Accelerating spin-glass simulations using quantum annealers through deep learning

Sebastiano Pilati

University of Camerino

COLLABORATORS:
G. Scriva, A. Spano, Prof. A. Perali (UniCam)
B. McNaughton (UniCam & U. Antwerp)
Prof. M. V. Milošević (U. Antwerp)

HPC-QC CINECA, December 15-16, 2021

Classical spin glasses

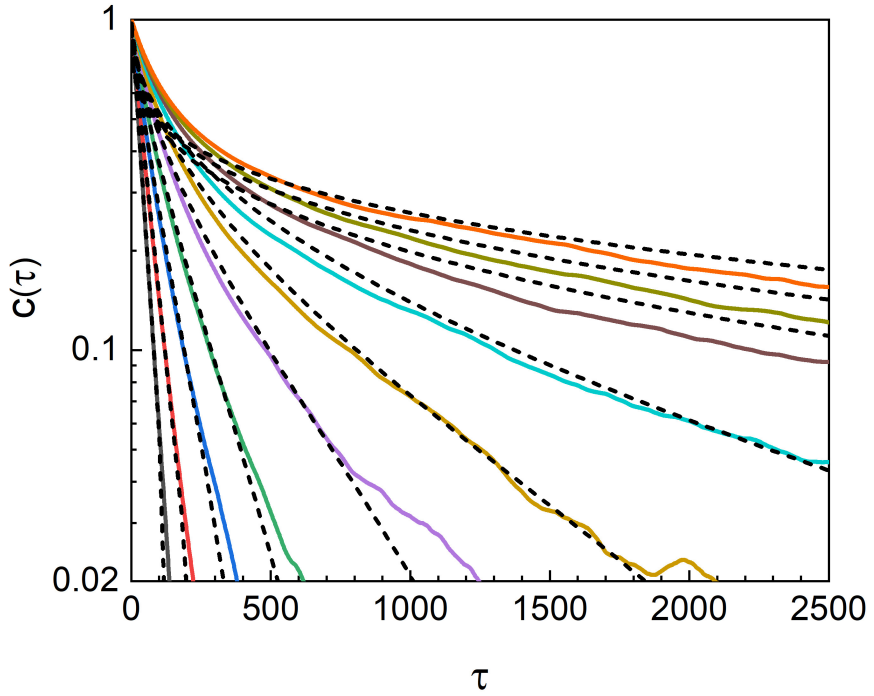
$$H = \sum_{ij} J_{ij} x_i x_j$$

$x_i = \pm 1$ binary variables on, e.g., a 2D square lattice, nearest neighbor interaction

J_{ij}~standard normal distribution

MC sim. with single-spin flip Metropolis algorithm

Autocorrelation: $c(\tau) = \frac{\langle E_{t+\tau} E_t \rangle - \langle E \rangle^2}{\langle E^2 \rangle - \langle E \rangle^2}$



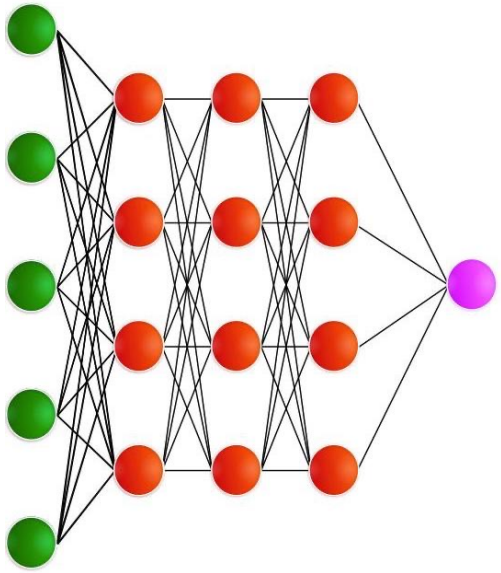
$\beta = 1$
$\beta = 0.9$
⋮
$\beta = 0.1$

- Spin-glass phase at low T $(T = 0$ in 2D square latt.$)$
- Finding ground-state is NP-complete on non-planar graphs
- Computing partition function in D>2 is NP-hard

Quantum annealers

$$H_{\mathrm{cl}} = -\sum_{ij} J_{ij}\sigma_i^z\sigma_j^z - \sum_i h_i\sigma_i^z \qquad H_{\mathrm{kin}} = -\Gamma(t)\sum_{ij}\sigma_i^x$$

Classical deep learning



Low-T equilibrium properties of spin glasses

$$H = \sum_{ij} J_{ij} x_i x_j$$

Generative neural networks

- Infer probability distribution of sampled data
- Trained by maximizing data log-likelihood
- Equivalent to minimize Kullback-Leibler divergence

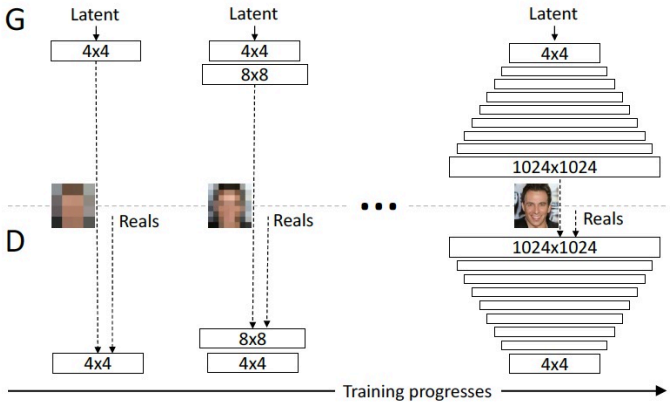
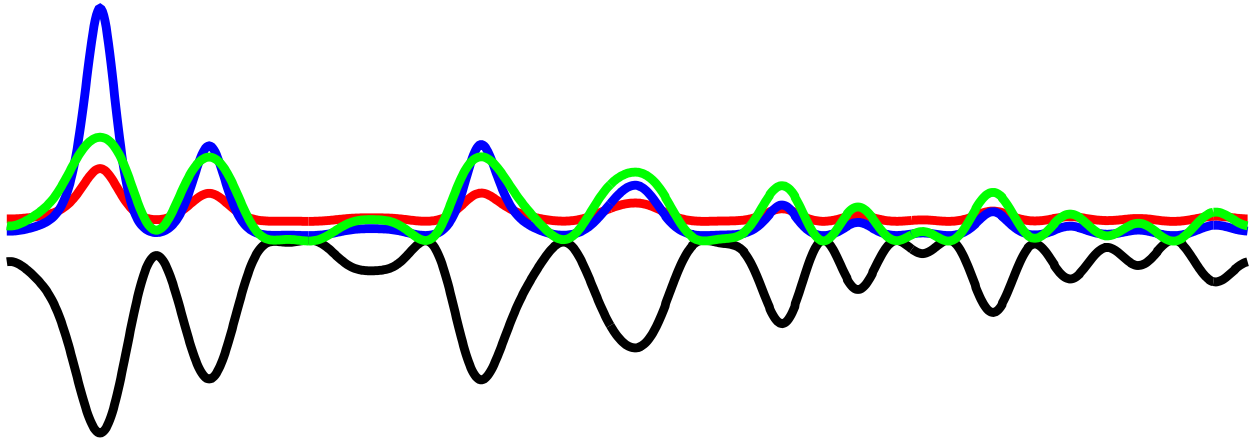


Image from: Karras et al., arXiv:1710.10196 (2017)

Can a generative NN learn the Boltzmann distribution of a spin glass?

Boltzmann distribution: $p_{\mathrm{BOLTZ}}(\boldsymbol{x}) = \frac{\exp[-\beta H(\mathbf{x})]}{Z}$

Partition function: $\mathrm{Z} = \sum_{\boldsymbol{x}} \exp[-\beta H(\mathbf{x})]$



Autoregressive neural networks

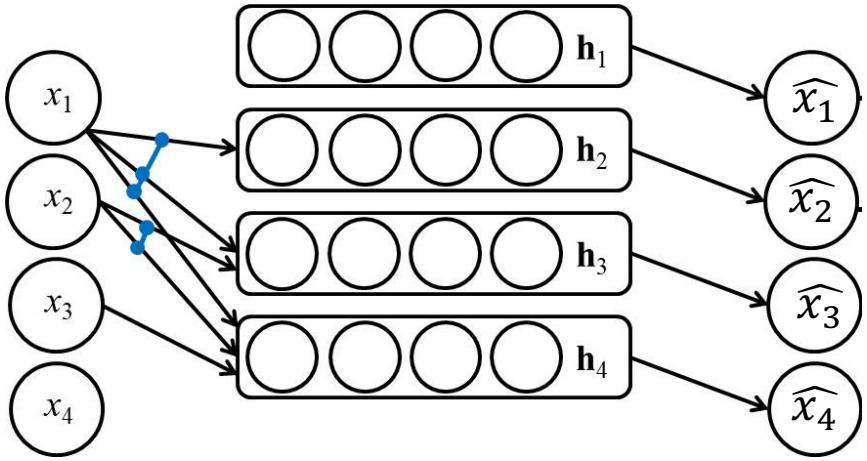
Chain of conditional distributions: $p_{\text{NETWORK}}(\mathbf{x}) = \prod_{d=1}^{N} p_d(x_d|\mathbf{x}_{<d})$

$\mathbf{x} = (x_1, \ldots, x_N)$

$\mathbf{x}_{<d} = (x_1, \ldots, x_{d-1})$

Direct samples via ancestral sampling: given $\mathbf{x}_{<d}$, sample $x_d \backsim p_d(x_d|\mathbf{x}_{<d})$

Normalized probability distribution: $\mathrm{Z} = \sum_{\boldsymbol{x}} p_{\text{NETWORK}}(\boldsymbol{x}) = 1$



Neural autoregressive distribution estimator:
H. Larochelle, I. JMLR: W&CP 15 (2011).

Masked Autoencoder for Distribution Estimation:
M. Germain et al., PMLR 37, 2015 (2015).

PixelCNN:
A. v. d. Oord et al., arXiv:1601.06759 (2016).

NEURAL MONTE CARLO SIMULATIONS

McNaughton, Milošević, Perali, SP, PRE (2020)

Proposal matrix in Metropolis-Hastings algorithm: $w_{\mathbf{x}'\mathbf{x}} = p_{\mathrm{NETWORK}}(\mathbf{x}')$ (use ancestral sampling)

Acceptance probability: $A_{\mathbf{x}'\mathbf{x}} = \min\left(1, \frac{p_{\mathrm{BOLTZ}}(\mathbf{x}')w_{\mathbf{x}\mathbf{x}'}}{p_{\mathrm{BOLTZ}}(\mathbf{x})w_{\mathbf{x}'\mathbf{x}}}\right)$

We only need: $p_{\mathrm{BOLTZ}}(\mathbf{x}) > 0 \implies p_{\mathrm{NETWORK}}(\mathbf{x}) > 0$

if: $p_{\mathrm{NETWORK}}(\mathbf{x}) \cong p_{\mathrm{BOLTZ}}(\mathbf{x})$ → $A_{\mathbf{x}'\mathbf{x}} \cong \mathbf{1}$ → efficient simulation!

Related work:

- D. Wu et al., PRL 122, 080602 (2019): variational optimization of autoregressive networks.
- K. A. Nicoli et al., PRE 101, 023304 (2020): MCMC+autoregressive n. for ferromagnetic models.

- F. Noè et al., Science 365, 1147 (2019): normalizing flows for complex-molecule simulations.
- X. Ding et al., J. Phys. Chem. B, 124, 10166 (2020): normalizing flows for free-energy computations.
- M. Gabrié et al., arXiv:2105.12603 (2021): adaptive MCMC via normalizing flows.
- G. S. Hartnett, M. Mohseni,arXiv:2001.00585v2 (2020): spin-glass simulations via normalizing flows.

Two training strategies

1) Reverse Kullback-Leibler divergence

$$\mathrm{KL}(p_{\mathrm{NETWORK}}|p_{\mathrm{BOLTZ}}) = \langle \log(p_{\mathrm{NETWORK}}) \rangle_{p_{\mathrm{NETWORK}}} - \langle \log(p_{\mathrm{BOLTZ}}) \rangle_{p_{\mathrm{NETWORK}}}$$

D. Wu et al., Phys. Rev. Lett. 122, 080602 (2019)

- Trained on data generated by network
- Reinforcement learning
- Prone to localization in metastable states

X. Ding et al., J. Phys. Chem. B, 124, 10166−10172 (2020)
F. Noè et al., Science 365, 1147 (2019)
M. Gabrié et al., arXiv:2105.12603 (2021)
H. Wu, arXiv:2002.06707 (2020)

---

2) Forward Kullback-Leibler divergence

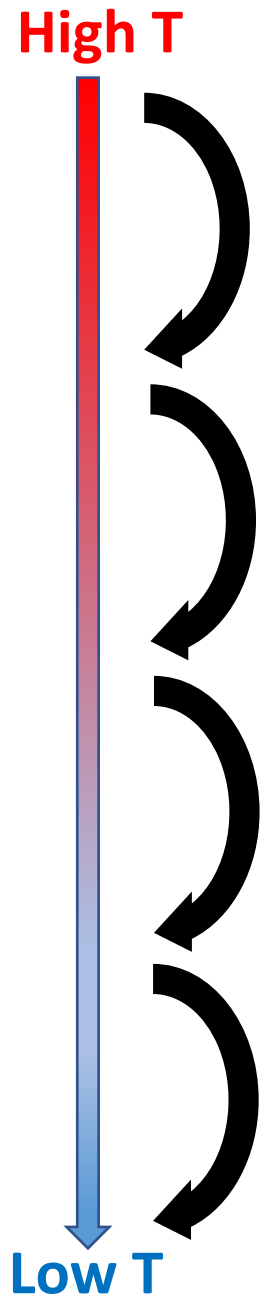
$$\mathrm{KL}(p_{\mathrm{BOLTZ}}|p_{\mathrm{NETWORK}}) = \langle \log(p_{\mathrm{BOLTZ}}) \rangle_{p_{\mathrm{BOLTZ}}} - \langle \log(p_{\mathrm{NETWORK}}) \rangle_{p_{\mathrm{BOLTZ}}}$$

- **Needs a-priori data for training**

Sequential tempering

D-Wave quantum annealer

NEURAL SEQUENTIAL TEMPERING



Run conventional MC (easy), save data, train network…

…run neural MC using previous network, save data, train new network…

…run neural MC using previous network, save data, train new network…

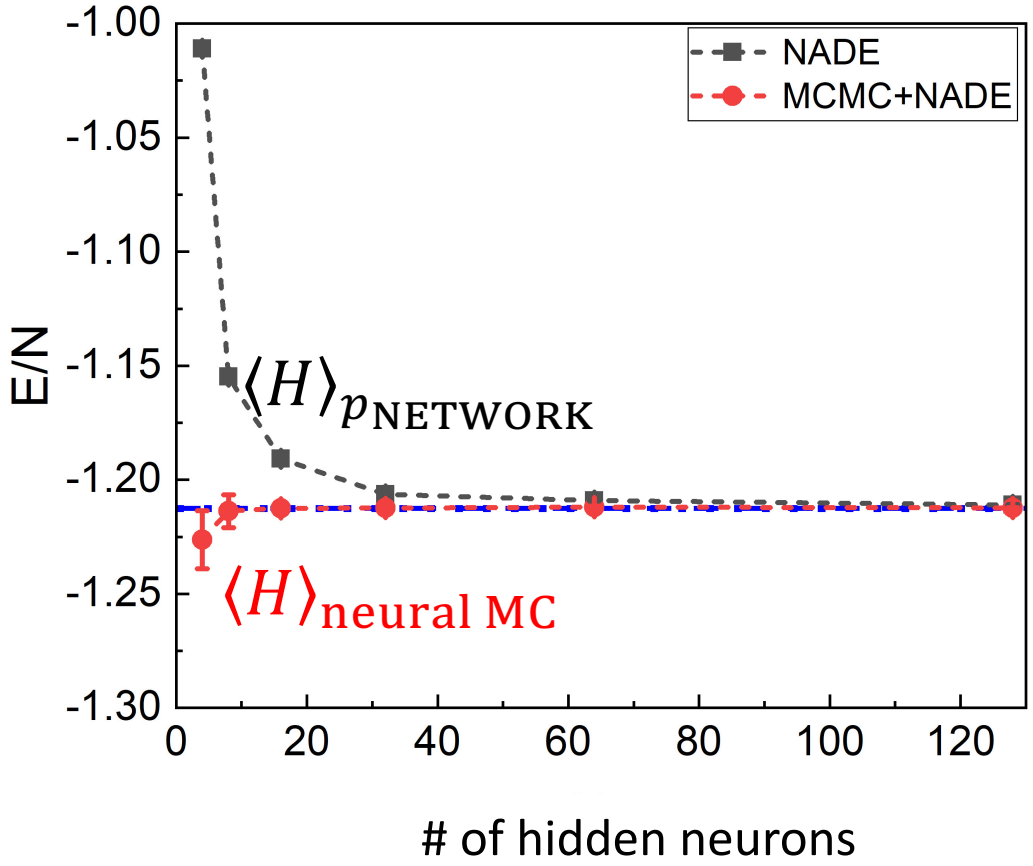
…

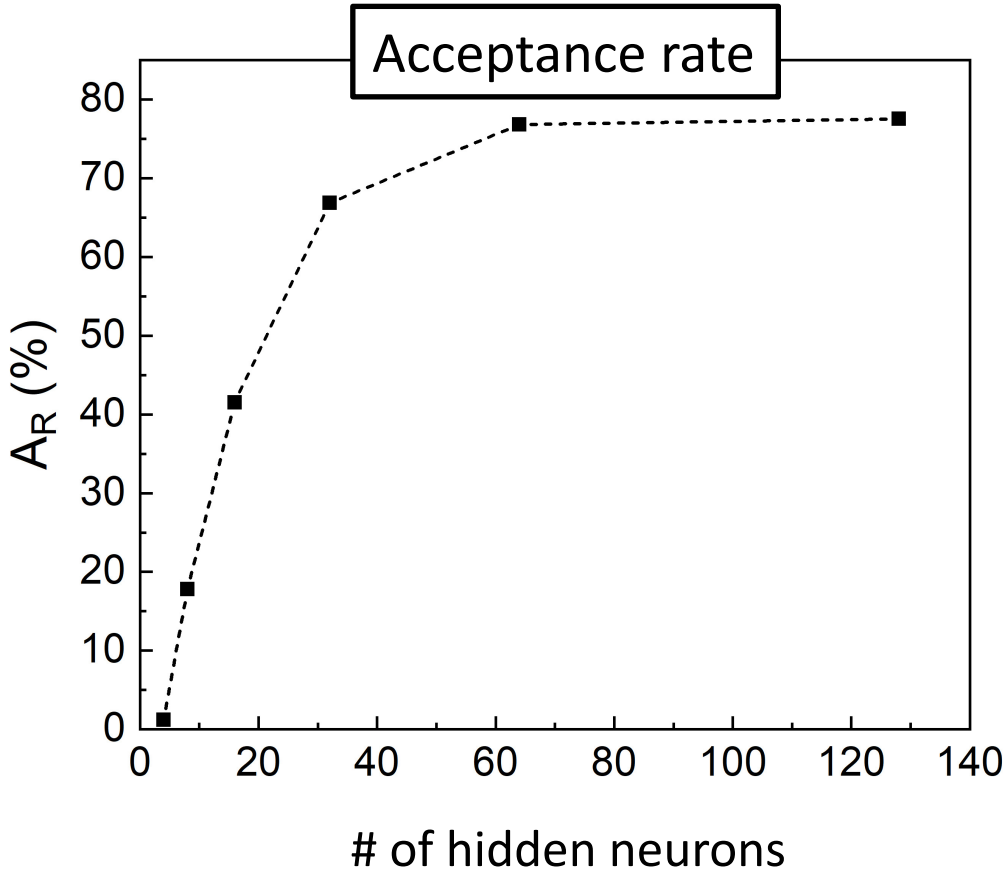
Our goal: use a quantum annealer to start directly from low T

→ simulate spin-glass at low T

TEST ON SYNTHETIC DATA

2D square lattice N=100
Nearest-neighbor gaussian couplings





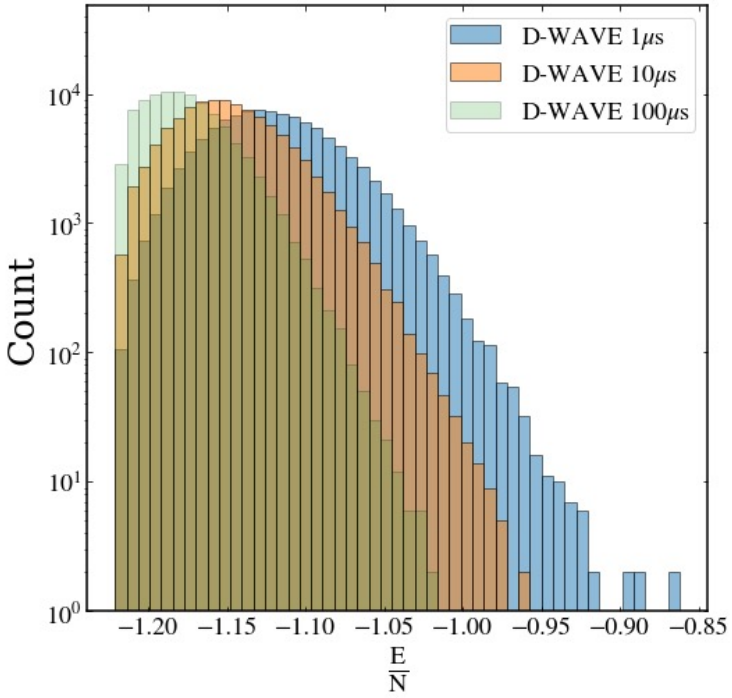
- Autoregressive network approximates Boltzmann distribution.
- Neural MC is “always” unbiased.
- Efficiency increases with network depth.

McNaughton, Milošević, Perali, SP, PRE (2020)

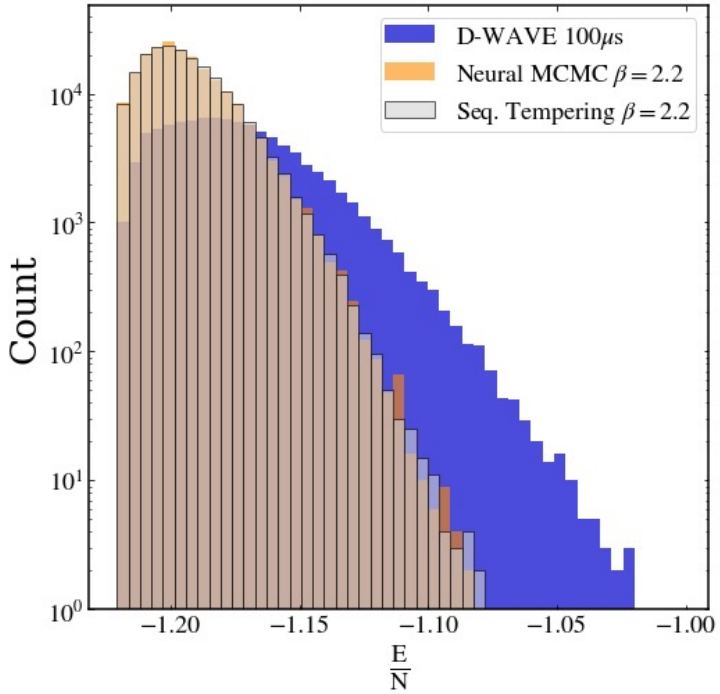
Neural MCMC with networks trained on D-WAVE data

2D square lattice N=100
Nearest-neighbor gaussian couplings

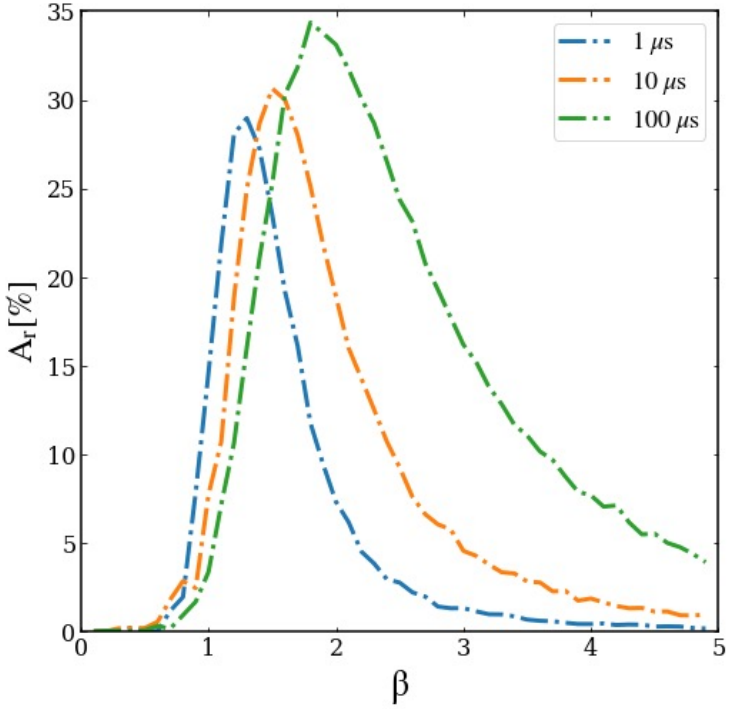
Energy histogram of samples from D-WAVE 2000Q



Histogram after neural MCMC

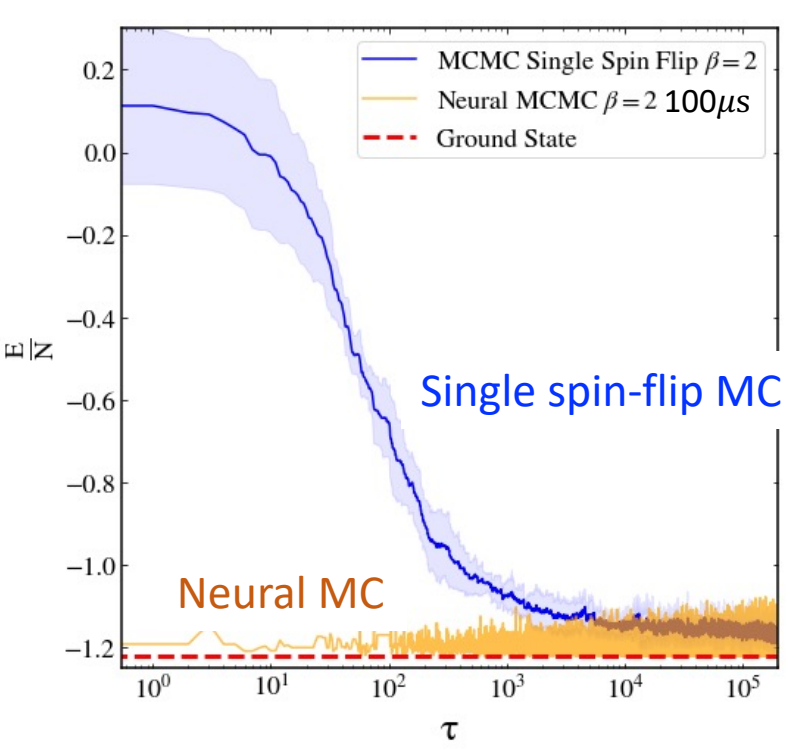


Acceptance rate

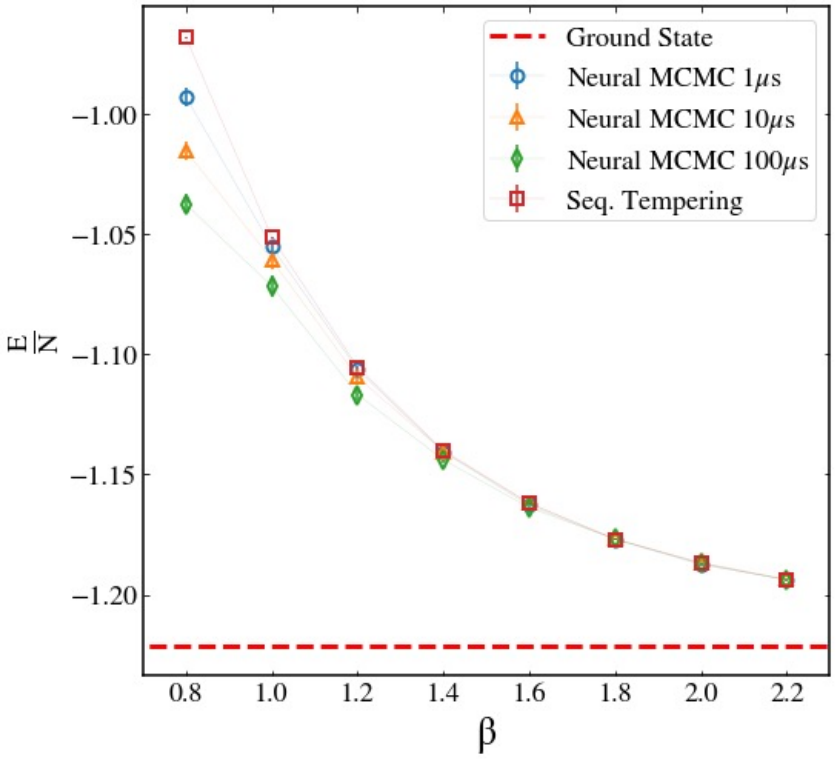


Slow annealing → higher acc. rate at low T

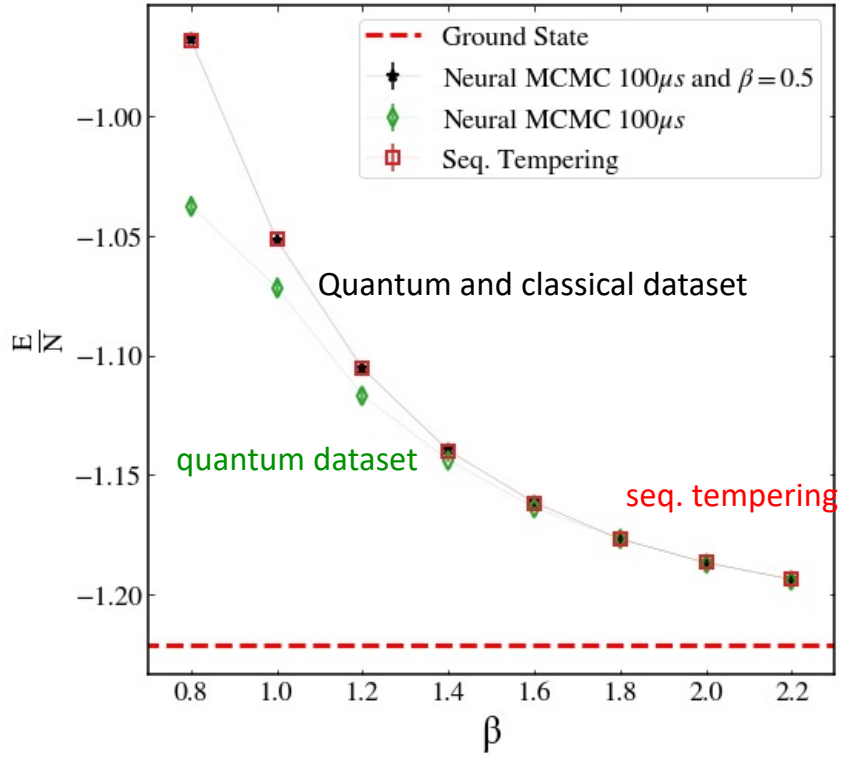
Example of low-T simulation: β = 2



# of MC steps



Inverse temperature



Inverse temperature

- Unbiased and efficient neural MC simulation at low T
- Breakdown of ergodicity at high T → hybrid algorithm
- Hybrid classical-quantum dataset restores ergodicity

2D square lattice N=100
Nearest-neighbor gaussian couplings

Conclusions:

McNaughton, Milošević, Perali, SP, PRE (2020)

- Autoregressive NNs can approximate spin-glasses.
- **Proof of principle**: data from D-WAVE can be used to accelerate equilibrium MC simulations of spin-glasses at low-T.
- **Work in progress**: simulating larger systems (N ≈ 1000) and higher connectivity. Hybrid MC algorithm implemented for high-T.
- Study of "ergodicity" of autoregressive neural networks, avoiding *weight concentration* as in., e.g., *random circuit sampling*.

ISCRA CINECA

CINECA

PRIN MIUR