Unlocking the power of the variational free-energy principle with deep generative models

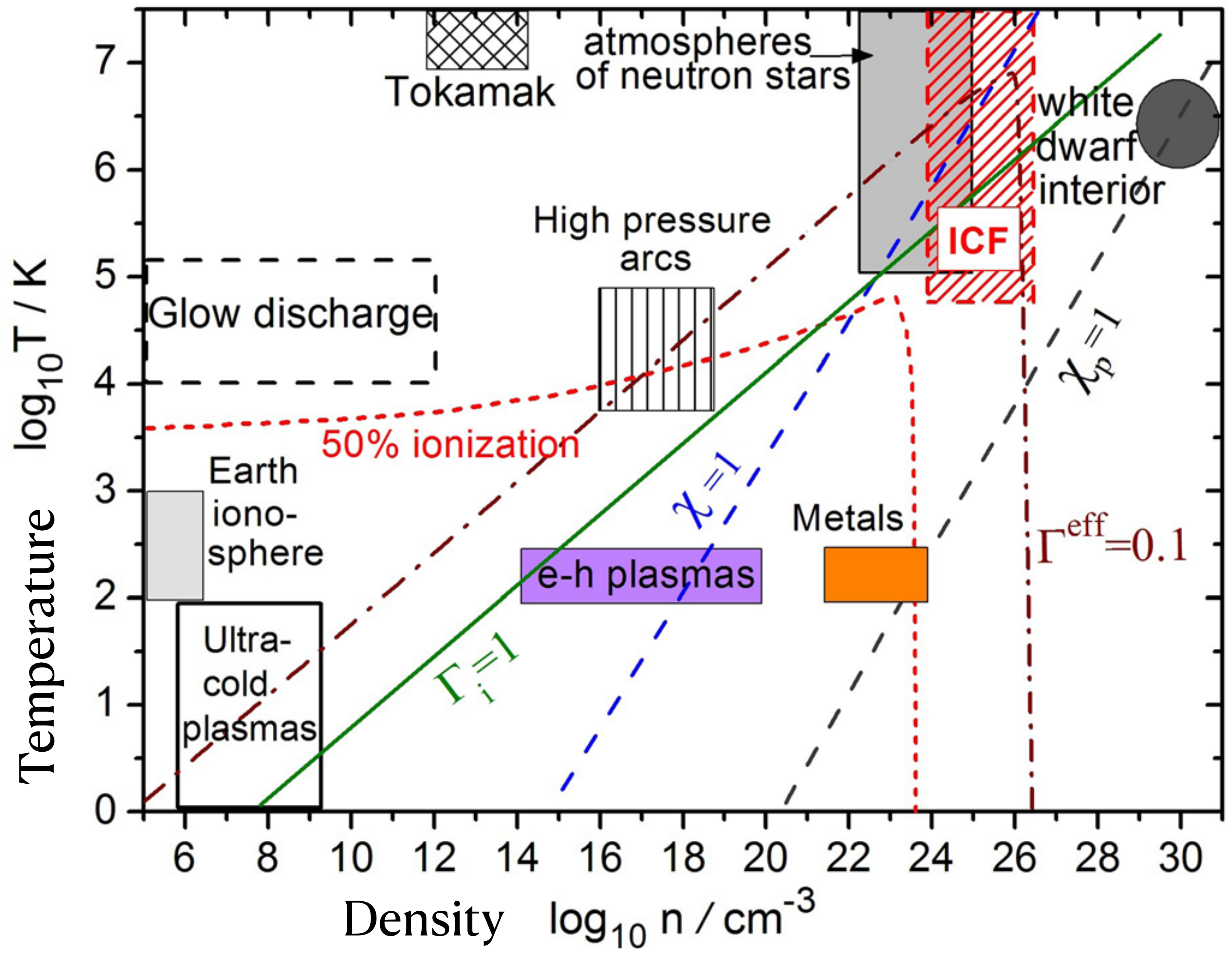
Lei Wang (王磊)
Institute of Physics, CAS
https://wangleiphy.github.io

“Using AI to accelerate scientific discovery” Demis Hassabis, co-founder and CEO of DeepMind 2021

Ab-initio study of quantum matters at T>0

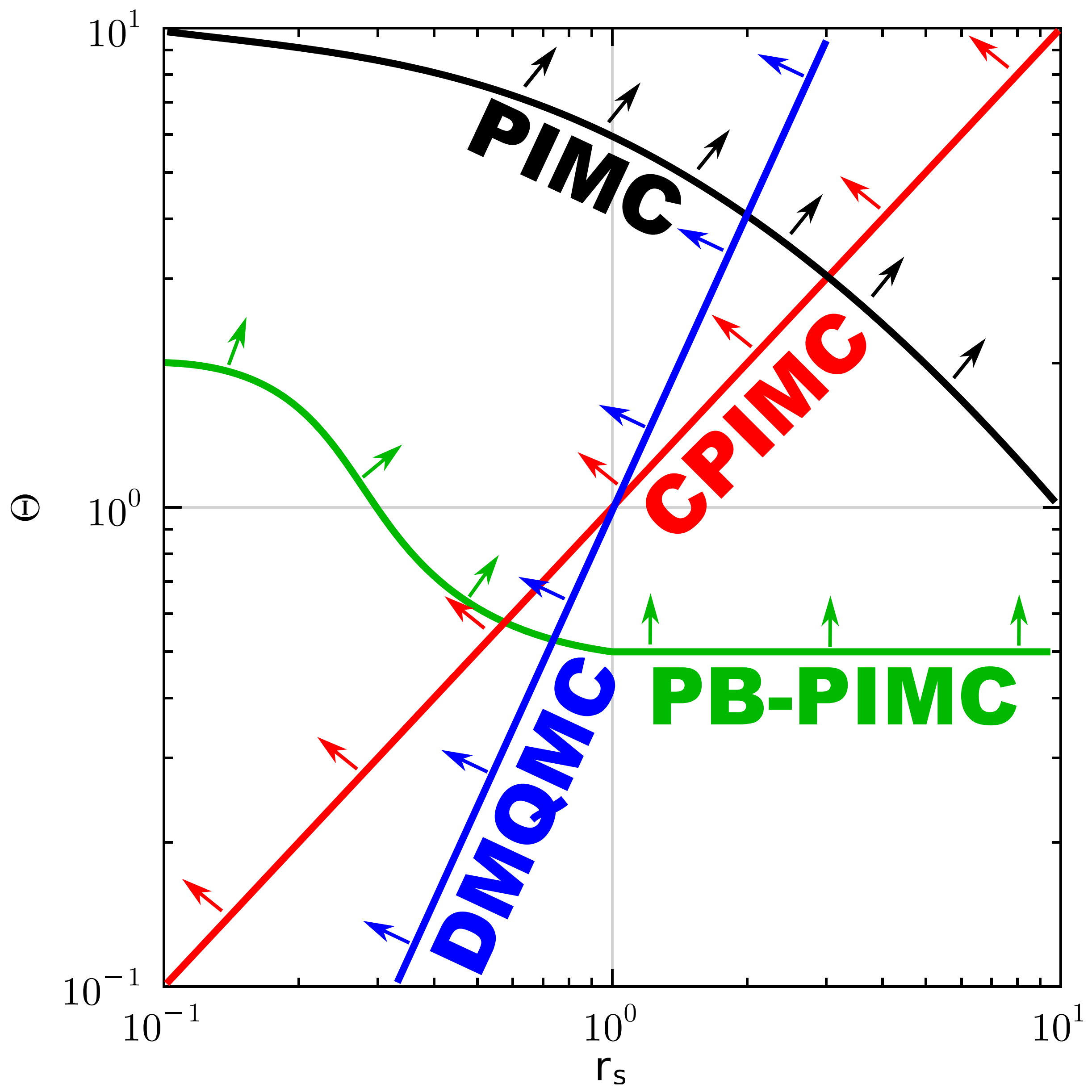
$$H = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_I \frac{\hbar^2}{2m_I}\nabla_I^2 - \sum_{I,i} \frac{Z_I e^2}{|R_I - r_i|} + \frac{1}{2}\sum_{i\neq j} \frac{e^2}{|r_i - r_j|} + \frac{1}{2}\sum_{I\neq J} \frac{Z_I Z_J e^2}{|R_I - R_J|}$$

$$Z = \mathrm{Tr}(e^{-H/k_B T})$$



Bonitz et al, Phys. Plasmas '20

Quantum Monte Carlo
is limited by the sign problem



Dornheim et al, Phys. Plasmas '17

Warmup: $\hbar = 0$

$$Z = \int d\mathbf{X}\, e^{-H(\mathbf{X})/k_B T}$$

The Gibbs-Bogolyubov-Feynman variational free energy principle

$$F = \int dX\, p(X) \left[\underbrace{k_B T \ln p(X)}_{\text{entropy}} + \underbrace{H(X)}_{\text{energy}} \right] \geq -k_B T \ln Z$$

Difficulties in Applying the Variational Principle to Quantum Field Theories[1]

Richard P. Feynman

deep generative models !

[1]transcript of Professor Feynman's talk in 1987

Discriminative learning

Generative learning

$$y = f(\boldsymbol{x})$$

or $p(y \mid \boldsymbol{x})$

$$p(\boldsymbol{x}, y)$$

Generative models and their physics genes

Goodfellow,
NIPS tutorial, 1701.00160

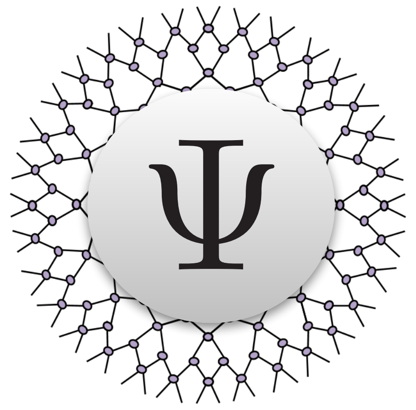
$p(\mathbf{X})$

Explicit density

Implicit density

Direct

GAN



Tensor Networks

Han et al, PRX 18

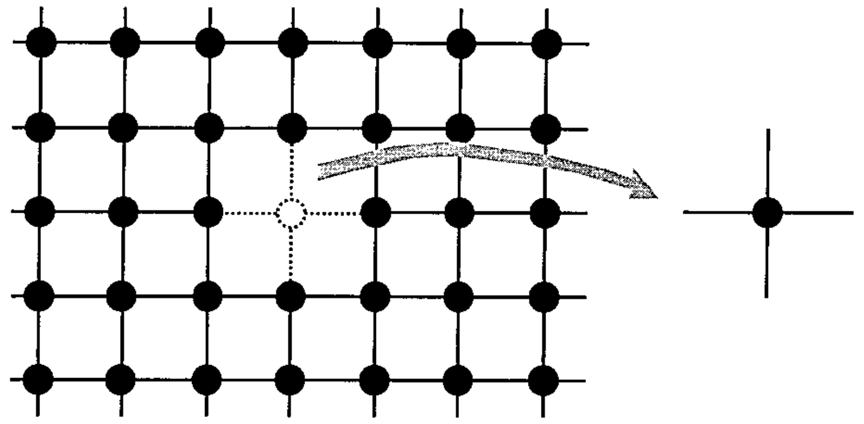
Tractable density

-Fully visible belief nets
 -NADE
 -MADE
 -PixelRNN
-Change of variables models (nonlinear ICA)

Approximate density

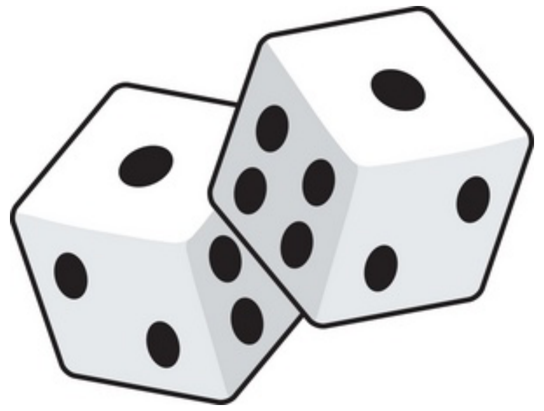
Variational

Variational autoencoder



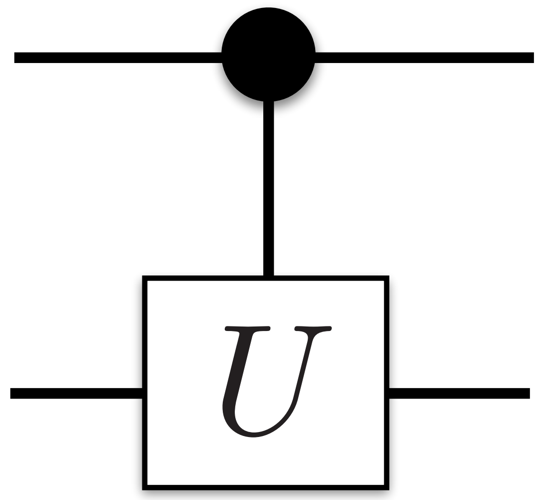
Markov Chain

Boltzmann machine



Markov Chain

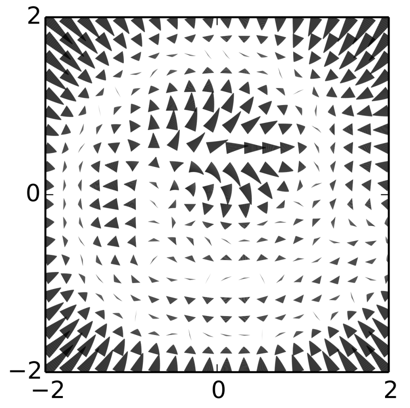
GSN



Quantum Circuits

Liu et al, PRA 18

+Diffusion models



Generative models	Statistical physics
Negative log-likelihood	Energy function
Score function	Force
Latent variables	Collective variables/coarse graining/renormalization group
Partition function	Free energy calculation
Sample diversity	Enhanced sampling

Two sides of the same coin

Generative modeling

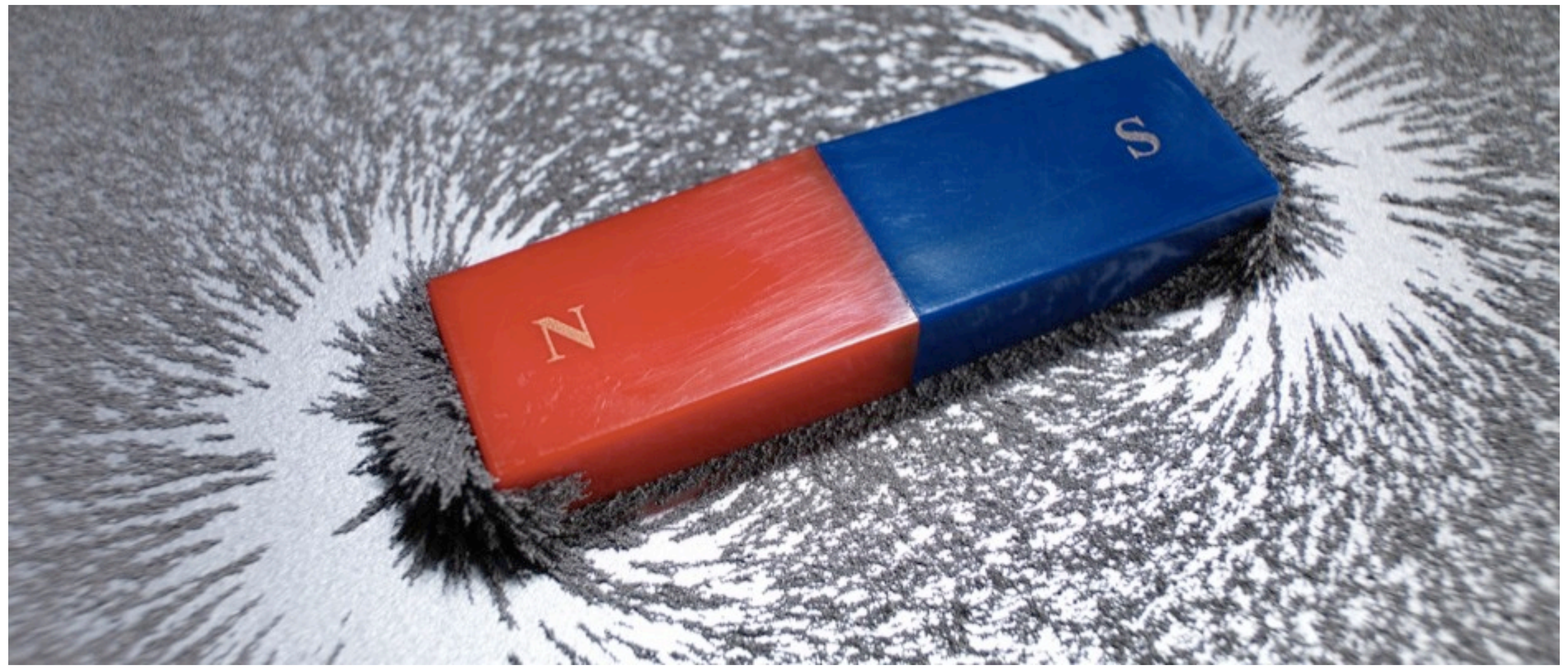
Known: samples

Unknown: generating distribution

“learn from data”

$$\mathscr{L} = -\,\mathbb{E}_{X\sim\text{data}}\left[\ln p(X)\right]$$

Statistical physics



Known: energy function

Unknown: samples, partition function

“learn from Hamiltonian”

$$F = \mathop{\mathbb{E}}_{X\sim p(X)}\left[H(X) + k_B T \ln p(X)\right]$$

$$\mathbb{KL}(\text{data} \,\|\, p) \quad \text{vs} \quad \mathbb{KL}(p \,\|\, e^{-H/k_B T})$$

Deep variational free energy approach

Deep generative models unlocks the power of
the Gibbs-Bogolyubov-Feynman variational principle

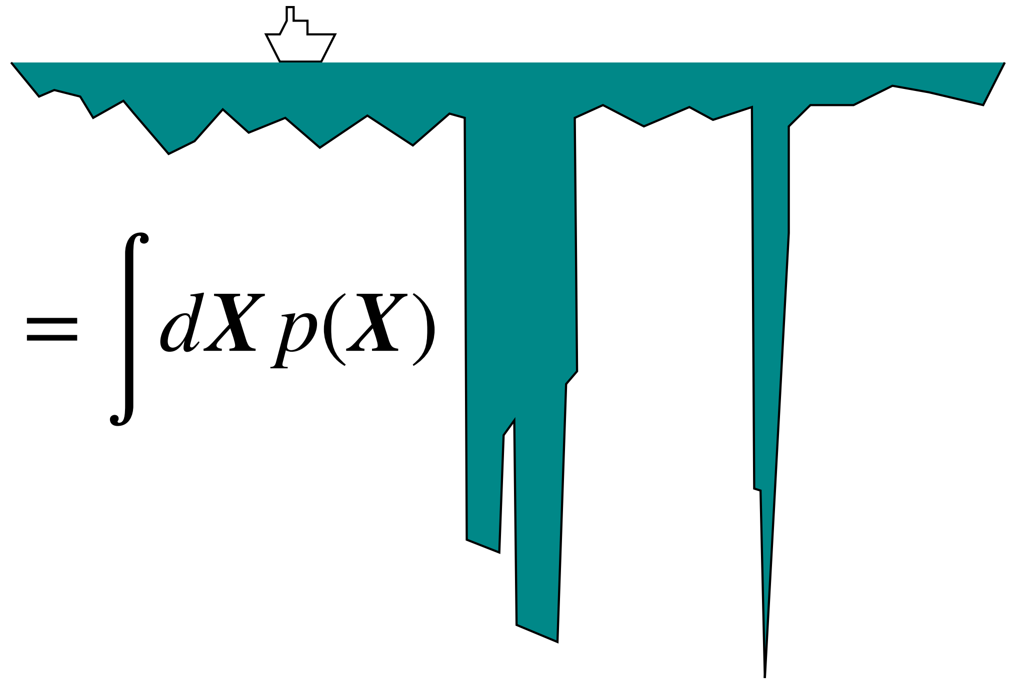
$$F[p] = \mathop{\mathbb{E}}_{\boldsymbol{X} \sim p(\boldsymbol{X})} \left[\underbrace{k_B T \ln p(\boldsymbol{X})}_{\text{entropy}} + \underbrace{E(\boldsymbol{X})}_{\text{energy}} \right] \geq -k_B T \ln Z$$

😁 entropy   energy

Li and LW, PRL '18
Wu, LW, Zhang, PRL '19

Tractable normalization

Mackay, Information Theory, Inference, and Learning Algorithms



$$1 = \int d\boldsymbol{X}\, p(\boldsymbol{X})$$

Direct sampling

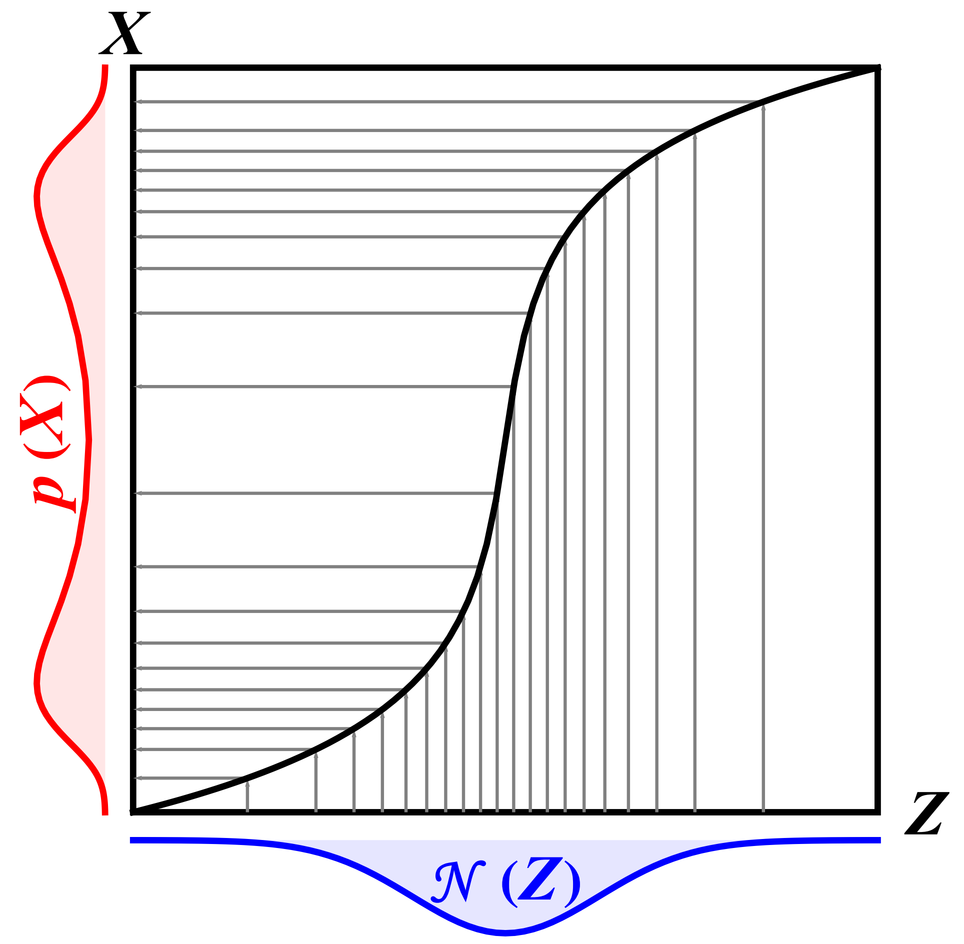
$$\mathop{\mathbb{E}}_{\boldsymbol{X} \sim p(\boldsymbol{X})}$$

Krauth, Statistical Mechanics: Algorithms and Computations

Examples of deep generative models

Normalizing flow

$$p(\boldsymbol{X}) = \mathcal{N}(\boldsymbol{Z}) \left| \det\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{X}}\right) \right|$$



Implementation: invertible Resnet (backflow)...

Autoregressive model

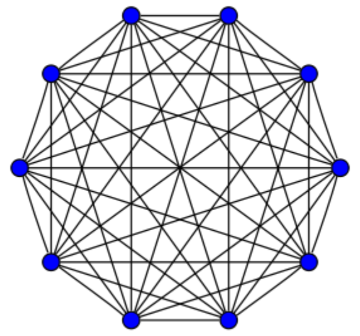
$$p(\boldsymbol{X}) = p(x_1)p(x_2 \,|\, x_1)p(x_3 \,|\, x_1, x_2)\cdots$$

"... the murderer is ___"

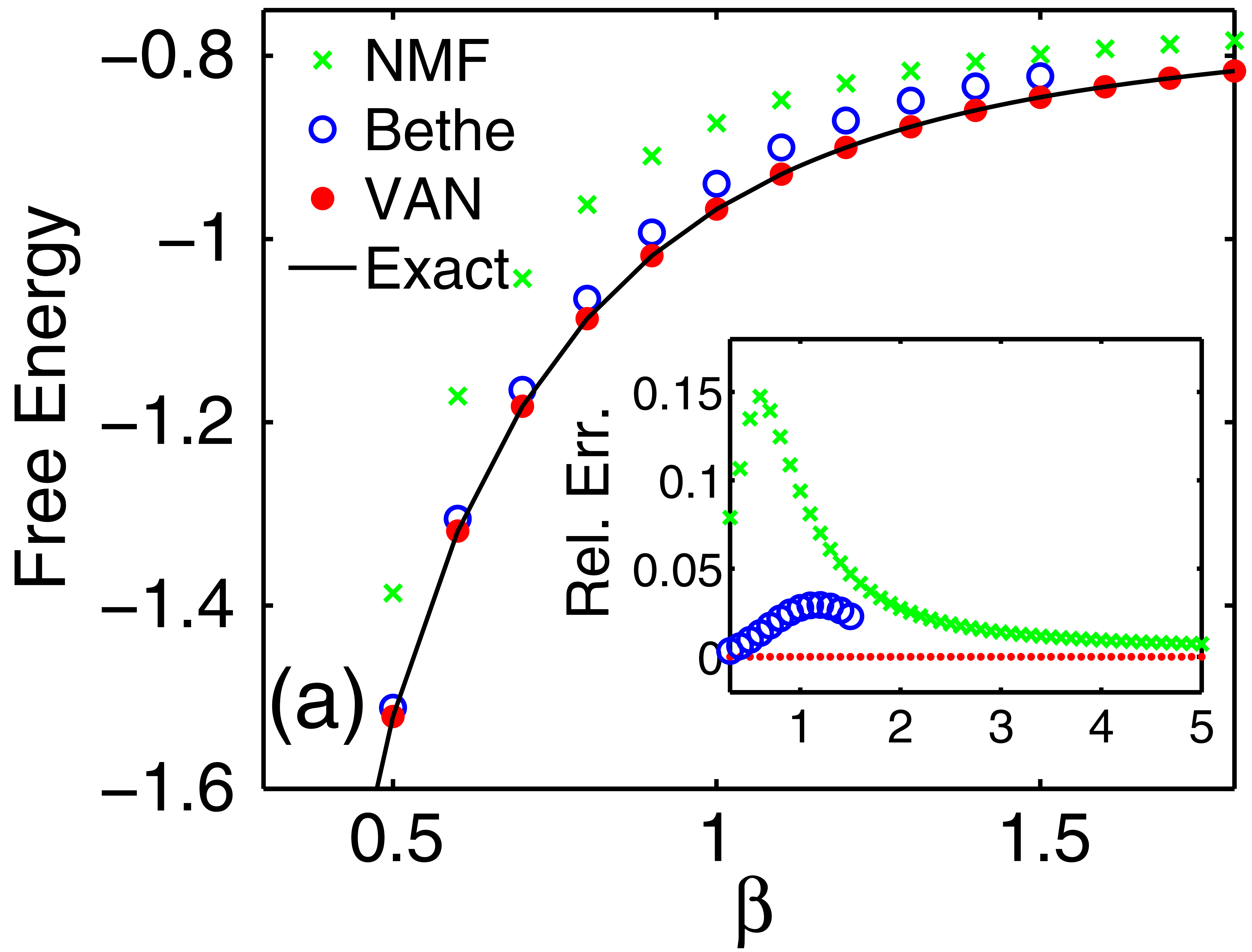
$p(\_ \,|\, \ldots)$

Implementation: transformer with causal mask...

Variational autoregressive networks



Sherrington-Kirkpatrick spin glass



Naive mean-field factorized probability $p(\boldsymbol{X}) = \prod_i p(x_i)$

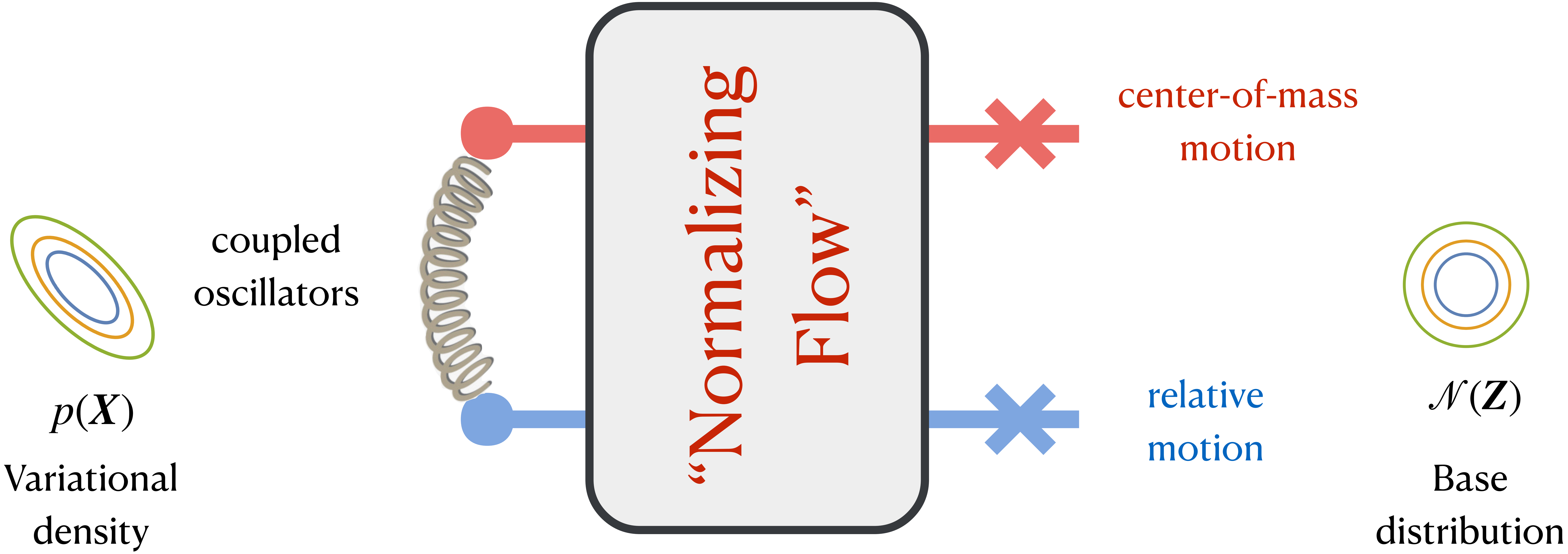
Bethe approximation pairwise interaction $p(\boldsymbol{X}) = \prod_i p(x_i) \prod_{(i,j)\in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$

Variational autoregressive network $p(\boldsymbol{X}) = \prod_i p(x_i | \boldsymbol{x}_{<i})$

Wu, LW, Zhang, PRL '19

github.com/wdphy16/stat-mech-van

Physics intuition of normalizing flow

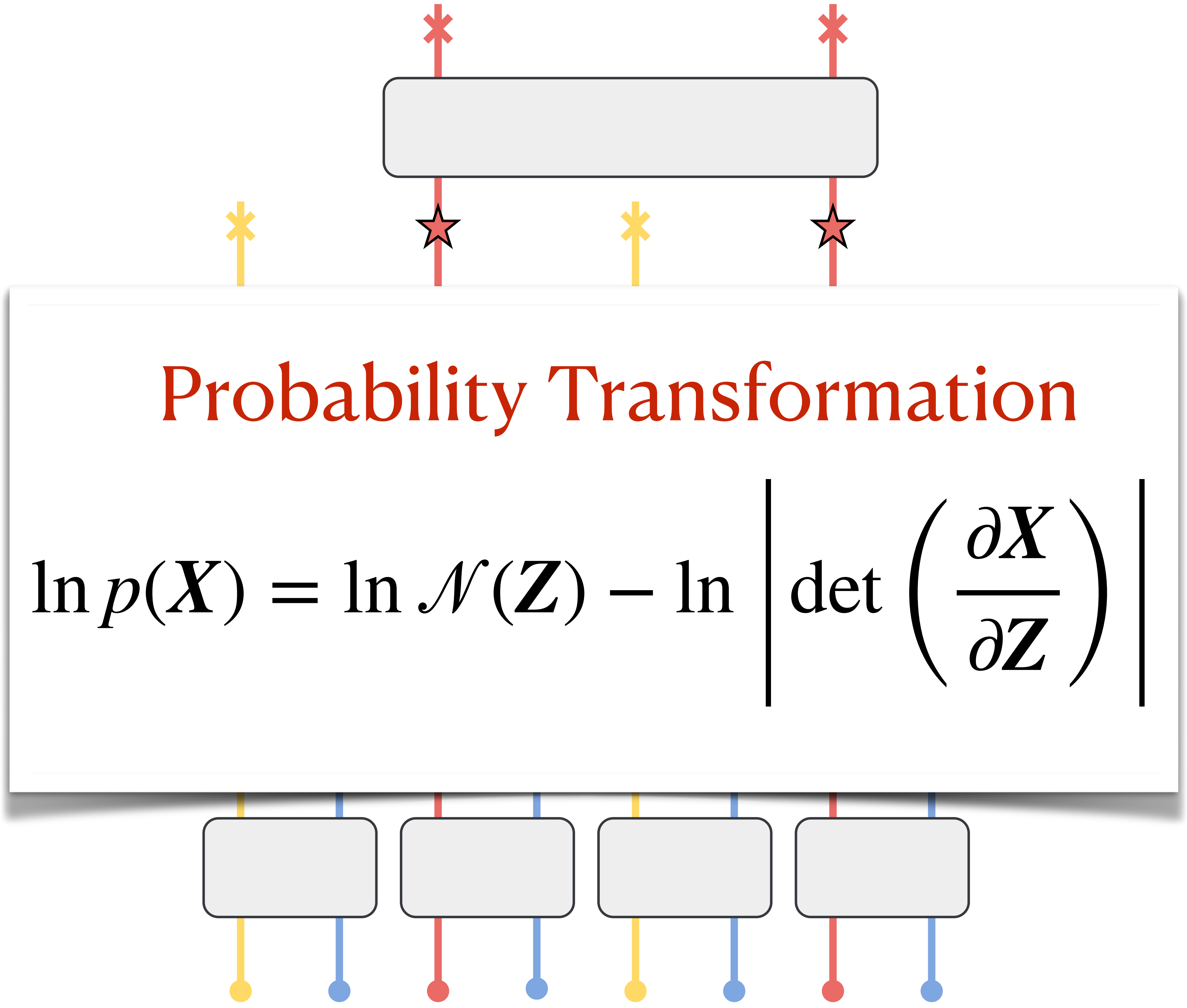


High-dimensional, nonlinear, learnable, composable transformations

Neural network renormalization group

Li, LW, PRL '18 [li012589/NeuralRG]

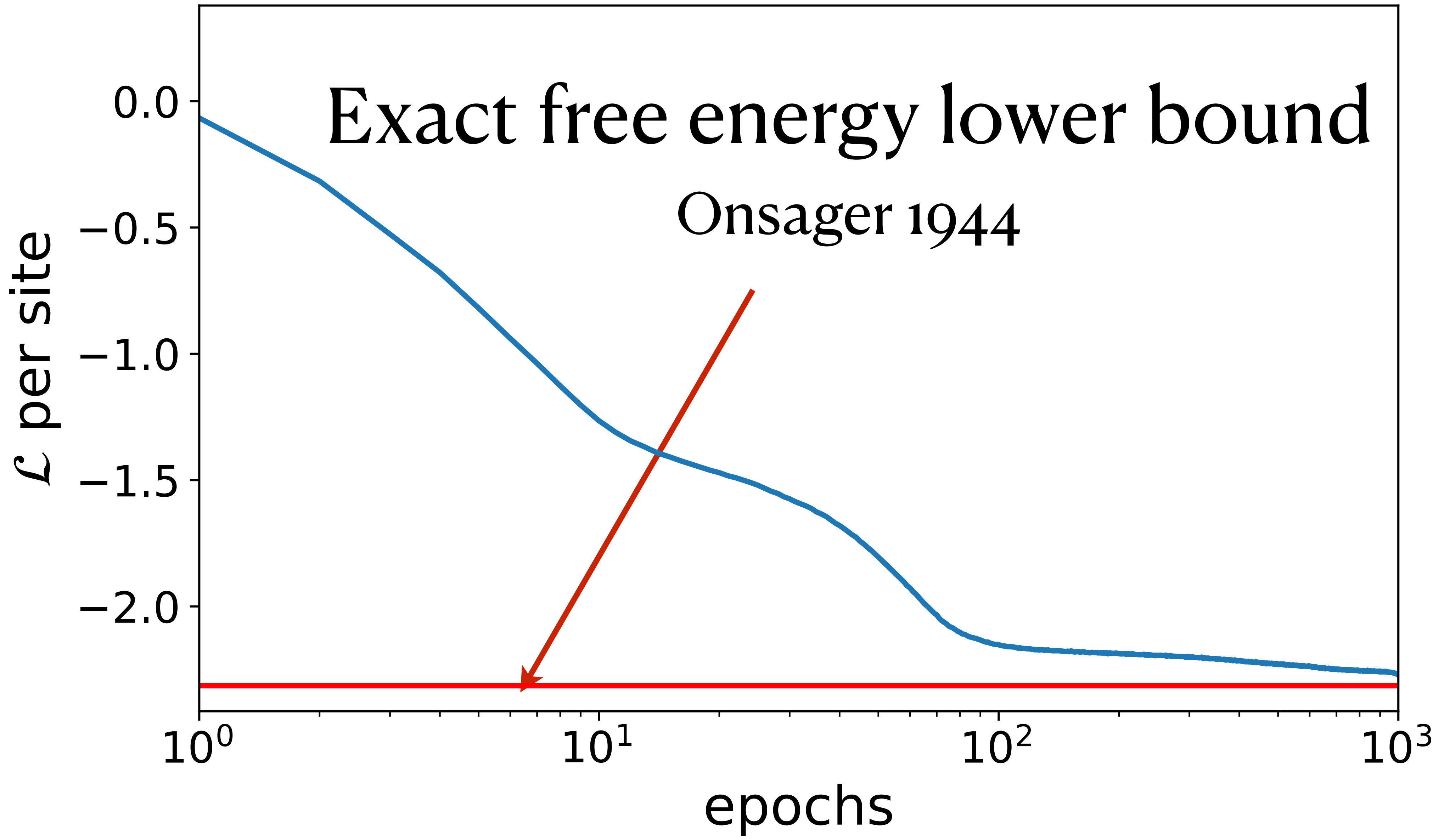
Collective variables

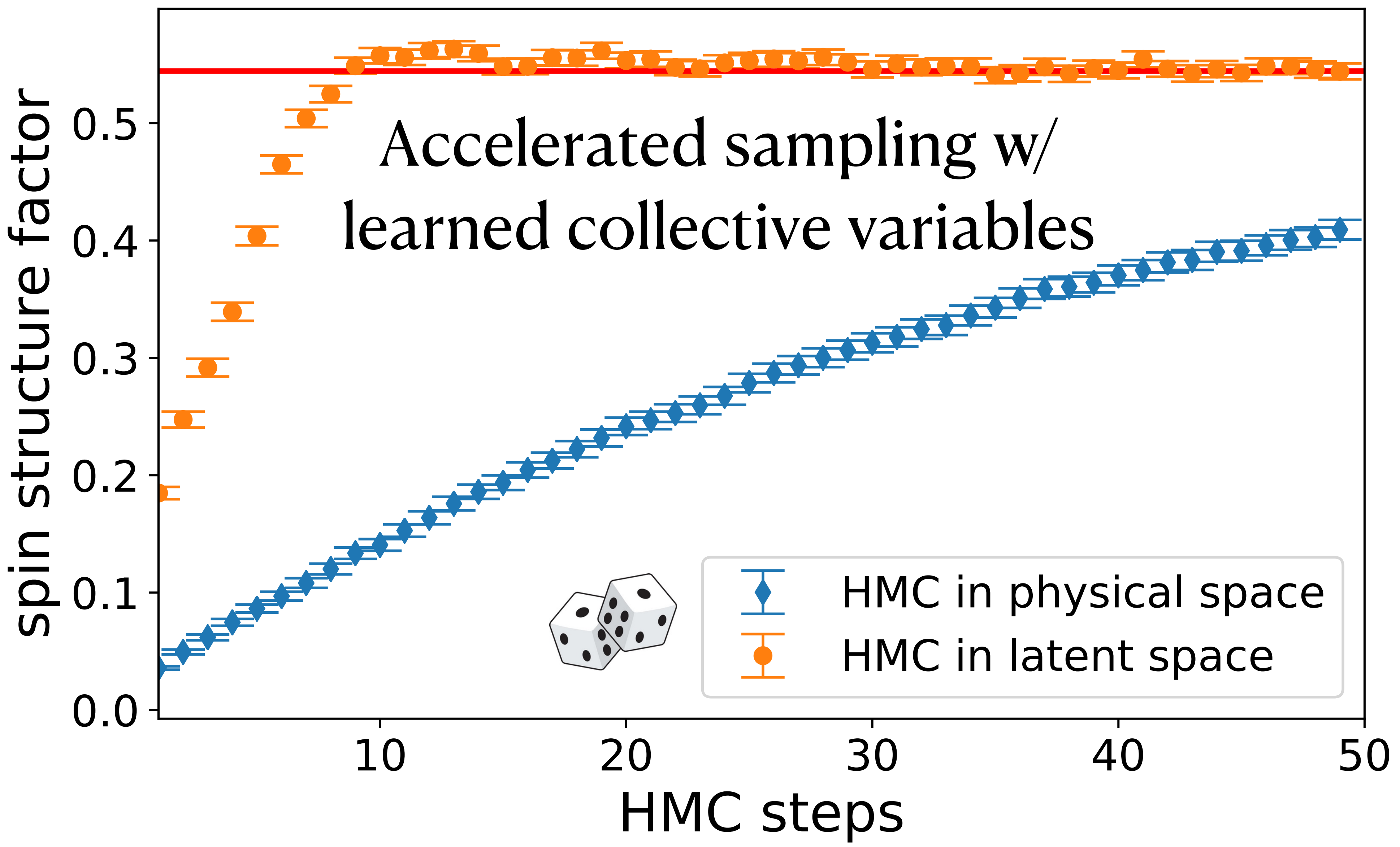


Probability Transformation

$$\ln p(\boldsymbol{X}) = \ln \mathcal{N}(\boldsymbol{Z}) - \ln \left| \det \left(\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Z}} \right) \right|$$

Physical variables





Now, move on to the quantum case

$$Z = \mathrm{Tr}(e^{-H/k_BT})$$

Gibbs–Bogolyubov-Feynman-Delbrück–Molière variational principle

$$\min\; F[\rho] = k_BT\,\mathrm{Tr}(\rho\ln\rho) + \mathrm{Tr}(H\rho)\;\geq -k_BT\ln Z$$

$$\text{s.t.}\quad \mathrm{Tr}\rho = 1 \qquad \rho \succ 0 \qquad \rho^\dagger = \rho \qquad \langle \boldsymbol{X}|\rho|\boldsymbol{X}'\rangle = (-)^{\mathscr{P}}\langle \mathscr{P}\boldsymbol{X}|\rho|\boldsymbol{X}'\rangle$$

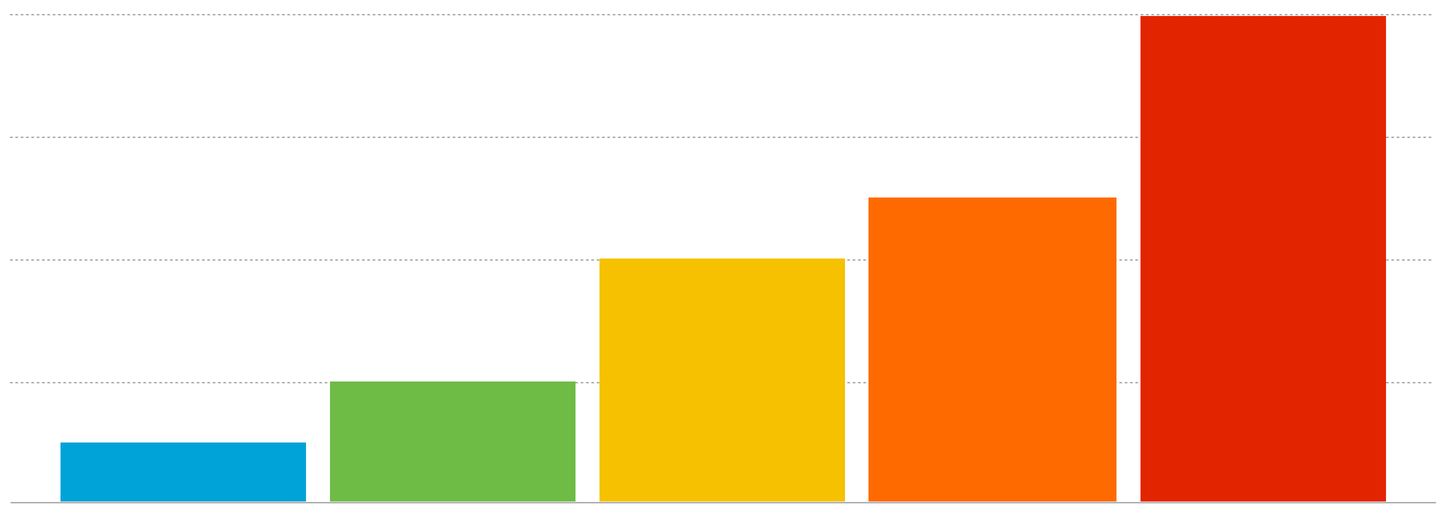
Q: How to parametrize ρ ?

A: Use TWO deep generative models !!

Variational density matrix

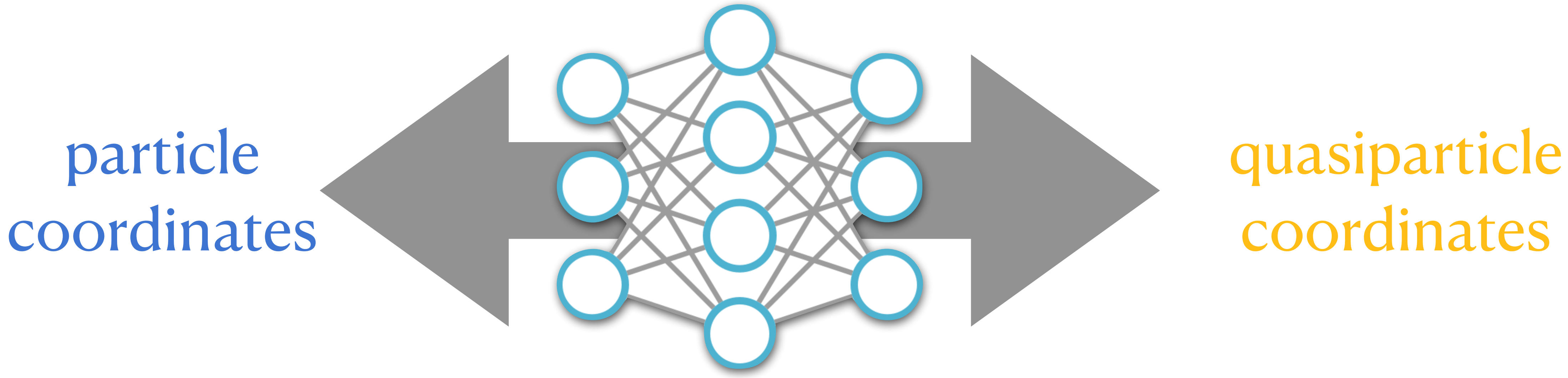
$$\rho = \sum_n p_n |\Psi_n\rangle\langle\Psi_n|$$

Classical probability p_n



Discrete probabilistic models
e.g. an autoregressive model

Quantum state basis $|\Psi_n\rangle$



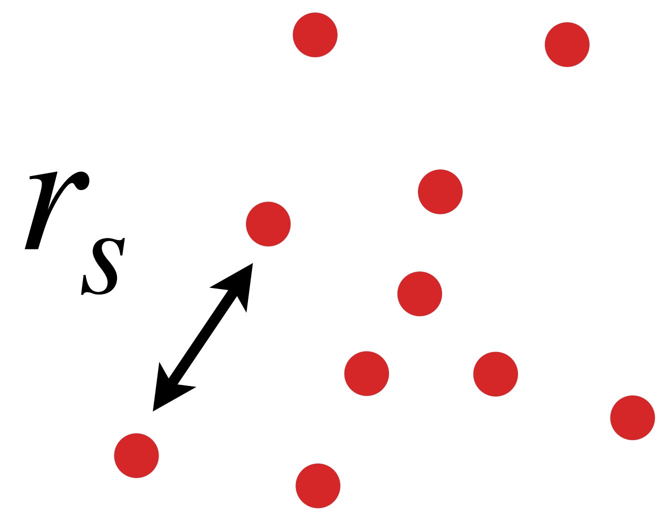
$\sqrt{\text{Normalizing flow}}$

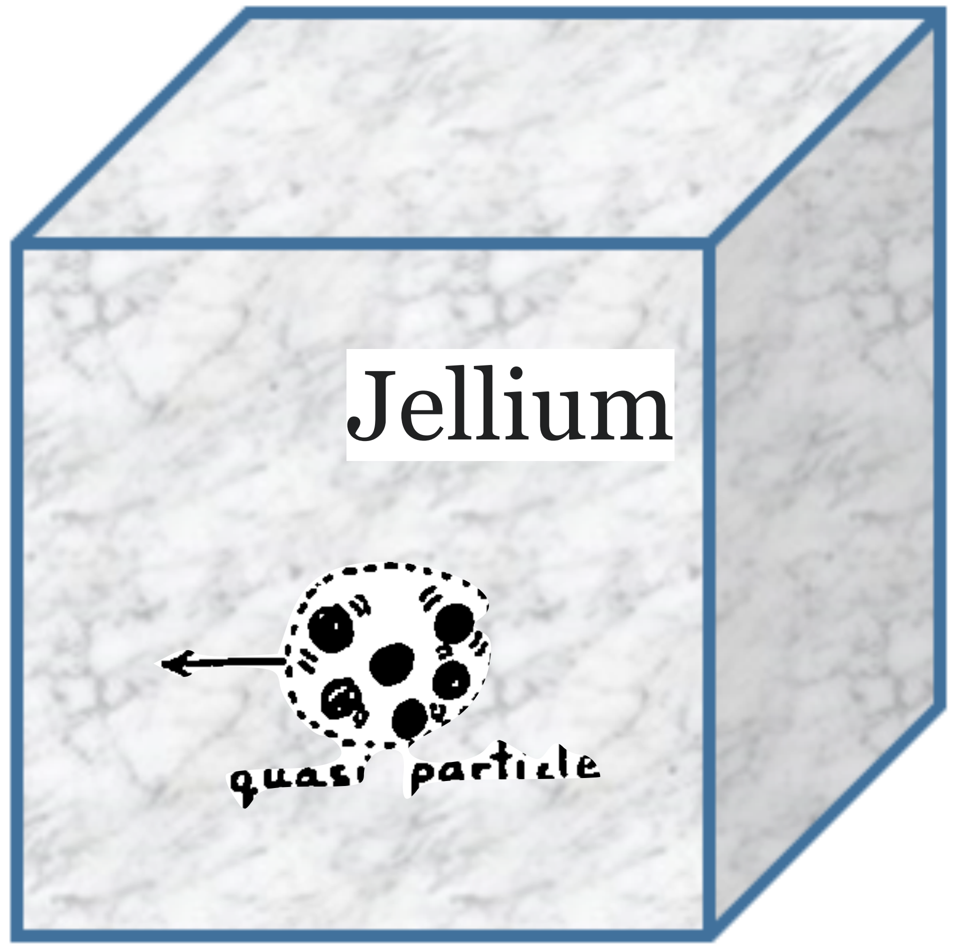
Xie, Zhang, LW, JML '22

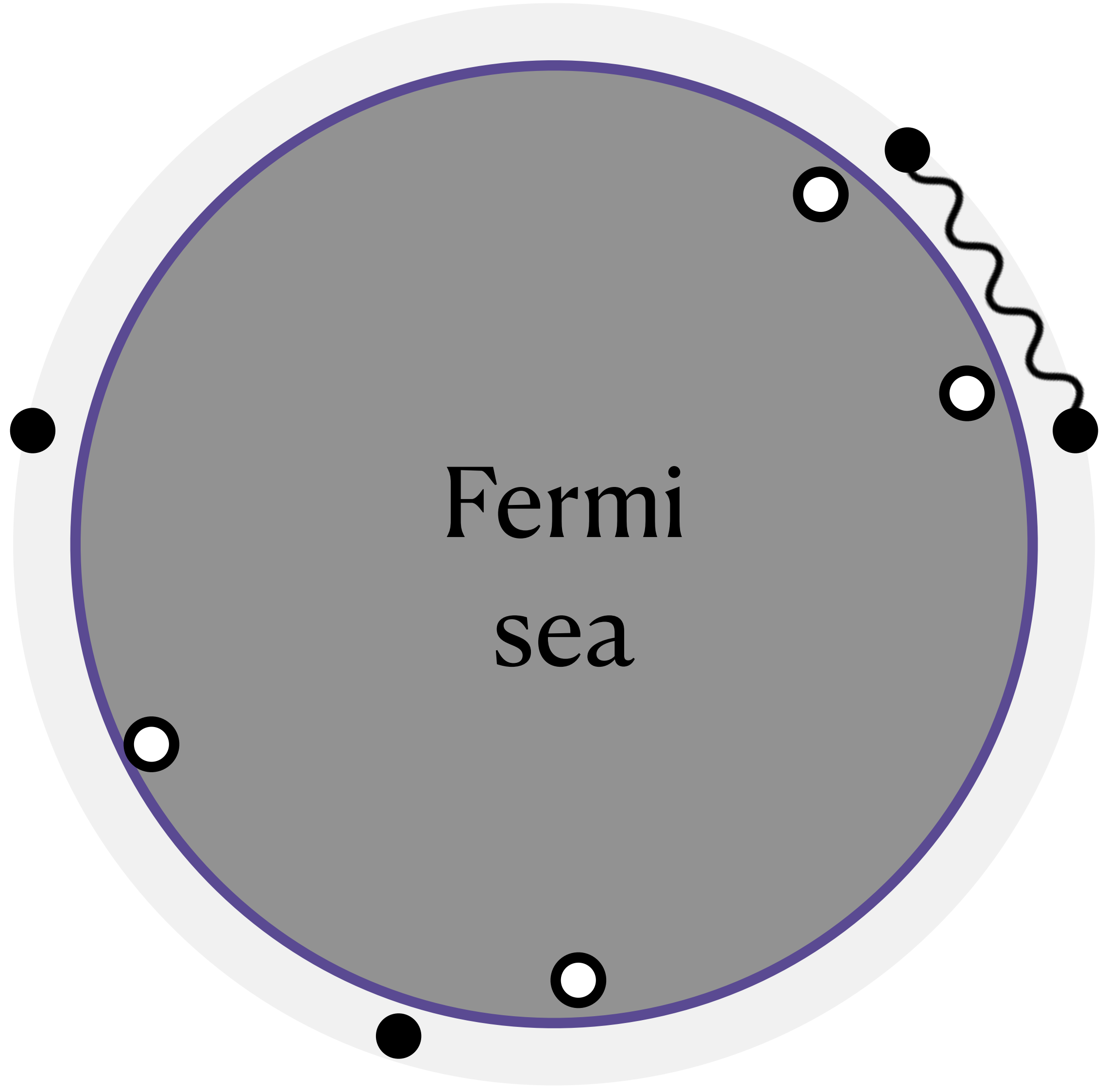
Example: uniform electron gas

Xie, Zhang, LW,
2201.03156, SciPost ‘23

$$H = -\sum_{i=1}^{N} \frac{\hbar^2 \nabla_i^2}{2m} + \sum_{i<j} \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$$







$$T \ll T_F \lesssim \frac{e^2}{r_s}$$

Fundamental model in condensed matter physics: metals $2 < r_s < 6$

Low energy excited states labeled in the same way as ideal Fermi gas $\boldsymbol{K} = \{\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_N\}$

Deep generative models for the variational density matrix

$$\rho = \sum_{\boldsymbol{K}} p(\boldsymbol{K}) \big| \Psi_{\boldsymbol{K}} \rangle \langle \Psi_{\boldsymbol{K}} \big|$$

Normalized probability distribution

Orthonormal many-electron basis

❶ $\sum_{\boldsymbol{K}} p(\boldsymbol{K}) = 1$

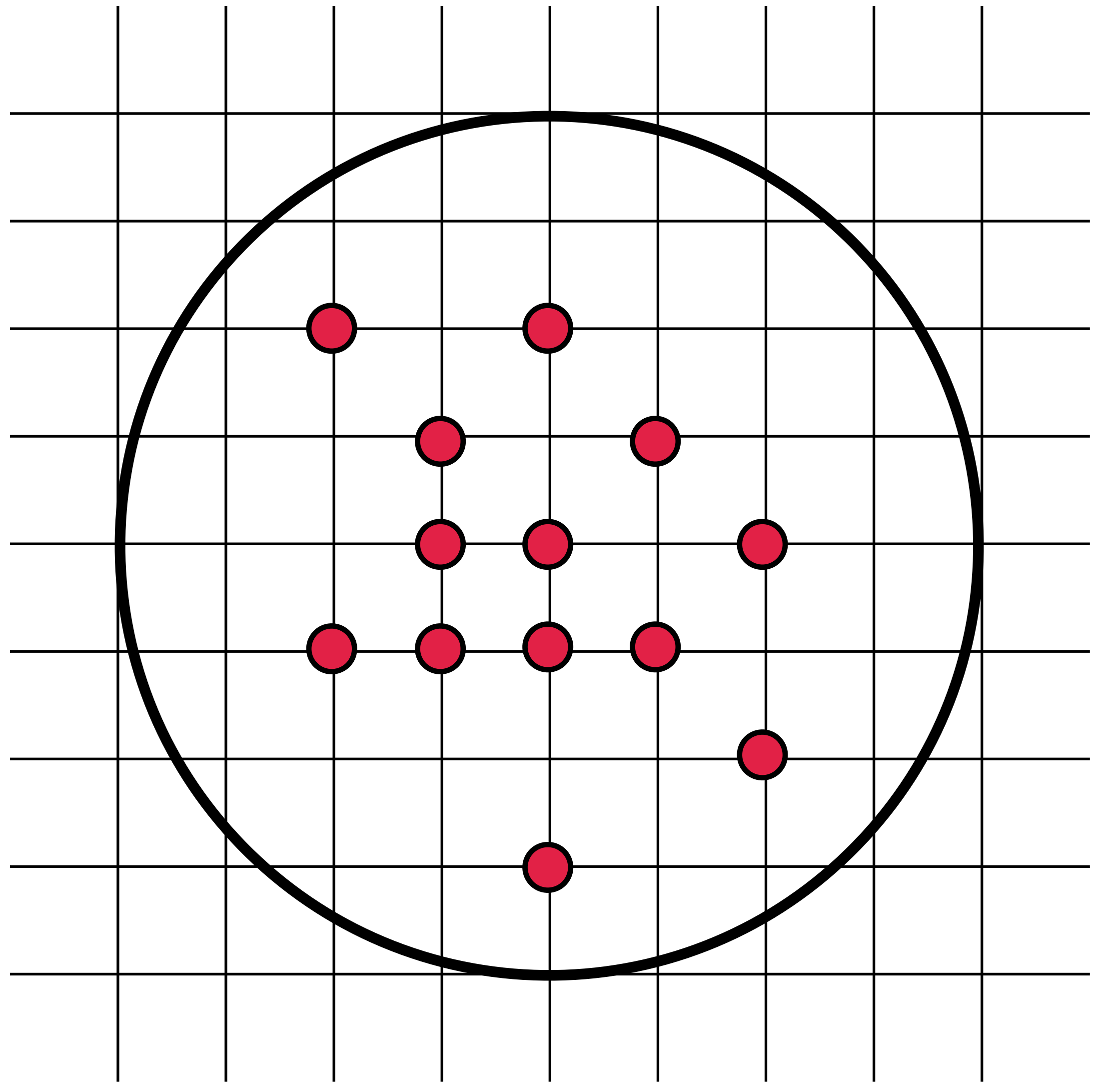
❷ $\langle \Psi_{\boldsymbol{K}} | \Psi_{\boldsymbol{K}'} \rangle = \delta_{\boldsymbol{K},\boldsymbol{K}'}$

Design deep generative models with physics constraints

❶ Autoregressive model for $p(\boldsymbol{K})$

Fermionic occupation in k-space

$$p(\boldsymbol{K}) = p(\boldsymbol{k}_1)p(\boldsymbol{k}_2\,|\,\boldsymbol{k}_1)p(\boldsymbol{k}_3\,|\,\boldsymbol{k}_1,\boldsymbol{k}_2)\cdots$$



$\binom{M}{N}$ probability space

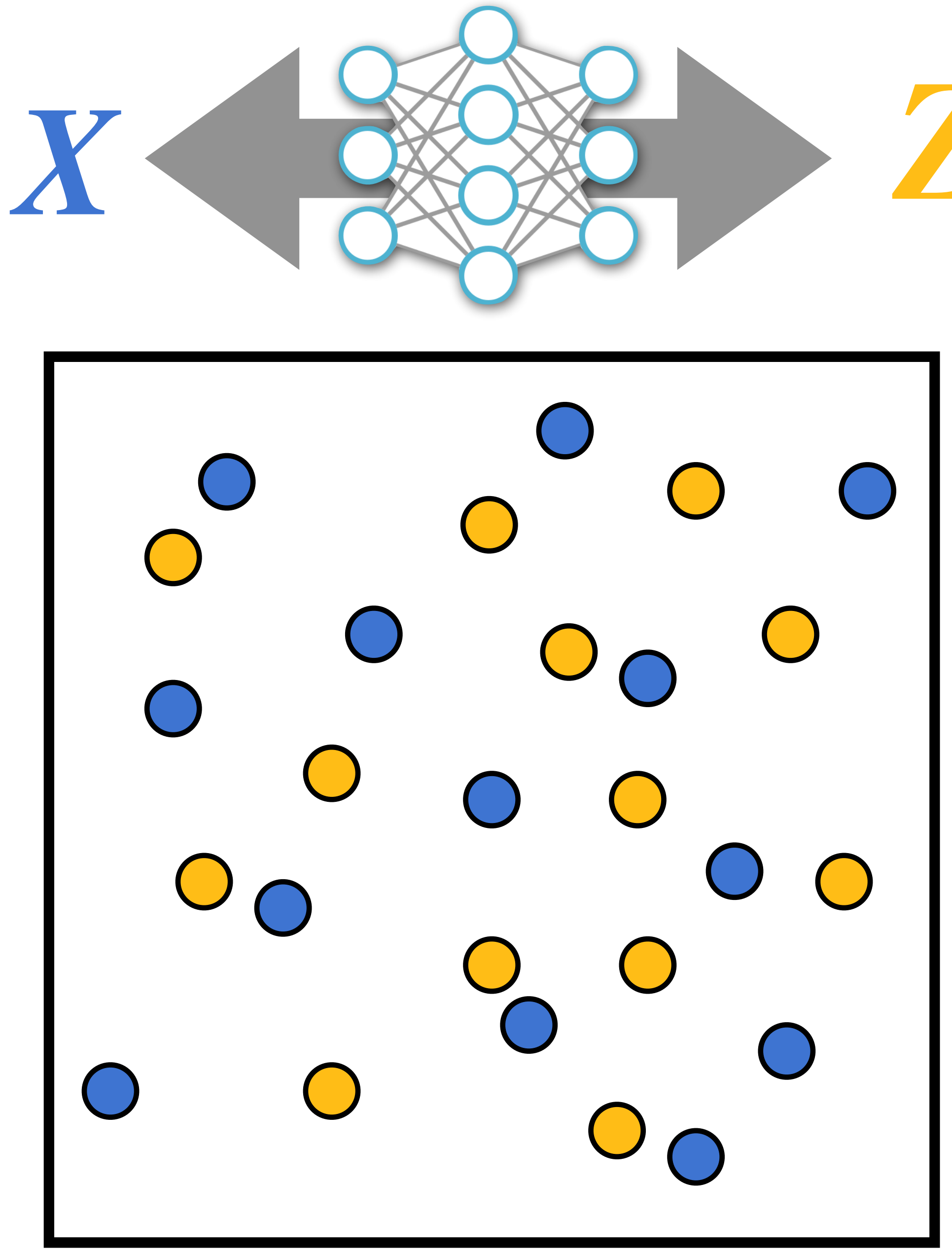
N	# of fermions	# of words
M	Momentum cutoff	Vocabulary

Pauli exclusion: we are modeling a *set of words* with no repetitions and no order

We use masked casual self-attention Vaswani et al 1706.03762; Alternative solution: Hibat-Allah et al, 2002.02793, Barrett et al, 2109.12606

❷ $\sqrt{\text{Normalizing flow}}$ for $|\Psi_K\rangle$

Electron coordinates X



Z Quasi-particle coordinates

$$\Psi_K(X) = \frac{\det(e^{ik_i \cdot z_j})}{\sqrt{N!}} \cdot \left| \det\left(\frac{\partial Z}{\partial X}\right) \right|^{\frac{1}{2}}$$

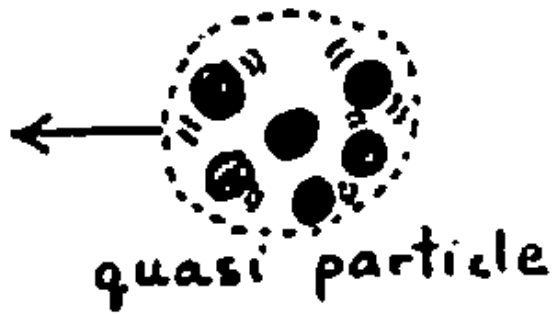
Orthonormal many-body states

Jacobian of the transformation

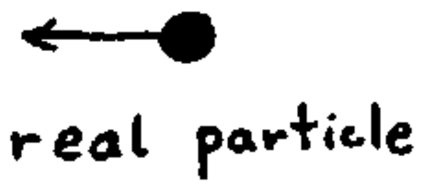
Fermion statistics: the flow should be permutation equivariant

we use FermiNet layer Pfau et al, 1909.02487

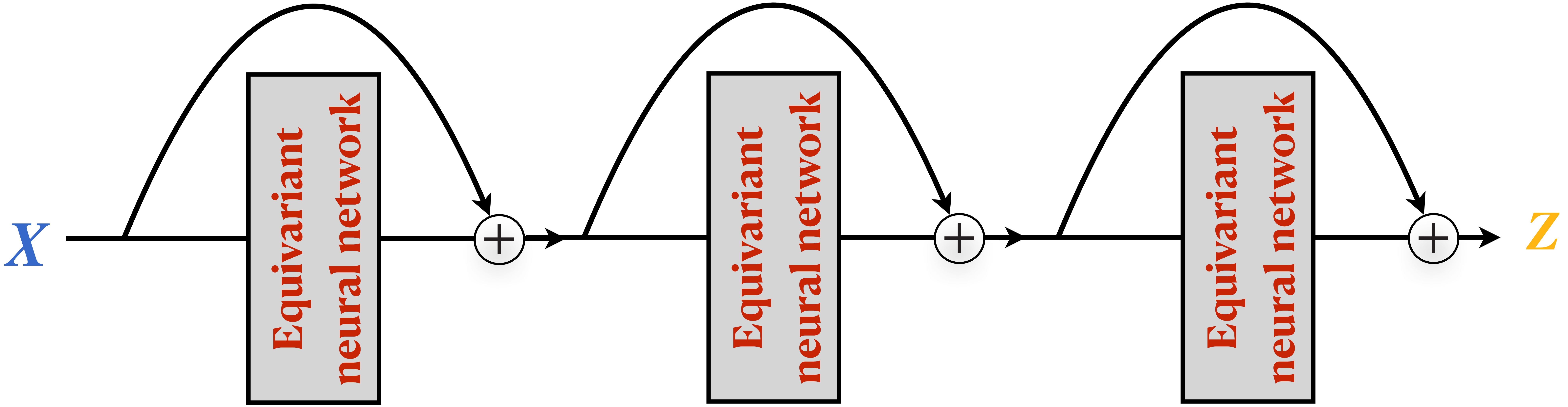
Feynman's backflow in the deep learning era



$$z_i = \boldsymbol{x}_i + \sum_{j \neq i} \eta(|\boldsymbol{x}_i - \boldsymbol{x}_j|)(\boldsymbol{x}_j - \boldsymbol{x}_i)$$



Feynman & Cohen 1956
wavefunction for liquid Helium



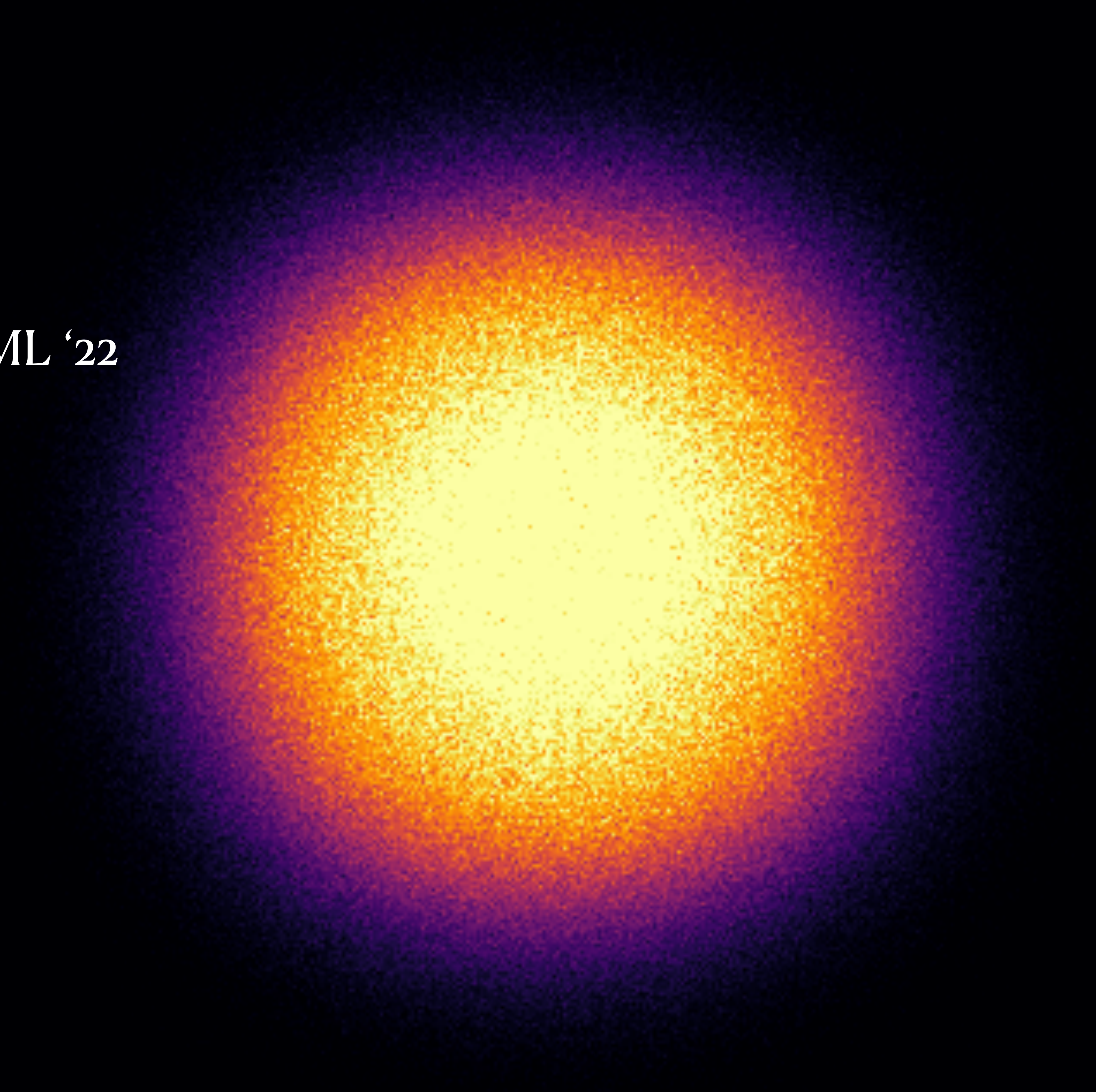
Iterative backflow → deep residual network → continuous normalizing flow

Taddei et al, PRB '15 E Commun. Math. Stat 17', Harbor el al 1705.03341, Lu et al 1710.10121, Chen et al, 1806.07366

Fermi Flow

Xie, Zhang, LW, 2105.08644, JML ‘22

github.com/fermiflow



Continuous flow of electron density in a quantum dot

The objective function

$$F = \underset{\boldsymbol{K}\sim p(\boldsymbol{K})}{\mathbb{E}} \left[k_B T \ln p(\boldsymbol{K}) + \underset{\boldsymbol{X}\sim \left|\langle \boldsymbol{X} | \Psi_{\boldsymbol{K}} \rangle\right|^2}{\mathbb{E}} \left[\frac{\langle \boldsymbol{X} | H | \Psi_{\boldsymbol{K}} \rangle}{\langle \boldsymbol{X} | \Psi_{\boldsymbol{K}} \rangle} \right] \right]$$

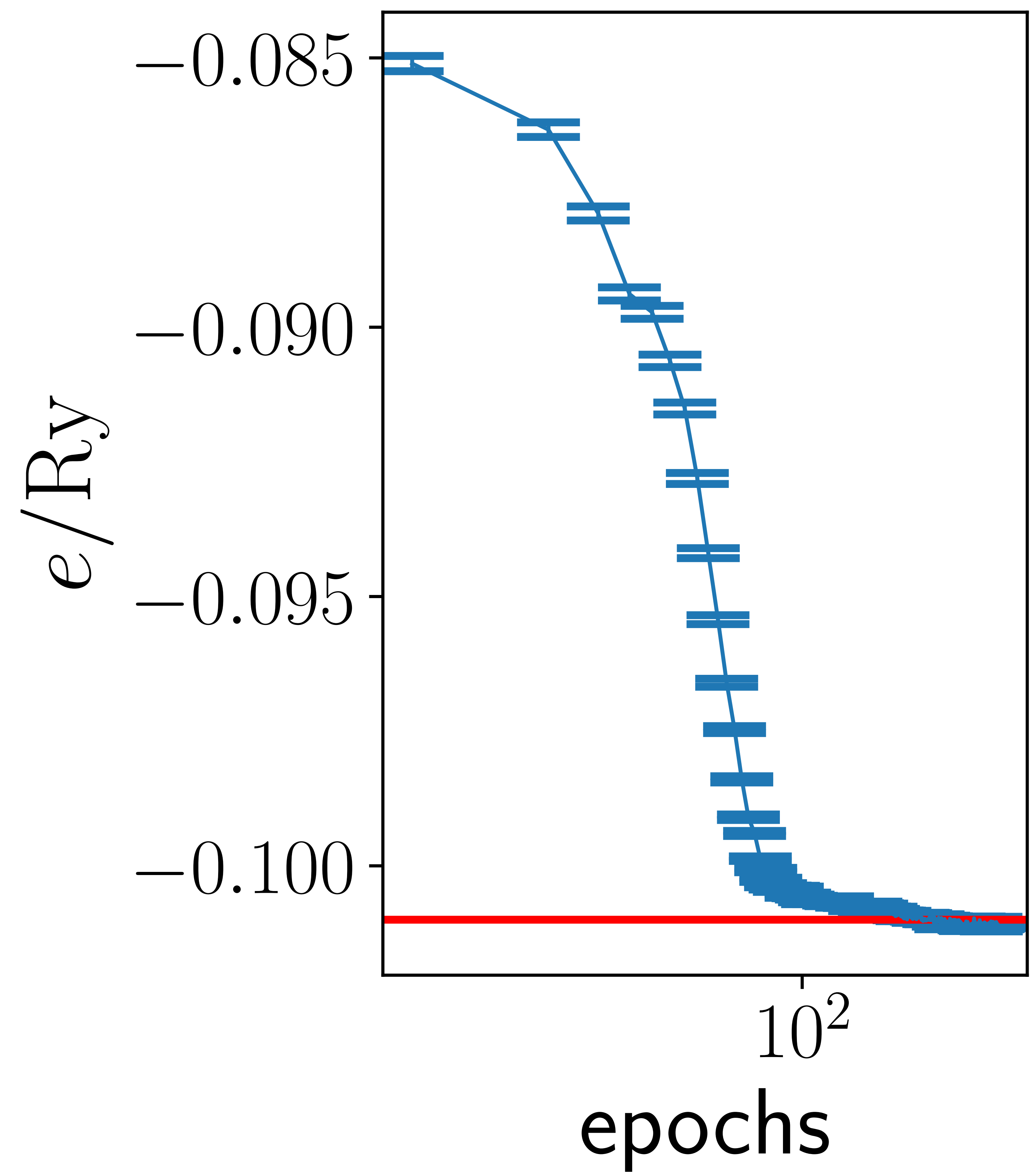
$\downarrow$ Boltzmann distribution

$\downarrow$ Born probability

Jointly optimize $|\Psi_{\boldsymbol{K}}\rangle$ and $p(\boldsymbol{K})$ to minimize the variational free energy

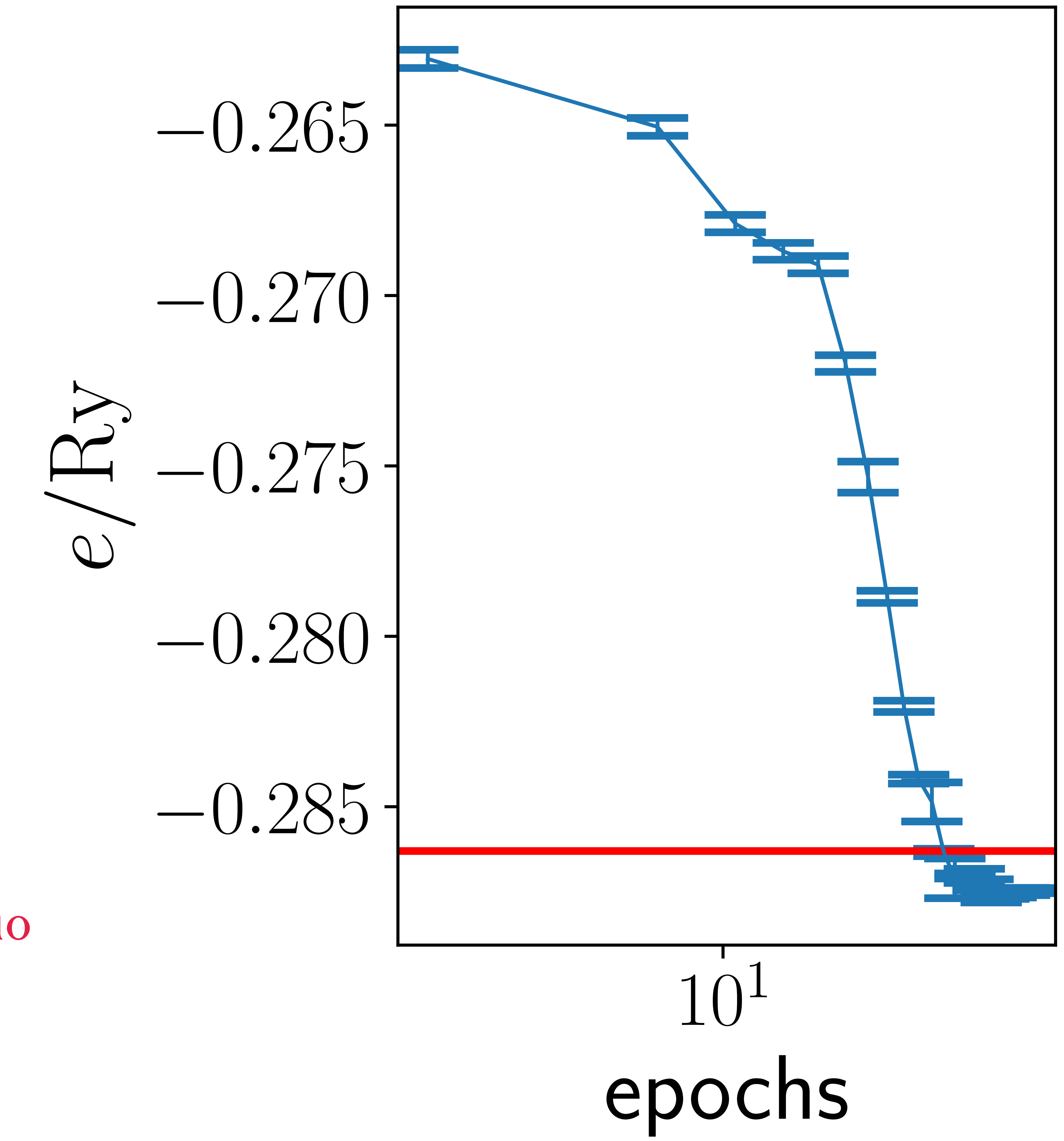
Benchmarks on spin-polarized electron gases

3D electron gas T/T_F=0.0625



Brown et al, PRL '13
restricted PIMC N=33, r_s=10

2D electron gas T=0



Tanatar, Ceperley, PRB, '89
Slater-Jastrow VMC N=37, r_s=5

Application: m^* from low temperature entropy

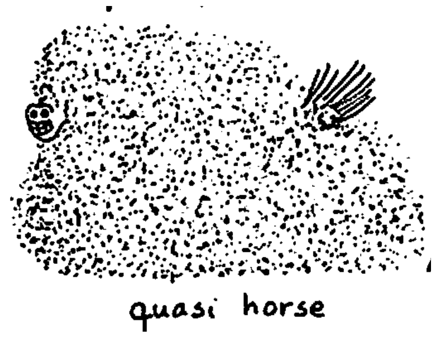
Eich, Holzmann, Vignale, PRB '17

$$s = \frac{\pi^2 k_B}{3} \frac{m^*}{m} \frac{T}{T_F}$$

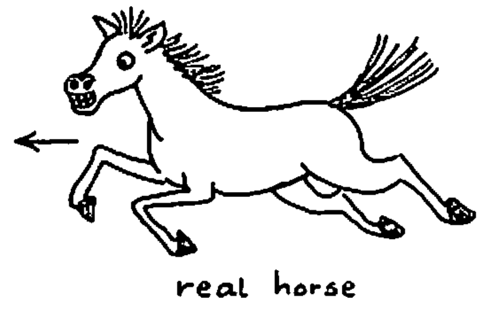
Richard D. Mattuck
A Guide to Feynman Diagrams in the Many-body Problem

$$\Rightarrow \frac{m^*}{m} = \frac{s}{s_0}$$

interacting electrons



noninteracting electrons



A fundamental quantity appears in nearly all physical properties of a Fermi liquid

Has been some debate despite its fundamental role and long history of research

Two-dimensional electron gas experiments

VOLUME 91, NUMBER 4 PHYSICAL REVIEW LETTERS week ending 25 JULY 2003

Spin-Independent Origin of the Strongly Enhanced Effective Mass in a Dilute 2D Electron System

A. A. Shashkin,* Maryam Rahimi, S. Anissimova, and S.V. Kravchenko
Physics Department, Northeastern University, Boston, Massachusetts 02115, USA

V.T. Dolgopolov
Institute of Solid State Physics, Chernogolovka, Moscow District 142432, Russia

T. M. Klapwijk
Department of Applied Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands
(Received 13 January 2003; published 24 July 2003)

$m^*/m > 1$

PRL **101**, 026402 (2008) PHYSICAL REVIEW LETTERS week ending 11 JULY 2008

Effective Mass Suppression in Dilute, Spin-Polarized Two-Dimensional Electron Systems

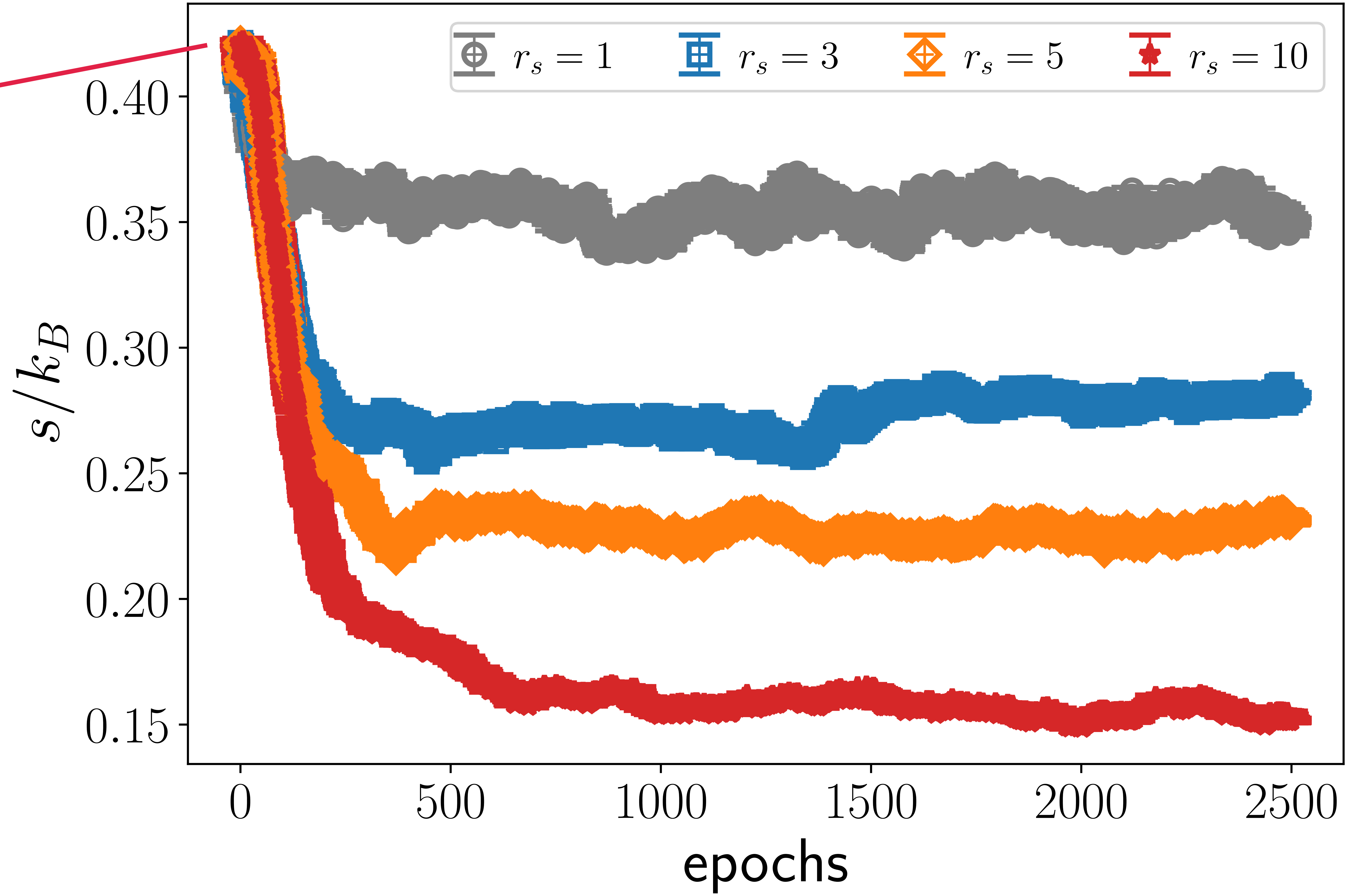
Medini Padmanabhan, T. Gokmen, N. C. Bishop, and M. Shayegan
Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA
(Received 19 September 2007; published 7 July 2008)

$m^*/m < 1$

Layer thickness, valley, disorder, spin-orbit coupling...

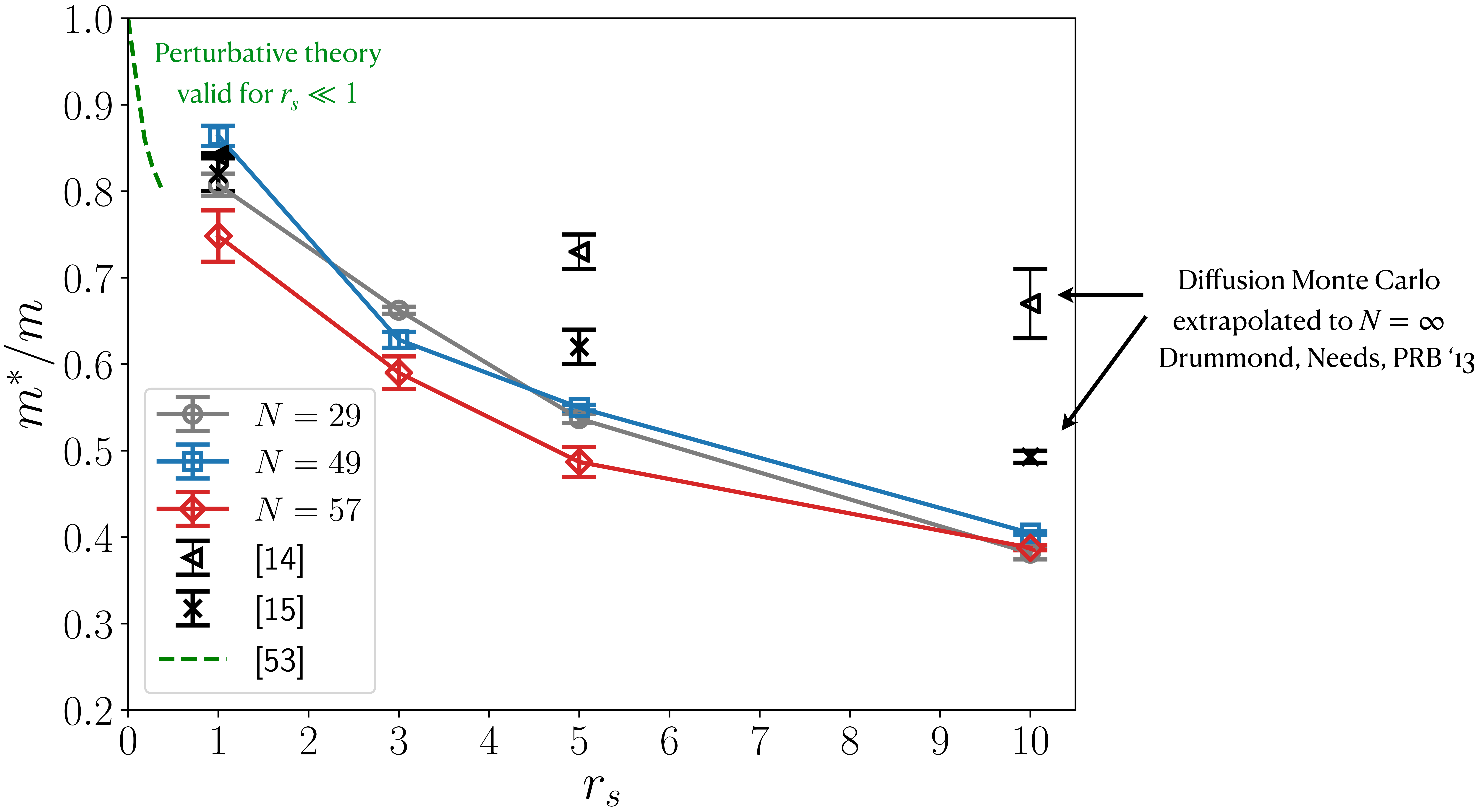
37 spin-polarized electrons in 2D @ T/T_F=0.15

Starting point:
Ideal gas entropy s_0



$$\frac{m^*}{m} = \frac{s}{s_0} < 1$$

Effective mass of spin-polarized 2DEG



More pronounced suppression of m^* in the low-density strong-coupling region

FAQs

Where to get training data ?

No training data. Data are self-generated from the generative model.

How do we know it is correct ?

Variational principle: lower free-energy is better.

Do I understand the “black box” model ?

a) I don’t care (as long as it is sufficiently accurate).

b) $\ln p(\boldsymbol{K})$ contains the Landau energy functional

$$E[\delta n_k] = E_0 + \sum_k \epsilon_k \delta n_k + \frac{1}{2} \sum_{k,k'} f_{k,k'} \delta n_k \delta n_{k'}$$

$\boldsymbol{Z} \leftrightarrow \boldsymbol{X}$ vividly illustrates adiabatic continuity.

Thank you!

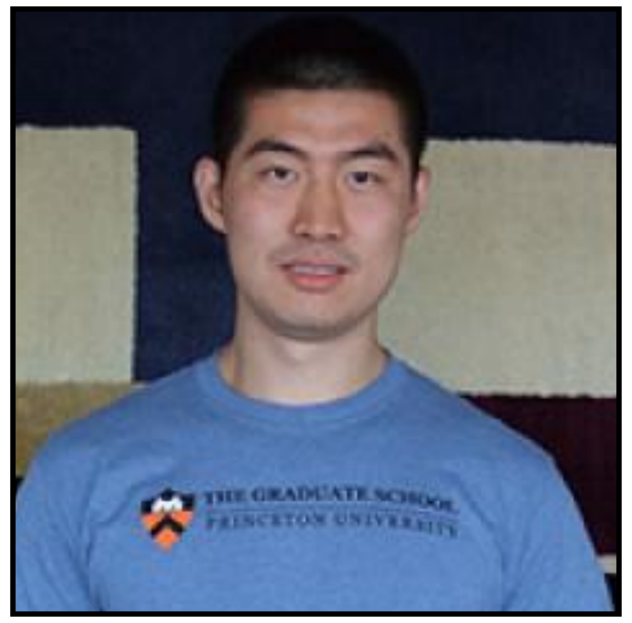
Thanks to deep generative models, the variational free-energy principle has become a practical computational tool for T>0 quantum matter

Shuo-Hui Li
IOP →HKUST

Dian Wu
PKU →EPFL

Hao Xie
IOP →UZH

Pan Zhang
ITP



Linfeng Zhang
DP/AISI

1802.02840, PRL '18
1809.10606, PRL '19
2105.08644, JML '22
2201.03156, SciPost Physics '23

li012589/NeuralRG
wdphy16/stat-mech-van
fermiflow/fermiflow
fermiflow/CoulombGas