ANTN: Bridging Autoregressive Neural Networks and Tensor Networks for Quantum Many-Body Simulation

Zhuo Chen[12], Laker Newhouse[1*], Eddie Chen[1*], Di Luo[123†], Marin Soljačić[12]

Dec. 12, 2023

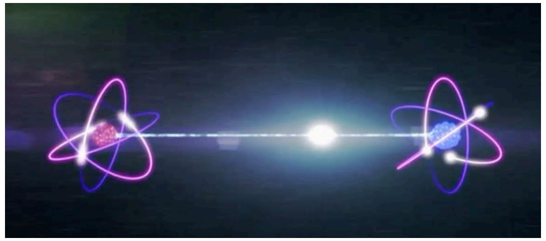
[1]Massachusetts Institute of Technology
[2]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
[3]Harvard University

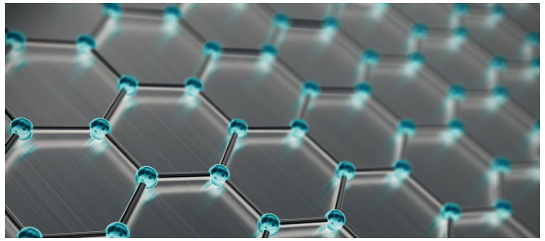
{chenzhuo,lakern,ezchen,diluo,soljacic}@mit.edu *equal contribution †corresponding author

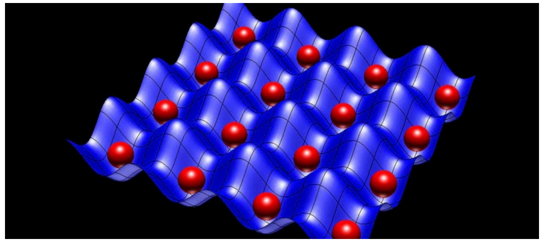
Quantum Many-Body Physics is Important but Challenging

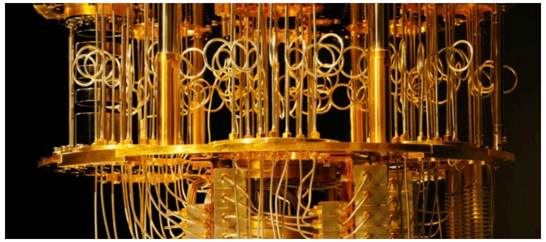
Crucial to modern quantum technologies

- Understand complex quantum interactions
- Study and develop new quantum materials
- Design quantum computers and other quantum devices









Difficult to study

- Curse of dimensionality
  - Exponential cost
- Complex-valued sign or phase structure
  - Sign (phase) problem

Single classical bit	**Single qubit**
0 or 1	$\alpha\vert 0\rangle + \beta\vert 1\rangle$
N **classical bits**	***N*** **qubits**
$b = b_0 b_1 \cdots b_{N-1}$	$\Sigma_{b=0}^{2^N} \psi(b)\vert b\rangle$

$\psi(b)$: complex-valued distribution(wavefunction) with 2^N components (intractable)

Question: How to efficiently represent $\boldsymbol{\psi(b)}$?

Existing Methods

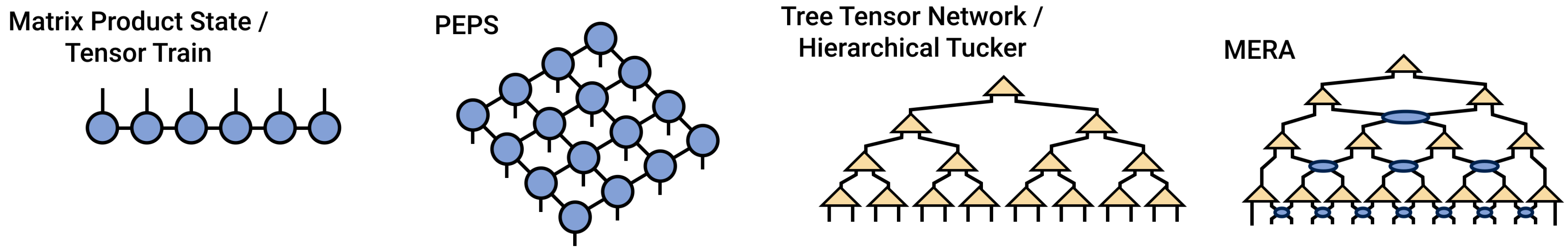
Tensor Network (TN)

- Matrix product state (MPS), Tensor Train
- Projected entangled pair state (PEPS)
- Tree tensor network (TTN)
- Multi-scale entanglement renormalization ansatz (MERA)

Pro:
- Physics prior/inductive bias
- Flexible sign structure
- Customized optimization algorithm such as DMRG

Con:
- Limited representation power

$\psi(x) =$



Low rank tensor decomposition to represent $\psi(x)$

Existing Methods

Neural Network (NN)

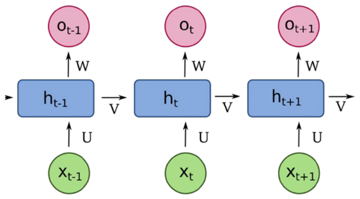
- Restricted Boltzmann machine (RBM)
- Autoregressive neural network (ARNN)
    - Recurrent neural network (RNN)
    - Pixel convolutional neural network (PixelCNN)
    - Transformer

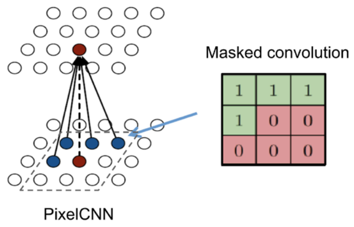
Pro:
- Expressive

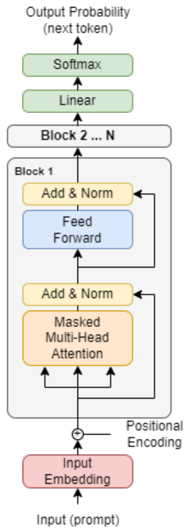
Con:
- Lack of physics prior/inductive bias
- Hard to learn sign structure

$$\psi(x) = \psi(x_0)\psi(x_1|x_0)\cdots\psi(x_N|x_{<N}) =$$

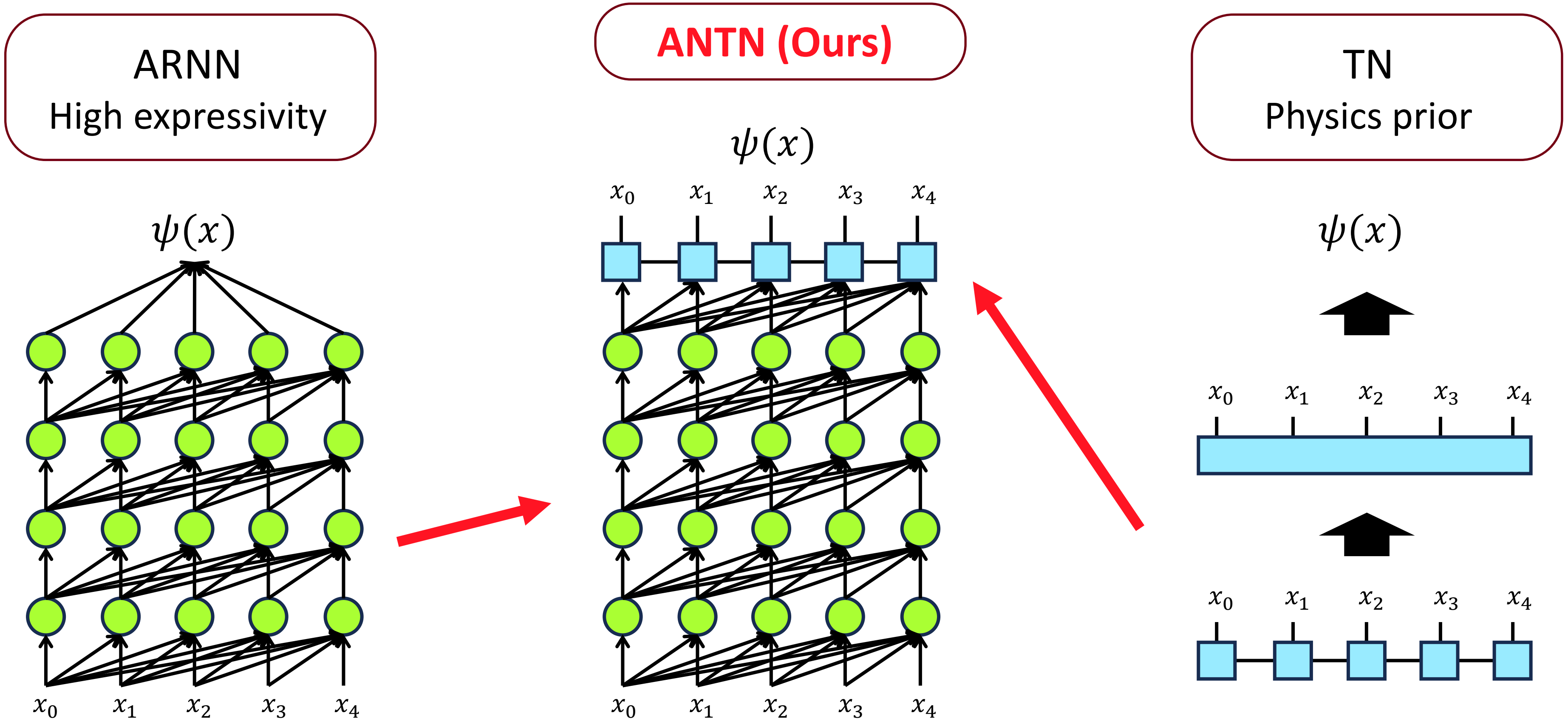
Compact NN representation of $\psi(x)$ via conditional wavefunctions







Our Method: Autoregressive Neural TensorNet (ANTN)



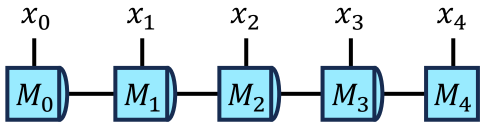
ANTN bridges ARNN and TN, achieving the best of both worlds

Detailed Construction of ANTN

Conditional probability of MPS

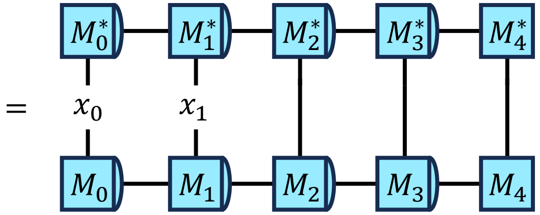
- MPS in right canonical form

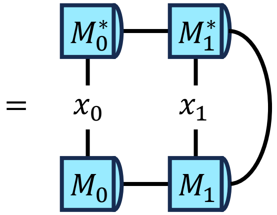
  $\psi(x_0, x_1, x_2, x_3, x_4)$



- Marginal probability

  $p(x_0, x_1)$





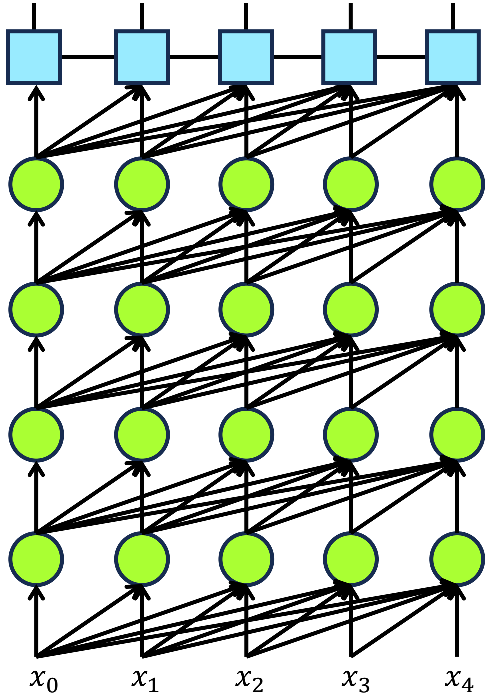
- Conditional probability

  $p(x_0|x_1) = p(x_0, x_1)/p(x_0)$

Only depends on M_0 and M_1!

Construction of ANTN

Generalize conditional wavefunction to conditional tensors $\widetilde{M}(x_i|x_{<i})$



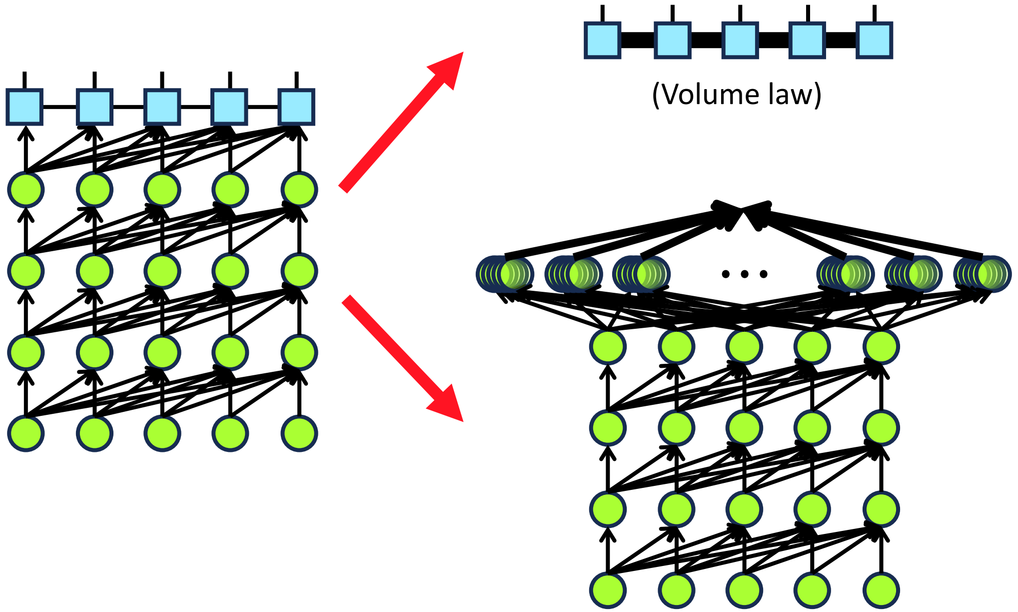
Conditional probability and phase defined analogous to MPS!

Two specific constructions

- **Elementwise construction**
  - Each tensor element $\widetilde{M}^{\alpha_{i-1}\alpha_i}(x_i|x_{<i})$ gets a unique output from the ARNN
  - Pro: flexible representation from ARNN
  - Con: higher cost → smaller maximum bond dimension
- **Blockwise construction**
  - Tensor elements share output across the bond dimension (each element still gets a unique bias)
  - Pro: reduced cost → larger maximum bond dimension
  - Con: less flexible representation

Crucial Properties of ANTN

ANTN has generalized expressivity over both TN and ARNN



ANTN can be written as either TN or ARNN with exponentially many (in system size) parameters

ANTN inherits properties from TN and ARNN

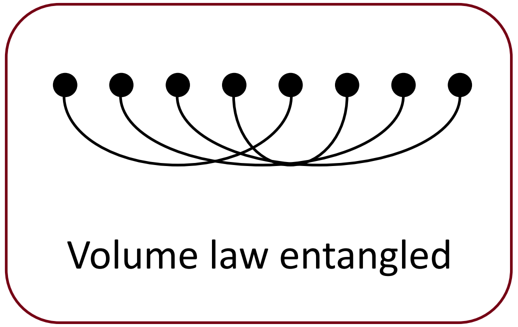
From TN:
- Exact sampling $x \sim |\psi(x)|^2$
- DMRG (initialization)
- Physics inductive bias
- Flexible sign structures

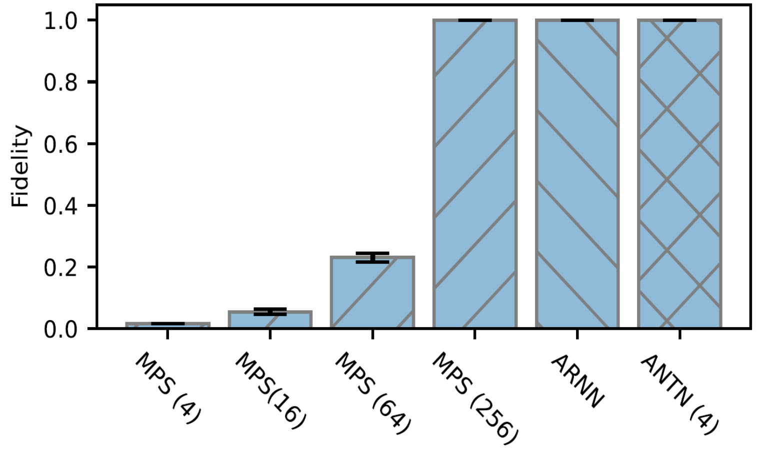
From ARNN:
- Exact sampling $x \sim |\psi(x)|^2$
- Expressivity (volume law)
- Various symmetries
    - Global $\mathrm{U}(1)$ symmetry
    - $\mathbb{Z}_2$ spin flip symmetry
    - Discrete Abelian and non-Abelian symmetries
    - ……

Numerical Experiments (Quantum State Learning)

Synthetic benchmark on expressivity

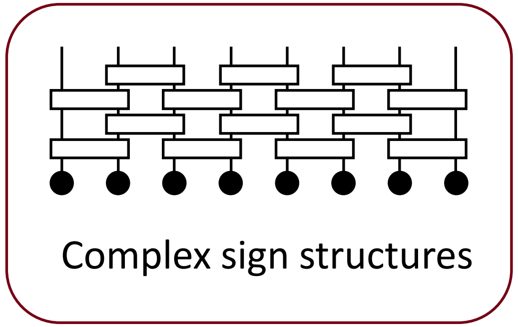
Random Bell states

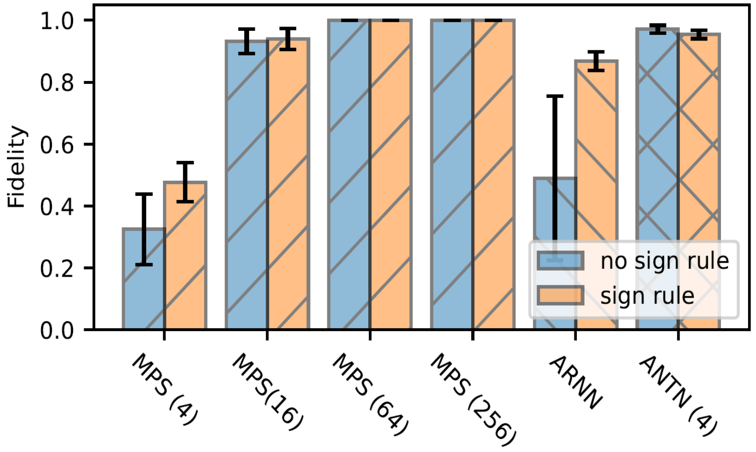




Synthetic benchmark on sign structure

Shallow random circuits



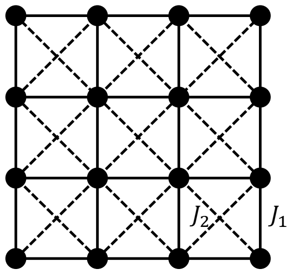


ANTN has good performance in both benchmarks

Numerical Experiments (Ground State Optimization)

Benchmark on the challenging 2D J_1-J_2 Heisenberg Model

$$\hat{\mathcal{H}} = J_1 \sum_{\langle i,j \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle i,j \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j$$



Energy per site ↓ 10 × 10							
Algorithms	$J_2 = 0.2$	$J_2 = 0.3$	$J_2 = 0.4$	$J_2 = 0.5$	$J_2 = 0.6$	$J_2 = 0.7$	$J_2 = 0.8$
MPS (8)	-1.997537	-1.893753	-1.797675	-1.734326	-1.716253	-1.768225	-1.869871
MPS (70)	-2.191048	-2.069029	-1.956480	-1.866159	-1.816249	-1.854296	-2.007845
MPS (1024)	-2.255633	-2.138591	-2.031681	*-1.938770*	**-1.865561**	**-1.894371**	**-2.062730**
PixelCNN	-2.22462(24)	-2.12873(14)	-2.02053(14)	-1.74098(29)	-1.71885(27)	-1.81800(13)	-1.98331(17)
Elementwise (8)	**-2.26034(6)**	**-2.14450(4)**	**-2.03727(7)**	**-1.94001(6)**	*-1.85684(10)*	*-1.88643(7)*	*-2.05707(43)*
Blockwise (70)	*-2.25755(8)*	*-2.14152(8)*	*-2.03319(10)*	-1.93842(42)	-1.85270(12)	-1.87853(13)	-2.05088(14)

Energy per site ↓ 12 × 12							
Algorithms	$J_2 = 0.2$	$J_2 = 0.3$	$J_2 = 0.4$	$J_2 = 0.5$	$J_2 = 0.6$	$J_2 = 0.7$	$J_2 = 0.8$
MPS (8)	-1.998207	-1.887531	-1.800784	-1.735906	-1.720619	-1.788652	-1.893916
MPS (70)	-2.185071	-2.059443	-1.944832	-1.851954	-1.812450	-1.853650	-2.030131
MPS (1024)	*-2.264084*	-2.141043	-2.027736	-1.931243	**-1.858846**	*-1.913483*	*-2.093013*
PixelCNN	-2.24311(102)	-2.12616(23)	-2.01768(21)	-1.74282(30)	-1.72637(16)	-1.85239(13)	-2.03226(59)
Elementwise (8)	**-2.27446(27)**	**-2.15537(6)**	**-2.04437(7)**	**-1.94686(6)**	*-1.85443(15)*	**-1.91391(10)**	**-2.09457(10)**
Blockwise (70)	-2.26152(50)	*-2.15395(7)*	*-2.04225(8)*	*-1.94298(43)*	-1.85176(15)	-1.90571(12)	-2.09113(43)

ANTN achieves state-of-the-art performance

Conclusion and Outlook

Conclusion:

- ANTN bridges TN and ARNN
- ANTN generalizes the expressivity of both TN and ARNN
- ANTN inherits various properties from TN and ARNN
- ANTN achieves better performance than TN and ARNN on both quantum state learning and the challenging 2D J_1-J_2 Heisenberg model

Outlook:

- Generalization of ANTN to incorporate other TNs and in higher dimensions
- Applications beyond scientific applications such as supervised learning and generative modeling

{chenzhuo,lakern,ezchen,diluo,soljacic}@mit.edu