Title: Neural annealing and visualization of autoregressive neural networks

Speakers: Estelle Inack

Collection: Quantum Criticality: Gauge Fields and Matter

Date: May 18, 2022 - 4:30 PM

URL: https://pirsa.org/22050040

Abstract: Artificial neural networks have been widely adopted as ansatzes to study classical and quantum systems. However, some notably hard systems such as those exhibiting glassiness and frustration have mainly achieved unsatisfactory results despite their representational power and entanglement content, thus, suggesting a potential conservation of computational complexity in the learning process. We explore this possibility by implementing the neural annealing method with autoregressive neural networks on a model that exhibits glassy and fractal dynamics: the two-dimensional Newman-Moore model on a triangular lattice. We find that the annealing dynamics is globally unstable because of highly chaotic loss landscapes. Furthermore, even when the correct ground state energy is found, the neural network generally cannot find degenerate ground-state configurations due to mode collapse. These findings indicate that the glassy dynamics exhibited by the Newman-Moore model caused by the presence of fracton excitations in the configurational space likely manifests itself through trainability issues and mode collapse in the optimization landscape.

Neural annealing and visualization of autoregressive neural networks

Estelle Inack
Research Scientist at the Perimeter Institute
Cofounder & CTO of yiyaniQ

Quantum Criticality: Gauge Fields and Matter

18th May 2022

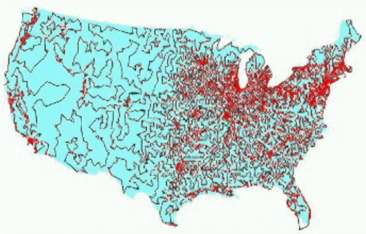
Hibat-Allah, E. M. I., Wiersema, Melko, Carrasquilla, *NMI volume 3, pages 952–961 (2021)*

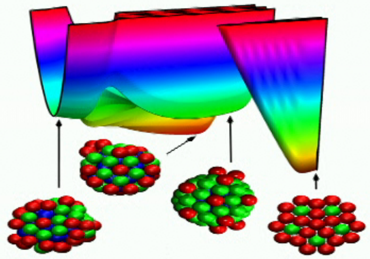
E. M. I., Morawetz and Melko, arXiv:2204.11272

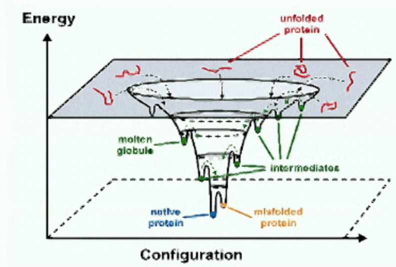
QUANTUM INTELLIGENCE LAB

yiyaniQ

Optimization problems







And many more

$$H_p = \sum_p \sum_{j_1,\ldots,j_p} J_{j_1,\ldots,j_p} \sigma_{j_1}^z \ldots \sigma_{j_p}^z$$

E.g: Disordered Ising glass in a random longitudinal field

$$H_p = -\sum_{i,j} J_{i,j}\, \sigma_i^z \sigma_j^z - \sum_{i=1}^{N} h_i\, \sigma_i^z$$

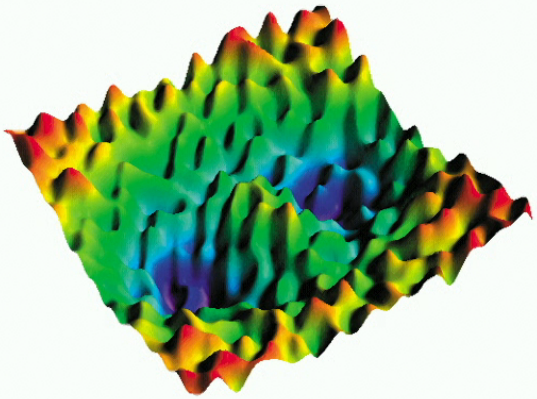
Finding the ground state among 2^N possible solutions is an **NP-hard** problem.

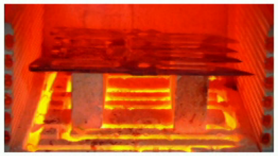
Lucas, Front. Phys. (2014)

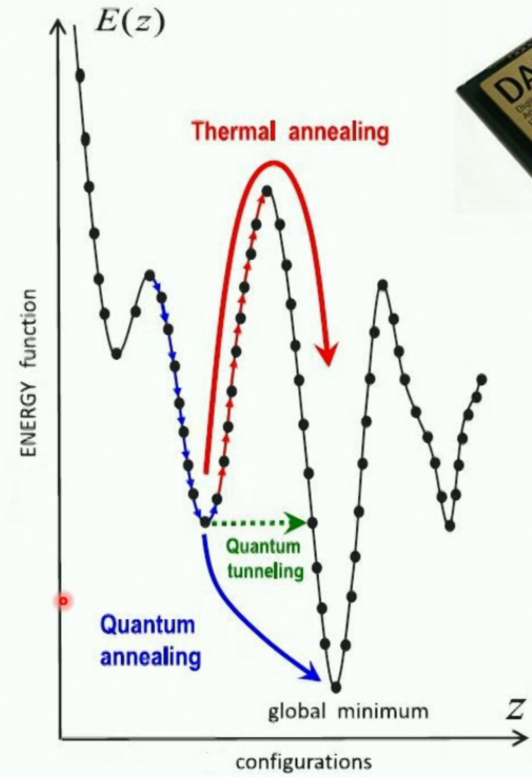
Heuristic methods

Aramon et al., Frontiers in Physics 7, 48 (2019).

Rugged an exponentially large





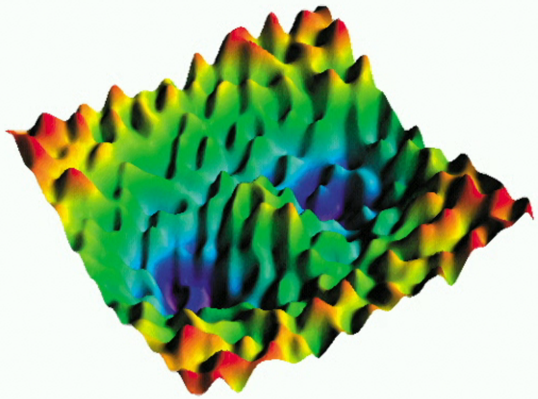


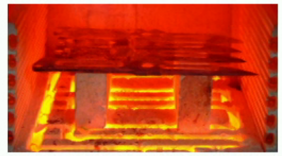
- Kirkpatrick, Gelatt, Vecchi, Science 220, 671 (1983)
- Kadowaki and Nishimori, PRE 58, 5355 (1998)
- Farhi, Goldstone, Gutmann, Lapan, Lundgren, and Preda, Science 292, 472 (2001)
- Santoro, Martonak, Tosatti, and Car, Science 295, 2427 (2002)

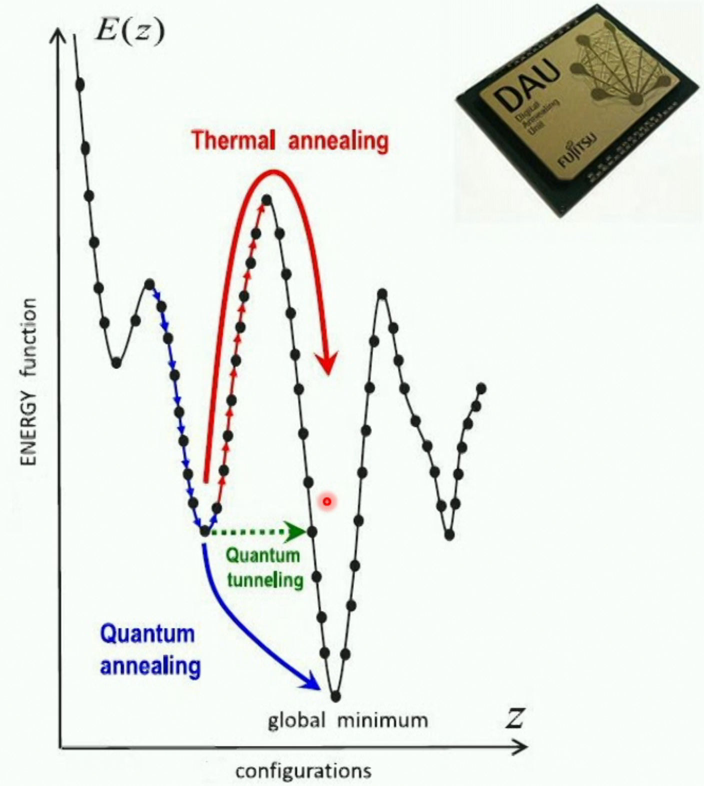
Heuristic methods

Aramon et al., Frontiers in Physics 7, 48 (2019).

Rugged an exponentially large







- Kirkpatrick, Gelatt, Vecchi, Science 220, 671 (1983)
- Kadowaki and Nishimori, PRE 58, 5355 (1998)
- Farhi, Goldstone, Gutmann, Lapan, Lundgren, and Preda, Science 292, 472 (2001)
- Santoro, Martonak, Tosatti, and Car, Science 295, 2427 (2002)

Simulating annealing paradigm with Monte Carlo methods

Classical systems	Quantum systems
• **Metropolis Monte Carlo** *Kirkpatrick, Gelatt, and Vecchi, Science 220, 671 (1983)* • **Replica Exchange/Parralel Tempering Monte Carlo** *Mandrà et al., PRA 94, 022337 (2016)* • **Population Annealing Monte Carlo** *Wang, Machta, and Katzgraber, PRE* ***92****, 013303 (2015)* • **Spin Vector Monte Carlo** *Shin et al., arXiv:1401.7087*	• **Path-integral Monte Carlo** *Santoro, et al., Science 295, 2427 (2002)* *Heim et al.,* Science 348, 215 (2015) • **Diffusion Monte Carlo** *Inack and Pilati, PRE 92, 053304 (2015)* *Jarret, Jordan, and Lackey, PRA 94, 042318 (2016)* • **Projector Monte Carlo** *Liu, Polkovnikov, and Sandvik, PRB 87, 174302 (2013)* • **Green's function Monte Carlo** *Stella and Santoro, PRE 75, 036703 (2007)*

Simulating classical and quantum systems with Artificial Neural networks

RESEARCH ARTICLE

MANY-BODY PHYSICS

Solving the quantum many-body problem with artificial neural networks

Giuseppe Carleo[1*] and Matthias Troyer[1,2]

Editors' Suggestion

Solving Statistical Mechanics Using Variational Autoregressive Networks

Dian Wu, Lei Wang, and Pan Zhang
Phys. Rev. Lett. **122**, 080602 – Published 28 February 2019

- Variational principles used to estimate equilibrium properties
- Neural networks used as variational ansatzes

CAN WE SIMULATE ANNEALING PARADIGM IN A VARIATIONAL FRAMEWORK?

Variational Monte Carlo method

VMC finds good approximations of ground state wave-functions of quantum systems by minimizing E_{var}.

$$E_{var} = \frac{\langle \Psi_\lambda | \hat{H} | \Psi_\lambda \rangle}{\langle \Psi_\lambda | \Psi_\lambda \rangle}$$

$$E_{var} \approx \frac{1}{\mathcal{N}_s} \sum_{\sigma \sim |\Psi_\lambda(\sigma)|^2} E_{loc}(\sigma)$$

$$E_{loc}(\sigma) = \frac{\langle \sigma | \hat{H} | \Psi_\lambda \rangle}{\langle \sigma | \Psi_\lambda \rangle}$$

OPTIMIZING THE ANSATZ

$$\frac{\partial E_{var}}{\partial \lambda_n} = 2\Re[\langle E_{loc}(\sigma) F_n^*(\sigma) \rangle - \langle E_{loc}(\sigma) \rangle \langle F_n^*(\sigma) \rangle]$$

$$F_n^*(\sigma) = \frac{\partial}{\partial_n} \log \Psi_\lambda^*(\sigma)$$

$$\lambda_n \leftarrow \lambda_n - \alpha \frac{\partial E_{var}}{\partial \lambda_n}$$

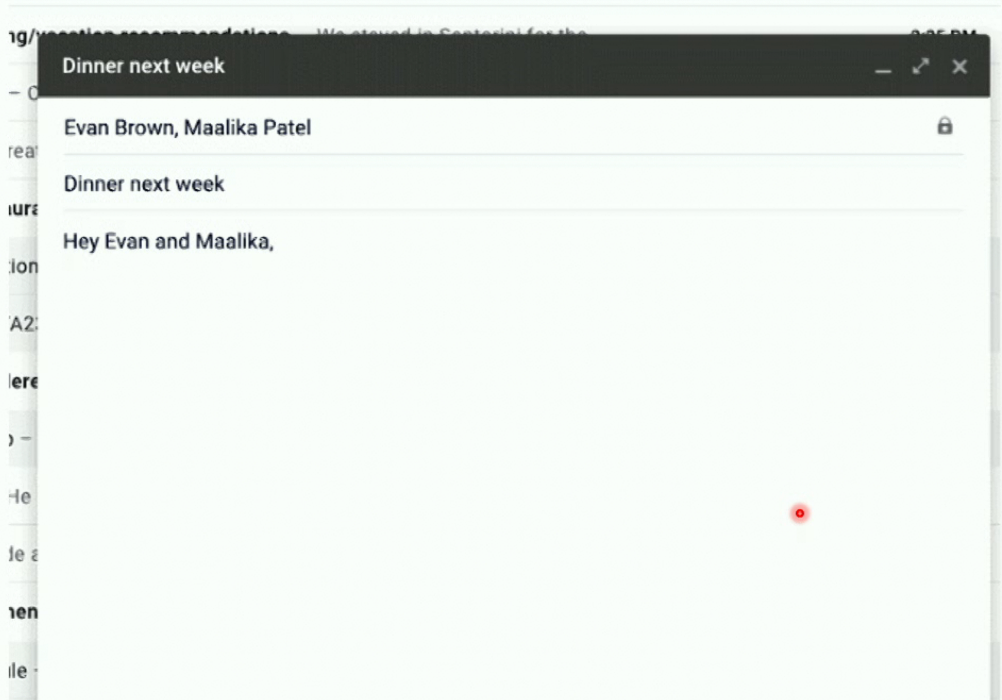
SGD, Momentum, SR, Adam,…

- Any ansatz gives an **upper bound** to the exact ground state energy.
- No **sign-problem** in VMC.

Becca and Sorella (2017)

NEURAL LANGUAGE MODELS

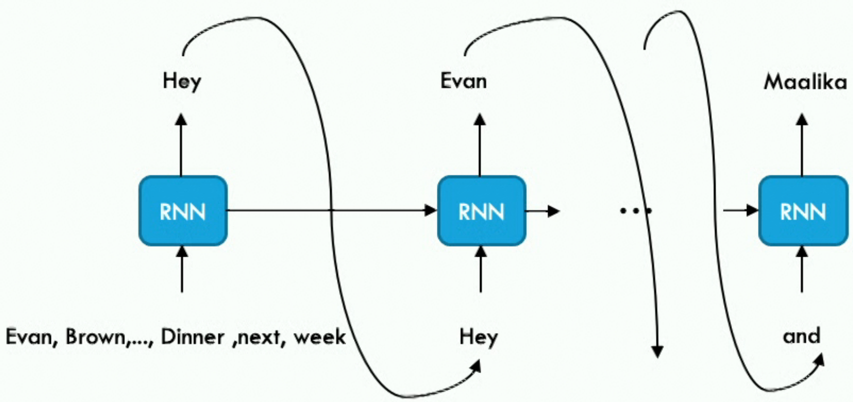
NEURAL LANGUAGE MODELS

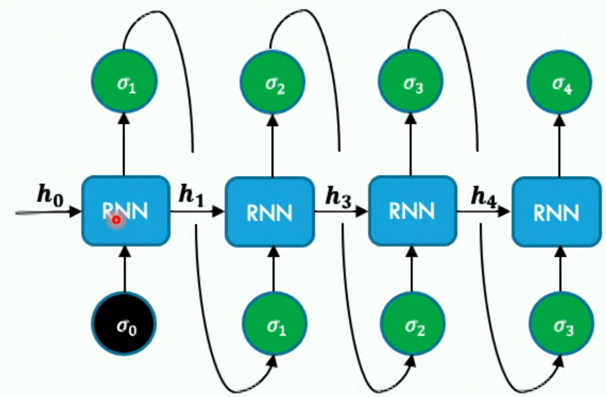


Smart Compose, Google AI

Recurrent Neural Network ansatzes

$\sigma_1\sigma_2\sigma_3\sigma_4$





P(Hey | Evan, Brown,..., Dinner, next, week) P(Evan | Evan, Brown,..., Dinner ,next, week, Hey) ... P(Maalika | Evan, Brown,...week, Hey, and)

= P(Hey, Evan, and, Maalika | Evan, Brown, Maalika, Patel, Dinner, next, week)

$P(\sigma_1,\sigma_2,\sigma_3,\sigma_4) = P(\sigma_1)P(\sigma_2|\sigma_1)P(\sigma_3|\sigma_1,\sigma_2)\,P(\sigma_4|\sigma_1,\sigma_2,\sigma_3)$

$\Psi(\sigma_1,\sigma_2,\sigma_3,\sigma_4)=\sqrt{P(\sigma_1,\sigma_2,\sigma_3,\sigma_4)}$

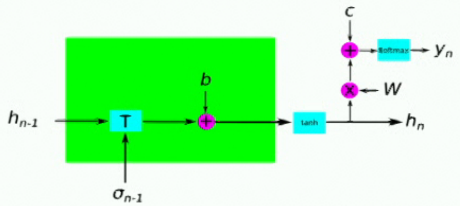
- h_n is called the hidden state and encode information about previous spins $\sigma_{n'<n}$
- Samples are generated exactly, **no autocorrelation time** and directly parallelizable
- Probability/wavefunction is **normalized** to unity.

Hibat-Allah, et al. Phys. Rev. Research **2,** 023358 (2020)

RNN architectures

[1] R Kelley (2016)
[2] Chang et al., arXiv:1710.02224 (2017)

Tensorized RNN cell [1]

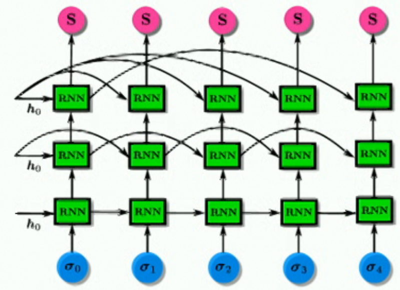


- $\{b, W, T, c\}$ *network parameters*
- f(x) is a non-linear activation function
- T is a tensor of dim $2 * d_h * d_h$

$$h_n = f(\sigma_{n-1}{}^{\intercal} T_n h_{n-1} + b_n)$$

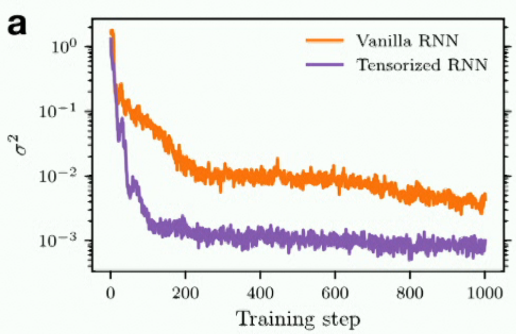
$$h_n = f(W[h_{n-1}; \sigma_{n-1}] + b) \quad \text{(Vanilla RNN)}$$

Dilated RNN architecture [2]

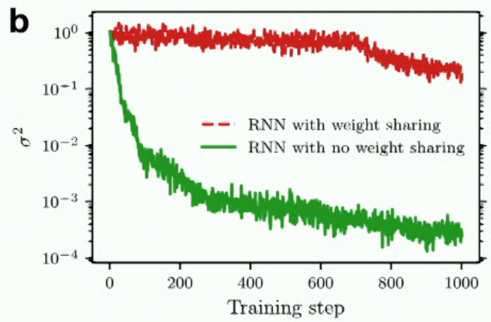


$$h_n = f\left(W^{(l)}[h^{(l)}_{\max(0,n-2^{l-1})}; h_n^{(l-1)}] + b\right)$$

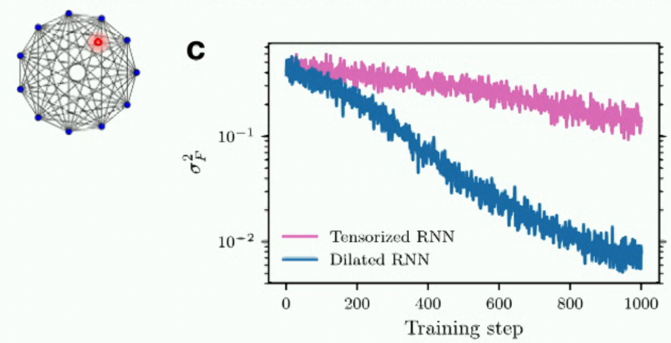
Quantum Ising chain (QIC) at criticality, N=100 spins



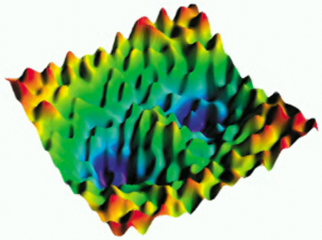
Random QIC at criticality, N=20 spins

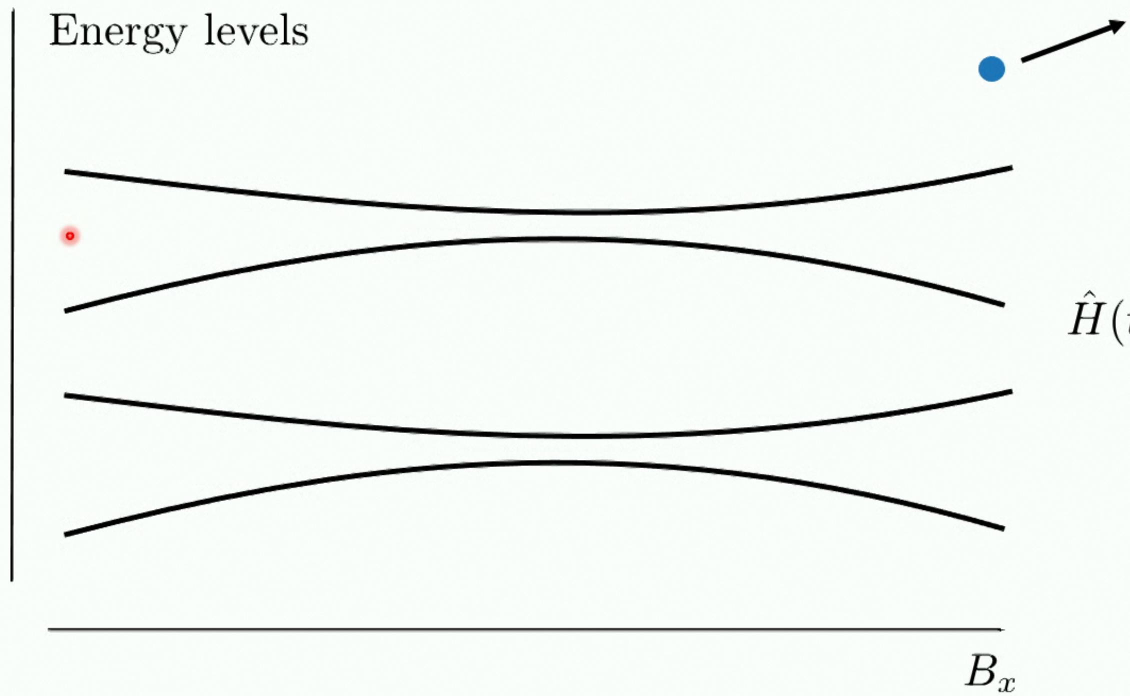


Sherrington-Kirkpatrick model at criticality, N=20 spins



VARIATIONAL QUANTUM ANNEALING





RNN wavefunction

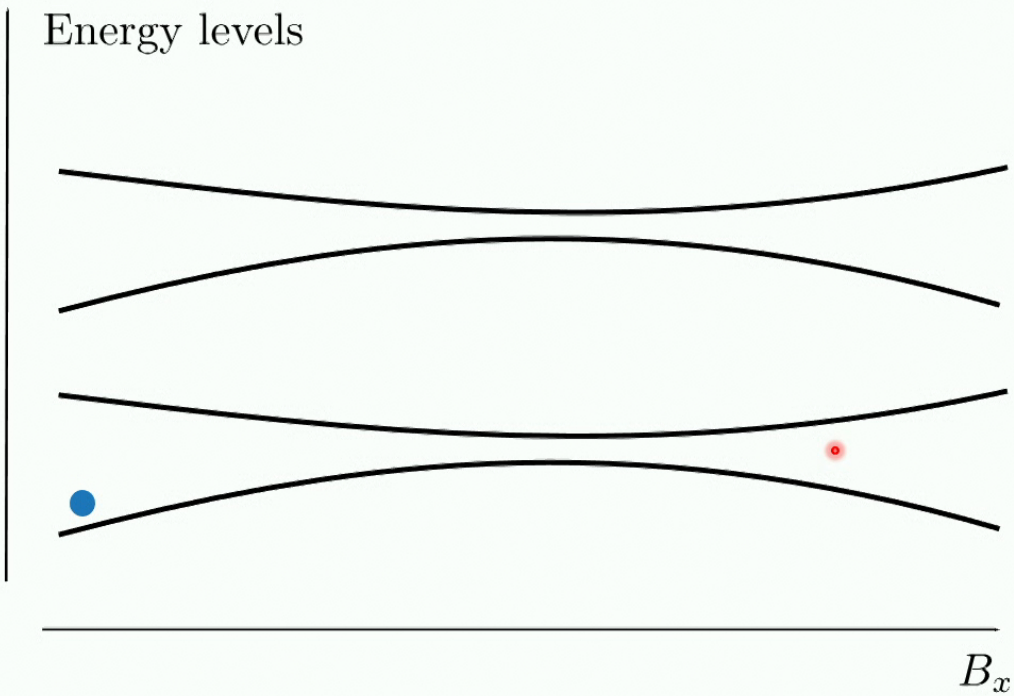
$$\hat{H}(t=0) = -B_{x0}\sum_i \hat{\sigma}_i^x$$

Slide by M. Hibat-Allah

VARIATIONAL QUANTUM ANNEALING

Ground state of problem Hamiltonian is found!

$$H_p = -\sum_{i,j} J_{i,j}\, \sigma_i^z \sigma_j^z$$

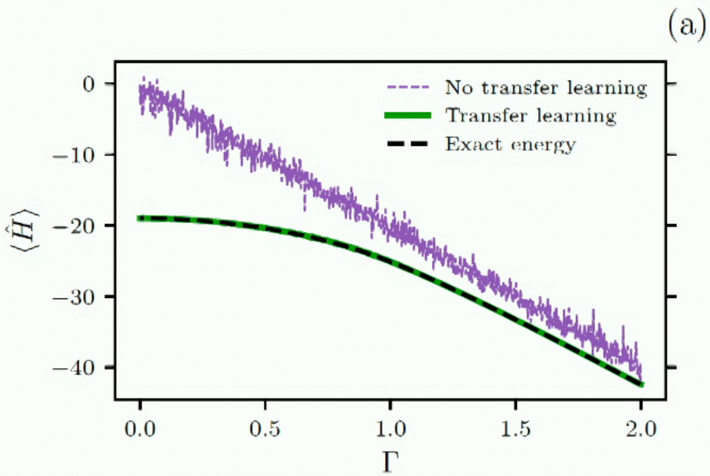


Variational adiabatic theorem

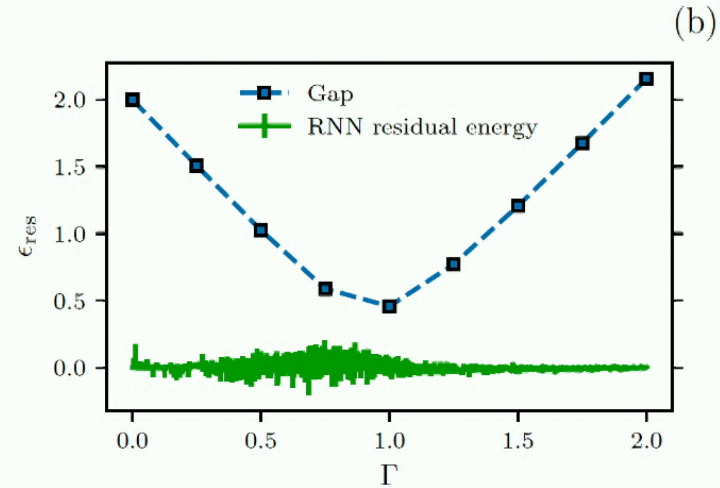
$$\mathcal{O}\left(\frac{poly(N)}{\epsilon \Delta_m}\right) \leq N_{steps} \leq \mathcal{O}\left(\frac{poly(N)}{\epsilon^2 \Delta_m^2}\right)$$

Slide by M. Hibat-Allah

Variational quantum annealing (proof of principle)



N = 20 spins



$$\hat{H}(t) = -J \sum_{i=1}^{N-1} \sigma_i^z \sigma_{i+1}^z - \Gamma(t) \sum_{i=1}^{N} \sigma_i^x$$

$$\mathrm{E} = \frac{\langle \Psi_\lambda(t) | \hat{H}(t) | \Psi_\lambda(t) \rangle}{\langle \Psi_\lambda(t) | \Psi_\lambda(t) \rangle}$$

$$\epsilon_{res} = E - E_{ED}$$

- Positive RNN wavefunction with Tensorized RNN cell used
- Quantum annealing dynamics well captured

Quantum and Classical variational formalism

Variational Monte Carlo

$$E_{var} = \frac{\langle \Psi_\lambda | \hat{H} | \Psi_\lambda \rangle}{\langle \Psi_\lambda | \Psi_\lambda \rangle}$$

$$E_{var} \approx \frac{1}{\mathcal{N}_s} \sum_{\sigma \sim |\Psi_\lambda(\sigma)|^2} E_{loc}(\sigma)$$

$$E_{loc}(\sigma) = \frac{\langle \sigma | \hat{H} | \Psi_\lambda \rangle}{\langle \sigma | \Psi_\lambda \rangle}$$

Variational free energy

$$F_{var} = \langle H_p \rangle_\lambda - T S_{classical}(p_\lambda)$$

$$F_{var} \approx \frac{1}{\mathcal{N}_s} \sum_{\sigma \sim p_\lambda(\sigma)} F_{loc}(\sigma)$$

$$F_{loc}(\sigma) = H_p(\sigma) + T \log(p_\lambda(\sigma))$$

- Variational free energy upper bound to the true free energy
- RNNs used to model both quantum and classical probabilities

Wu, Wang and Zhang, PRL 122 (2019)

VARIATIONAL CLASSICAL ANNEALING

$$F(t) = \langle \sum_{i,j} J_{ij} \hat{\sigma}_i^z \hat{\sigma}_j^z \rangle - (1 - \frac{t}{T}) T_0 S$$

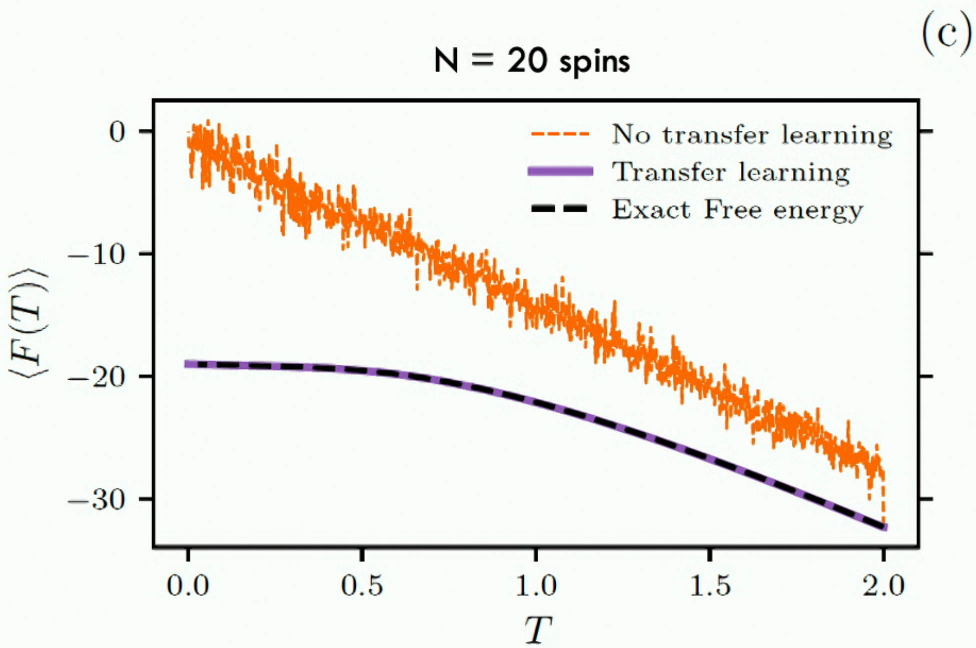
Expectation value over RNN distribution
$P_\lambda(\sigma)$

Von-Neumann entropy

$$S = -\sum_\sigma P_\lambda(\sigma) \log(P_\lambda(\sigma)) = \langle -\log(P_\lambda(\sigma)) \rangle$$

Slide by M. Hibat-Allah

Variational classical annealing (proof of principle)



$$H_p = -J\sum_{i=1}^{N-1}\sigma_i\sigma_{i+1}$$

$$F_{var}(T) = \langle H_p\rangle_\lambda - T(t)S_{classical}(p_\lambda)$$

$$S_{classical}(p_\lambda) = -\sum_{\sigma} p_\lambda(\sigma)\log(p_\lambda(\sigma))$$

Tensorized RNN cell used

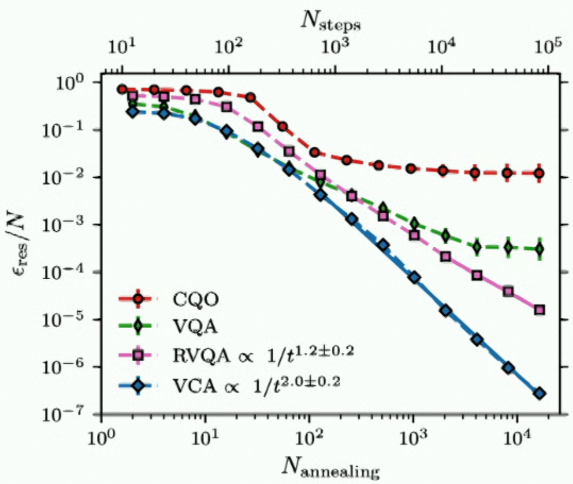
RECAP OF VARIATIONAL NEURAL ANNEALING

Set the initial temperature at $T = T_o$(for VCA) or $\Gamma = \Gamma_o$ (for VQA).

1. Train the system to equilibrate at T (or Γ) via gradient descent steps.
2. Perform annealing via $T \leftarrow T - \delta T$ (or $\Gamma \leftarrow \Gamma - \delta\Gamma$). Use transfer learning of parameters.
3. Repeat 1. and 2. until $T = 0$ (or $\Gamma = 0$).
4. Autoregressive samples give solution to the optimization problem.

Variational annealing on 2D Edwards-Anderson spin glass

$N = 10 \times 10$ spins

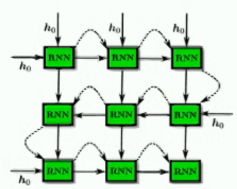


$$H_p = -\sum_{\langle i,j \rangle} J_{i,j}\sigma_i^z \sigma_j^z$$

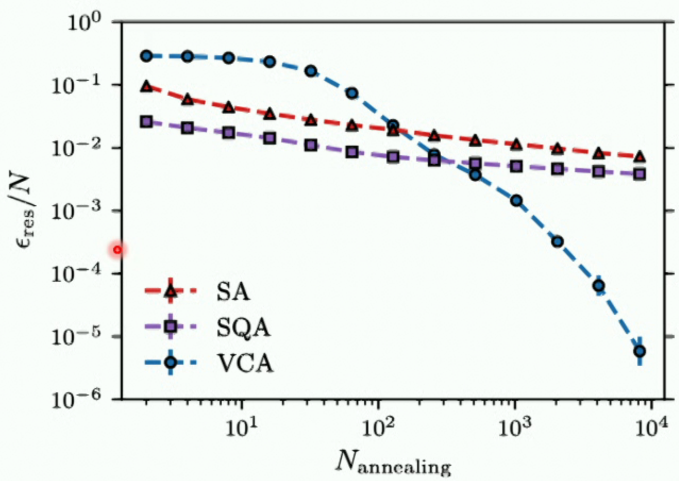
10^6 samples

25 random realizations

Positive Tensorized RNNs



$N = 40 \times 40$ spins

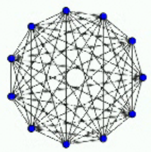


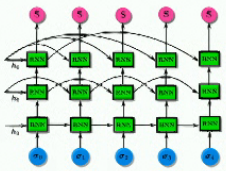
- VCA is superior to VQA.
- Annealing paradigm is more efficient to direct optimization [1]
- VCA is more efficient than SA and SQA [2]

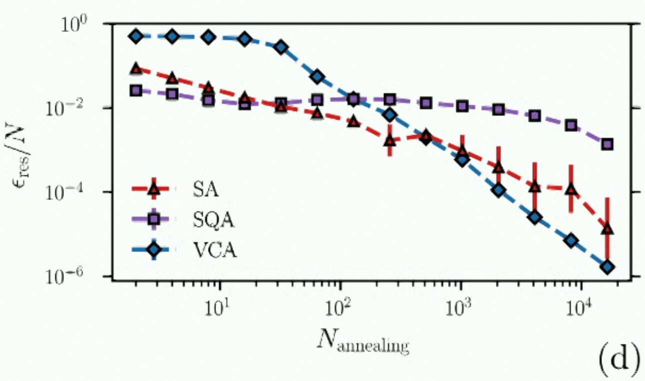
[1] Gomes et al (2019), Sinchenko and Bazhanov (2019), Zhao et al. (2020)

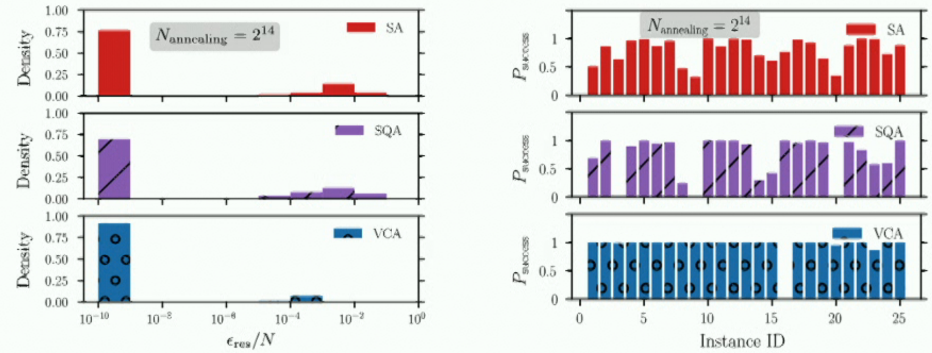
[2] Santoro, et al., Science 295, 2427 (2002)

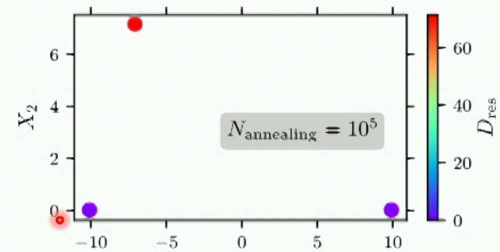
VCA on Sherrington-Kirkpatrick model











$N = 100$ spins

10^6 samples

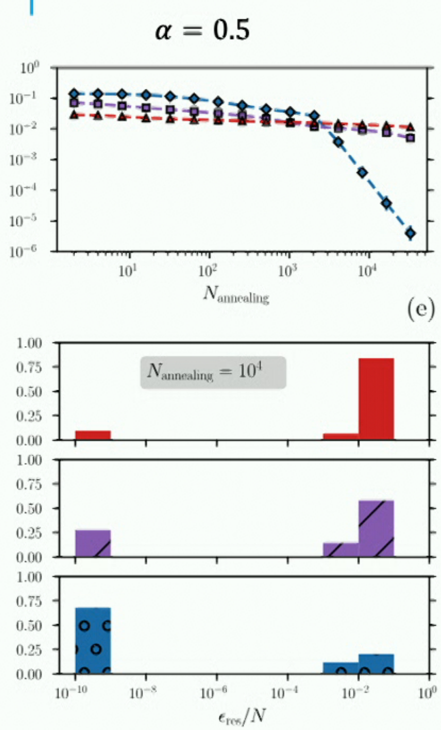
25 random realizations

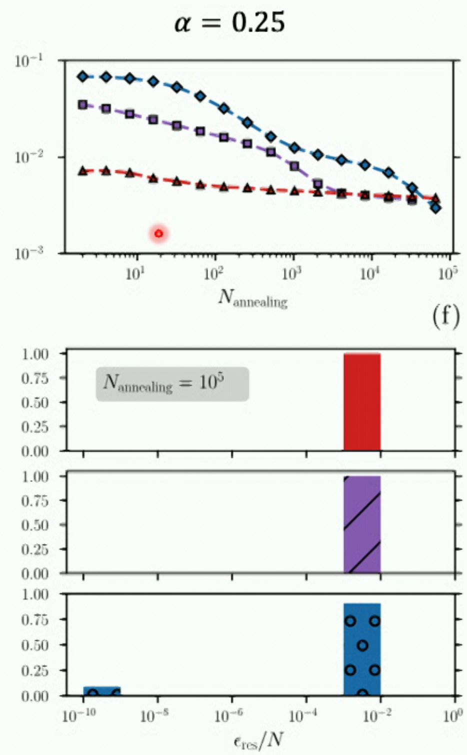
25,000 data points

Positive Tensorized RNNs

- VCA is superior to both SA and SQA
- Autoregressive sampling and entropy seems to give that advantage

VCA on Wishart planted ensemble model [1]





$$H^{\alpha}_{WPE} = -\frac{1}{2}\sum_{i \neq j} J^{\alpha}_{ij}\sigma^z_i \sigma^z_j$$

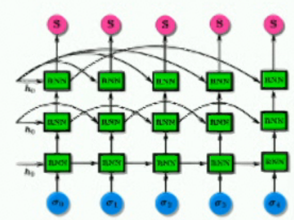
$N = 32$ spins

10^6 samples

25 random realizations

25,000 data points

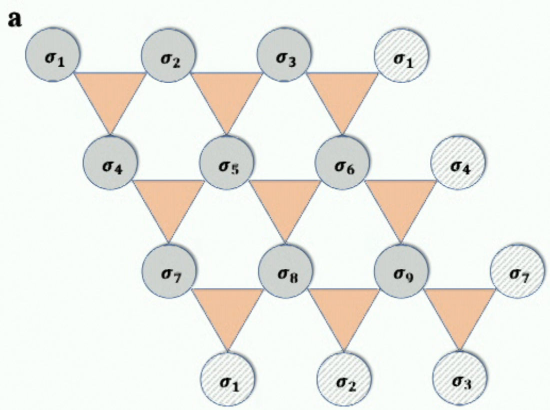
Positive Tensorized RNNs



VCA competes with both SA and SQA

[1] Hamze et al., PRE 101 (2020)

The Newman-Moore (NM) model

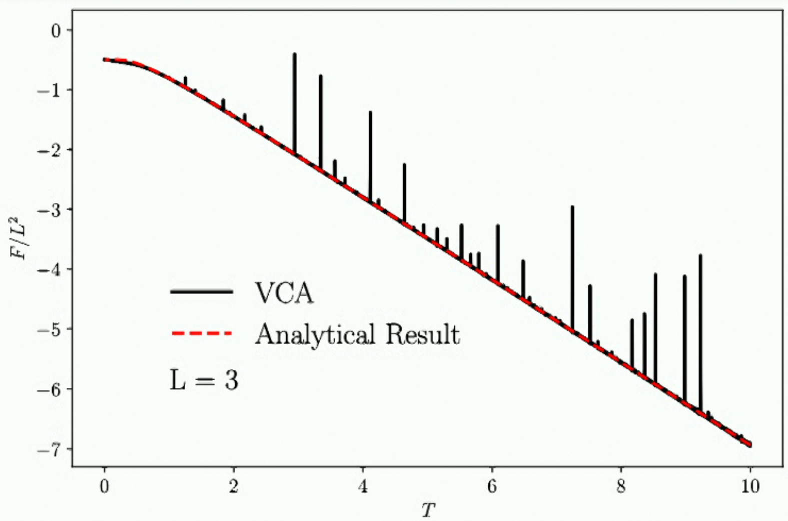


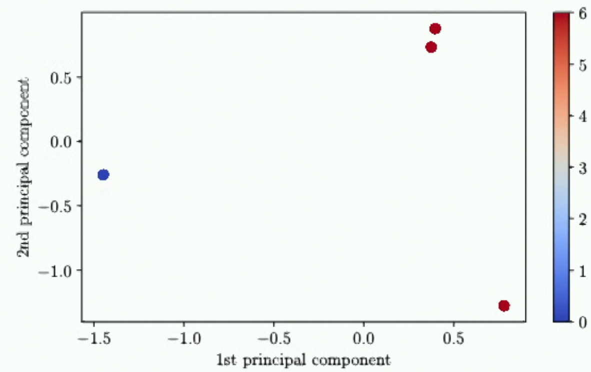
- The NM exhibits glassy dynamics [1]
- Signatures of 1st order quantum phase transition [2]
- Degenerate ground states and immobile excitations [1, 3]

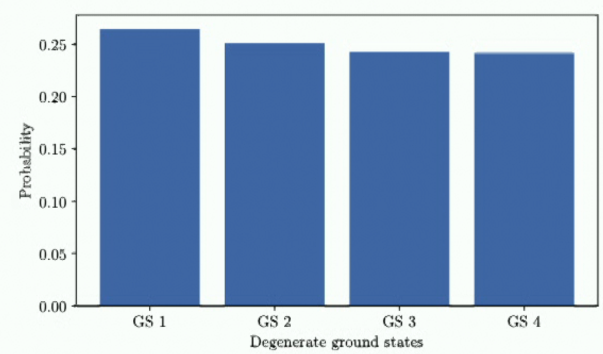
$$H_p = \frac{J}{2}\sum_{i,j,k \text{ in } \nabla} \sigma_i^z \sigma_j^z \sigma_k^z$$

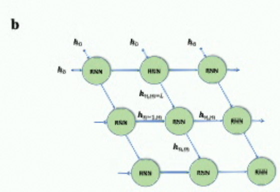
[1] Newman and Moore (1999), Garrahan and Newman (2000)
[2] Vasiloiu et al, Phys. Rev. E **101**, 042115 (2020)
[3] Zhou et al, arXiv:2105.05851

Neural Annealing on the Newman-Moore model



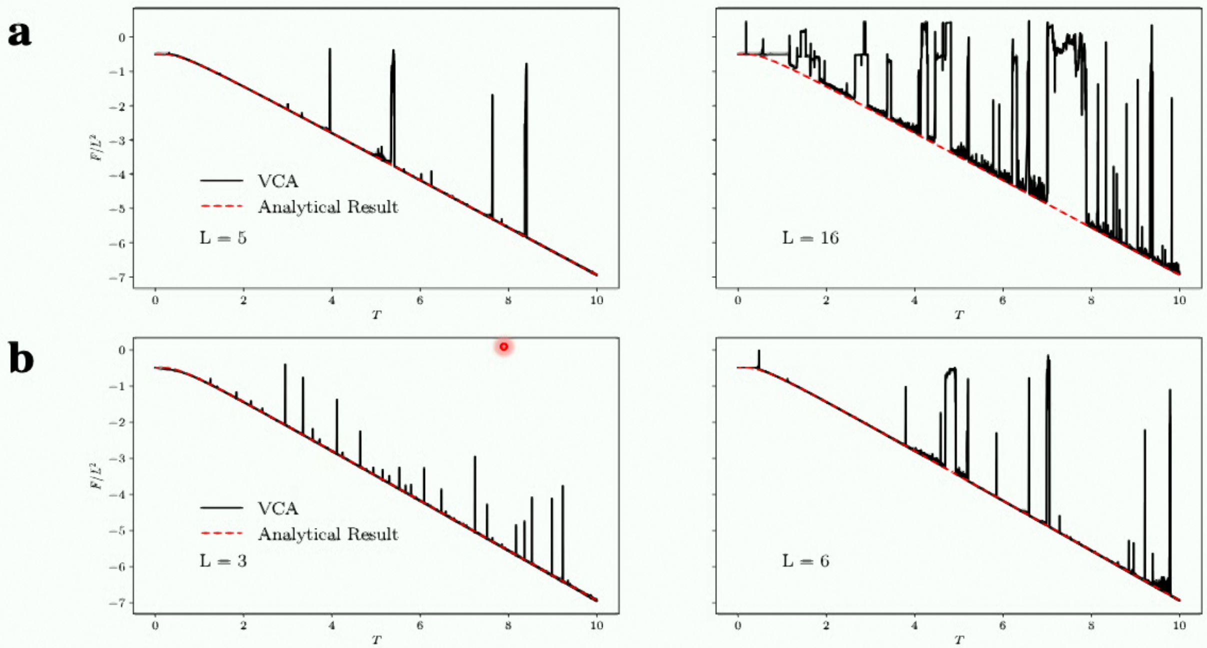






$N = 3{\times}3$ spins
$\sim 10^4$ samples
10,000 annealing steps
Vanilla RNNs

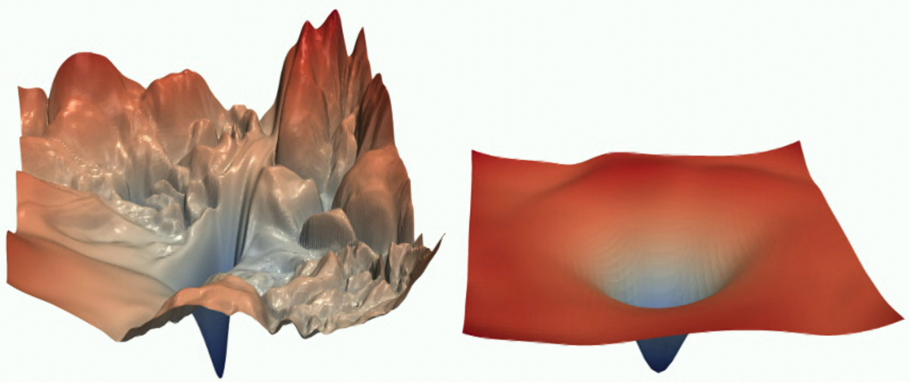
VCA on the Newman-Moore model



- Unstable annealing dynamics
- Strong mode-collapse

Visualizing the Loss Landscape of Neural Nets

Hao Li[1], Zheng Xu[1], Gavin Taylor[2], Christoph Studer[3], Tom Goldstein[1]
[1]University of Maryland, College Park [2]United States Naval Academy [3]Cornell University
{haoli,xuzh,tomg}@cs.umd.edu, taylor@usna.edu, studer@cornell.edu

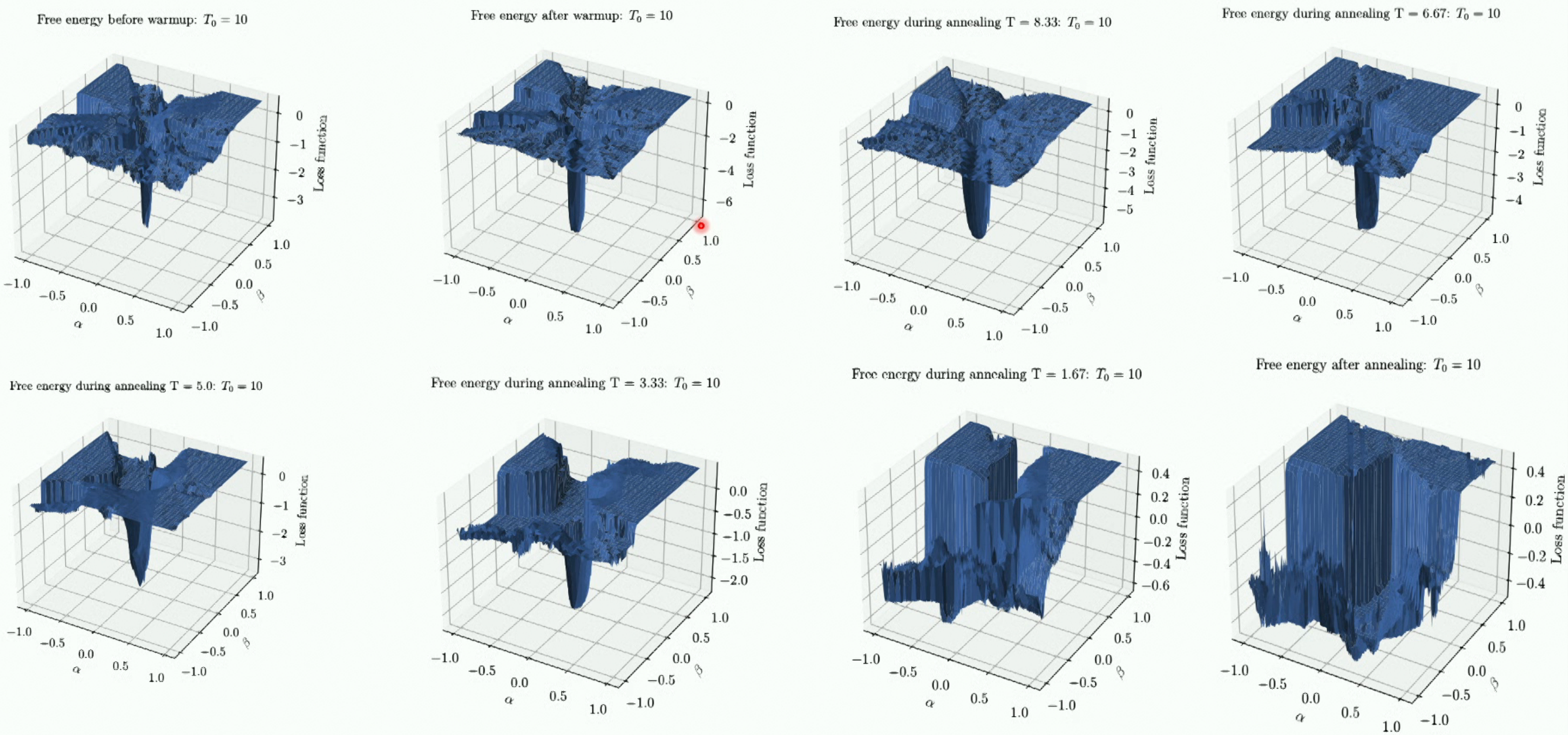


(a) without skip connections (b) with skip connections

Figure 1: The loss surfaces of ResNet-56 with/without skip connections. The proposed filter normalization scheme is used to enable comparisons of sharpness/flatness between the two figures.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

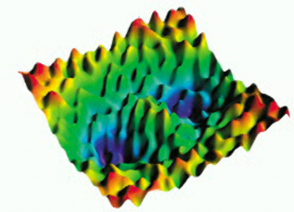
Loss landscapes visualization the Newman-Moore model



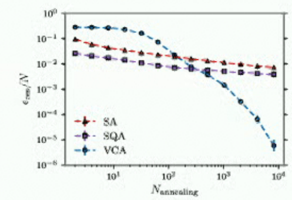
$N = 16 \times 16$ **spins**

Conclusion and perspectives

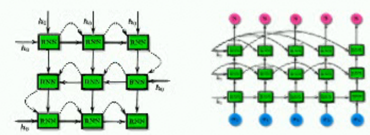
Annealing search in variational landscape seems to outperform search in the configurational landscape



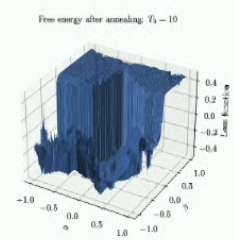
VCA is more efficient than VQA, SA and SQA. Good candidate for real world optimization problems



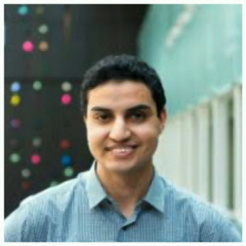
Rigorous relationship between RNN architecture and underlying Hamiltonian



Trainability issues elude to conservation of computational complexity - visualizing landscape may help design better ansatzes.



THANK YOU!



Mohamed Hibat-Allah
(Vector Institute/UW)

Roeland Wiersema
(Vector Institute/UW)

Roger Melko
(Perimeter Institute/UW)

Juan Carrasquilla
(Vector Institute/UW)

Stewart Morawetz
(UW/Boston University)