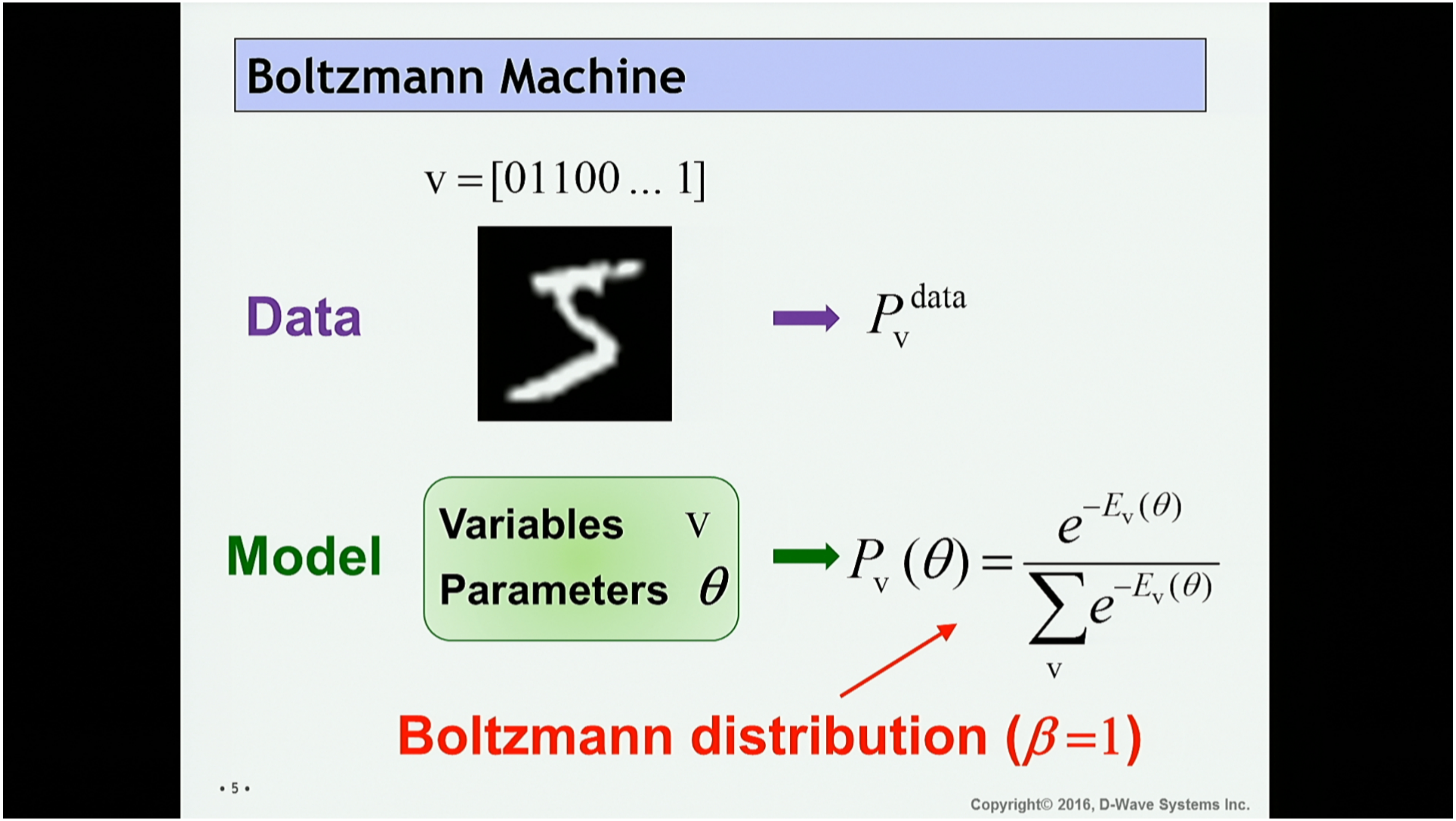
Unseen data → **Model**



Introduction to Machine Learning

Data → **Model**

Unseen data → **Model** → 3



Boltzmann Machine

$$v = [01100 \dots 1]$$

Data → P_{v}^{data}

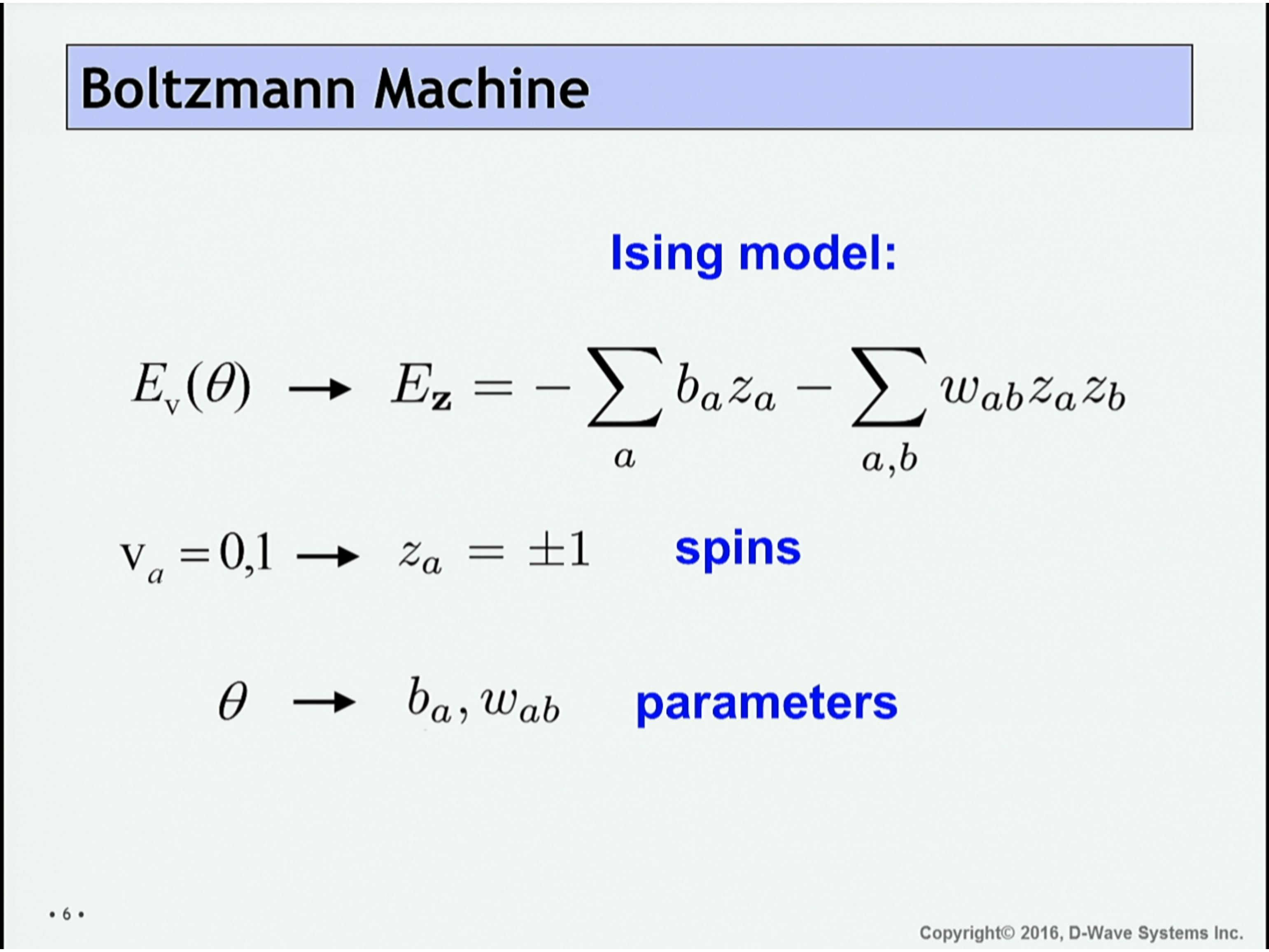
Model

Variables v

Parameters θ

$$\rightarrow P_{v}(\theta) = \frac{e^{-E_{v}(\theta)}}{\sum_{v} e^{-E_{v}(\theta)}}$$

Boltzmann distribution ($\beta=1$)



Boltzmann Machine

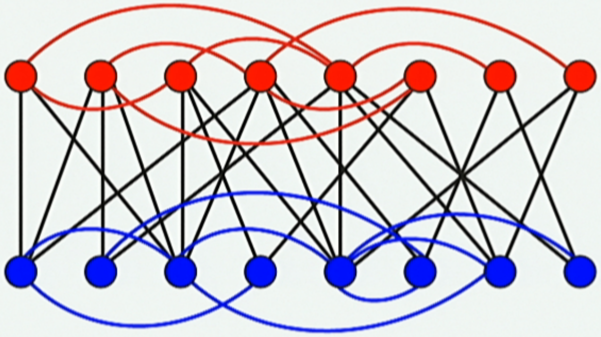
Ising model:

$$E_{\mathrm{v}}(\theta) \rightarrow E_{\mathbf{z}} = -\sum_{a} b_a z_a - \sum_{a,b} w_{ab} z_a z_b$$

$\mathrm{v}_a = 0{,}1 \rightarrow z_a = \pm 1$ **spins**

$\theta \rightarrow b_a, w_{ab}$ **parameters**

Adding Hidden Variables



hidden z_i

visible z_ν

$$E_{\mathbf{z}} = -\sum_a b_a z_a - \sum_{a,b} w_{ab} z_a z_b$$

$$z_a = (\underset{\text{visible}}{z_\nu}\,,\ \underset{\text{hidden}}{z_i})$$

$$\underset{\text{visible}}{P_{\mathbf{v}}} = Z^{-1} \sum_{\underset{\text{hidden}}{\mathbf{h}}} e^{-E_{\mathbf{z}}}, \qquad Z = \sum_{\mathbf{z}} e^{-E_{\mathbf{z}}}$$

Training a BM

Tune $\theta \in \{b_a, w_{ab}\}$ **such that** $P_{\mathrm{v}} \approx P_{\mathrm{v}}^{\mathrm{data}}$

Maximize log-likelihood: $\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log P_{\mathbf{v}}$

Or minimize: $\mathcal{L} = -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log P_{\mathbf{v}}$

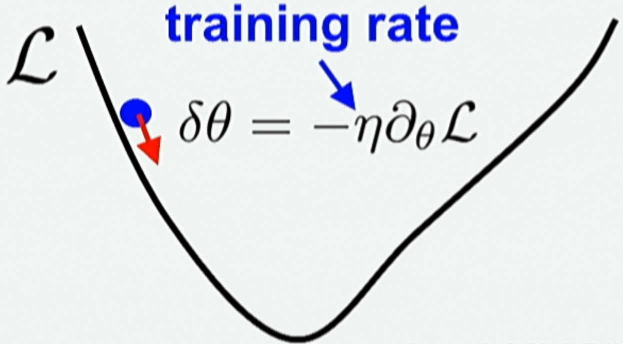
Training a BM

Tune $\theta \in \{b_a, w_{ab}\}$ **such that** $P_{\text{v}} \approx P_{\text{v}}^{\text{data}}$

Maximize log-likelihood: $\sum_{\mathbf{v}} P_{\mathbf{v}}^{\text{data}} \log P_{\mathbf{v}}$

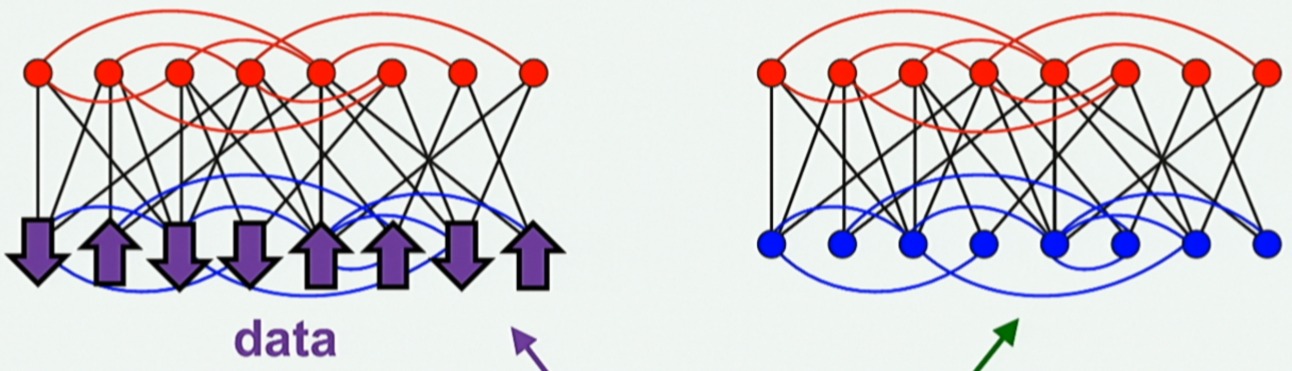
Or minimize: $\mathcal{L} = -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\text{data}} \log P_{\mathbf{v}}$

gradient descent technique



$$\delta\theta = -\eta \partial_\theta \mathcal{L}$$

Training Ising Hamiltonian Parameters



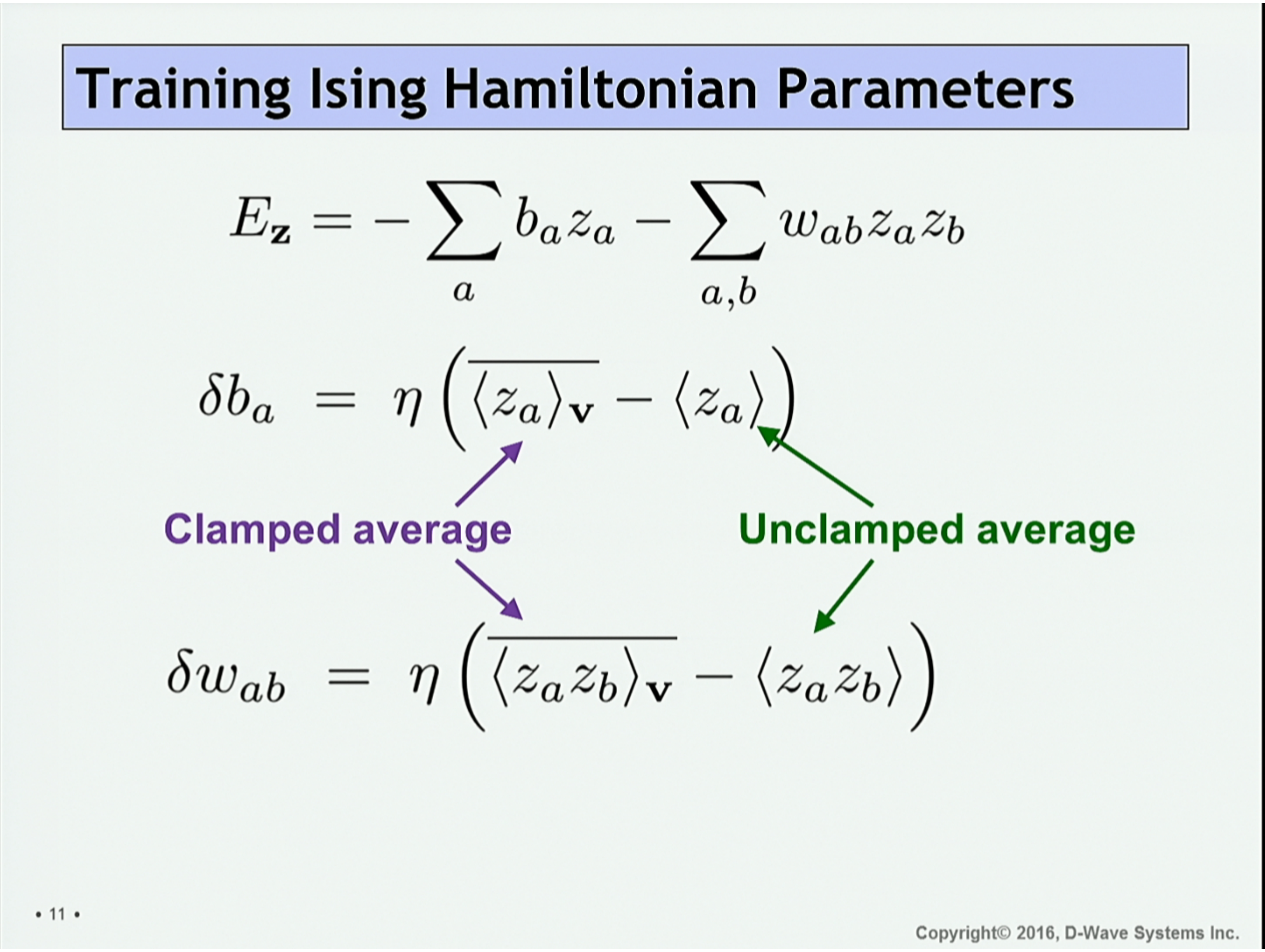
clamped

unclamped

$$\partial_\theta \mathcal{L} = \overline{\langle \partial_\theta E_{\mathbf{z}} \rangle_{\mathbf{v}}} - \langle \partial_\theta E_{\mathbf{z}} \rangle$$

Average with clamped visibles

Unclamped average



Training Ising Hamiltonian Parameters

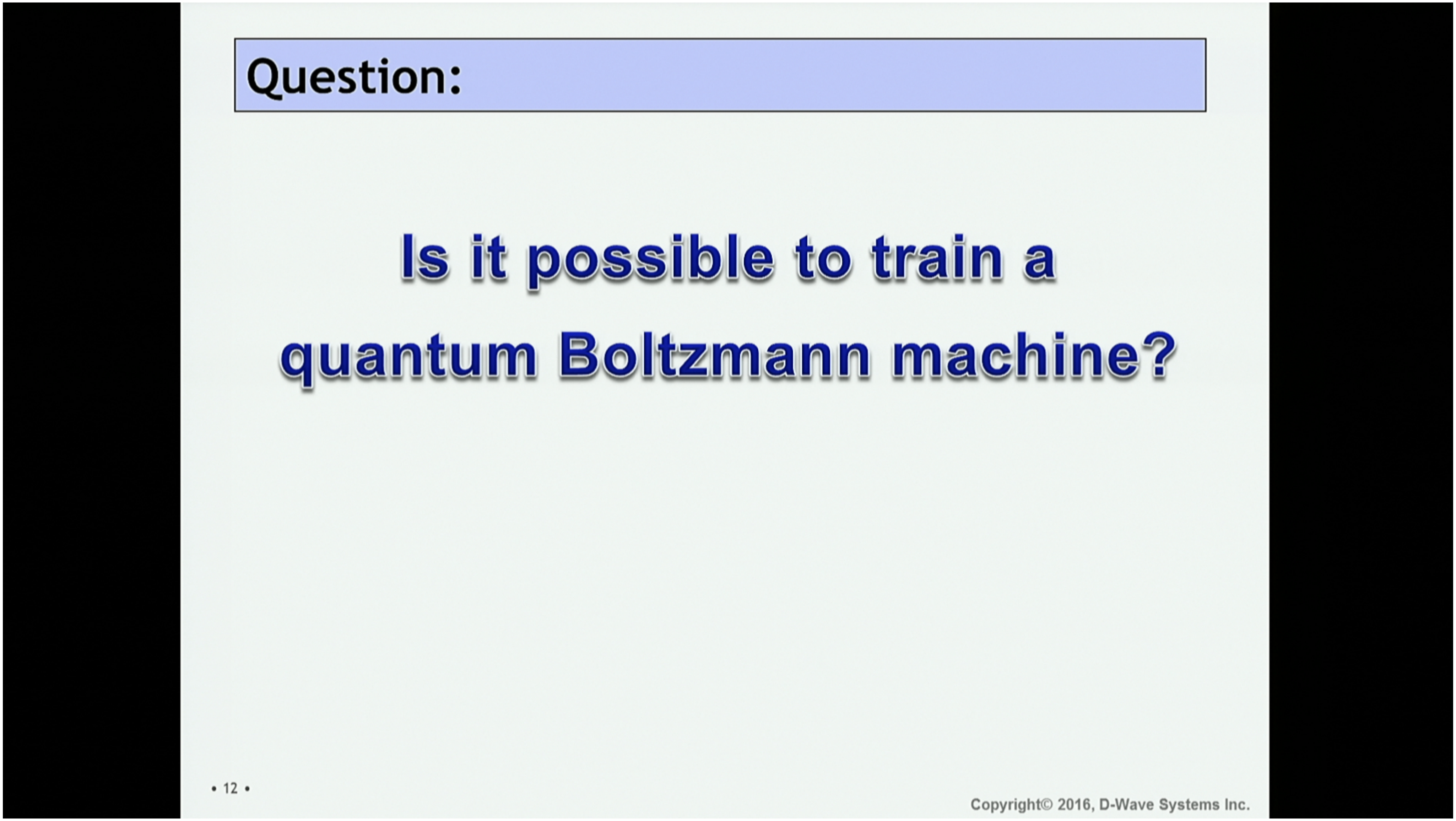
$$E_{\mathbf{z}} = -\sum_{a} b_a z_a - \sum_{a,b} w_{ab} z_a z_b$$

$$\delta b_a \;=\; \eta \left(\overline{\langle z_a \rangle_{\mathbf{v}}} - \langle z_a \rangle \right)$$

Clamped average

Unclamped average

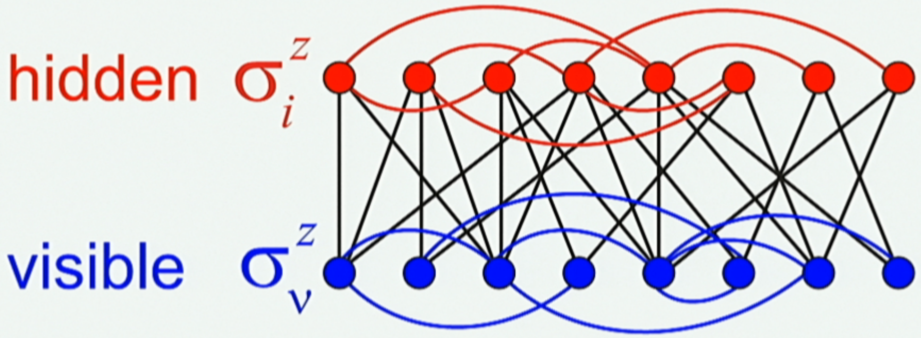
$$\delta w_{ab} \;=\; \eta \left(\overline{\langle z_a z_b \rangle_{\mathbf{v}}} - \langle z_a z_b \rangle \right)$$



Question:

Is it possible to train a quantum Boltzmann machine?

Transverse Ising Hamiltonian

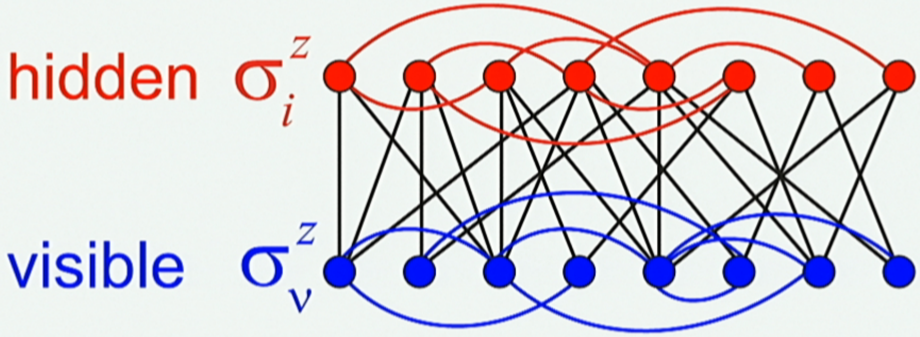


$$H = -\sum_a \Gamma_a \sigma_a^x - \sum_a b_a \sigma_a^z - \sum_{a,b} w_{ab} \sigma_a^z \sigma_b^z$$

$$\sigma_a^z \equiv \overbrace{I \otimes \ldots \otimes I}^{a-1} \otimes \sigma_z \otimes \overbrace{I \otimes \ldots \otimes I}^{N-a}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_a^x \equiv \overbrace{I \otimes \ldots \otimes I}^{a-1} \otimes \sigma_x \otimes \overbrace{I \otimes \ldots \otimes I}^{N-a}, \qquad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Quantum Boltzmann Distribution

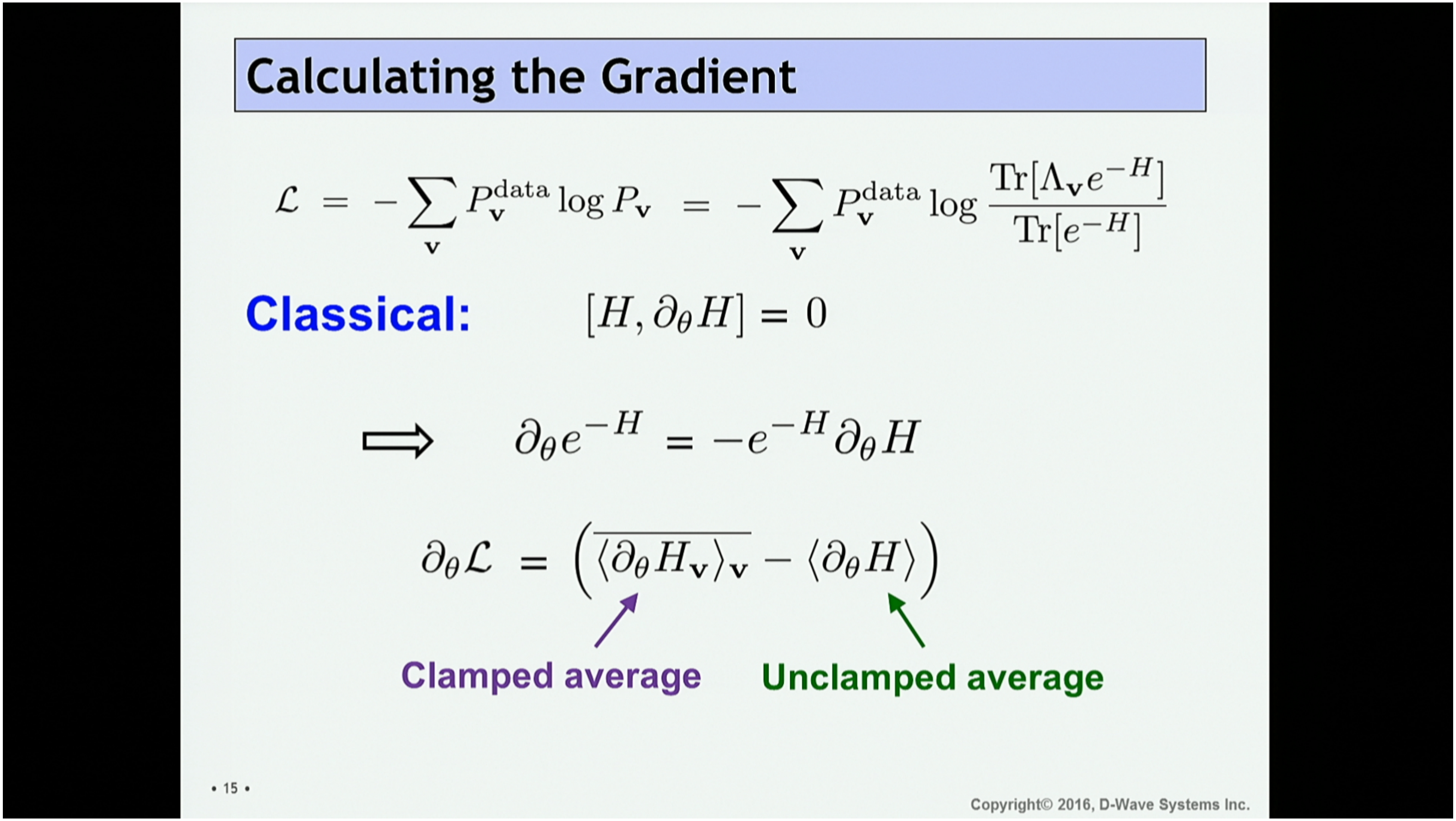


Marginal distribution: $P_{\mathbf{v}} = \mathrm{Tr}[\Lambda_{\mathbf{v}} \rho]$

Density matrix: $\rho = e^{-H} / \mathrm{Tr}[e^{-H}]$

$$\Lambda_{\mathbf{v}} = |\mathbf{v}\rangle \langle \mathbf{v}| \otimes \mathcal{I}_{\mathbf{h}}$$

Projection operator **Identity matrix**



Calculating the Gradient

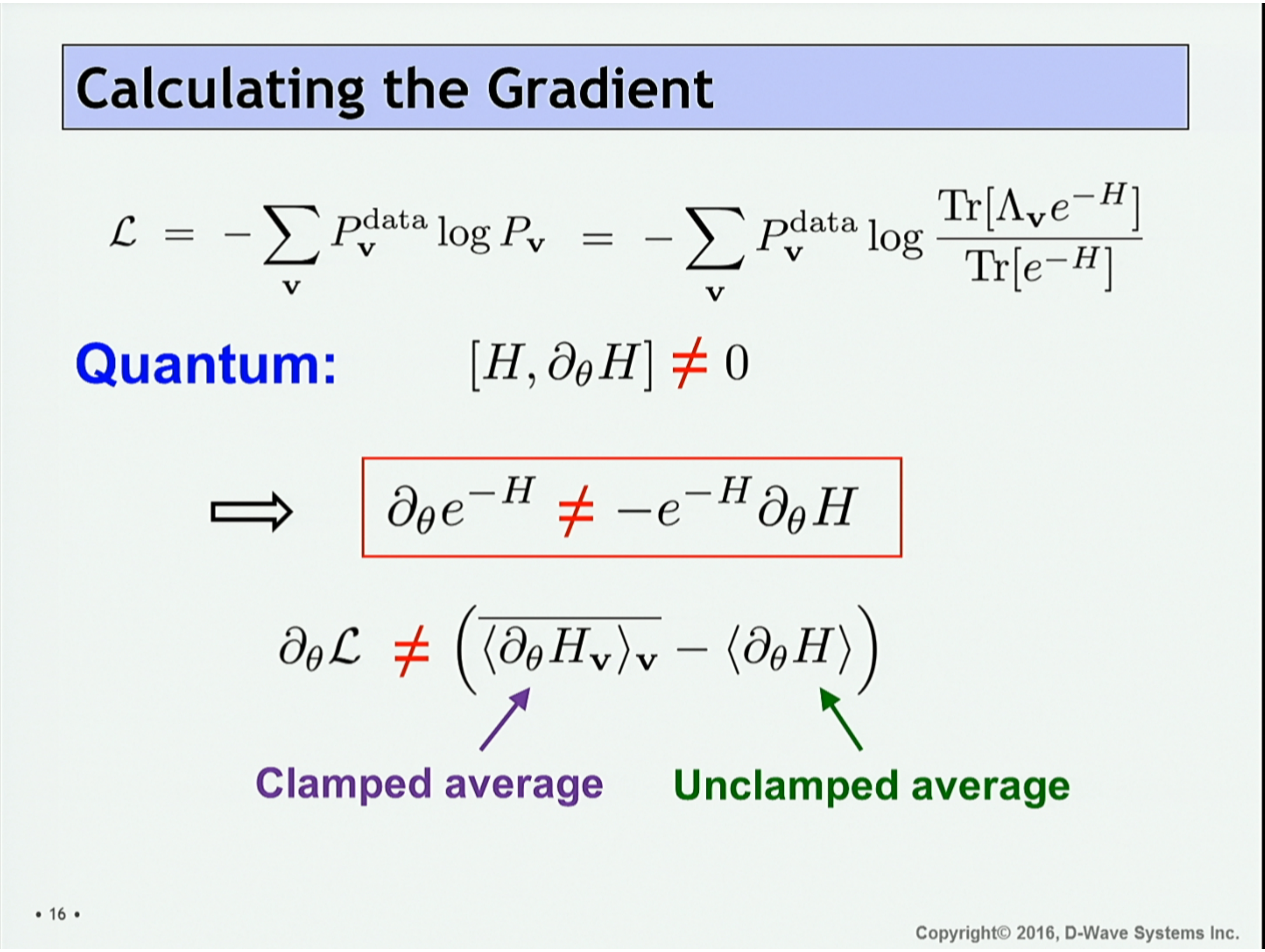
$$\mathcal{L} = -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log P_{\mathbf{v}} = -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log \frac{\mathrm{Tr}[\Lambda_{\mathbf{v}} e^{-H}]}{\mathrm{Tr}[e^{-H}]}$$

Classical: $[H, \partial_\theta H] = 0$

$$\Longrightarrow \quad \partial_\theta e^{-H} = -e^{-H} \partial_\theta H$$

$$\partial_\theta \mathcal{L} = \left(\overline{\langle \partial_\theta H_{\mathbf{v}} \rangle_{\mathbf{v}}} - \langle \partial_\theta H \rangle \right)$$

Clamped average **Unclamped average**



Calculating the Gradient

$$\mathcal{L} = -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log P_{\mathbf{v}} = -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log \frac{\mathrm{Tr}[\Lambda_{\mathbf{v}} e^{-H}]}{\mathrm{Tr}[e^{-H}]}$$

Quantum: $[H, \partial_\theta H] \neq 0$

$$\Longrightarrow \quad \boxed{\partial_\theta e^{-H} \neq -e^{-H} \partial_\theta H}$$

$$\partial_\theta \mathcal{L} \neq \left(\overline{\langle \partial_\theta H_{\mathbf{v}} \rangle_{\mathbf{v}}} - \langle \partial_\theta H \rangle \right)$$

Clamped average **Unclamped average**

Two Useful Properties of Trace

$$\partial_\theta e^{-H} \neq -e^{-H}\partial_\theta H$$

Two Useful Properties of Trace

$$\mathrm{Tr}[\partial_\theta e^{-H}] = -\mathrm{Tr}[e^{-H}\partial_\theta H]$$

Two Useful Properties of Trace

$$\mathrm{Tr}[\partial_\theta e^{-H}] = -\mathrm{Tr}[e^{-H}\partial_\theta H]$$

Golden-Thompson inequality:

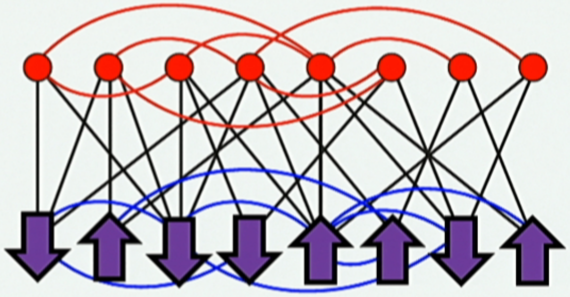
For Hermitian matrices A and B

$$\mathrm{Tr}[e^{A}e^{B}] \geq \mathrm{Tr}[e^{A+B}]$$

Finding lower bounds

$$P_{\mathbf{v}} = \frac{\mathrm{Tr}[\Lambda_{\mathbf{v}} e^{-H}]}{\mathrm{Tr}[e^{-H}]} = \frac{\mathrm{Tr}[e^{-H} e^{\ln \Lambda_{\mathbf{v}}}]}{\mathrm{Tr}[e^{-H}]} \geq \frac{\mathrm{Tr}[e^{-H+\ln \Lambda_{\mathbf{v}}}]}{\mathrm{Tr}[e^{-H}]}$$

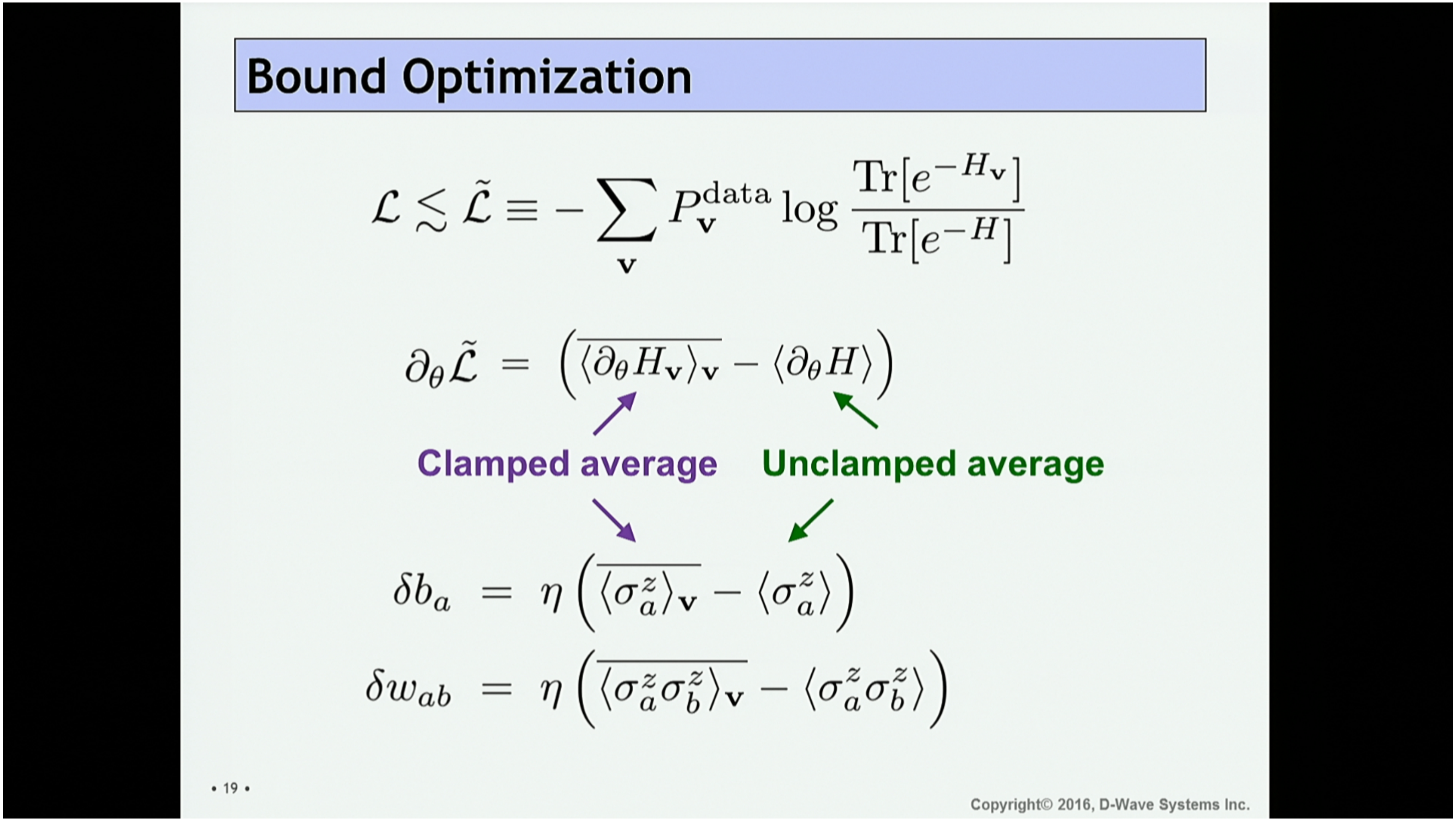
$$\boxed{P_{\mathbf{v}} \geq \frac{\mathrm{Tr}[e^{-H_{\mathbf{v}}}]}{\mathrm{Tr}[e^{-H}]} \qquad H_{\mathbf{v}} = H - \ln \Lambda_{\mathbf{v}}}$$



←

Clamped Hamiltonian

Visibles **are clamped to** **data**



Bound Optimization

$$\mathcal{L} \lesssim \tilde{\mathcal{L}} \equiv -\sum_{\mathbf{v}} P_{\mathbf{v}}^{\mathrm{data}} \log \frac{\mathrm{Tr}[e^{-H_{\mathbf{v}}}]}{\mathrm{Tr}[e^{-H}]}$$

$$\partial_\theta \tilde{\mathcal{L}} \;=\; \left(\overline{\langle \partial_\theta H_{\mathbf{v}} \rangle_{\mathbf{v}}} - \langle \partial_\theta H \rangle\right)$$

Clamped average **Unclamped average**

$$\delta b_a \;=\; \eta \left(\overline{\langle \sigma_a^z \rangle_{\mathbf{v}}} - \langle \sigma_a^z \rangle\right)$$

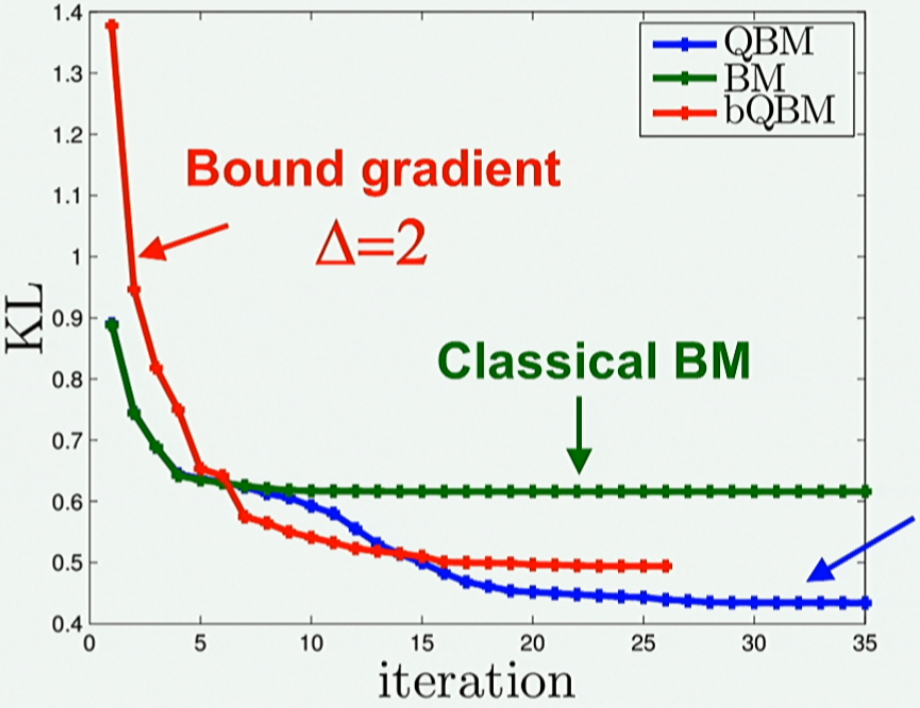
$$\delta w_{ab} \;=\; \eta \left(\overline{\langle \sigma_a^z \sigma_b^z \rangle_{\mathbf{v}}} - \langle \sigma_a^z \sigma_b^z \rangle\right)$$

Exact Diagonalization Results

Amin, Andriyash, Rolfe, Kulchytskyy, *Melko, arXiv:1601.02036*

Exact Diagonalization Results

Amin, Andriyash, Rolfe,* Kulchytskyy, *Melko, arXiv:1601.02036



Bound gradient

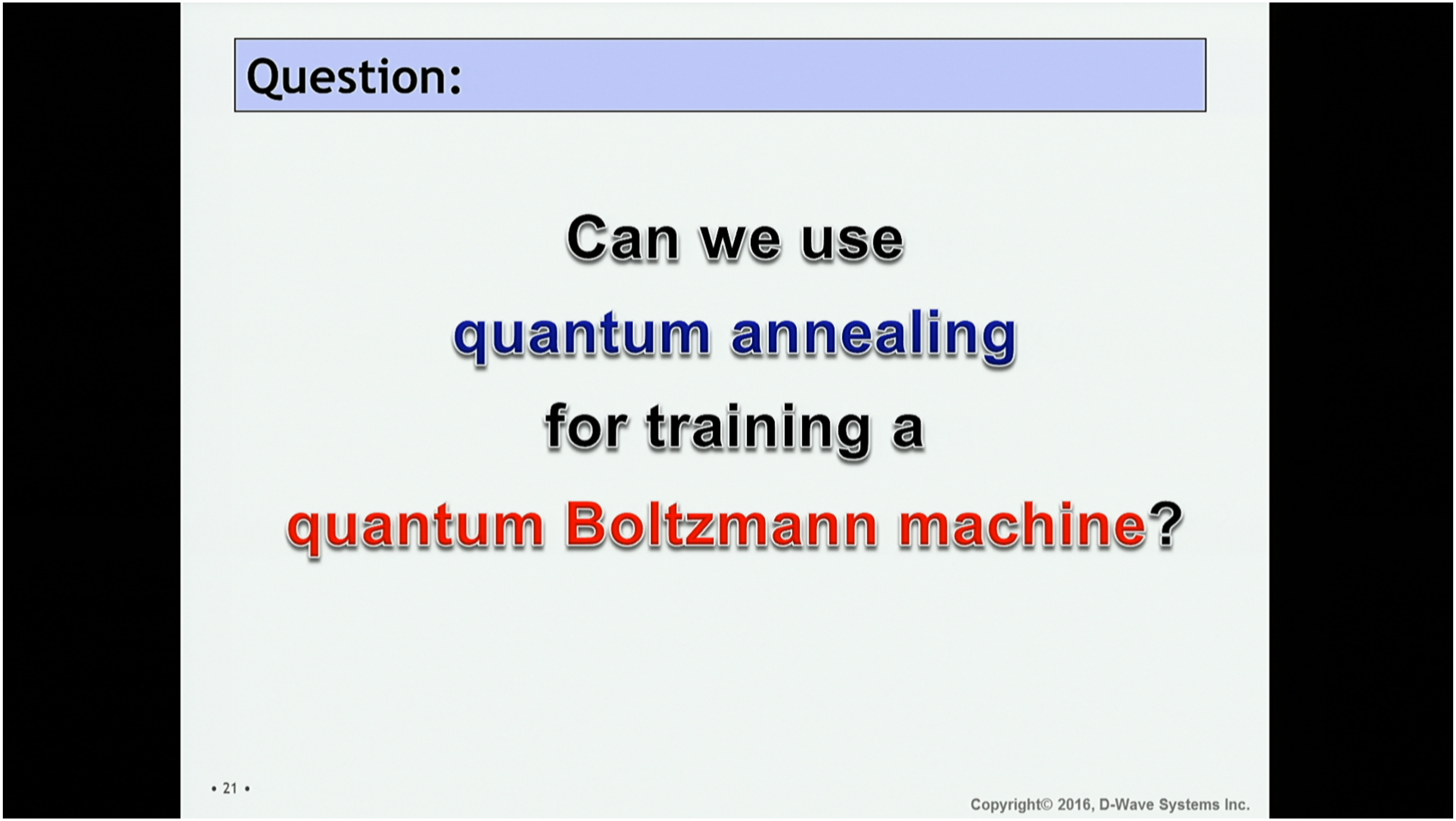
Δ=2

Classical BM

Exact gradient

(Δ is trained)

$\Delta_{\text{final}} = 2.5$



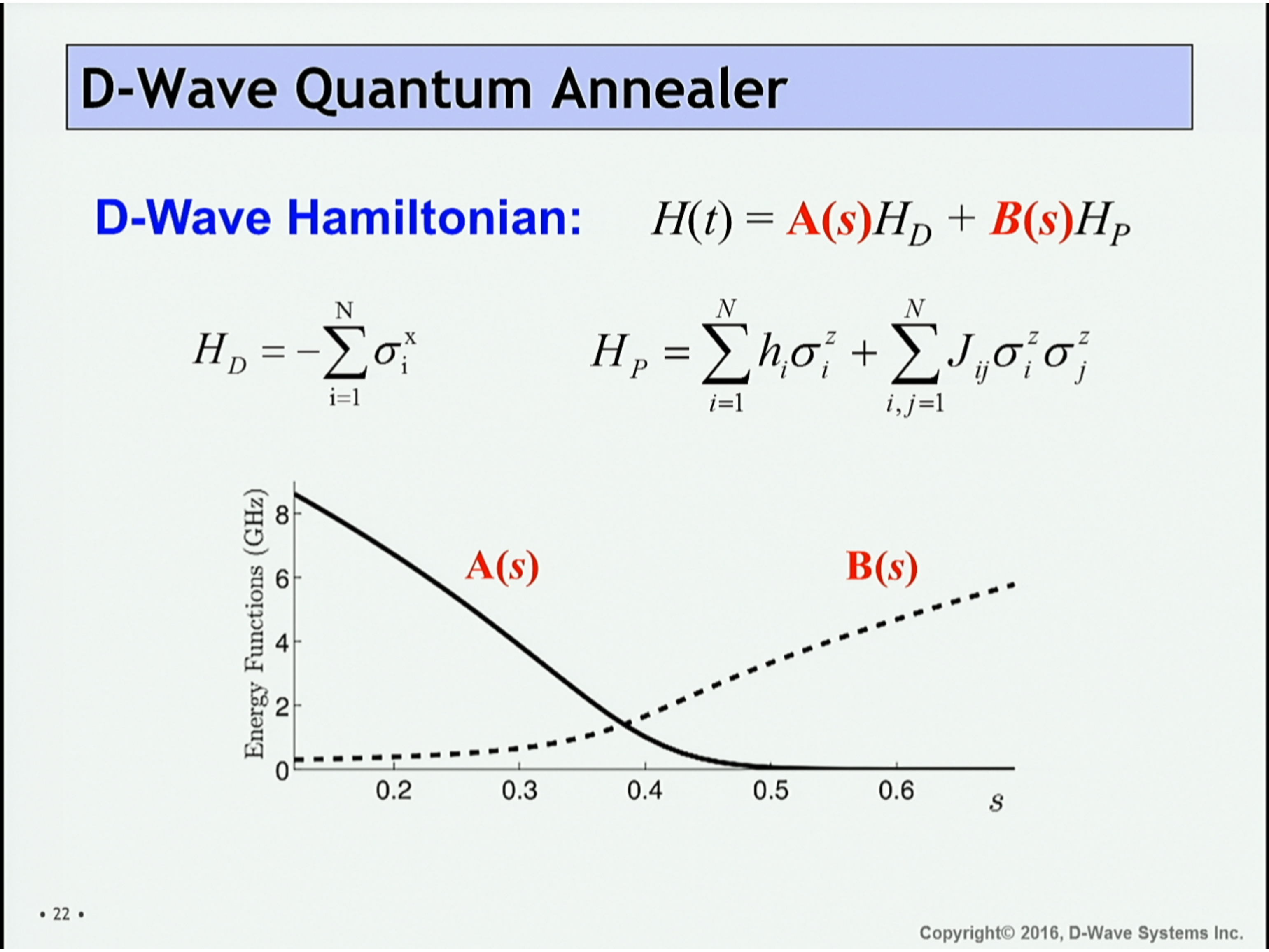
Question:

Can we use

quantum annealing

for training a

quantum Boltzmann machine?



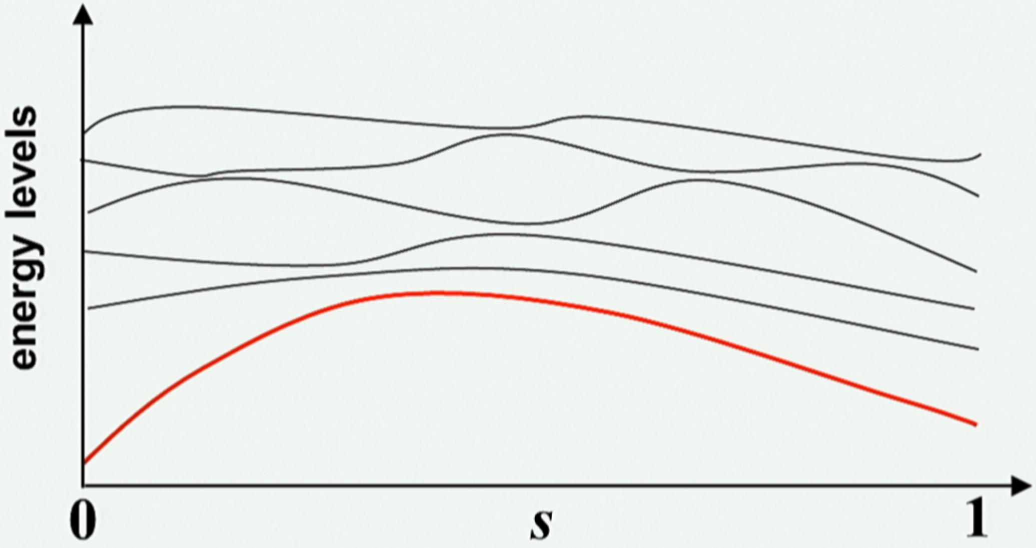
D-Wave Quantum Annealer

D-Wave Hamiltonian: $H(t) = \mathbf{A}(s)H_D + \mathbf{B}(s)H_P$

$$H_D = -\sum_{i=1}^{N} \sigma_i^x \qquad H_P = \sum_{i=1}^{N} h_i \sigma_i^z + \sum_{i,j=1}^{N} J_{ij} \sigma_i^z \sigma_j^z$$

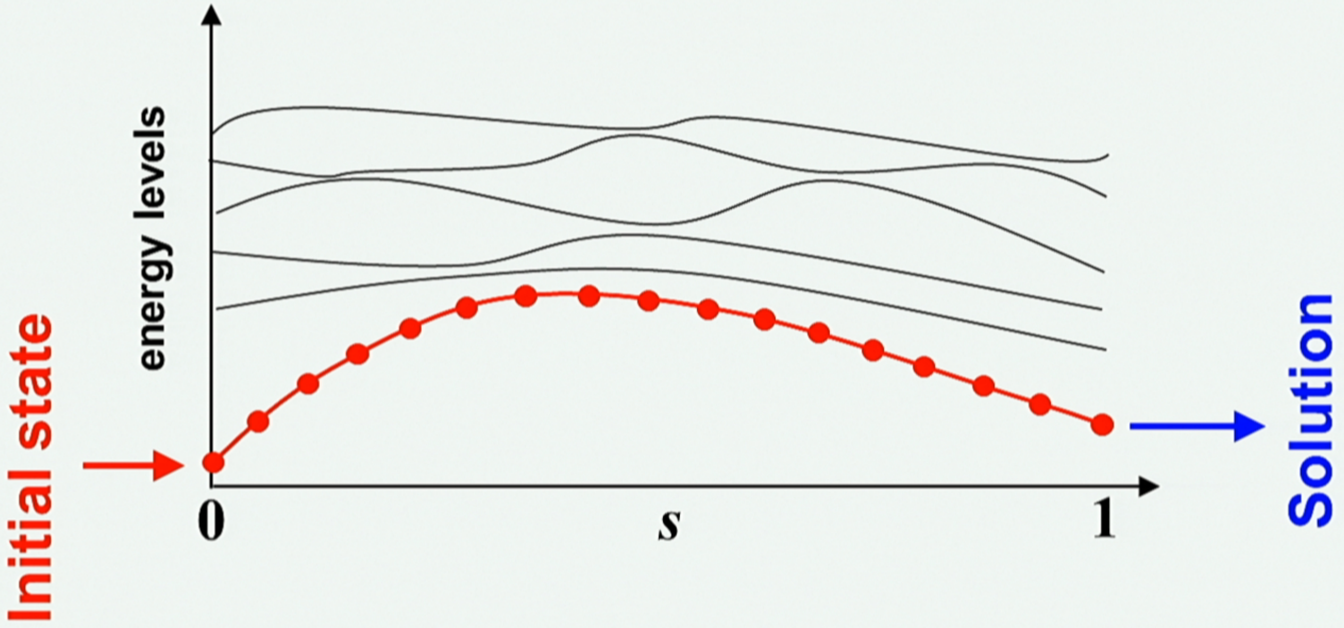
Adiabatic Quantum Computation

$$H(t) = (1-s)H_D + sH_P\,, \quad s = t/t_f$$



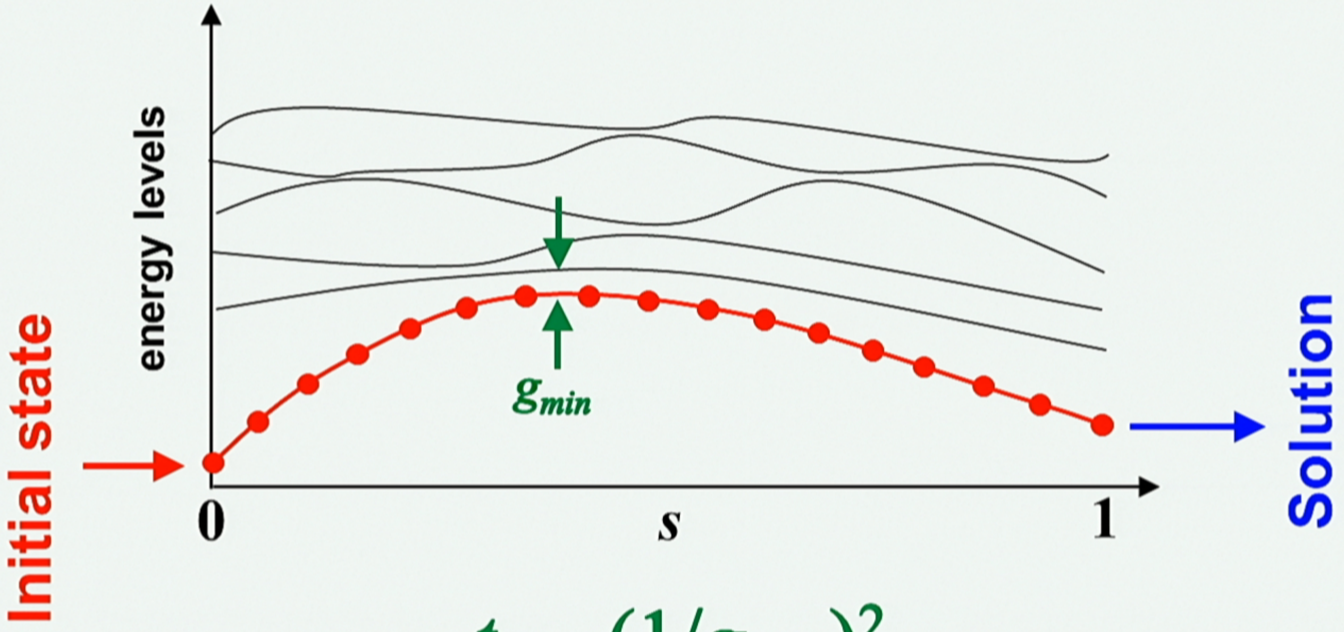
Adiabatic Quantum Computation

$$H(t) = (1-s)H_D + sH_P, \quad s = t/t_f$$



Adiabatic Quantum Computation

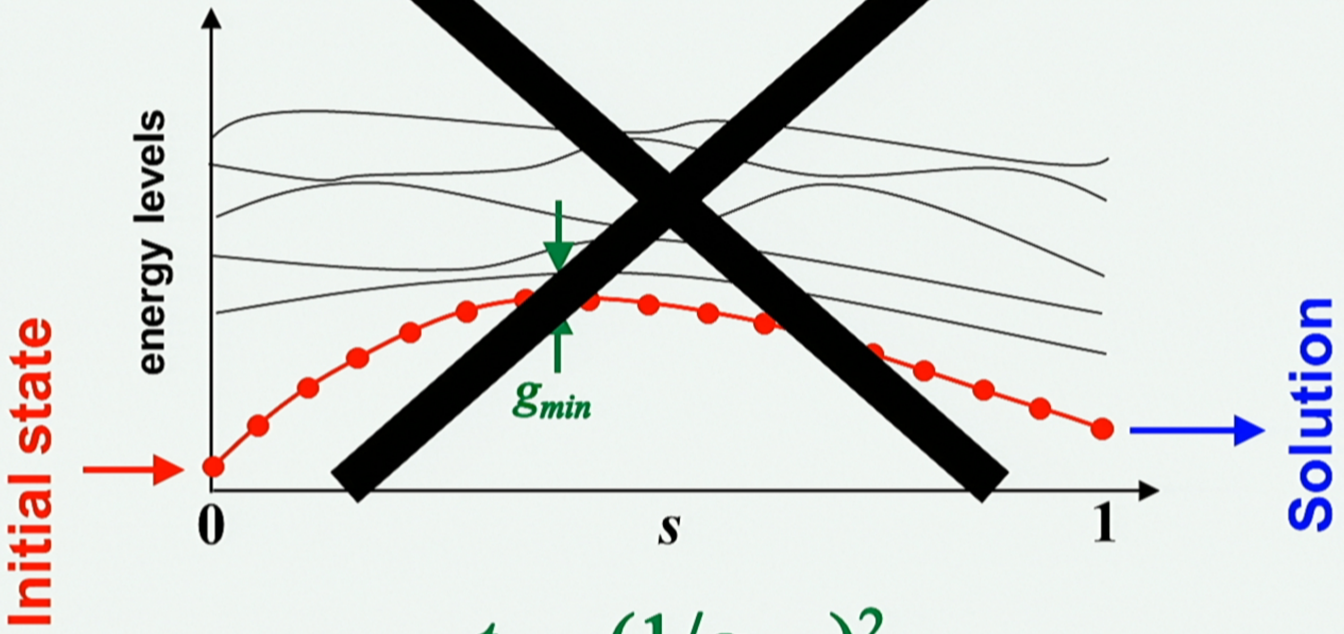
$$H(t) = (1-s)H_D + sH_P, \quad s = t/t_f$$



$$t_f \sim (1/g_{min})^2$$

Adiabatic Quantum Computation

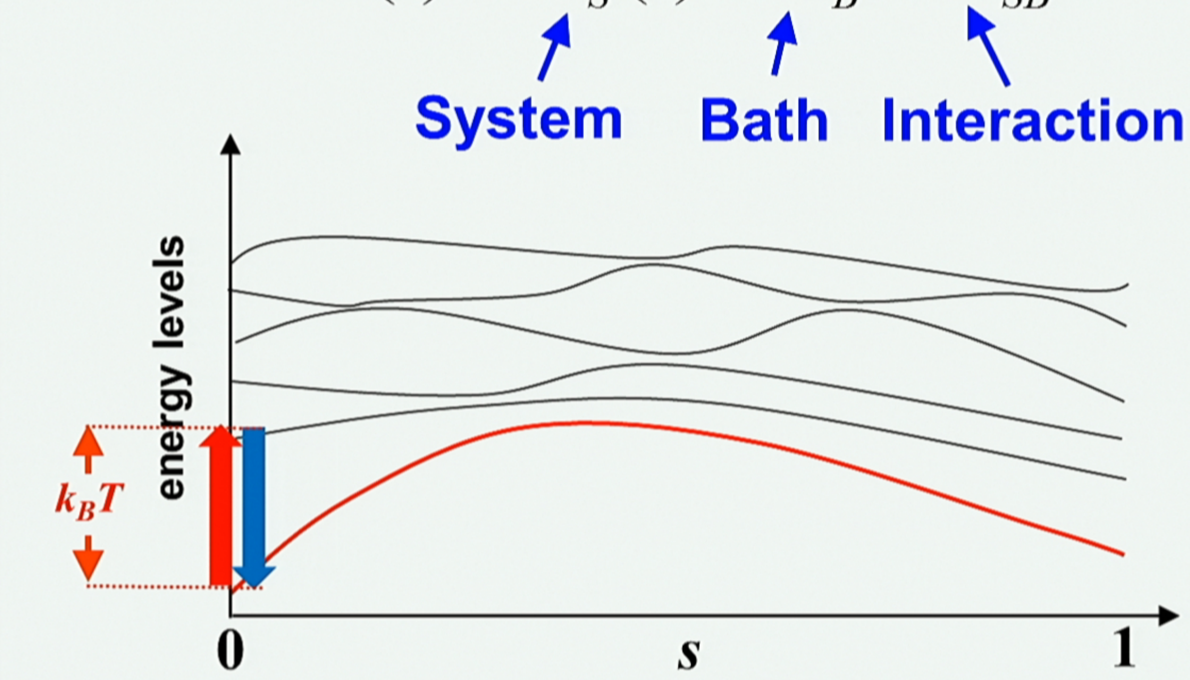
$$H(t) = (1-s)H_D + sH_P\,, \quad s = t/t_f$$



$$t_f \sim (1/g_{min})^2$$

Thermal Noise

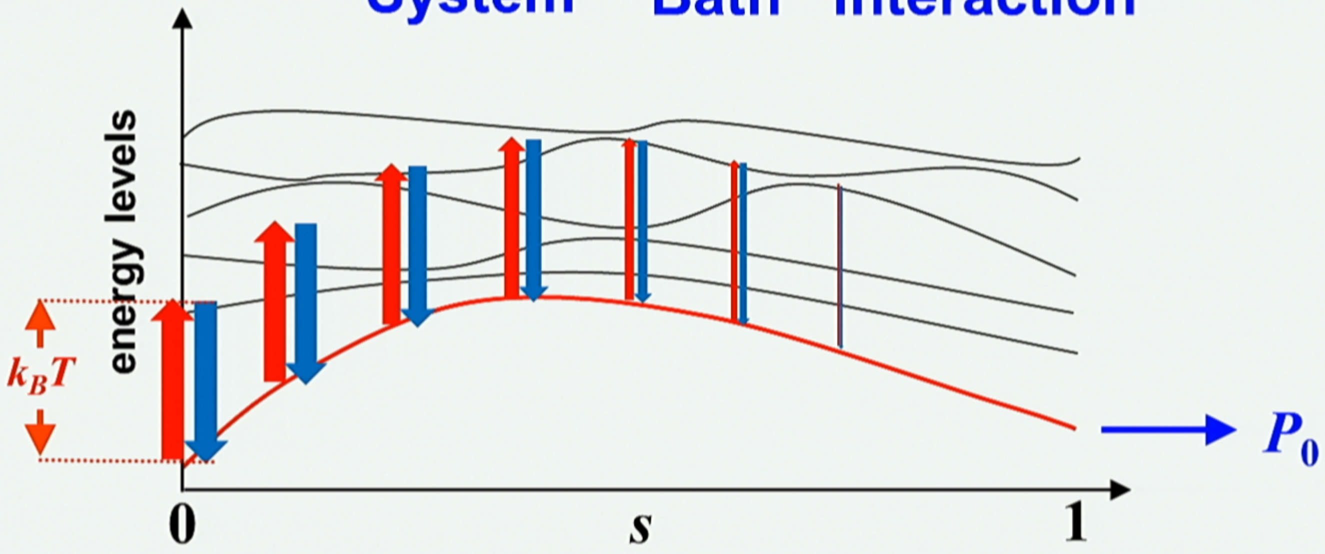
$$H(t) = H_S(t) + H_B + H_{SB}$$



Thermal Noise

$$H(t) = H_S(t) + H_B + H_{SB}$$

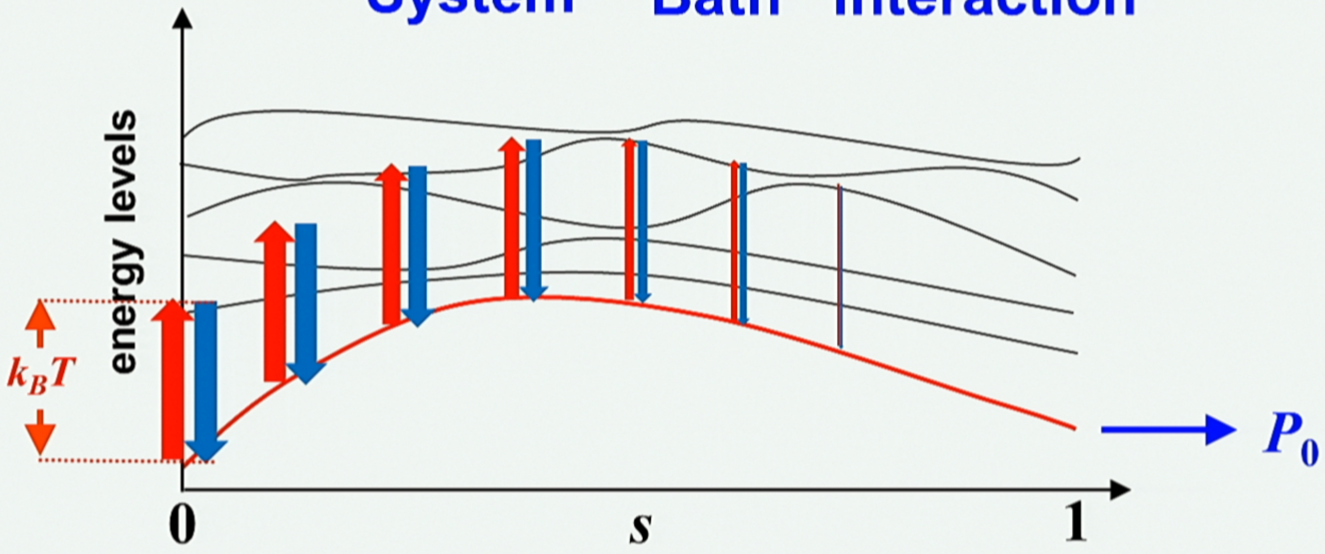
System **Bath** **Interaction**



Thermal Noise

$$H(t) = H_S(t) + H_B + H_{SB}$$

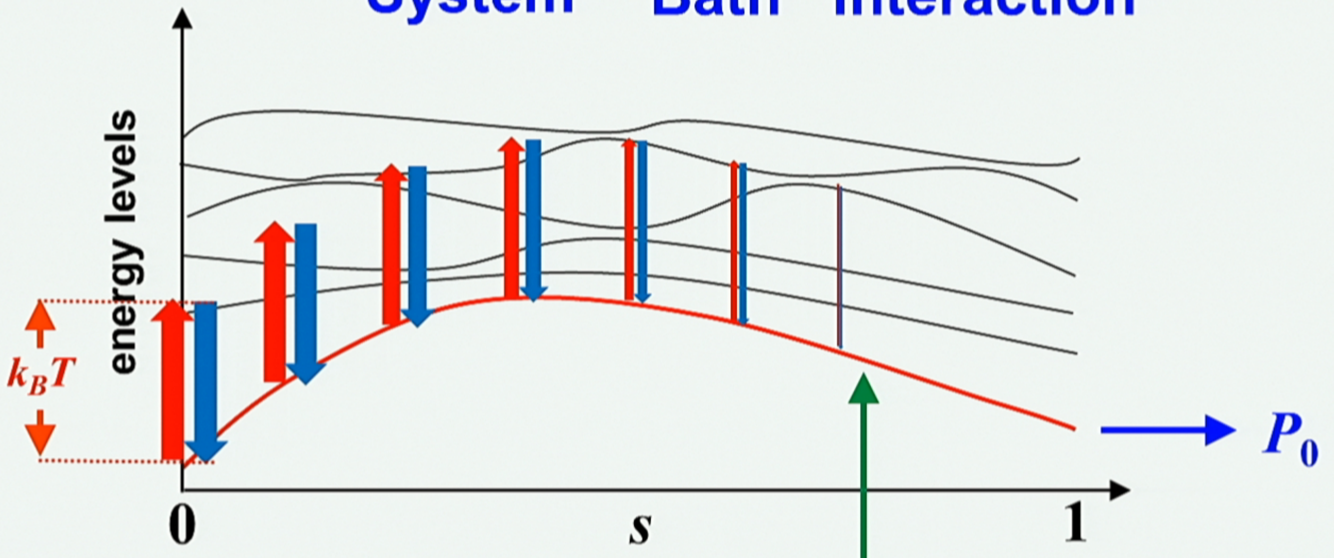
System **Bath** **Interaction**



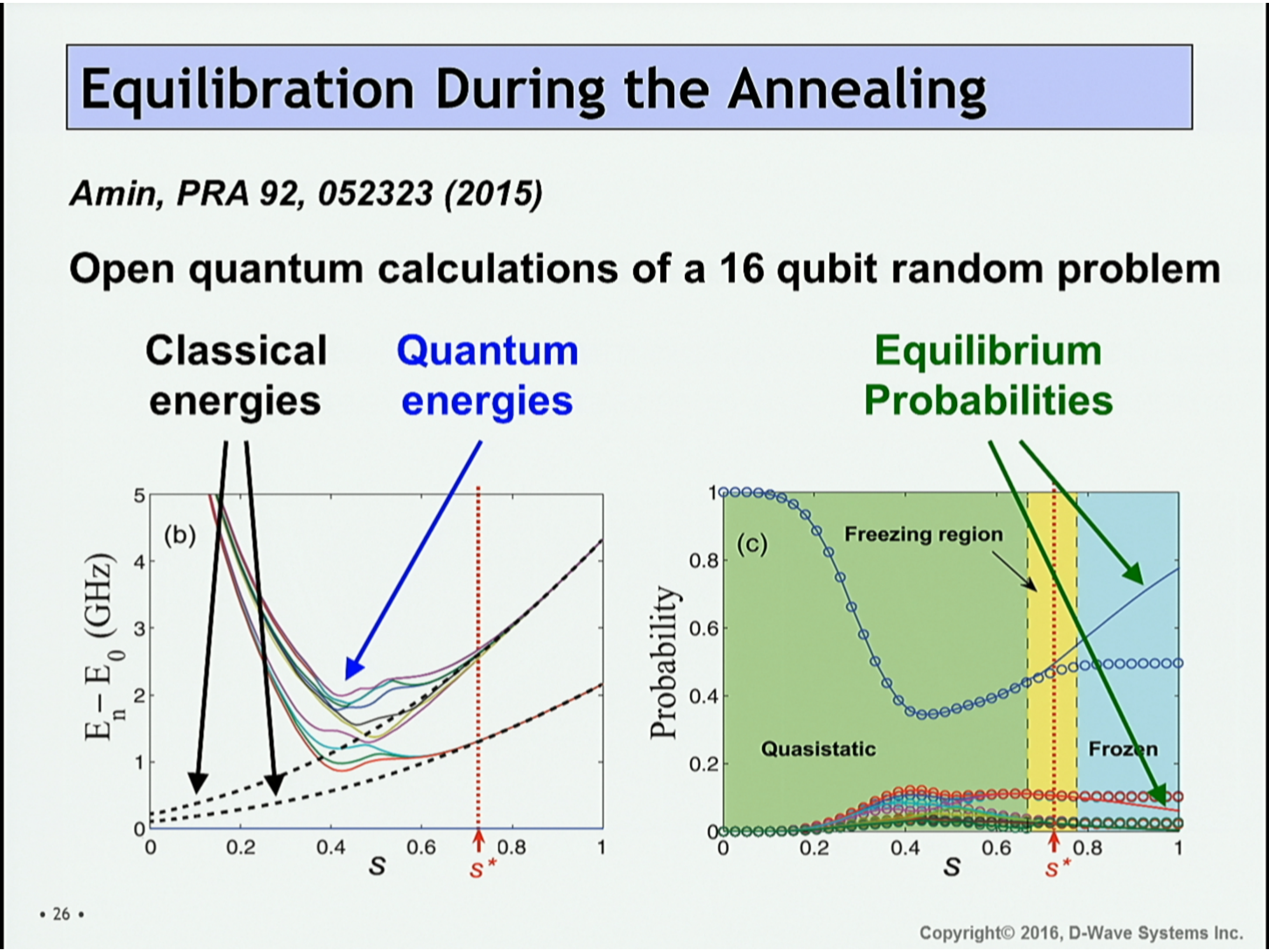
Thermal Noise

$$H(t) = H_S(t) + H_B + H_{SB}$$

System **Bath** **Interaction**



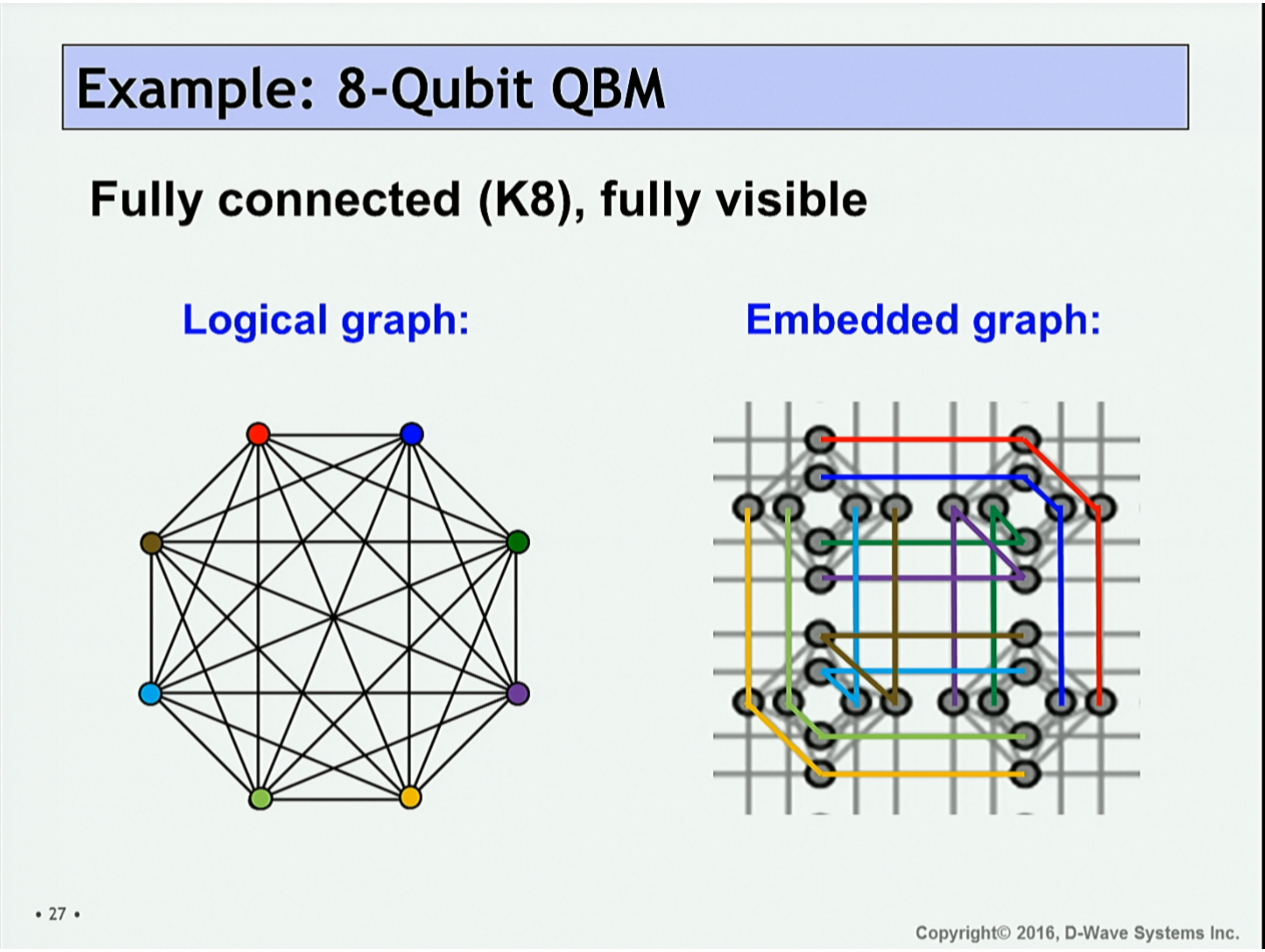
Dynamical freeze-out



Equilibration During the Annealing

Amin, PRA 92, 052323 (2015)

Open quantum calculations of a 16 qubit random problem



Example: 8-Qubit QBM

Fully connected (K8), fully visible

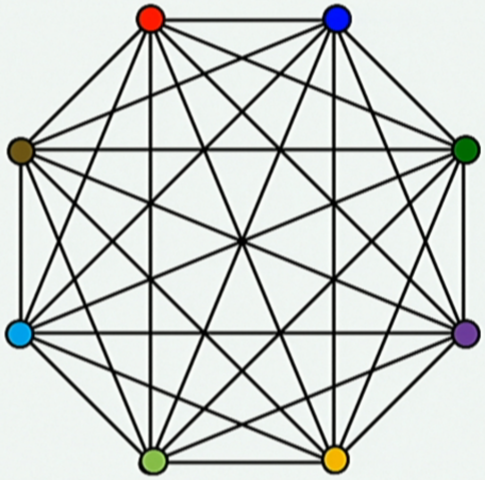
Logical graph:

Embedded graph:

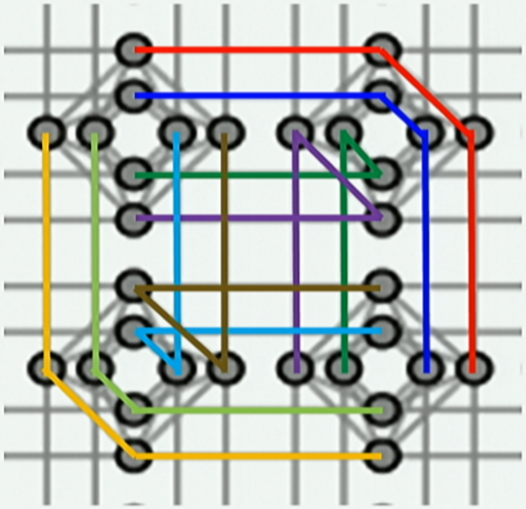
Example: 8-Qubit QBM

Fully connected (K8), fully visible

Logical graph:



Embedded graph:



Logical Hamiltonian

D-Wave Hamiltonian: $H(t) = \mathbf{A(s)}H_D + \boldsymbol{B(s)}H_P$

$$H_D = -\sum_{\mathrm{i}=1}^{\mathrm{N}} \sigma_{\mathrm{i}}^{\mathrm{x}} \qquad H_P = \sum_{i=1}^{N} h_i \sigma_i^z + \sum_{i,j=1}^{N} J_{ij} \sigma_i^z \sigma_j^z$$

Logical Hamiltonian (dimensionless):

$$H = -\sum_a \Gamma_a \sigma_a^x - \sum_a b_a \sigma_a^z - \sum_{a,b} w_{ab} \sigma_a^z \sigma_b^z$$

Chain's effective tunneling amplitude

$$b_a = \frac{4h_a B(s^*)}{k_B T} \qquad w_{ab} = \frac{2J_{ab} B(s^*)}{k_B T} \qquad \Gamma_a = \frac{\Delta_{eff}(s^*)}{2k_B T}$$

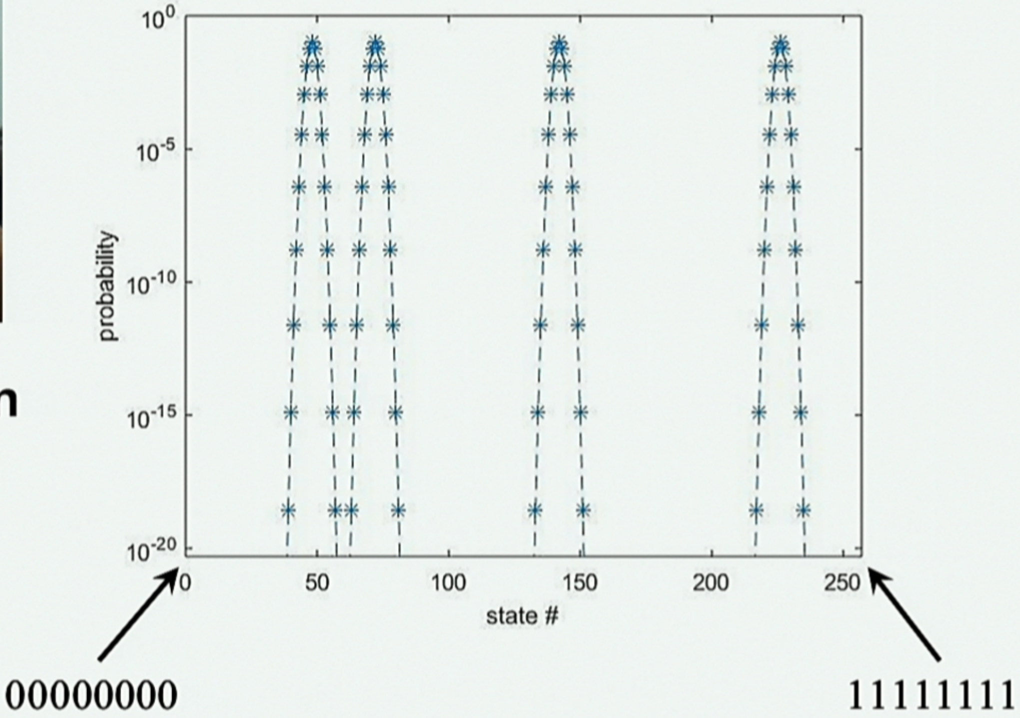
Training QBM

Amir Khoshaman

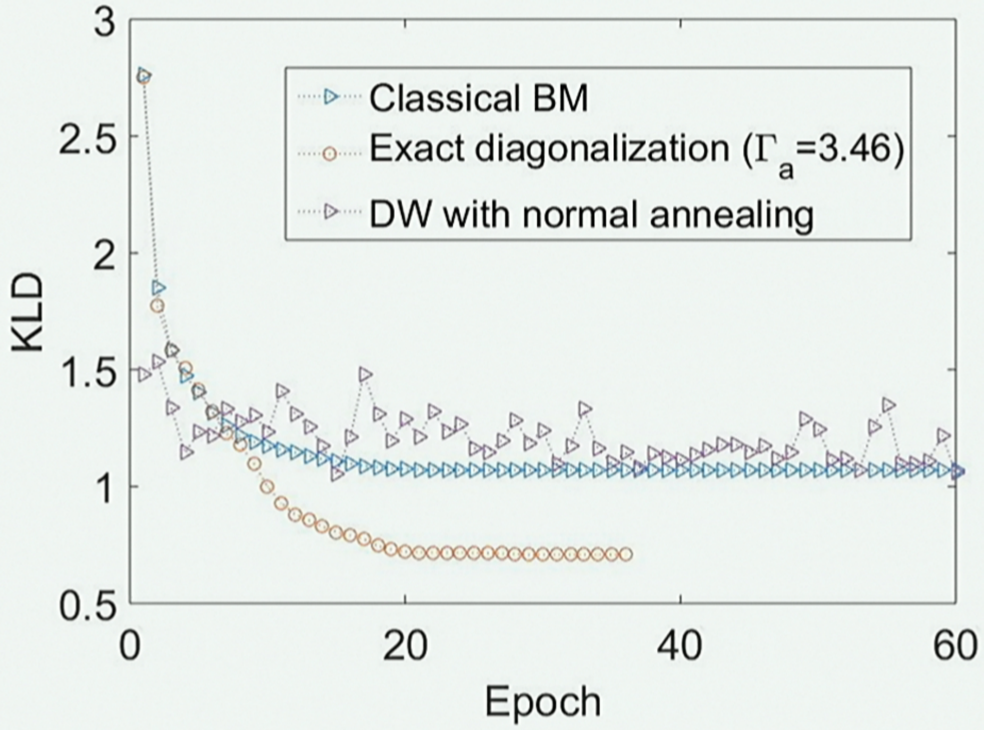
Training QBM

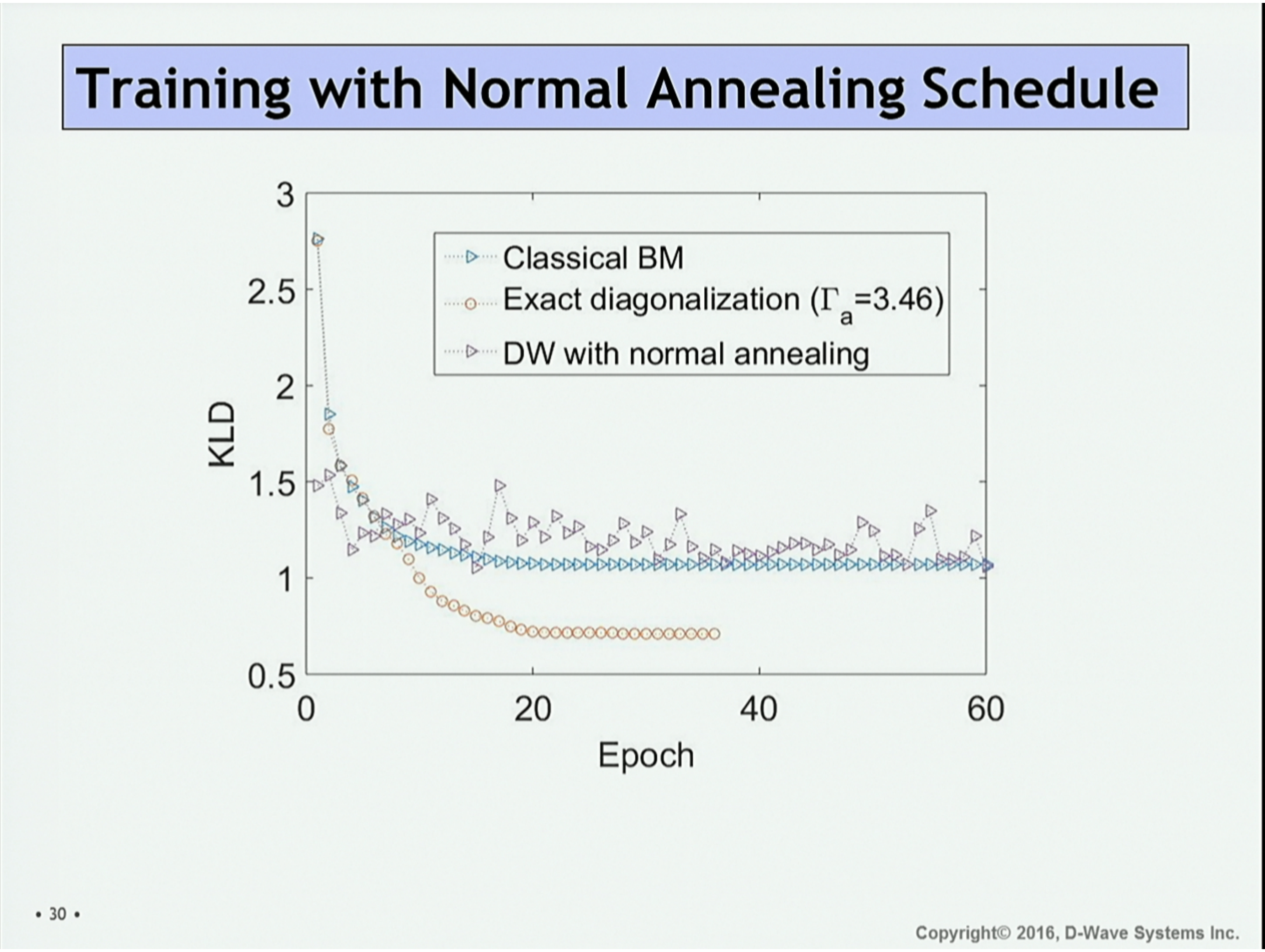
Amir Khoshaman

Training set: 4 Gaussian peaks



Training with Normal Annealing Schedule

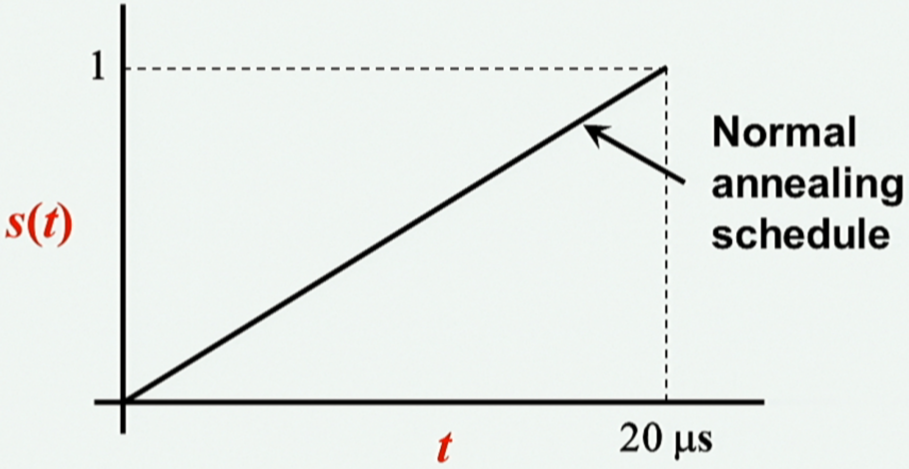




Training with Normal Annealing Schedule

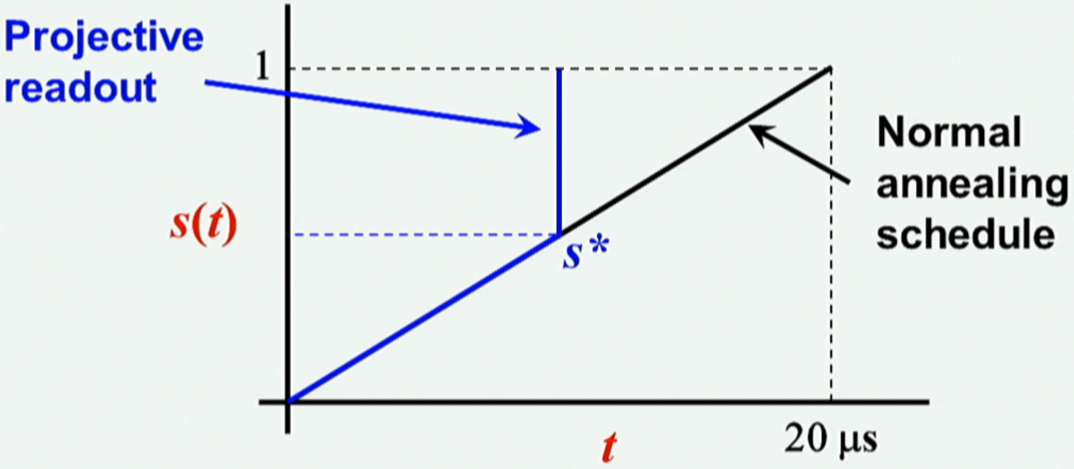
Annealing with Ramp

D-Wave Hamiltonian: $H(t) = \mathbf{A(s)}\, H_D + \mathbf{B(s)}\, H_P$



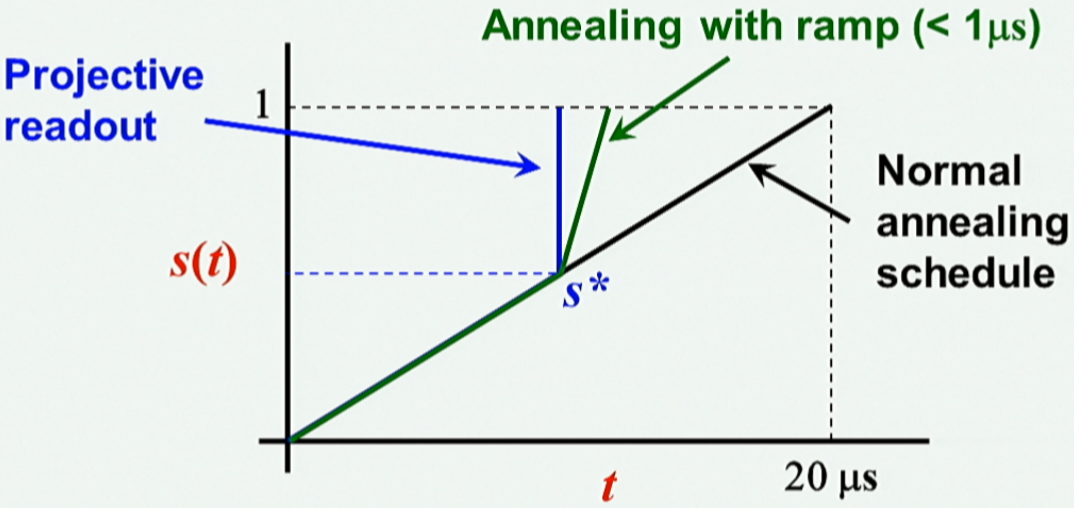
Annealing with Ramp

D-Wave Hamiltonian: $H(t) = \mathbf{A(s)}\, H_D + \mathbf{B(s)}\, H_P$



Annealing with Ramp

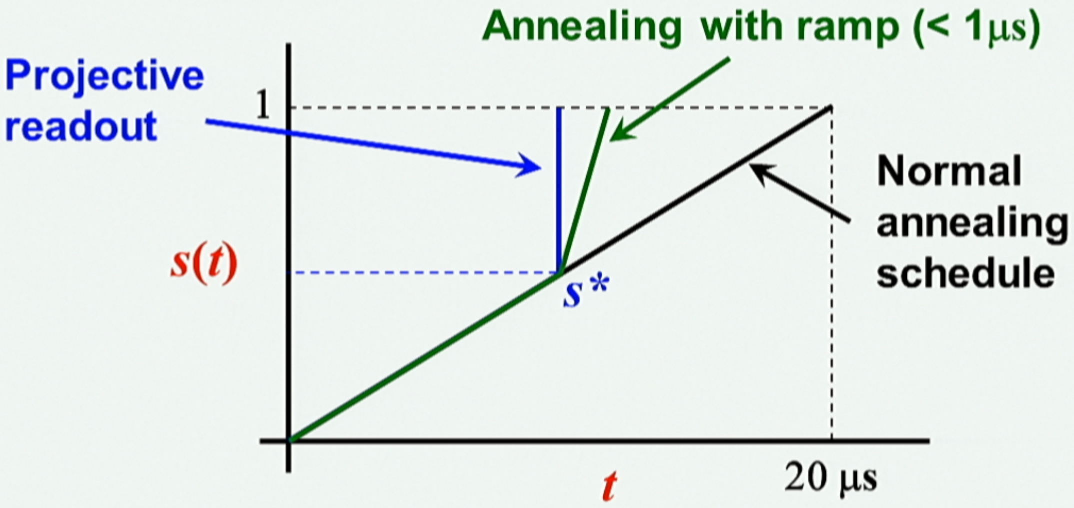
D-Wave Hamiltonian: $H(t) = \mathbf{A}(s)\, H_D + \mathbf{B}(s)\, H_P$



Dickson et al., Nat. Commun. 4, 1903 (2013)

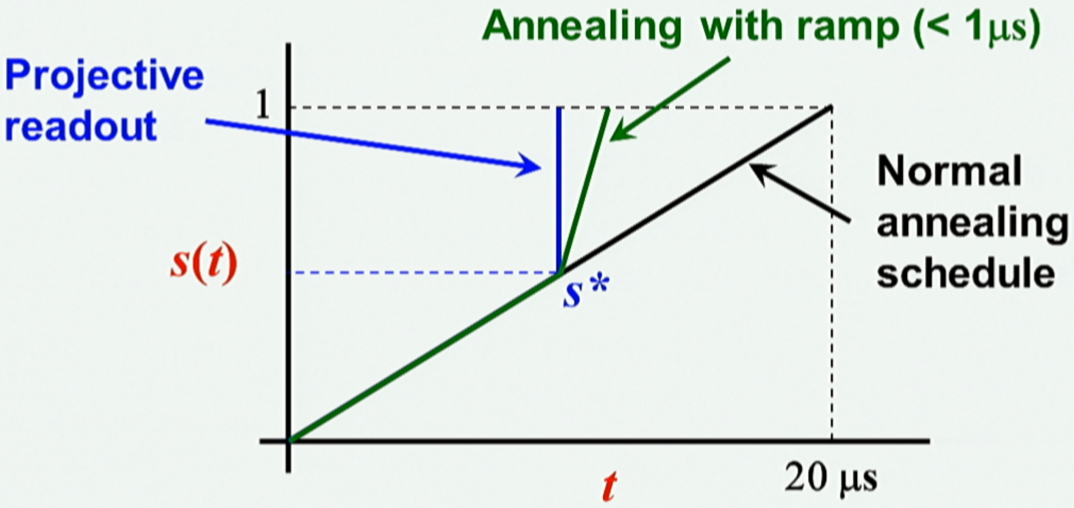
Correcting Distortions Due to the Ramp

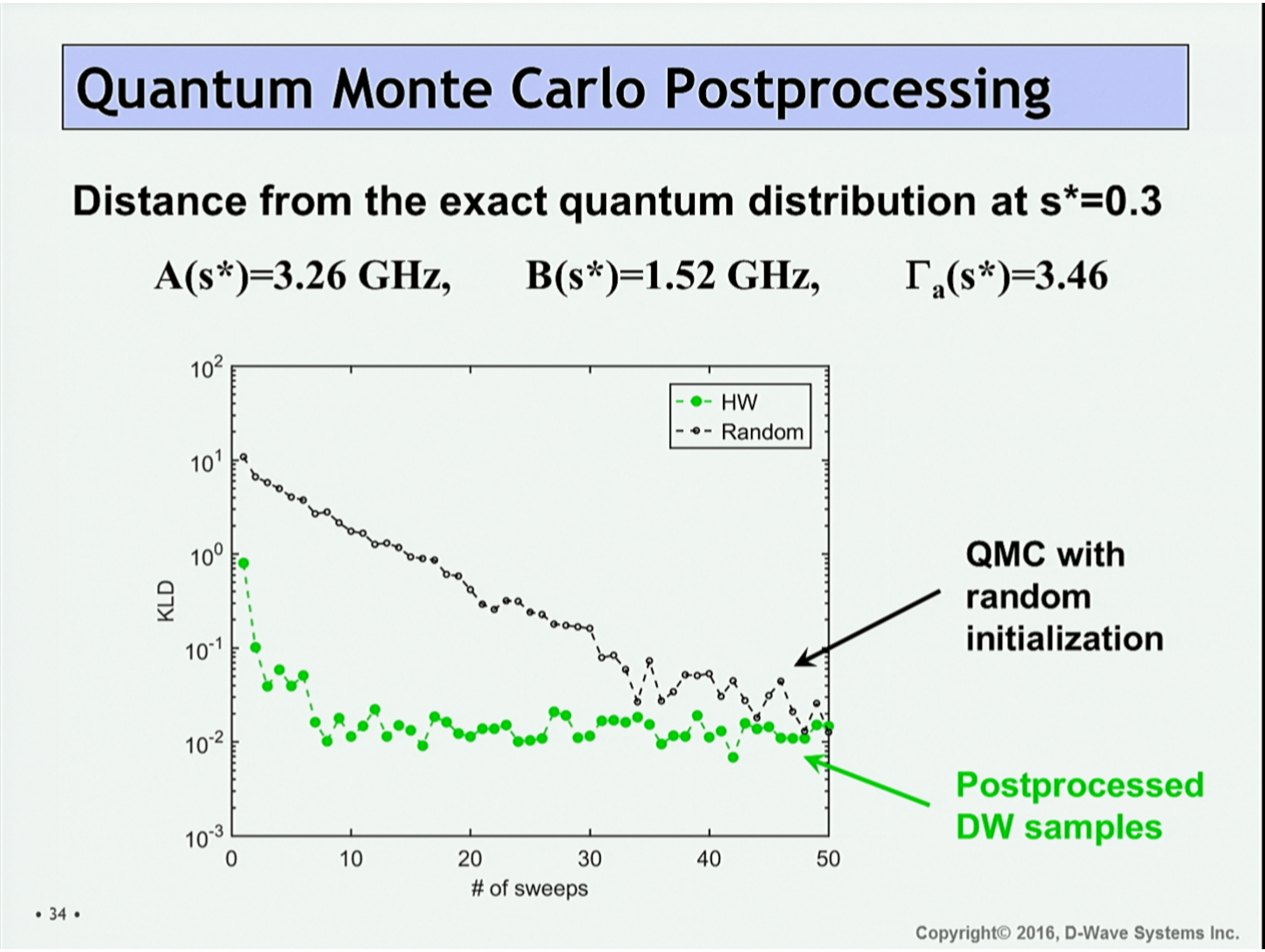
If evolution during the ramp is local (1-2 qubits), a few sweeps of QMC can restore the distribution



Correcting Distortions Due to the Ramp

If evolution during the ramp is local (1-2 qubits), a few sweeps of QMC can restore the distribution





Quantum Monte Carlo Postprocessing

Distance from the exact quantum distribution at s*=0.3

A(s*)=3.26 GHz, B(s*)=1.52 GHz, Γ_a(s*)=3.46

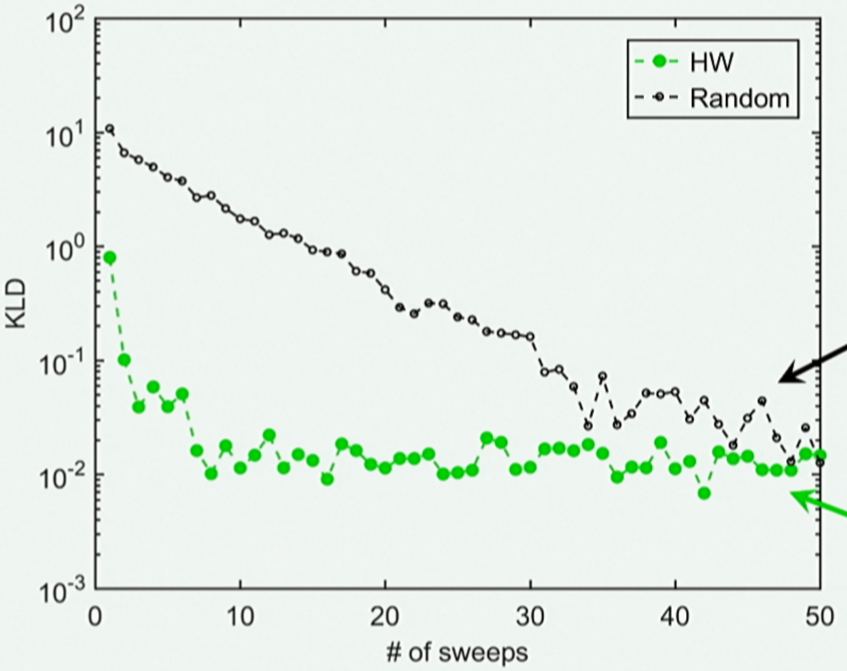
QMC with random initialization

Postprocessed DW samples

Quantum Monte Carlo Postprocessing

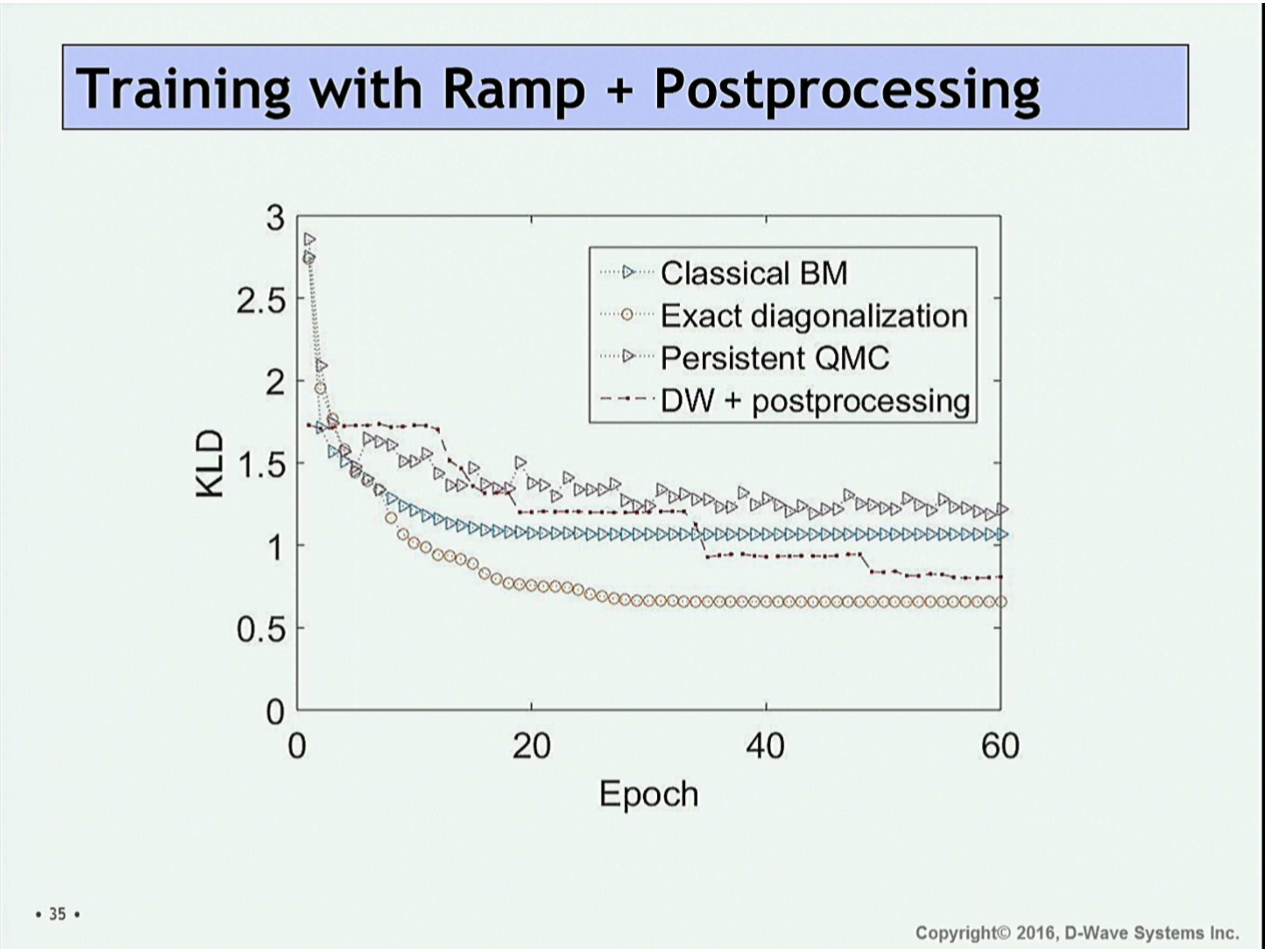
Distance from the exact quantum distribution at s*=0.3

$A(s^*)=3.26$ GHz, $B(s^*)=1.52$ GHz, $\Gamma_a(s^*)=3.46$

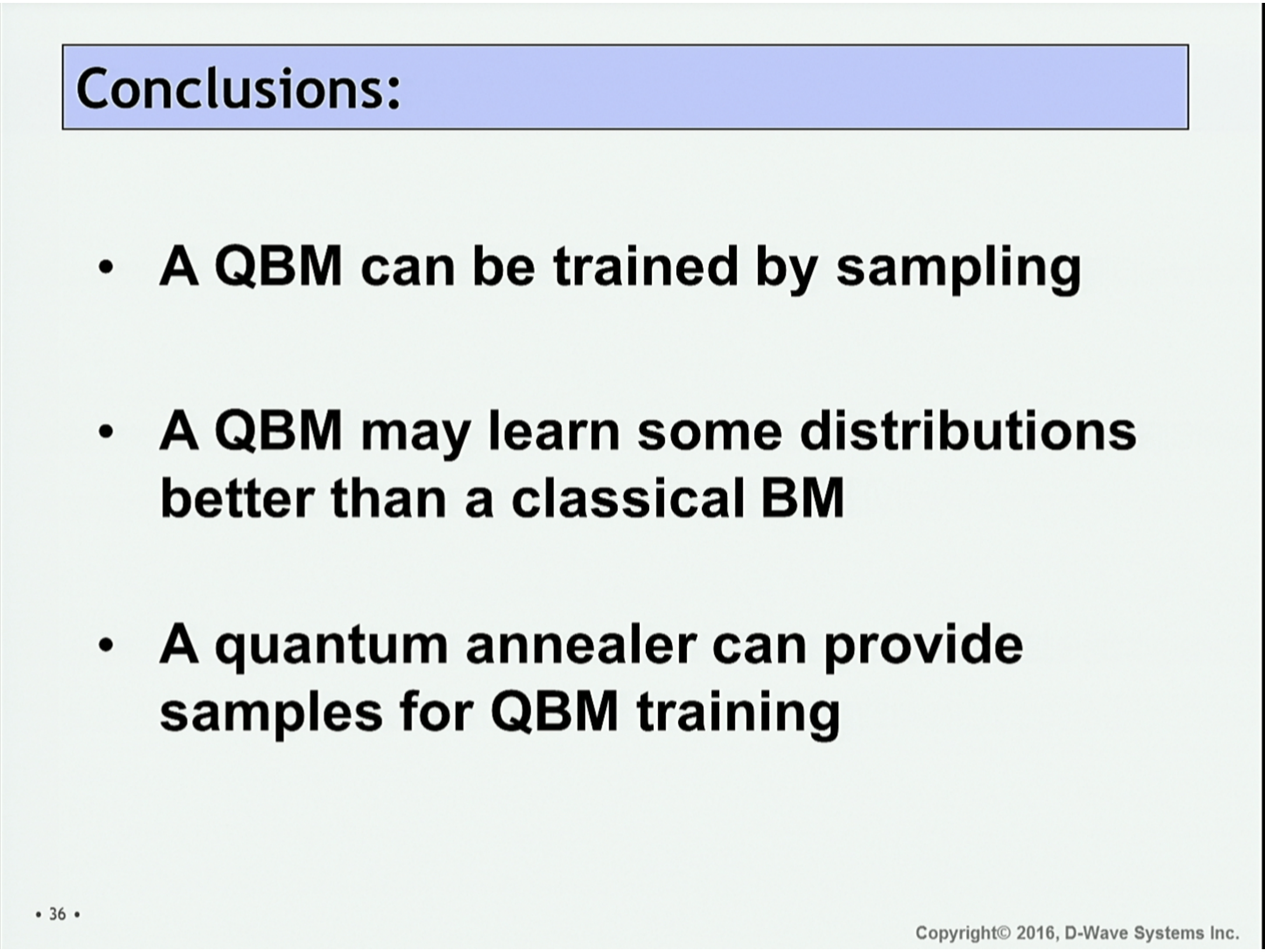


QMC with random initialization

Postprocessed DW samples



Training with Ramp + Postprocessing



Conclusions:

- A QBM can be trained by sampling
- A QBM may learn some distributions better than a classical BM
- A quantum annealer can provide samples for QBM training