Field Theory and Neural Networks

Strings 2025 @ NYU Abu Dhabi

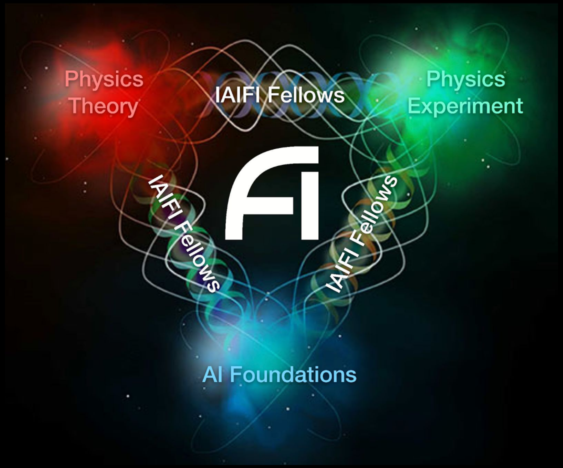
based on works with Maiti, Stoner, Demirtas, Schwartz, Tian, Naskar, and Ferko

Northeastern University

Jim Halverson

NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI /aɪ-faɪ/)

Advance physics knowledge—from the smallest building blocks of nature to the largest structures in the universe—and galvanize AI research innovation



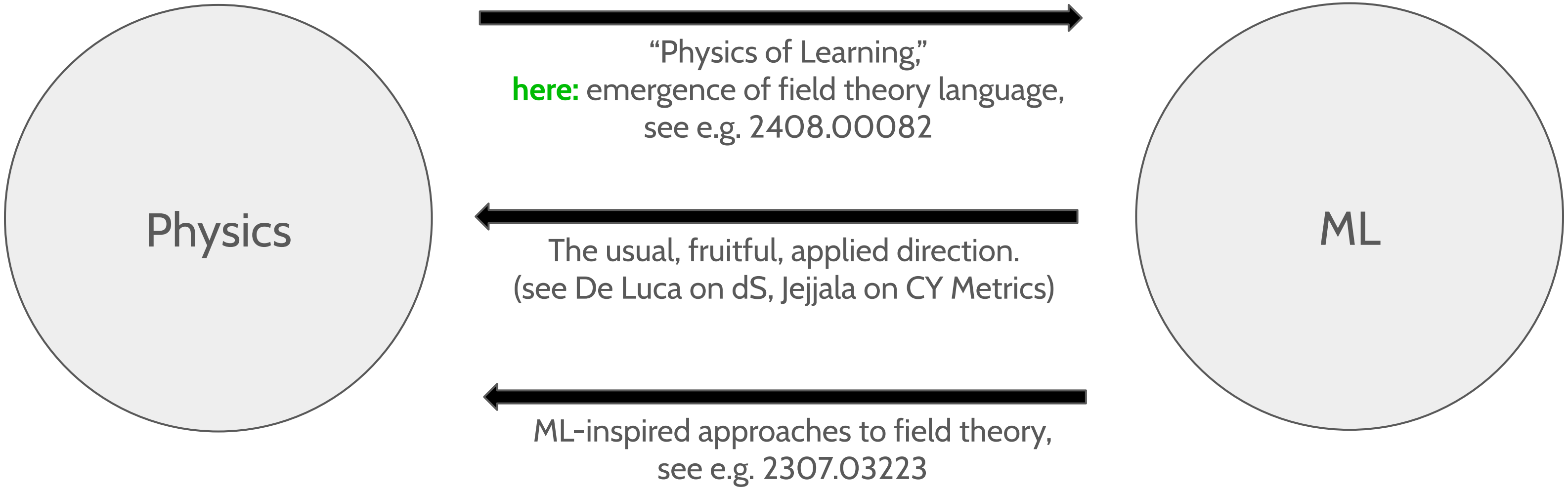
iaifi@mit.edu

https://iaifi.org

@iaifi-news

TASI Lectures on Physics for Machine Learning

Jim Halverson



Neural Network Field Theories: Non-Gaussianity, Actions, and Locality

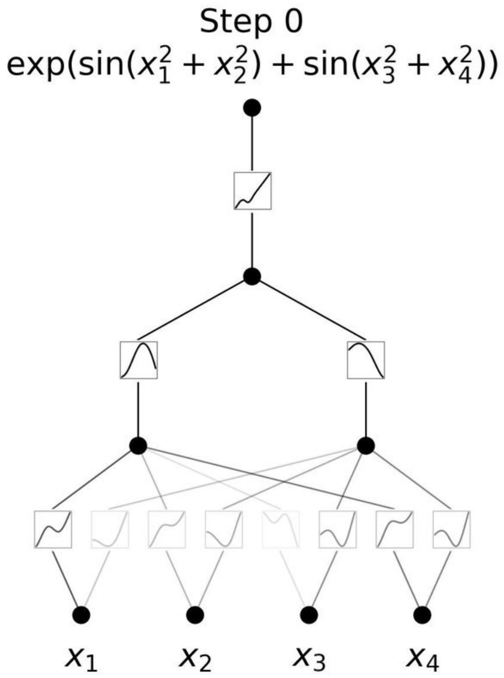
Mehmet Demirtas[1,2i], James Halverson[1,2ii], Anindita Maiti[1,2,4iii], Matthew D. Schwartz[1,3 iv], and Keegan Stoner[1,2v]

Understanding ML

Deep neural networks are compositions
of simpler parametrized functions.

Source of recent breakthroughs in ML,
so we should understand them.

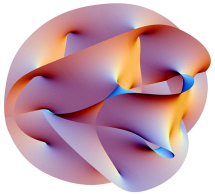
Optimizing NN Learning: Some Ideas from HET Community



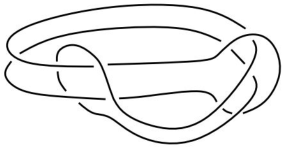
e.g. Kolmogorov-Arnold Network.

[Liu, J.H., et al], 2404.19756

- **Data:**

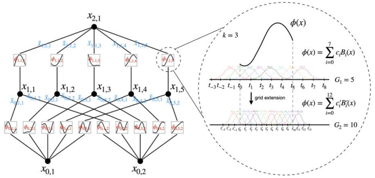


see Jejjala's talk



Hughes, Jejjala, Gukov, J.H., Ruehle, Manolescu
+ a number of others

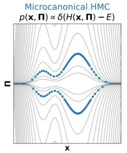
- **Architecture:**



Kolmogorov-Arnold Nets

Above: GDL textbook
e.g. also Group Equivariant Nets [Cohen, Welling]
SU(N) equivariance for lattice [Boyda et. al]

- **Optimization Dynamics:**



Energy Conserving Descent
[De Luca, Silverstein]

see also: [Klinger, Berman],
[Gerdes, Cheng, Welling]
[Tovey, Krippendorf et al]

- **Statistics:** e.g. works in this talk, [Dyer, Gur-Ari] [Yaida] [Hanin, Roberts, Yaida] (book!), [Erdmenger, Grosvenor, Jefferson], [Erbin, Lahoche, Samary] . . .

Big Takeaway:

Field Theory is a Natural Language for Neural Networks

What does a neural network predict? Two Complications.

Network:

$$\phi_\theta \in \mathrm{Maps}(\mathbb{R}^d, \mathbb{R})$$

simple answer: for fixed θ, x, predicts $\phi_\theta(x)$

But, dynamics: networks evolve along trajectories associated to a fixed architecture, data, and optimization algorithm. **Trajectories:**

$$\begin{aligned}
\text{Parameter Space:}\quad & \theta(t) \in \mathbb{R}^{|\theta|} \\
\text{Output Space:}\quad & \phi_{\theta(t)}(x) \in \mathbb{R} \\
\text{Function Space:}\quad & \phi_{\theta(t)} \in \mathrm{Maps}(\mathbb{R}^d, \mathbb{R}).
\end{aligned}$$

But, statistics: $\theta \sim P(\theta)$

~ means "drawn from"

$$\begin{aligned}
\mathbb{E}[\phi_\theta(x)] &= \int d\theta P(\theta)\, \phi_\theta(x) \\
\mathbb{E}[\phi_\theta(x)\phi_\theta(y)] &= \int d\theta P(\theta)\, \phi_\theta(x)\phi_\theta(y)
\end{aligned}
\quad \Longrightarrow \quad
\begin{aligned}
G^{(1)}(x) &= \langle \phi_\theta(x) \rangle \\
G^{(2)}(x, y) &= \langle \phi_\theta(x)\phi_\theta(y) \rangle
\end{aligned}$$

NNs are random functions at init, should compute expectations, **i.e. 1-pt and 2-pt functions are the average NN prediction and covariance, respectively.**

Dynamics + Statistics: to understand NNs as they learn, understand flow of 1-pt and 2-pt functions.

Notable: detailed theory exists, even exact result, **e.g.** NTK, mu-P, DMFT, etc.

[Jacot et al.]
[Lee et al.]
[Hu, Yang],
[Pehlevan et. al]

Quick Recap:

Understanding NN is essential to understanding ML.
There's an ensemble of them, evolving.

How does the average prediction evolve? The covariance?
These are questions about t-dependent 1-pt and 2-pt functions.

Question: do these FTs satisfy any properties we know and love?
can we engineer them to?

Outline: Field Theory and Neural Networks

- **Field Theory for NNs:** a natural language

- **Free Theories and NNs:** a physics surprise from ML theory

- **Neural Networks and Field Theory**
  i) generalities
  ii) symmetries
  iii) interactions
  iv) conformal fields
  v) unitarity

Free Theories and Neural Nets

A Physics Surprise from ML Theory

Example: Infinite Width Single-Layer Networks

[Neal], 90's.

a single-layer feedforward network is just

$$\phi_{\theta,N} : \mathbb{R}^d \xrightarrow{W_0} \mathbb{R}^N \xrightarrow{\sigma} \mathbb{R}^N \xrightarrow{W_1} \mathbb{R}$$

$$\phi_{\theta,N}(x) = W_1(\sigma(W_0 x))$$

Weight matrices W drawn i.i.d.

Consider $N \to \infty$ limit

Output adds an infinite number of i.i.d. entries from W_1 matrix, so CLT applies, NN drawn from Gaussian!

Free Theory Mechanism: Central Limit Theorem

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

where Φ_i are "neurons", i.i.d. of any arch.

Free Theory Limit: $$P[\phi] = e^{-S[\phi]}$$

$$S[\phi] = \frac{1}{2} \int \int d^dx d^dy \; \phi(x) \, G^{(2)}(x,y)^{-1} \, \phi(y)$$

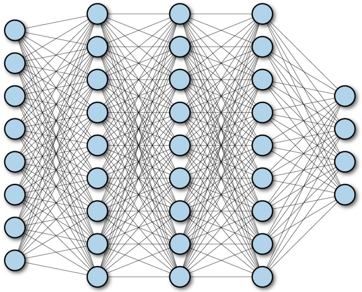
a $N \rightarrow \infty$, obtain free theory,
fields are Gaussian distributed by central limit thm.

Note: Gaussianity can persist during training.

Arises Everywhere: e.g.

$$z_i^l(x) = b_i^l + \sum_{j=1}^{N} W_{ij}^l x_j^l(x)$$

$$x_j^l = \sigma(z_j^{l-1}(x))$$



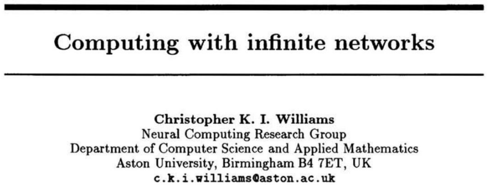
Deep FC nets, N = width.
Transformers, N = # attention heads.
Conv-nets, N = # channels.

Many, many architectures have free limit.

recent:
[Lee et al],
[Matthews et al]
[Yang], many refs therein.

Compute correlators in parameter space:

[Williams] 1996



Computing with infinite networks

Christopher K. I. Williams
Neural Computing Research Group
Department of Computer Science and Applied Mathematics
Aston University, Birmingham B4 7ET, UK
c.k.i.williams@aston.ac.uk

Neural Networks and Field Theory

i) generalities
ii) symmetries
iii) interactions
iv) conformal fields
v) unitarity

NN-FT: Generalities

e.g. [Demirtas, J.H., Maiti, Schwartz, Stoner]

Network:

$$\phi_\theta \in \mathrm{Maps}(\mathbb{R}^d, \mathbb{R})$$

Parameters Drawn at Init:

$$\theta \sim P(\theta)$$

Statistics → Correlators

$$\mathbb{E}[\phi_\theta(x)] = \int d\theta P(\theta)\, \phi_\theta(x)$$

$$\mathbb{E}[\phi_\theta(x)\phi_\theta(y)] = \int d\theta P(\theta)\, \phi_\theta(x)\phi_\theta(y)$$

Add dynamics for learning.

NN-FT Correspondence:

essential NN information is $(\phi_\theta, P(\theta))$
defines a field theory with
partition function given by

$$Z[J] = \int d\theta P(\theta) e^{\int d^d x J(x)\phi_\theta(x)}$$

can be related to Feynman's path integral

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

A different way to define a field theory.
Sometimes compute exact correlators, a la Williams.

NN-FT: Symmetries

$$Z[J] = \int d\theta P(\theta) e^{\int d^d x J(x) \phi_\theta(x)}$$

from invariance of the partition function.

Mechanism: [J.H., Maiti, Stoner]

1) transform field.
2) absorb transformation into parameters, redefine accordingly.
3) check invariance of Z[J], generally requires invariance of P(θ).

Examples: space symmetries on input (e.g. Euclidean), internal symmetries on output.

Example: Rotation Invariance

$$\phi_\theta(x) = g_{\theta_g} \circ W_{ij} x_j$$

$$\theta = \{W_{ij}, \theta_g\}$$

network is *any* NN g appended to linear layer Wx, where weights W are specific rot-invt P(w), i.i.d. Gaussian.

$$P(W) \propto \exp\left(-\frac{\mathrm{Tr}(W^T W)}{\sigma^2}\right)$$

NN-FT: Interactions

e.g. [J.H.], [Demirtas, J.H., Maiti, Schwartz, Stoner]

Key: central limit theorem yields free theories, violate its assumptions to get interactions, e.g. N → ∞ or stat. independence.

Interactions from 1/N-corrections:

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

$$G^{(2k)}(x_1, \ldots, x_{2k})\big|_{\text{connected}} \propto \frac{1}{N^{k-1}}$$

observed N-dependence of connected correlators

Note: Edgeworth expansion → action in 1/N.

Interactions from Independence Breaking:

same architecture as rotationally inv't example

$$\phi_\theta(x) = g_{\theta_g} \circ W_{ij} x_j$$

but λ-deformed param. density

$$P(W) \propto \exp\left(-\frac{\mathrm{Tr}(W^T W)}{\sigma^2} - \lambda\, \mathrm{Tr}(W^T W)^2\right)$$

that preserves rotational invariance but turns on interactions.

NN-FT: Local Interactions and ϕ^4 Theory

[Demirtas, J.H., Maiti, Schwartz, Stoner]

Engineer the free theory:

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

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

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

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

where i = 1, ..., N. in N $\to \infty$ limit get NNGP with

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

Introduce the Operator Insertion:

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

Absorb into Param. Density Deformation:

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

Write the Partition Function:

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

this is ϕ^4 theory as an infinite width NN-FT.

local interactions are from ***independence breaking***.

NN-FT: Conformal Fields

[J.H., Naskar, Tian]

Key Fact: Lorentz transformations in D+2 dimensions induce non-linearly realized conformal transformations on the D-dim projective null cone.

$$X_\mu = (X_0, x, X_{d+1}) \in \mathbb{R}^{D+1,1}$$

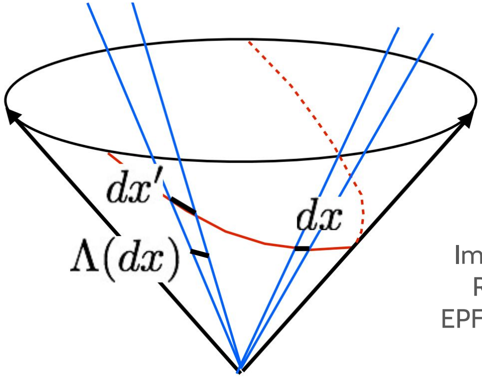


Image from Rychkov EPFL Lectures

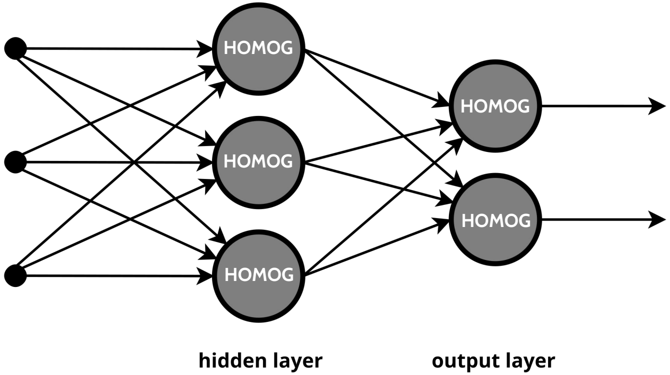
Take null cone, projectivize, choose Poincare

$$X_\mu = (X_+, x, X_-) = (1, x, x^2) \in \mathbb{R}^D \subsetneq \mathbb{R}^{D+1,1}$$

Construction idea:

Define Lorentz SO(D+1,1) invariant homogeneous theory on (D+1)-Minkowski, push to proj. null cone.

- **Homogeneity.**



- **Lorentz Invariance.** By previous mechanism.
- **Correlators.** Ensure well-behaved

NN-FT: Conformal Fields Pedagogical Example

[J.H., Naskar, Tian]

One Approach:

$$\begin{array}{ccc} \begin{array}{c} SO(D+2)\text{-symmetric} \\ \text{Homogeneous Theory} \\ \text{on } \mathbb{R}^{D+2} \end{array} & \xrightarrow{\text{Wick}} & \begin{array}{c} SO(D+1,1)\text{-symmetric} \\ \text{Homogeneous Theory} \\ \text{on } \mathbb{R}^{D+1,1} \end{array} \\ & & \Big\downarrow \text{Restrict} \\ & & \text{CFT on } \mathbb{R}^D \end{array}$$

Solve a Euclidean D+2 theory,
Wick rotate correlators to D+2 Lorentzian,
push down to null cone.

One Potential Euclidean D+2 Theory:

$$\Phi_E(X) = \Theta \cdot X$$

$$P(\Theta) \;\; \text{rotationally invariant}$$

yields Lorentzian theory with

$$G^{(2)}(X_1, X_2) = X_1 \cdot X_2$$

$$G^{(4)}(X_1, X_2, X_3, X_4) = \frac{\mu_4}{3} \left[(X_1 \cdot X_2)(X_3 \cdot X_4) + \text{perms} \right]$$

pushes to interacting conformal fields with

$$G^{(4)}(x_1, x_2, x_3, x_4) = \frac{1}{x_{12}^{2\Delta} x_{34}^{2\Delta}} g(u, v)$$

$$g(u,v) = \frac{\mu_4}{3} \left(1 + \frac{1}{u} + \frac{v}{u} \right), \qquad u = \frac{x_{12}^2 x_{34}^2}{x_{13}^2 x_{24}^2}, \qquad v = \frac{x_{14}^2 x_{23}^2}{x_{13}^2 x_{24}^2}$$

NN-QFT: Quantum Field Theories

Question: When is one of these Euclidean NN-FTs a *quantum* field theory? [J.H.] for a first example

Rely on Osterwalder-Schrader reconstruction theorem, constraints on Euclidean correlators sufficient to ensure a Lorentzian QFT.

Osterwalder-Schrader Axioms:
Constraints on Euclidean Correlators

1) Euclidean Invariance
2) Permutation Symmetry
3) Cluster Property
4) **Reflection Positivity**

$$\left\langle \mathcal{F}[\phi(Tx_1), \ldots, \phi(Tx_k)]^* \, \mathcal{F}[\phi(x_1), \ldots, \phi(x_k)] \right\rangle \geq 0$$

crucial for unitarity, absence of negative norm states.

Facts: Gaussian theories easy to check, Lagrangian defs of RP theories are RP, by perfect square mech. Have examples of both. **Outside those cases?**

NN-FT: Unitarity and Reflection Positivity

[Ferko, J.H.] WIP x 2, QM + QFT

In neural networks, the condition for RP is

$$\int d\theta P(\theta)\,\mathcal{F}_-^*\,\mathcal{F}_+ \geq 0$$

If we have a partition of parameters

$$\theta = \theta_0 \cup \theta_+ \cup \theta_-$$

$$P(\theta) = P(\theta_0)P_+(\theta_+,\theta_0)P_-(\theta_-,\theta_0)$$

s.t. $\mathcal{F}_\pm$ depends only on $\theta_\pm$ θ_0 then RP is

$$\int d\theta_0 P(\theta_0)\left(\int d\theta_-\, P_-(\theta_-,\theta_0)\,\mathcal{F}_-\right)^* \left(\int d\theta_+\, P_+(\theta_+,\theta_0)\mathcal{F}_+\right) \geq 0$$

If

$$\int d\theta_-\, P_-(\theta_-,\theta_0)\,\mathcal{F}_- = \int d\theta_+\, P_+(\theta_+,\theta_0)\mathcal{F}_+$$

then **integrand is perfect square, RP holds.** Can happen if architecture in F_- can absorb sign to become F_+ after change of variables, cond on P's.

Can realize this in simple architectures, but translation invariance requires more.

More generally:

Markov processes → RP, are useful.

NN acting on Markov process preserves RP.

Wide class of models, still exploring.

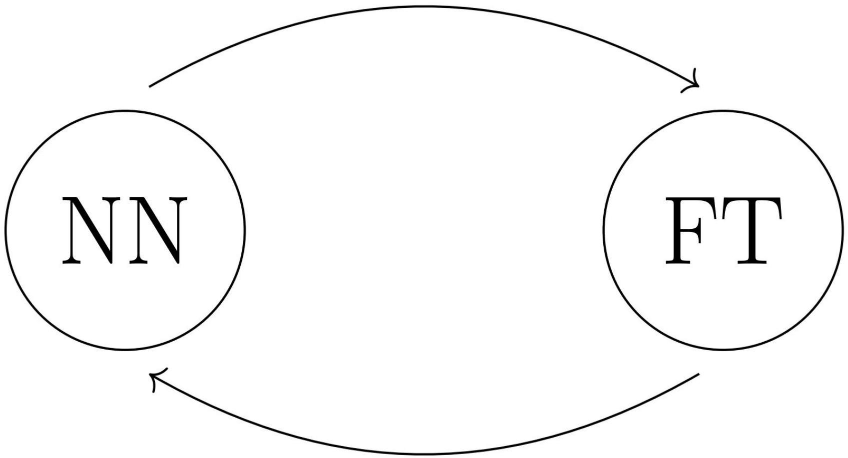
Recap: Field Theory and Neural Networks

- **Field Theory for NNs:** a natural language

- **Free Theories and NNs:** a physics surprise from ML theory

- **Neural Networks and Field Theory**
  i) generalities
  ii) symmetries
  iii) interactions
  iv) conformal fields
  v) unitarity

Outlook

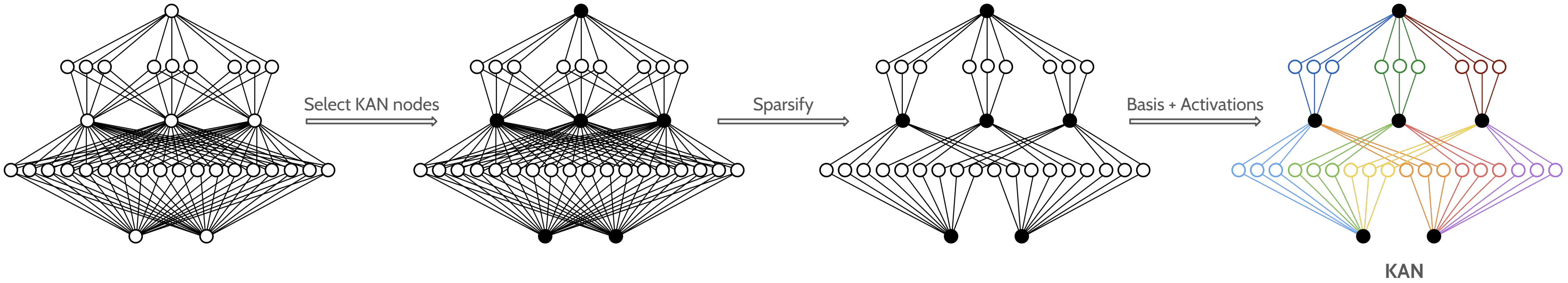


This talk: mostly about developing FT
from NN perspective, cherished physics principles.

Happy to give more physics outlook.

An ML Outlook

Sparsity, the development and understanding of smaller-but-powerful neural networks, is an extremely important direction.



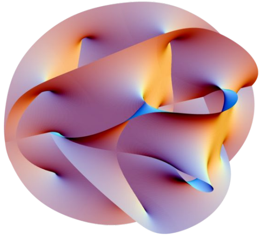
Can we sparsify models that have powerful approximation theorems while still retaining theoretical guarantees?

This is what a Kolmogorov-Arnold network does!
UAT → Sparsify → Kolmogorov-Arnold theorem.

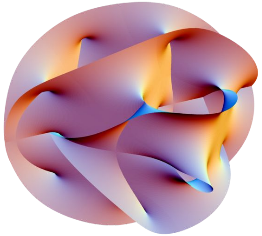
What is the field theory on each side? How does sparsification relate them?
Does this give new insights into architecture design?

IAIFI

Summer Workshop
August 11–August 15
2025

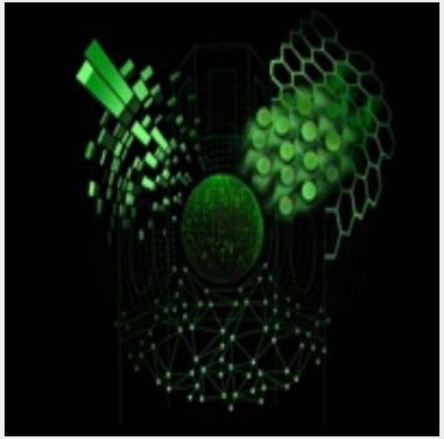


String Data 2025, December, London
Cambridge-Infosys AI Labs & LIMS



String Pheno 2025, July 7-11
Northeastern University

@ KITP



Generative AI for High & Low Energy Physics

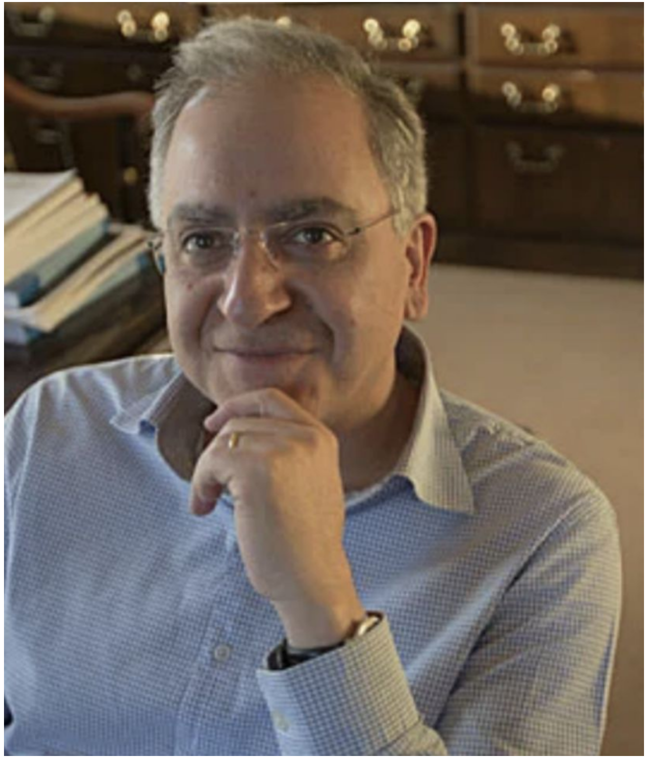
Nov 3, 2025 - Dec 19, 2025

Thanks!

Questions?

Feel free to get in touch:

e-mail: jhh@neu.edu
Twitter: @jhhalverson
web: www.jhhalverson.com



But what does this have to do with string theory?

Of course, I have dreams of this yielding a useful different perspective on quantum systems, and have a student thinking about the bosonic string.

Existing result: Used this type of theory to understand NN approximations of CY metrics, how Perelman's Ricci-Flow is realized in infinite limit, and *why finite NN learning of CY metrics is better*.

"Metric flows with Neural Networks" with Ruehle.