Weaving together machine learning, theoretical physics and neuroscience through mathematics

Surya Ganguli

Dept. of Applied Physics,
Neurobiology,
and Electrical Engineering

Stanford University

Funding:

Bio-X Neuroventures
Burroughs Wellcome
Genentech Foundation
James S. McDonnell Foundation
McKnight Foundation
National Science Foundation

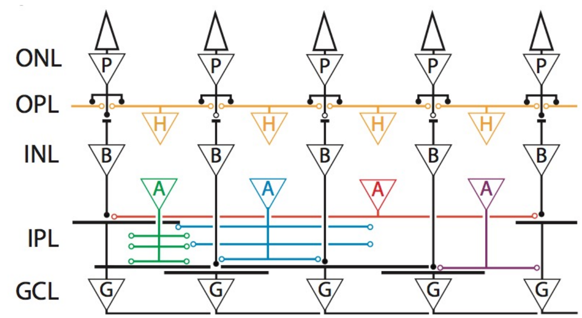
NIH
Office of Naval Research
Simons Foundation
Sloan Foundation
Swartz Foundation
Stanford Terman Award

http://ganguli-gang.stanford.edu

Twitter: @SuryaGanguli

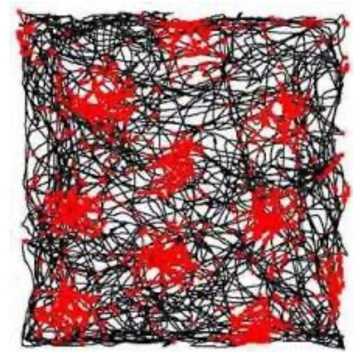
Neural circuits and behavior: theory, computation and experiment

with Baccus lab: inferring hidden circuits in the retina
w/ Niru Maheswaranathan and Lane McIntosh

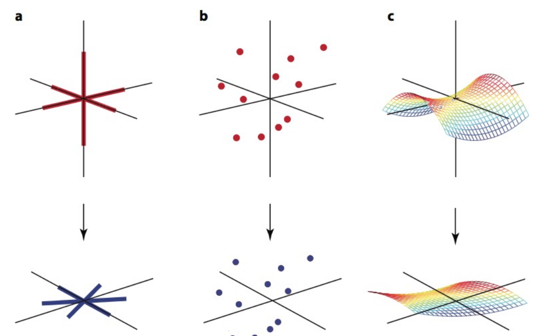


with Clandinin lab: unraveling the computations underlying fly motion vision from whole brain optical imaging
w/ Jonathan Leong, Ben Poole and Jennifer Esch

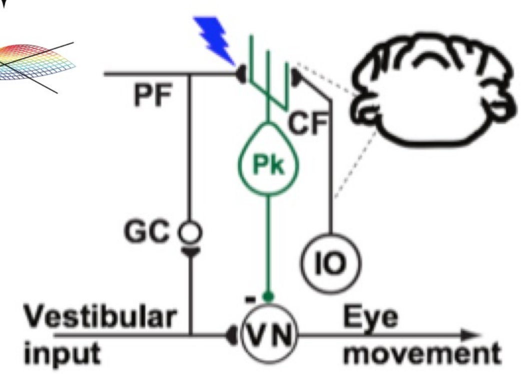
with the Giocomo lab: understanding the internal representations of space in the mouse entorhinal cortex
w/ Kiah Hardcastle and Sam Ocko



with the Shenoy lab: a theory of neural dimensionality, dynamics and measurement
w/ Peiran Gao, Eric Trautmann



with the Raymond lab: theories of how enhanced plasticity can either enhance or impair learning depending on experience
w/ Subhaniel Lahiri, Barbara Vu, Grace Zhao



Motivations for alliances between theoretical neuroscience machine learning and physics

- What does it mean to understand the brain (or a neural circuit?)
- We understand how the connectivity and dynamics of a neural circuit gives rise to behavior.
- And also how neural activity and synaptic learning rules conspire to self-organize useful connectivity that subserves behavior.
- The field of machine learning has generated a plethora of learned neural networks that accomplish interesting functions.
- We know their connectivity, dynamics, learning rule, and developmental experience, *yet*, we do not have a meaningful understanding of how they learn and work!
- Can we exploit novel physics to implement neuromorphic computations at even lower power and higher speed?

On simplicity and complexity in the brave new world of large scale neuroscience, Peiran Gao and S. Ganguli, Curr. Op. in Neurobiology, 2015.

References on deep learning theory: http://ganguli-gang.stanford.edu

- M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.
- M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.
- A. Saxe, J. McClelland, S. Ganguli, Learning hierarchical category structure in deep neural networks, Proc. of the 35th Cognitive Science Society, pp. 1271-1276, 2013.
- A. Saxe, J. McClelland, S. Ganguli, Exact solutions to the nonlinear dynamics of learning in deep neural networks, ICLR 2014.
- Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, Y. Bengio, Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, NIPS 2014.
- B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, Exponential expressivity in deep neural networks through transient chaos, NIPS 2016.
- S. Schoenholz, J. Gilmer, S. Ganguli, and J. Sohl-Dickstein, Deep information propagation, ICLR 2017.
- S. Lahiri, J. Sohl-Dickstein and S. Ganguli, A universal tradeoff between energy speed and accuracy in physical communication, arxiv 1603.07758
- A memory frontier for complex synapses, S. Lahiri and S. Ganguli, NIPS 2013.
- Continual learning through synaptic intelligence, F. Zenke, B. Poole, S. Ganguli, ICML 2017.
- Modelling arbitrary probability distributions using non-equilibrium thermodynamics, J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, ICML 2015.
- Deep Knowledge Tracing, C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, J. Sohl-Dickstein, NIPS 2015.
- Deep learning models of the retinal response to natural scenes, L. McIntosh, N. Maheswaranathan, S. Ganguli, S. Baccus, NIPS 2016.
- Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice, J. Pennington, S. Schloenholz, and S. Ganguli, NIPS 2017.
- Variational walkback: learning a transition operator as a recurrent stochastic neural net, A. Goyal, N.R. Ke, S. Ganguli, Y. Bengio, NIPS 2017.
- The emergence of spectral universality in deep networks, J. Pennington, S. Schloenholz, and S. Ganguli, AISTATS 2018.
- The statistical mechanics of low-rank tensor decompositions, J. Kadmon, S. Ganguli, NIPS 2018.
- A mathematical theory of semantic development in deep neural networks, A. Saxe, J. McClelland, S. Ganguli, PNAS 2019.
- An analytic theory of generalization dynamics and transfer learning in deep linear networks, ICLR 2019.
- The emergence of multiple retinal cell-types through the efficient coding of natural movies, S. Ocko, J. Lindsey, S. Ganguli, S. Deny, NeurIPS 2018.
- A computational origin for grid cells, B. Sorscher, G. Mel, S.Ganguli, S. Ocko. Neurips 2019
- Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev,
- Enhancing associative memory recall and storage capacity using confocal cavity QED, Physical Review X, 2021.
- Y. Yamamoto, T. Leleu, S. Ganguli, H. Mabuchi, Coherent Ising Machines: quantum optics and neural network perspectives, Applied Physics Letters, 2020.

Tools: Non-equilibrium statistical mechanics
Dynamical mean field theory
Statistical mechanics of random landscapes
Riemannian geometry
Random matrix theory
Free probability theory

Statistical mechanics of deep learning

Yasaman Bahri,[1] Jonathan Kadmon,[2] Jeffrey Pennington,[1] Sam Schoenholz,[1], Jascha Sohl-Dickstein,[1] and Surya Ganguli[1,2]

[1]Google Brain, Google Inc., Mountain View, CA 94043
[2]Department of Applied Physics, Stanford University, Stanford, CA 94035

Contents

Outline

What is the optimal way to do regression in high dimensions?

M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.

M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.

Deriving the first steps of vision from the statistics of natural movies

Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, The emergence of multiple retinal cell-types through the efficient coding of natural movies. Neural Information Processing Systems 2018.

Quantum neuromorphic computing with atoms and photons

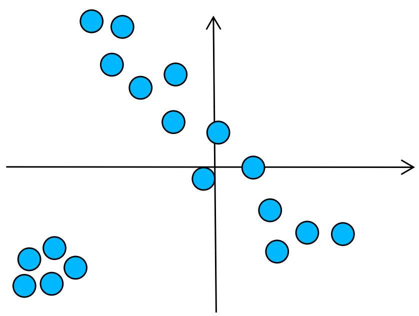
Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev, Enhancing associative memory recall and storage capacity using confocal cavity QED, Physical Review X 2021.

Yoshi Yamamoto, Timothee Leleu, S. Ganguli, Hideo Mabuchi, Coherent Ising Machines: quantum optics and neural network perspectives, Applied Physics Letters, 2020.

Statistical mechanics of high dimensional data analysis

P = dimensionality of data N = number of data points $\alpha = N / P$

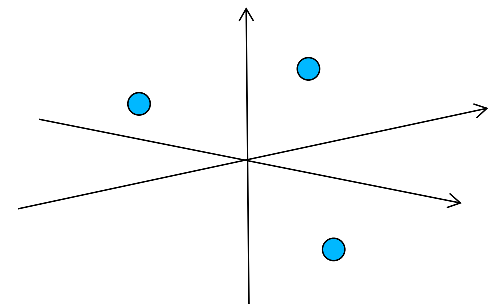
Classical Statistics



$P \sim O(1)$
$N \to \infty$
$\alpha \to \infty$

Modern Statistics

$P \to \infty$
$N \to \infty$
$\alpha \sim O(1)$



Machine Learning and Data Analysis
Learn statistical parameters by maximizing log likelihood of data given parameters.

Statistical Physics of Quenched Disorder
Energy = - log Prob (data l parameters)
Data = quenched disorder
Parameters = thermal degrees of freedom

Statistical mechanics of compressed sensing, S. Ganguli and H. Sompolinsky, PRL 2010.

Short-term memory in neuronal networks through dynamical compressed sensing, NIPS 2010.

Compressed sensing, sparsity and dimensionality in neuronal information processing and data analysis, S. Ganguli and H. Sompolinsky, Annual Reviews of Neuroscience, 2012

Statistical mechanics of high dimensional inference, M. Advani and S. Ganguli, Physical Review X, 2016.

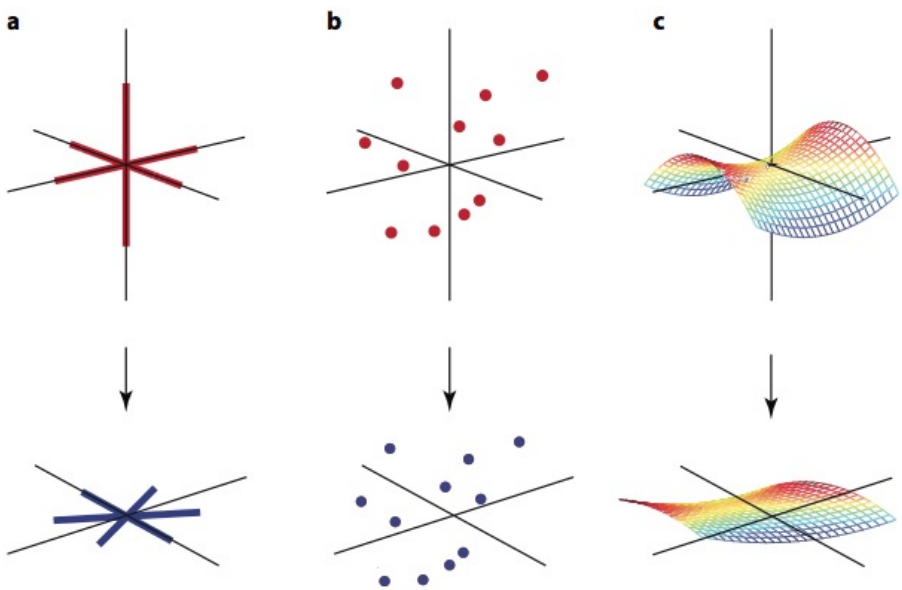
An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.

Random projections of random manifolds, S. Lahiri, P. Gao, S. Ganguli, http://arxiv.org/abs/1607.04331, under review at JMLR.

Statistical mechanics of complex neural systems and high dimensional data

Madhu Advani, Subhaneil Lahiri and Surya Ganguli
Hide affiliations
msadvani@stanford.edu sulahiri@stanford.edu sganguli@stanford.edu
Department of Applied Physics, Stanford University, Stanford, CA, USA

Madhu Advani *et al J. Stat. Mech.* (2013) P03014. doi:10.1088/1742-5468/2013/03/P03014
Received 9 October 2012, accepted for publication 14 January 2013. Published 12 March 20



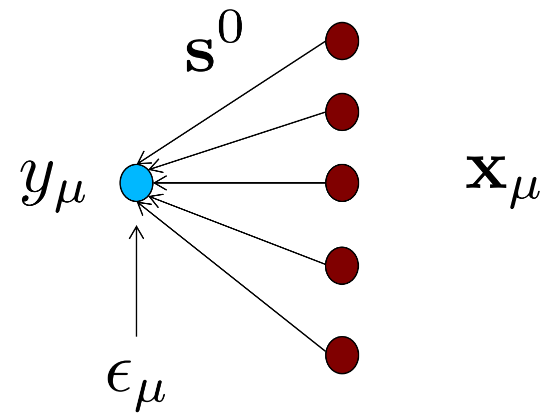
Optimal inference in high dimensions

Estimation algorithm

ρ = loss function
σ = regularizer
q_s = L_2 estimation error

Generative model and measurements

P dim signal $s^0 \sim P_s$
N measurements with noise $\varepsilon \sim P_\varepsilon$
α = N/P = measurement density



$$\hat{\mathbf{s}} \longleftarrow \boxed{\rho, \sigma} \longleftarrow \{\mathbf{x}_\mu, y_\mu\}$$

$$\mathbf{s}^0 \longrightarrow \mathbf{x}_\mu \longrightarrow y_\mu = \mathbf{x}_\mu \cdot \mathbf{s}^0 + \epsilon_\mu$$

$$\underbrace{\mathbf{x}_\mu \longrightarrow y_\mu}_{\downarrow\ \hat{\mathbf{s}}}$$

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \sum_\mu \rho(\mathbf{y}_\mu - \mathbf{x}_\mu \cdot \mathbf{s}) + \sum_j \sigma(s_j)$$

$$\frac{1}{P} \sum_j (\hat{s}_j - s_j^0)^2 = q_s(\alpha, \rho, \sigma, P_\epsilon, P_s)$$

Question: For a given: signal distribution P_s , noise distribution P_ε , and measurement density α,

What is the best: loss function ρ and regularizer σ?

Optimal inference in high dimensions

$$\mathbf{s}^0 \longrightarrow \underbrace{\mathbf{x}_\mu \longrightarrow y_\mu}_{\downarrow\ \hat{\mathbf{s}}} = \mathbf{x}_\mu \cdot \mathbf{s}^0 + \epsilon_\mu$$

Generative model and measurements

P dim signal $\mathbf{s}^0 \sim P_s$
N measurements with noise $\varepsilon \sim P_\varepsilon$
α = N/P = measurement density

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \sum_\mu \rho(\mathbf{y}_\mu - \mathbf{x}_\mu \cdot \mathbf{s}) + \sum_j \sigma(s_j)$$

Estimation algorithm

ρ = loss function
σ = regularizer
q_s = L_2 estimation error

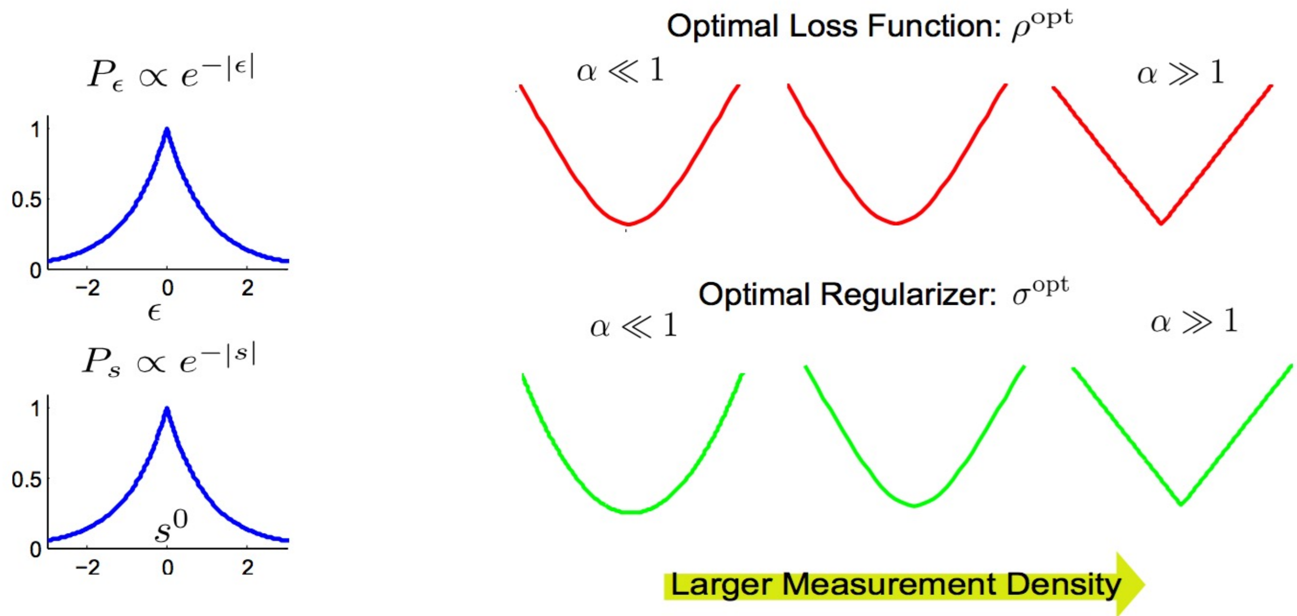
$$\frac{1}{P}\sum_j (\hat{s}_j - s_j^0)^2 = q_s(\alpha, \rho, \sigma, P_\epsilon, P_s)$$

Example algorithms

Least squares:	$\rho(\epsilon) = \epsilon^2$	$\sigma(s) = 0$
Maximum likelihood:	$\rho(\epsilon) = -\log P_\epsilon(\epsilon)$	$\sigma(s) = 0$
Ridge regression:	$\rho(\epsilon) = \epsilon^2$	$\sigma(s) = s^2$
LASSO:	$\rho(\epsilon) = \epsilon^2$	$\sigma(s) = \lambda_1 \lvert s \rvert$
Elastic Net:	$\rho(\epsilon) = \epsilon^2$	$\sigma(s) = \lambda_1 \lvert s \rvert + \lambda_2 s^2$
MAP:	$\rho(\epsilon) = -\log P_\epsilon(\epsilon)$	$\sigma(s) = -\log P_s(s)$

Optimal inference in high dimensions

Question: For a given signal distribution P_s , noise distribution P_ε , and measurement density α, what is the best loss function ρ and regularizer σ?



For log-concave signal and noise: the optimal loss and regularizer are nonlinearly smoothed versions of MAP where the smoothing increases as the measurement density decreases.

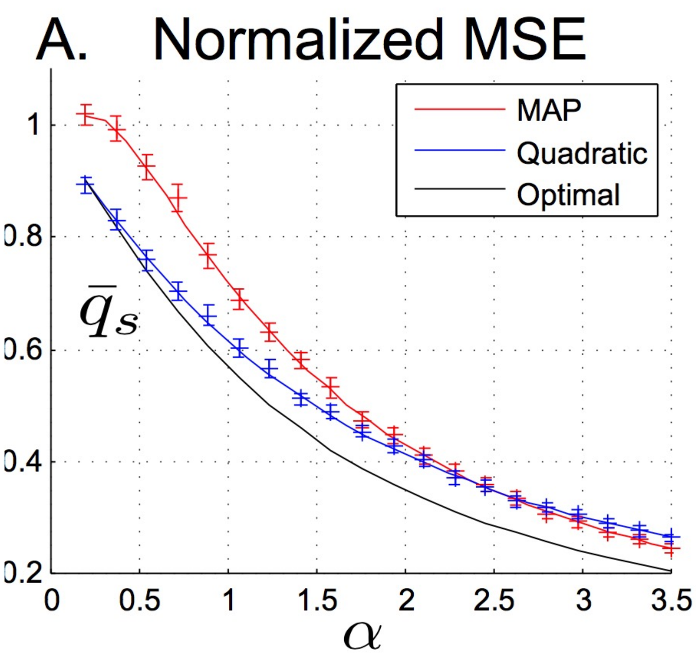
MAP is optimal at high measurement density.

Ridge regression is optimal at low measurement density independent of signal and noise!

No inference algorithm can out-perform our optimal algorithm!

Optimal inference in high dimensions

Question: For a given signal distribution P_s , noise distribution P_ε , and measurement density α, what is the best loss function ρ and regularizer σ?



M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.

M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.

For log-concave signal and noise: the optimal loss and regularizer are nonlinearly smoothed versions of MAP where the smoothing increases as the measurement density decreases.

MAP is optimal at high measurement density.

Ridge regression is optimal at low measurement density independent of signal and noise!

No inference algorithm can out-perform our optimal algorithm!

Outline

What is the optimal way to do regression in high dimensions?

M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.

M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.

Deriving the first steps of vision from the statistics of natural movies

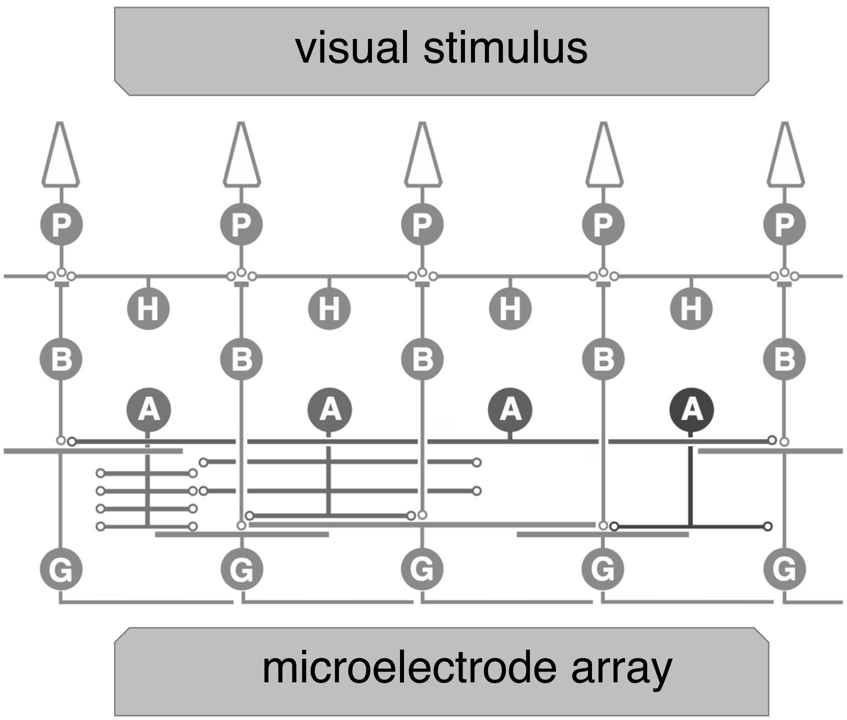
Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, The emergence of multiple retinal cell-types through the efficient coding of natural movies. Neural Information Processing Systems 2018.

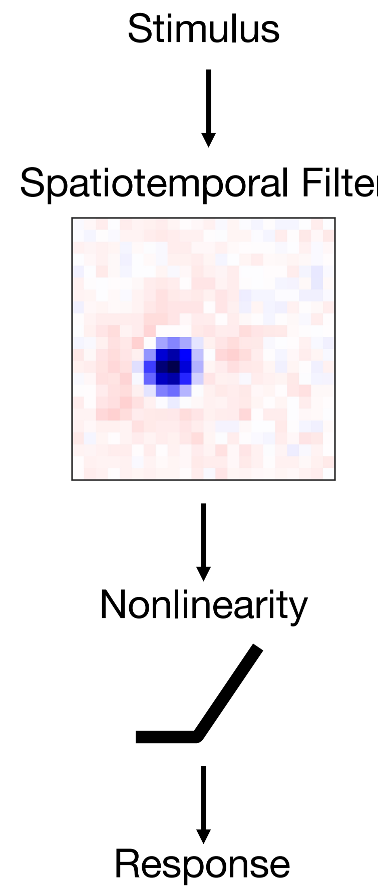
Quantum neuromorphic computing with atoms and photons

Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev, Enhancing associative memory recall and storage capacity using confocal cavity QED, Physical Review X 2021.

Yoshi Yamamoto, Timothee Leleu, S. Ganguli, Hideo Mabuchi, Coherent Ising Machines: quantum optics and neural network perspectives, Applied Physics Letters, 2020.

The retina 101





Chichilnisky 2001
Baccus and Meister 2002
Pillow et al 2005, 2008

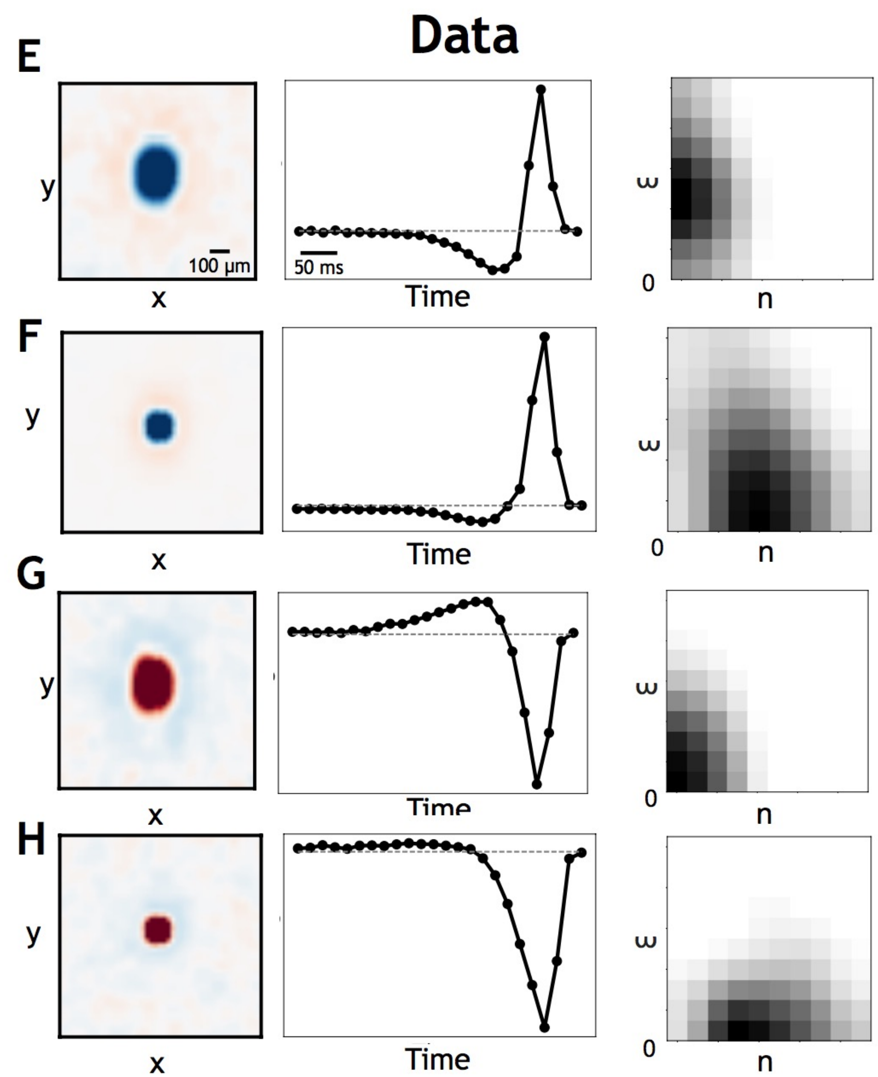
Emergence of multiple retinal cell types through the efficient coding of natural movies

Off
Parasol

Off
Midget

On
Parasol

On
Midget



Data from E.J. Chichilinsky's lab.

Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, Neural Information Processing Systems 2018

Emergence of multiple retinal cell types through the efficient coding of natural movies: an alternate normative approach to retinal function

- A major experimental effort in experimental neuroscience is going towards cataloguing all cell types in the brain.

- E.g.: Allen Brain Institute
  Chan-Zuckerberg cell type atlas
  US Brain Initiative

However, there is no theory for why we have so many cell-types.

We attempt to address this in the retina.

Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, Neural Information Processing Systems 2018

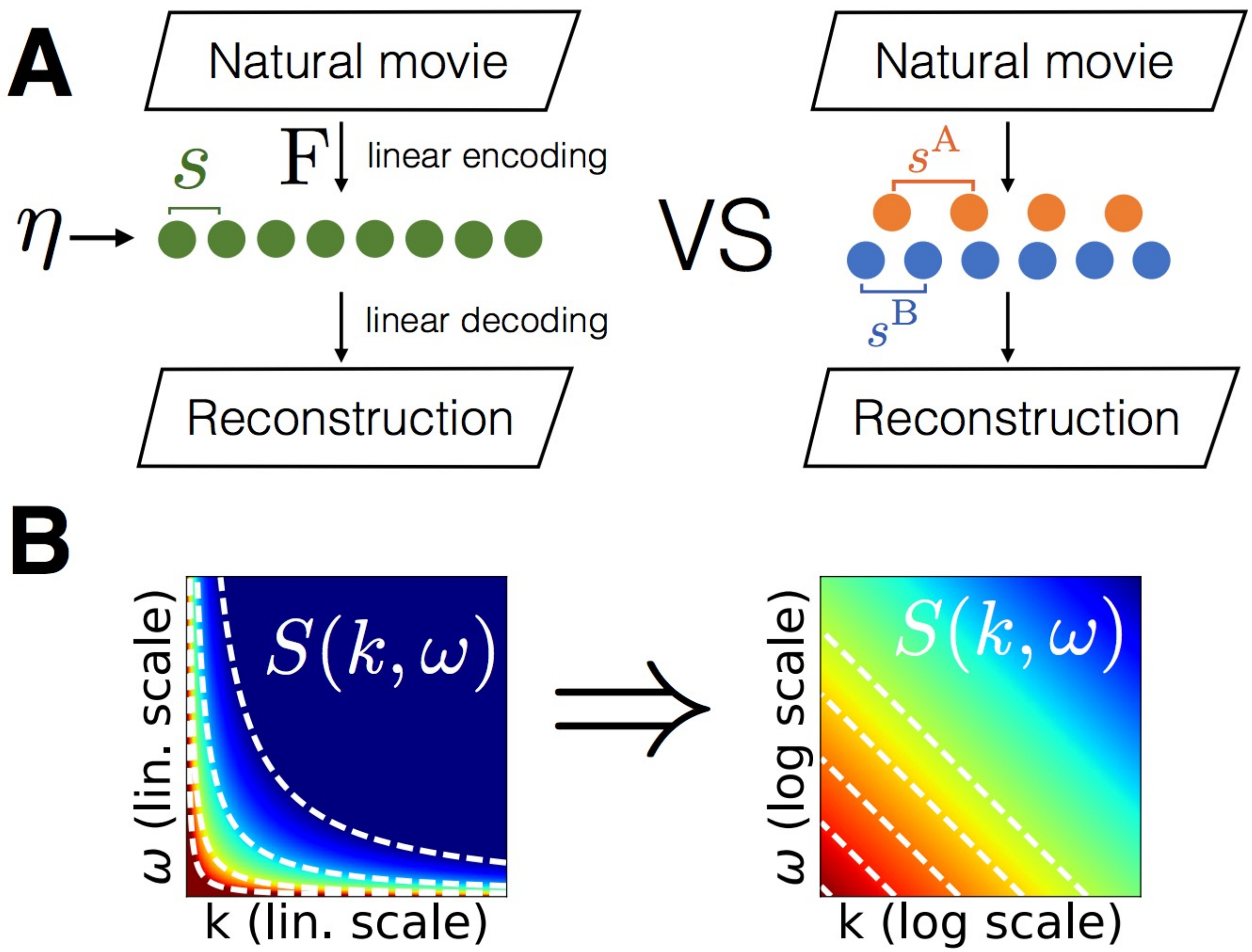
Emergence of multiple retinal cell types through the efficient coding of natural movies

70% of all cell-types in the primate retina are comprised of:

	Midget Cells	Parasol cells
Spatial RF Size	Small	Large
Temporal filtering speed	Slow	Fast
Number density	High	Low
Gain / sensitivity	Low	High

Emergence of multiple retinal cell types through the efficient coding of natural movies

Theoretical principle: accurately encode natural movies using low firing rates via a convolutional neural network

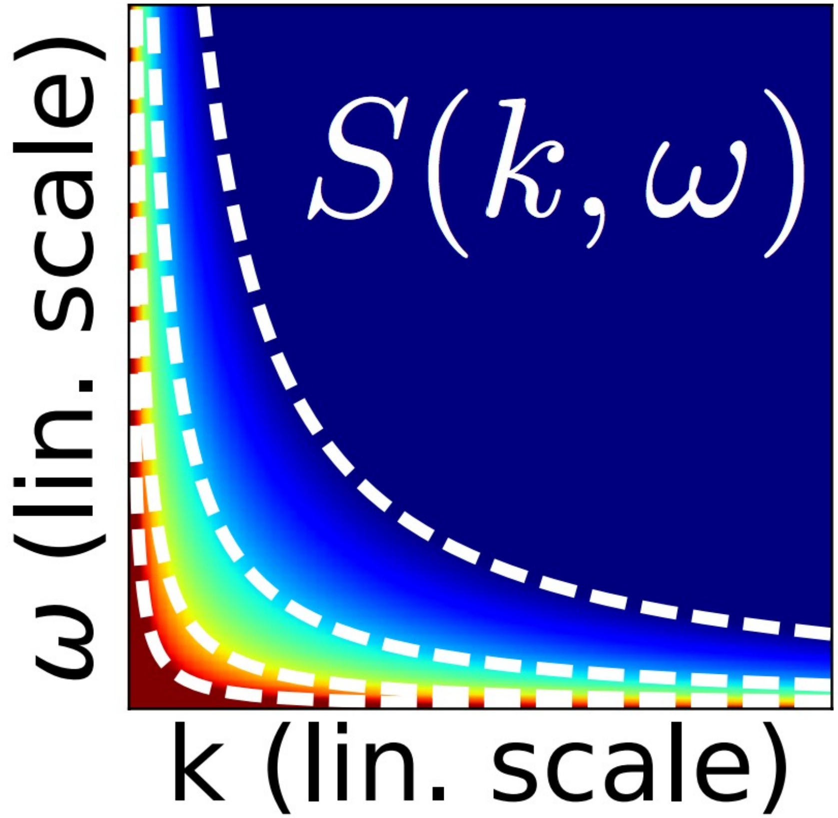


Emergence of multiple retinal cell types through the efficient coding of natural movies

Key idea: two cell types can more accurately encode natural movies with lower firing rates by specializing to encode different parts of the natural movie power spectrum!

Parasol cells

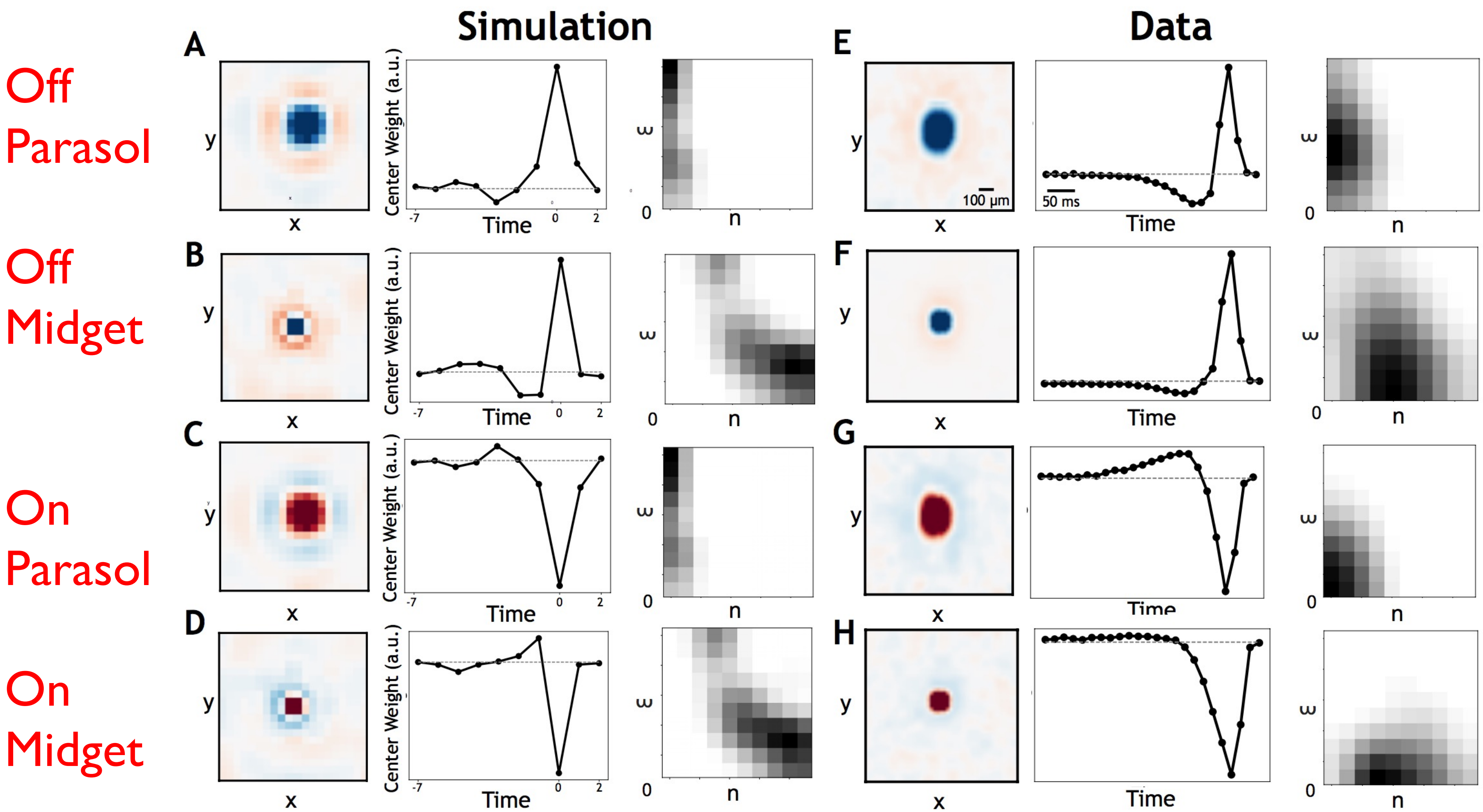
Low spatial freq
High temporal freq
few cells
with high firing



Midget Cells

high spatial freq
low temporal freq
many cells
with low firing

Emergence of multiple retinal cell types through the efficient coding of natural movies



Data from E.J. Chichilinsky's lab.

Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, Neural Information Processing Systems 2018

Outline

What is the optimal way to do regression in high dimensions?

M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.

M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.

Deriving the first steps of vision from the statistics of natural movies

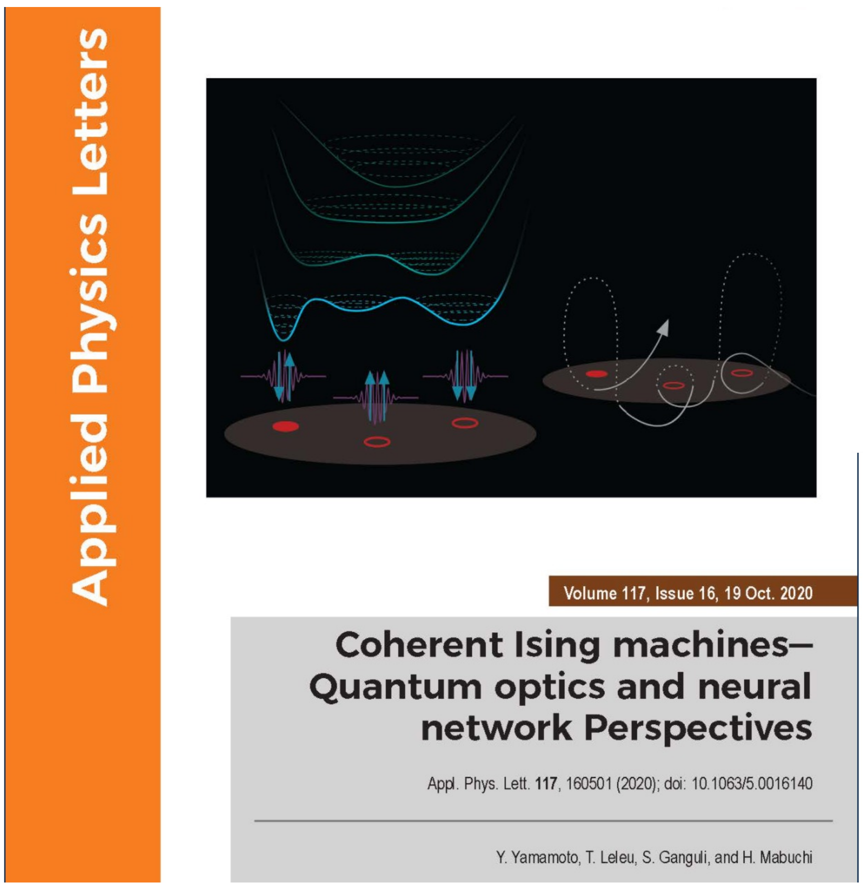
Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, The emergence of multiple retinal cell-types through the efficient coding of natural movies. Neural Information Processing Systems 2018.

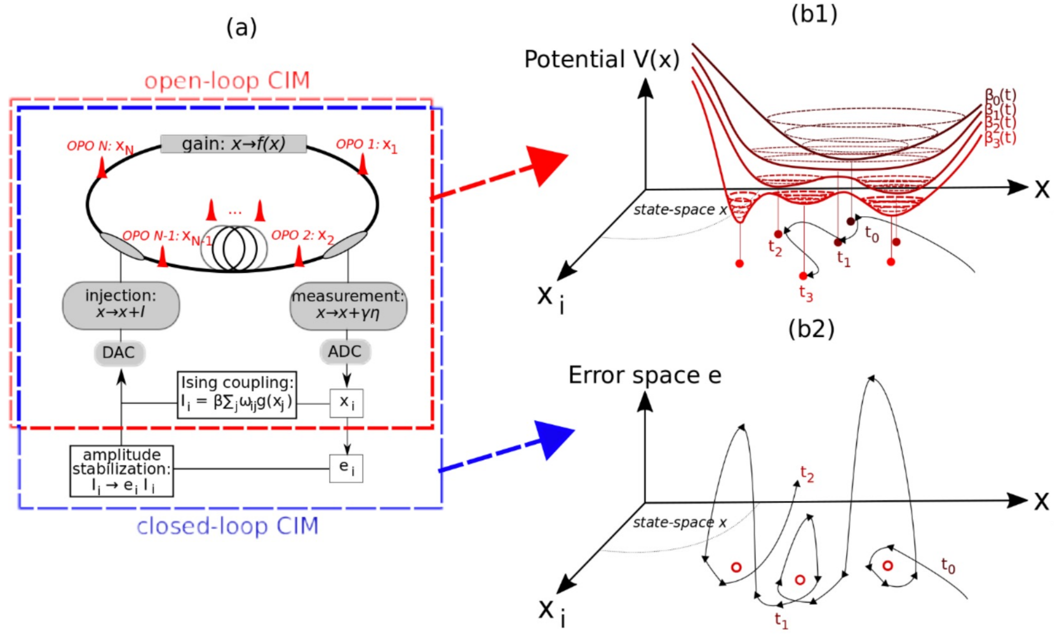
Quantum neuromorphic computing with atoms and photons

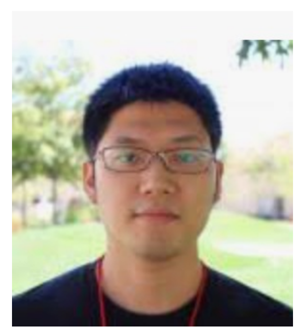
Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev, Enhancing associative memory recall and storage capacity using confocal cavity QED, Physical Review X 2021.

Yoshi Yamamoto, Timothee Leleu, S. Ganguli, Hideo Mabuchi, Coherent Ising Machines: quantum optics and neural network perspectives, Appli
Physics Letters, 2020.

Statistical mechanics of high dimensional optimization landscapes in the CIM







Joint work with: Atushi Yamamura Hideo Mabuchi

Three energy functions of interest

$$E(x) = -\sum_{ij} J_{ij} x_i x_j \qquad x_i \in \{\pm 1\}$$

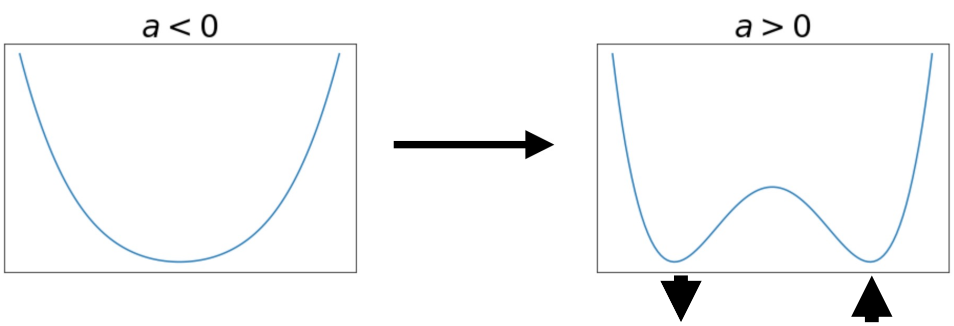
Ising energy
Encodes combinatorial problem of interest

$$E(x) = -\sum_{ij} J_{ij} x_i x_j \qquad x_i \in \mathbb{R} \quad x^T x = 1$$

Spectral relaxation; solution: *maximal* eigenvector of J

$$E(x) = \frac{1}{4}\sum_i x_i^4 - \frac{a}{2}\sum_i x_i^2 - \sum_{ij} J_{ij} x_i x_j \qquad x_i \in \mathbb{R}$$

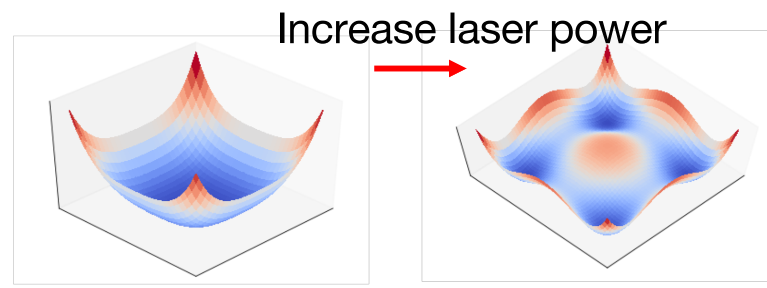
Classical version of CIM energy
a = laser pump power parameter
x = x quadrature of an OPO



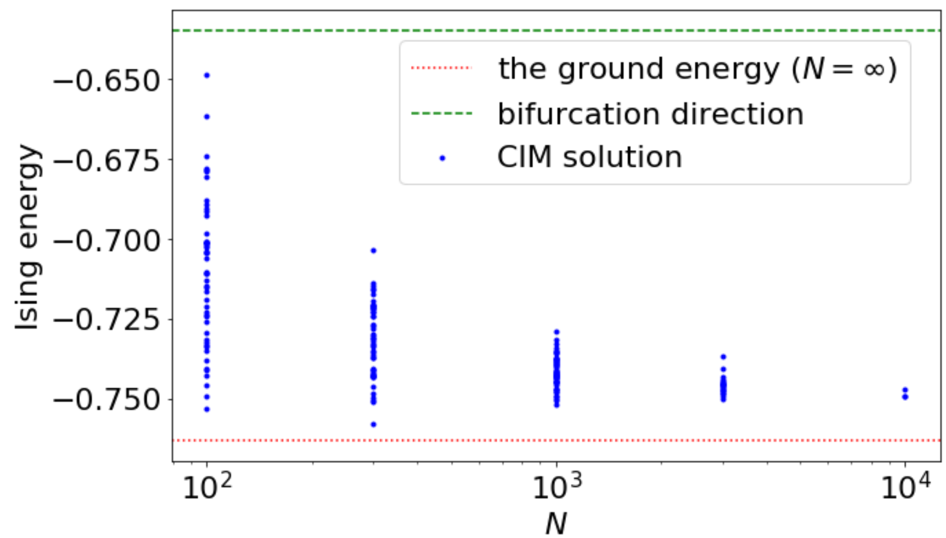
Approaching the SK spin glass solution through adiabatic evolution of the CIM

Sherrington Kirkpatrick (SK) spin glass: J_{ij} drawn as a Gaussian: $\mathcal{N}(0, 1/N)$

A result: adiabatic evolution of CIM to find SK low energy state



1) Start from the origin at a small value of laser pump power
2) Increase the pump power a small amount
3) Minimize energy via gradient descent
4) Repeat

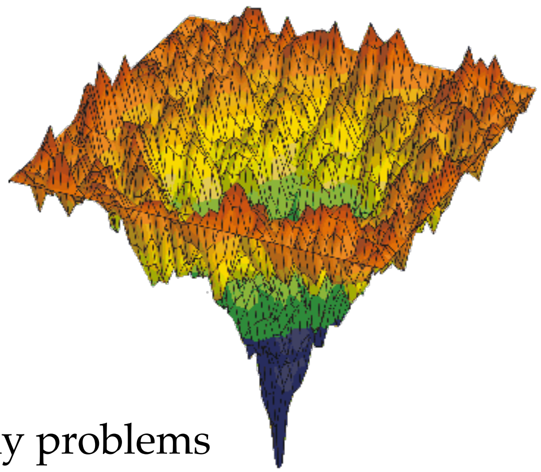


CIM adiabatic evolution outperforms the spectral solution and comes close to the theoretically predicted SK ground state energy

What makes this possible?
What is the shape of the energy landscape?
How does it change with laser pump power?
How can exploit our understanding of this changing geometry to determine optimal annealing schedule?

High dimensional nonconvex optimization

It is often thought that local minima at high error stand as
as a major impediment to non-convex optimization.



In random non-convex error surfaces over
high dimensional spaces, local minima at high
error are exponentially rare in the dimensionality.

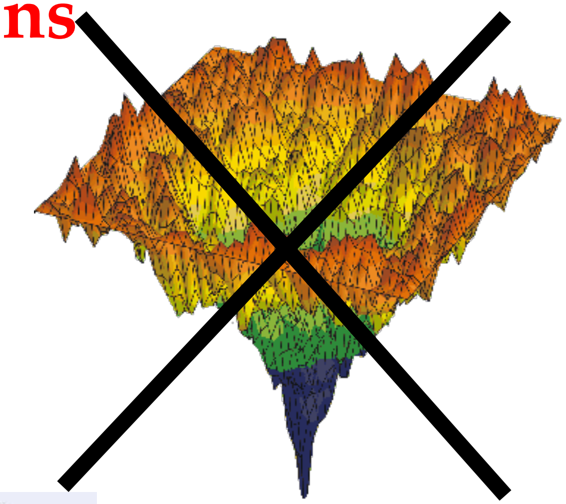
Instead saddle points proliferate.

We demonstrated this picture indeed occurs in many problems
of relevance to deep learning and artificial intelligence,

And we developed an algorithm that rapidly escapes saddle
points in high dimensional spaces.

Identifying and attacking the saddle point problem in high dimensional non-convex optimization.
Yann Dauphin, Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Surya Ganguli, Yoshua Bengio. NIPS 2014

General properties of error landscapes in high dimensions



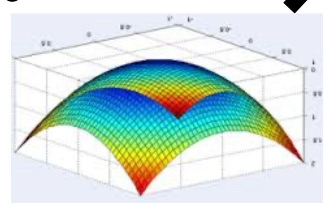
From statistical physics:

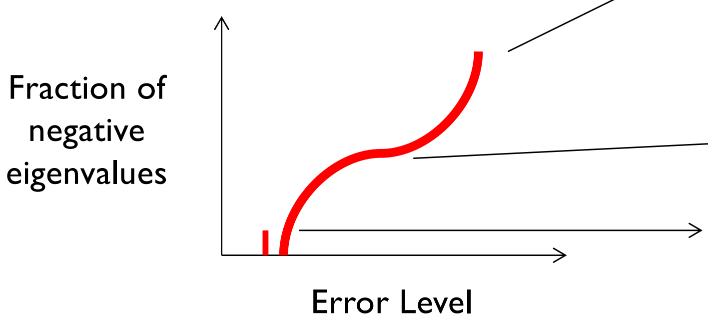
Consider a random Gaussian error landscape over N variables.

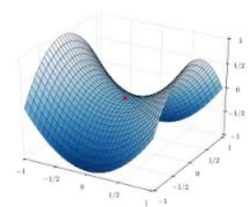
Let x be a critical point.

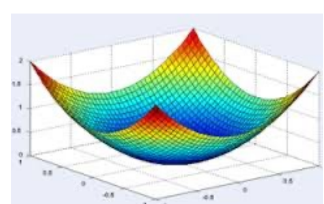
Let E be its error level.

Let f be the fraction of negative curvature directions.



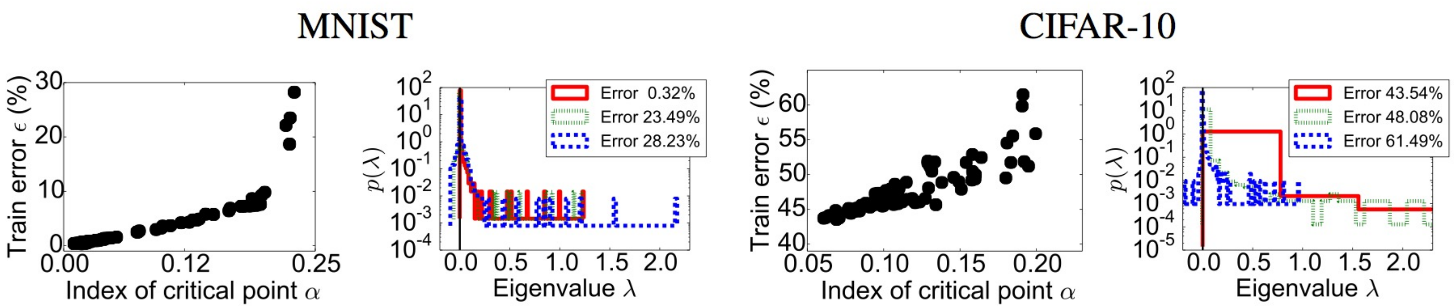






Bray and Dean, Physical Review Letters, 2007

Properties of Error Landscapes on the Synaptic Weight Space of a Deep Neural Net



Qualitatively consistent with the
statistical physics theory of random error landscapes

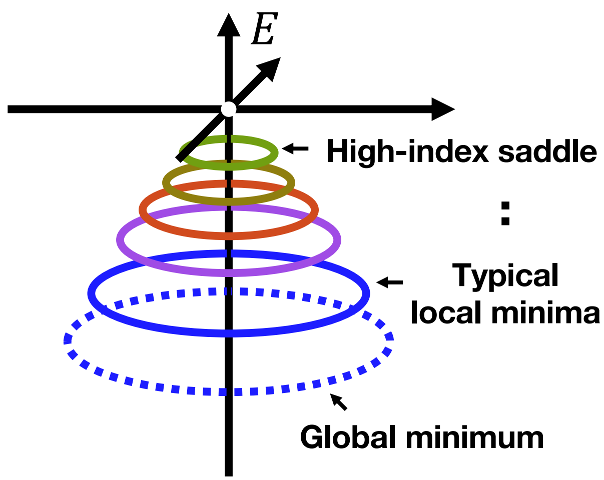
Understanding the changing energy landscape of the CIM under adiabatic evolution of the pump parameter

$$E(x) = \frac{1}{4}\sum_i x_i^4 - \frac{a}{2}\sum_i x_i^2 - \sum_{ij} J_{ij}x_ix_j \qquad x_i \in \mathbb{R}$$

Sherrington Kirkpatrick (SK) spin glass: J_{ij} drawn as a Gaussian: $\mathcal{N}(0, 1/N)$

Questions: In terms of:

1) CIM energy E; 2) radial distance from origin; 3) pump power a:



Where does the global minimum lie?

Where do the most likely local minima lie?

Where do the most likely saddle points of a given index lie?

Where do the lowest energy saddle points of a given index lie?

Given a critical point (minimum or saddle) what is the distribution of OPO quadrature x_i?

What is the eigenvalue spectrum of the Hessian of a critical point
as a function of its index, energy and radius?

Theory for the complexity of critical points

Theoretical techniques:

Kac-Rice Formula

Replica Method

Supersymmetry Breaking

Definition [The complexity of random critical points]

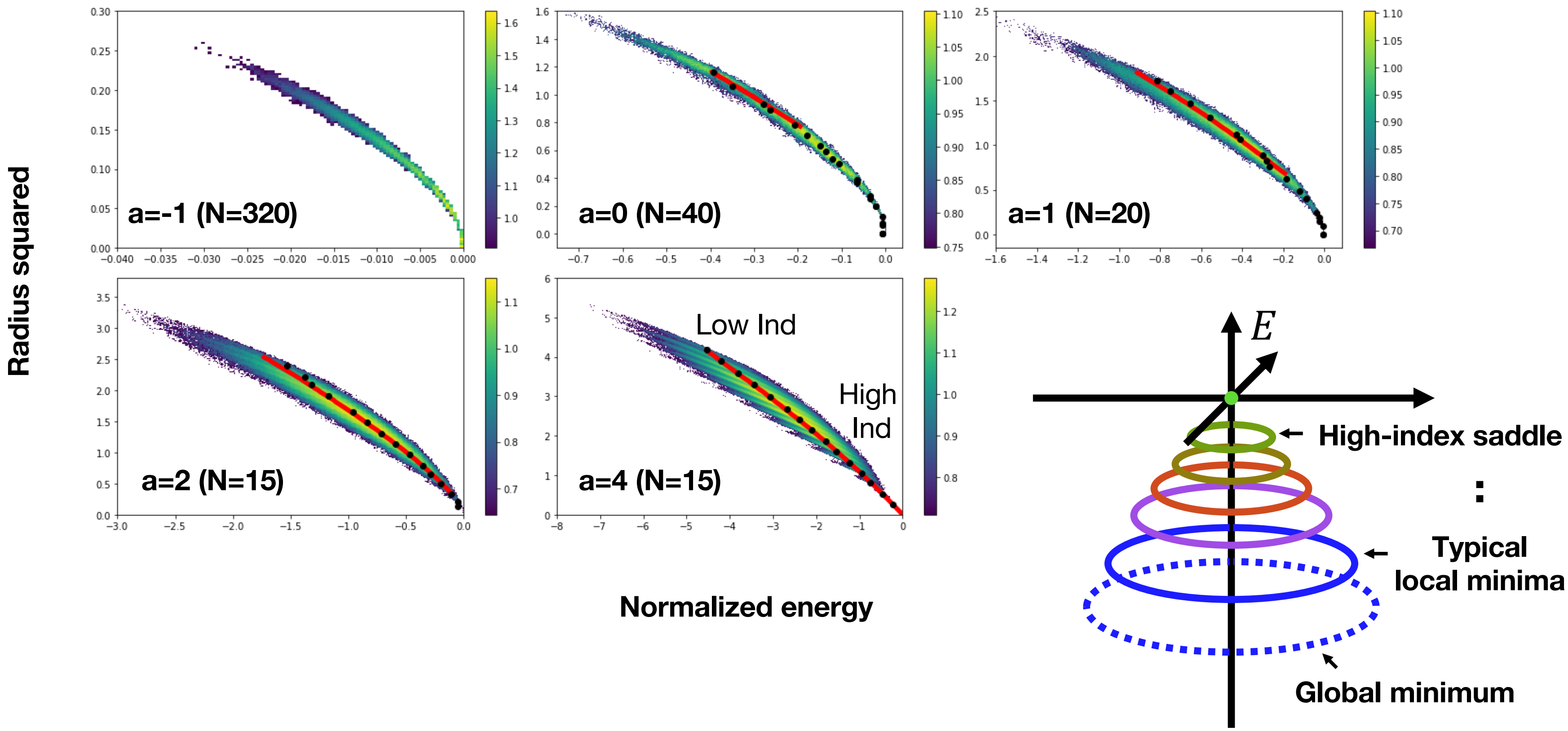
$$N\Sigma(E,r) = \langle \ln \sum_{\alpha}^{\mathcal{N}} \delta(NE - E(\mathrm{x}^{\alpha}))\delta(Nr - \mathcal{I}(\mathrm{x}^{\alpha}))\rangle \quad (1)$$

Definition [The grand potential of random critical points]

$$N\Omega(\beta,\mu) = \langle \ln \sum_{\alpha}^{\mathcal{N}} e^{-\beta E(x^{\alpha})+\mu\mathcal{I}(x^{\alpha})}\rangle \quad (2)$$

- The grand potential $\Omega(\beta,\mu)$: the Legendre transform of the complexity
- The effective temperature β : the Legendre dual of energy E
- The chemical potential μ: the Legendre dual of normalized index I/N

Geometry of the changing CIM Energy landscape as a function of laser pump power a

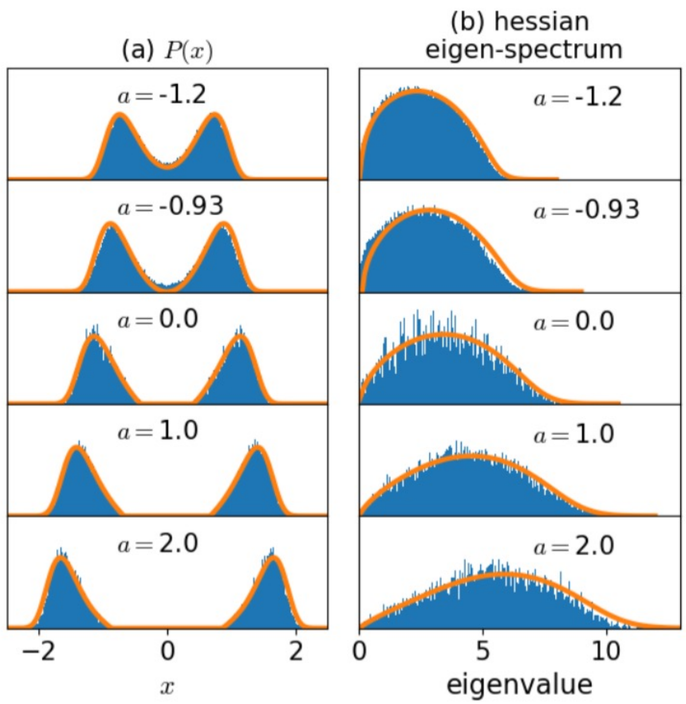


Geometric phase transitions in the low energy structure of the landscape as laser power is increased

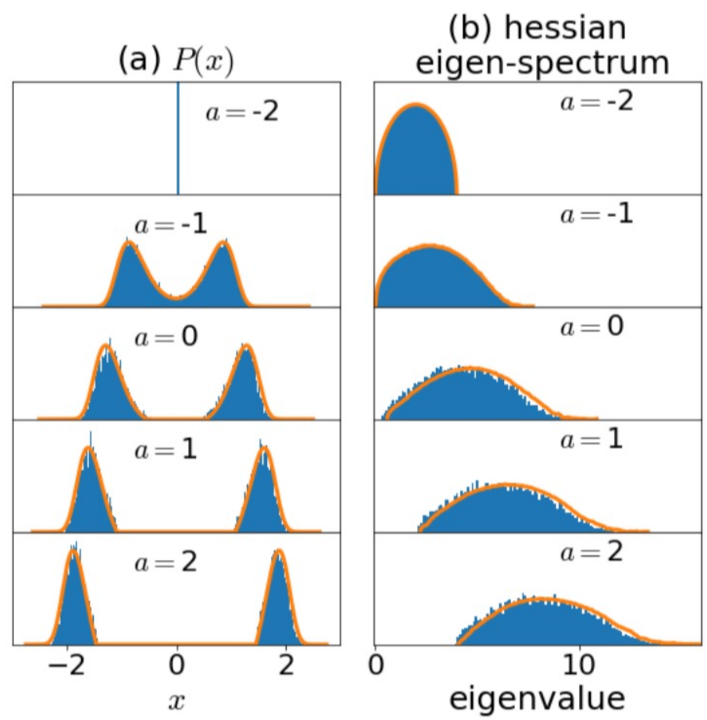
$$E(x) = \frac{1}{4}\sum_i x_i^4 - \frac{a}{2}\sum_i x_i^2 - \sum_{ij} J_{ij} x_i x_j \qquad x_i \in \mathbb{R}$$

Interested in empirical histograms of x_i at and the Hessian eigenspectrum at:

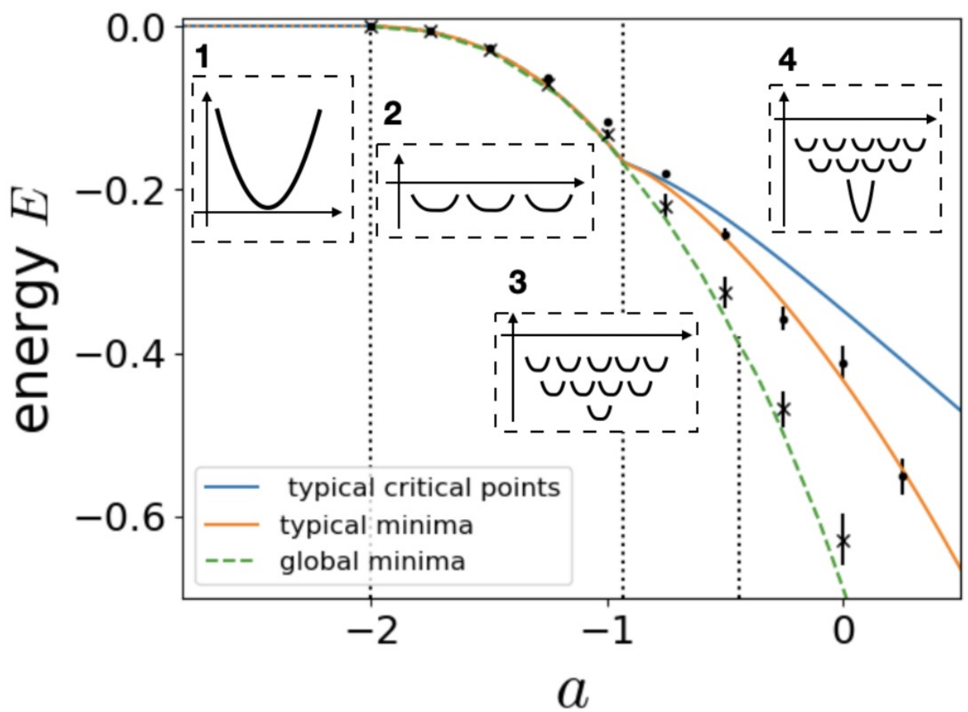
Typical local minima:



Global minimum:

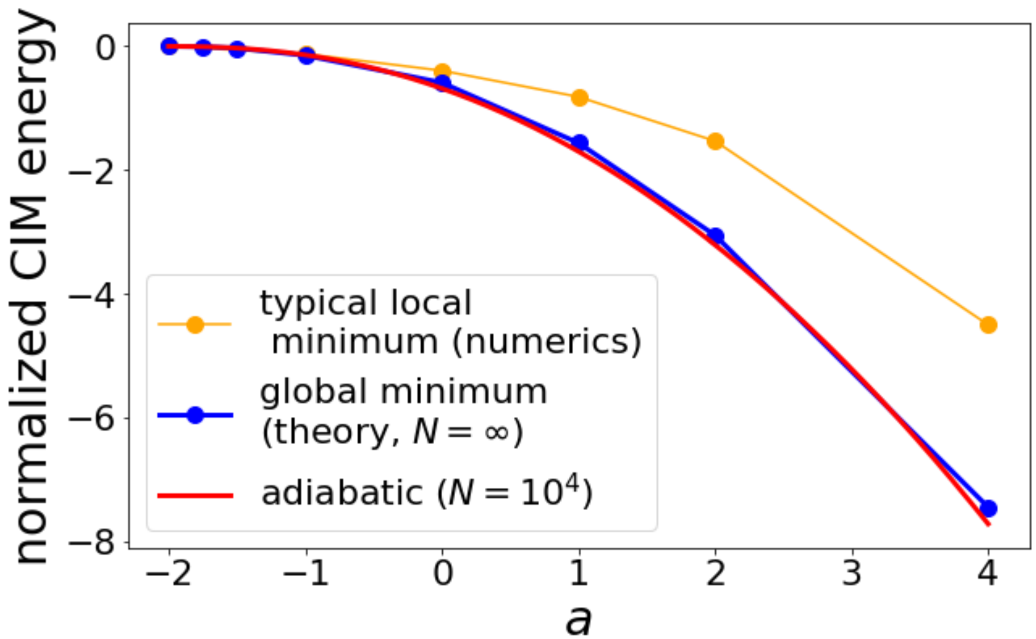


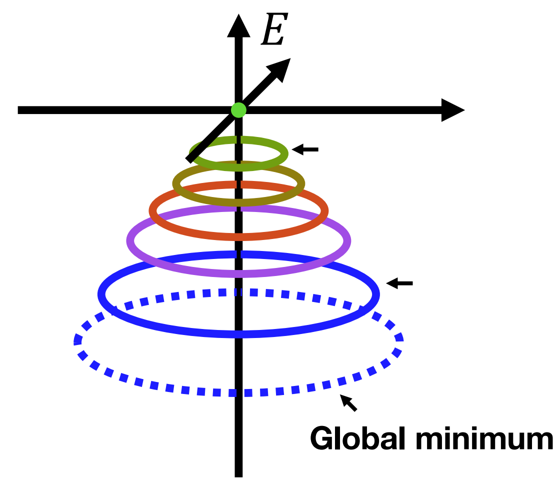
Phase transitions in the geometry of the CIM landscape



$a < -2$	$-2 < a < -0.93$	$-0.93 < a < -0.45$	$-0.45 < a$
Replica symmetric phase:	Replica symmetric breaking:	SUSY breaking:	Freezing:
A single stable local minimum	Exponentially many local minima All of similar energy densities All are marginally stable “Flat landscape”	Exponentially many local minima Most local minima marginally stable Global minimum marginally stable Range of energies Global min < Typical local min	Global minimum is no Longer marginally stable And is fully stable.

Adiabatic evolution tracks the global minimum of CIM energy

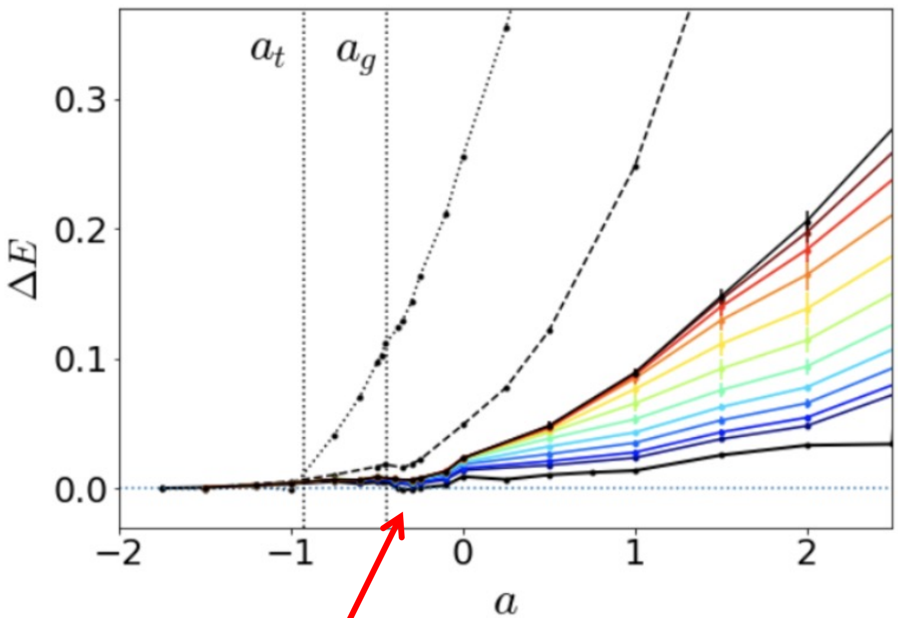




Our geometric understanding of the landscape suggests the optimal annealing schedule:

Energy of typical local minima:

Energy obtained with a rapid quench to a starting from a random initial condition

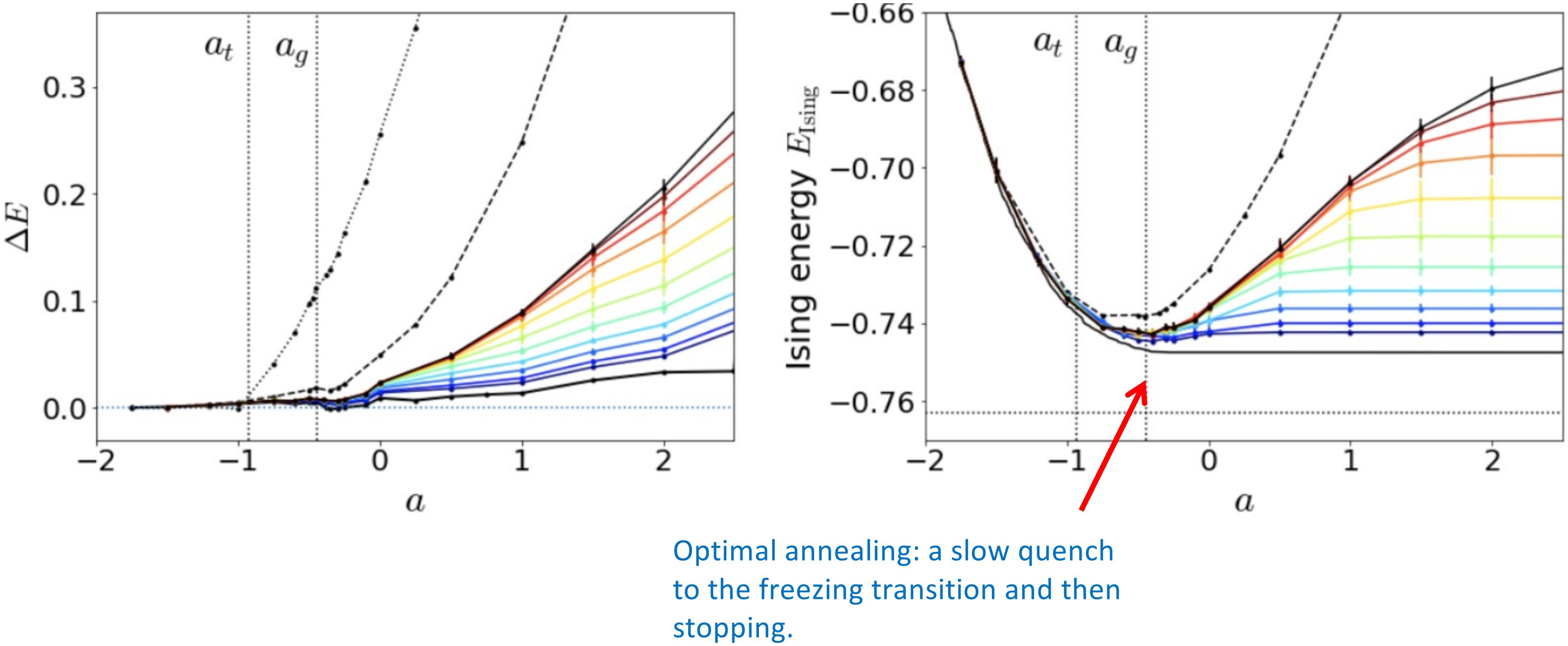


Energy obtained with a rapid quench to a final value of laser power a, starting near origin

Energy obtained with a slow quench to A final value of the laser power

Optimal annealing: increase a linearly and stop at the freezing transition

Our geometric understanding of the landscape suggests the optimal annealing schedule:



Going beyond gradient descent on an energy function: nonconservative forces in the CIM

$$E(x) = \frac{1}{4}\sum_i x_i^4 - \frac{a}{2}\sum_i x_i^2 - \frac{1}{2}\sum_{ij} J_{ij} x_i x_j \qquad x_i \in \mathbb{R}$$

$$\frac{dx_i}{dt} = -\frac{\partial E}{\partial x} = -x_i^3 + a x_i^2 + \sum_j J_{ij} x_j$$

Non conservative forces: make J asymmetric - can introduce chaos, to destabilize bad fixed points?

Critical points of an energy landscape	⟷	Fixed points of a dynamical system
Spectrum of Hessian		Spectrum Jacobian of vector field

Mapping more naturally occurring optimization problems to the CIM and understanding the energy landscape

SOLVING CONSTRAINED QUADRATIC BINARY PROBLEMS VIA QUANTUM ADIABATIC EVOLUTION,
POOYA RONAGH,
BRAD WOODS, AND
EHSAN IRANMANESH,
arxiv: 1509.0500

$$\begin{aligned} \text{Minimize} \quad & x^T Q x \\ \text{subject to} \quad & Ax \leq b, \\ & x \in \{0,1\}^n \end{aligned}$$

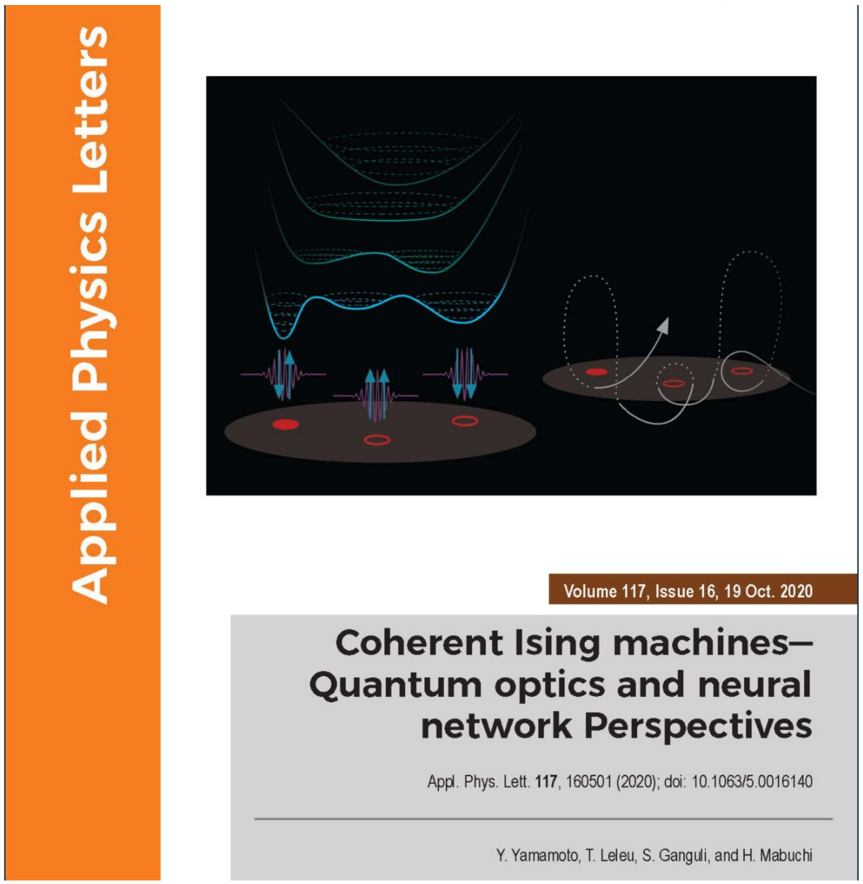
Quadratic assignment problem with constraints

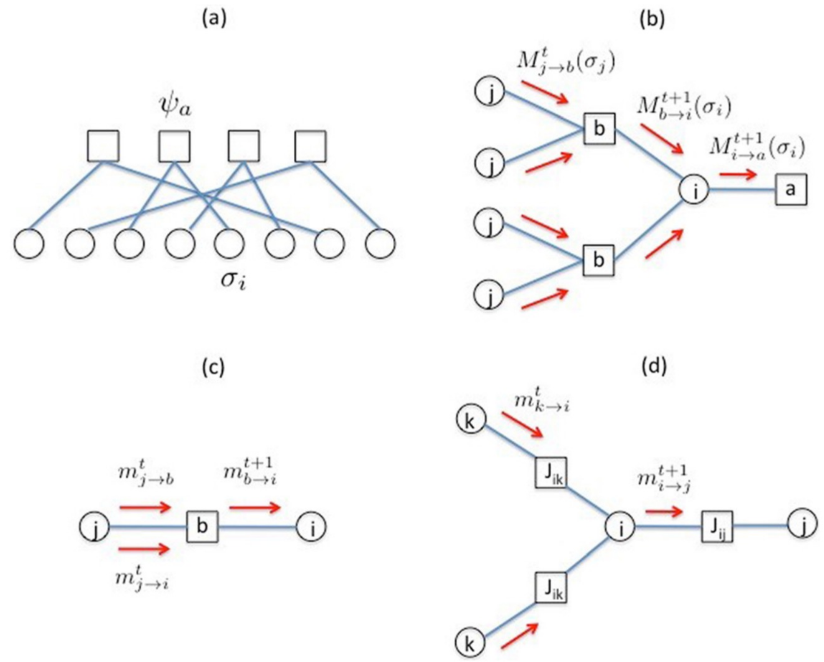
Quantum bridge analytics:
Fred Glover1 , Gary Kochenberger , Yu Du
Arxiv:1811.11538

- Quadratic Assignment Problems
- Capital Budgeting Problems
- Multiple Knapsack Problems
- Task Allocation Problems (distributed computer systems)
- Maximum Diversity Problems
- P-Median Problems
- Asymmetric Assignment Problems
- Symmetric Assignment Problems
- Side Constrained Assignment Problems
- Quadratic Knapsack Problems

- Constraint Satisfaction Problems (CSPs)
- Discrete Tomography Problems
- Set Partitioning Problems
- Set Packing Problems
- Warehouse Location Problems
- Maximum Clique Problems
- Maximum Independent Set Problems
- Maximum Cut Problems
- Graph Coloring Problems
- Number Partitioning Problems
- Linear Ordering Problems
- Clique Partitioning Problems
- SAT problems

Relations belief and survey propagation





Discrete variables	Ising Spins	Comb Opt
Real variables.	OPO quadrature x	Belief prop
Functions over Real variables	Quantum wave. functions	Survey prop

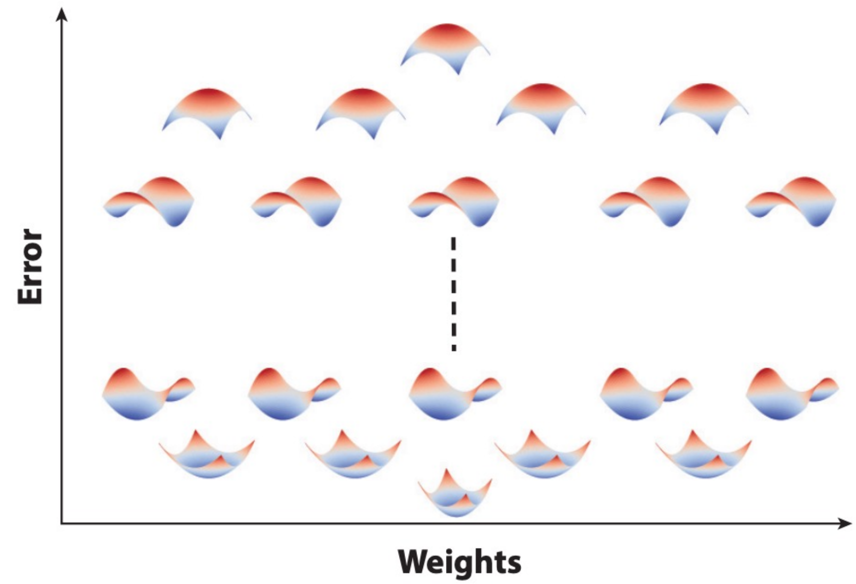
Deterministic annealing in other areas of artificial intelligence

Statistical mechanics of deep learning

Yasaman Bahri,[1] Jonathan Kadmon,[2] Jeffrey Pennington,[1] Sam Schoenholz,[1], Jascha Sohl-Dickstein,[1] and Surya Ganguli[1,2]

[1]Google Brain, Google Inc., Mountain View, CA 94043
[2]Department of Applied Physics, Stanford University, Stanford, CA 94035

Contents



Deterministic annealing of a broader class
Of hyperparameters? i.e. the nonlinearity?

Outline

What is the optimal way to do regression in high dimensions?

M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.

M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.

Deriving the first steps of vision from the statistics of natural movies

Sam Ocko, Jack Lindsey, S. Ganguli, Stephane Deny, The emergence of multiple retinal cell-types through the efficient coding of natural movies. Neural Information Processing Systems 2018.

Quantum neuromorphic computing with atoms and photons

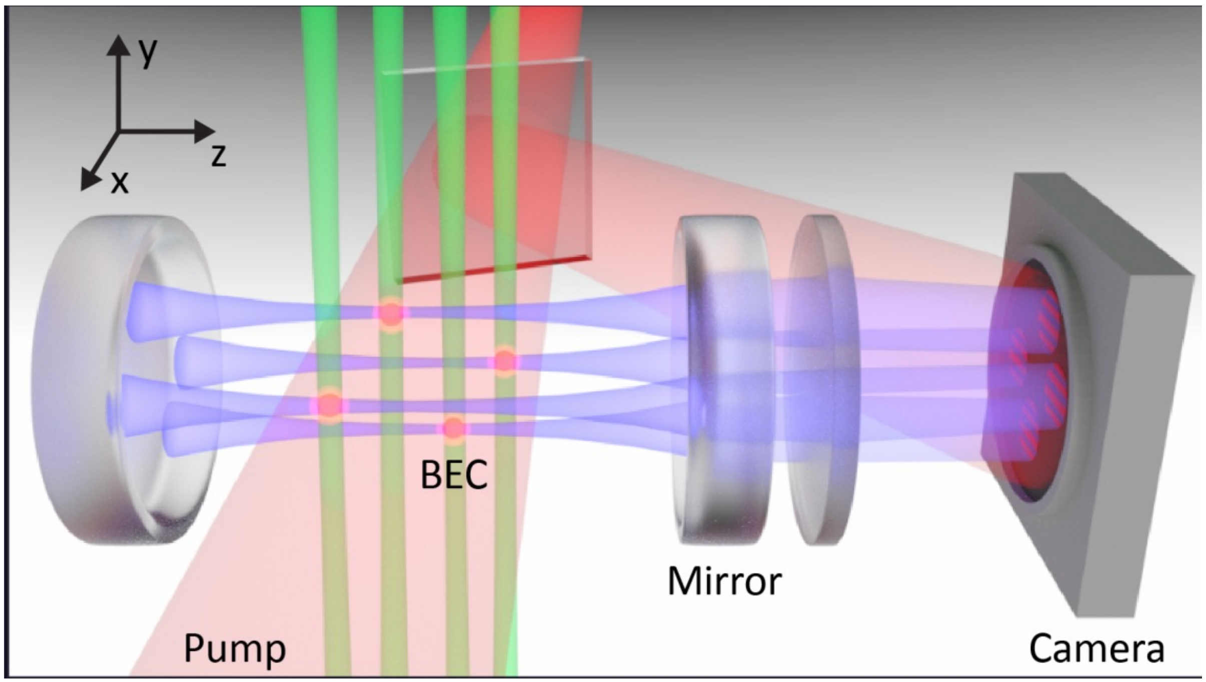
Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev, Enhancing associative memory recall and storage capacity using confocal cavity QED, Physical Review X 2021.

Yoshi Yamamoto, Timothee Leleu, S. Ganguli, Hideo Mabuchi, Coherent Ising Machines: quantum optics and neural network perspectives, Applied Physics Letters, 2020.

Enhancing associative memory recall and storage capacity using confocal cavity QED

Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev

We introduce a near-term experimental platform for realizing an associative memory. It can simultaneously store many memories by using spinful bosons coupled to a degenerate multimode optical cavity. The associative memory is realized by a confocal cavity QED neural network, with the cavity modes serving as the synapses, connecting a network of superradiant atomic spin ensembles, which serve as the neurons. Memories are encoded in the connectivity matrix between the spins, and can be accessed through the input and output of patterns of light. Each aspect of the scheme is based on recently demonstrated technology using a confocal cavity and Bose-condensed atoms. Our scheme has two conceptually novel elements. First, it introduces a new form of random spin system that interpolates between a ferromagnetic and a spin-glass regime as a physical parameter is tuned---the positions of ensembles within the cavity. Second, and more importantly, the spins relax via deterministic steepest-descent dynamics, rather than Glauber dynamics. We show that this nonequilibrium quantum-optical scheme has significant advantages for associative memory over Glauber dynamics: These dynamics can enhance the network's ability to store and recall memories beyond that of the standard Hopfield model. Surprisingly, the cavity QED dynamics can retrieve memories even when the system is in the spin glass phase. Thus, the experimental platform provides a novel physical instantiation of associative memories and spin glasses as well as provides an unusual form of relaxational dynamics that is conducive to memory recall even in regimes where it was thought to be impossible.

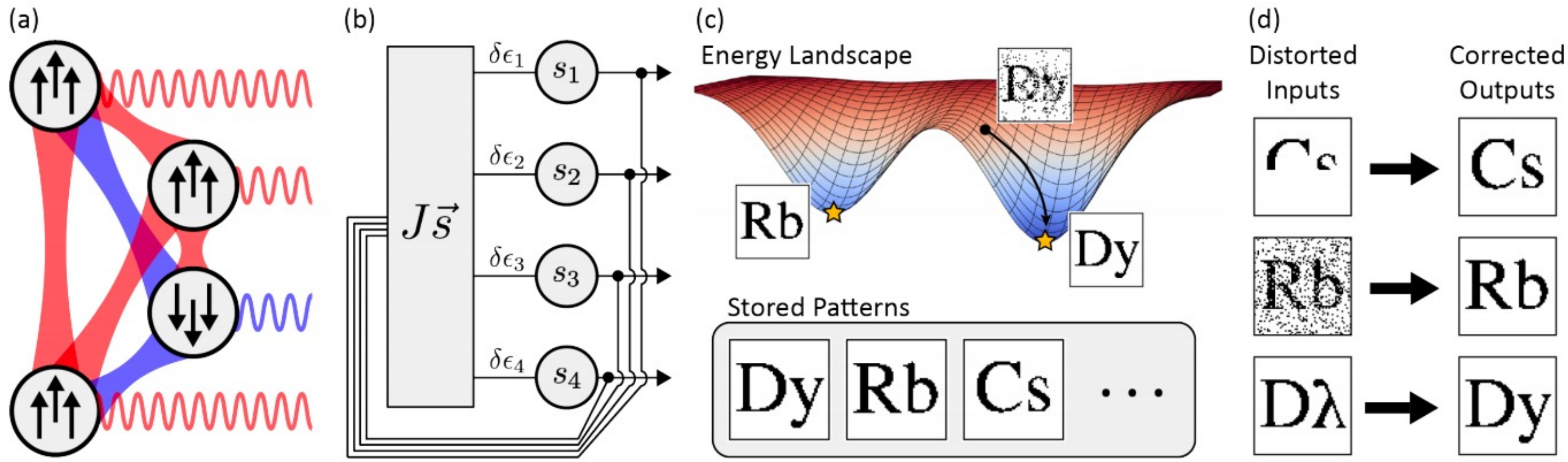


Atomic spins <-> Neurons <-> State $s_i = +/- 1$

Photons <-> Synapses <-> Connectivity J_{ij}

Energy function: $$E(s_1, \ldots, s_N) = -\sum_{ij=1}^{N} J_{ij} s_i s_j$$

Memories = minima of energy function
Memory recall = energy minimizing dynamics from a partial cue



Classical Glauber dynamics:

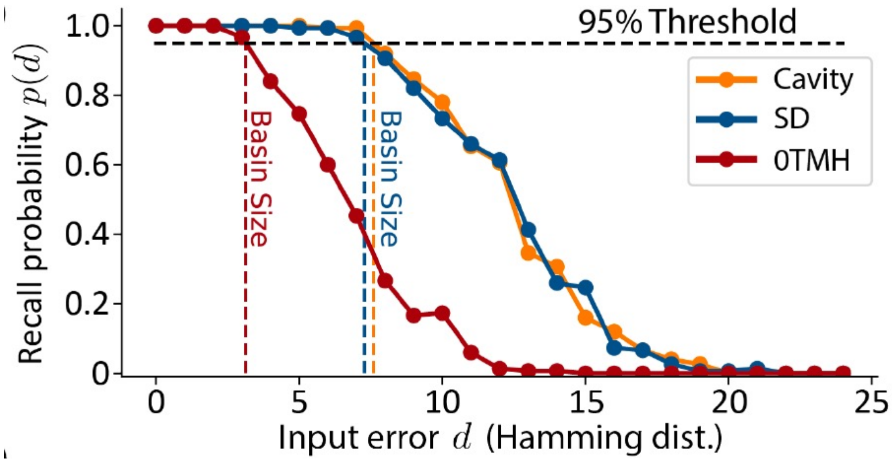
2 ways to minimize energy

1) Flip a spin at random
2) Compute dE = change in energy
3) If dE < 0 accept the spin flip
4) If dE >=0 accept the spin flip with probability exp(-dE/T) where T = temperature

Classical limit of emergent quantum dynamics (steepest descent)

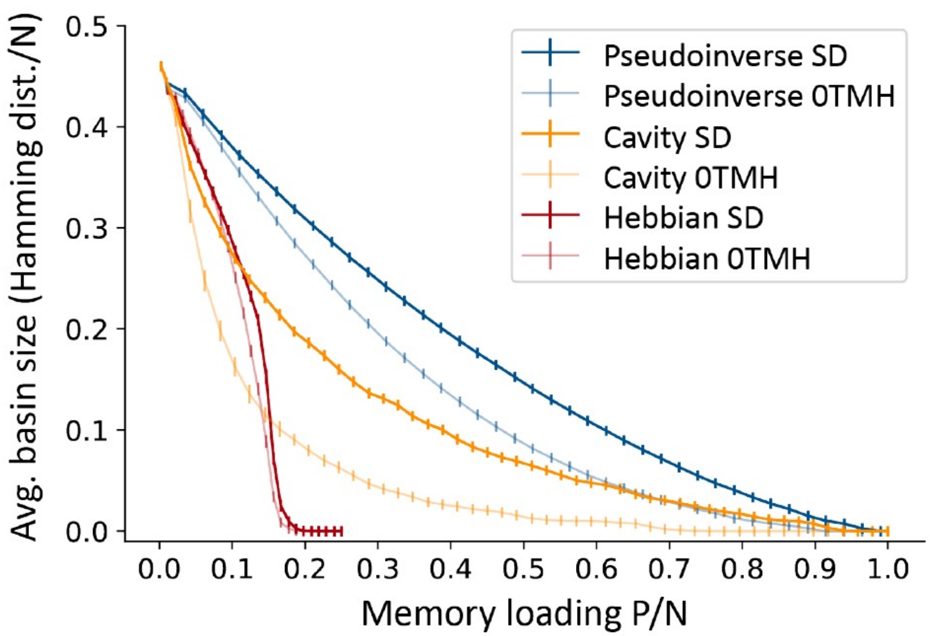
1) Do a quantum parallel search over all N spins to find the spin whose flip would lower the energy the most.
2) Flip that spin.

Steepest descent has better memory recall characteristics



Steepest descent can tolerate more errors in the initial pattern.

This leads to increased basin size of memories (energy minima).



Steepest descent yields a better tradeoff between:

capacity (number of memories) and

robustness (basin size)

for 3 different connectivity types

References on deep learning theory: http://ganguli-gang.stanford.edu

- M. Advani and S. Ganguli, An equivalence between high dimensional Bayes optimal inference and M-estimation, NIPS 2016.
- M. Advani and S. Ganguli, Statistical mechanics of optimal convex inference in high dimensions, Physical Review X, 6, 031034, 2016.
- A. Saxe, J. McClelland, S. Ganguli, Learning hierarchical category structure in deep neural networks, Proc. of the 35th Cognitive Science Society, pp. 1271-1276, 2013.
- A. Saxe, J. McClelland, S. Ganguli, Exact solutions to the nonlinear dynamics of learning in deep neural networks, ICLR 2014.
- Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, Y. Bengio, Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, NIPS 2014.
- B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, Exponential expressivity in deep neural networks through transient chaos, NIPS 2016.
- S. Schoenholz, J. Gilmer, S. Ganguli, and J. Sohl-Dickstein, Deep information propagation, ICLR 2017.
- S. Lahiri, J. Sohl-Dickstein and S. Ganguli, A universal tradeoff between energy speed and accuracy in physical communication, arxiv 1603.07758
- A memory frontier for complex synapses, S. Lahiri and S. Ganguli, NIPS 2013.
- Continual learning through synaptic intelligence, F. Zenke, B. Poole, S. Ganguli, ICML 2017.
- Modelling arbitrary probability distributions using non-equilibrium thermodynamics, J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, ICML 2015.
- Deep Knowledge Tracing, C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, J. Sohl-Dickstein, NIPS 2015.
- Deep learning models of the retinal response to natural scenes, L. McIntosh, N. Maheswaranathan, S. Ganguli, S. Baccus, NIPS 2016.
- Resurrecting the sigmoid in deep learning through dynamical isometry: theory and practice, J. Pennington, S. Schloenholz, and S. Ganguli, NIPS 2017.
- Variational walkback: learning a transition operator as a recurrent stochastic neural net, A. Goyal, N.R. Ke, S. Ganguli, Y. Bengio, NIPS 2017.
- The emergence of spectral universality in deep networks, J. Pennington, S. Schloenholz, and S. Ganguli, AISTATS 2018.
- The statistical mechanics of low-rank tensor decompositions, J. Kadmon, S. Ganguli, NIPS 2018.
- A mathematical theory of semantic development in deep neural networks, A. Saxe, J. McClelland, S. Ganguli, PNAS 2019.
- An analytic theory of generalization dynamics and transfer learning in deep linear networks, ICLR 2019.
- The emergence of multiple retinal cell-types through the efficient coding of natural movies, S. Ocko, J. Lindsey, S. Ganguli, S. Deny, NeurIPS 2018.
- A computational origin for grid cells, B. Sorscher, G. Mel, S.Ganguli, S. Ocko. Neurips 2019
- Brendan P. Marsh, Yudan Guo, Ronen M. Kroeze, Sarang Gopalakrishnan, Surya Ganguli, Jonathan Keeling, Benjamin L. Lev,
- Enhancing associative memory recall and storage capacity using confocal cavity QED, Phyiscal Review X, 2021.
- Y. Yamamoto, T. Leleu, S. Ganguli, H. Mabuchi, Coherent Ising Machines: quantum optics and neural network perspectives, Applied Physics Letters, 2020.

Tools:

- Non-equilibrium statistical mechanics
- Dynamical mean field theory
- Statistical mechanics of random landscapes
- Riemannian geometry
- Random matrix theory
- Free probability theory