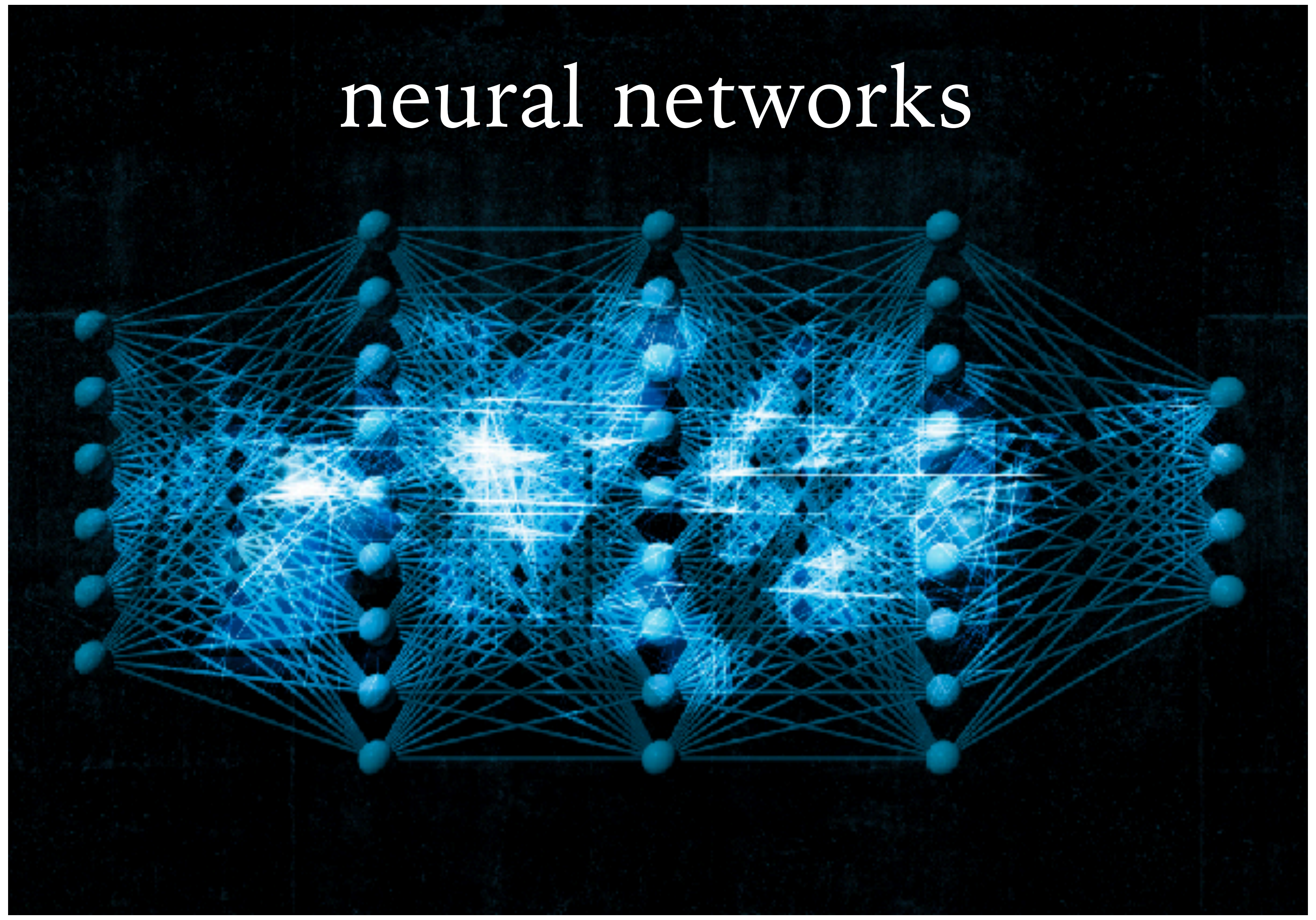
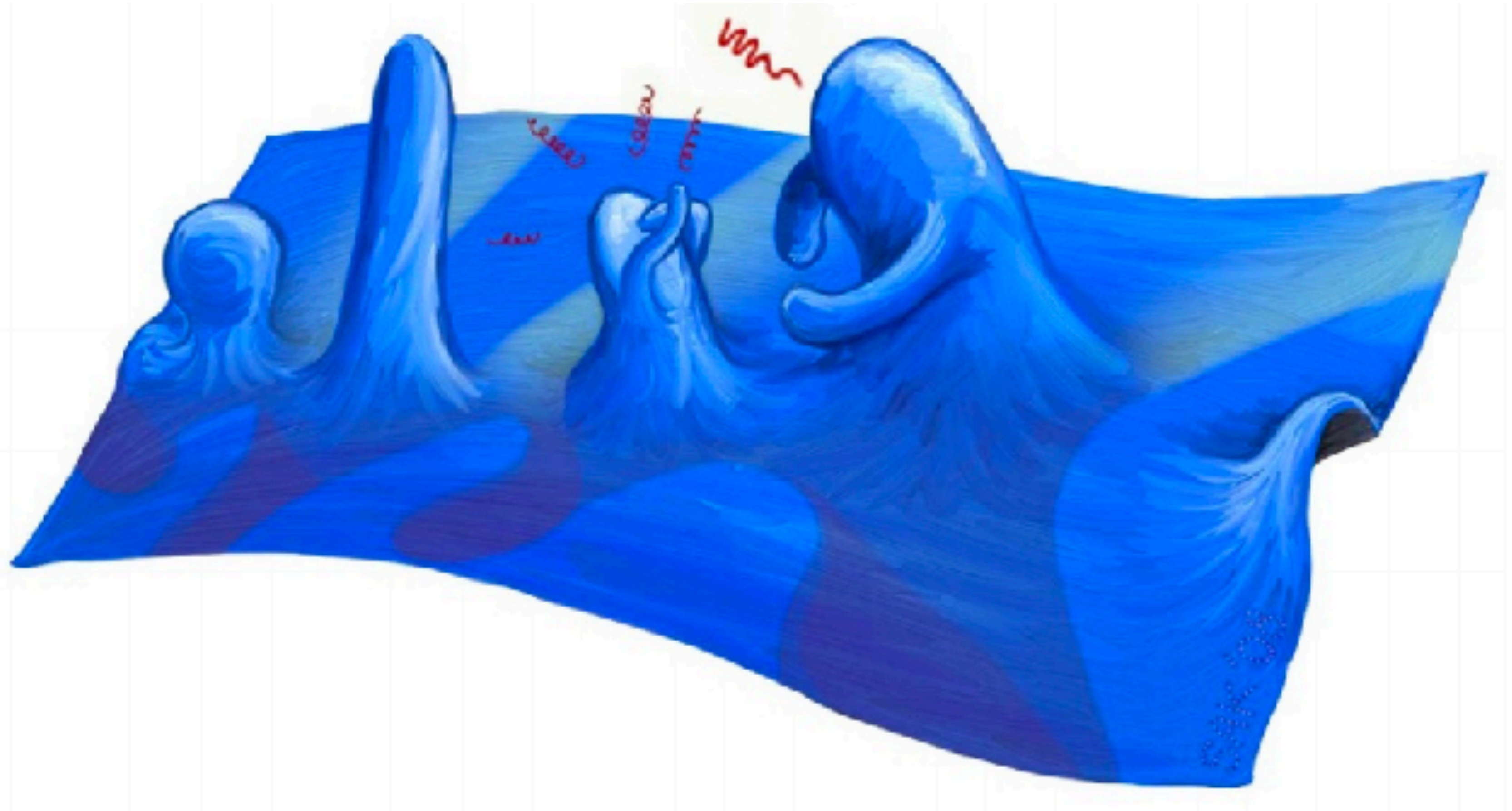


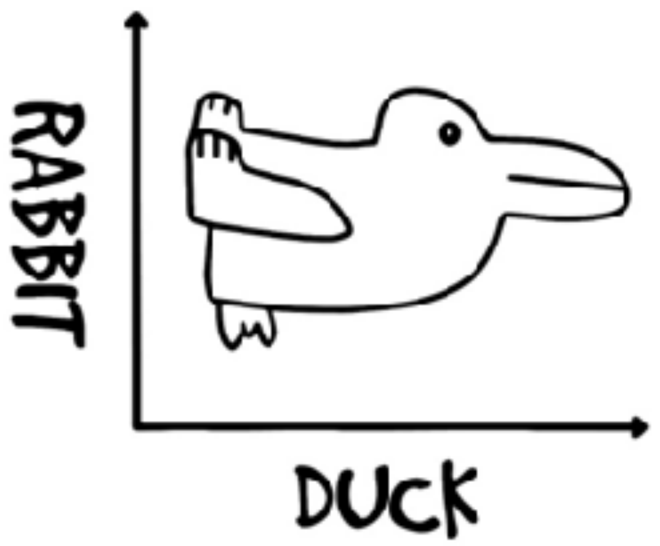
field theories



Structures of Neural Network Effective Theories

Zhengkang "Kevin" Zhang (University of Utah)

I. Banta, T. Cai, N. Craig, ZZ, 2305.02334.



Outline

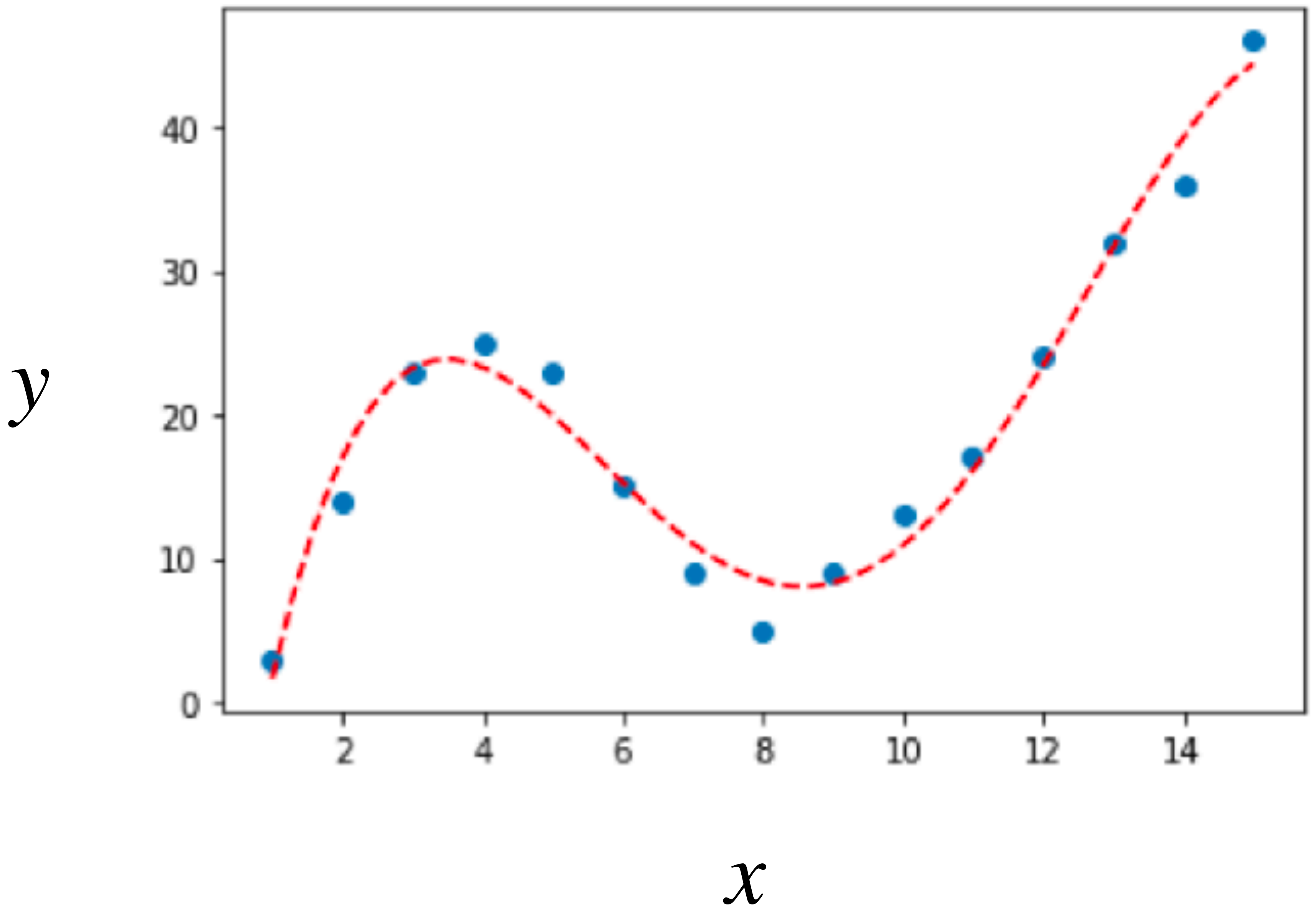
1. Neural networks $\leftrightarrow$ field theories (high-level summary).
2. EFT of deep neural networks (at initialization).
3. Diagrammatic approach.
4. Structures of neural network EFTs and criticality.

Outline

1. Neural networks $\leftrightarrow$ field theories (high-level summary).
2. EFT of deep neural networks (at initialization).
3. Diagrammatic approach.
4. Structures of neural network EFTs and criticality.

What is a (deep) neural network?

Goal (supervised learning): learn a function $y = f(\vec{x})$ from training dataset $(\vec{x}_\alpha, y_\alpha)$.

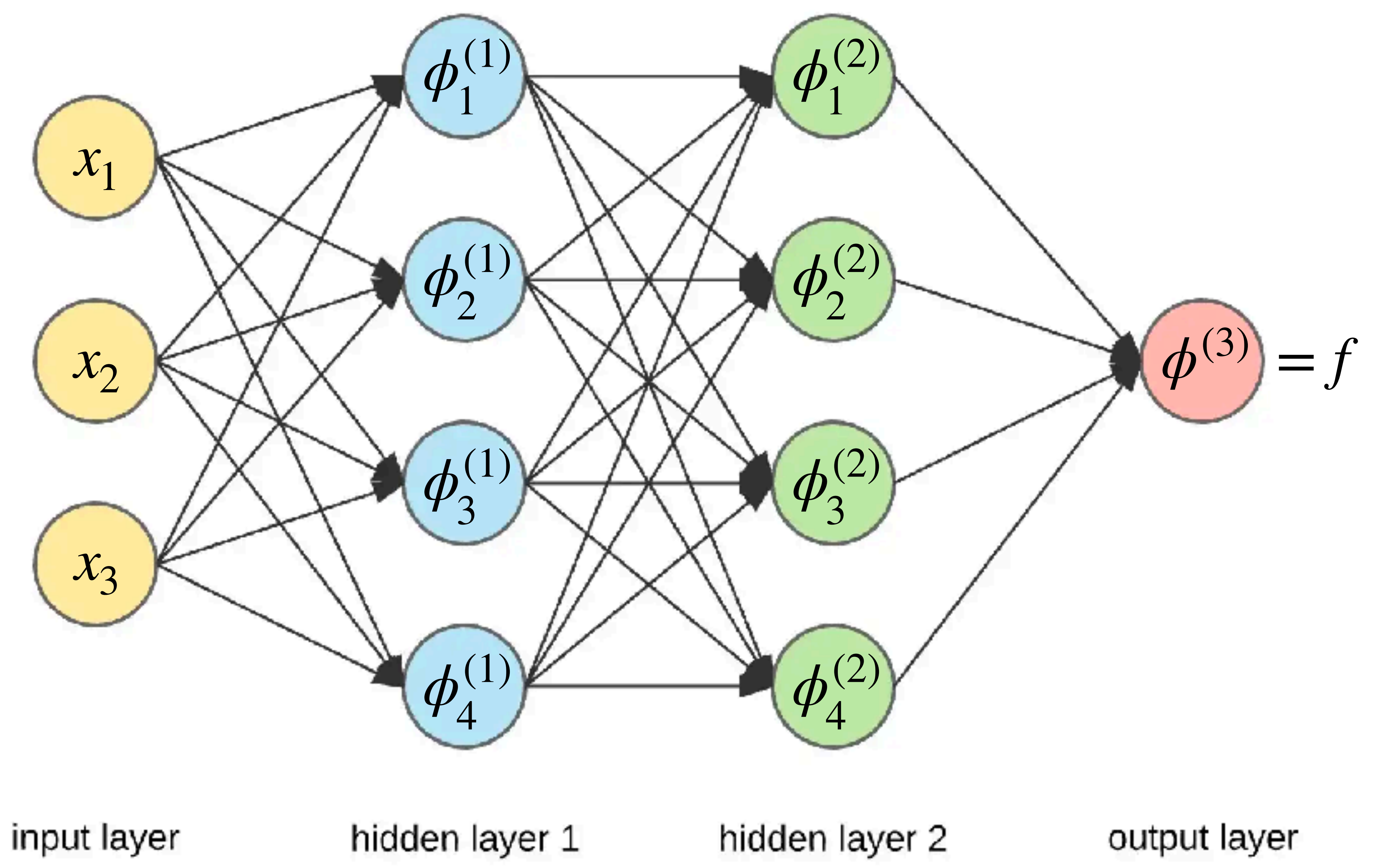


x Image	y Label
	Cat
	Cat
	Dog
	Dog

What is a (deep) neural network?

Goal (supervised learning): learn a function $y = f(\vec{x})$ from training dataset $(\vec{x}_\alpha, y_\alpha)$.

Archetype: multilayer perceptron.



$$\phi_i^{(1)}(\vec{x}) = \sum_{j=1}^{n_0} W_{ij}^{(1)} x_j + b_i^{(1)},$$

$$\phi_i^{(\ell)}(\vec{x}) = \sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} \sigma\left(\phi_j^{(\ell-1)}(\vec{x})\right) + b_i^{(\ell)} \quad (\ell \geq 2).$$

nonlinear activation function (e.g. tanh)

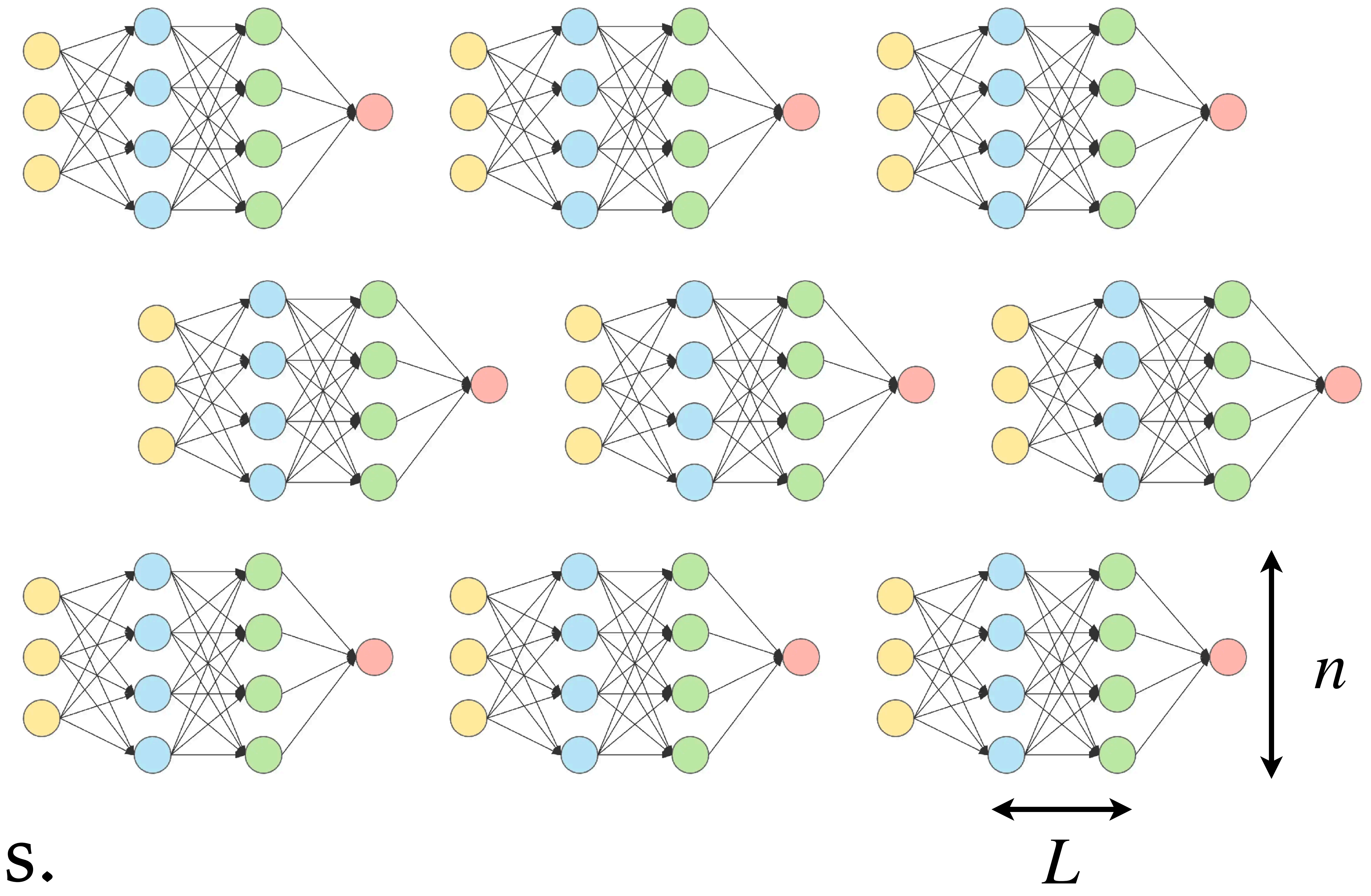
weights biases

trainable parameters:

- randomly initialized
- then updated by gradient descent to minimize a loss, e.g. $\sum_\alpha \left(f(x_\alpha) - y_\alpha\right)^2$

Neural networks $\leftrightarrow$ field theories (1/2)

Ensemble of networks, randomly initialized.



Neurons $\leftrightarrow$ scalar fields $\phi(\vec{x})$.

Ensemble statistics $\leftrightarrow$ action: $P(\phi) = e^{-S[\phi]}$.

Infinitely-wide networks* $(n \to \infty)$ $\leftrightarrow$ free theories.

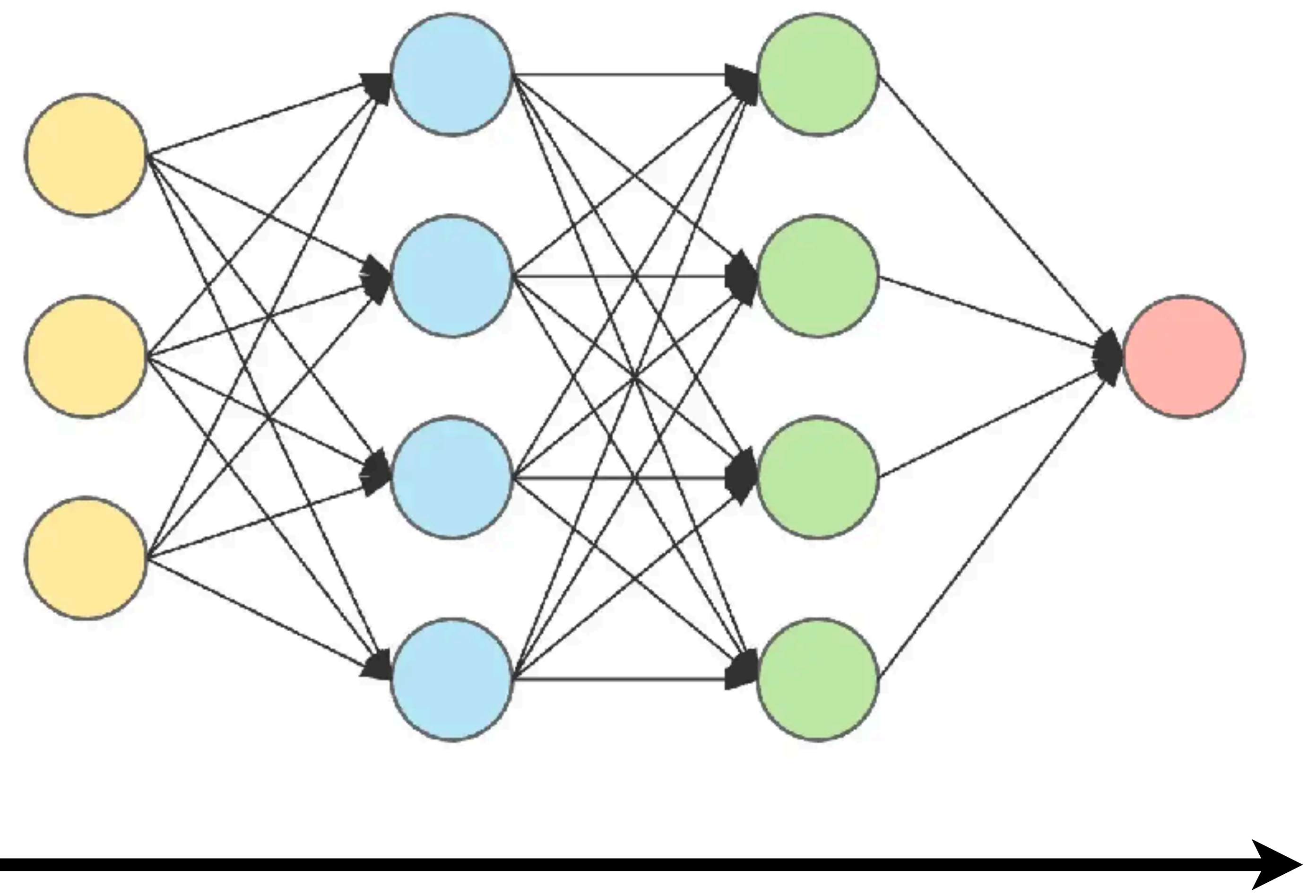
Wide networks $(n \gg L)$ $\leftrightarrow$ weakly-interacting theories (perturbation theory!).

* Neal '96. Williams '96.

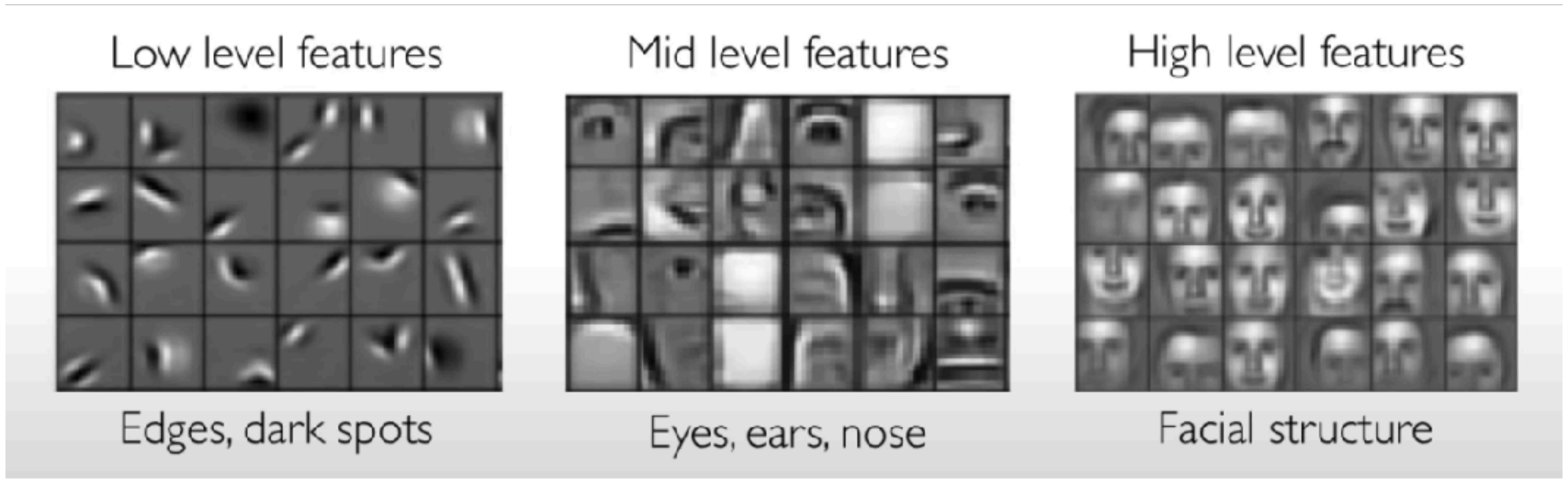
Neural networks ↔ field theories (2/2)

Information flow ↔ RG flow.

input (UV)



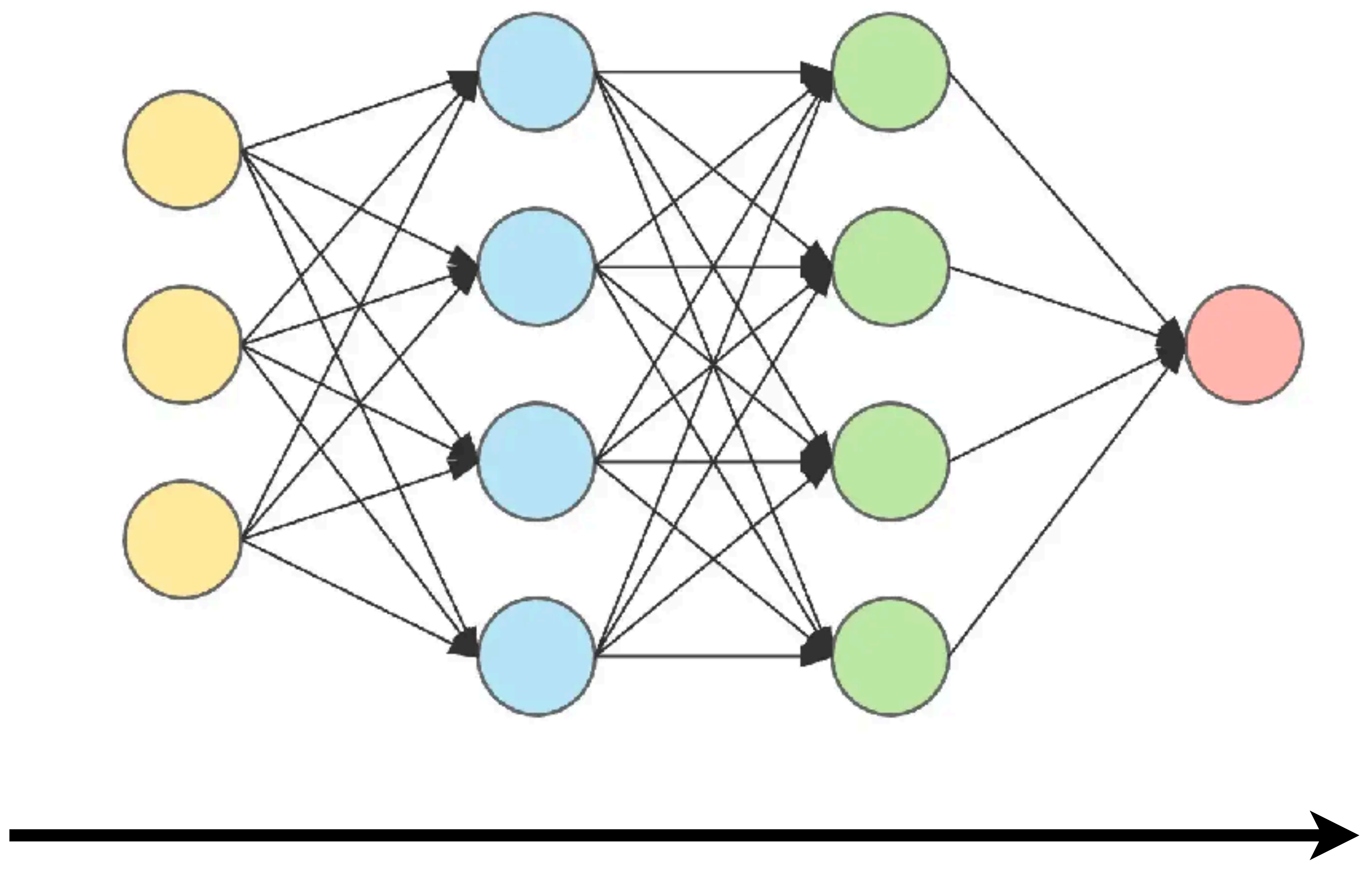
output (IR)



Neural networks ↔ field theories (2/2)

Information flow ↔ RG flow.

input (UV)



output (IR)

Exponential scaling (generic) ↔ flow to trivial fixed point.

Tune to criticality* ⇒ power-law scaling ↔ nontrivial fixed point.

* Raghu et al '16. Poole et al '16. Schoenholz et al '16.

Dreams

A theory of ~~everything~~ deep learning (opening the black box)?

Lee et al '17-19. Matthews et al '18. Yang '19-23.

Jacot, Gabriel, Hongler '18.

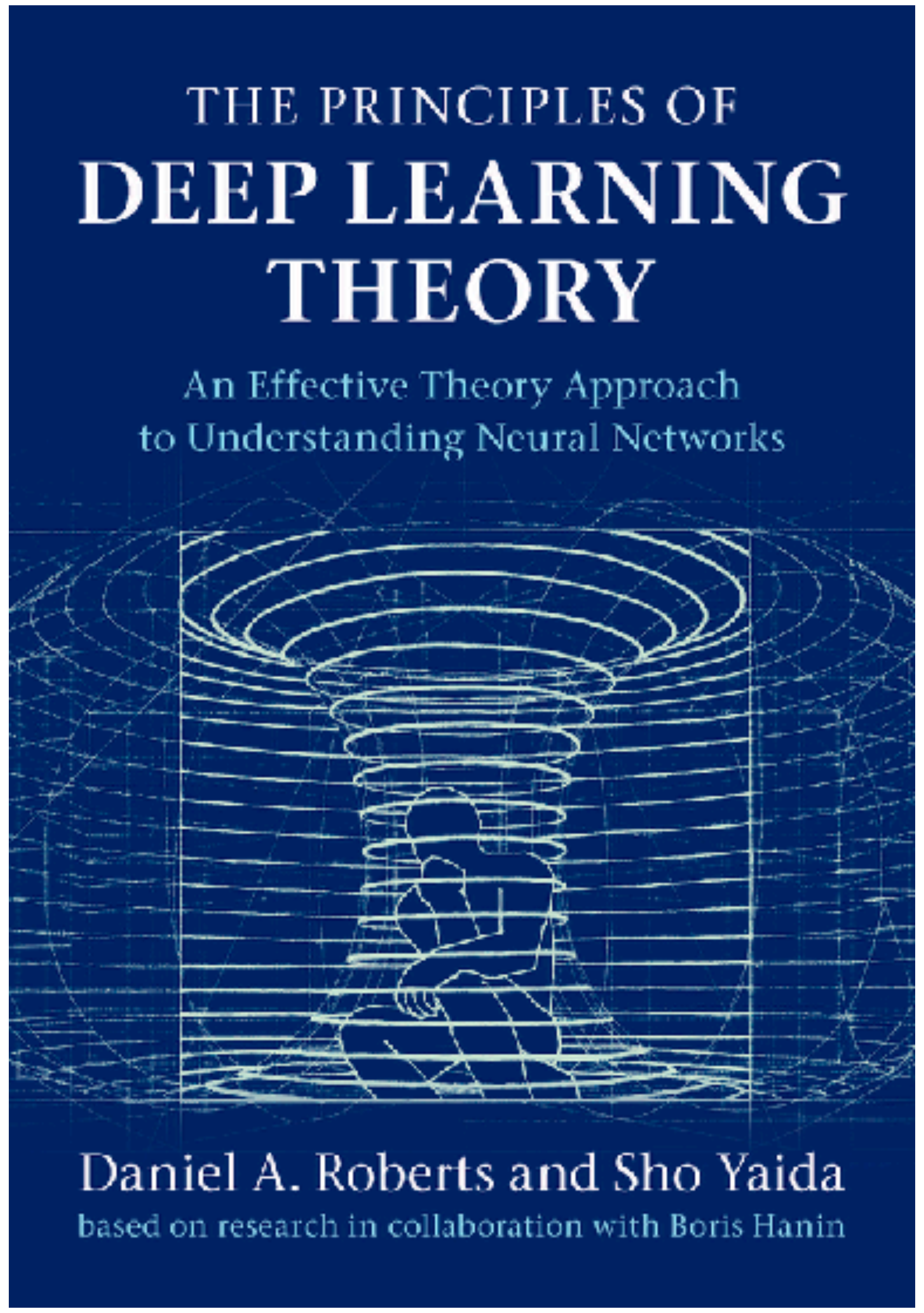
Antognini '19. Huang, Yau '19.

Yaida '19, '22. Hanin, Nica '19. Hanin '21, '22.

Dyer, Gur-Ari '19. Aitken, Gur-Ari '20. Andreassen, Dyer '20.

Naveh, Ringel et al '20, '21. Zavatone-Veth et al '21.

Roberts, Yaida, Hanin '21. (Our work is largely inspired by this book.)



Dreams

A theory of ~~everything~~ deep learning (opening the black box)?

A new angle to learn about field theories?

Schoenholz, Pennington, Sohl-Dickstein '17.

Cohen, Malka, Ringel '19.

Erbin, Lahoche, Samary '21, '22.

Bachtis, Aarts, Lucini '21.

Grosvenor, Jefferson (+ Erdmenger) '21.

Halverson '21; + Maiti, Stoner '20, '21; + Demirtas, Schwartz '23.

Outline

1. Neural networks $\leftrightarrow$ field theories (high-level summary).

2. EFT of deep neural networks (at initialization).

3. Diagrammatic approach.

4. Structures of neural network EFTs and criticality.

Initializing a deep neural network

Network depth (number of layers): L .

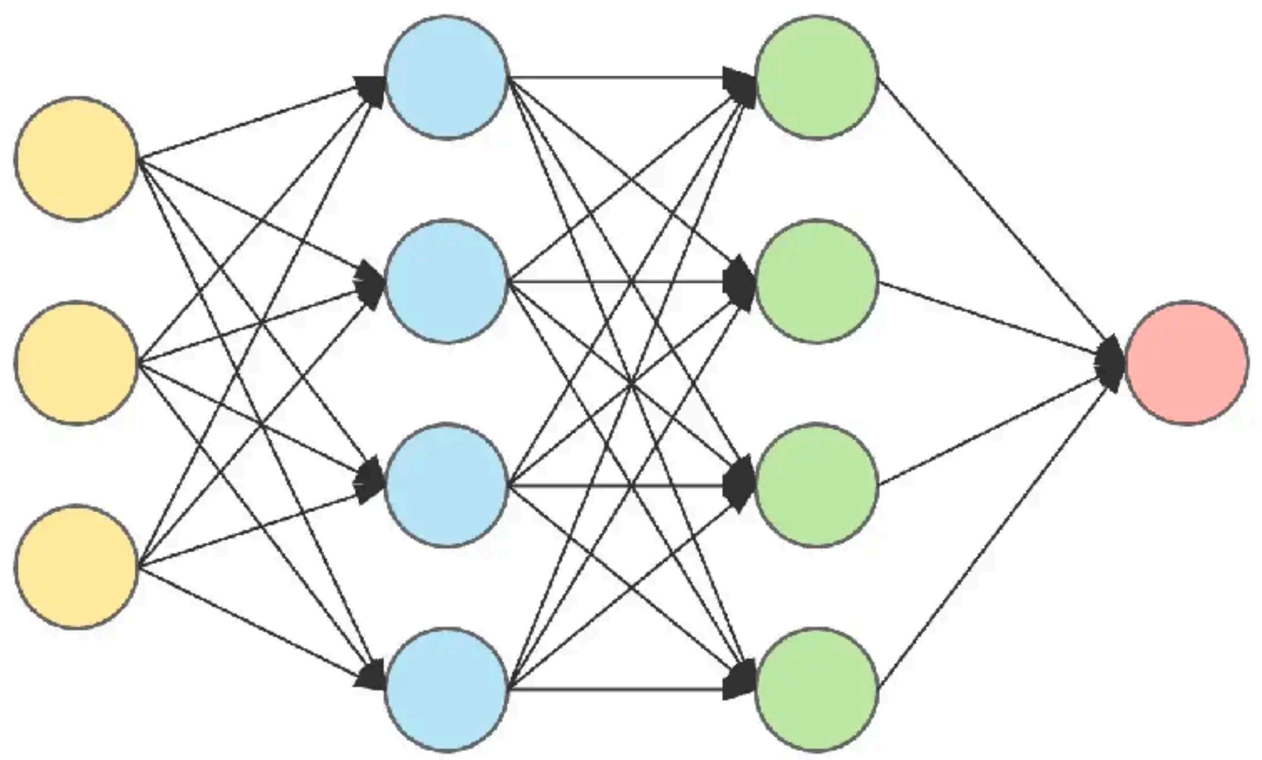
Widths (number of neurons per layer): $n_0, n_1, \ldots, n_{L-1}, n_L$.

← *architecture hyperparameters*

input $x \in \mathbb{R}^{n_0}$

hidden layer widths $\gg 1$

output $y \in \mathbb{R}^{n_L}$

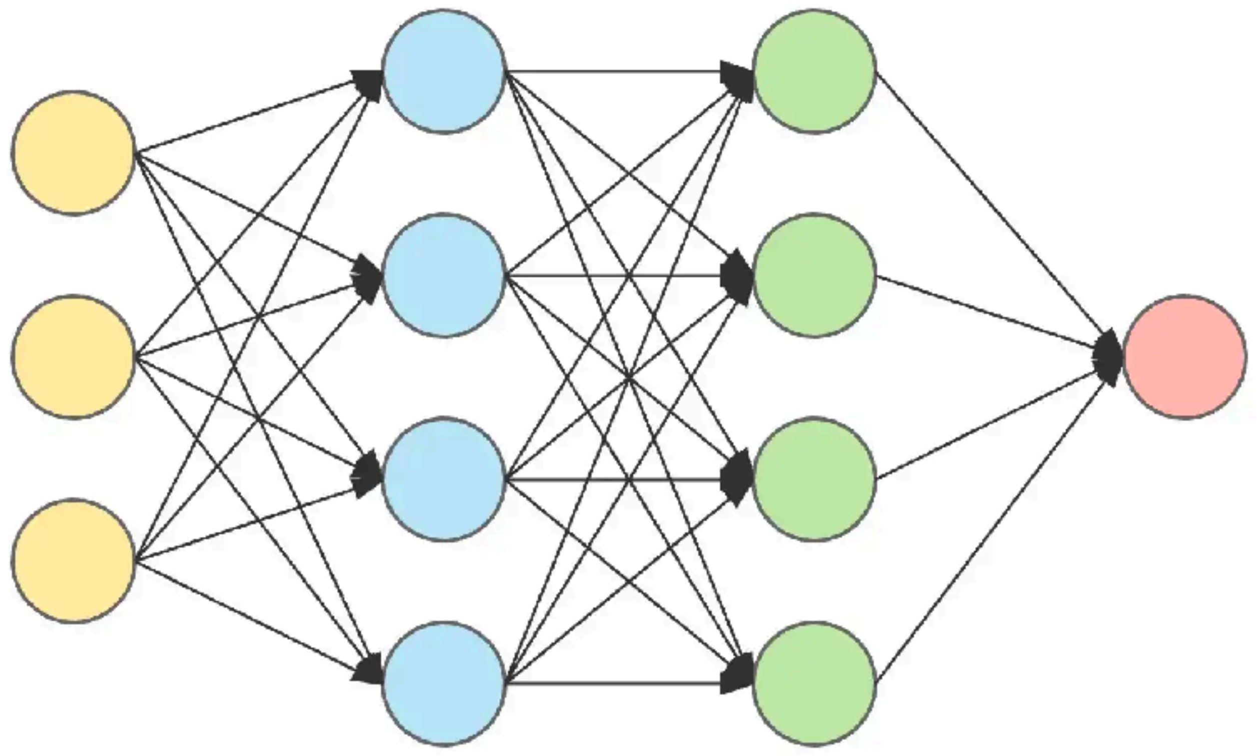


Initializing a deep neural network

Network depth (number of layers): L .

Widths (number of neurons per layer): $n_0, n_1, \ldots, n_{L-1}, n_L$.

architecture hyperparameters



$$\phi_i^{(\ell)}(\vec{x}) = \sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} \sigma\left(\phi_j^{(\ell-1)}(\vec{x})\right) + b_i^{(\ell)}$$

Weights and biases drawn from Gaussian distributions with mean 0, variances $C_W^{(\ell)}/n_{\ell-1}$, $C_b^{(\ell)}$.

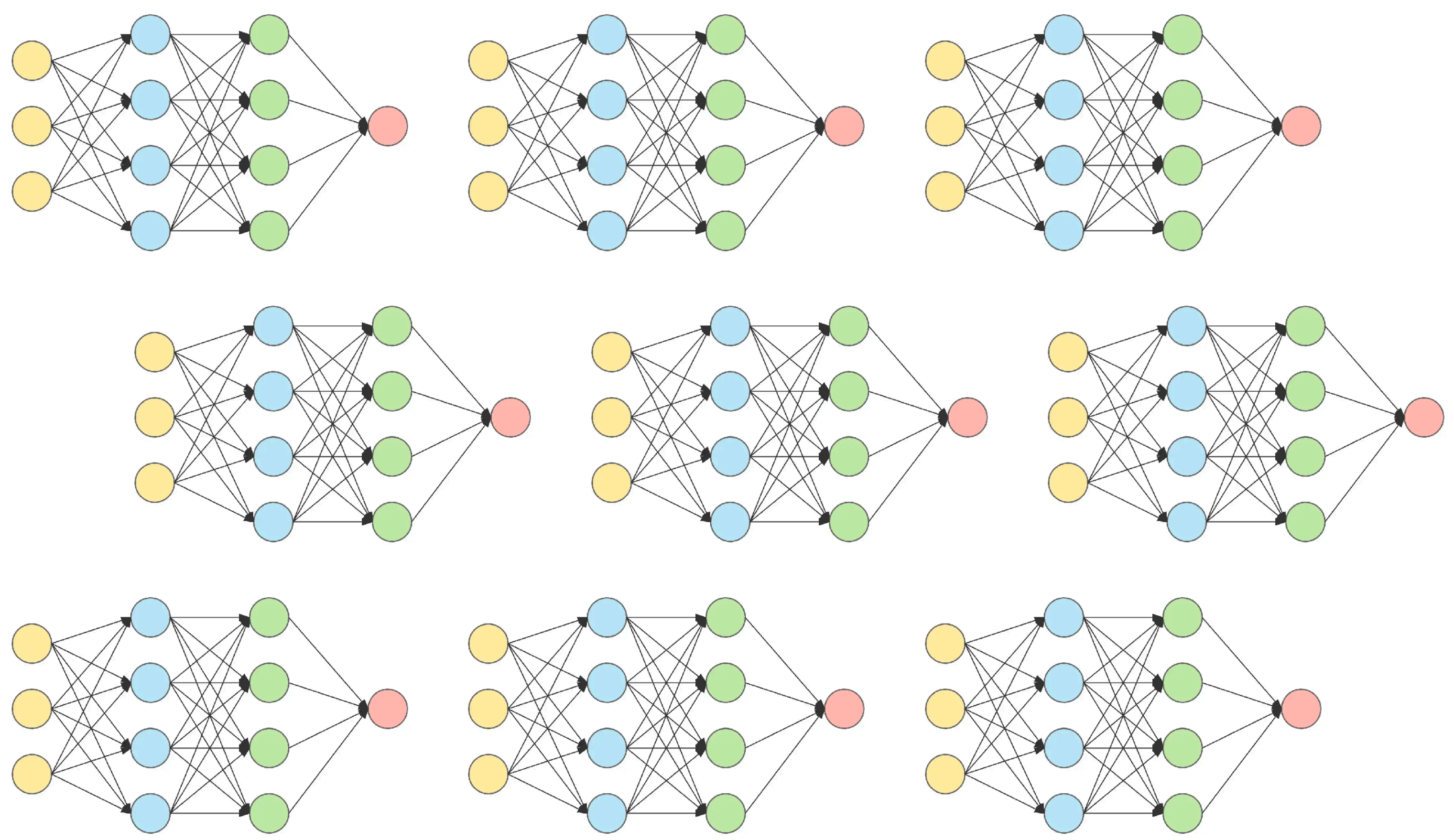
initialization hyperparameters

Initializing ~~a deep neural network~~ an ensemble of networks

Network depth (number of layers): L .

Widths (number of neurons per layer): $n_0, n_1, \ldots, n_{L-1}, n_L$.

← *architecture hyperparameters*



$$\phi_i^{(\ell)}(\vec{x}) = \sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} \sigma\left(\phi_j^{(\ell-1)}(\vec{x})\right) + b_i^{(\ell)}$$

Weights and biases drawn from Gaussian distributions with mean 0, variances $C_W^{(\ell)}/n_{\ell-1}$, $C_b^{(\ell)}$.

← *initialization hyperparameters*

Statistics of the ensemble (at initialization)

$$P(\phi) = e^{-S[\phi]}$$

We can derive the field theory action $S[\phi]$ (next slide).

Then observables (neuron correlators) can be calculated as in field theory:

$$\left\langle \phi_{i_1}^{(\ell)}(\vec{x}_1) \;\dots\; \phi_{i_{2k}}^{(\ell)}(\vec{x}_{2k}) \right\rangle = \int \mathcal{D}\phi\, \phi_{i_1}^{(\ell)}(\vec{x}_1) \;\dots\; \phi_{i_{2k}}^{(\ell)}(\vec{x}_{2k})\, e^{-S[\phi]} \,.$$

And we can study e.g. how they evolve from layer to layer $\Rightarrow$ RG flow.

(which can tell us about how deep neural networks process information)

Deriving the EFT action $S[\phi]$

$$e^{-\mathcal{S}} = P\big(\phi^{(1)}, \ldots, \phi^{(L)}\big) = P\big(\phi^{(1)}\big)\, P\big(\phi^{(2)}|\phi^{(1)}\big)\ \ldots\ P\big(\phi^{(L)}|\phi^{(L\text{-}1)}\big)$$

$$P\big(\phi^{(\ell)}|\phi^{(\ell\text{-}1)}\big) = \prod_{i,j} \int dW_{ij}\, P_W^{(\ell)}(W_{ij}) \prod_i \int db_i\, P_b^{(\ell)}(b_i) \prod_{i,\vec{x}} \delta\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell\text{-}1}} W_{ij}\, \sigma\big(\phi_j^{(\ell\text{-}1)}(\vec{x})\big) - b_i\Big)$$

$$\frac{1}{\sqrt{2\pi C_W^{(\ell)}/n_{\ell\text{-}1}}} \exp\left(-\frac{W^2}{2C_W^{(\ell)}/n_{\ell\text{-}1}}\right) \qquad \frac{1}{\sqrt{2\pi C_b^{(\ell)}}} \exp\left(-\frac{b^2}{2C_b^{(\ell)}}\right) \qquad \int \frac{d\Lambda_i(\vec{x})}{2\pi}\, \exp\Big[i\,\Lambda_i(\vec{x})\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell\text{-}1}} W_{ij}\sigma\big(\phi_j^{(\ell\text{-}1)}(\vec{x})\big) - b_i\Big)\Big]$$

Deriving the EFT action $S[\phi]$

$$e^{-S} = P(\phi^{(1)}, \ldots, \phi^{(L)}) = P(\phi^{(1)})\, P(\phi^{(2)}|\phi^{(1)}) \,\ldots\, P(\phi^{(L)}|\phi^{(L-1)})$$

$$P(\phi^{(\ell)}|\phi^{(\ell-1)}) = \prod_{i,j} \int dW_{ij}\, P_W^{(\ell)}(W_{ij}) \prod_i \int db_i\, P_b^{(\ell)}(b_i) \prod_{i,\vec{x}} \delta\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell-1}} W_{ij}\, \sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big) - b_i\Big)$$

$$\frac{1}{\sqrt{2\pi C_W^{(\ell)}/n_{\ell-1}}} \exp\Big(-\frac{W^2}{2C_W^{(\ell)}/n_{\ell-1}}\Big) \qquad \frac{1}{\sqrt{2\pi C_b^{(\ell)}}} \exp\Big(-\frac{b^2}{2C_b^{(\ell)}}\Big) \qquad \int \frac{d\Lambda_i(\vec{x})}{2\pi} \exp\Big[i\,\Lambda_i(\vec{x})\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell-1}} W_{ij}\,\underbrace{\sigma\big(\phi_j^{(\ell-1)}(\vec{x})\big)}_{\sigma_{j,\vec{x}}^{(\ell-1)}} - b_i\Big)\Big]$$

Complete the squares, integrate out W, b, then integrate out Λ (~F.T. of Gaussians) $\Rightarrow$

$$P(\phi^{(\ell)}|\phi^{(\ell-1)}) = \Big[\det\big(2\pi \mathcal{G}^{(\ell)}\big)\Big]^{-\frac{n_\ell}{2}} \exp\Big[-\int d\vec{x}_1 d\vec{x}_2\, \frac{1}{2} \sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1) \big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\, \phi_i^{(\ell)}(\vec{x}_2)\Big]$$

$$\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2) \equiv \frac{1}{n_{\ell-1}} \sum_{j=1}^{n_{\ell-1}} \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2)\,, \qquad \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2) \equiv C_b^{(\ell)} + C_W^{(\ell)}\, \sigma_{j,\vec{x}_1}^{(\ell-1)}\, \sigma_{j,\vec{x}_2}^{(\ell-1)}$$

operator built from $\phi^{(\ell-1)}$ $\Rightarrow$ interactions between adjacent-layer neurons!

Deriving the EFT action $S[\phi]$

Faddeev-Popov?

(this looks like gauge fixing...)

$$e^{-S} = P(\phi^{(1)},\dots,\phi^{(L)}) = P(\phi^{(1)})\, P(\phi^{(2)}|\phi^{(1)}) \,\dots\, P(\cdots|\phi\cdots)$$

$$P(\phi^{(\ell)}|\phi^{(\ell\text{-}1)}) = \prod_{i,j}\int dW_{ij}\, P_W^{(\ell)}(W_{ij}) \prod_i \int db_i\, P_b^{(\ell)}(b_i) \prod_{i,\vec{x}} \delta\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell\text{-}1}} W_{ij}\,\sigma\big(\phi_j^{(\ell\text{-}1)}(\vec{x})\big) - b_i\Big)$$

$$P_W^{(\ell)}(W) = \frac{1}{\sqrt{2\pi C_W^{(\ell)}/n_{\ell\text{-}1}}}\exp\left(-\frac{W^2}{2C_W^{(\ell)}/n_{\ell\text{-}1}}\right) \qquad P_b^{(\ell)}(b) = \frac{1}{\sqrt{2\pi C_b^{(\ell)}}}\exp\left(-\frac{b^2}{2C_b^{(\ell)}}\right)$$

$$\delta(\cdots) = \int \frac{d\Lambda_i(\vec{x})}{2\pi}\exp\Big[i\,\Lambda_i(\vec{x})\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell\text{-}1}} W_{ij}\underbrace{\sigma\big(\phi_j^{(\ell\text{-}1)}(\vec{x})\big)}_{\sigma_{j,\vec{x}}^{(\ell\text{-}1)}} - b_i\Big)\Big]$$

Complete the squares, integrate out W, b, then integrate out Λ (~F.T. of Gaussians) ⇒

$$P(\phi^{(\ell)}|\phi^{(\ell\text{-}1)}) = \Big[\det\big(2\pi\mathcal{G}^{(\ell)}\big)\Big]^{-\frac{n_\ell}{2}}\exp\Big[-\int d\vec{x}_1 d\vec{x}_2\,\frac{1}{2}\sum_{i=1}^{n_\ell}\phi_i^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\phi_i^{(\ell)}(\vec{x}_2)\Big]$$

$$\mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2) \equiv \frac{1}{n_{\ell\text{-}1}}\sum_{j=1}^{n_{\ell\text{-}1}}\mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2)\,, \qquad \mathcal{G}_j^{(\ell)}(\vec{x}_1,\vec{x}_2) \equiv C_b^{(\ell)} + C_W^{(\ell)}\,\sigma_{j,\vec{x}_1}^{(\ell\text{-}1)}\,\sigma_{j,\vec{x}_2}^{(\ell\text{-}1)}$$

operator built from $\phi^{(\ell-1)}$ ⇒ interactions between adjacent-layer neurons!

Deriving the EFT action $S[\phi]$

Faddeev-Popov?

(this looks like gauge fixing...)

$$e^{-\mathcal{S}} = P\big(\phi^{(1)},\dots,\phi^{(L)}\big) = P\big(\phi^{(1)}\big)\, P\big(\phi^{(2)}|\phi^{(1)}\big)\ \dots\ P\big($$

$$P\big(\phi^{(\ell)}|\phi^{(\ell\text{-}1)}\big) = \prod_{i,j}\int dW_{ij}\, P_W^{(\ell)}(W_{ij}) \prod_i \int db_i\, P_b^{(\ell)}(b_i) \prod_{i,\vec{x}} \delta\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell\text{-}1}} W_{ij}\,\sigma\big(\phi_j^{(\ell\text{-}1)}(\vec{x})\big) - b_i\Big)$$

$$\frac{1}{\sqrt{2\pi C_W^{(\ell)}/n_{\ell\text{-}1}}}\exp\left(-\frac{W^2}{2C_W^{(\ell)}/n_{\ell\text{-}1}}\right) \qquad \frac{1}{\sqrt{2\pi C_b^{(\ell)}}}\exp\left(-\frac{b^2}{2C_b^{(\ell)}}\right) \qquad \int\frac{d\Lambda_i(\vec{x})}{2\pi}\exp\left[i\,\Lambda_i(\vec{x})\Big(\phi_i^{(\ell)}(\vec{x}) - \sum_{j=1}^{n_{\ell\text{-}1}} W_{ij}\sigma\big(\phi_j^{(\ell\text{-}1)}(\vec{x})\big) - b_i\Big)\right]$$

Complete the squares, integrate out W, b, then integrate out Λ (~F.T. of Gaussians) $\Rightarrow$

$$P\big(\phi^{(\ell)}|\phi^{(\ell\text{-}1)}\big) = \Big[\det\big(2\pi\mathcal{G}^{(\ell)}\big)\Big]^{-\frac{n_\ell}{2}} \exp\left[-\int d\vec{x}_1 d\vec{x}_2\, \frac{1}{2}\sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\phi_i^{(\ell)}(\vec{x}_2)\right]$$

$$\int \mathcal{D}\psi\mathcal{D}\bar{\psi}\ \exp\left[\sum_{i'=1}^{n_\ell/2} \bar{\psi}_{i'}^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\psi_{i'}^{(\ell)}(\vec{x}_2)\right]$$

ghosts!

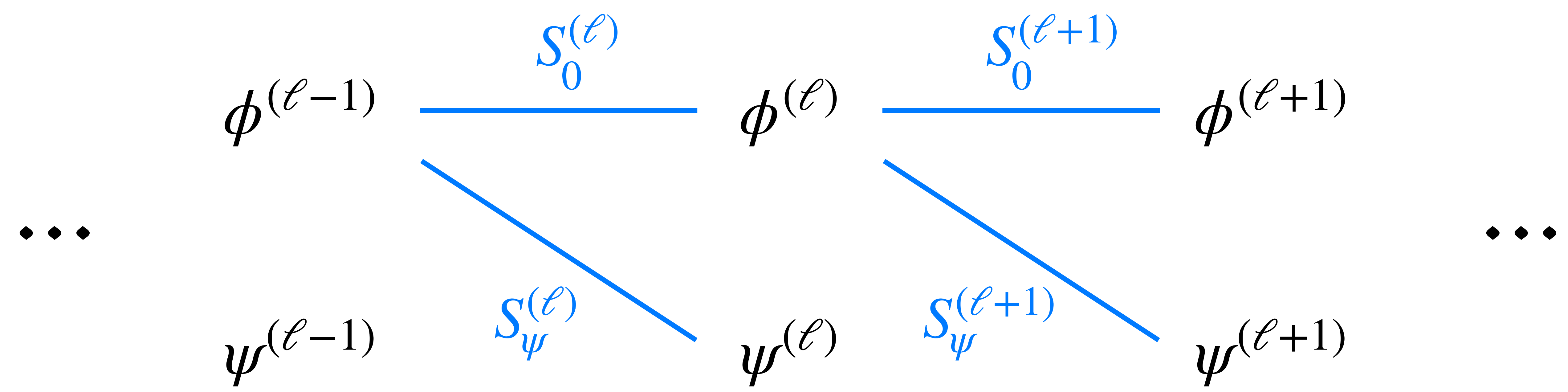
Deriving the EFT action $S[\phi]$

$$e^{-\mathcal{S}} = P\big(\phi^{(1)}, \ldots, \phi^{(L)}\big) = P\big(\phi^{(1)}\big)\, P\big(\phi^{(2)} | \phi^{(1)}\big) \;\ldots\; P\big(\phi^{(L)} | \phi^{(L\text{-}1)}\big)$$

$$= \int \mathcal{D}\psi \mathcal{D}\bar{\psi}\; e^{-\sum_{\ell=1}^{L}\left(\mathcal{S}_0^{(\ell)}[\phi] + \mathcal{S}_\psi^{(\ell)}[\phi,\psi,\bar{\psi}]\right)}$$

$$\mathcal{S}_0^{(\ell)} = \int d\vec{x}_1 d\vec{x}_2\, \frac{1}{2} \sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1) \big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1, \vec{x}_2)\, \phi_i^{(\ell)}(\vec{x}_2)$$

$$\mathcal{S}_\psi^{(\ell)} = -\int d\vec{x}_1 d\vec{x}_2 \sum_{i'=1}^{n_\ell/2} \bar{\psi}_{i'}^{(\ell)}(\vec{x}_1) \big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1, \vec{x}_2)\, \psi_{i'}^{(\ell)}(\vec{x}_2)$$

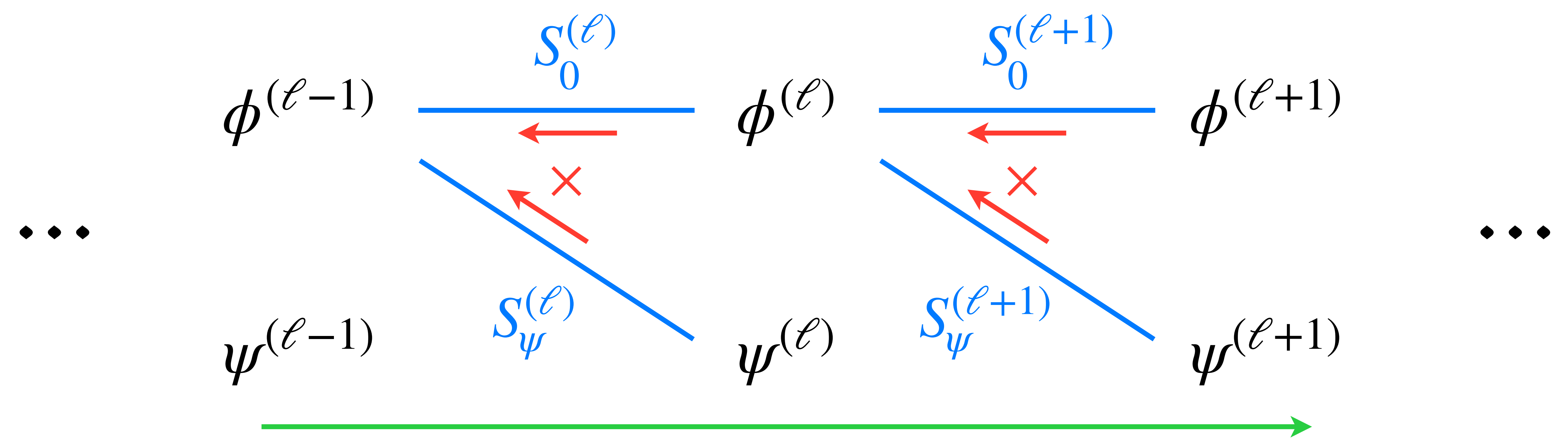


Deriving the EFT action $S[\phi]$

$$e^{-\mathcal{S}} = P\big(\phi^{(1)},\dots,\phi^{(L)}\big) = P\big(\phi^{(1)}\big)\, P\big(\phi^{(2)}|\phi^{(1)}\big)\ \dots\ P\big(\phi^{(L)}|\phi^{(L\text{-}1)}\big)$$

$$= \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\ e^{-\sum_{\ell=1}^{L}\left(\mathcal{S}_0^{(\ell)}[\phi]+\mathcal{S}_\psi^{(\ell)}[\phi,\psi,\bar{\psi}]\right)}$$

$$\mathcal{S}_0^{(\ell)} = \int d\vec{x}_1 d\vec{x}_2\, \frac{1}{2}\sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\phi_i^{(\ell)}(\vec{x}_2) \qquad \mathcal{S}_\psi^{(\ell)} = -\int d\vec{x}_1 d\vec{x}_2 \sum_{i'=1}^{n_\ell/2} \bar{\psi}_{i'}^{(\ell)}(\vec{x}_1)\big(\mathcal{G}^{(\ell)}\big)^{-1}(\vec{x}_1,\vec{x}_2)\,\psi_{i'}^{(\ell)}(\vec{x}_2)$$



Network has directionality! (Loop corrections cancel between ϕ and ψ when going backward.)

When calculating neuron correlators $\langle \phi_{i_1}^{(\ell)}(\vec{x}_1)\ \dots\ \phi_{i_{2k}}^{(\ell)}(\vec{x}_{2k})\rangle$, ghosts do not enter.

Outline

1. Neural networks ↔ field theories (high-level summary).

2. EFT of deep neural networks (at initialization).

3. Diagrammatic approach.

4. Structures of neural network EFTs and criticality.

Deriving Feynman rules

$$\mathcal{S}_0^{(\ell)} = \int d\vec{x}_1 d\vec{x}_2 \, \frac{1}{2} \sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1) \left(\mathcal{G}^{(\ell)}\right)^{-1}(\vec{x}_1, \vec{x}_2) \, \phi_i^{(\ell)}(\vec{x}_2)$$

$$\mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \equiv \frac{1}{n_{\ell-1}} \sum_{j=1}^{n_{\ell-1}} \mathcal{G}_j^{(\ell)}(\vec{x}_1, \vec{x}_2)\,, \qquad \mathcal{G}_j^{(\ell)}(\vec{x}_1, \vec{x}_2) \equiv C_b^{(\ell)} + C_W^{(\ell)} \, \sigma_{j,\vec{x}_1}^{(\ell-1)} \, \sigma_{j,\vec{x}_2}^{(\ell-1)} = \mathcal{G}_j^{(\ell)}(\vec{x}_2, \vec{x}_1)$$

↑ operator built from $\phi^{(\ell-1)}$

If $\phi^{(\ell-1)}$ were classical background ⇒ free theory for $\phi^{(\ell)}$.

$$\left\langle \phi_{i_1}^{(\ell)}(\vec{x}_1) \, \phi_{i_2}^{(\ell)}(\vec{x}_2) \right\rangle = \delta_{i_1 i_2} \; \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2)$$

$$\left\langle \phi_{i_1}^{(\ell)}(\vec{x}_1) \, \phi_{i_2}^{(\ell)}(\vec{x}_2) \, \phi_{i_3}^{(\ell)}(\vec{x}_3) \, \phi_{i_4}^{(\ell)}(\vec{x}_4) \right\rangle = \delta_{i_1 i_2} \delta_{i_3 i_4} \; \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \, \mathcal{G}^{(\ell)}(\vec{x}_3, \vec{x}_4) \; + \text{perms.}$$

Deriving Feynman rules

$$\mathcal{S}_0^{(\ell)} = \int d\vec{x}_1 d\vec{x}_2 \, \frac{1}{2} \sum_{i=1}^{n_\ell} \phi_i^{(\ell)}(\vec{x}_1) \left(\mathcal{G}^{(\ell)}\right)^{-1}(\vec{x}_1, \vec{x}_2) \, \phi_i^{(\ell)}(\vec{x}_2)$$

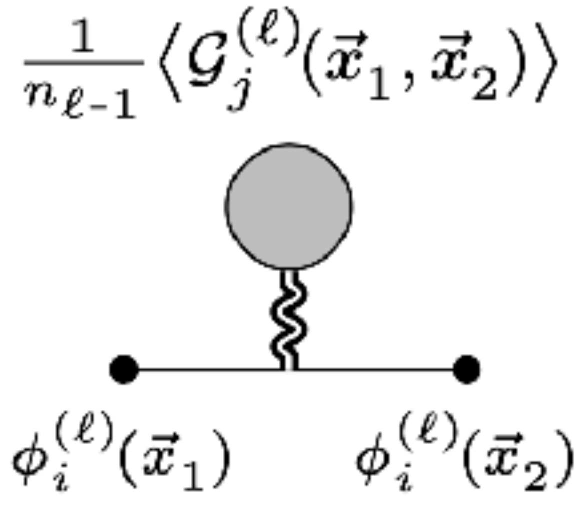
$$\mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \equiv \frac{1}{n_{\ell-1}} \sum_{j=1}^{n_{\ell-1}} \mathcal{G}_j^{(\ell)}(\vec{x}_1, \vec{x}_2) \,, \qquad \mathcal{G}_j^{(\ell)}(\vec{x}_1, \vec{x}_2) \equiv C_b^{(\ell)} + C_W^{(\ell)} \, \sigma_{j,\vec{x}_1}^{(\ell-1)} \, \sigma_{j,\vec{x}_2}^{(\ell-1)} = \mathcal{G}_j^{(\ell)}(\vec{x}_2, \vec{x}_1)$$

↑ operator built from $\phi^{(\ell-1)}$

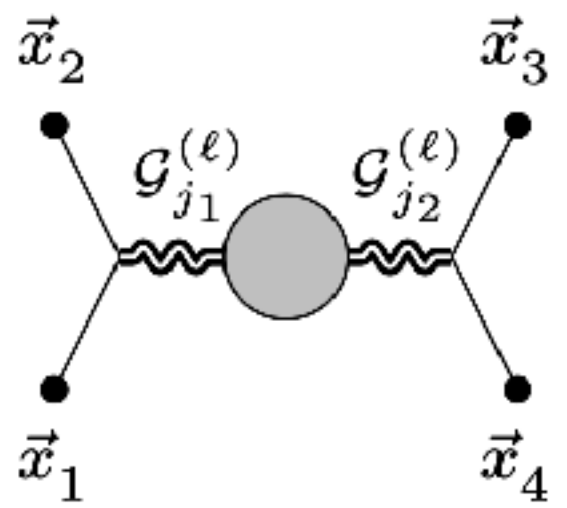
If $\phi^{(\ell-1)}$ were classical background $\Rightarrow$ free theory for $\phi^{(\ell)}$.

In reality $\phi^{(\ell-1)}$ have statistical fluctuations.

$$\langle \phi_{i_1}^{(\ell)}(\vec{x}_1) \, \phi_{i_2}^{(\ell)}(\vec{x}_2) \rangle = \delta_{i_1 i_2} \langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \rangle = \delta_{i_1 i_2} \sum_j$$



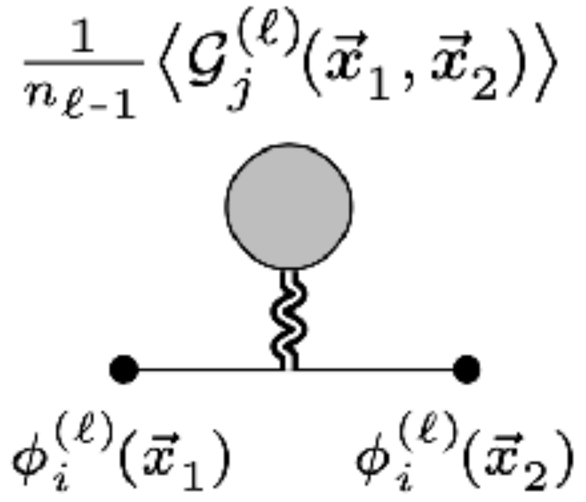
$$\langle \phi_{i_1}^{(\ell)}(\vec{x}_1) \, \phi_{i_2}^{(\ell)}(\vec{x}_2) \, \phi_{i_3}^{(\ell)}(\vec{x}_3) \, \phi_{i_4}^{(\ell)}(\vec{x}_4) \rangle = \delta_{i_1 i_2} \delta_{i_3 i_4} \langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \, \mathcal{G}^{(\ell)}(\vec{x}_3, \vec{x}_4) \rangle + \text{perms.} \quad = \delta_{i_1 i_2} \delta_{i_3 i_4} \sum_{j_1, j_2}$$



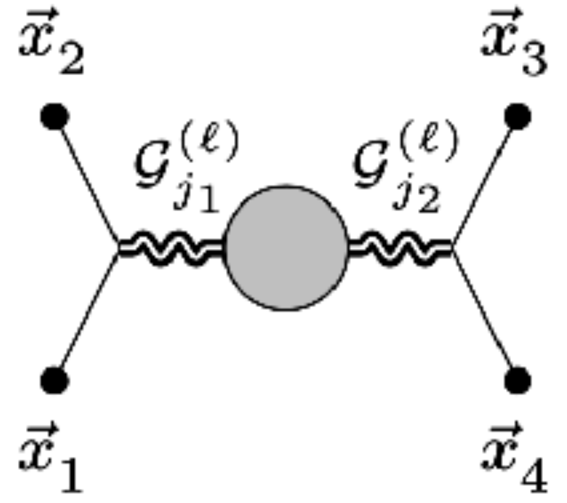
$+ \text{perms.}$

Deriving Feynman rules

$$\left\langle \phi^{(\ell)}_{i_1}(\vec{x}_1)\,\phi^{(\ell)}_{i_2}(\vec{x}_2)\right\rangle = \delta_{i_1 i_2}\left\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\right\rangle = \delta_{i_1 i_2}\sum_{j}$$

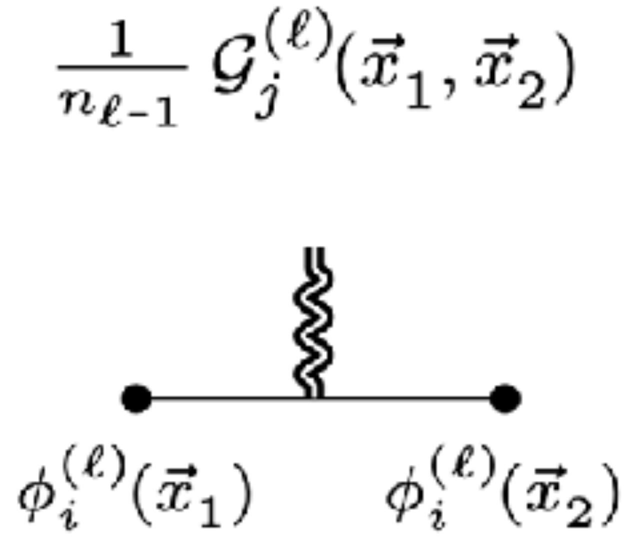


$$\left\langle \phi^{(\ell)}_{i_1}(\vec{x}_1)\,\phi^{(\ell)}_{i_2}(\vec{x}_2)\,\phi^{(\ell)}_{i_3}(\vec{x}_3)\,\phi^{(\ell)}_{i_4}(\vec{x}_4)\right\rangle = \delta_{i_1 i_2}\delta_{i_3 i_4}\left\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\,\mathcal{G}^{(\ell)}(\vec{x}_3,\vec{x}_4)\right\rangle + \text{perms.} \qquad = \delta_{i_1 i_2}\delta_{i_3 i_4}\sum_{j_1,j_2}$$



$+ \text{perms.}$

Effectively, we can simply use the following Feynman rule to build up diagrams.

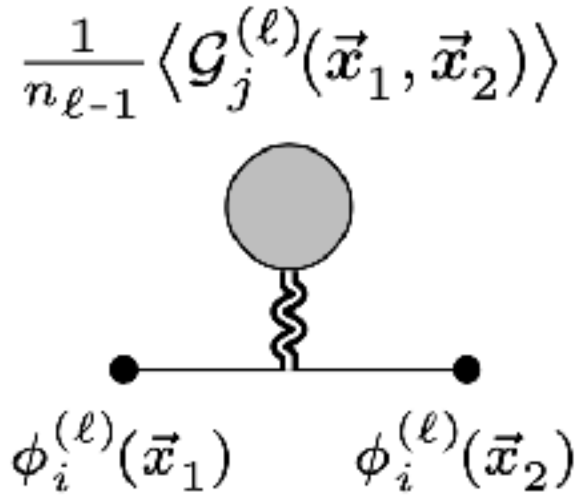


⟵ just be sure to attach this to a blob
(no external wavy lines)

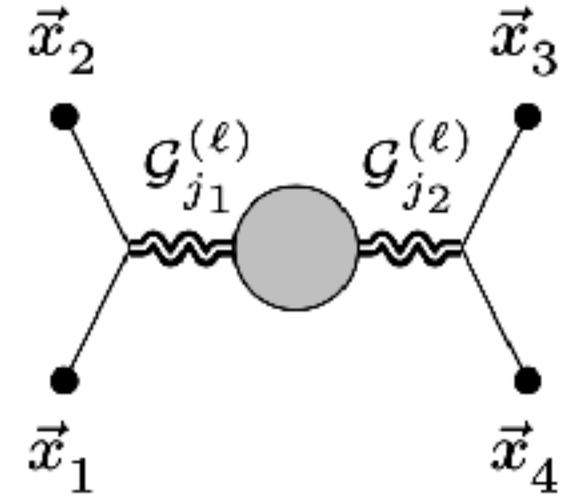
$$\mathcal{G}^{(\ell)}_j(\vec{x}_1,\vec{x}_2) \equiv C^{(\ell)}_b + C^{(\ell)}_W\,\sigma^{(\ell-1)}_{j,\vec{x}_1}\,\sigma^{(\ell-1)}_{j,\vec{x}_2}$$

Deriving Feynman rules

$$\left\langle \phi^{(\ell)}_{i_1}(\vec{x}_1)\, \phi^{(\ell)}_{i_2}(\vec{x}_2) \right\rangle = \delta_{i_1 i_2} \left\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \right\rangle = \delta_{i_1 i_2} \sum_j$$



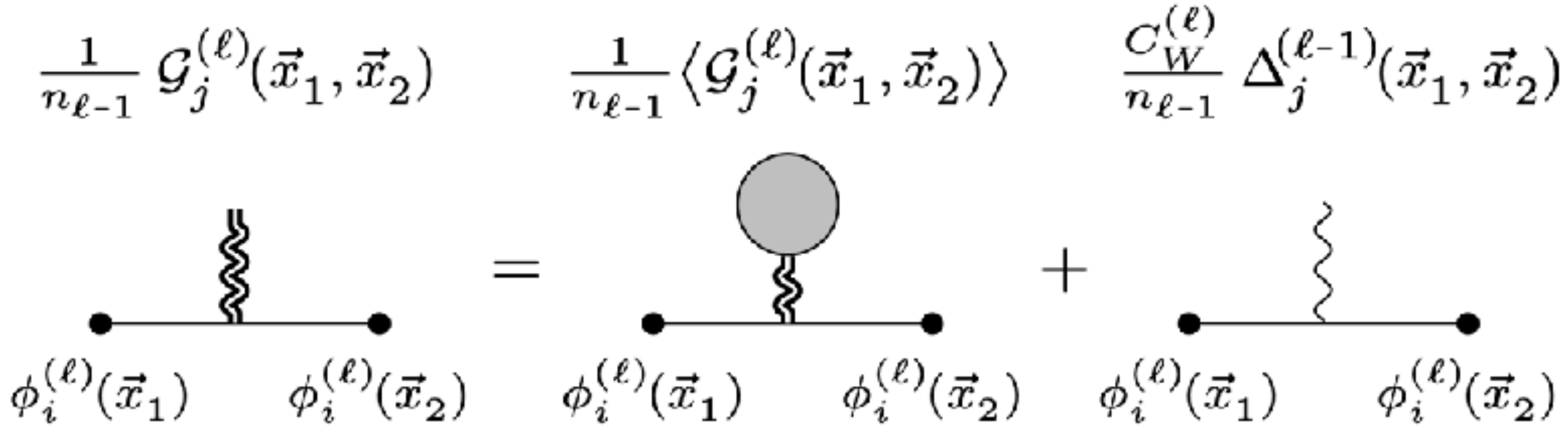
$$\left\langle \phi^{(\ell)}_{i_1}(\vec{x}_1)\, \phi^{(\ell)}_{i_2}(\vec{x}_2)\, \phi^{(\ell)}_{i_3}(\vec{x}_3)\, \phi^{(\ell)}_{i_4}(\vec{x}_4) \right\rangle = \delta_{i_1 i_2} \delta_{i_3 i_4} \left\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2)\, \mathcal{G}^{(\ell)}(\vec{x}_3, \vec{x}_4) \right\rangle + \text{perms.} \quad = \delta_{i_1 i_2} \delta_{i_3 i_4} \sum_{j_1, j_2}$$



$$+ \text{perms.}$$

Effectively, we can simply use the following Feynman rule to build up diagrams.

Further decompose into vev + fluctuation.

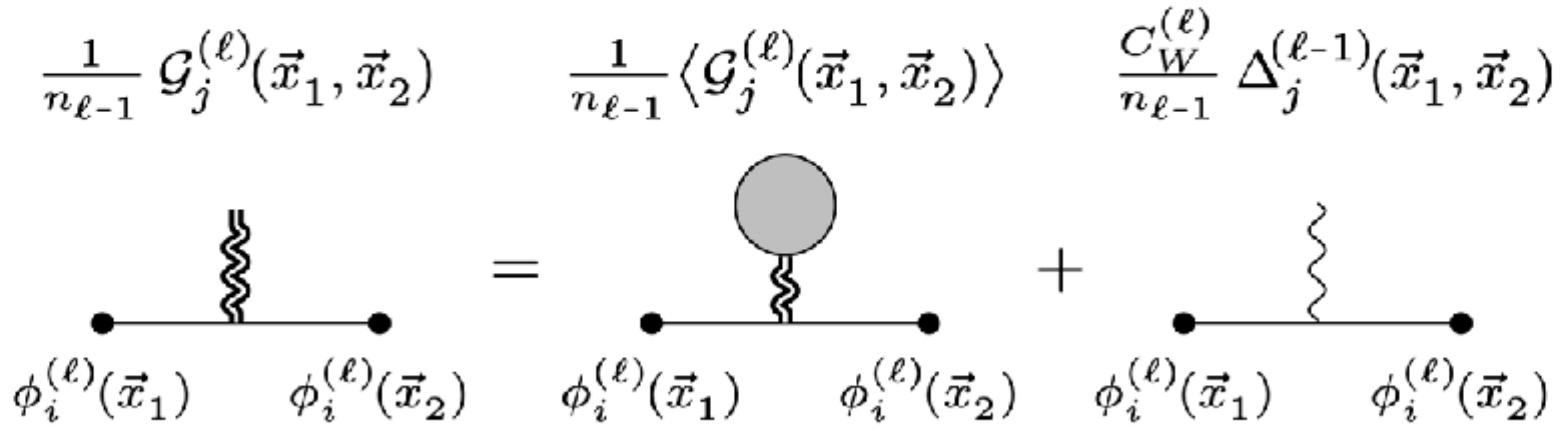


$$\mathcal{G}^{(\ell)}_j(\vec{x}_1, \vec{x}_2) \equiv C^{(\ell)}_b + C^{(\ell)}_W\, \sigma^{(\ell\text{-}1)}_{j,\vec{x}_1}\, \sigma^{(\ell\text{-}1)}_{j,\vec{x}_2}$$

$$\Delta^{(\ell\text{-}1)}_j(\vec{x}_1, \vec{x}_2) \equiv \sigma^{(\ell\text{-}1)}_{j,\vec{x}_1}\, \sigma^{(\ell\text{-}1)}_{j,\vec{x}_2} - \left\langle \sigma^{(\ell\text{-}1)}_{j,\vec{x}_1}\, \sigma^{(\ell\text{-}1)}_{j,\vec{x}_2} \right\rangle$$

Only fluctuation piece (Δ, single wavy line) contributes to connected correlators.

1/n expansion



$\longleftarrow$ Each interaction vertex is $\sim \dfrac{1}{n}$

Infinitely-wide network ($n \to \infty$) $\Rightarrow$ free theory.

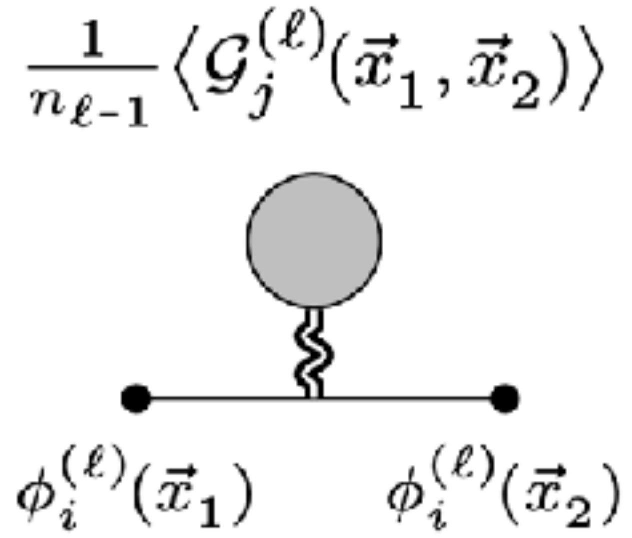
Finitely-wide network (most relevant in practice) $\Rightarrow$ weakly-interacting theory.

Observables calculated order by order in *1/n*.

Interested in RG flows of connected correlators $\langle \phi^{2k} \rangle_c \sim \mathcal{O}(n^{1-k})$.

2-point correlator

$$\langle \phi^{(\ell)}_{i_1}(\vec{x}_1)\, \phi^{(\ell)}_{i_2}(\vec{x}_2) \rangle = \delta_{i_1 i_2} \sum_j$$



$$= \delta_{i_1 i_2} \langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \rangle$$

Expand in *1/n*: $\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \rangle = \sum_{p=0}^{\infty} \frac{1}{n^p_{\ell-1}} \mathcal{K}^{(\ell)}_p(\vec{x}_1, \vec{x}_2)$

Recall: $\mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \equiv \frac{1}{n_{\ell-1}} \sum_{j=1}^{n_{\ell-1}} \mathcal{G}^{(\ell)}_j(\vec{x}_1, \vec{x}_2)\,,$ $\quad \mathcal{G}^{(\ell)}_j(\vec{x}_1, \vec{x}_2) \equiv C^{(\ell)}_b + C^{(\ell)}_W\, \sigma^{(\ell-1)}_{j,\vec{x}_1}\, \sigma^{(\ell-1)}_{j,\vec{x}_2}$ (operators of $\phi^{(\ell-1)}$)

Leading order (LO): use LO (free-theory) propagators to evaluate $\langle \cdots \rangle$.

$$\mathcal{K}^{(\ell)}_0(\vec{x}_1, \vec{x}_2) = \sum_j \frac{1}{n_{\ell-1}} \langle \mathcal{G}^{(\ell)}_j(\vec{x}_1, \vec{x}_2) \rangle_{\mathcal{K}^{(\ell-1)}_0} = C^{(\ell)}_b + C^{(\ell)}_W \langle \sigma_{\vec{x}_1} \sigma_{\vec{x}_2} \rangle_{\mathcal{K}^{(\ell-1)}_0}$$

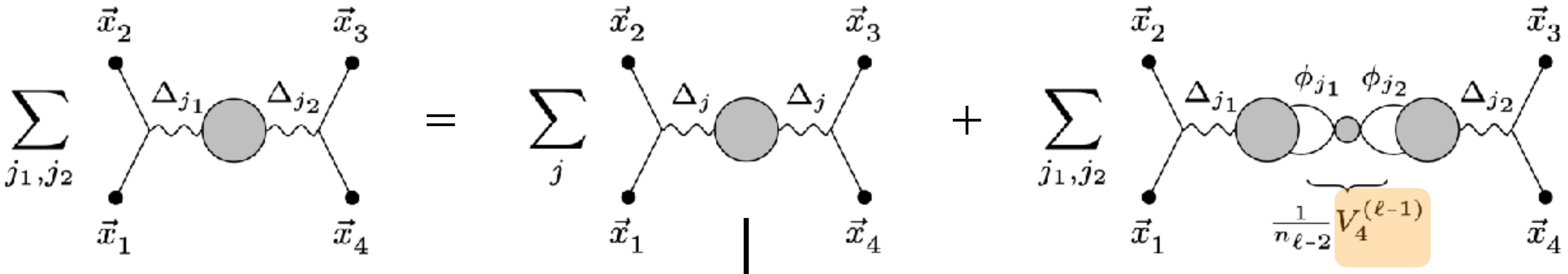
"Kernel recursion" (RG flow of $\mathcal{K}_0$, UV boundary condition $\mathcal{K}^{(1)}_0(\vec{x}_1, \vec{x}_2) = C^{(1)}_b + \frac{C^{(1)}_W}{n_0} \vec{x}_1 \cdot \vec{x}_2$).

(well-known in ML literature)

Connected 4-point correlator

$$\left\langle \phi^{(\ell)}_{i_1}(\vec{x}_1)\, \phi^{(\ell)}_{i_2}(\vec{x}_2)\, \phi^{(\ell)}_{i_3}(\vec{x}_3)\, \phi^{(\ell)}_{i_4}(\vec{x}_4) \right\rangle_{\mathrm{C}} = \delta_{i_1 i_2}\delta_{i_3 i_4}\, \frac{1}{n_{\ell-1}}\, V^{(\ell)}_4(\vec{x}_1, \vec{x}_2; \vec{x}_3, \vec{x}_4) + \text{perms.}$$

$$\frac{1}{n_{\ell-1}} V^{(\ell)}_4(\vec{x}_1, \vec{x}_2; \vec{x}_3, \vec{x}_4) = \sum_{j_1, j_2} \;[\text{diagram}]\; = \sum_{j} \;[\text{diagram}]\; + \sum_{j_1, j_2} \;[\text{diagram}]$$



$$\frac{\left(C^{(\ell)}_W\right)^2}{n_{\ell-1}} \left\langle \Delta(\vec{x}_1, \vec{x}_2) \Delta(\vec{x}_3, \vec{x}_4) \right\rangle_{\mathcal{K}^{(\ell-1)}_0} + \mathcal{O}\!\left(\frac{1}{n^2}\right)$$

$$\frac{\left(C^{(\ell)}_W\right)^2}{4\, n_{\ell-2}} \prod_{\alpha=1}^{4} \int d\vec{y}_\alpha d\vec{z}_\alpha \left(\mathcal{K}^{(\ell-1)}_0\right)^{-1}(\vec{y}_\alpha, \vec{z}_\alpha)\, V^{(\ell-1)}_4(\vec{y}_1, \vec{y}_2; \vec{y}_3, \vec{y}_4) \left\langle \Delta(\vec{x}_1, \vec{x}_2)\, \phi(\vec{z}_1)\, \phi(\vec{z}_2) \right\rangle_{\mathcal{K}^{(\ell-1)}_0} \left\langle \Delta(\vec{x}_3, \vec{x}_4)\, \phi(\vec{z}_3)\, \phi(\vec{z}_4) \right\rangle_{\mathcal{K}^{(\ell-1)}_0} + \mathcal{O}\!\left(\frac{1}{n^2}\right)$$

symmetry factor

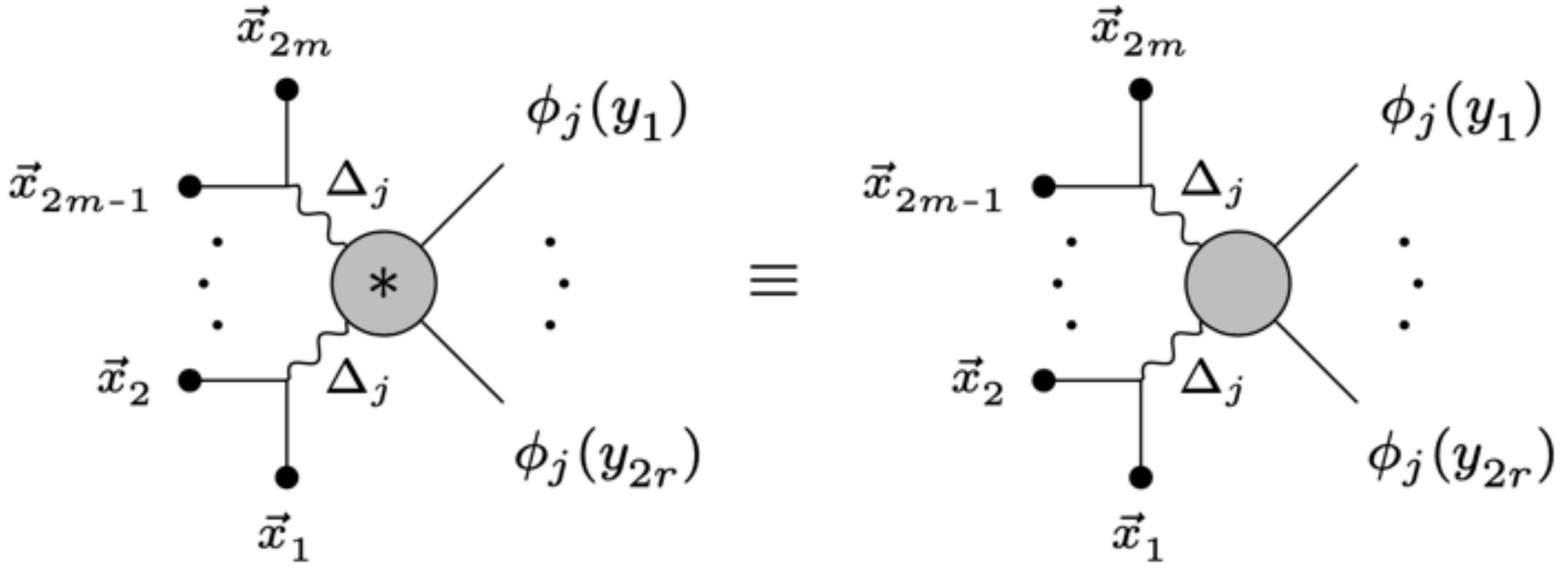
$$= \frac{\left(C^{(\ell)}_W\right)^2}{4\, n_{\ell-2}} \prod_{\alpha=1}^{4} \int d\vec{y}_\alpha\, V^{(\ell-1)}_4(\vec{y}_1, \vec{y}_2; \vec{y}_3, \vec{y}_4) \left\langle \frac{\delta^2 \Delta(\vec{x}_1, \vec{x}_2)}{\delta\phi(\vec{y}_1)\delta\phi(\vec{y}_2)} \right\rangle_{\mathcal{K}^{(\ell-1)}_0} \left\langle \frac{\delta^2 \Delta(\vec{x}_3, \vec{x}_4)}{\delta\phi(\vec{y}_3)\delta\phi(\vec{y}_4)} \right\rangle_{\mathcal{K}^{(\ell-1)}_0} + \mathcal{O}\!\left(\frac{1}{n^2}\right)$$

Wick contraction

(in agreement with Yaida '19, Roberts, Yaida, Hanin '21)

Progressing to higher orders

Basic building blocks are $*$-blobs:



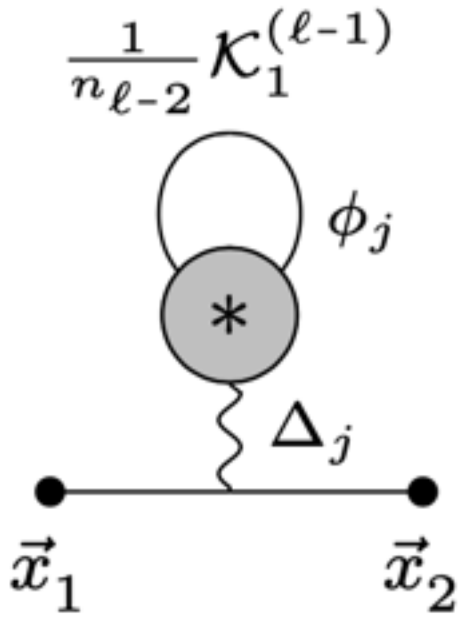
$$- \left(\begin{array}{c} \text{diagrams where the } \phi^{2r}\Delta^m \text{ blob becomes} \\ \text{disconnected due to contractions among } \phi\text{'s} \end{array} \right)$$

$$= \left(\frac{C_W^{(\ell)}}{n_{\ell\text{-}1}} \right)^m \int \prod_{\alpha=1}^{2r} d\vec{z}_\alpha \, \mathcal{K}_0^{(\ell\text{-}1)}(\vec{y}_\alpha, \vec{z}_\alpha) \left\langle \frac{\delta^{2r} \left(\Delta(\vec{x}_1, \vec{x}_2) \cdots \Delta(\vec{x}_{2m\text{-}1}, \vec{x}_{2m}) \right)}{\delta\phi(\vec{z}_1) \cdots \delta\phi(\vec{z}_{2r})} \right\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}}$$

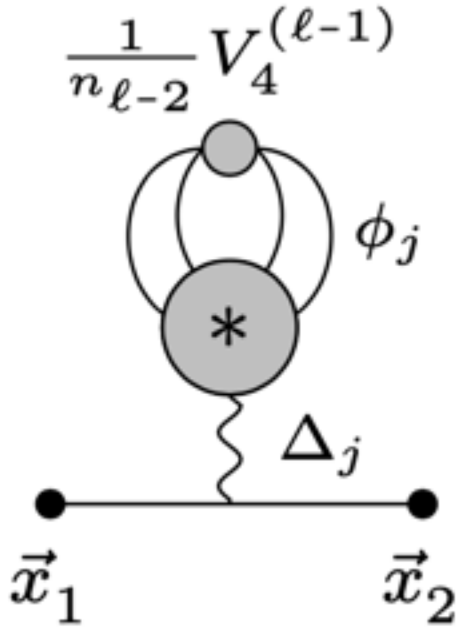
2-point correlator, NLO

$$\left\langle \mathcal{G}^{(\ell)}(\vec{x}_1, \vec{x}_2) \right\rangle = \sum_{p=0}^{\infty} \frac{1}{n_{\ell-1}^{p}} \, \mathcal{K}_p^{(\ell)}(\vec{x}_1, \vec{x}_2)$$

$$\frac{1}{n_{\ell-1}} \, \mathcal{K}_1^{(\ell)}(\vec{x}_1, \vec{x}_2) = \sum_j$$



$$+ \sum_j$$



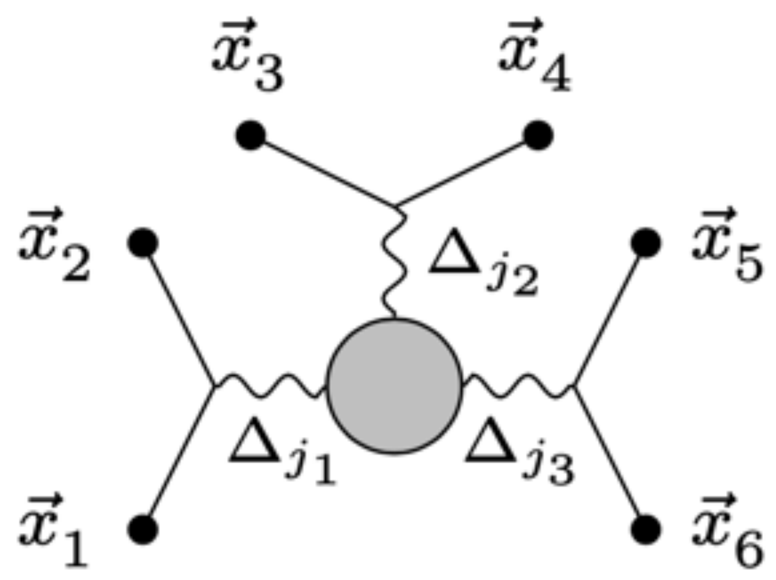
$$\frac{C_W^{(\ell)}}{2\, n_{\ell-2}} \int d\vec{y}_1 \, d\vec{y}_2 \; \mathcal{K}_1^{(\ell-1)}(\vec{y}_1, \vec{y}_2) \left\langle \frac{\delta^2 \Delta(\vec{x}_1, \vec{x}_2)}{\delta\phi(\vec{y}_1)\, \delta\phi(\vec{y}_2)} \right\rangle_{\mathcal{K}_0^{(\ell-1)}}$$

$$\frac{C_W^{(\ell)}}{8\, n_{\ell-2}} \int \prod_{\alpha=1}^{4} d\vec{y}_\alpha \; V_4^{(\ell-1)}(\vec{y}_1, \vec{y}_2; \vec{y}_3, \vec{y}_4) \left\langle \frac{\delta^4 \Delta(\vec{x}_1, \vec{x}_2)}{\delta\phi(\vec{y}_1)\, \delta\phi(\vec{y}_2)\, \delta\phi(\vec{y}_3)\, \delta\phi(\vec{y}_4)} \right\rangle_{\mathcal{K}_0^{(\ell-1)}}$$

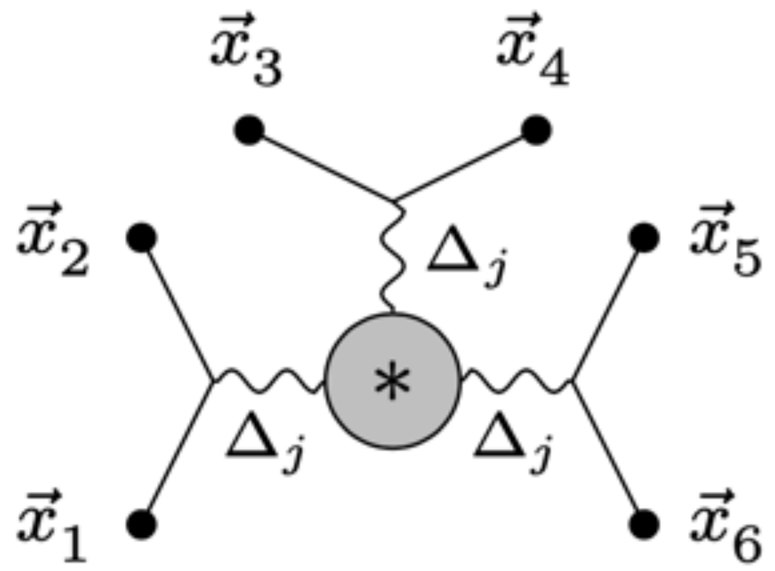
(in agreement with Roberts, Yaida, Hanin '21)

Connected 6-point correlator

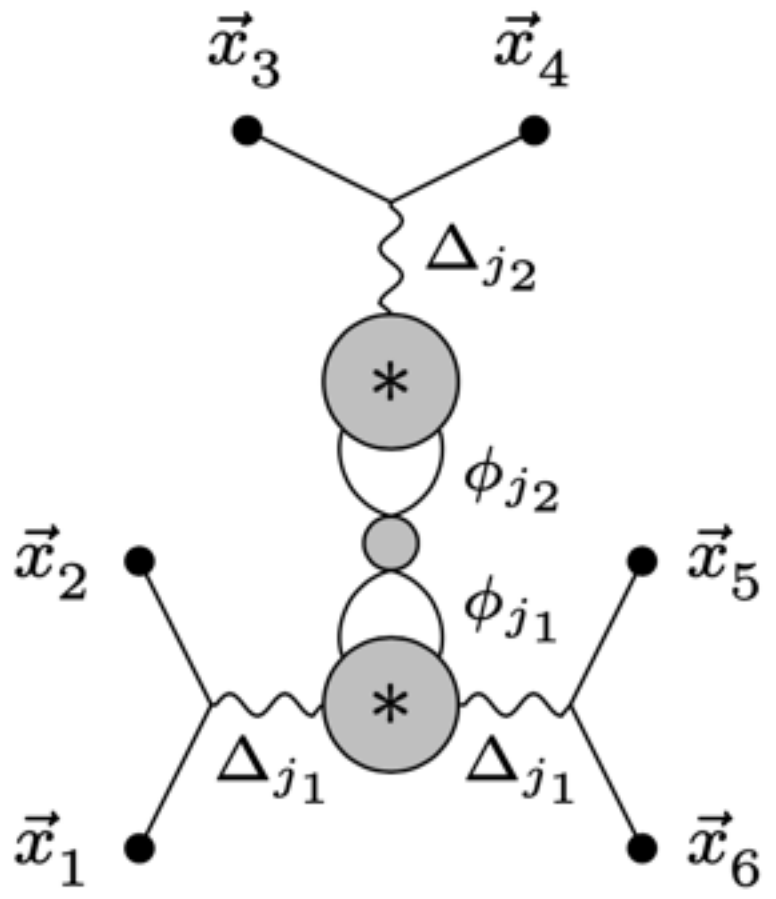
$$\frac{1}{n_{\ell-1}^2} V_6^{(\ell)}(\vec{x}_1, \vec{x}_2; \vec{x}_3, \vec{x}_4; \vec{x}_5, \vec{x}_6) = \sum_{j_1, j_2, j_3}$$



$$= \sum_{j}$$

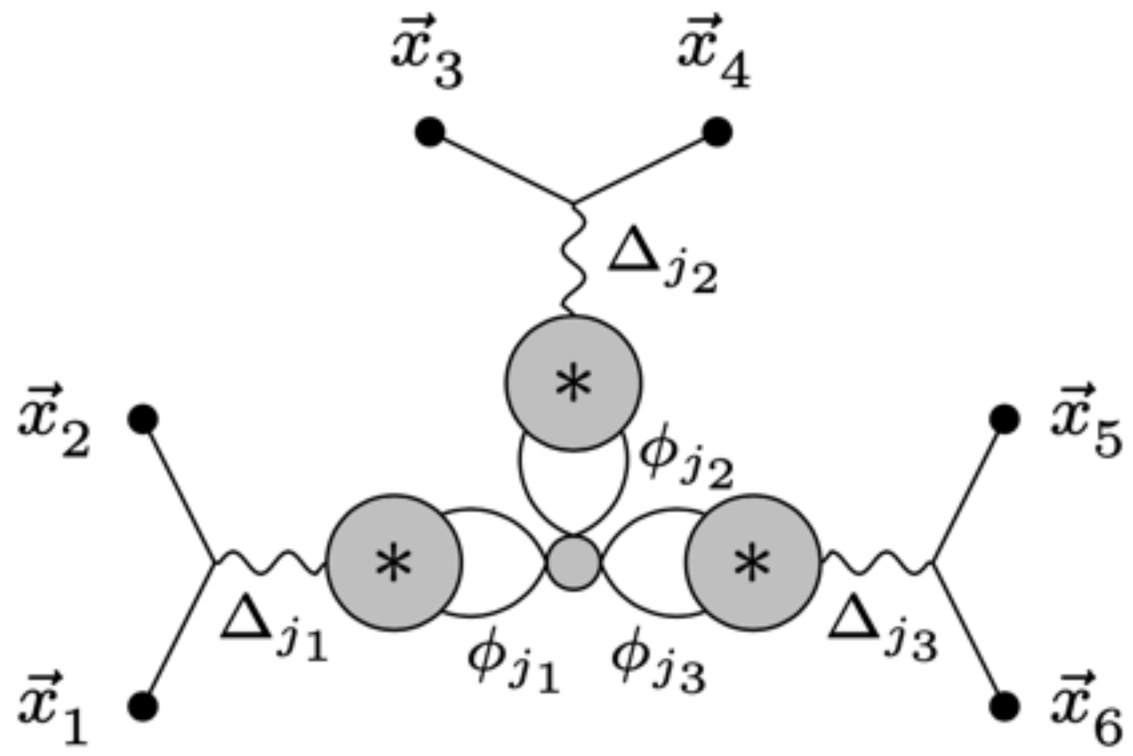


$$+ \sum_{j_1, j_2}$$

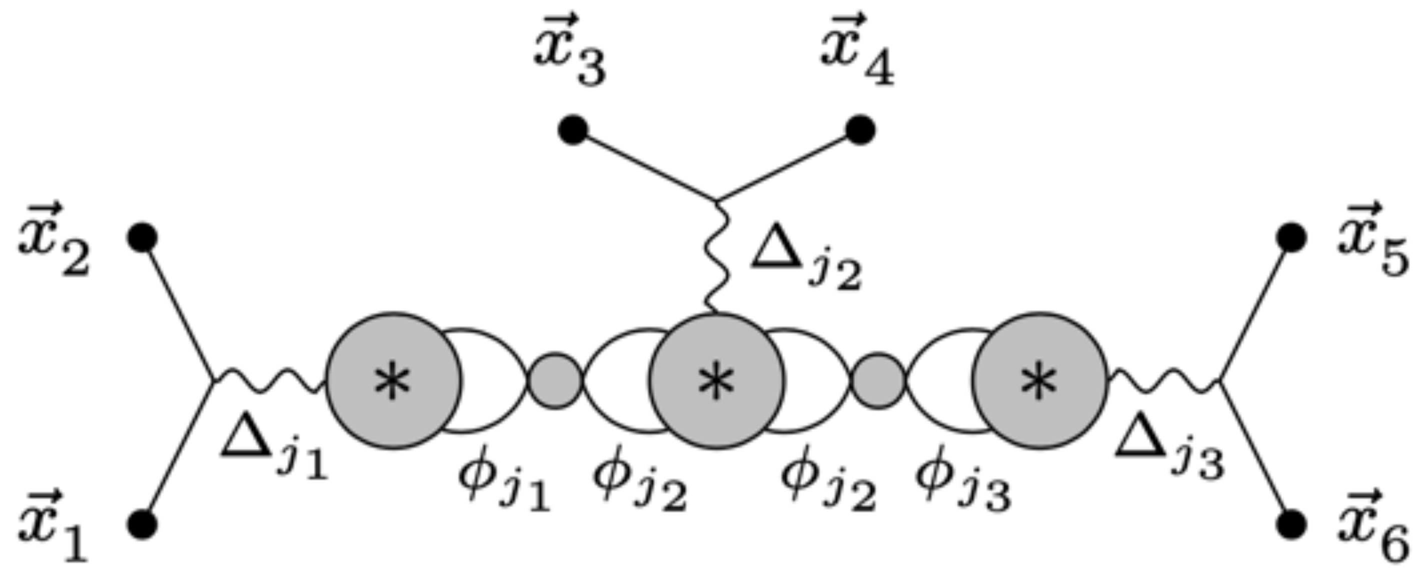


$$+ \text{ perms.}$$

$$+ \sum_{j_1, j_2, j_3}$$

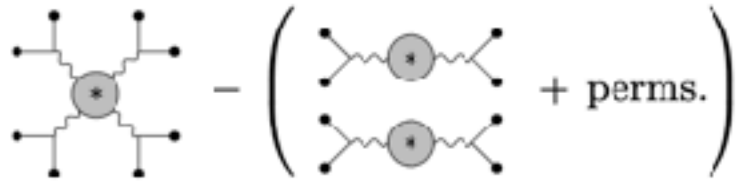
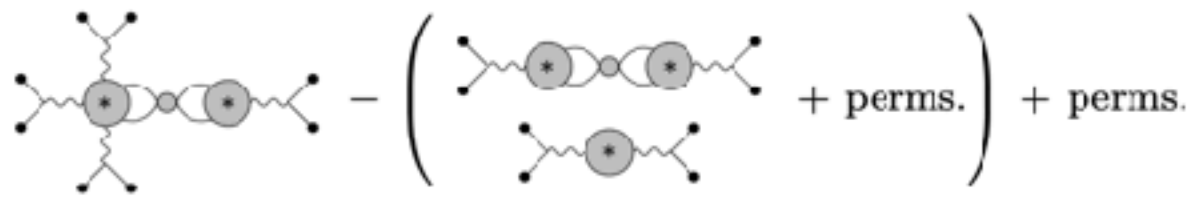
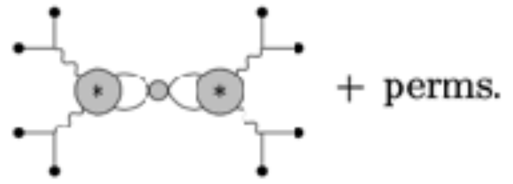
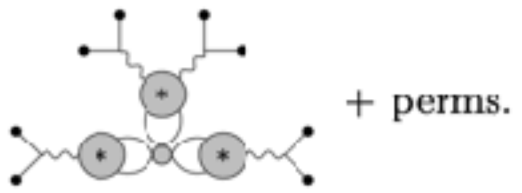
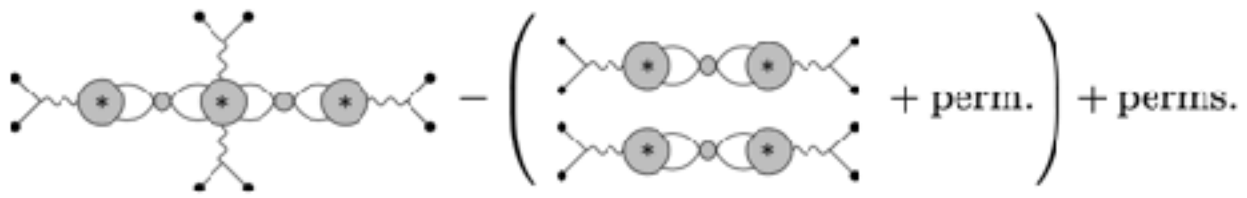
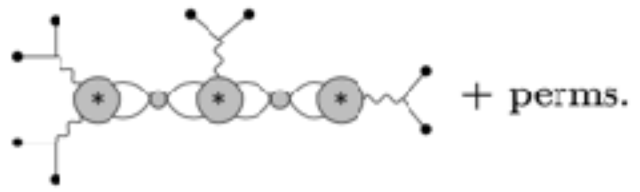
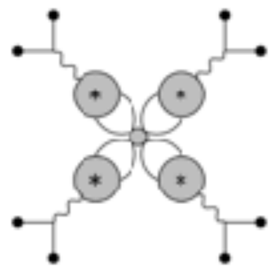
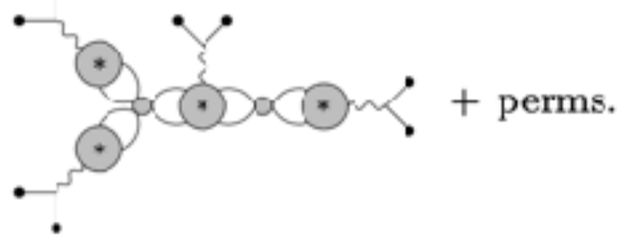
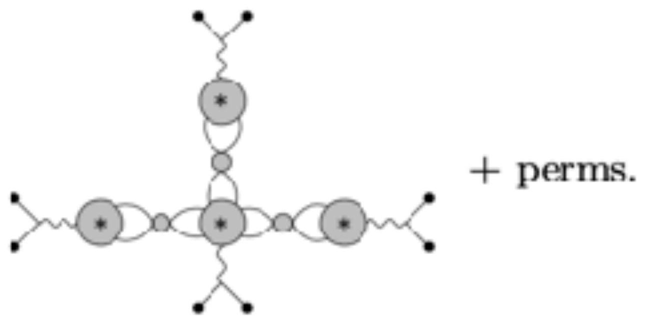
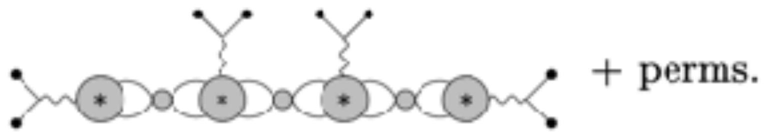


$$+ \sum_{j_1, j_2, j_3}$$



$$+ \text{ perms.}$$

Connected 8-point correlator

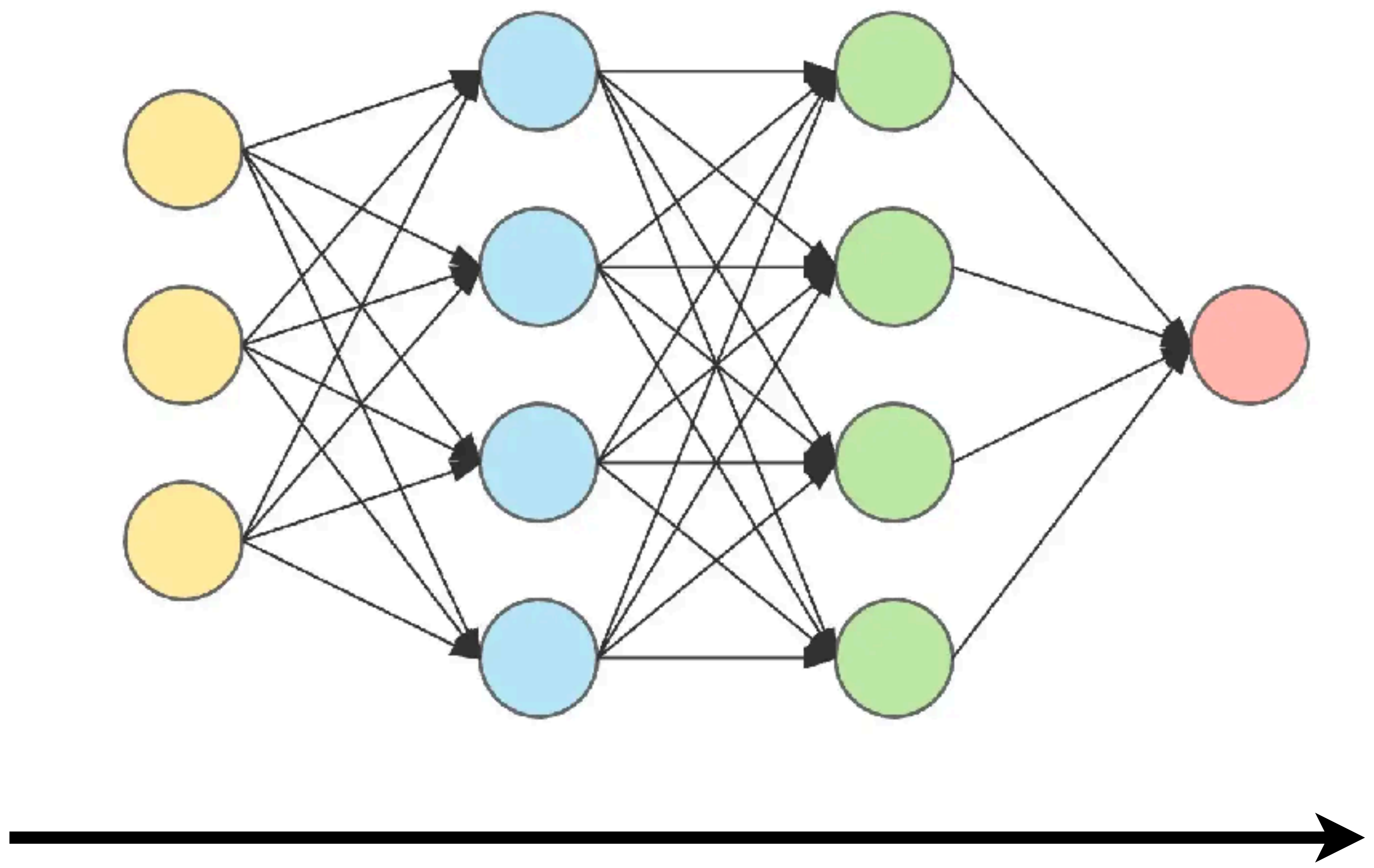
# of j sums	diagrams	degenerate limit result $/\,(C_W^{(\ell)})^4$
1		$\frac{1}{n_{\ell-1}^3}\left(\langle\Delta^4\rangle - 3\langle\Delta^2\rangle^2\right)$
2		$\frac{1}{4}\cdot\frac{V_4^{(\ell-1)}}{n_{\ell-1}^2 n_{\ell-2}}\cdot 4\left(\langle\partial^2\Delta^3\rangle\langle\partial^2\Delta\rangle - 3\langle\partial^2\Delta\rangle^2\langle\Delta^2\rangle\right)$
		$\frac{1}{4}\cdot\frac{V_4^{(\ell-1)}}{n_{\ell-1}^2 n_{\ell-2}}\cdot 3\langle\partial^2\Delta^2\rangle^2$
3		$\frac{1}{8}\cdot\frac{V_6^{(\ell-1)}}{n_{\ell-1} n_{\ell-2}^2}\cdot 6\langle\partial^2\Delta^2\rangle\langle\partial^2\Delta\rangle^2$
		$\frac{1}{16}\cdot\frac{\left(V_4^{(\ell-1)}\right)^2}{n_{\ell-1} n_{\ell-2}^2}\cdot 6\left(\langle\partial^4\Delta^2\rangle\langle\partial^2\Delta\rangle^2 - 2\langle\partial^2\Delta\rangle^4\right)$
		$\frac{1}{16}\cdot\frac{\left(V_4^{(\ell-1)}\right)^2}{n_{\ell-1} n_{\ell-2}^2}\cdot 12\langle\partial^4\Delta\rangle\langle\partial^2\Delta^2\rangle\langle\partial^2\Delta\rangle$
4		$\frac{1}{16}\cdot\frac{V_8^{(\ell-1)}}{n_{\ell-2}^3}\langle\partial^2\Delta\rangle^4$
		$\frac{1}{32}\cdot\frac{V_6^{(\ell-1)}V_4^{(\ell-1)}}{n_{\ell-2}^3}\cdot 12\langle\partial^4\Delta\rangle\langle\partial^2\Delta\rangle^3$
		$\frac{1}{64}\cdot\frac{\left(V_4^{(\ell-1)}\right)^3}{n_{\ell-2}^3}\cdot 4\langle\partial^6\Delta\rangle\langle\partial^2\Delta\rangle^3$
		$\frac{1}{64}\cdot\frac{\left(V_4^{(\ell-1)}\right)^3}{n_{\ell-2}^3}\cdot 12\langle\partial^4\Delta\rangle^2\langle\partial^2\Delta\rangle^2$

Outline

1. Neural networks ↔ field theories (high-level summary).
2. EFT of deep neural networks (at initialization).
3. Diagrammatic approach.
4. Structures of neural network EFTs and criticality.

Criticality

input (UV)



output (IR)

$$\phi_i^{(\ell)}(\vec{x}) = \sum_{j=1}^{n_{\ell-1}} W_{ij}^{(\ell)} \sigma\left(\phi_j^{(\ell-1)}(\vec{x})\right) + b_i^{(\ell)}$$

Exponential behavior is generic ⇒ numerical instability or loss of information.

To avoid this, need to fine-tune network hyperparameters to critical values.

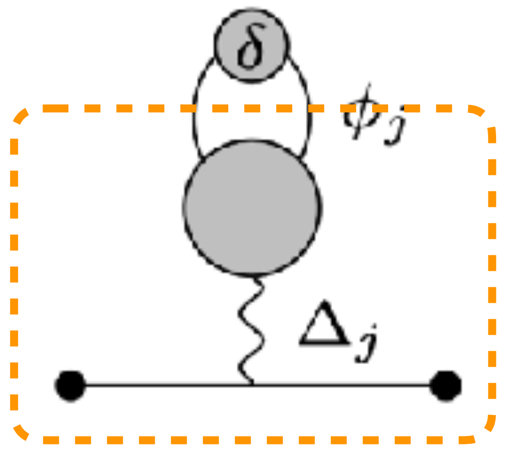
2-point correlator

$$\langle \mathcal{G}^{(\ell\text{-}1)}(\vec{x}_1,\vec{x}_2)\rangle \rightarrow \langle \mathcal{G}^{(\ell\text{-}1)}(\vec{x}_1,\vec{x}_2)\rangle + \delta\langle \mathcal{G}^{(\ell\text{-}1)}(\vec{x}_1,\vec{x}_2)\rangle$$

susceptibility

$$\Rightarrow \quad \delta\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\rangle =$$

$$= \sum_j$$



$$= \int d\vec{y}_1 d\vec{y}_2\, \chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\, \delta\langle \mathcal{G}^{(\ell\text{-}1)}(\vec{y}_1,\vec{y}_2)\rangle$$

$$= \frac{C_W^{(\ell)}}{2}\left\langle \frac{\delta^2\Delta(\vec{x}_1,\vec{x}_2)}{\delta\phi(\vec{y}_1)\delta\phi(\vec{y}_2)}\right\rangle_{\mathcal{K}_0^{(\ell\text{-}1)}} + \mathcal{O}\left(\frac{1}{n}\right)$$

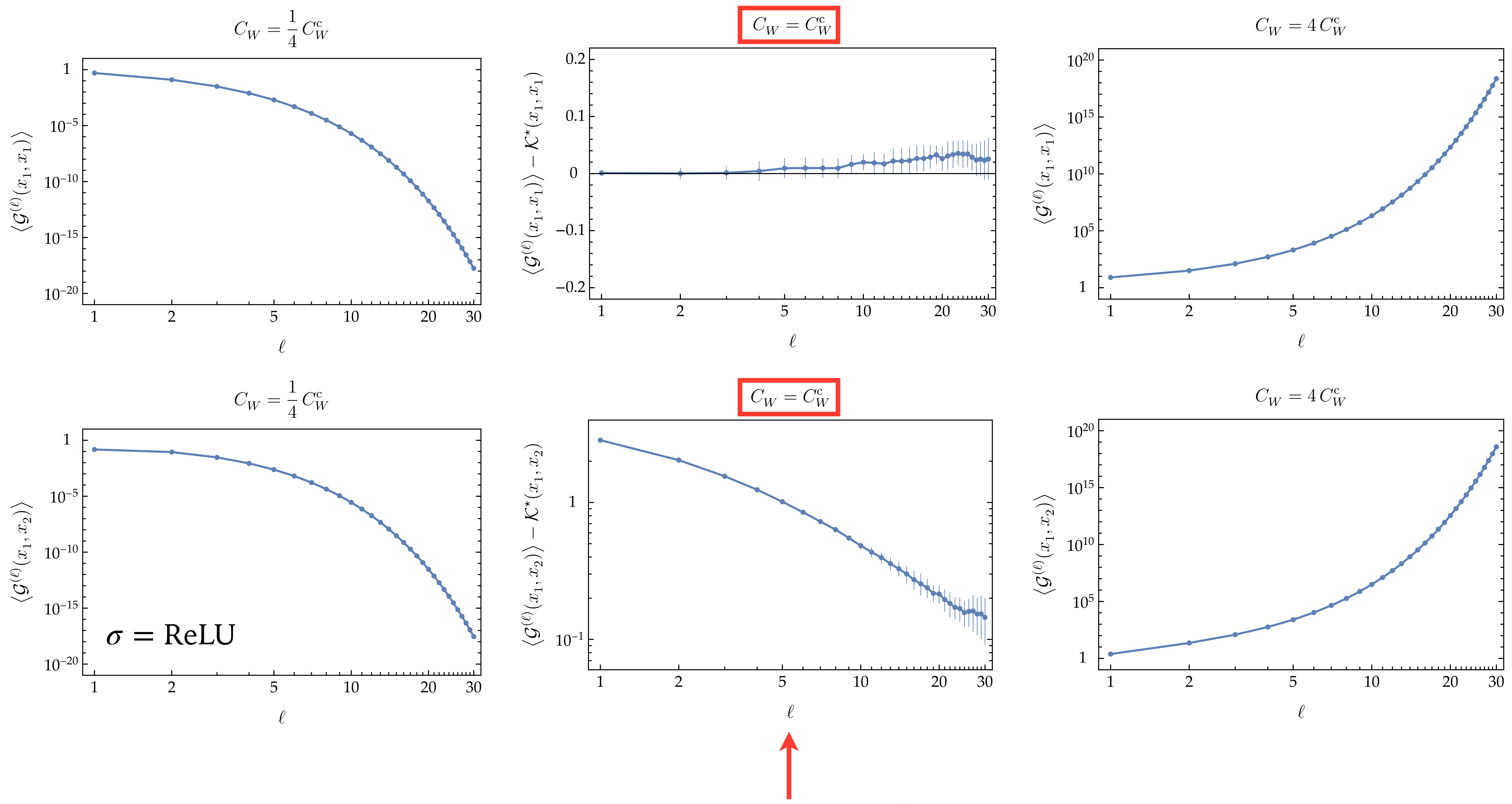
Roughly speaking, $\begin{cases} \chi > 1 \Rightarrow |\langle G^{(\ell)}\rangle - K^\star| \sim e^{\ell} \\ \chi < 1 \Rightarrow |\langle G^{(\ell)}\rangle - K^\star| \sim e^{-\ell} \end{cases}$

RG fixed point

Tune to criticality: $\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\Big|_{\mathcal{K}_0^{(\ell\text{-}1)}=\mathcal{K}^\star} = \frac{1}{2}\Big[\delta(\vec{x}_1-\vec{y}_1)\,\delta(\vec{x}_2-\vec{y}_2)+\delta(\vec{x}_1-\vec{y}_2)\,\delta(\vec{x}_2-\vec{y}_1)\Big]$

$\Rightarrow\ \delta\langle \mathcal{G}^{(\ell)}(\vec{x}_1,\vec{x}_2)\rangle = \delta\langle \mathcal{G}^{(\ell\text{-}1)}(\vec{x}_1,\vec{x}_2)\rangle \Rightarrow$ power-law scaling: $|\langle G^{(\ell)}\rangle - K^\star| \sim \ell^{\gamma}$ ← critical exponent

Hyperparameter tuning



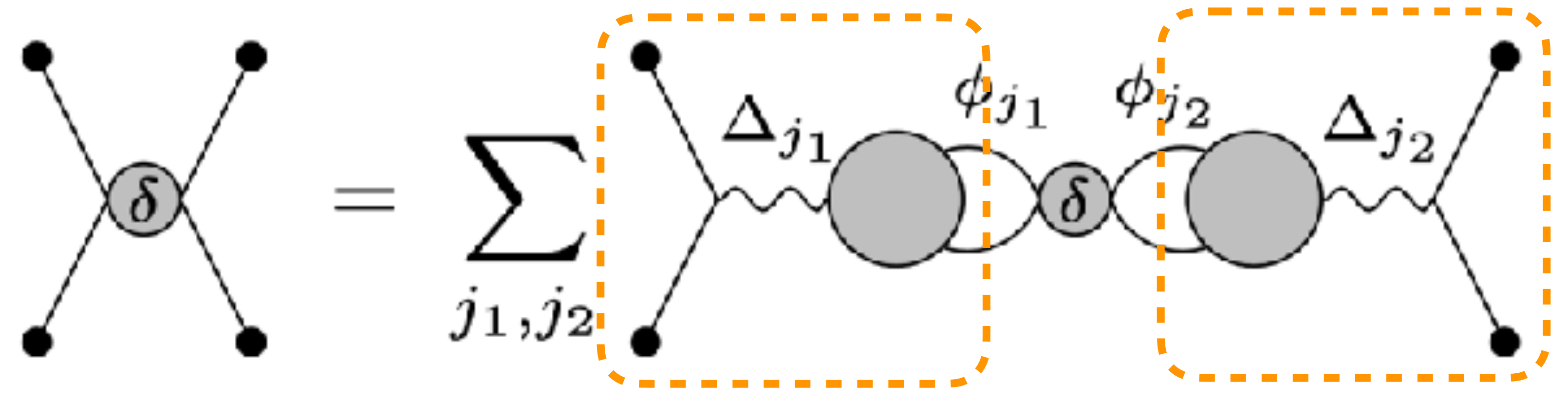
critical point where $\chi^{(\ell)}(\vec{x}_1, \vec{x}_2; \vec{y}_1, \vec{y}_2)\Big|_{\mathcal{K}_0^{(\ell-1)} = \mathcal{K}^\star} = \frac{1}{2}\left[\delta(\vec{x}_1 - \vec{y}_1)\,\delta(\vec{x}_2 - \vec{y}_2) + \delta(\vec{x}_1 - \vec{y}_2)\,\delta(\vec{x}_2 - \vec{y}_1)\right]$

Higher-point connected correlators?

All of them must have power-law scaling.

Naively more constraints than tunable hyperparameters.

However, they have a common structure!



$$\Rightarrow \quad \frac{n_{\ell-2}}{n_{\ell-1}} \frac{\delta V_4^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{x}_3,\vec{x}_4)}{\delta V_4^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\vec{y}_3,\vec{y}_4)} = \frac{1}{2}\left[\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\,\chi^{(\ell)}(\vec{x}_3,\vec{x}_4;\vec{y}_3,\vec{y}_4) + \chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_3,\vec{y}_4)\,\chi^{(\ell)}(\vec{x}_3,\vec{x}_4;\vec{y}_1,\vec{y}_2)\right]$$

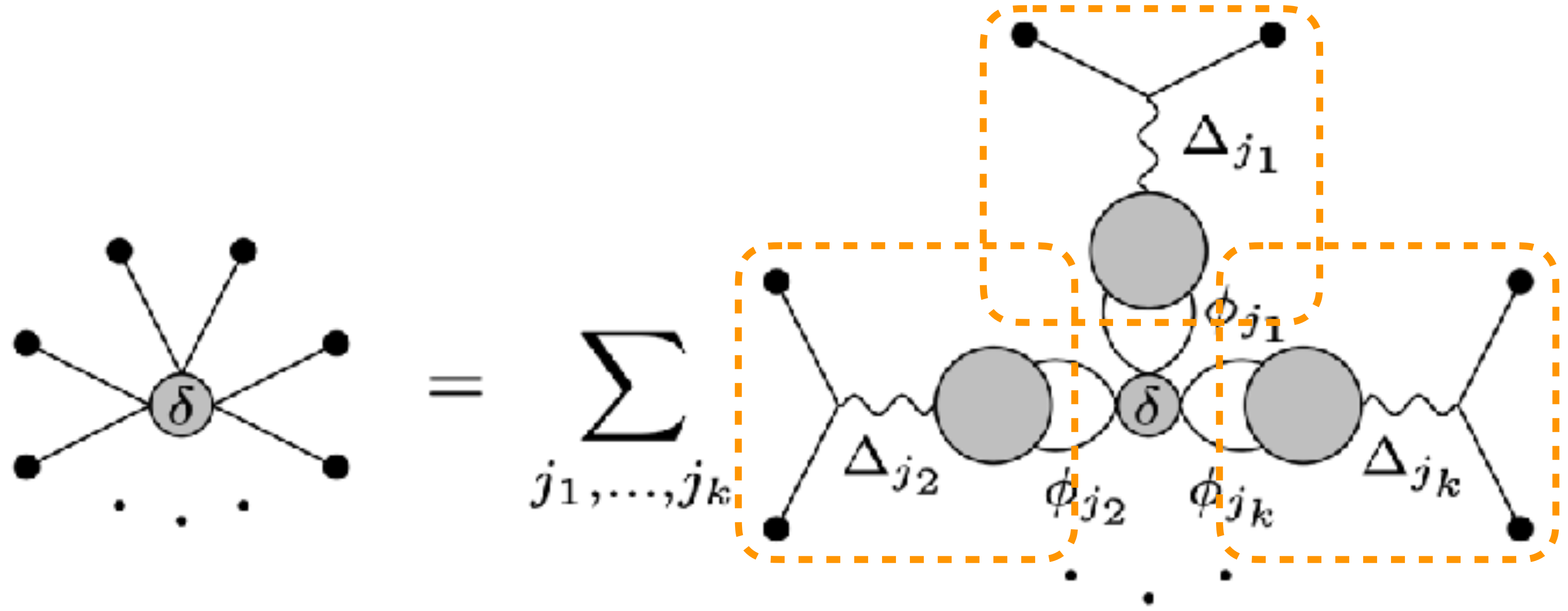
same susceptibility introduced in the 2-point correlator analysis!

Higher-point connected correlators?

All of them must have power-law scaling.

Naively more constraints than tunable hyperparameters.

However, they have a common structure!



$$\Rightarrow \quad \left(\frac{n_{\ell-2}}{n_{\ell-1}}\right)^{k-1} \frac{\delta V_{2k}^{(\ell)}(\vec{x}_1,\vec{x}_2;\ldots;\vec{x}_{2k-1},\vec{x}_{2k})}{\delta V_{2k}^{(\ell-1)}(\vec{y}_1,\vec{y}_2;\ldots;\vec{y}_{2k-1},\vec{y}_{2k})} = \text{sym.}\left[\prod_{k'=1}^{k} \chi^{(\ell)}(\vec{x}_{2k'-1},\vec{x}_{2k'};\vec{y}_{2k'-1},\vec{y}_{2k'})\right]$$

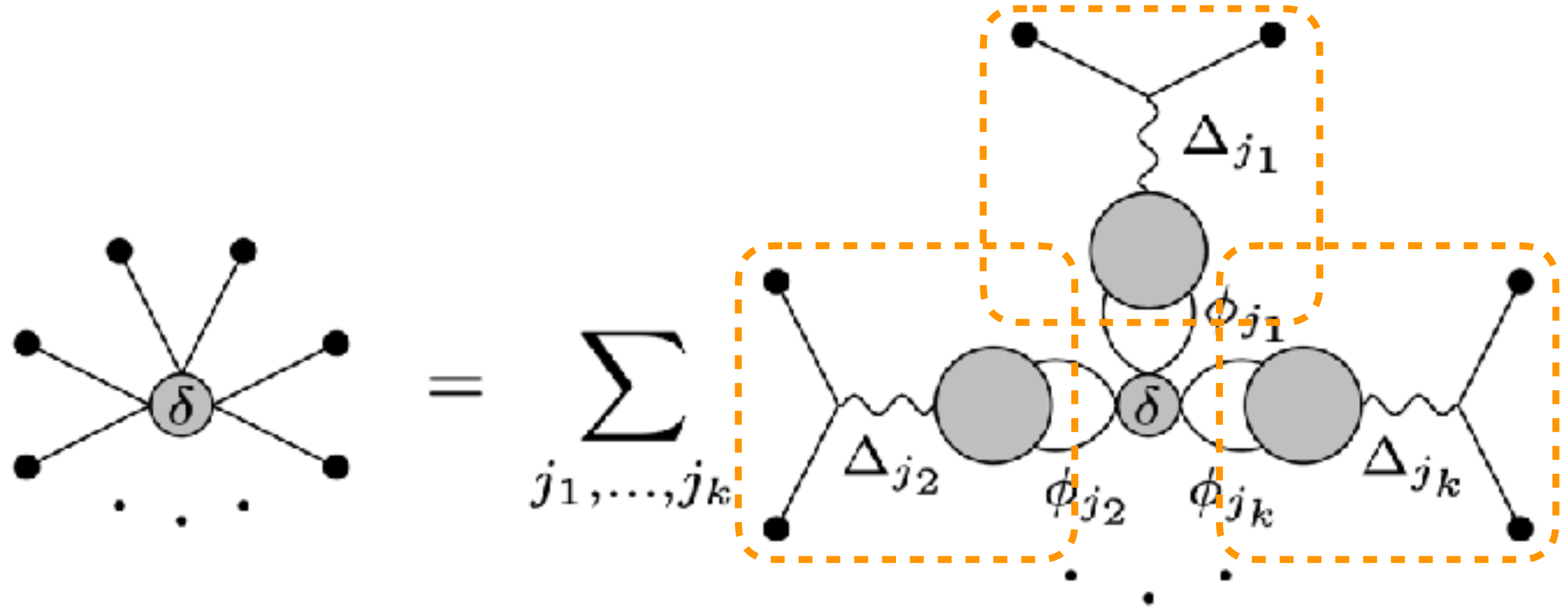
same susceptibility introduced in the 2-point correlator analysis!

Higher-point connected correlators?

All of them must have power-law scaling.

Naively more constraints than tunable hyperparameters.

However, they have a common structure!



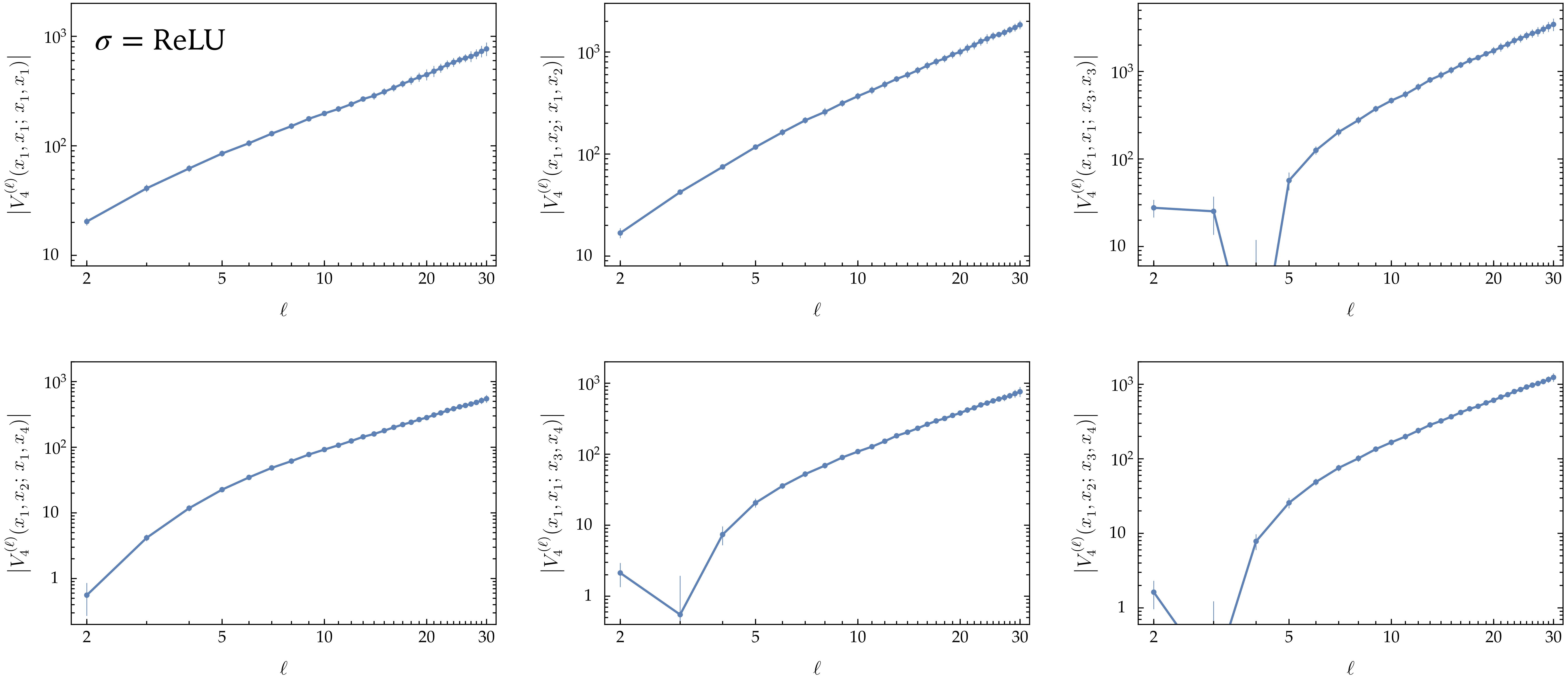
$$\Rightarrow \quad \left(\frac{n_{\ell\text{-}2}}{n_{\ell\text{-}1}}\right)^{k\text{-}1} \frac{\delta V^{(\ell)}_{2k}(\vec{x}_1,\vec{x}_2;\ldots;\vec{x}_{2k\text{-}1},\vec{x}_{2k})}{\delta V^{(\ell-1)}_{2k}(\vec{y}_1,\vec{y}_2;\ldots;\vec{y}_{2k\text{-}1},\vec{y}_{2k})} = \text{sym.}\left[\prod_{k'=1}^{k} \chi^{(\ell)}(\vec{x}_{2k'\text{-}1},\vec{x}_{2k'};\vec{y}_{2k'\text{-}1},\vec{y}_{2k'})\right]$$

Single criticality condition: $\chi^{(\ell)}(\vec{x}_1,\vec{x}_2;\vec{y}_1,\vec{y}_2)\Big|_{\mathcal{K}_0^{(\ell-1)}=\mathcal{K}^\star} = \frac{1}{2}\left[\delta(\vec{x}_1-\vec{y}_1)\,\delta(\vec{x}_2-\vec{y}_2)+\delta(\vec{x}_1-\vec{y}_2)\,\delta(\vec{x}_2-\vec{y}_1)\right]$

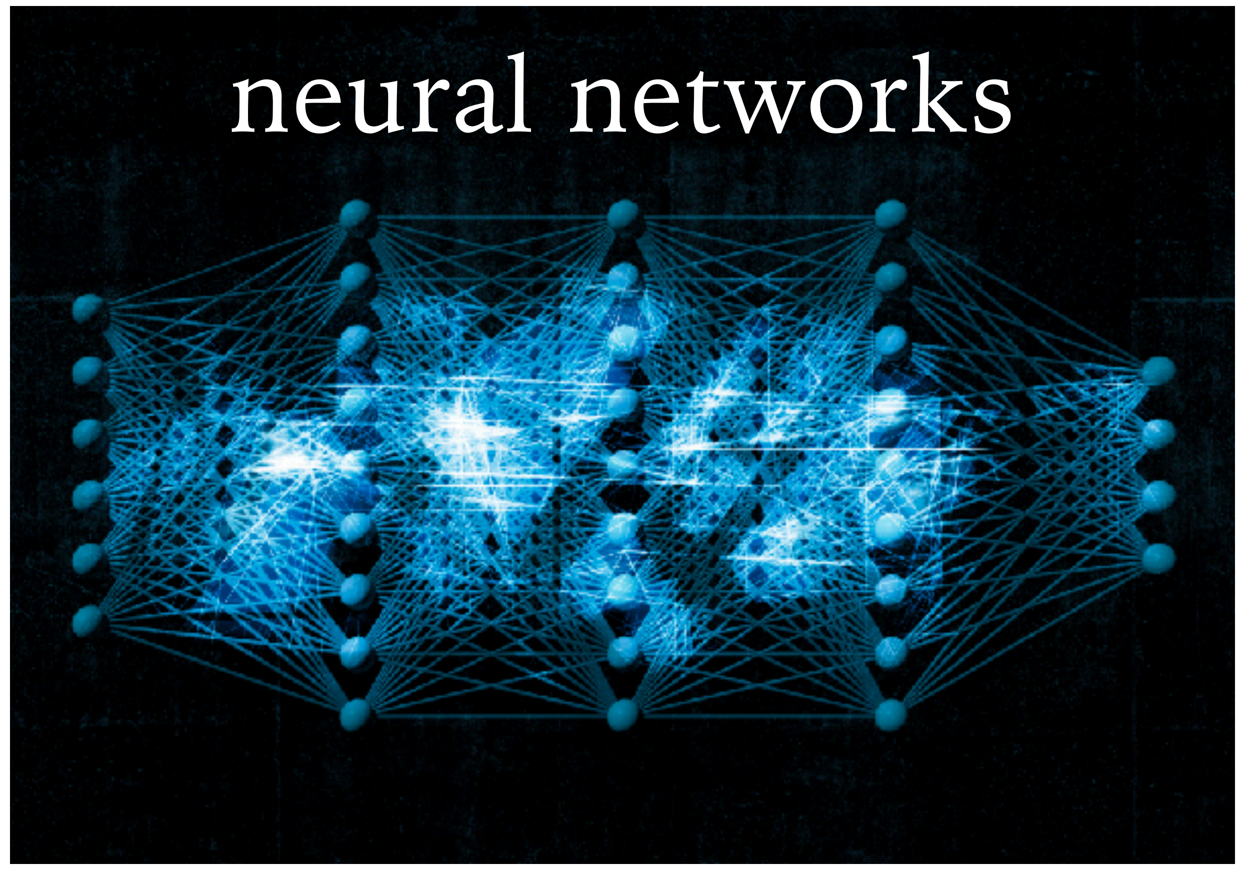
⇒ Power-law scaling for all connected correlators!

Numerical verification

Connected 4-pt correlator at criticality.

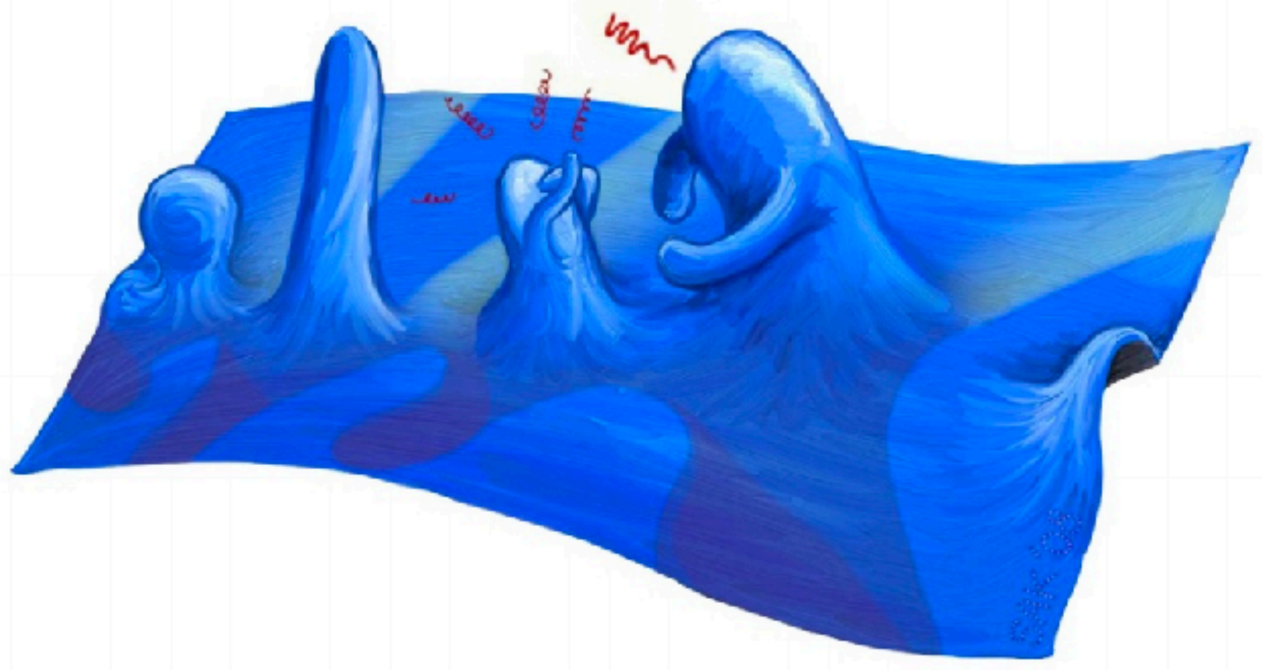


Summary



⟷

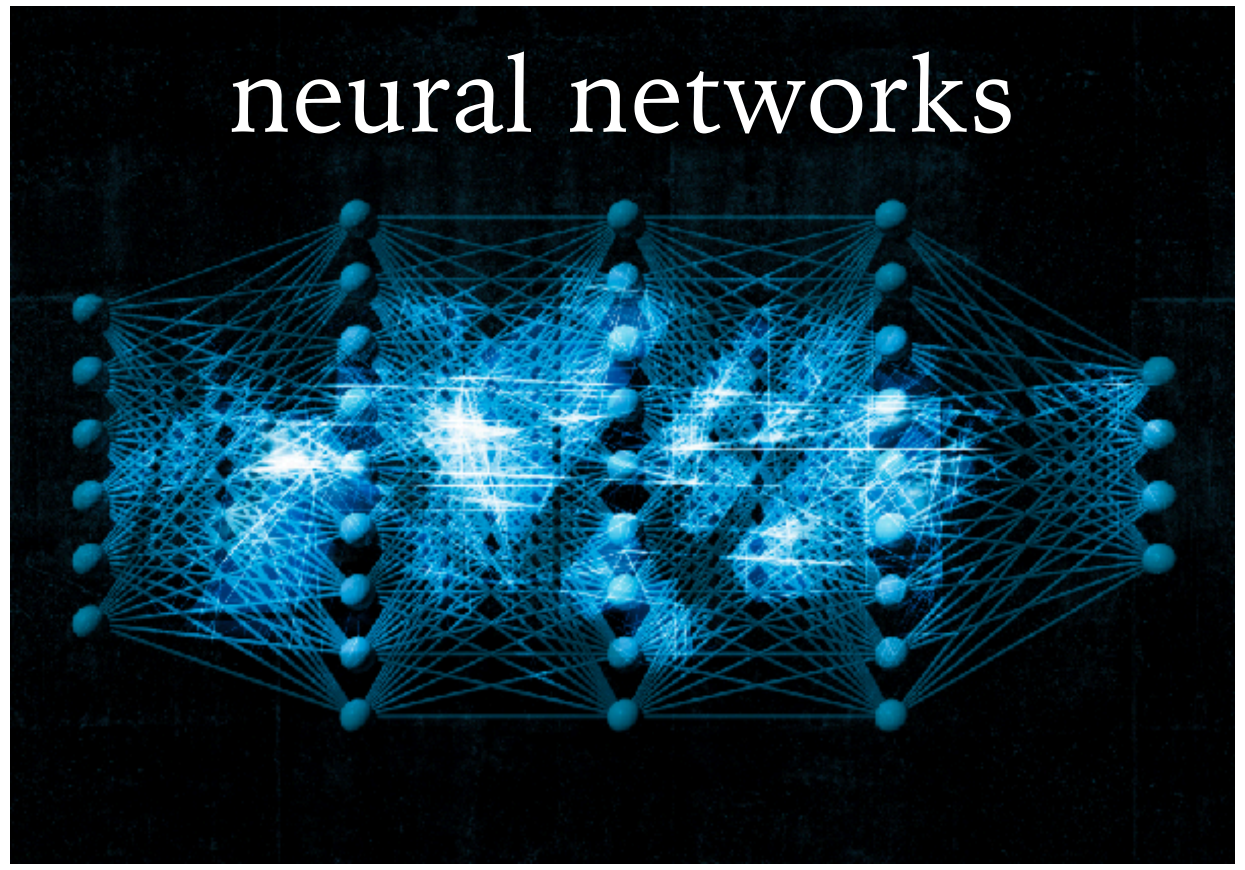
field theories



Diagrammatic approach to EFTs corresponding to neural networks.

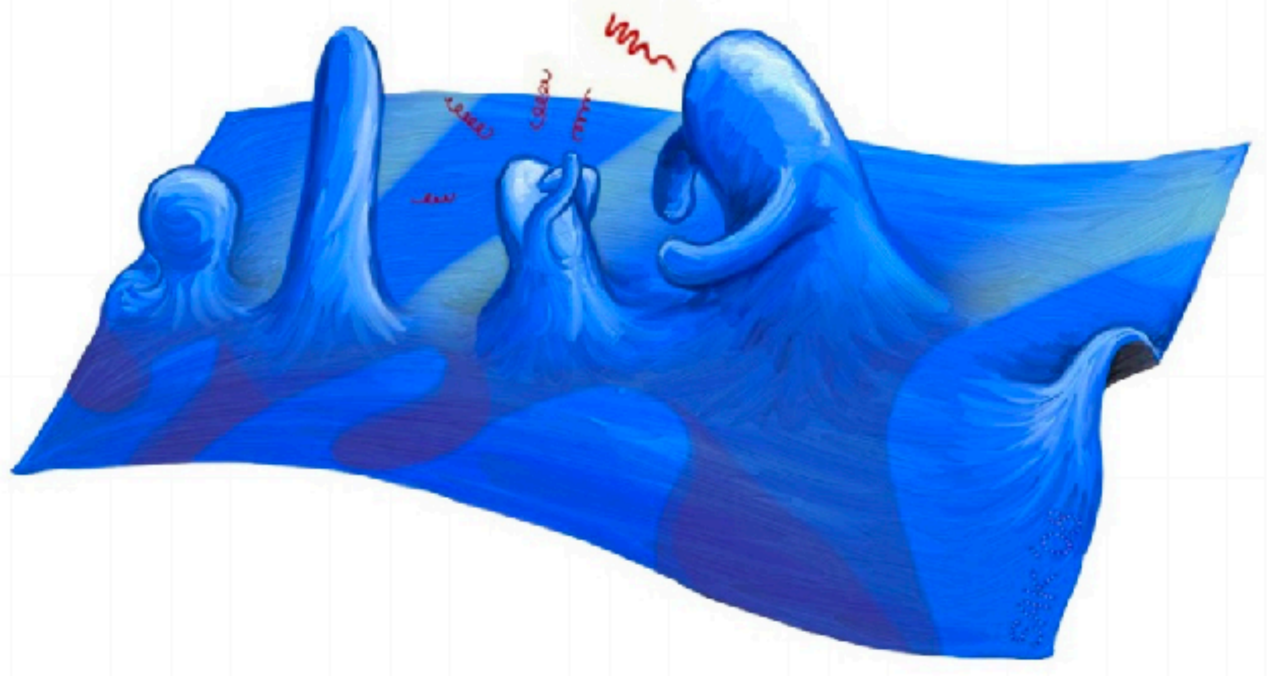
Structures of RG calculation ⇒ successful tuning to criticality.

Summary



⟷

field theories



Diagrammatic approach to EFTs corresponding to neural networks.

Structures of RG calculation ⇒ successful tuning to criticality.