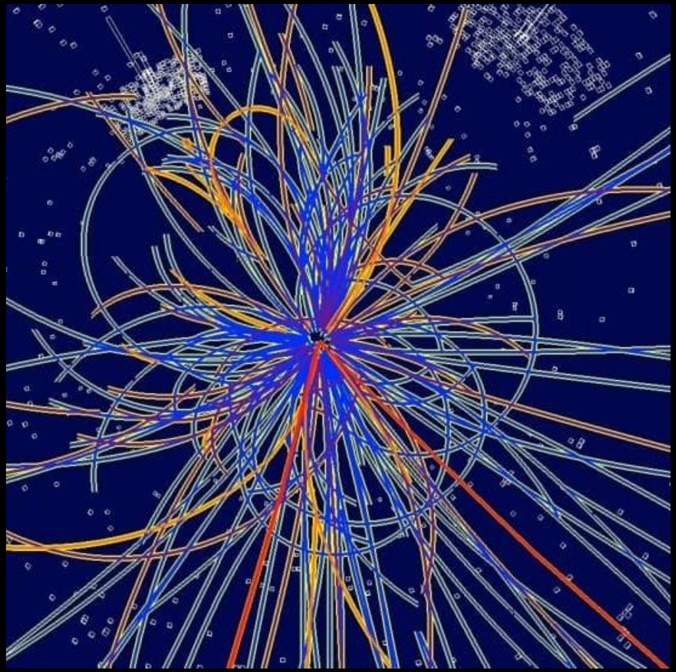
Quantum Field Theory in the Age of Data

Miranda C. N. Cheng

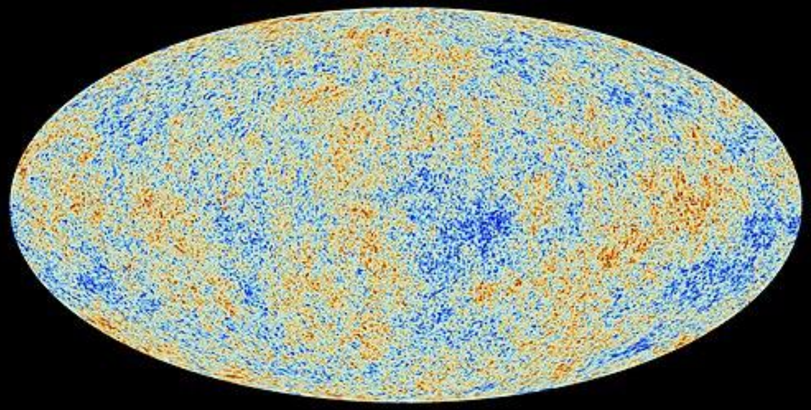
University of Amsterdam · Academia Sinica Taiwan

PASCOS 2026 · Sheffield · 22–26 June 2026

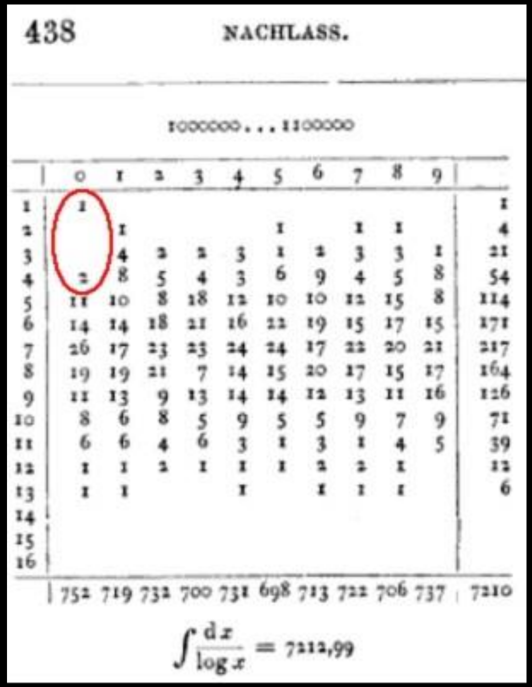
Data has long helped drive physical sciences.



The LHC
particle-collision data informs the most fundamental physics



Cosmic Microwave Background
thermal radiation data informs the origin of universe

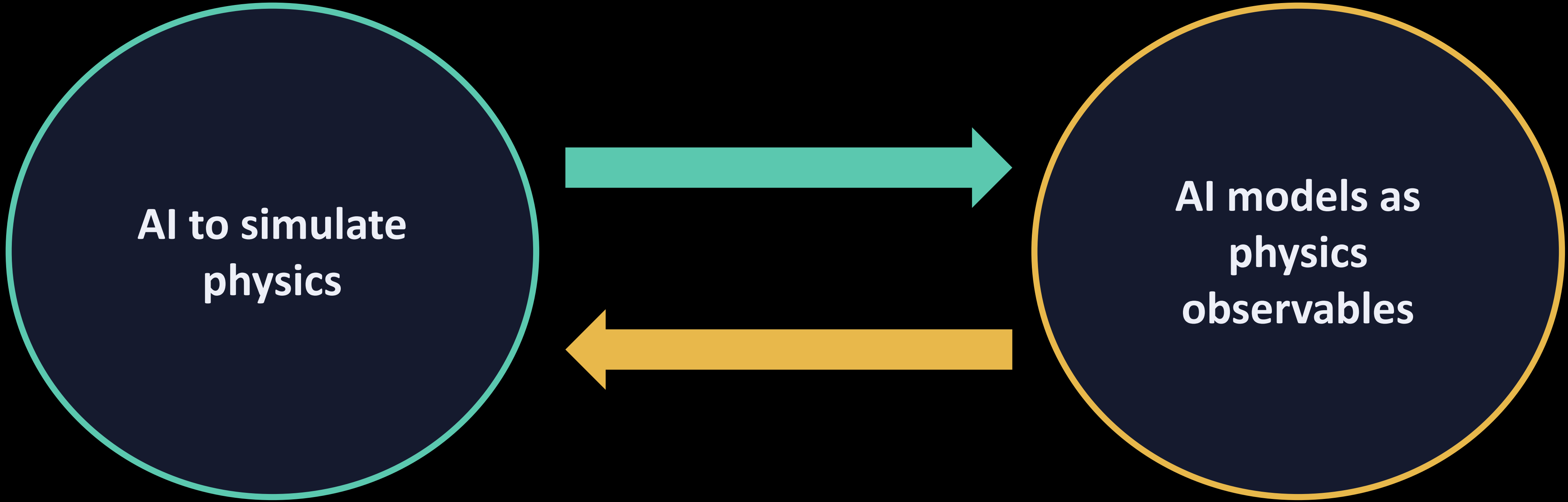


Gauss' Count of Prime Numbers
the birth of analytic number theory

Today, we discuss some new **AI-enabled** ways of treating and learning from scientific data.

AI Models as Data:

Use AI to simulate physics — and study the trained models as new physical observables.



How do we study a quantum field theory?

A quantum field theory

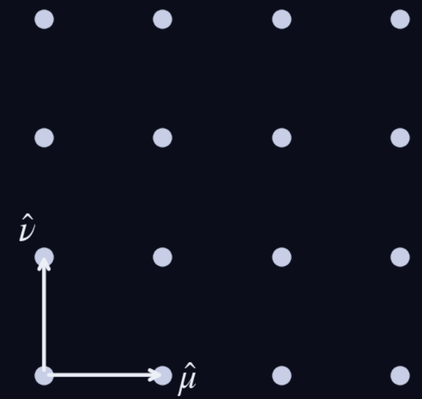
- is a pillar of modern physics
- defined by the data: the fields ϕ, a (Euclidean) action $S_g[\phi]$, the coupling constants g = (m², $g_2, g_4, \ldots$.)
- but **notoriously ill-defined** — the space of fields is infinite-dimensional, $\varphi \propto e^{-S_g(\varphi)}$

$$S_g[\varphi] = \sum_x \left[\varphi(-\Delta + m^2)\varphi + \lambda\varphi^4\right]$$

$g = (m^2, \lambda)$: two couplings

A lattice field theory

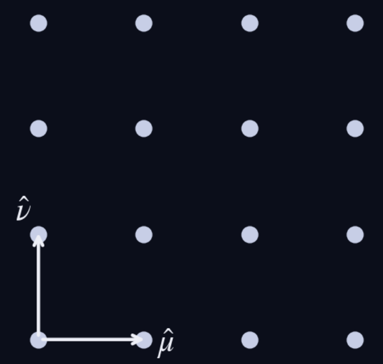
- the main first-principle method for non-perturbative QFT (e.g. QCD)



Lattice Field Theories

Workhorse for non-perturbative physics: from the strong force to phases in materials.

	QFT (continuum)	**lattice**
spacetime	$\int d^D x$	$a^D \sum_{x \in (\mathbb{Z}/L\mathbb{Z})^D}$
the field	$\varphi : \mathbb{R}^D \to \mathbb{R},\ x \mapsto \varphi(x)$	$\varphi : (\mathbb{Z}/L\mathbb{Z})^D \to \mathbb{R},\ x \mapsto \varphi_x$
observable	$\langle \varphi(x_1) \cdots \varphi(x_n) \rangle$	$\frac{1}{N} \sum_{i=1}^{N} \varphi_{x_1}^{(i)} \cdots \varphi_{x_n}^{(i)}$



↑ ***N samples (of field configurations)***

$$\varphi \sim e^{-S_g(\varphi)}/Z$$

The lattice field theory workflow

from couplings to physics, in four steps

The action

$$S_g[\varphi]$$

→

Samples

$$\varphi_x^{(i)}$$

→

Observables

$$\langle \mathcal{O} \rangle$$

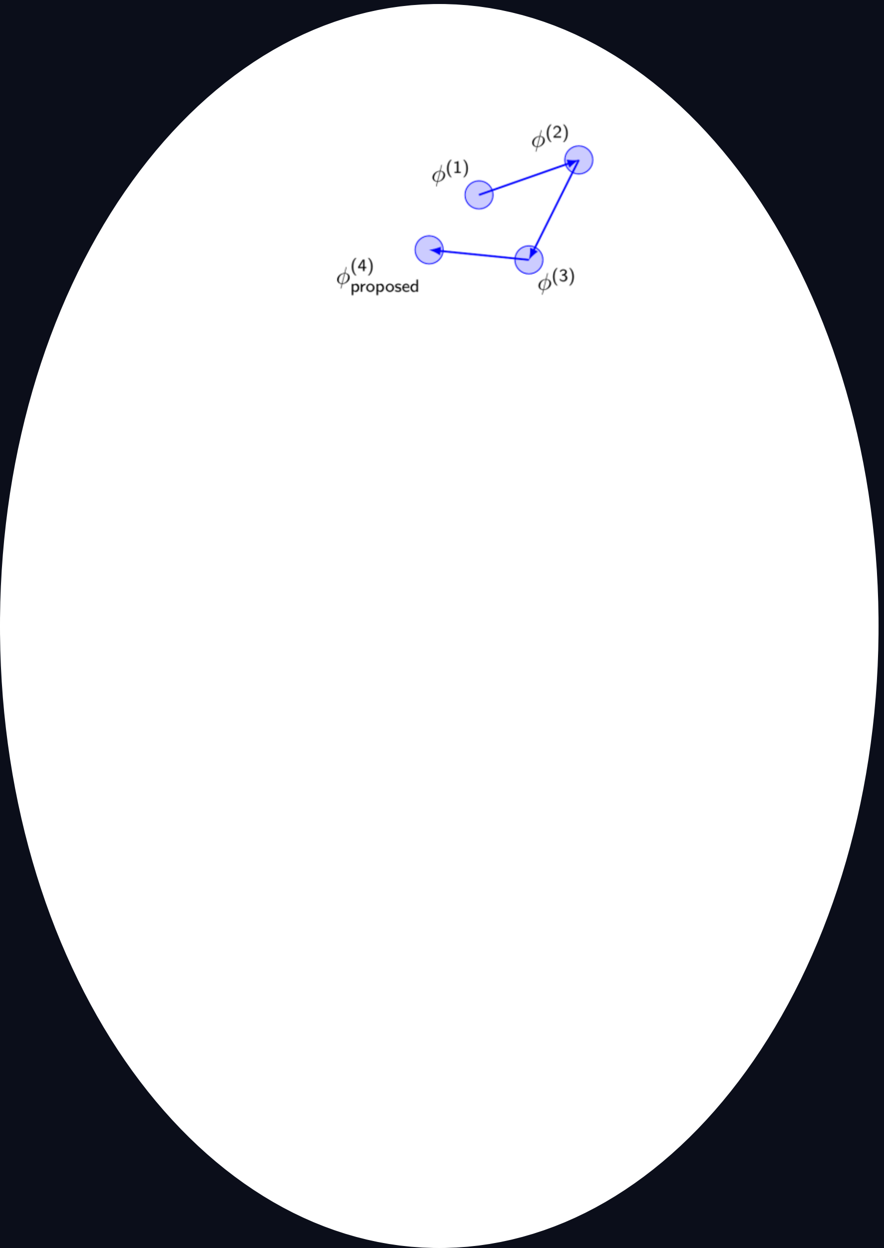
→

Physics

e.g. the phase diagram

Monte-Carlo sampling of field configurations turns the action into measurable physics.

Challenges: Critical Slowing Down



Slow exploration of the high-dimensional configuration space, particularly when using local updates, is a problem.

Interesting regimes are especially challenging:

- *continuous limit: physical size* $L \times a$ *fixed, lattice spacing* $a \to 0$.

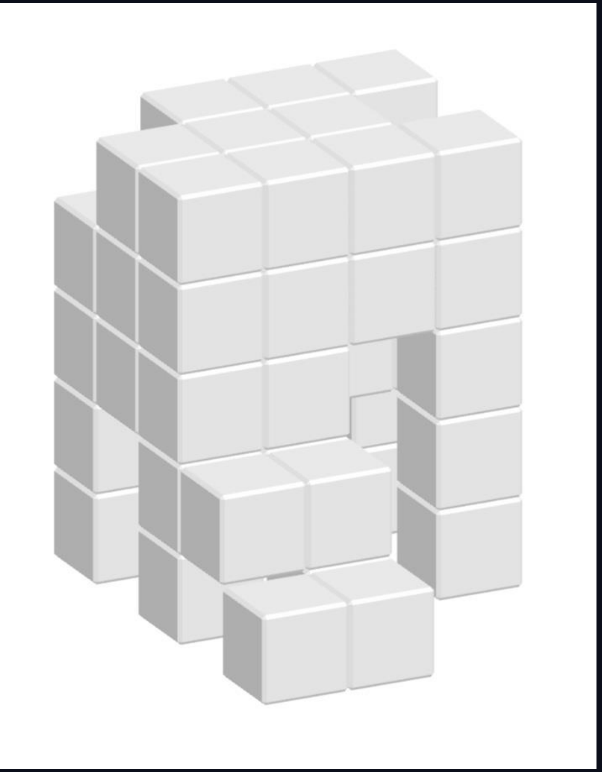
- *phase transition limit: long-range correlation* $\frac{\xi}{a} \to \infty$

- especially bad if there are *topological sectors*

Generative AI

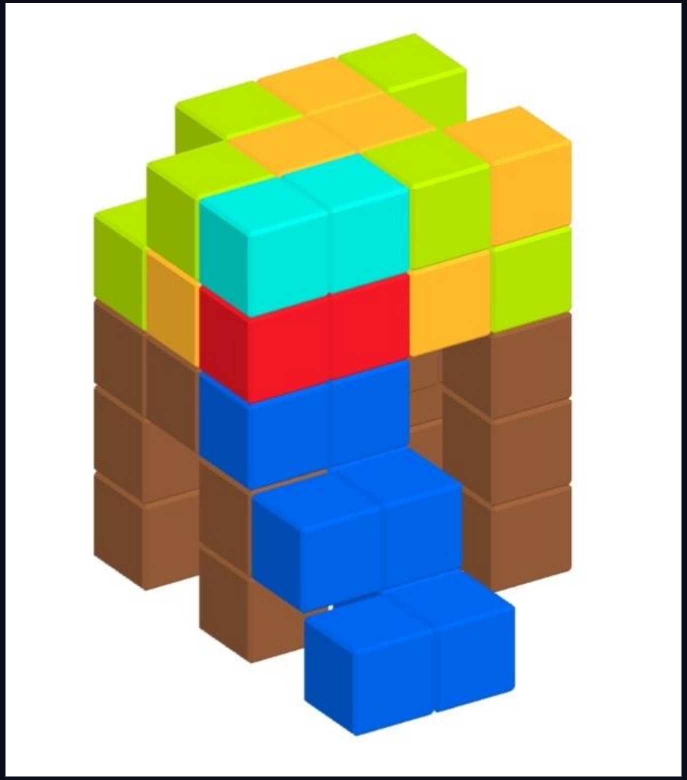
$f_{\theta^{ini}}$

$f_{\theta^*}: input \mapsto output$

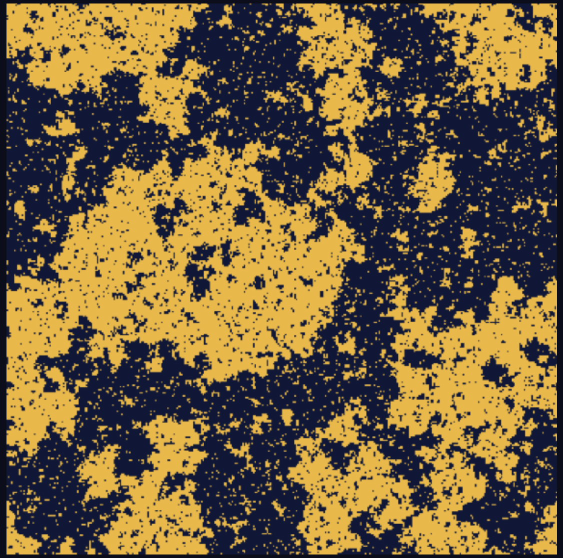


train

$\theta^{ini} \to \theta^*$



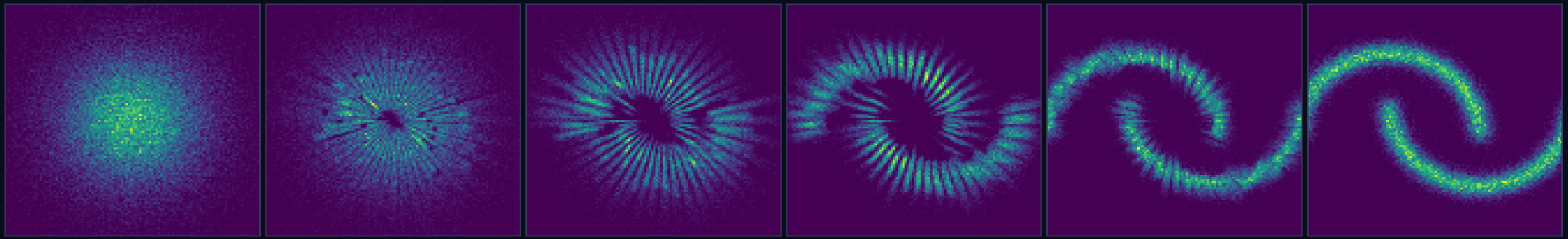
generate



Ising model configurations

The 30-second "flow" (generative) model

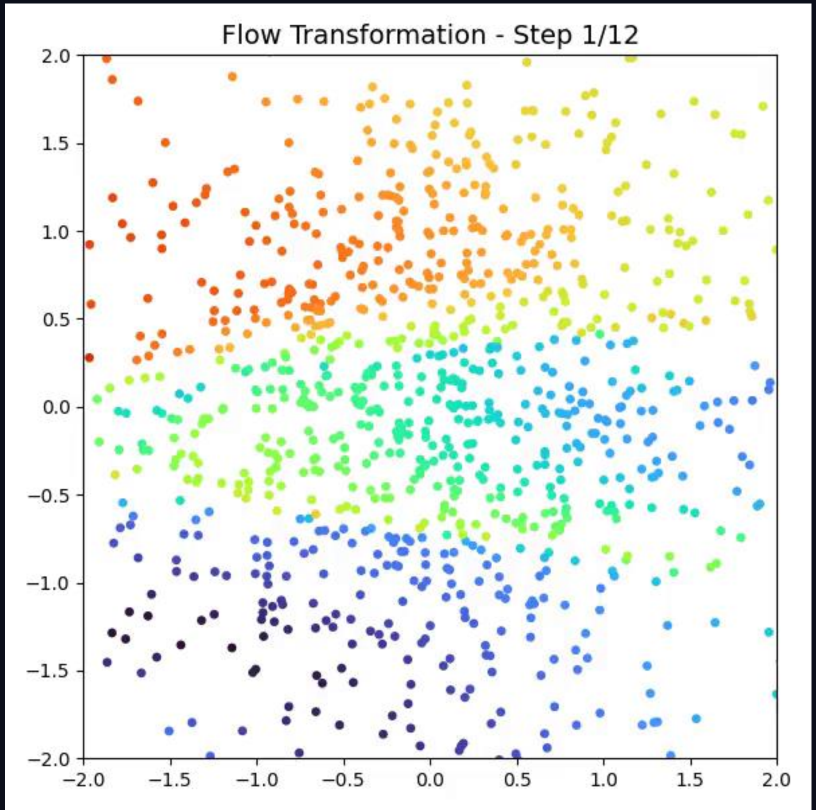
$$\text{simple noise} \longrightarrow \text{invertible map } f_\theta \longrightarrow p_g(\varphi) \propto e^{-S_g[\varphi]}$$



- The same machinery behind image generation.
- θ: network parameters parametrizing the map f_θ.
- can be trained against a **Boltzmann distribution** $p_g(\varphi) \sim e^{-S_g[\varphi]}$ (no data needed)
- Independent samples, sidestepping critical slowing down of MCMC. fields (cf. "trivializing flows" by Luscher 2009)

Normalizing flows, illustrated

an invertible map carries simple noise to the target, layer by layer



[M Gerdes, **MC**, *Analytic Bijections for Smooth and Interpretable Normalizing Flows*, ICML 2026]

A fast-moving field: generative samplers for the lattice

Coupling / gauge-equivariant flows — RealNVP for ϕ^4; U(1) & SU(N) gauge [Albergo–Kanwar–Shanahan ’19; Kanwar+ ’20; Boyda+ ’21, …]

Continuous & trivializing flows — neural-ODE flows; trivializing maps [Lüscher ’09; **de Haan+ ’21**; Del Debbio+ ’21; **Gerdes+ ’23,…**]

Stochastic normalizing flows — flows + non-equilibrium MCMC [Caselle+ ’22, ….]

Diffusion models — diffusion ≈ stochastic quantization; transferable across couplings [Wang–Aarts+ ’24,…]

………

Reviews: *[Cranmer+ ’23, Nat. Rev. Phys.; **Cheng–Stratikopoulou ’26 (lecture notes)**]*

Flows do more than sample: variance reduction of observables

[Kanwar+ ’26; Bacchio ’23, ….]

Then we asked: what did the network learn?

- So far: Choose couplings g, train the flow model; each trained network is a **point ($\boldsymbol{\theta}_g$) in weight space.**

- The questions: can we **learn this map**?

$$\text{coupling} \;\; g \;\; \longrightarrow \;\; \text{flow weights} \;\; \theta_g$$

Do the network weights **($\boldsymbol{\theta}_g$) know the physics**?

- Method: treat $\{(\boldsymbol{g}, \boldsymbol{\theta}_g)\}$ as data for training new models

JEPAWG (JEPA-based Weight Generator)

— treating trained networks as data

(Göbel, Ebelt, Mensch, Gerdes, MC, 2026 ICML AI4Phys Workshop)

The hypernetwork: a network that outputs networks

g
two numbers:
(m^2, λ)

HYPERNETWORK
a neural network
trained once

$\boldsymbol{\theta_g} \in \mathbb{R}^{\sim\mathbf{100{,}000}}$
network parameters

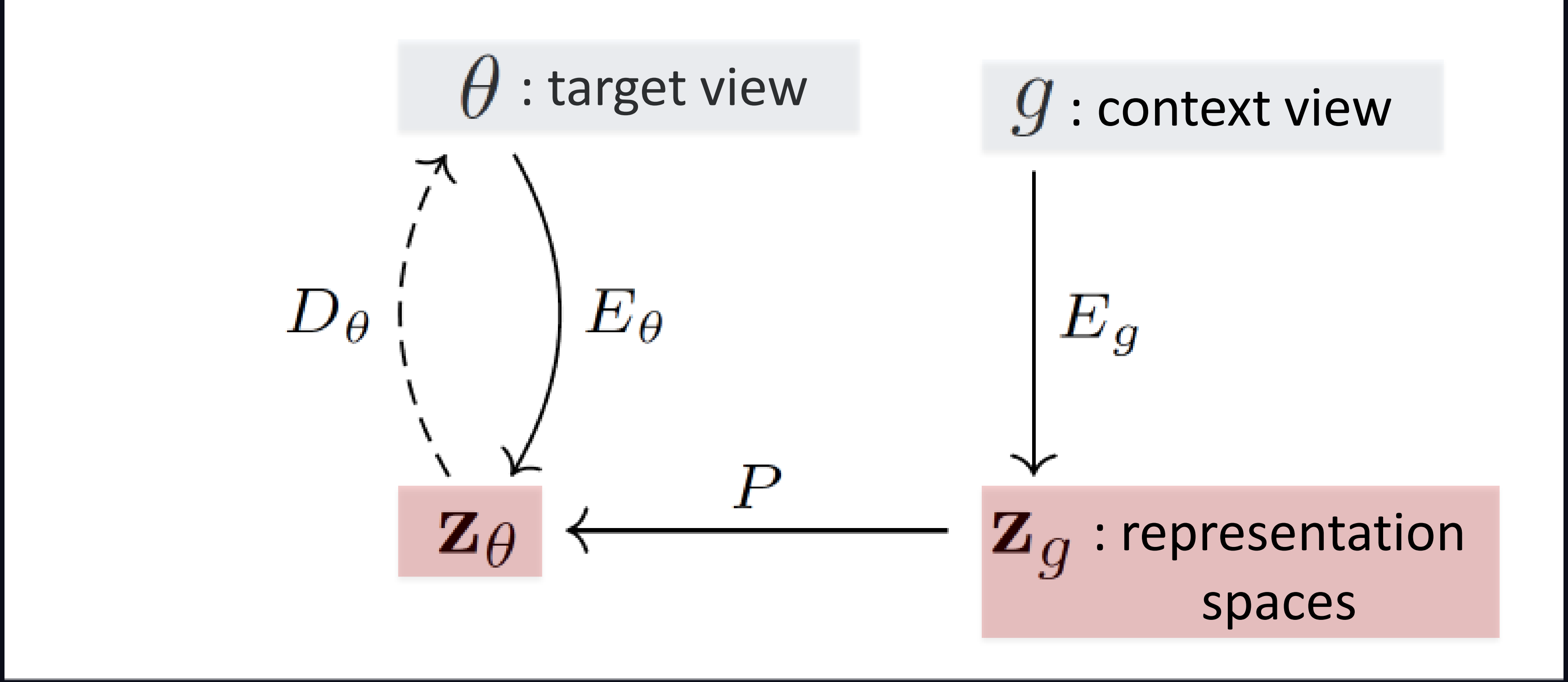
a flow f_{θ_g} to sample from
$p_g(\varphi) \propto e^{-S_g(\varphi)}$

- The hypernetwork learns the whole map **g ↦ θ(g)** — a network whose output layer is another network's weight vector.
- Physics analogy: learning the map from couplings to an observable— but a **hyper-observable** allowing to simulate the theory.
- Before: one coupling g → one training run → one flow θ_g. Starts anew for each g, no knowledge between couplings learned or retained.
- w. hypernet: exploit the fact that the physics, hence θ(g) , to vary **smoothly** with g **(except at criticality)**.

*Next: make this robust and interpretable with a **representation space** in the middle.*

JEPA: Joint-Embedding Predictive Architecture

predict in representation space, not input space (LeCun 2022)



- Encode two "views" into two learned "representation" spaces: $z_g = E_g(g), z_\theta = E_\theta(\theta)$.
- A predictor **P** maps $z \mapsto z'$, trained so $P(z_g) \approx z_{\theta_g}$ — entirely in representation space.
- Learns abstract, semantic features; ignores irrelevant detail.

JEPAWG: a hypernetwork with a feature window

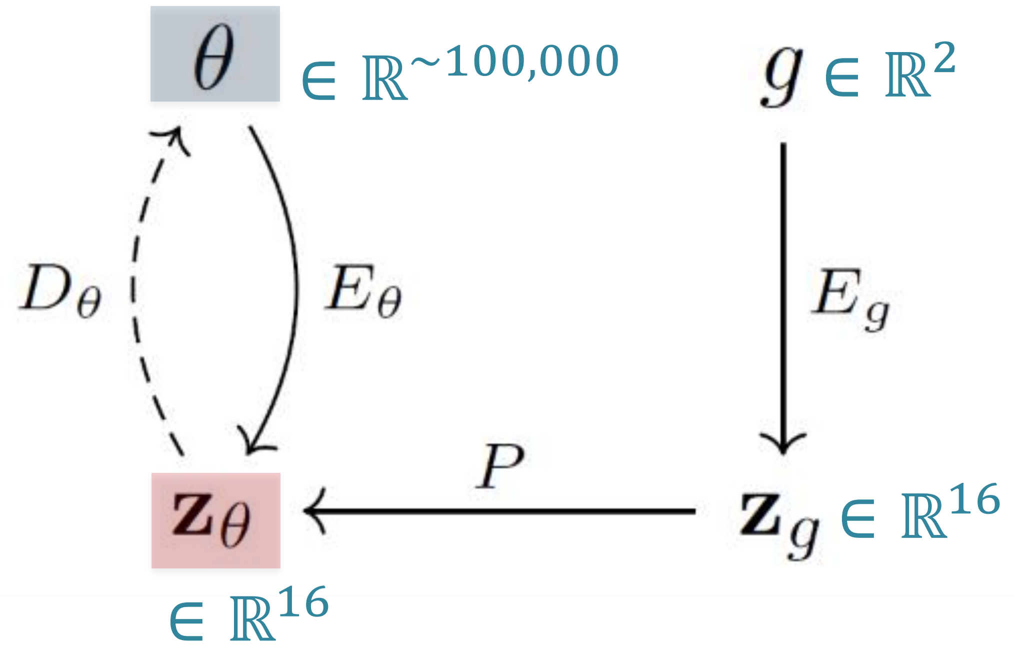
JEPA+ ***Decoder D_θ***

$$\begin{tikzcd}
\theta \in \mathbb{R}^{\sim 100{,}000} \arrow[d, bend left, "E_\theta"] & g \in \mathbb{R}^2 \arrow[d, "E_g"] \\
\mathbf{z}_\theta \in \mathbb{R}^{16} \arrow[u, bend left, dashed, "D_\theta"] & \mathbf{z}_g \in \mathbb{R}^{16} \arrow[l, "P"']
\end{tikzcd}$$

$$\mathcal{L}_{\text{pred}} = \frac{1}{N}\sum_i \|P(E_g(g_i)) - E_\theta(\theta_i)\|^2 + (\text{regularization term})$$

$$\mathcal{L}_{\text{dec}} = \frac{1}{N}\sum_i \|D_\theta(P(E_g(g_i))) - \theta_i\|^2$$

JEPAWG: how well does it work?



Quality: Are the generated weights fitting the reality (here: physics) ? ✓😊

Interpretability: Does the feature space capture the relevant structure (here: physics) ?

Interpretability : The feature space captures the physics

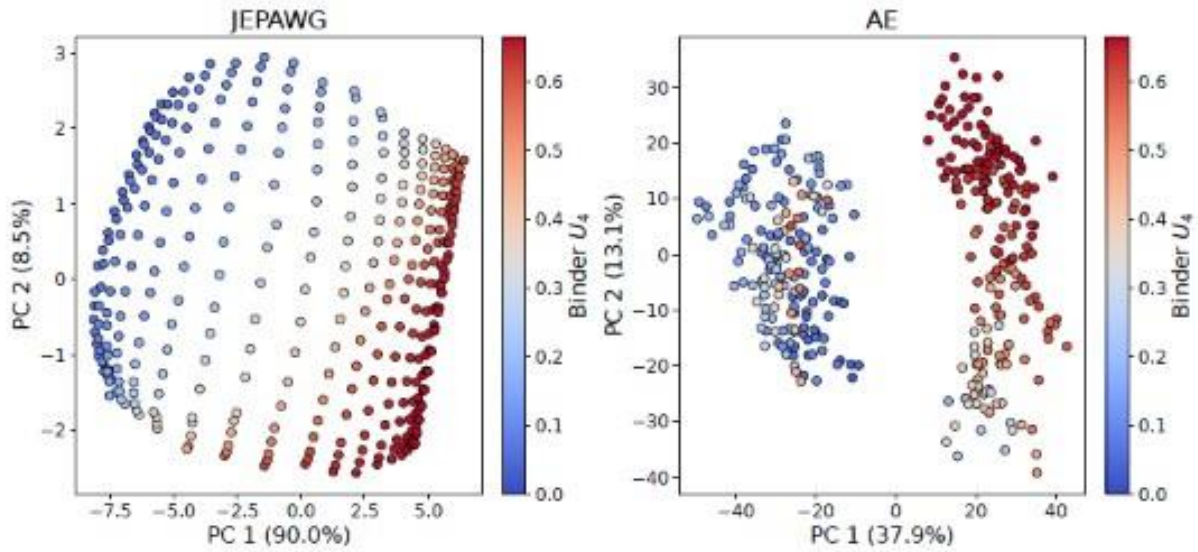
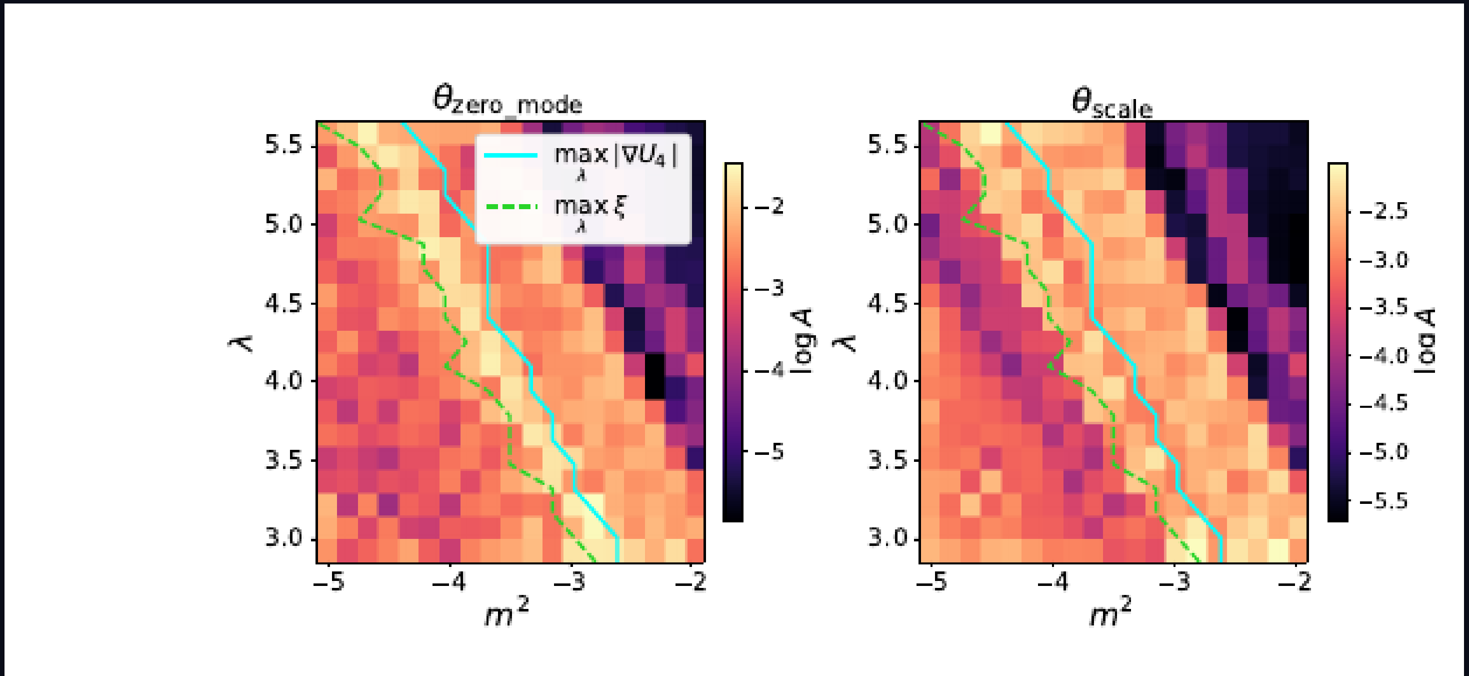


Table 2. Intrinsic dimension estimates of z_0 at encoder width H=128 and latent dimension d_z=16. Top-5 PCA denoising; mean ± standard deviation across 5 seeds.

	MLE (K=10)	PCA-95%
$\dim((m^2, \lambda))$	2.00	2
JEPAWG	**2.01±0.06**	**2.0±0.0**
JEPAWG*	2.36±0.04	13.2±0.8
VAE	4.88±0.15	14.8±0.4
AE	3.68±0.09	11.6±0.5
PCA	2.34±0.01	15

- shown in picture: images of a uniform grid of g
- Intrinsic dimension **= 2 = #couplings**, discovered blind
- AE baseline: learns the 2 origins of data, not physics
- JEPAWG: feature axis aligned with the **phase structure**

The critical line, from weights alone

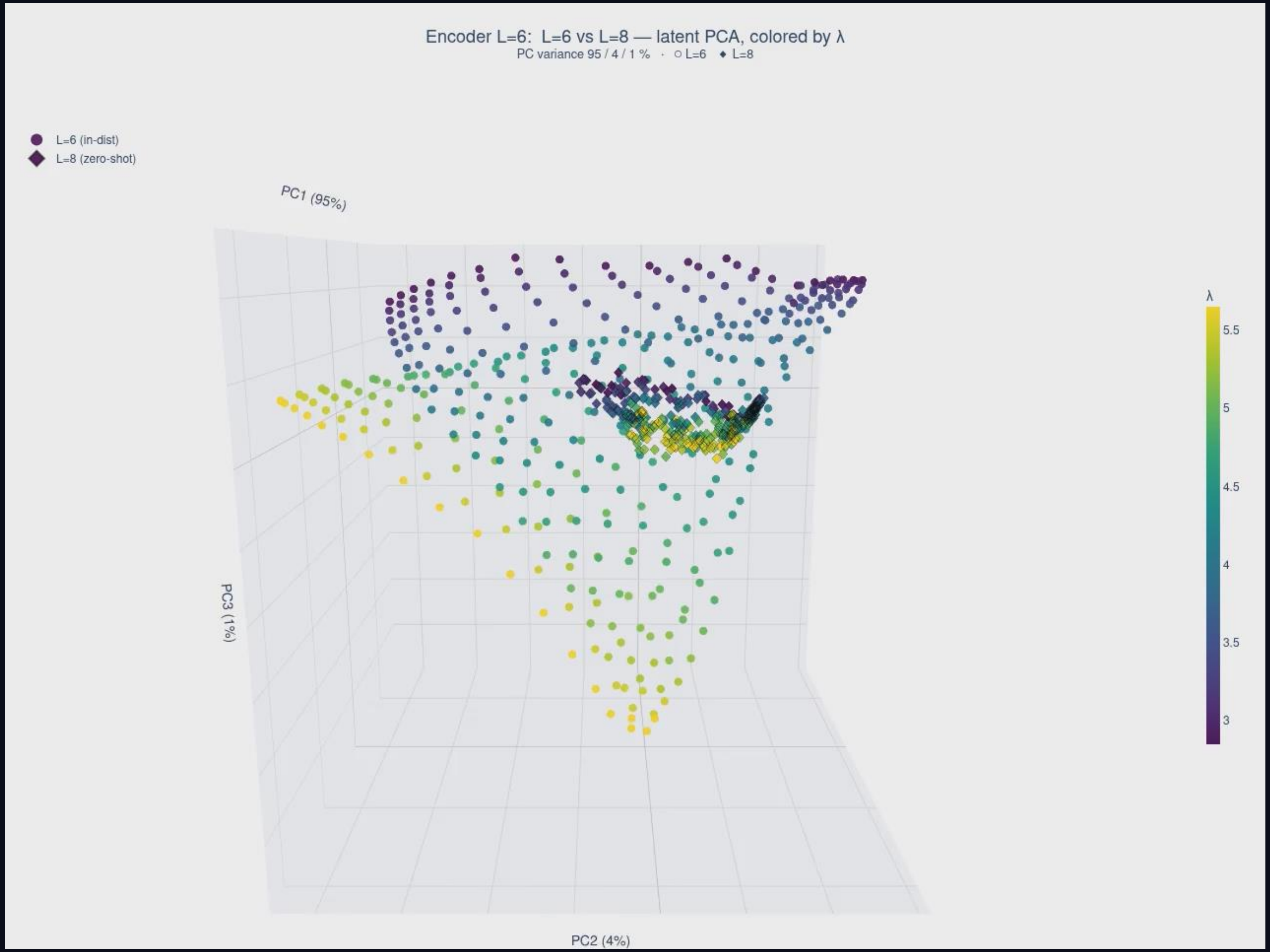


sensitivity of weight components to couplings; overlaid: standard critical-line proxies

- Sensitivity spikes **along the critical line**
- Matches max $|\nabla U_4|$ and ξ proxies
- No correlators computed — weights only

Finite-size scaling in weight space: ν ≈ 1

*Idea: look at networks for lattice sizes L and L + δL, and ask **how to pair points such that they are closest in feature space.***



Finite-size scaling in weight space: ν ≈ 1

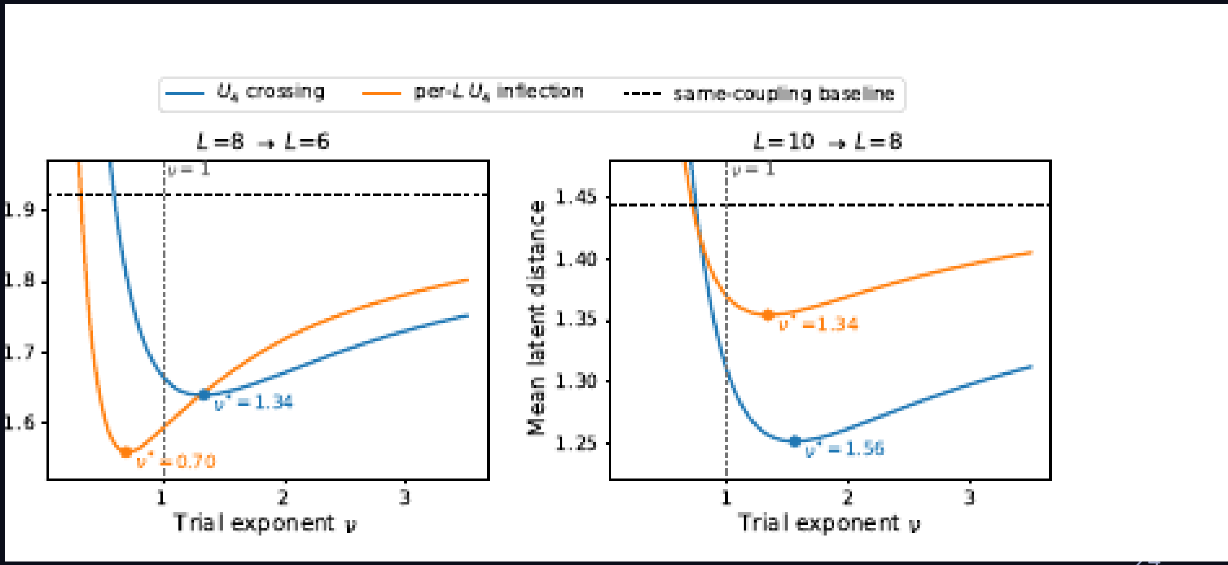
FSS — the workhorse of numerical criticality — reveals itself in network parameters.

$$t = a(\lambda)\,\big(m^2 - m_c^2(\lambda)\big)$$

$$x = t\,L^{1/\nu}$$

$$m^2_{\text{pred}} = m_c^2(\lambda; L_1) + \big(m_2^2 - m_c^2(\lambda; L_2)\big)\left(\frac{L_2}{L_1}\right)^{1/\nu}$$

- Pair couplings across L = 6, 8, 10 by the FSS map with trial exponent ν̂
- Feature distance minimized **near ν = 1** — the 2D Ising value
- **Weights as a brand *new type of (hyper)-observable***



A data-driven new paradigm for studying unknown physics

- pre-NN: 1. run MCMC chains (for a long time). Each chain likely to be stuck in a subsector. Collect all samples to compute observables. 2. Repeat that for each set of coupling constants g. Compare observables to see the phase diagram.

- with NN [2021----]: for each g, tain the neural network once. The samples it generates are all independent.

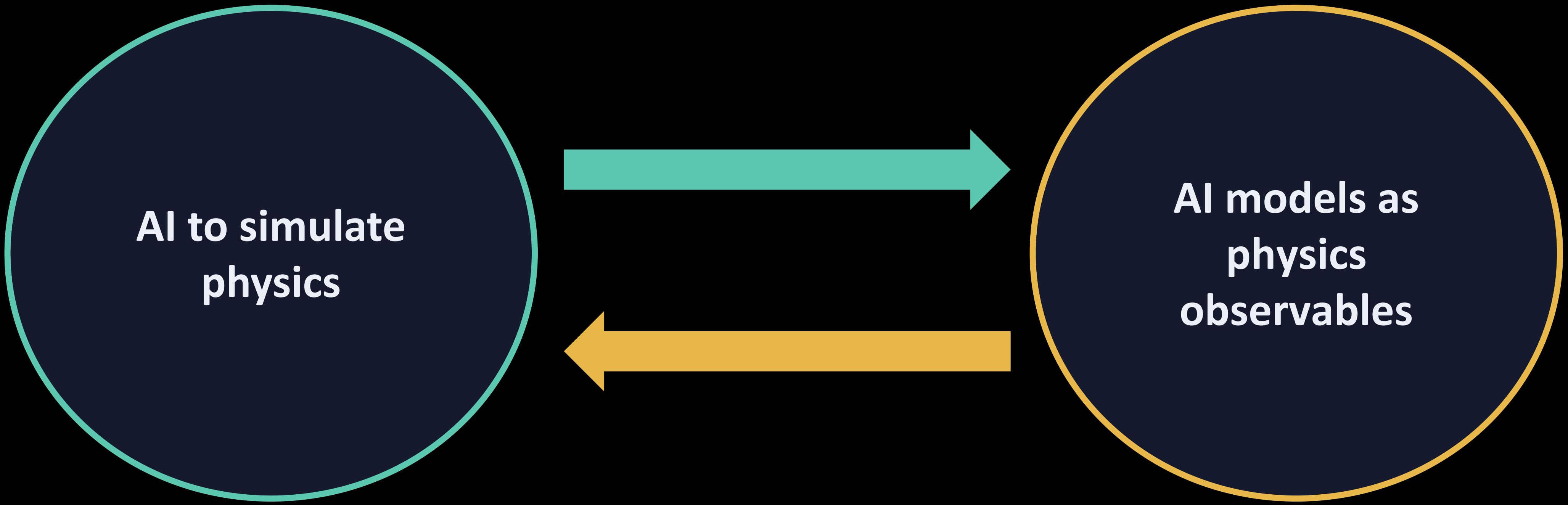
- with JEPAWG Hypernets [2026---]: for a set of g, tain the neural network once. Then train the hypernet. Generate new NN for unseen g.

- with JEPAWG Hypernets with interpretable feature space: Look at your feature space directly and read off the phase diagram. Qualitative understanding achieved.

Disclaimer: Currently, this paradigm can still be computationally heavy. But this can change soon (WIP).

What’s Next?

More types of fields and Larger system sizes.



Goal: a new standard tool for theoretical physicists in the age of AI.

Thank you!

with M. Gerdes, T. Göbel, J. R. Ebelt, Z. Mensch, R. Bondesan, P. de Haan, C. Rainone.