Scalar Field Theories via Neural Networks at Initialization

Anindita Maiti

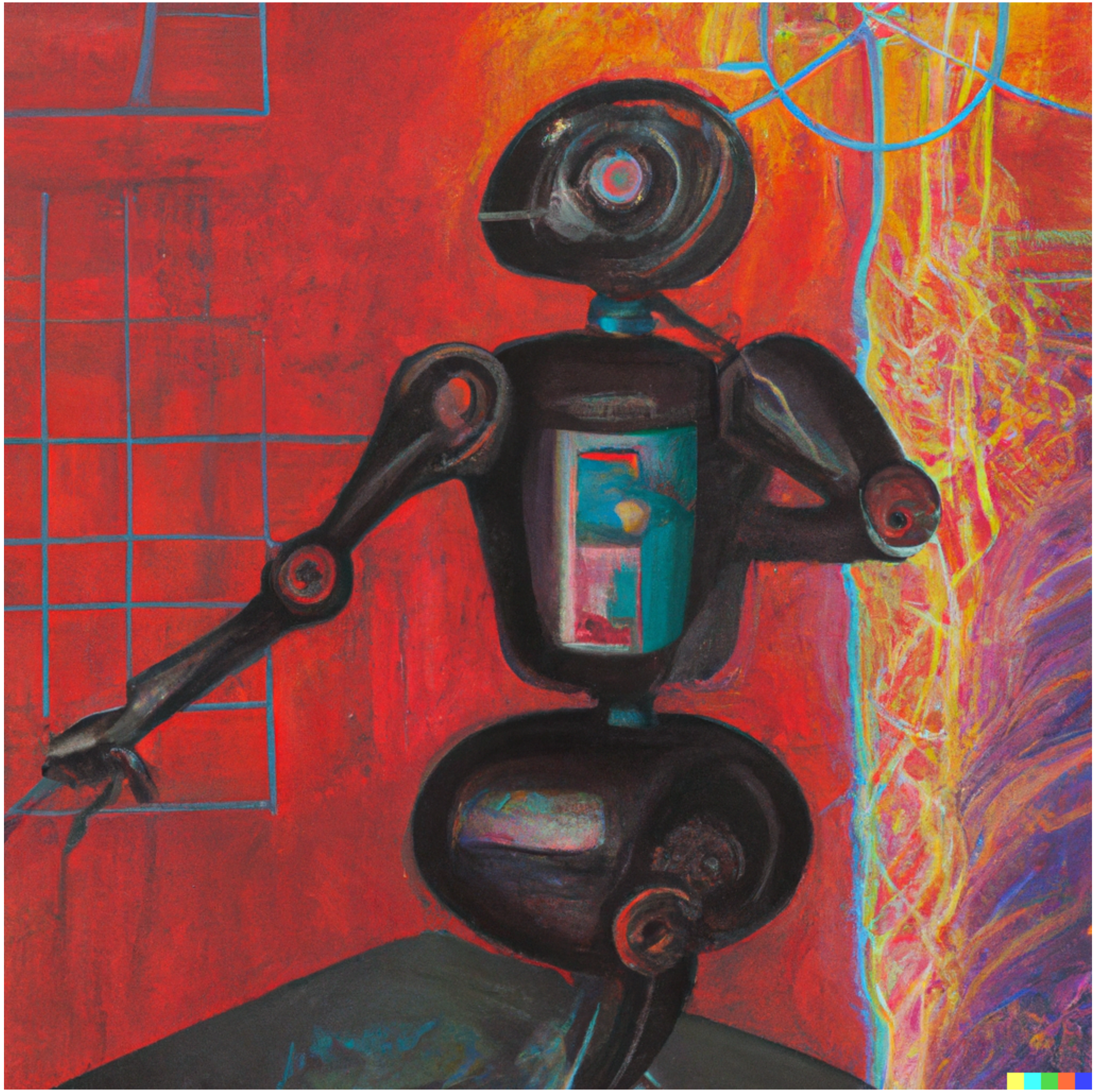
Email: amaiti@perimeterinstitute.ca

Workshop: Probing the Frontiers of Nuclear Physics with AI at the EIC

Sept 26, 2023

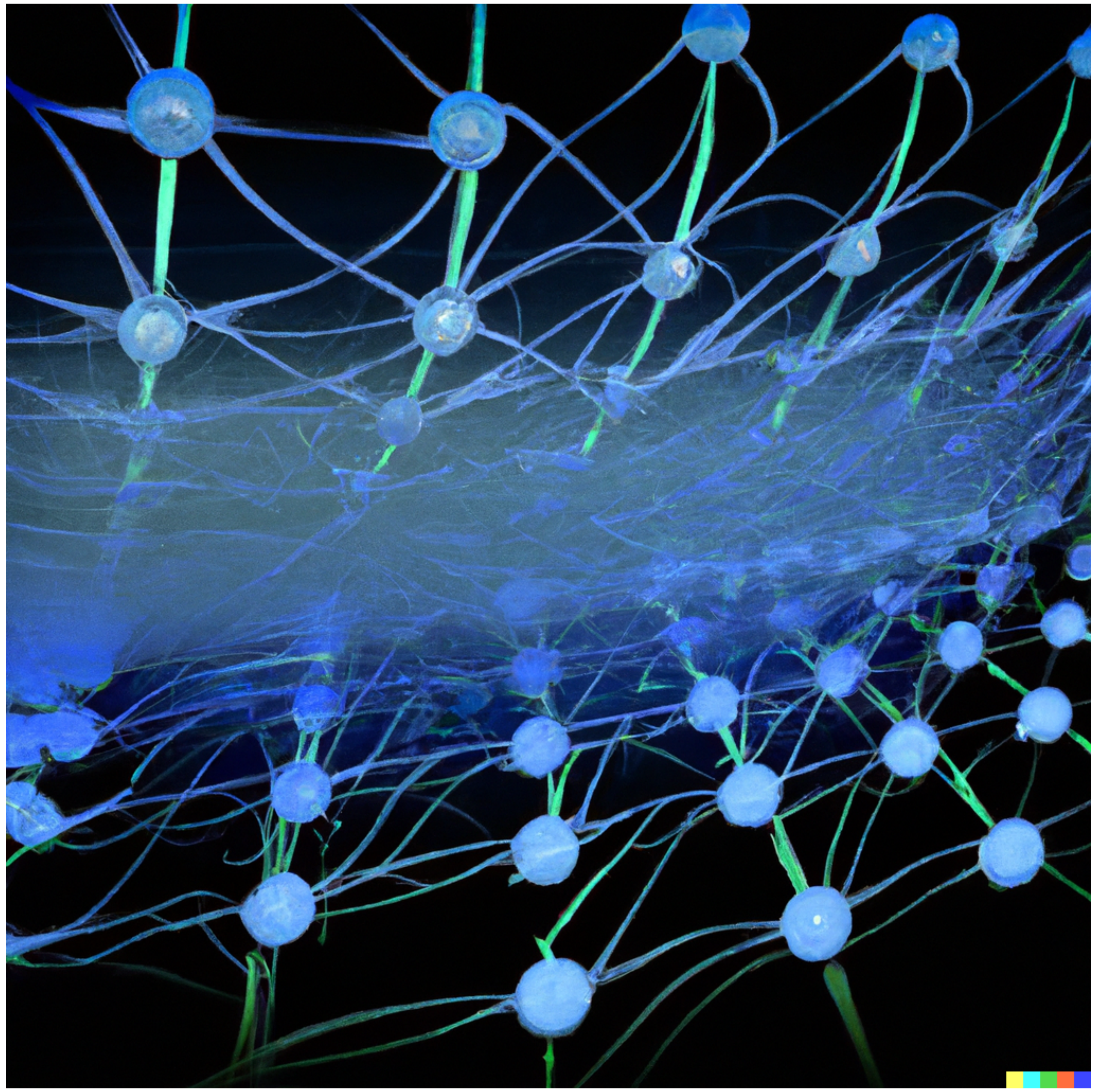
Based on [2307.03223], [2106.00694], [2008.08601] w/ Halverson, Schwartz, Demirtas, Stoner

Overview

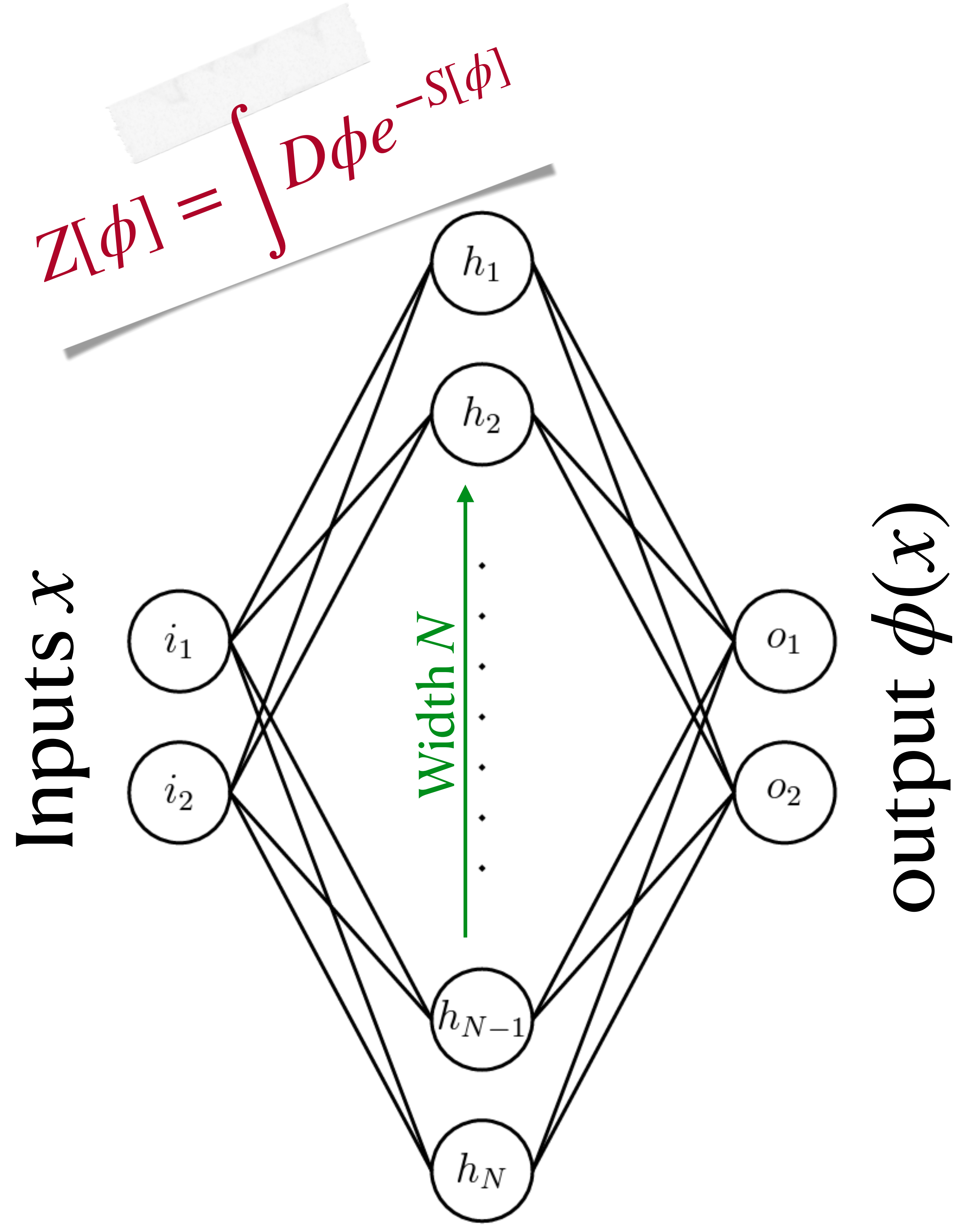


Motivation

- Deep Learning is widely used as a toolbox.
- E.g. take any neural network (NN), train it sufficiently; it can approximate a field configuration ϕ, from field action $S[\phi]$.
- But, NN training can find approximate answers.
- Can we design NNs to result in ϕ exactly, a step towards ab initio AI for QFT?



Introduction



$$\phi(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_i(x)$$

Notations: NN output $\phi(x)$ depends on input x & architecture details, e.g. width N, parameter distributions $P(h)$.

Initialize NN many times, do not train. NN output ensemble is a single distribution over functions $\phi(x)$.

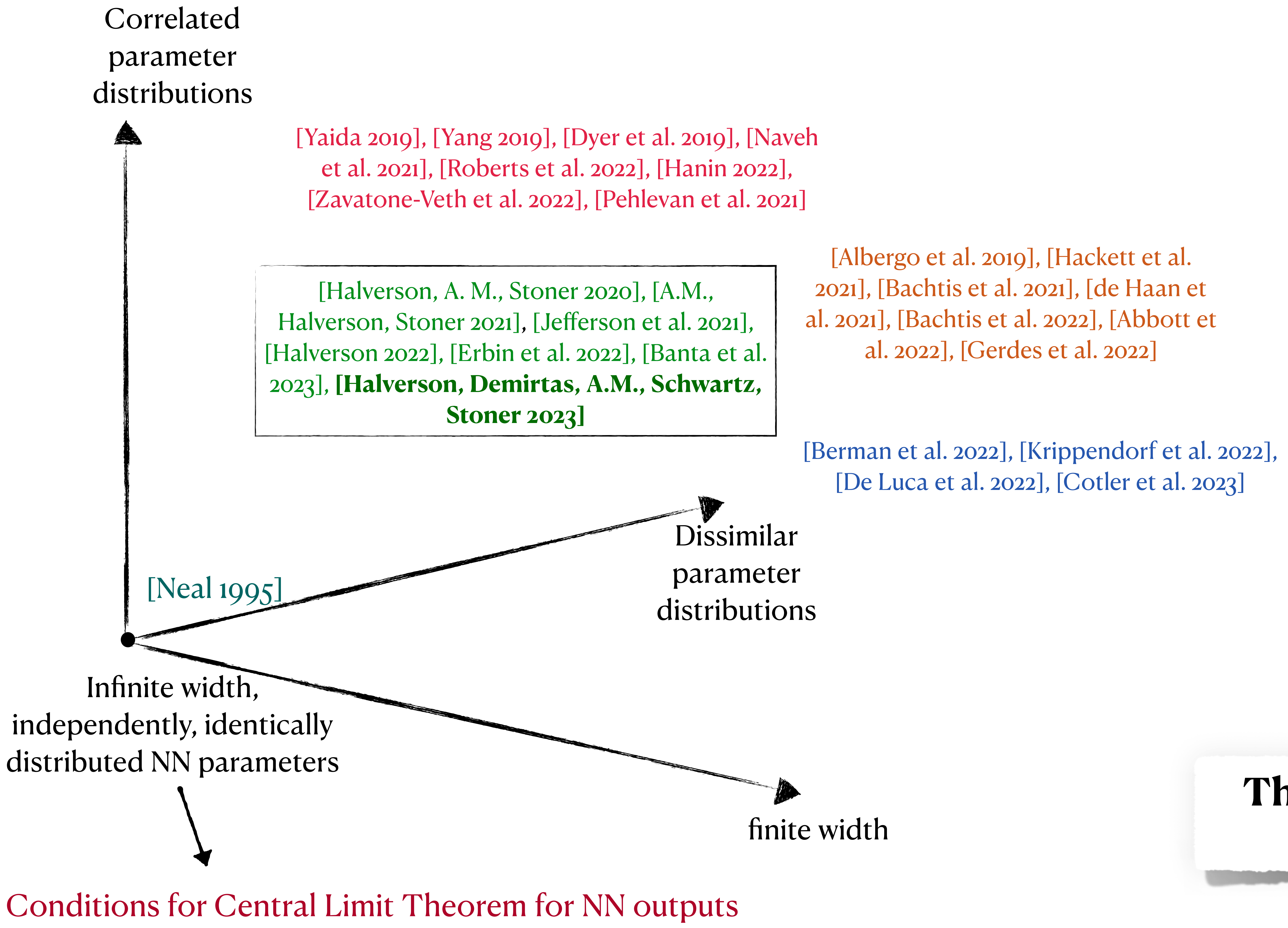
+

In path integral formalism, field theories are functional distributions, too.

↓

Each NN architecture, at initialization, corresponds to a unique distribution over fields / functions.

Related Works



- ML for field theories on lattice
- Correspondences between learning dynamics and RG and Hamiltonian evolution
- Correspondences between initialized NNs and field theory
- Theory of NN and deep learning at finite width, and other aspects of CLT breaking
- Theory of NN at asymptotically large width

The three axes depict three independent ways to violate CLT for NN outputs.

Talk Outline

- NN field theory correspondence (at initialization).

- Construction of field action $S[\phi]$ using NN architectures.

- Construction of NN architectures using certain $S[\phi]$.

NN Field Theory Correspondence

NN Field Theory Correspondence

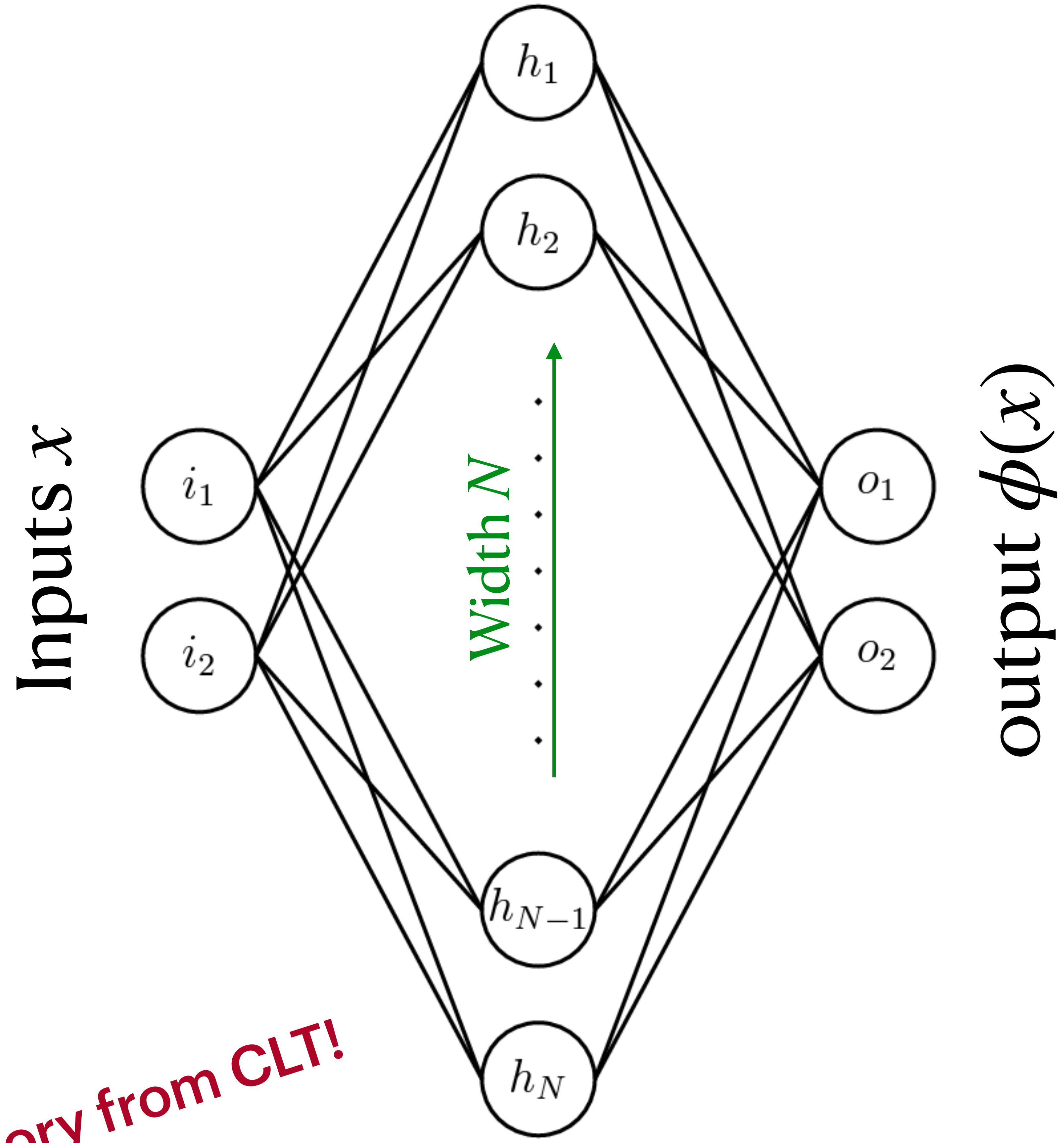
$$\phi(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_i(x)$$

- NN outputs are obtained by summing over last hidden layer neuron contributions $h_i(x)$.

- Choose NN parameters identically, independently distributed (i.i.d.), all $h_i(x)$ are i.i.d.

- At $N \to \infty$, such a sum, $\phi(x)$, is drawn from a Gaussian distribution, by **Central Limit Theorem (CLT).**

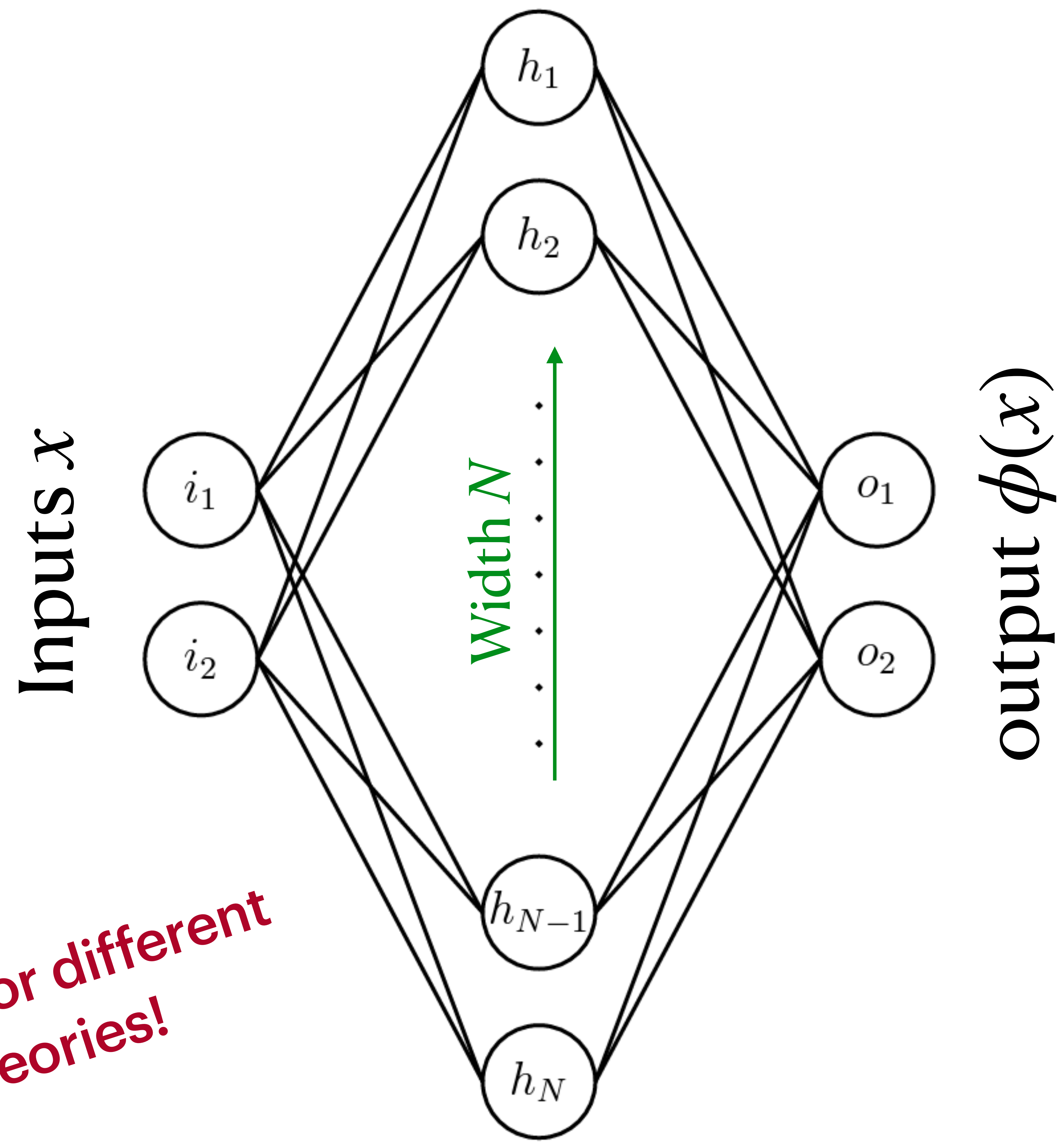
Free field theory from CLT!



NN Field Theory Correspondence

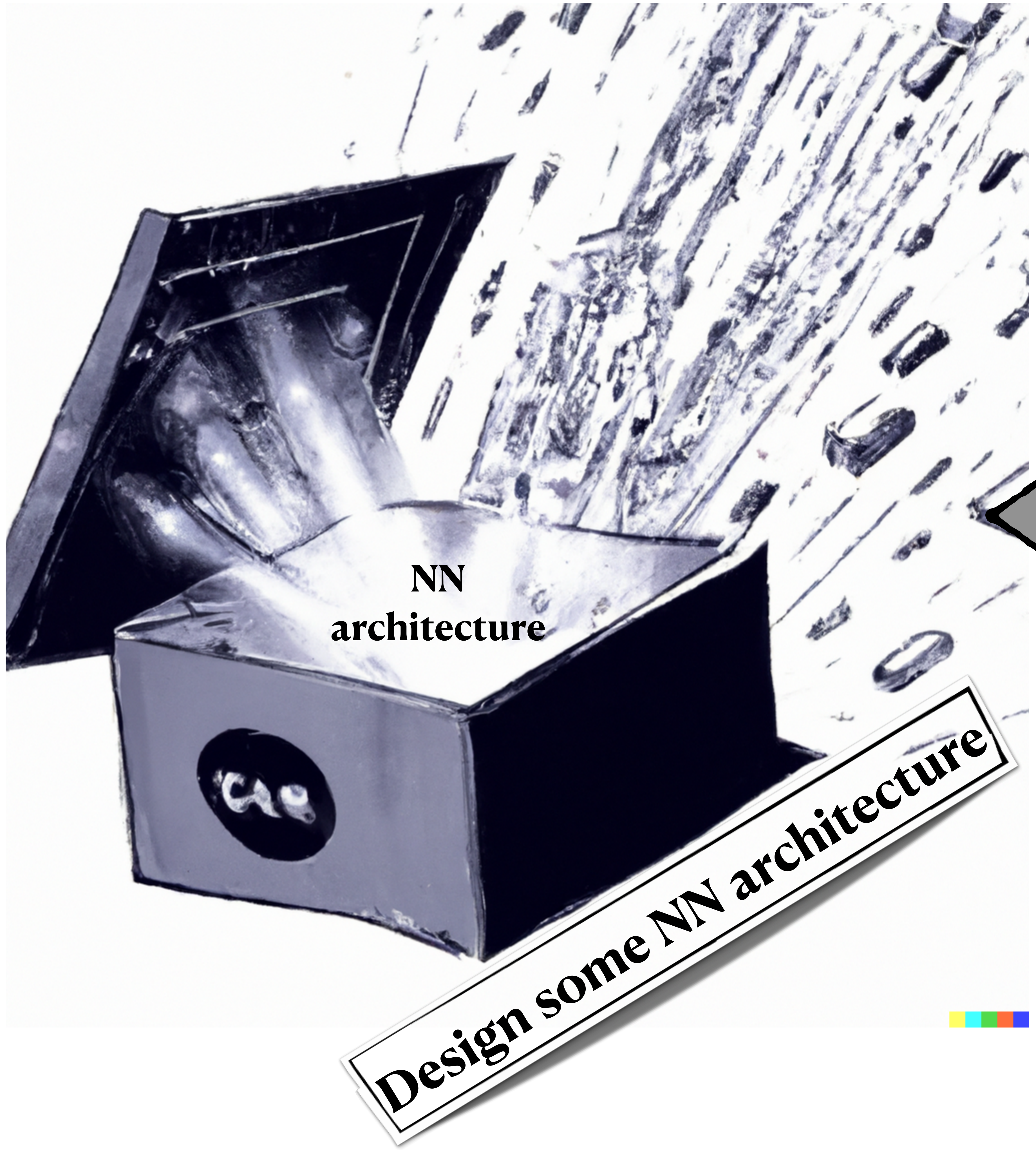
$$\phi(x) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} h_i(x)$$

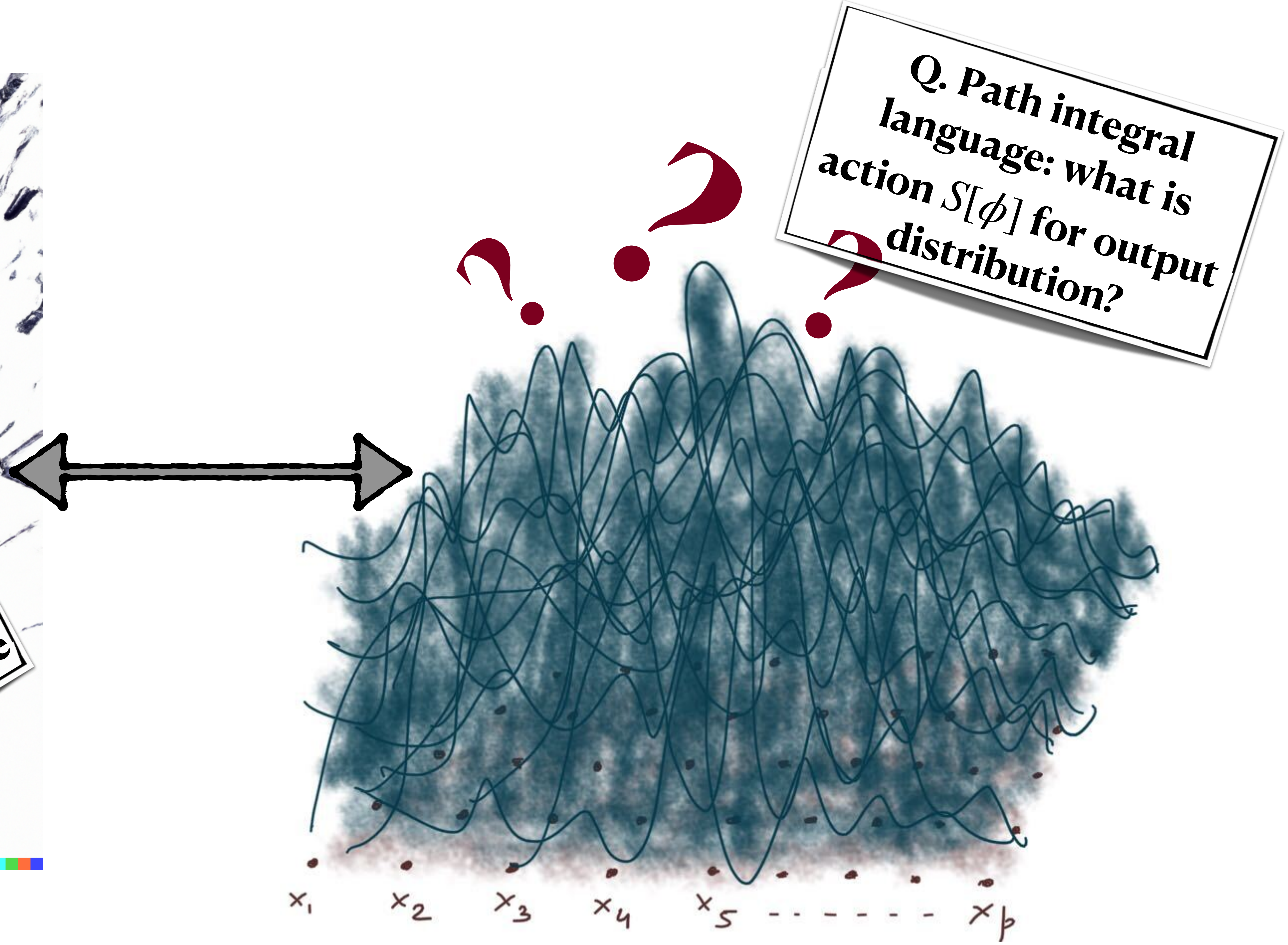
- Violate CLT on $\phi(x)$. Choose 1. Finite N, 2. Dissimilar NN parameters, 3. Correlated NN parameters.
- CLT is the origin of free field theory in NNs.
- Tuning NN initializations to violate CLT controls field interactions in $S[\phi]$.



Different initializations for different interacting field theories!

NN Field Theory Correspondence





Construct $S[\phi]$ for NN Architectures

Construct $S[\phi]$ for NN Architectures

- Fix NN architecture at initialization, we have exact probability density for output ensemble, in architecture language.

- In terms of connected correlators in architecture space.

$$P[\phi] = \frac{1}{Z} \exp\left(\sum_{r=3}^{\infty} \frac{(-1)^r}{r!} \int \prod_{i=1}^{r} d^d x_i G_c^{(r)}(x_1, \cdots, x_r) \frac{\delta}{\delta\phi(x_1)} \cdots \frac{\delta}{\delta\phi(x_r)}\right)$$
$$\times \exp\left(-\frac{1}{2} \int d^d x_1 d^d x_2\, \phi(x_1) G_c^{(2)}(x_1, x_2)^{-1} \phi(x_2)\right),$$

How do we infer $S[\phi]$ from this?

Connected r-pt function of NN outputs, in architecture space

Construct $S[\phi]$ for NN Architectures

- Want to systematically construct field theory from NN ensemble.

$$Z[J] = \int D\phi \, e^{-S_{\text{free}}[\phi] - S_{\text{int}} + \int d^d x J(x)\phi(x)}$$

- Free field action is easy to derive.

$$S_{\text{free}}[\phi] = \frac{1}{2} \int d^d x_1 \, d^d x_2 \, \phi(x_1) \, G_c^{(2)}(x_1, x_2)^{-1} \phi(x_2)$$

Construct $S[\phi]$ for NN Architectures

- Use NN architecture space to obtain connected parts of n-pt functions.

$$G^{(n)}(x_1, \cdots, x_n) := \mathbb{E}[\phi(x_1)\cdots\phi(x_n)] = \int dhP(h)\phi(x_1)\cdots\phi(x_n)$$

- CLT-violation at finite width, i.i.d. NN parameters

$$G_c^{(r)} \propto \frac{1}{N^{r/2-1}}.$$

- CLT-violation at finite width, non-i.i.d. NN parameters

$G_c^{(r)} \rightarrow$ mixed N-scalings.

Construct $S[\phi]$ for NN Architectures

- How to construct couplings $g_r(x_1, \cdots, x_r)$, for CLT-violating architectures?

$$S_{\text{int}} = \sum_{r=3}^{\infty} \int \prod_{i=1}^{r} d^d x_i \, g_r(x_1, \ldots, x_r) \, \phi(x_1) \ldots \phi(x_r)$$

- Introduce a new set of Feynman rules, to construct couplings $g_r(x_1, \cdots, x_r)$ in $S_{\text{int}}[\phi]$, using $G_c^{(r)}$ from architecture info.

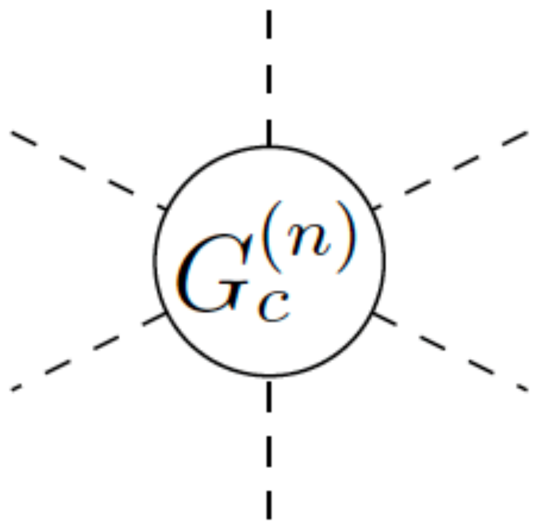
Feynman Rules for $g_r(x_1, \ldots, x_r)$.

1. Internal points associated to vertices are unlabeled, for diagrammatic simplicity. Propagators therefore connect to internal points in all possible ways.

2. For each propagator between z_i and z_j,

$$z_i \text{ - - - - } z_j \;= G_c^{(2)}(z_i, z_j)^{-1}.$$

3. For each vertex,



$$= (-1)^n \int d^d y_1 \cdots d^d y_n \, G_c^{(n)}(y_1, \cdots, y_n).$$

4. Divide by symmetry factor and insert overall $(-)$.

[Halverson, A.M., Schwartz, Demirtas, Stoner 2023]

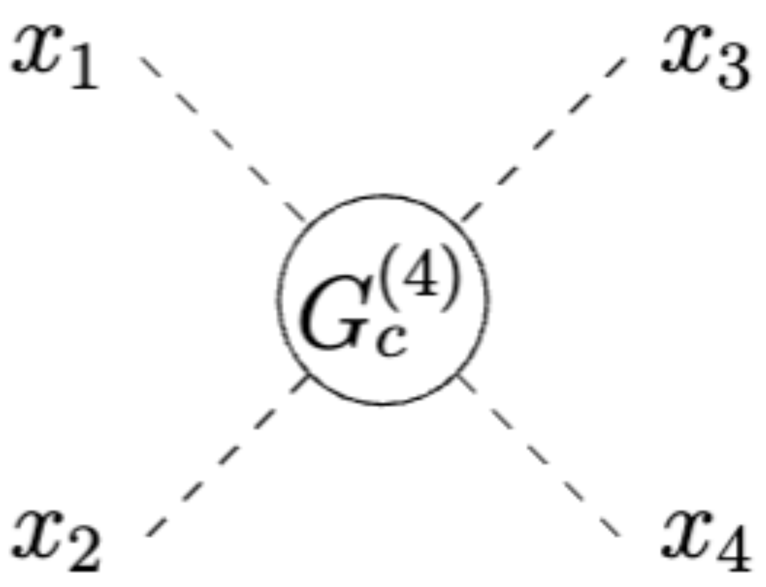
Table 2: Feynman rules for computing g_r from each connected diagram with $G_c^{(n)}$ vertices.

Example: Finite N, i.i.d. parameters

- Construct quartic coupling

$$g_4(x_1,\ldots,x_4) = \frac{1}{4!}\Bigg[\int dy_1dy_2dy_3dy_4\, G_c^{(4)}(y_1,y_2,y_3,y_4)\, G_c^{(2)}(y_1,x_1)^{-1}G_c^{(2)}(y_2,x_2)^{-1} G_c^{(2)}(y_3,x_3)^{-1}G_c^{(2)}(y_4,x_4)^{-1} + \text{Comb.}\Bigg] + O\left(\frac{1}{N^2}\right),$$

$$=$$



$$+\, O\left(\frac{1}{N^2}\right).$$

- Interacting NN field theory action

$$S[\phi] = S_{\text{free}}[\phi] - \int d^dx_1\cdots d^dx_4\, g_4(x_1,\cdots,x_4)\,\phi(x_1)\cdots\phi(x_4) + O\Big(\frac{1}{N^2}\Big)$$

Where's the 'Quantum'-ness?

- NN ensembles are Euclidean or statistical field theories.

- However, if all NN correlation functions satisfy Osterwalder-Schrader axioms, the field theory over output ensembles has analytic continuation to Lorentzian background.

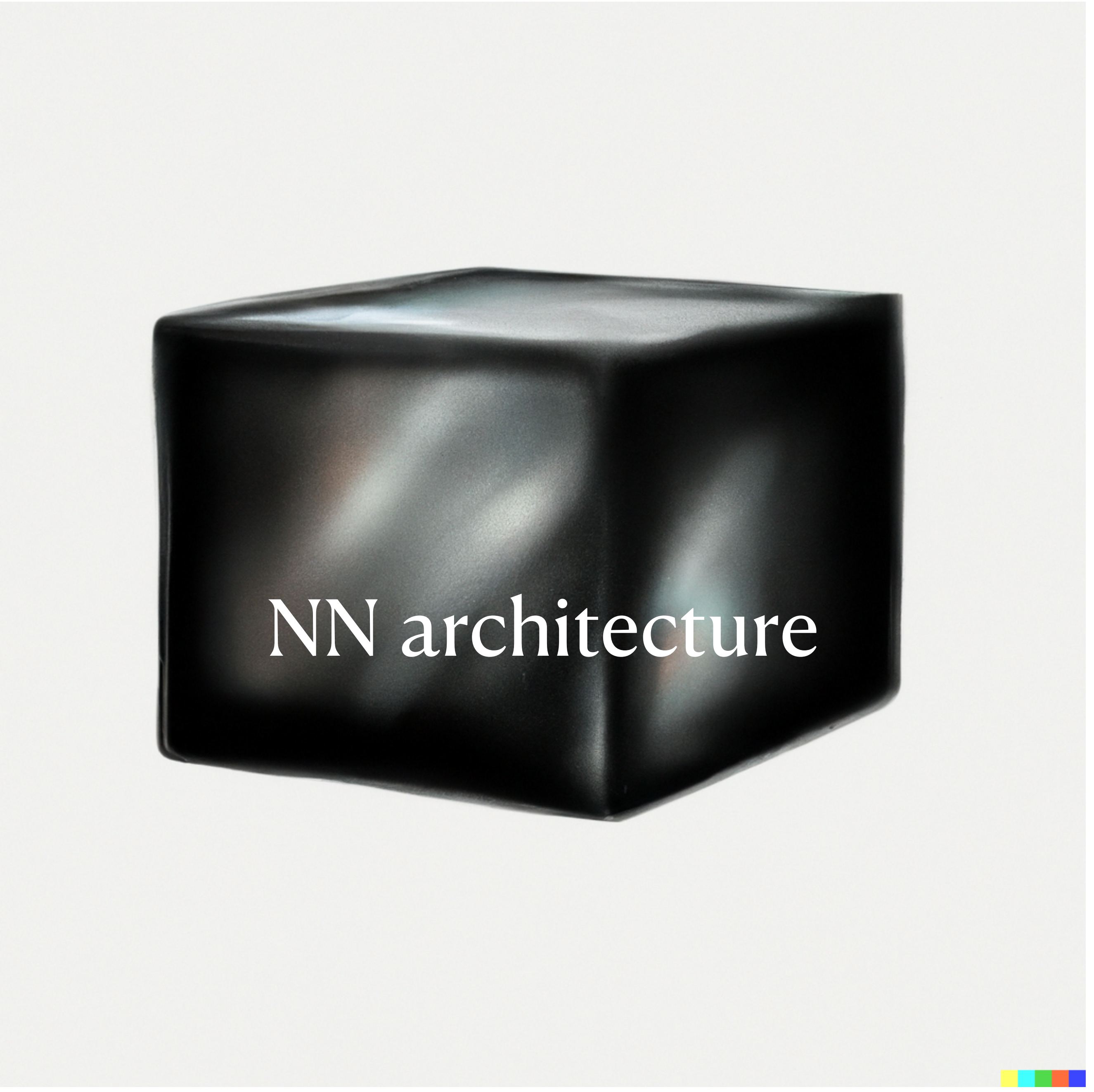
[Halverson 2021]

- Such NN architectures correspond to *quantum* field theories.

Construct NN Architectures for $S[\phi]$

Construct NN Architectures for $S[\phi]$

- We want rules for the following goal:



To get field interactions of our choice

$$Z[J] = \int D\phi \, e^{-S[\phi] + \int d^d x J(x)\phi(x)}$$

Come up with architecture details: width, depth, activation functions, parameters etc.

$$S[\phi] = S_{\text{free}}[\phi] + \lambda \int d^d x_1 \cdots d^d x_r \, \mathcal{O}_\phi(x_1, \cdots, x_r)$$

Construct NN Architectures for $S[\phi]$

- Getting correct free theory action $S_{\text{free}}[\phi]$ is easy.

  - $\lim N \to \infty$
  - i.i.d. parameters $P(h) := P_G(h)$

- Next, deform NN parameters at infinite N,

  - Redefine $P(h) := P_G(h)\, e^{-\lambda \int d^d x_1 \cdots d^d x_r\, \mathcal{O}_{\phi_h}(x_1,\cdots,x_r)}$

to insert $S_{\text{int}}[\phi] = \lambda \int d^d x_1 \cdots d^d x_r \mathcal{O}_\phi(x_1, \cdots, x_r)$ in NN ensemble action.

Example: Scalar $\lambda\phi^4$ theory

Can we design NN architecture for weakly interacting scalar field theory?

$$S[\phi] = \int d^d x \left[\phi(x)(\nabla^2 + m^2)\phi(x) + \frac{\lambda}{4!}\phi(x)^4\right] \qquad \nabla^2 := \frac{\partial^2}{\partial x^2}$$

Solution

Step A) Design an infinite width NN architecture for free scalar field theory

$$\phi_{a,b,c}(x) = \sqrt{\frac{2\,\mathrm{vol}(B^d_\Lambda)}{\sigma_a^2(2\pi)^d}} \sum_{i,j} \frac{a_i\ \cos(b_{ij}x_j + c_i)}{\sqrt{\mathbf{b}_i^2 + m^2}}$$

$$P_G(a) = \prod_i e^{-\frac{N}{2\sigma_a^2}a_i a_i}$$

$$P_G(b) = \prod_i P_G(\mathbf{b}_i) \ \text{ with } \ P_G(\mathbf{b}_i) = \mathrm{Unif}(B^d_\Lambda)$$

$$P_G(c) = \prod_i P_G(c_i) \ \text{ with } \ P_G(c_i) = \mathrm{Unif}([-\pi, \pi])$$

$\lim N \to \infty.$

$$G^{(2)}(p) = \frac{1}{p^2 + m^2}$$

Example: Scalar $\lambda\phi^4$ theory

Step B) Deform NN parameter distribution at infinite width

$$P(a,b,c) = P_G(a)P_G(b)P_G(c)\; e^{-\frac{\lambda}{4!}\int d^dx\, \phi_{a,b,c}(x)^4}$$

Insert operator $e^{-\frac{\lambda}{4!}\int d^dx\, \phi_{a,b,c}(x)^4}$ in previous partition function.

$$Z[J] = \int da\, db\, dc\;\; P(a,b,c)\; e^{\int d^dx J(x)\, \phi_{a,b,c}(x)}$$

An architecture for scalar $\lambda\phi^4$ theory on Euclidean space, at initialization!

- Symmetries of NN parameter distributions become symmetries of $S[\phi]$. ← Symmetry-via-duality method
- All correlators of deformed architecture still satisfy OS axioms.

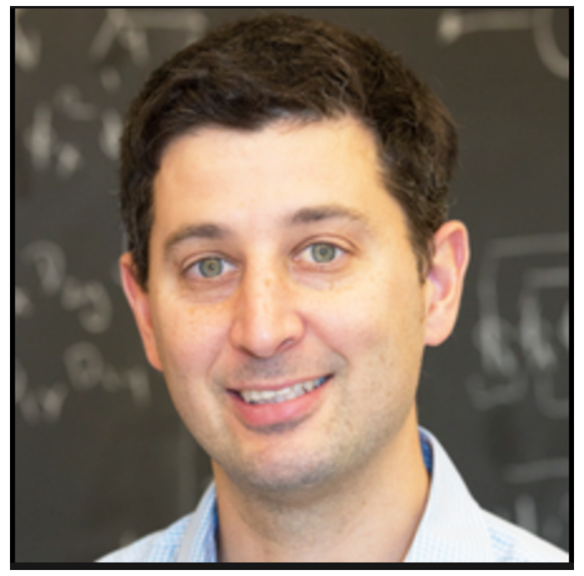
[Halverson, A.M., Stoner 2021]

Conclusion & Outlook

- Given NN architecture, we have rules to get NN field action $S[\phi]$, at leading order.
- Given an interacting field theory action $S[\phi]$, we have rules to design NN architecture for this, at initialization.
- As an example, we presented an architecture for $\lambda\phi^4$ scalar field theory at inf N, using parameter deformations from i.i.d.
- Currently, we are working on NN architectures at initialization for fermionic field theories.
- In the future, we want NN architectures for gauge theories, at initialization.

My amazing collaborators!

Jim Halverson



Matt Schwartz

Mehmet Demirtas

Keegan Stoner

Thank you!

Questions?

Feel free to get in touch:
email: amaiti@perimeterinstitute.ca
Twitter: @AninditaMaiti7
web: https://aninditamaiti.github.io/

Extra Slides

Technical Details

Consider a sum over N random variables:

$$\phi = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$$

Moment generating function of ϕ:

$$Z_\phi[J] := \mathbb{E}[e^{J\phi}] = \mathbb{E}[e^{J\sum_i X_i/\sqrt{N}}]$$

Cumulant generating functional of ϕ is the log of the moment generating function:

$$W_\phi[J] := \log\, \mathbb{E}[e^{J\phi}] = \log\, \mathbb{E}[e^{J\sum_i X_i/\sqrt{N}}] \;= \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} J^r$$

Cumulant or connected statistical moment of ϕ, as derivative of $W[J]$:

$$\kappa_r^\phi := \left(\frac{\partial}{\partial J}\right)^r W_\phi[J]$$

Edgeworth method for probability density of ϕ in terms of cumulants in architecture space:

$$P[\phi] = \exp\left[\sum_{r=3}^{\infty} \frac{\kappa_r}{r!}(-\partial_\phi)^r\right] \int dJ e^{\kappa_2 \frac{J^2}{2} + \kappa_1 J - J\phi}$$

$$= \exp\left[\sum_{r=3}^{\infty} \frac{\kappa_r}{r!}(-\partial_\phi)^r\right] e^{-\frac{(\phi-\kappa_1)^2}{2\kappa_2}},$$

NNFT quartic at finite N, non-i.i.d. parameters

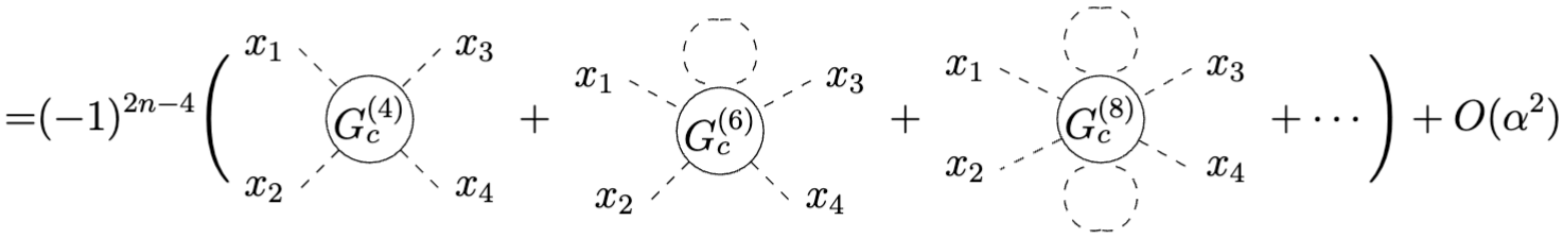
For statistical independence breaking hyperparameter α, connected correlators at leading order

$$G_c^{(r)}(x_1, \cdots, x_r) \propto \alpha \quad \forall r > 2$$

Quartic coupling at leading order

$$g_4(x_1, x_2, x_3, x_4) = \sum_{n=2}^{\infty} \frac{(-1)^{2n-4}}{(2n)!} \Big[\int dy_1 \cdots dy_{2n} \, G_c^{(2n)}(y_1, \cdots, y_{2n}) \, G_c^{(2)}(y_1, x_1)^{-1}$$
$$G_c^{(2)}(y_2, x_2)^{-1} \, G_c^{(2)}(y_3, x_3)^{-1} \, G_c^{(2)}(y_4, x_4)^{-1} \prod_{m=5}^{2n-1} G_c^{(2)}(y_m, y_{m+1})^{-1} + \text{Comb.} \Big] + O(\alpha^2)$$

It's an infinite series sum! Cannot be truncated unless we impose more conditions on NN architecture.



Compute $S[\phi]$ for (In Real Life) NN architecture

Network output $\phi(x) = W_i^1 \cos(W_{ij}^0 x_j + b_i^0)$

i.i.d. parameter distributions

$$W^0 \sim \mathcal{N}(0, \sigma_{W_0}^2/d)$$
$$W^1 \sim \mathcal{N}(0, \sigma_{W_1}^2/N)$$
$$b^0 \sim \text{Unif}[-\pi, \pi]$$

Interacting NN field theory action due to finite N, at leading order

$$S_{\text{Cos}}[\phi] = \frac{2\sigma_{W_0}^2}{\sigma_{W_1}^2 d} \int d^d x\, \phi(x)\, e^{-\frac{\sigma_{W_0}^2 \nabla_x^2}{2d}} \phi(x) \; - \int d^d x_1 \cdots d^d x_4 \left[\frac{4\sqrt{6}\pi^{3/2}\sigma_{W_0}^4}{N d^2 \sigma_{W_1}^4} \sum_{\mathcal{P}(abcd)} e^{-\frac{\sigma_{W_0}^2 \nabla_{r_{abcd}}^2}{6d}} \right.$$
$$\left. - \frac{8\pi\sigma_{W_0}^4}{N d^2 \sigma_{W_1}^4} \sum_{\mathcal{P}(ab,cd)} e^{-\frac{\sigma_{W_0}^2 (\nabla_{r_{ab}}^2 + \nabla_{r_{cd}}^2)}{2d}} \right] \phi(x_1) \cdots \phi(x_4) + O(1/N^2).$$