HAMMERS & NAILS SUM

@KyleCranmer
University of Wisconsin-Madison
American Family Insurance Data Science Institute

Thank you Tobias and Eilam!

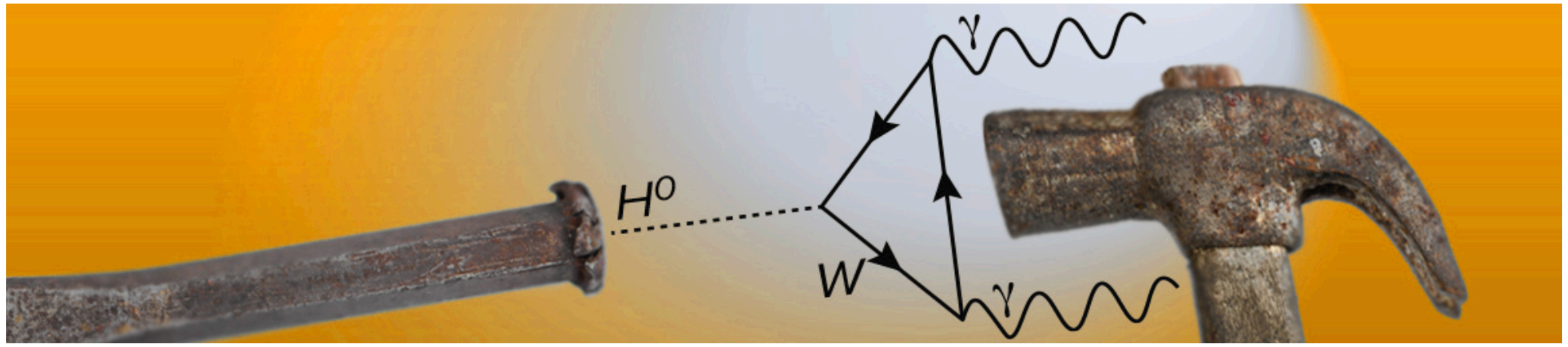
Hammers & Nails 2023 - Swiss Edition

29 Oct 2023, 09:00 → 3 Nov 2023, 19:00 Europe/Zurich

Congressi Stefano Franscini (CSF)

Description Frontiers in Machine Learning in Cosmology, Astro & Particle Physics

October 29 – November 3, 2023 | Conference center Congressi Stefano Franscini (CSF) in Monte Verità, Ascona, Switzerland



The Swiss Edition of Hammers & Nails in 2023 is following the success of the 2017, 2019 and 2022 Hammers & Nails workshops at Weizmann Institute of Science, Israel.

https://indico.cern.ch/event/1202995/timetable/

Major themes

- Inverse problems: Simulation-based Inference & Unfolding
- Anomaly detection
- Sampling high dimensional distributions
- Self-supervised learning — learning representations
- Leveraging physics knowledge / Inductive bias
- Multidisciplinary research and collaborations & cross-pollination
- Technical advances in deploying AI/ML in experiments

Look back at H&N
2017

Look back at H&N
2019

A photo of a robot that looks like Albert Einstein sitting under a tree with an apple falling on its head

H&N 2022

A photo of a robot that looks like Albert Einstein sitting under a tree with an apple falling on its head

H&N 2023

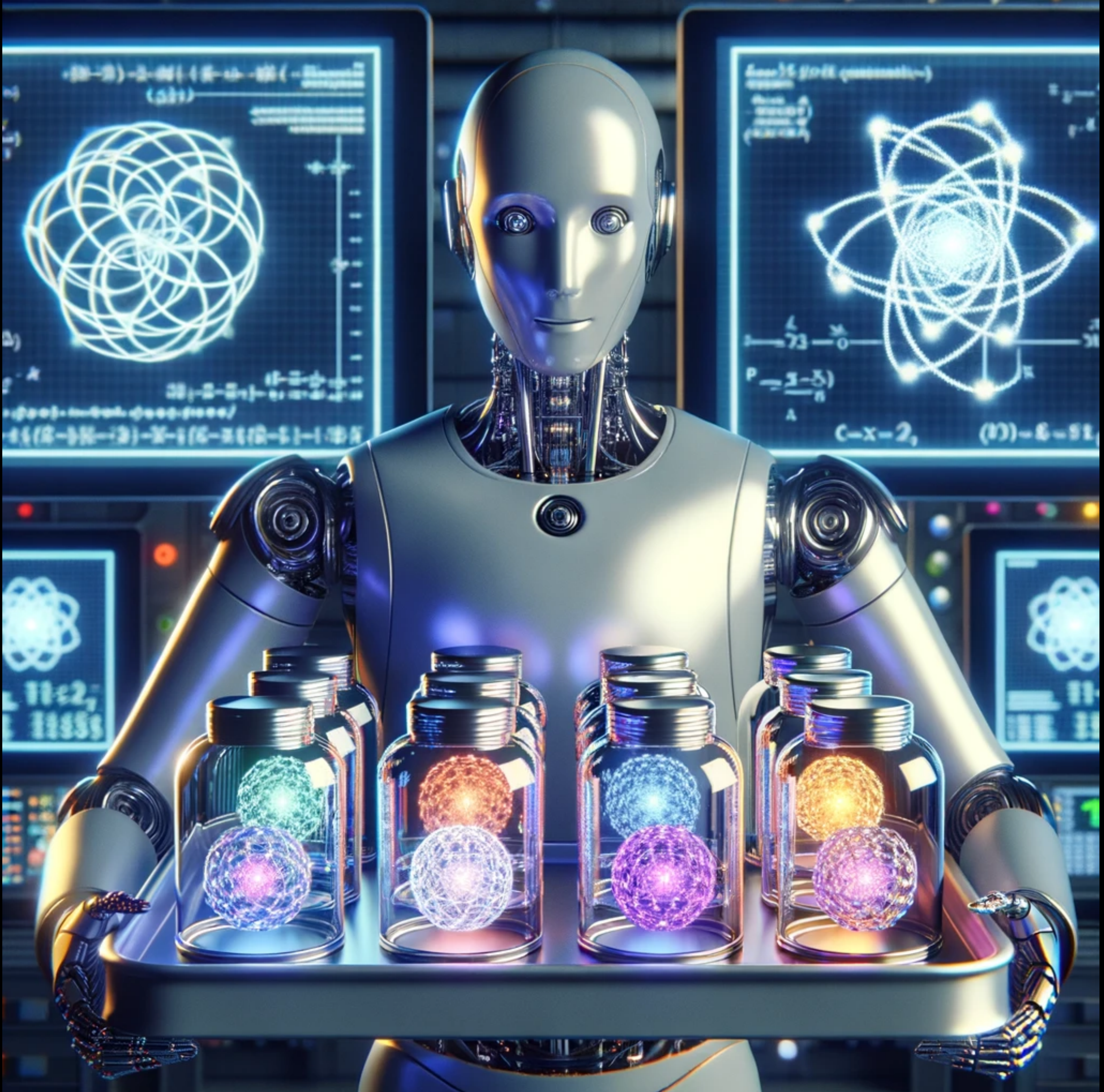
H&N 2022

"A robot with collection of glowing glass jars with Calabi-Yau manifolds inside"

H&N 2022

"A robot with collection
of glowing
glass jars with Calabi-Yau
manifolds inside"

H&N 2023

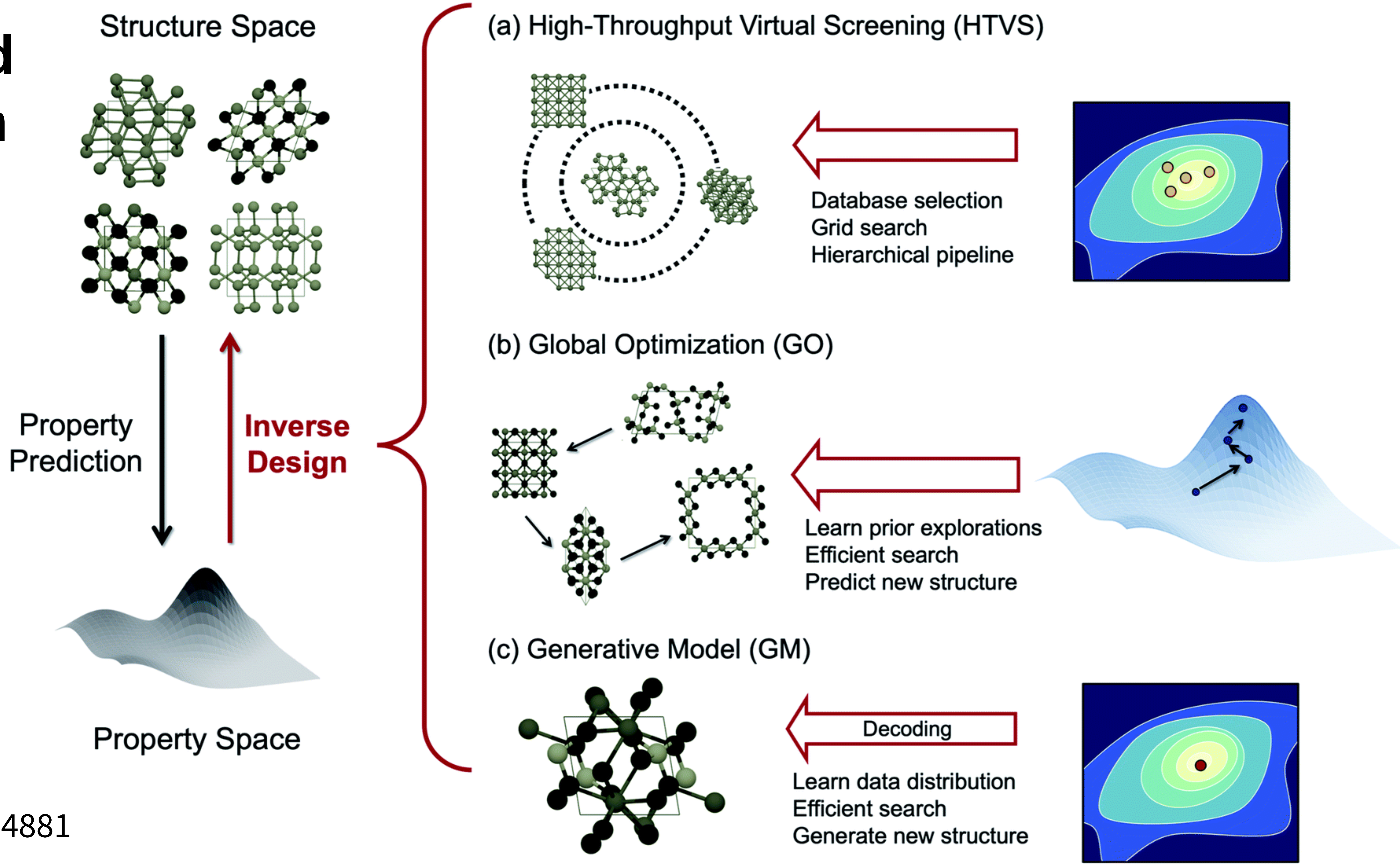


"A robot with collection of glowing glass jars with Calabi-Yau manifolds inside"

H&N 2023

Inverse Problems

Generative and Inverse Design

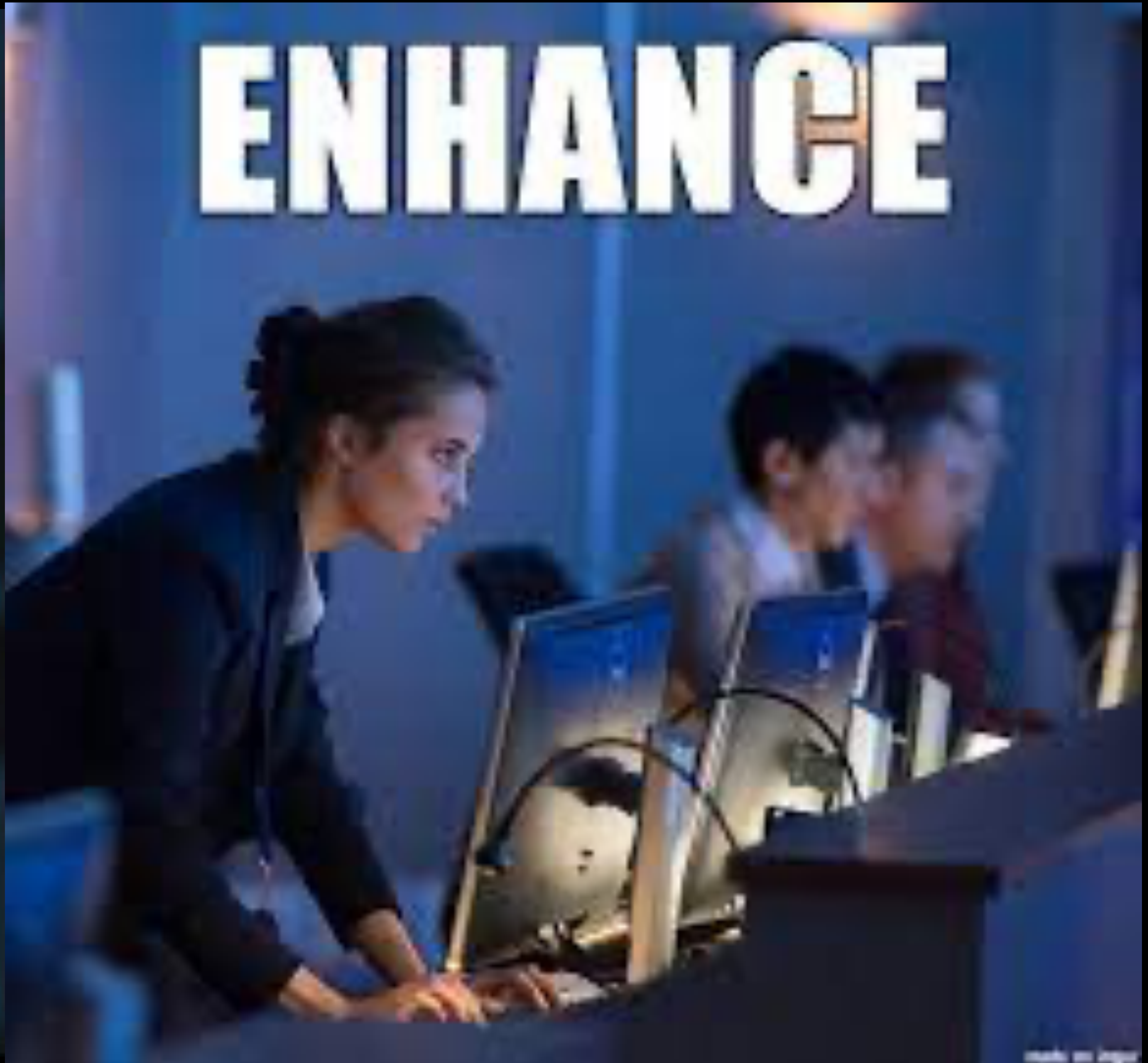


Chem. Sci., 2020,**11**, 4871-4881

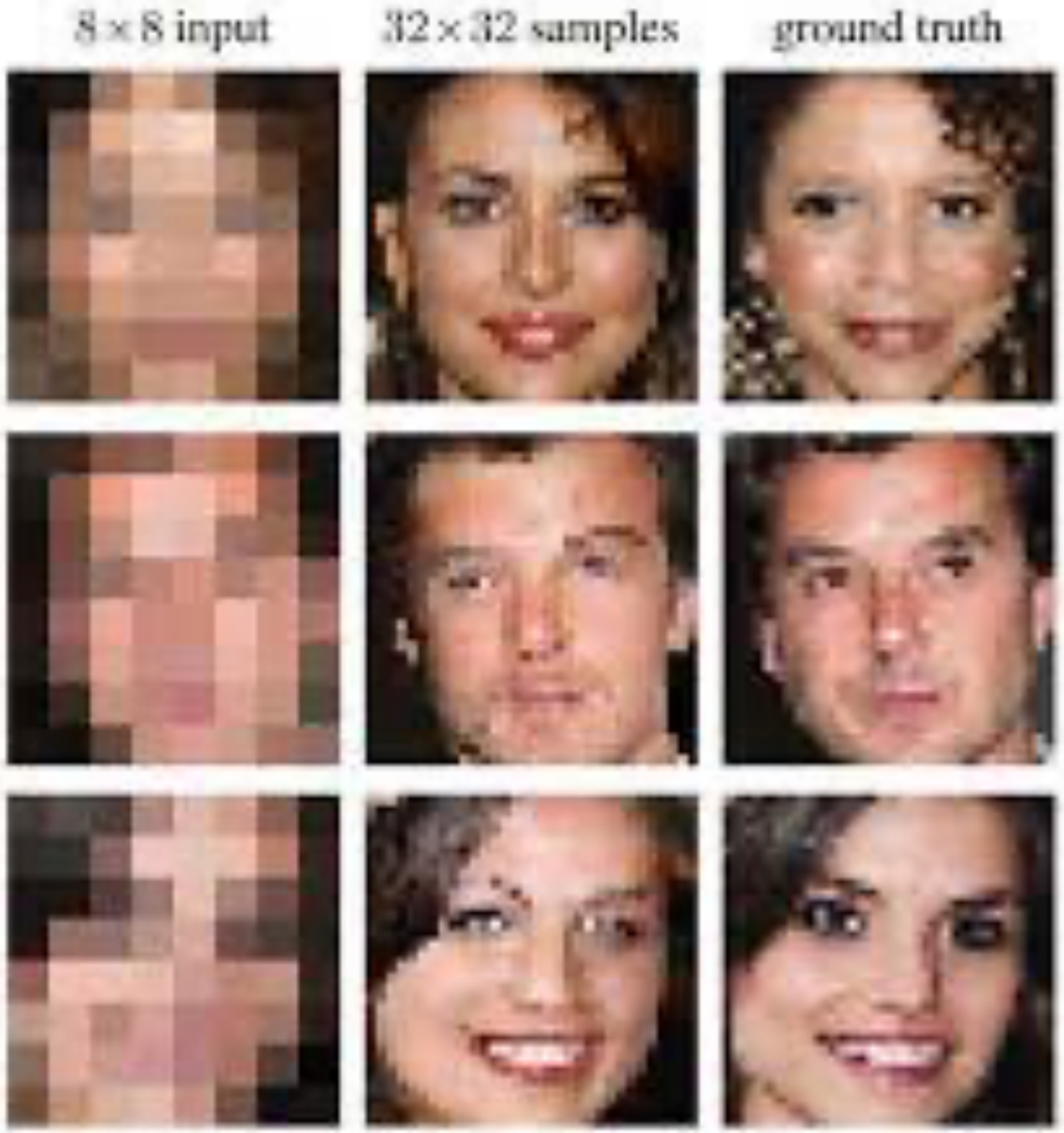
Simulation-based inference and the places it takes us

Jakob Macke, www.mackelab.org @mackelab
Machine Learning in Science, Tübingen University
Excellence Cluster Machine Learning & Tübingen AI Center
Bernstein Center for Computational Neuroscience Tübingen
Empirical Inference, Max Planck Institute for Intelligent Systems

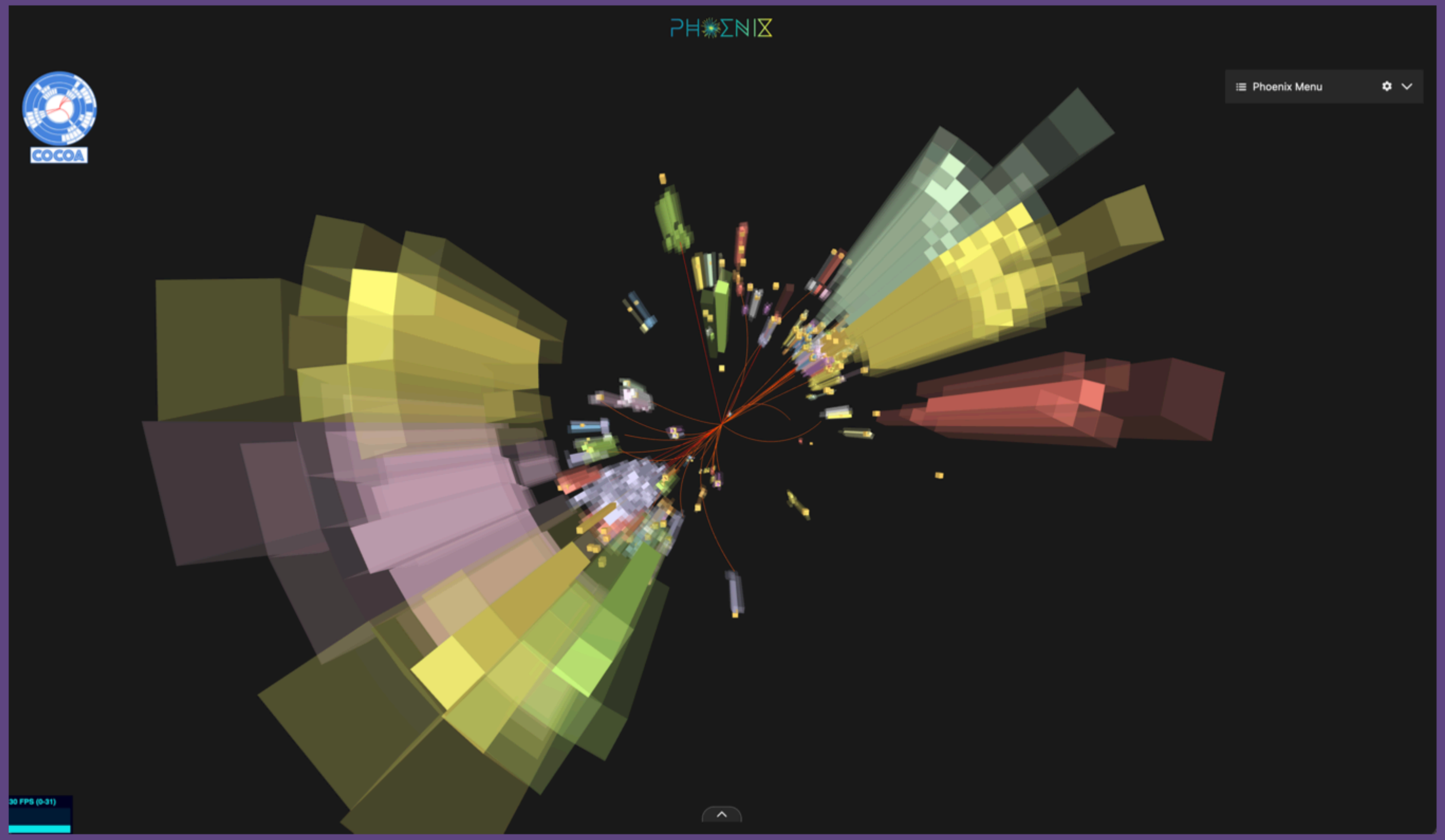
Super resolution



Super resolution



De-noising Graph Super Resolution with Diffusion Models and transformers



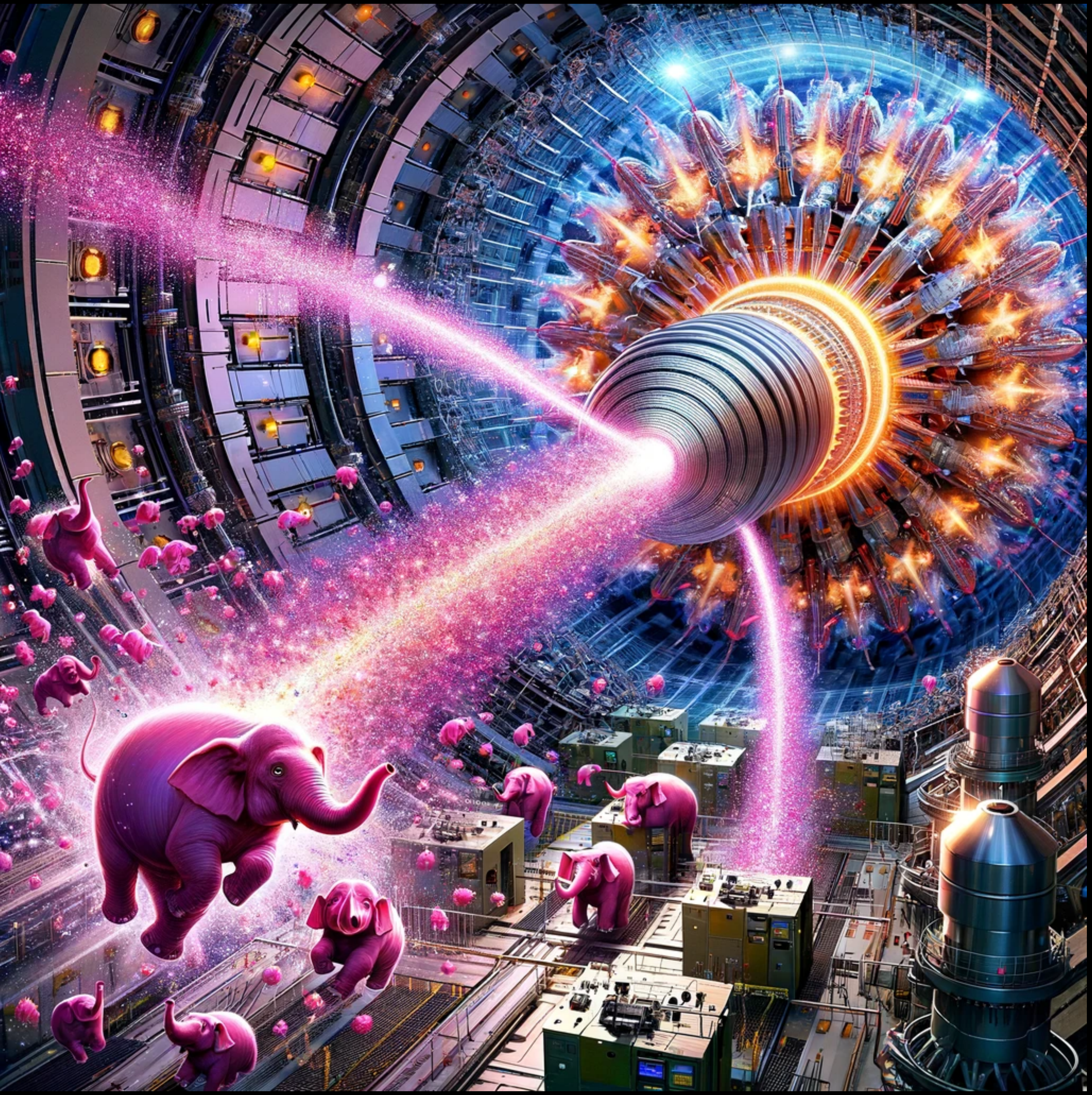
Hammers and Nails - Swiss Edition
30 November, 2023

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Nilotpal Kakati, *Etienne Dreyer, Eilam Gross*

(nilotpal.kakati@cern.ch)

Anomaly detection



"An image of particles colliding at the large hadron collider where the collisions are producing pink elephants"

H&N 2023

Anomaly detection

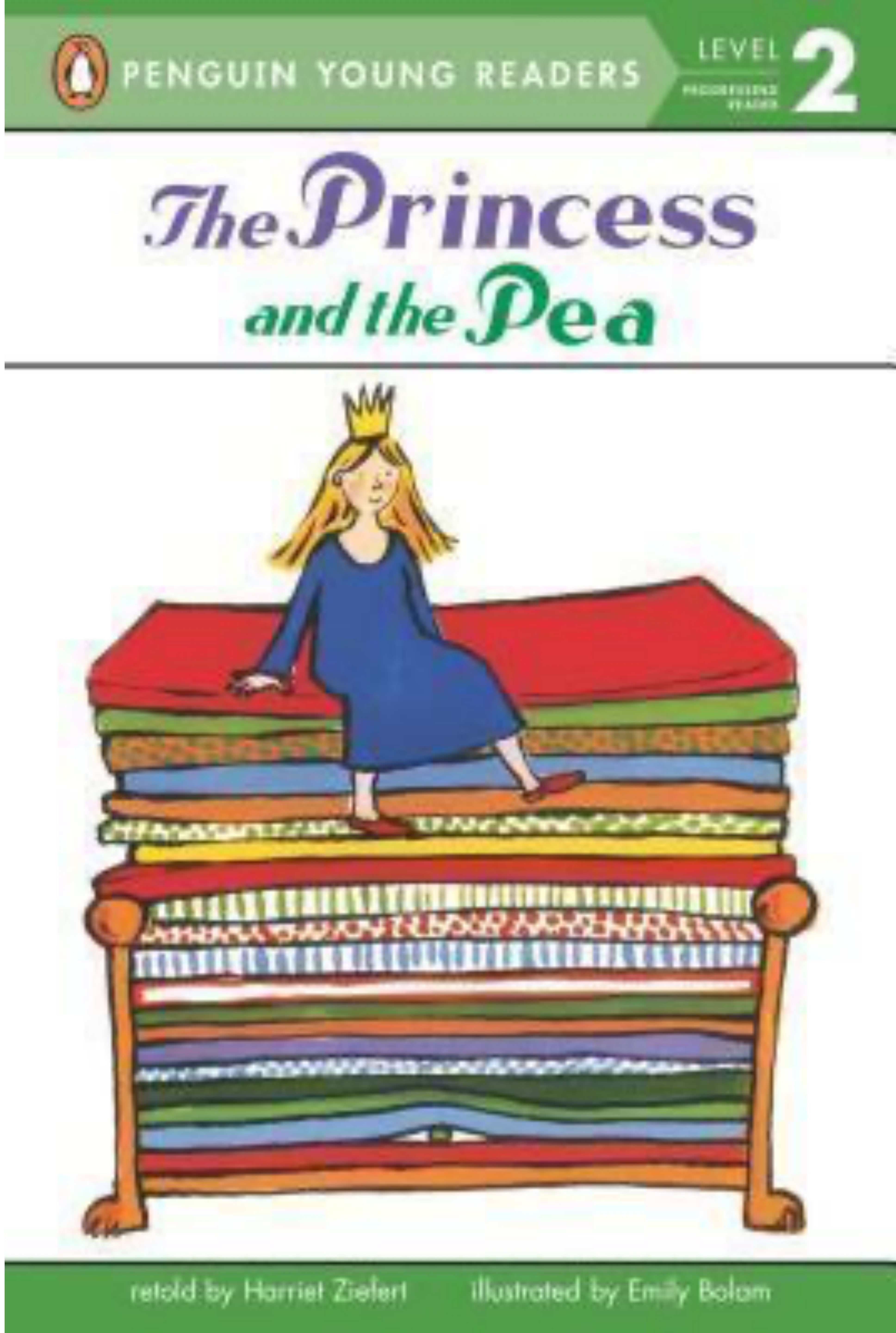
Classification **W**ith**o**ut **La**bels method (CWoLa)

Constructing **U**nobserved **R**egions by **T**ransforming **A**djacent **In**tervals (CURTAIns)

Anomaly **D**etection with Density **E**stimation (AnoDE)

Classifying **A**nomalies **TH**rough **O**uter **D**ensity **E**stimation (CATHODE)

Probabilistic **In**version **C**an **E**fficiently **S**pot **S**ignal (PrInCESS)



Some prompt based on kowala, Anode, cathode, curtains.

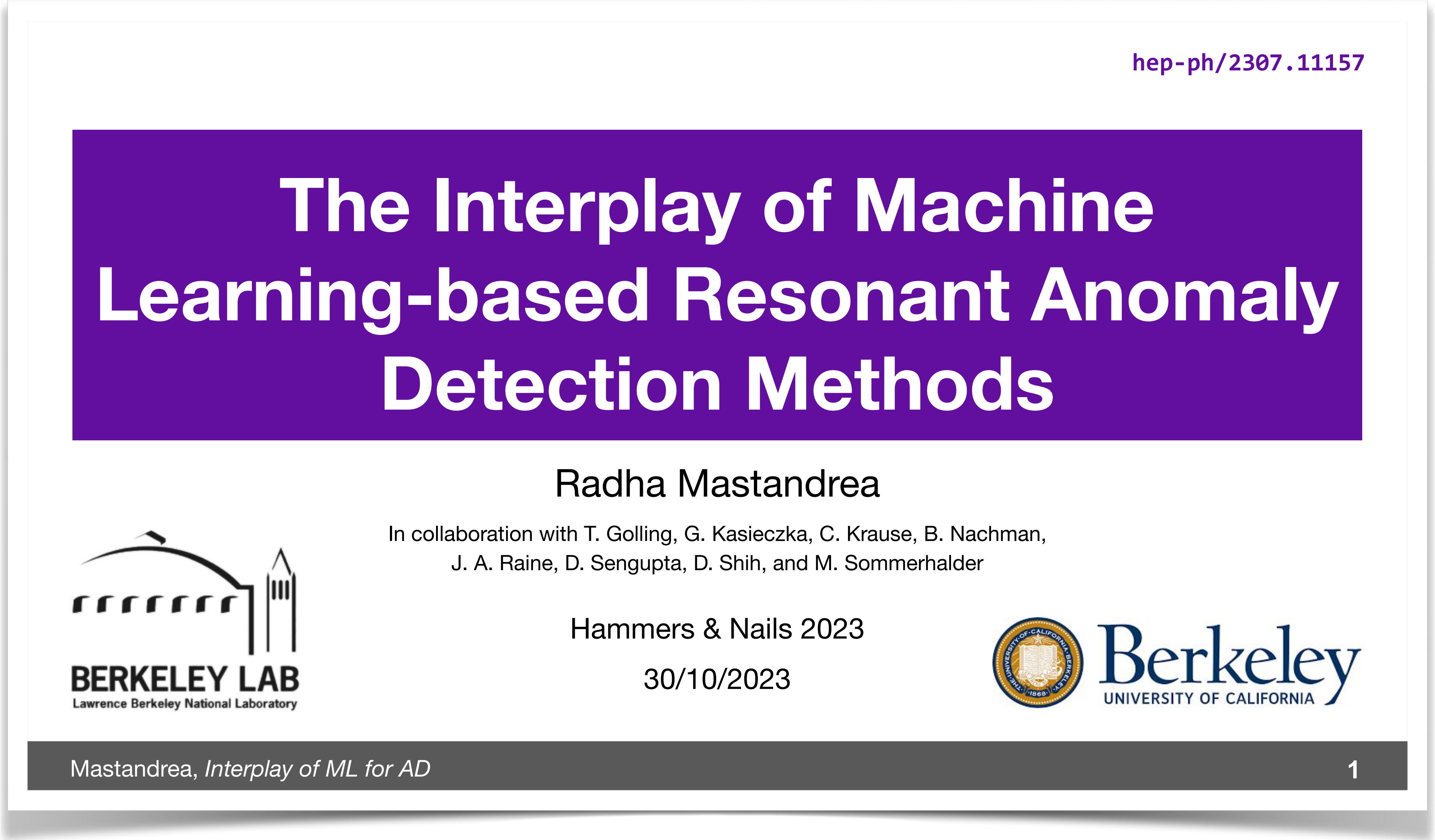
H&N 2022

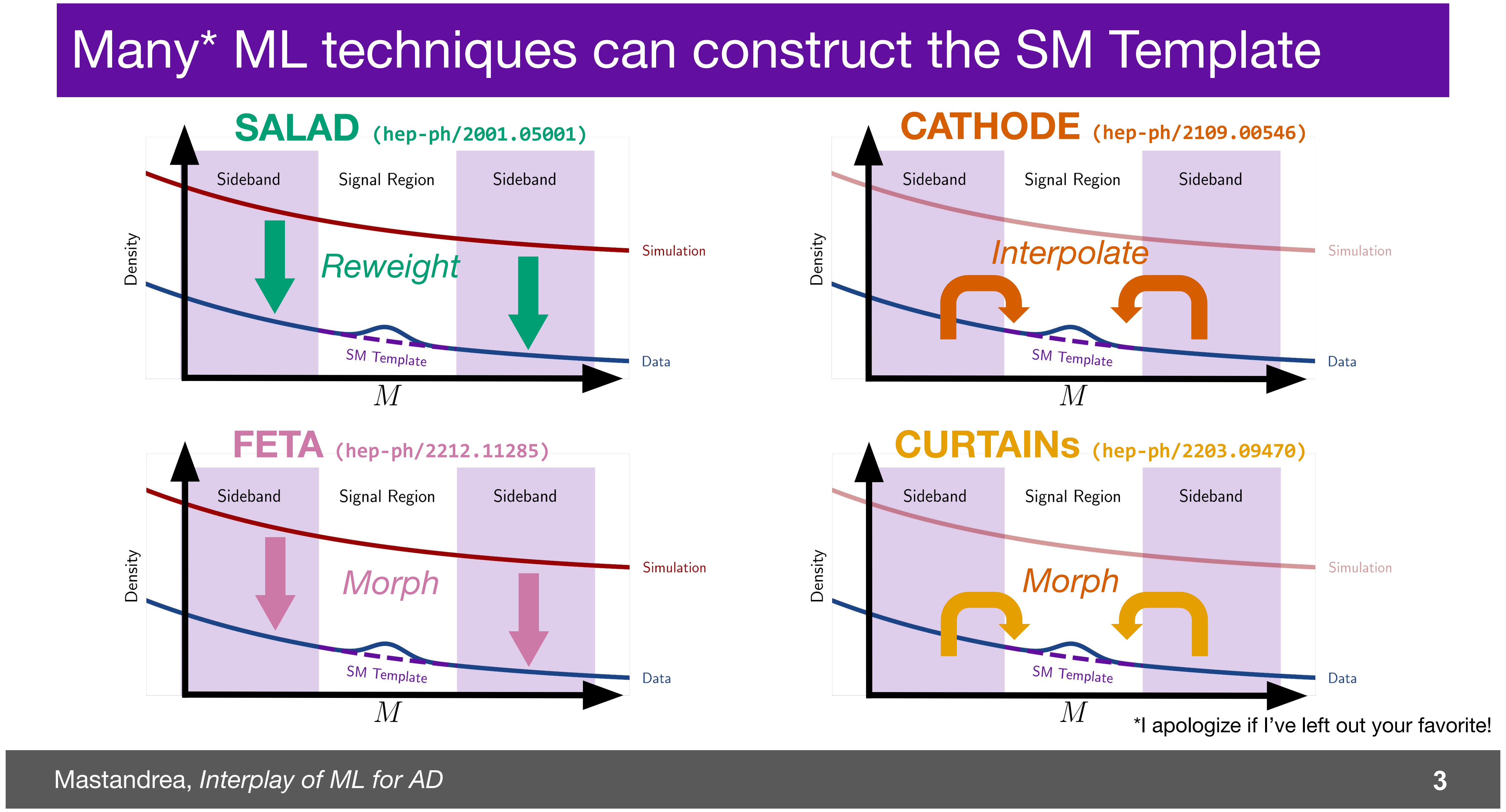
Fail!

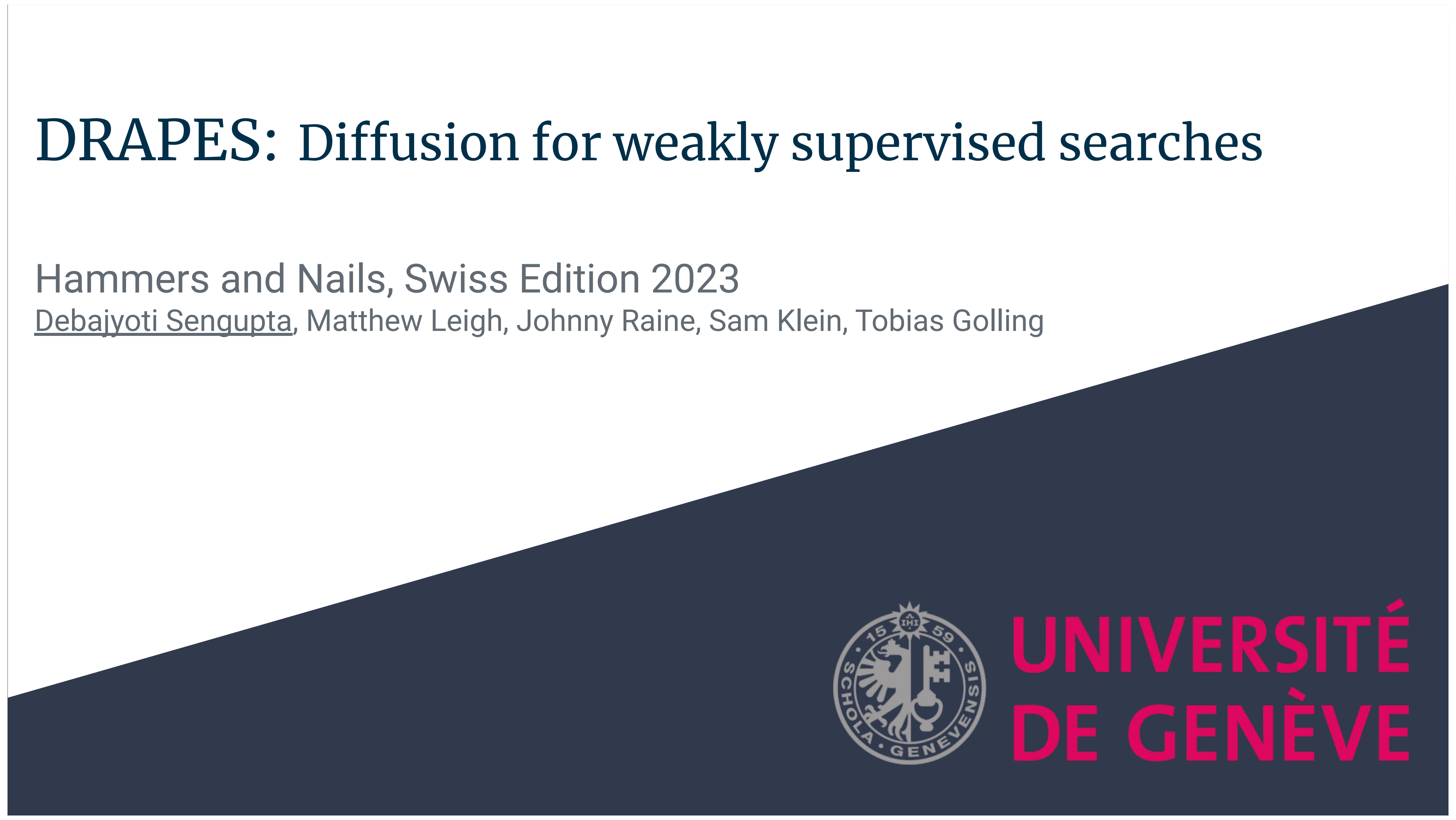
An image of a koala bear holding two electric cables with a large spark between the ends of the cable. The cables should be connected to a car battery. In the background there should be a window with curtains that are on fire.

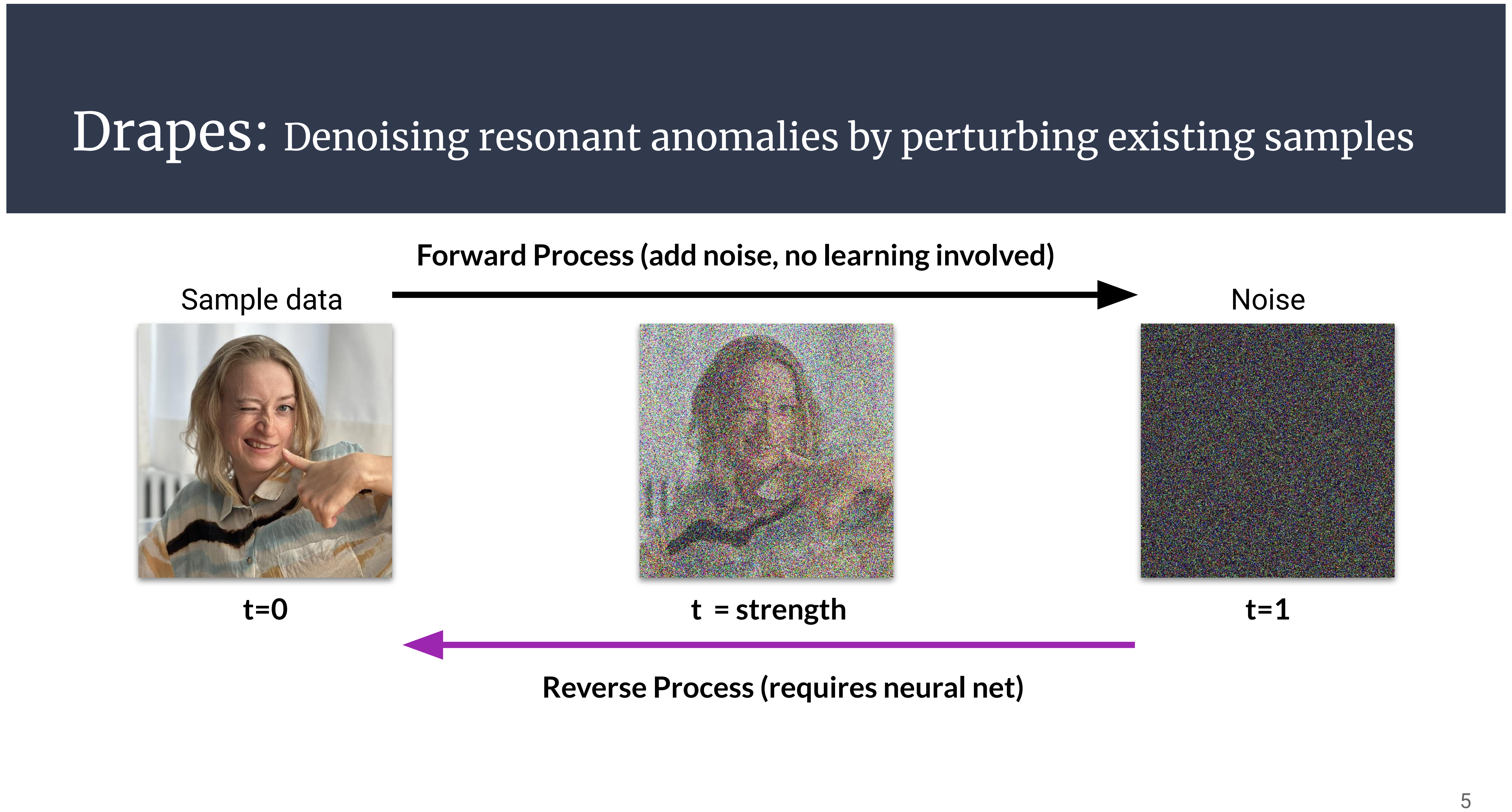
H&N 2023


Anomaly Detection









Sampling high dimensional distributions

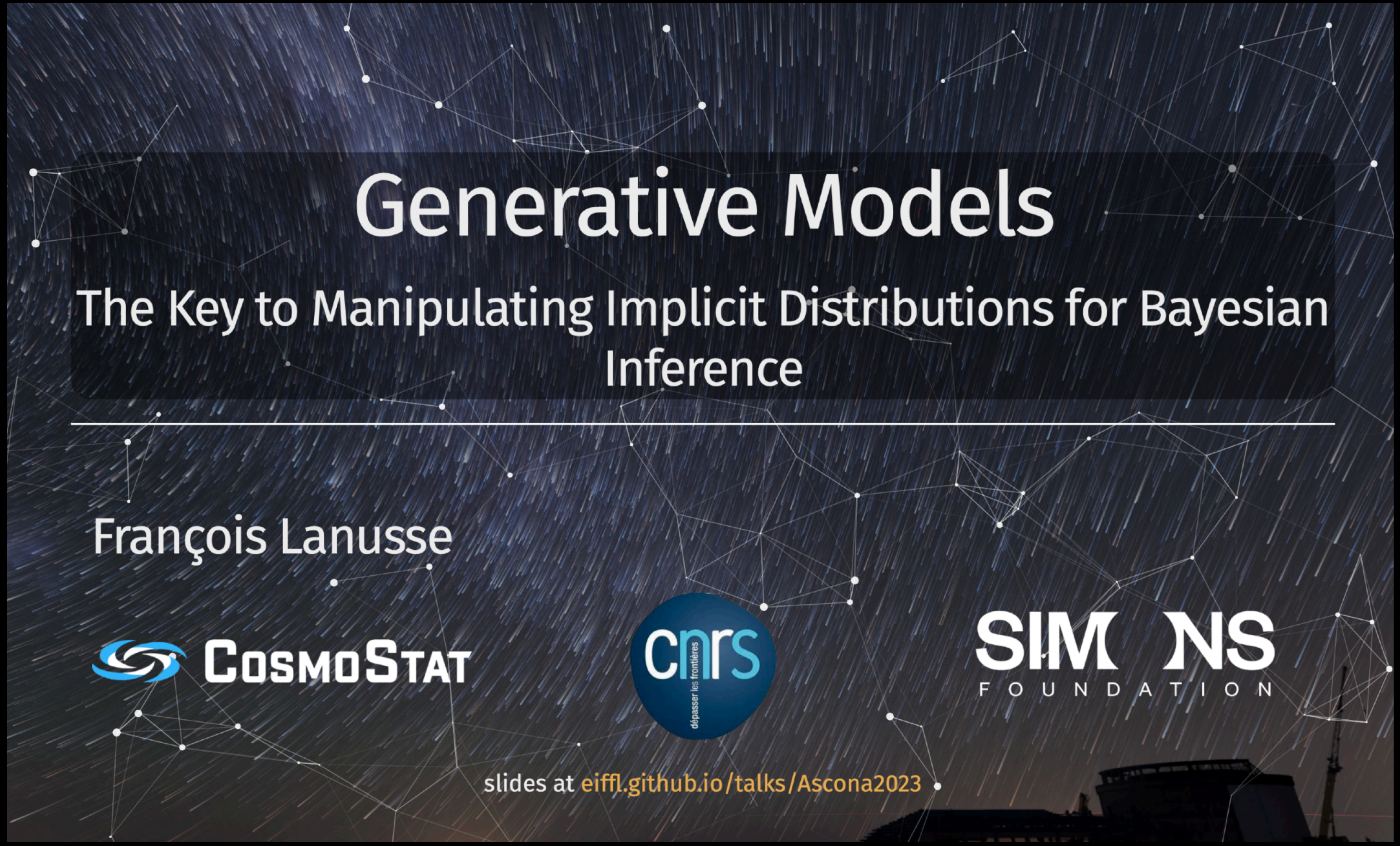
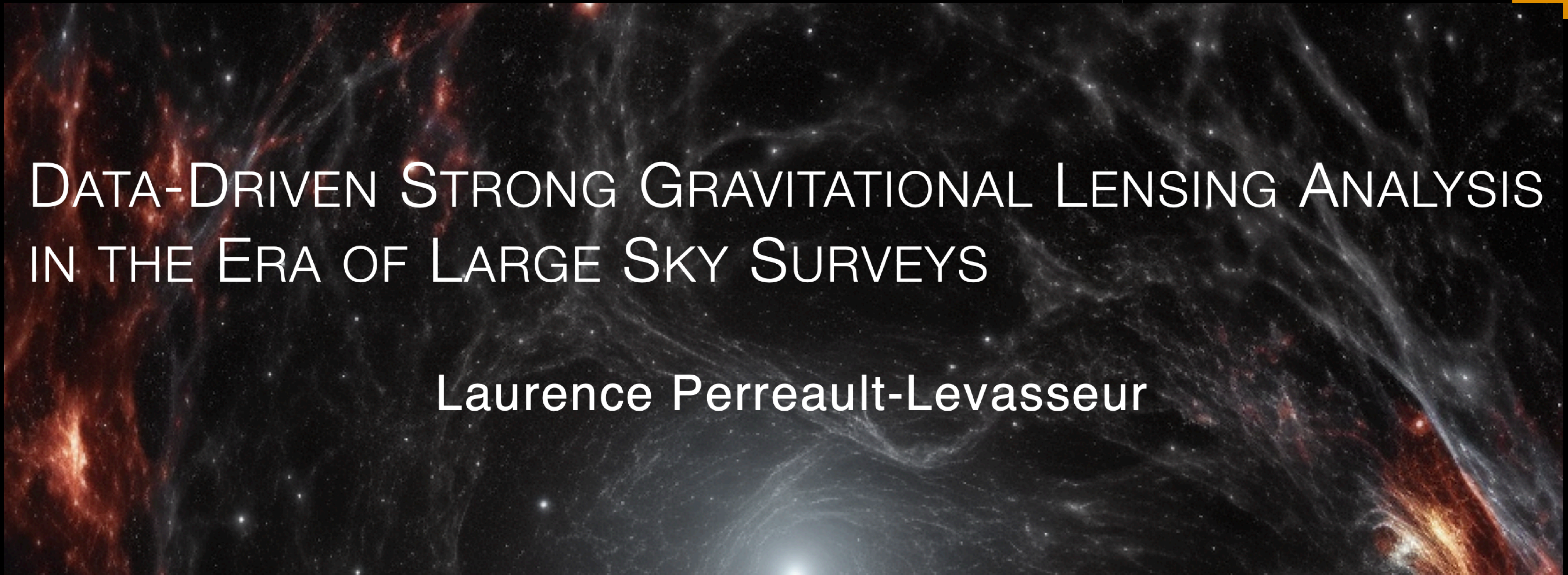

Generative Models
The Key to Manipulating Implicit Distributions for Bayesian Inference
François Lanusse
CosmoStat
cnrs
dépasser les frontières
SIMONS
FOUNDATION
slides at eiffl.github.io/talks/Ascona2023


Image Denoising
… Not What You Think
Michael Elad
Computer Science Department
The Technion - Israel Institute of Technology
Haifa 32000, Israel
Verily Research
November 1st 2023


Data-Driven Strong Gravitational Lensing Analysis in the Era of Large Sky Surveys
Laurence Perreault-Levasseur

Louis' Question

Say I only have a limited sample of training events, when will using these generative methods help?

- If I start with 100 events, can I really generate more?
- Often we are interested in tails of distributions

It's a good question… implicitly skeptical

- Can we get something from nothing?
  - No
- But can we get more from our samples if we have a model for what the data might be? Parametric or non-parametric with some inductive bias?
  - Yes

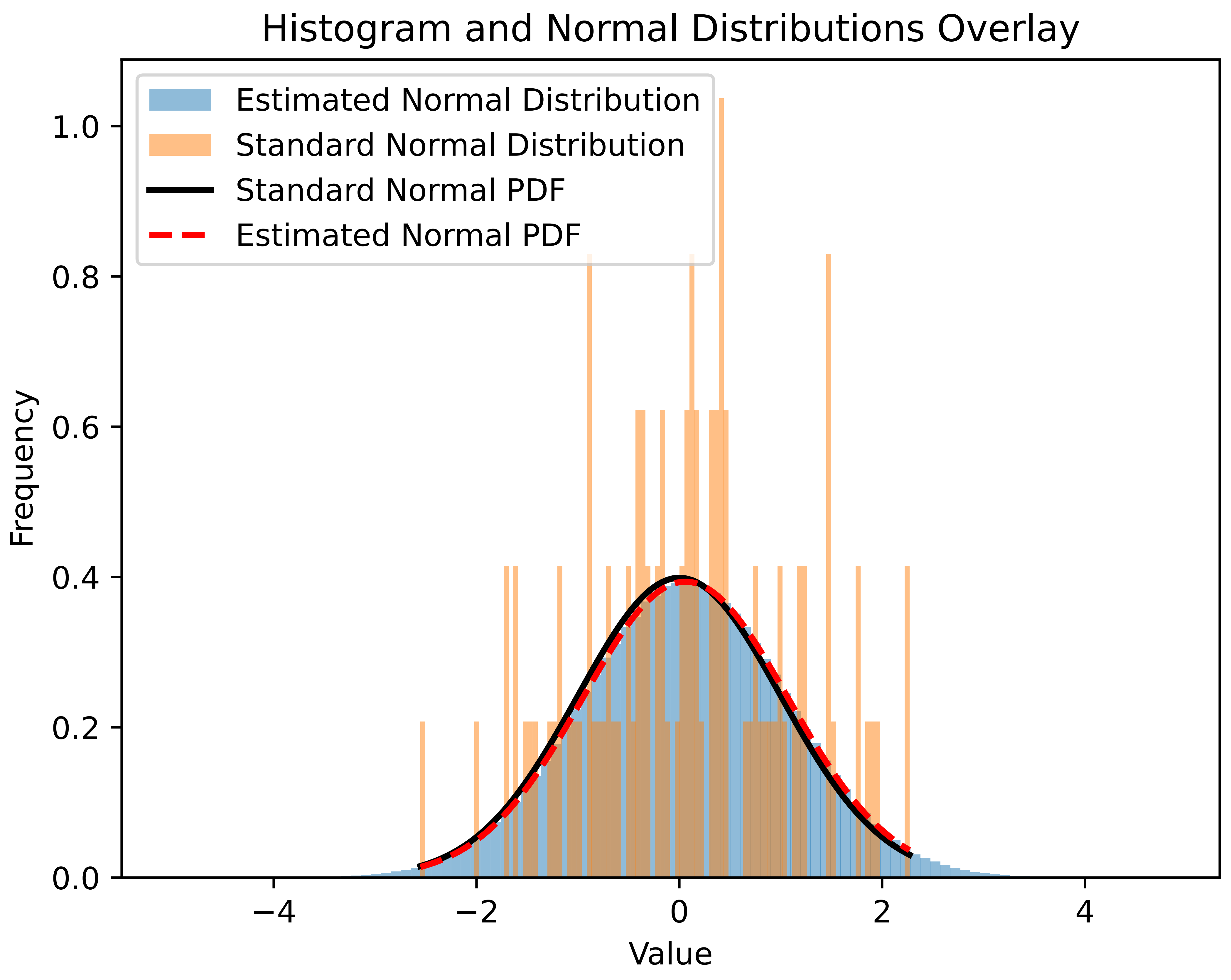
My example… I asked ChatGPT to code it up for me

Can you generate some python code for me that does the following.

- First, generate 100 samples from a standard normal distribution.
- Second, estimate the sample mean and sample standard deviation for the samples.
- Third, sample 1e6 samples from a normal distribution with this mean and standard deviation.
- Make a histogram of the samples and overlay the original standard normal and the normal with the estimated mean and standard deviation.

The code ChatGPT produced and the result

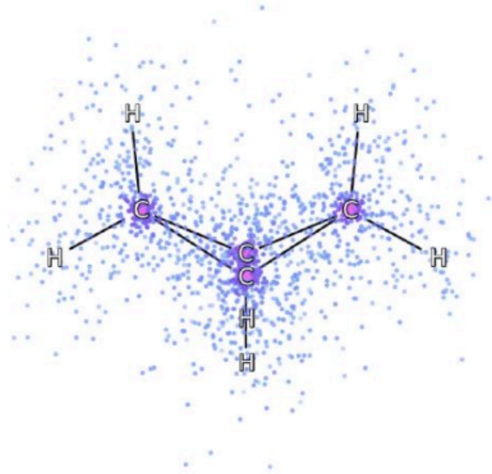
```
import numpy as np
import matplotlib.pyplot as plt

# Set a random seed for reproducibility
np.random.seed(0)

# Generate 100 samples from a standard normal distribution
samples_100 = np.random.normal(loc=0.0, scale=1.0, size=100)

# Estimate the sample mean and sample standard deviation
sample_mean = np.mean(samples_100)
sample_std = np.std(samples_100, ddof=1)  # ddof=1 provides an unbiased estimator

# Generate 1e6 samples from a normal distribution with the estimated mean and std deviation
samples_1e6 = np.random.normal(loc=sample_mean, scale=sample_std, size=int(1e6))

# Make histogram of the 1e6 samples
plt.hist(samples_1e6, bins=100, density=True, alpha=0.5, label='Estimated Normal Distribution')

# Overlay the histogram of the original standard normal distribution
count, bins, ignored = plt.hist(samples_100, bins=100, density=True, alpha=0.5, label='Standard Normal Distribution'

# Plot the probability density function of the standard normal
from scipy.stats import norm
x = np.linspace(min(bins), max(bins), 100)
p = norm.pdf(x, 0, 1)
plt.plot(x, p, 'k', linewidth=2, label='Standard Normal PDF')

# Plot the probability density function of the estimated normal
p_est = norm.pdf(x, sample_mean, sample_std)
plt.plot(x, p_est, 'r--', linewidth=2, label='Estimated Normal PDF')

# Title and labels
plt.title('Histogram and Normal Distributions Overlay')
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.legend()

# Show plot
plt.show()
```



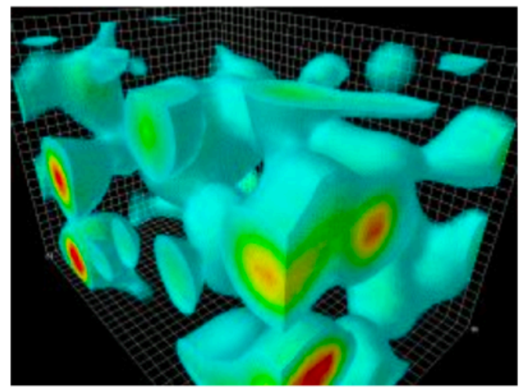
Yes, the model we fit on 100 examples is very useful for generating a million samples, even out in the tail.

- Uncertainty in model parameters will clearly propagate through. There is no magic.

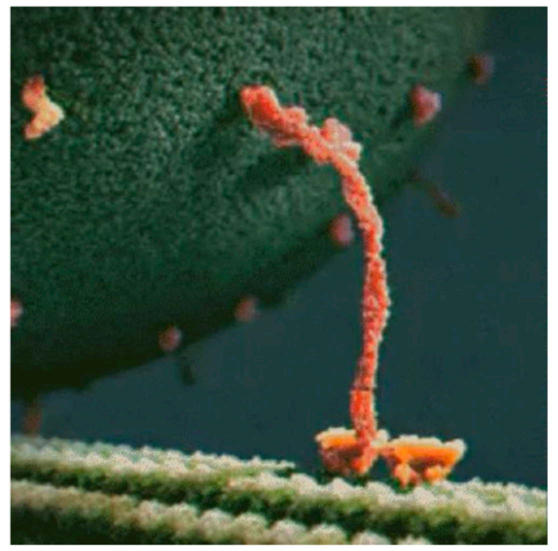
Alex Matthews this morning



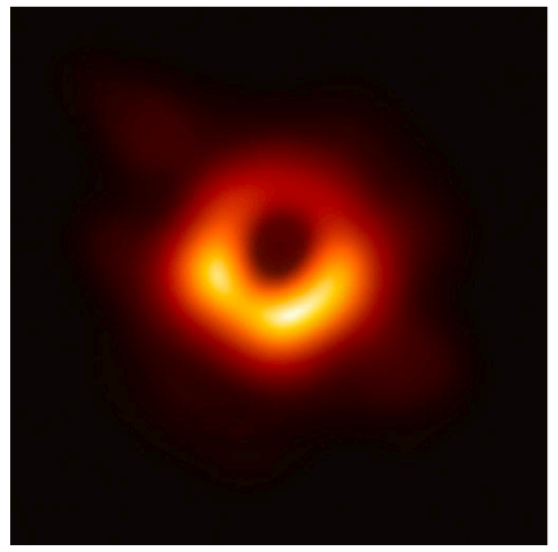
Quantum Monte Carlo



Lattice QCD



Protein physics



Black hole astronomy

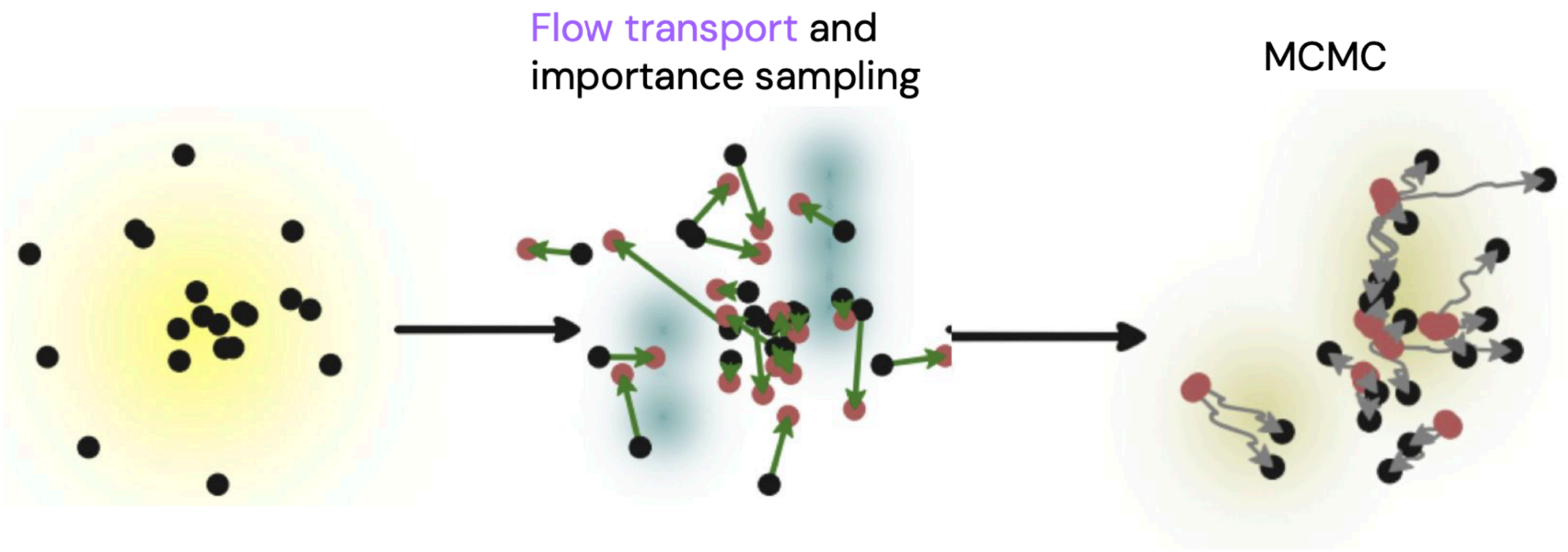
DeepMind

Normalizing flows, Diffusion and Annealed Importance Sampling

Alex Matthews
Hammer and Nails Conference.

3/11/2023

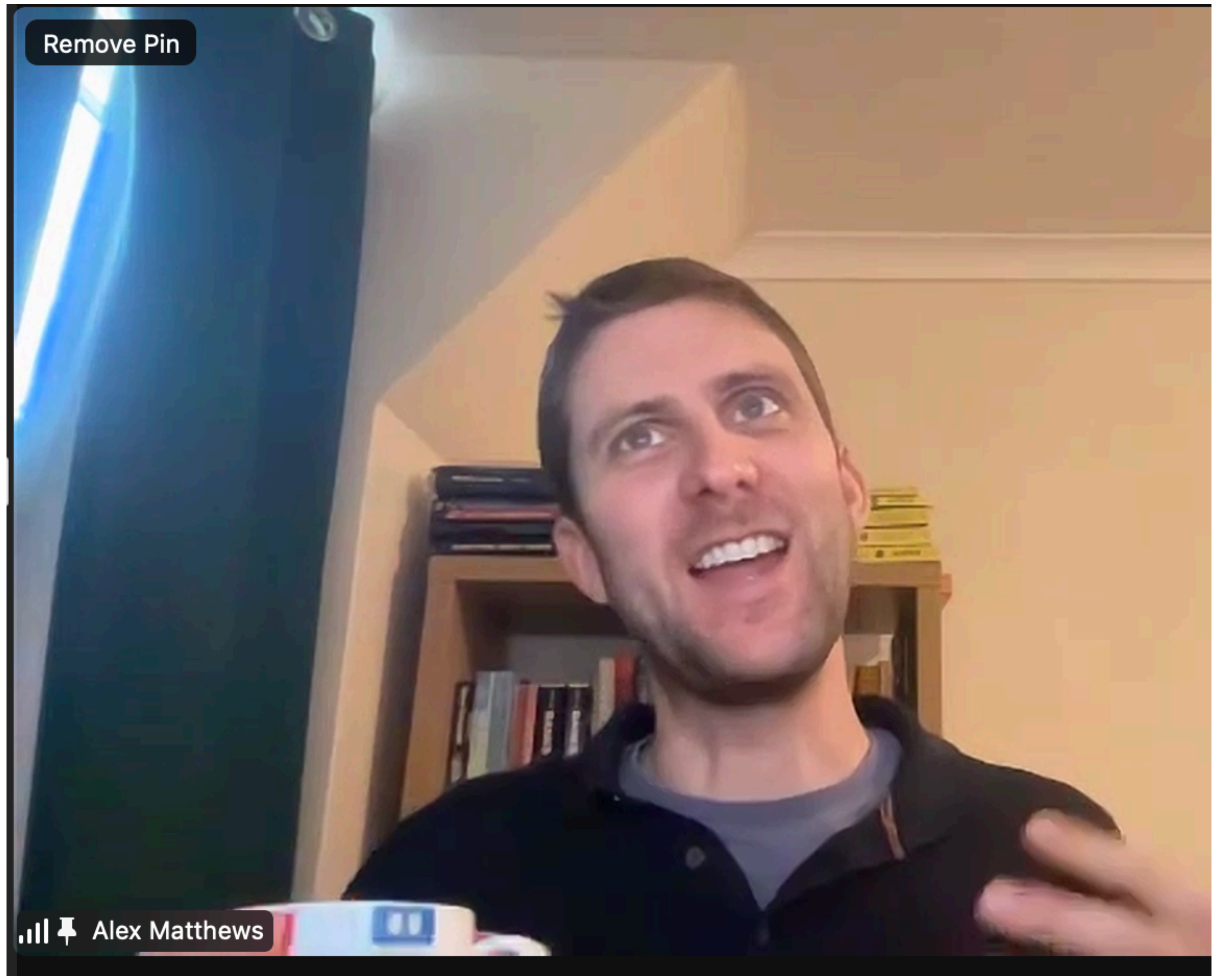
CRAFT one step with fixed normalizing flow(simplified)



$$w_k^{\mathrm{CRAFT}} = w_{k-1}^{\mathrm{CRAFT}} \frac{\gamma_k(T_k(x_{k-1}))}{\gamma_{k-1}(x_{k-1})} |\nabla T_k(x_{k-1})| \qquad x_k \sim F_k(\cdot | T_k(x_{k-1}))$$

AIS has identity flow.

Optimal and only valid reversal of a flow is its inverse.



Barnabas, this morning

- Entropy $-\int p \log p$
- KL Divergence $\int p \log \frac{p}{q}$
- Mutual Information $\int p_{XY} \log \frac{p_{XY}}{p_X p_Y}$

Fernandes & Gloor: Mutual information is critically dependent on prior assumptions: **would the correct estimate of mutual information please identify itself?**
BIOINFORMATICS Vol. 26 no. 9 2010, pages 1135–1139

Take me Home!

Some density functionals
(e.g entropy, mutual information, divergences)
can be estimated directly,
without estimating the densities first!

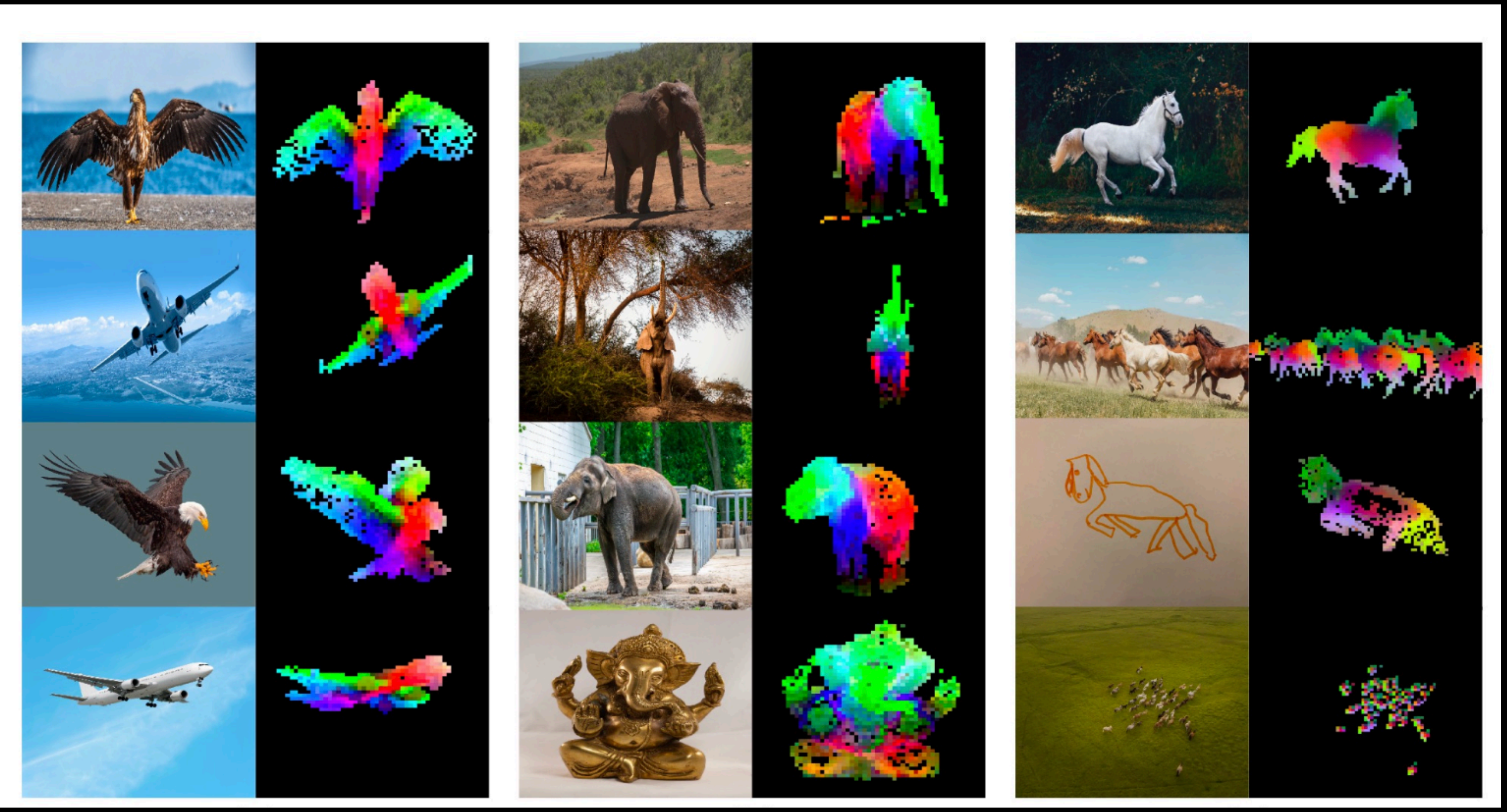
"When solving a problem of interest, do not solve a more general problem as an intermediate step"

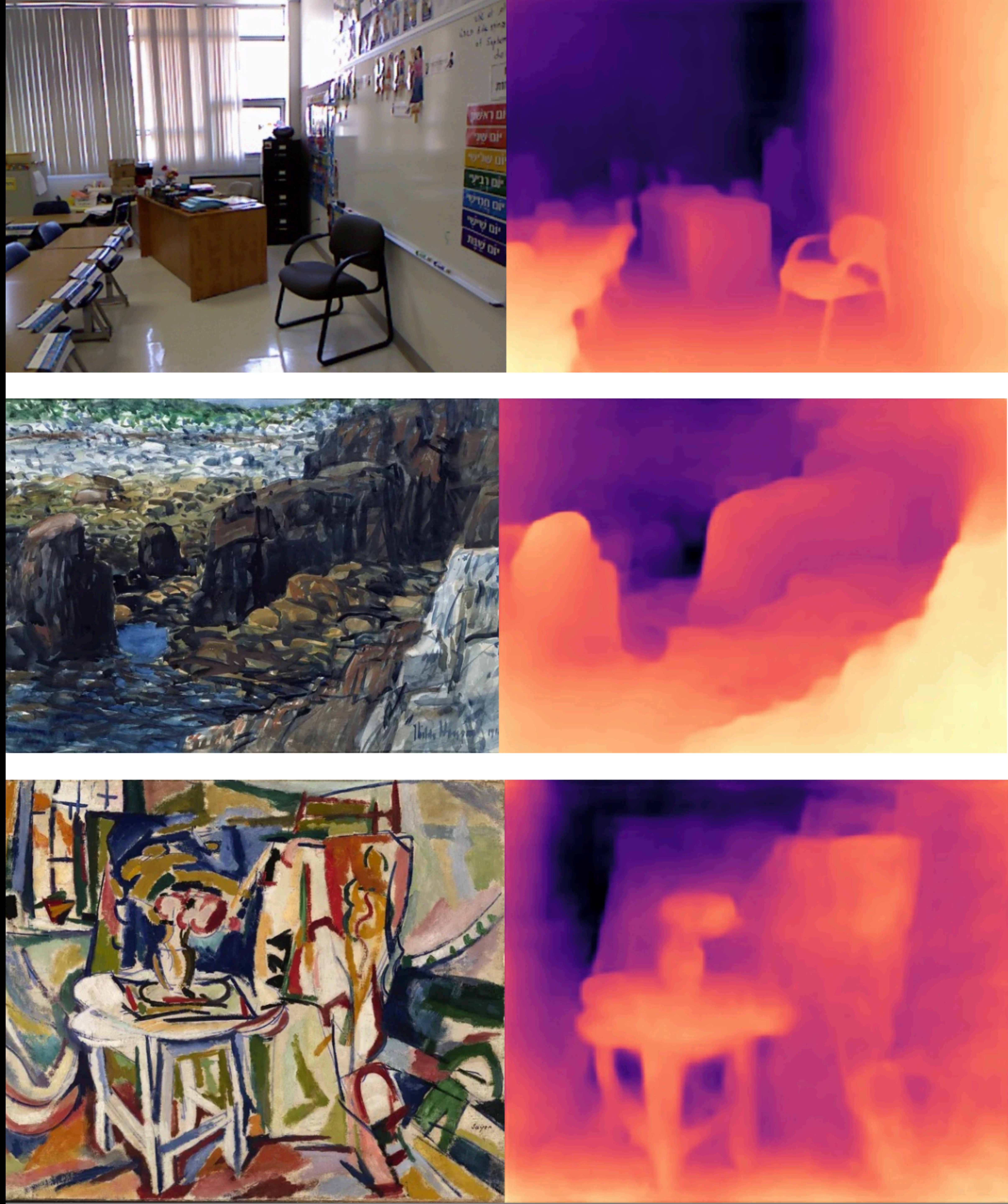
— Vladimir Vapnik

Self-supervised learning — learning representations

Learning Image Representations Without Manual Annotations and Related Applications

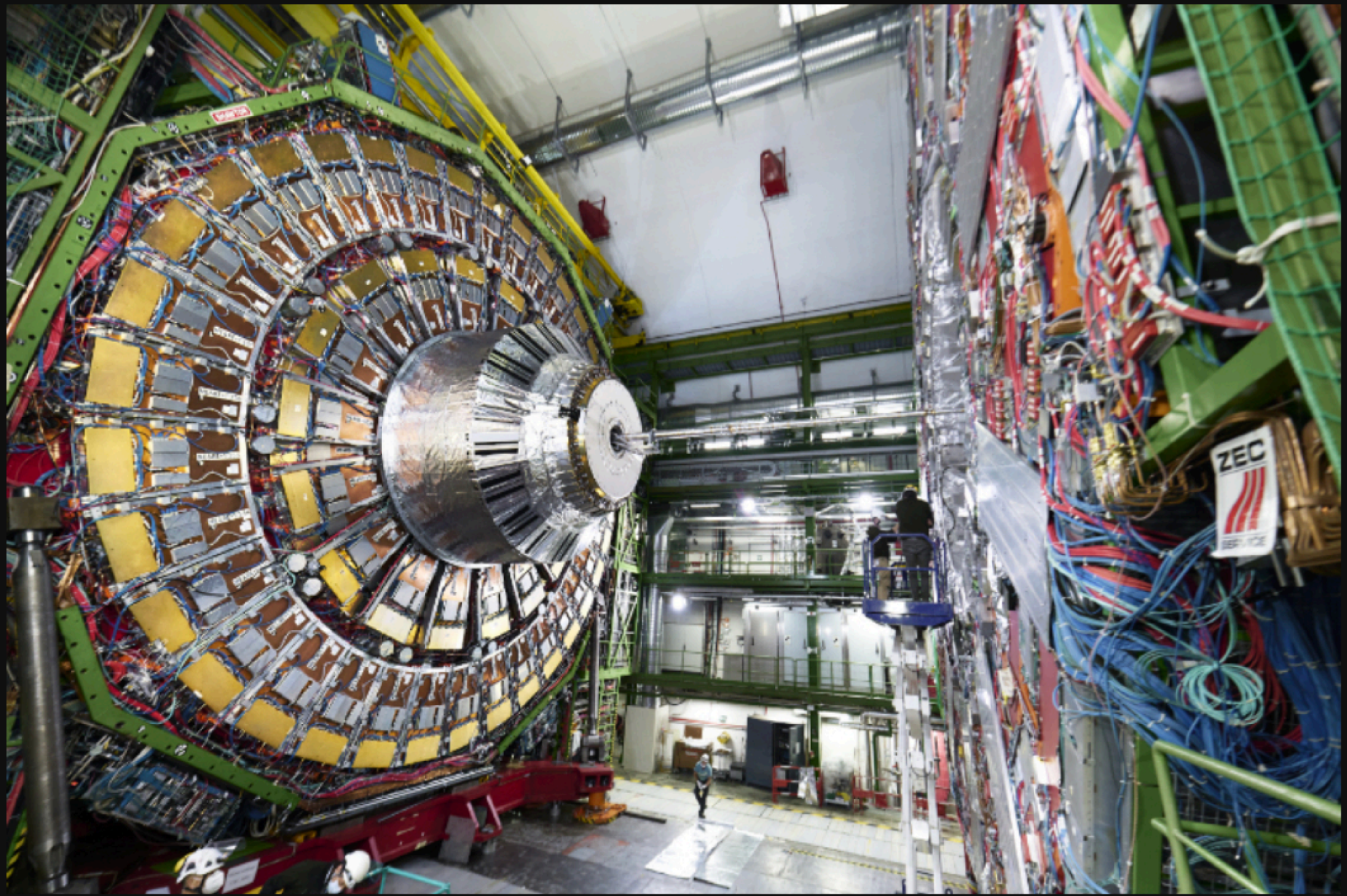
Piotr Bojanowski, Senior Research Scientist Manager, FAIR, Meta

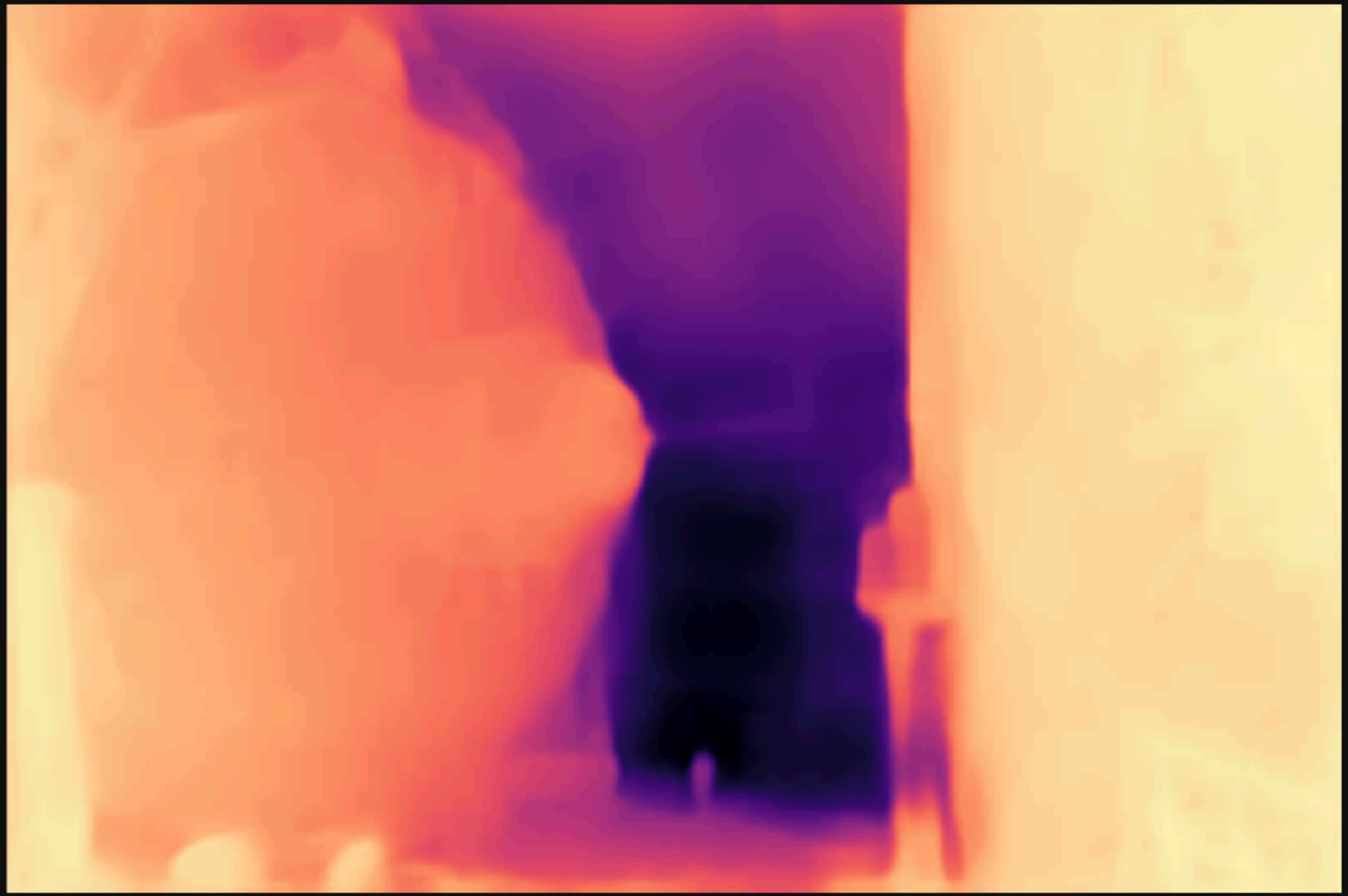




Depth Estimation

DINOv2's frozen features can readily be used in models predicting per-pixel depth from a single image, both in and out-of-distribution.





Try another image

Click to switch to parallax view, then move around the scene

HTTPS://DINOV2.METADEMOLAB.COM/

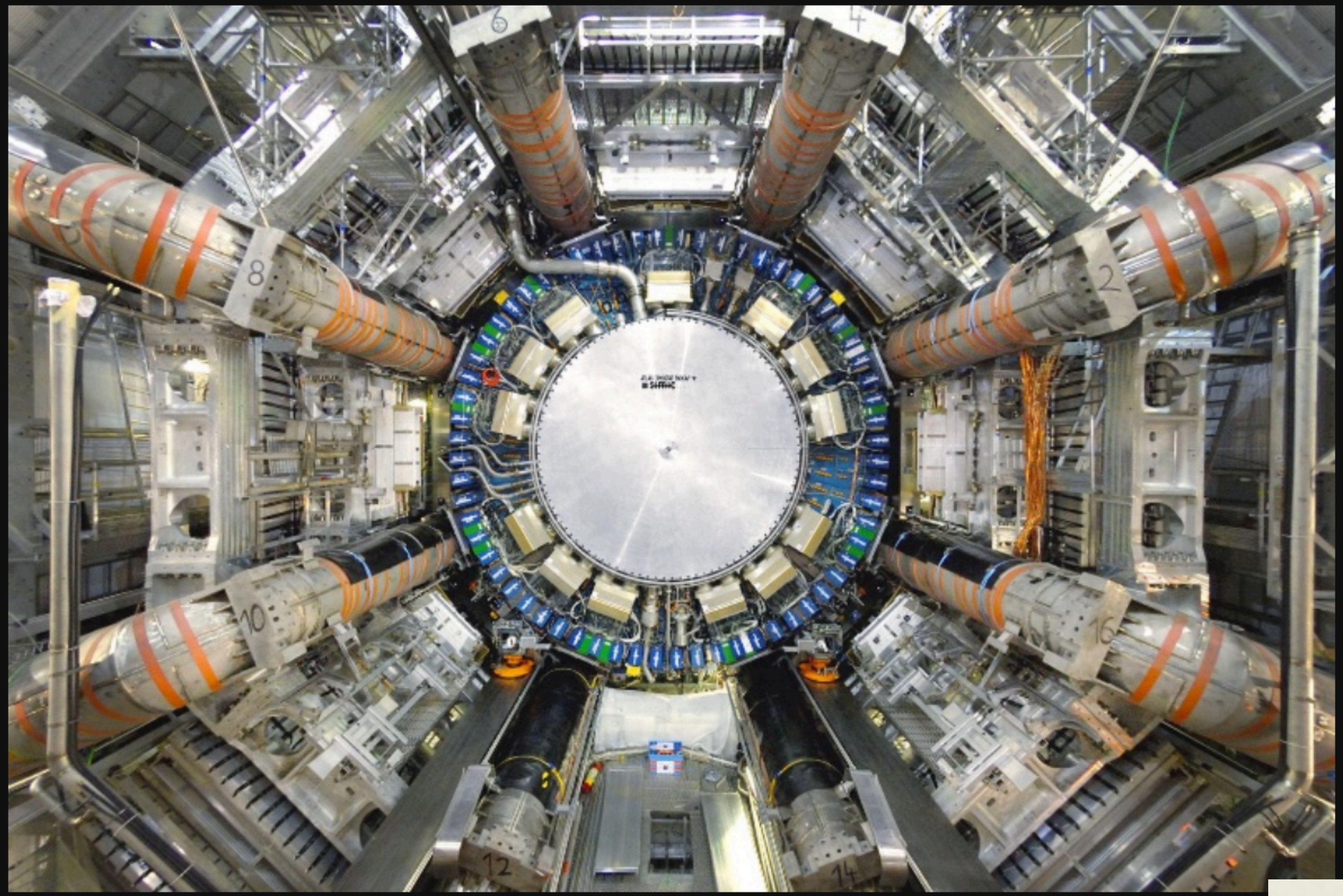
Instance Retrieval

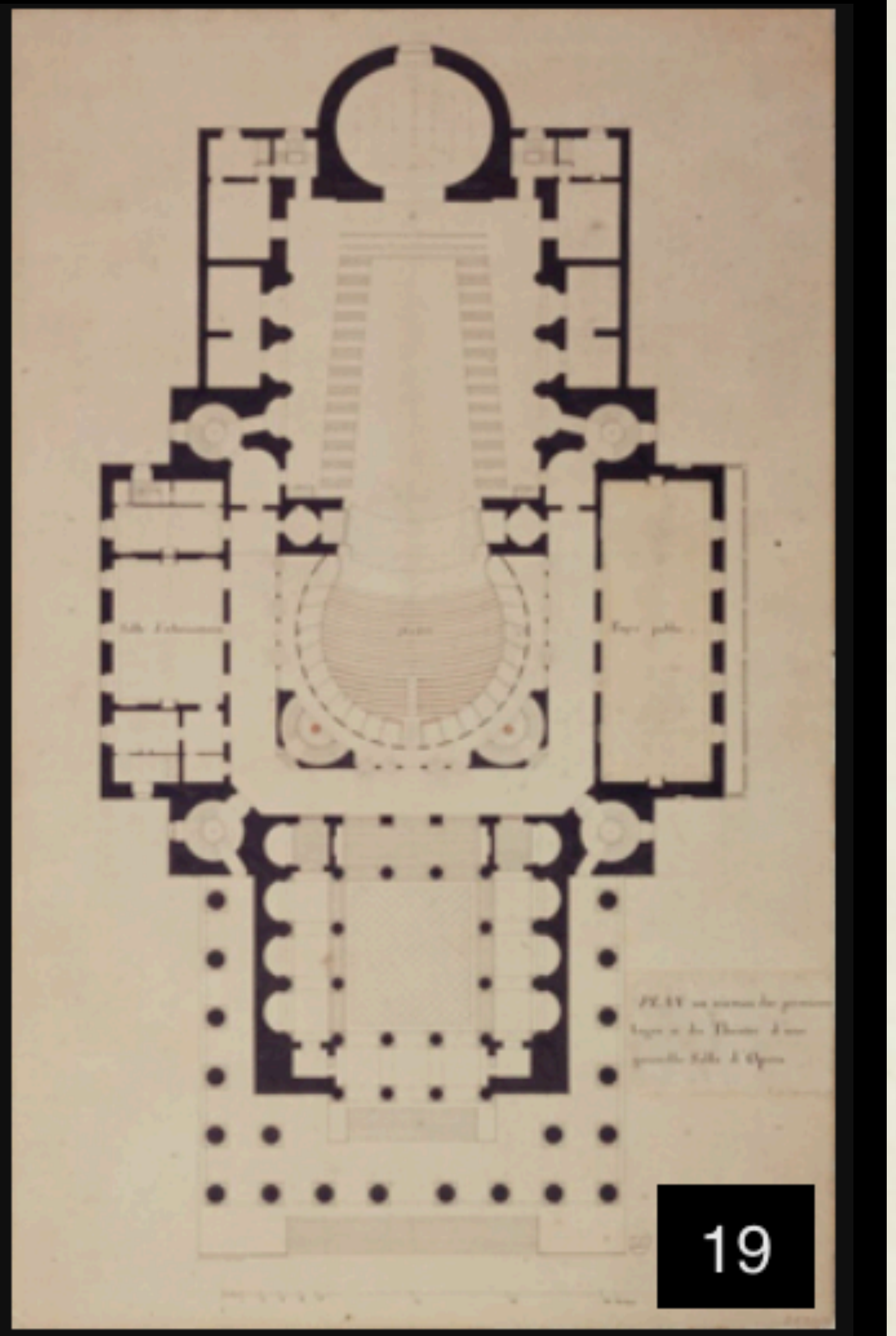
Find art pieces similar to a given image from a large collection of art images.

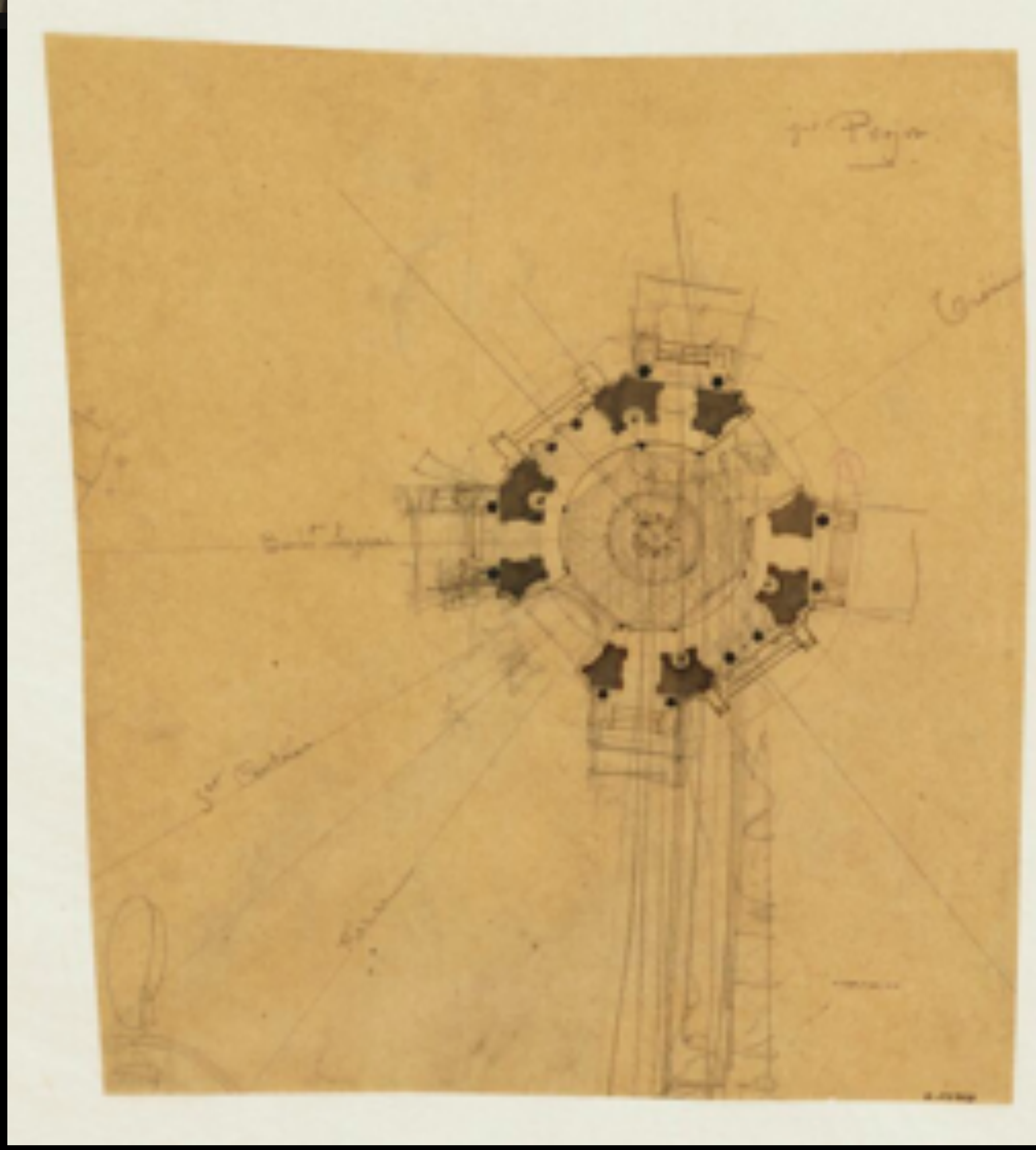
DINOv2's frozen features can readily be used to retrieve images similar to a query image using a non-parametric approach: database images are simply ranked according to the similarity of their features with those of the query image.

See results

Try another image







HTTPS://DINOV2.METADEMOLAB.COM/

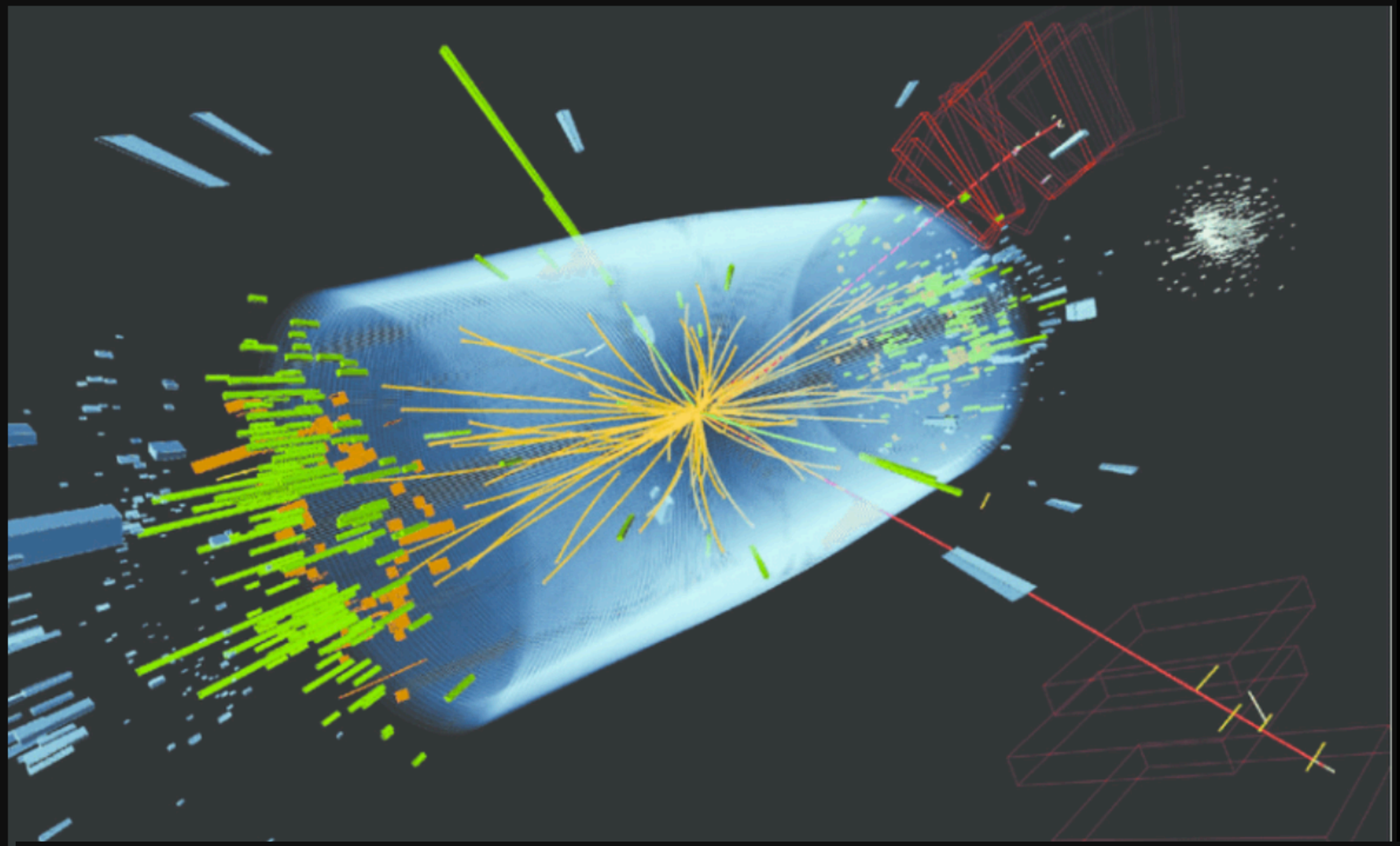
Instance Retrieval

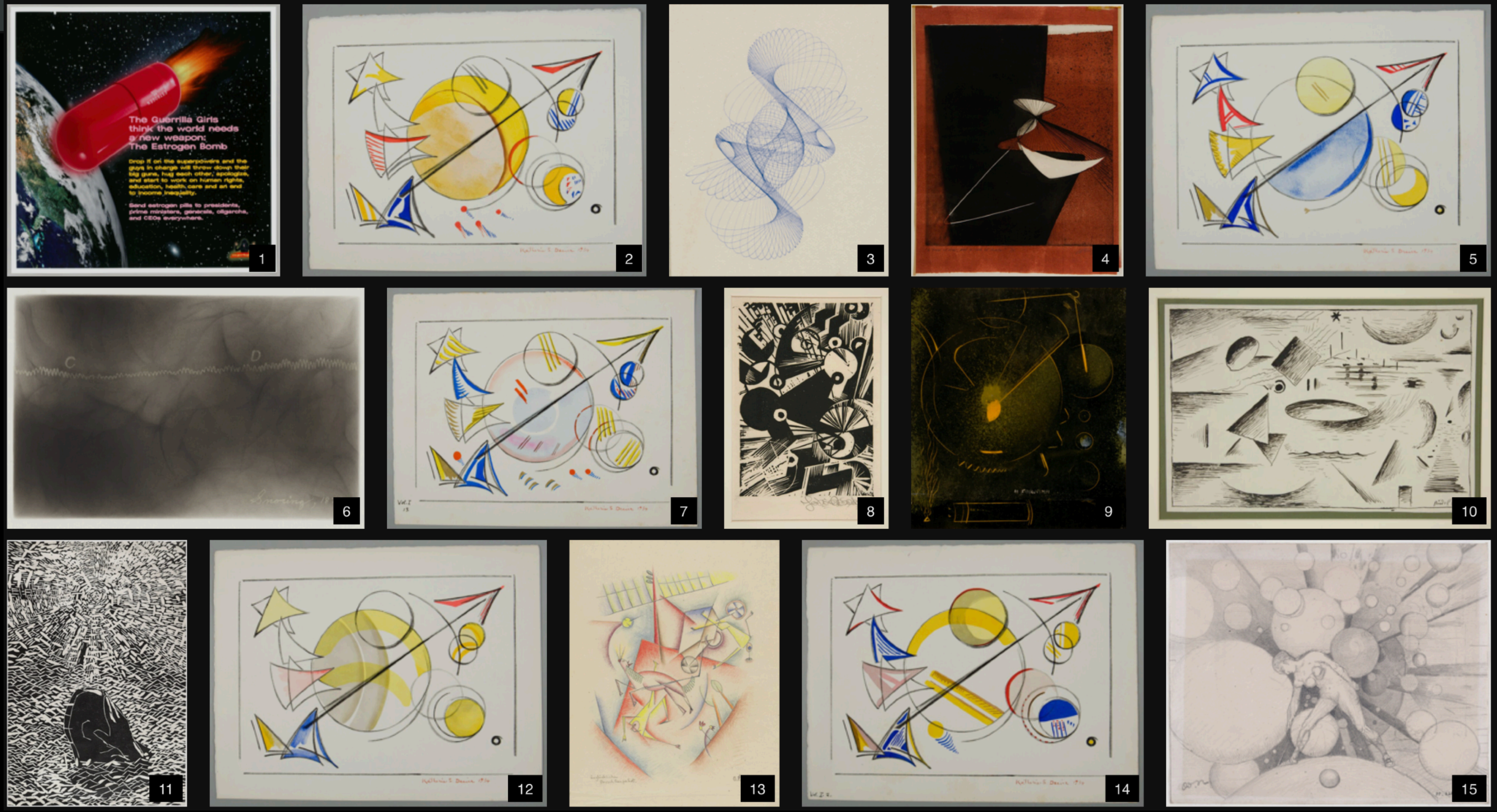
Find art pieces similar to a given image from a large collection of art images.

DINOv2's frozen features can readily be used to retrieve images similar to a query image using a non-parametric approach: database images are simply ranked according to the similarity of their features with those of the query image.

See results

Try another image

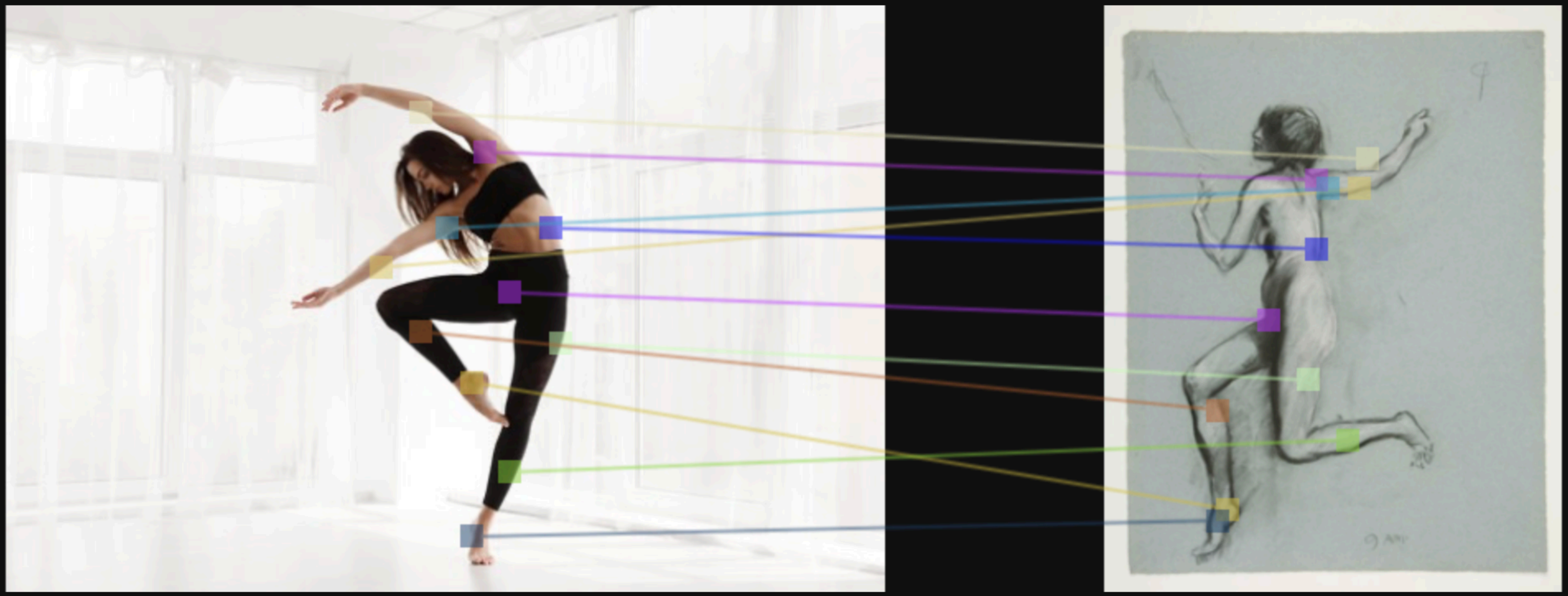


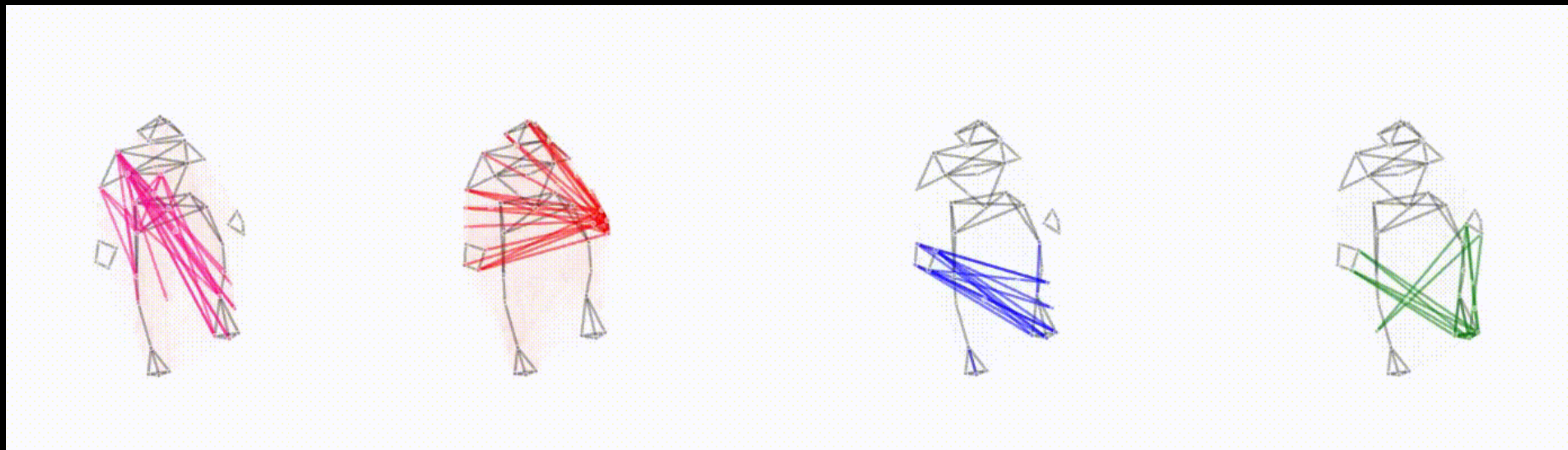


HTTPS://DINOV2.METADEMOLAB.COM/

Sparse Matching

DINOv2's frozen features are relevant to recognize the main objects in an image and to consistently encode similar parts across images. Here we match most similar patches across two images.





HTTPS://DINOV2.METADEMOLAB.COM/

Learning Image Representations Without Manual Annotations and Related Applications

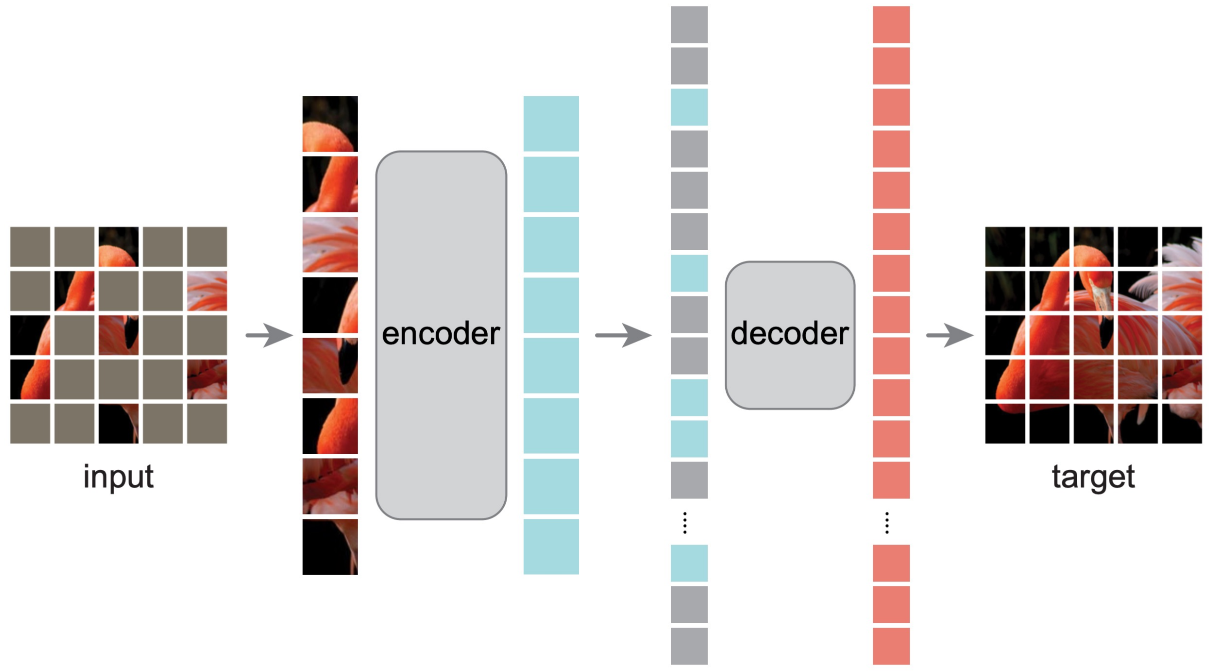
Piotr Bojanowski, Senior Research Scientist Manager, FAIR, Meta

History of Self-Supervised Learning

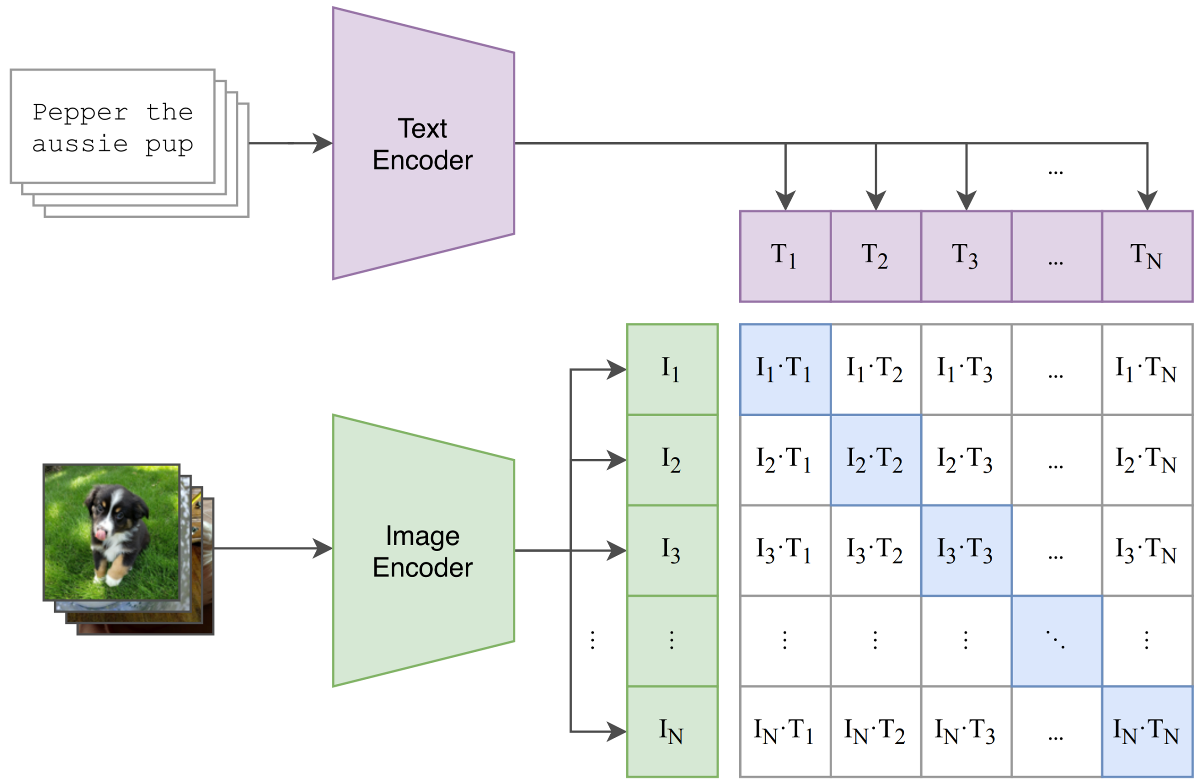
CPCv2, SELA, MoCo, PIRL, SimCLR, MoCov2, PCL, BYOL, Barlow Twins, SimCLRv2, NN-CLR, VicReg...

Meta AI

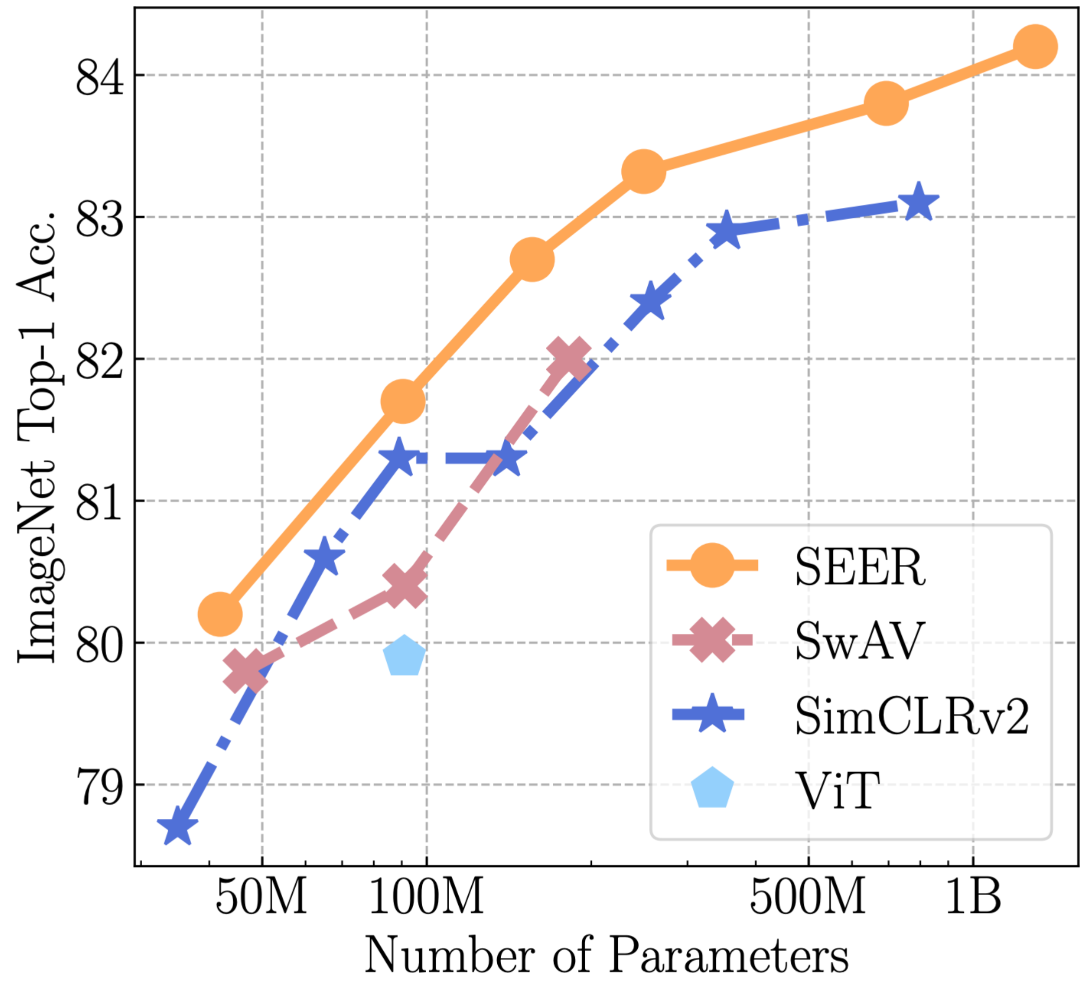
Motivations for DINOv2



He, Kaiming, et al. "Masked autoencoders are scalable vision learners." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.



Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International conference on machine learning*. PMLR, 2021.



Goyal, Priya, et al. "Self-supervised pretraining of visual features in the wild." *arXiv preprint arXiv:2103.01988* (2021).

Was the modeling effort worth it?

The Bitter Lesson

Rich Sutton

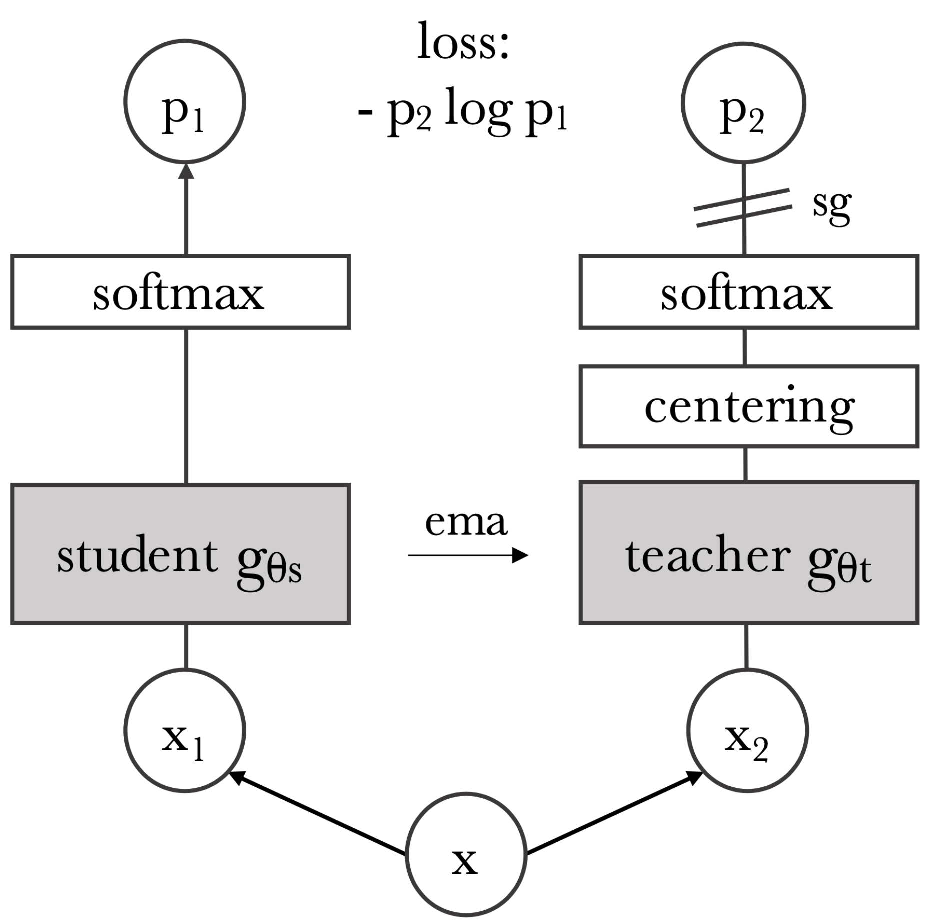
http://www.incompleteideas.net/IncIdeas/BitterLesson.html

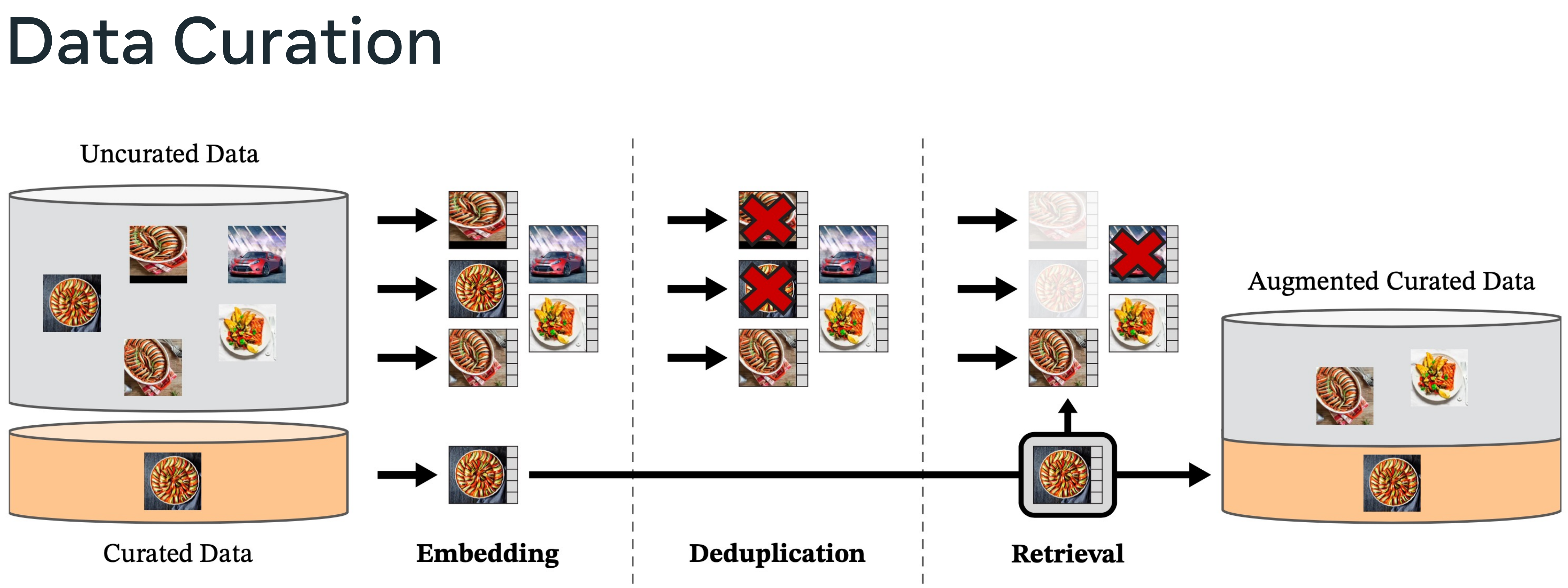
March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.

Effective data augmentation and curation is effectively a form of inductive bias

- Not on the model architecture, but part of the overarching ML strategy



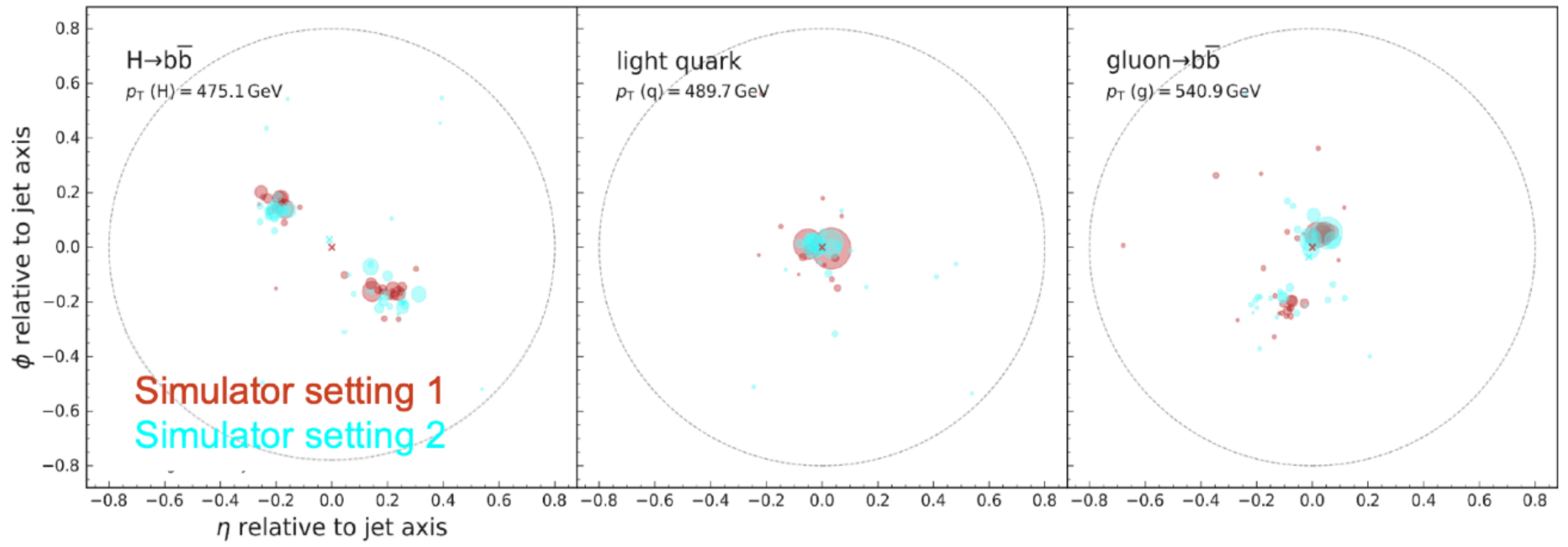


Self-supervised learning of jets using a realistic detector simulation

Kyle Cranmer, Etienne Dreyer, Eilam Gross, Nilotpal Kakati, Dmitrii Kobilianskii, Patrick Rieck, Nathalie Soybelman

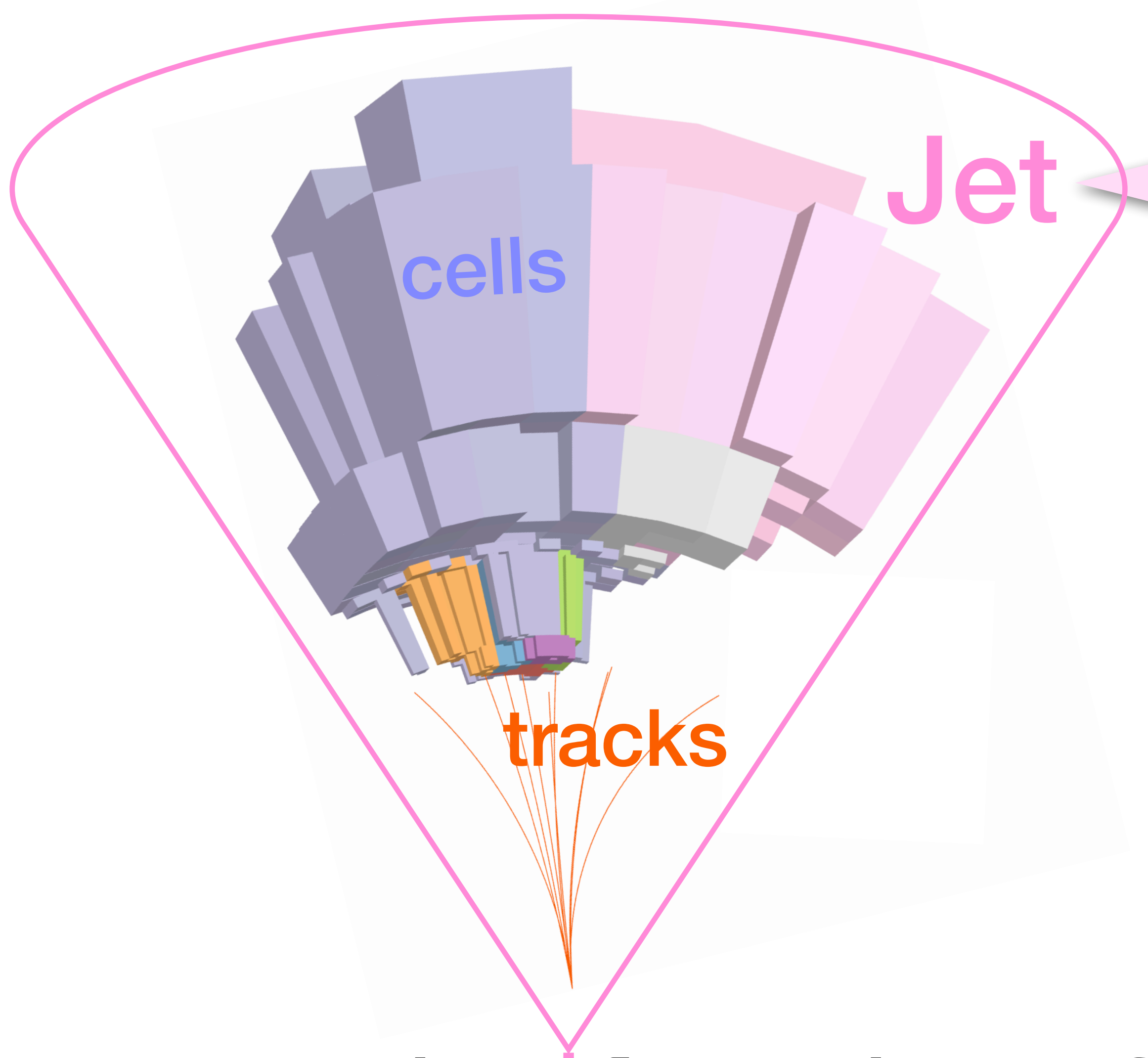
Hammers & Nails 2023
Ascona, Switzerland

Start from fixed interaction, re-run simulator downstream



Different simulator settings represent theory uncertainties and variation due to stochastic nature of underlying physics

Benedikt Maier, Hammers & Nails, Oct 30, 2023

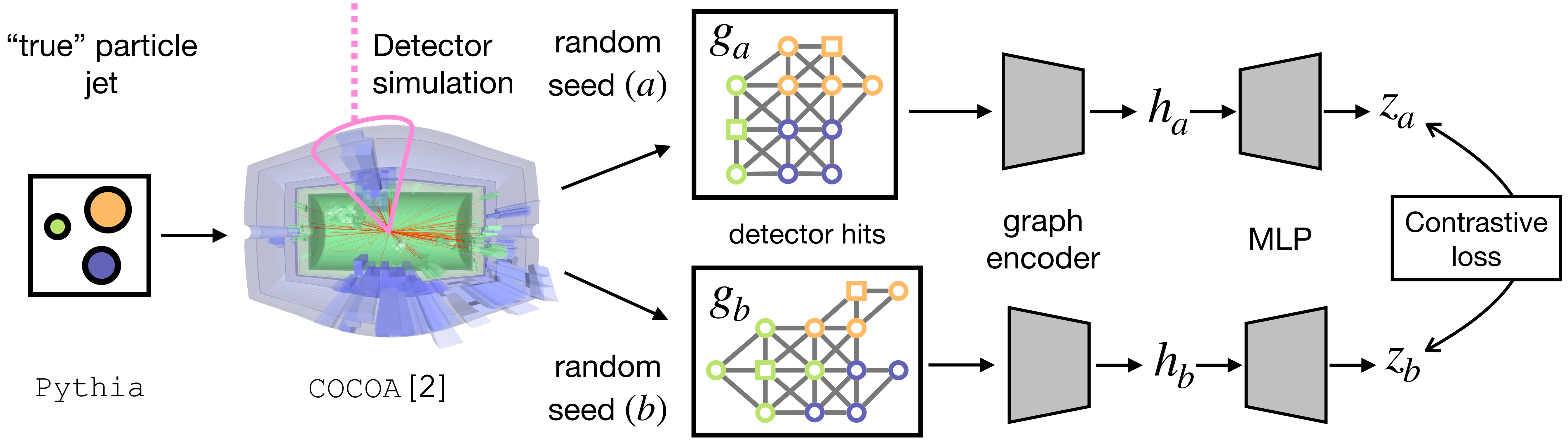


A correlated set of particles with several learnable properties:
(due to their composite nature)

- ✓ Process of origin (classification)
- ✓ Energy, mass (regression)
- ✓ Possible anomalous signatures

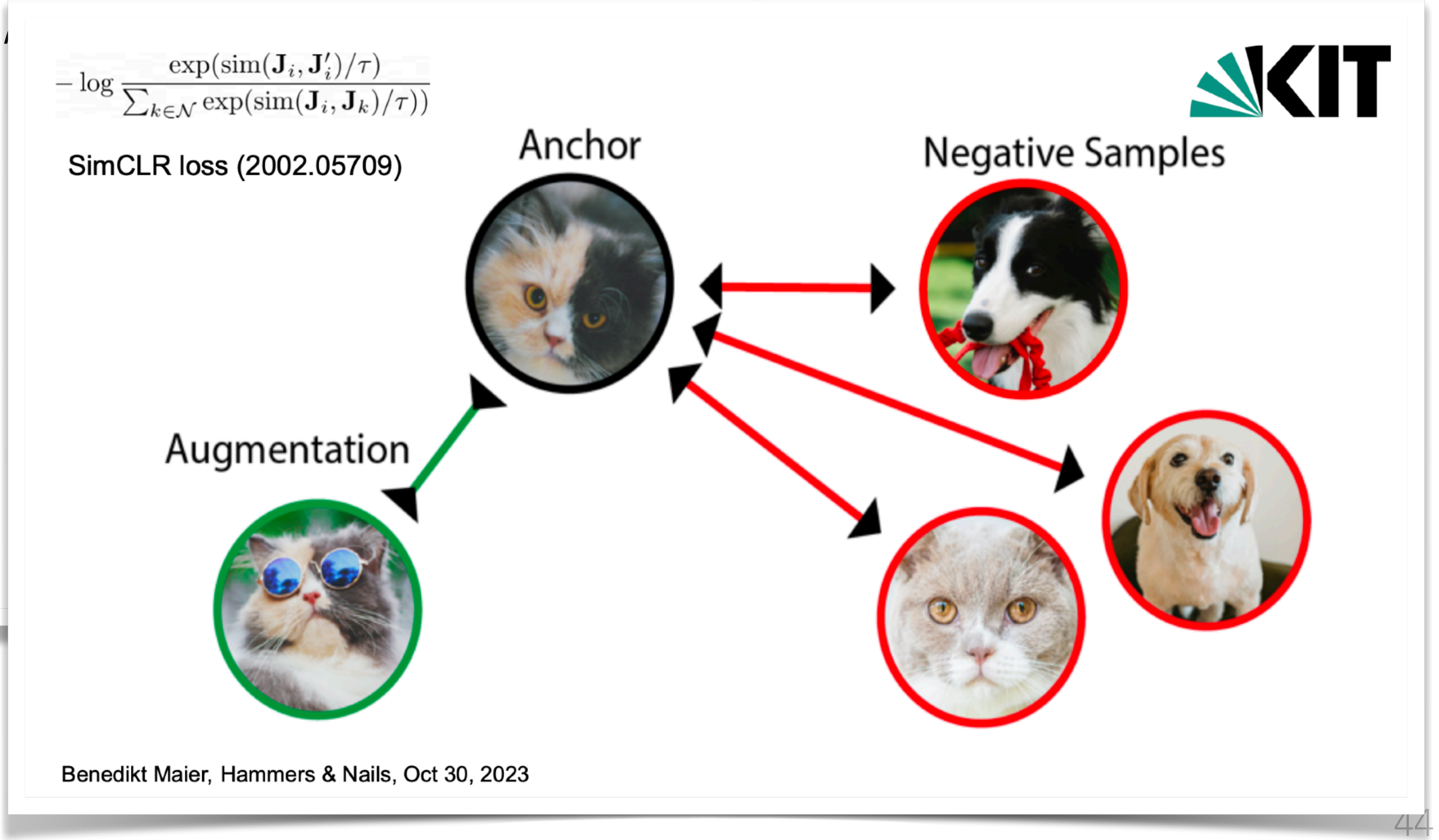
Note: similar effort by MIT/KIT/SLAC (see their talk at BOOST)

Contrastive learning of jets via detector replicas



[2] Configurable calorimeter simulation for AI applications A. Charkin-Gorbulin et al, Mach. Learn.: Sci. Tech. (2023)

What is the same and what's different?



Leveraging physics knowledge / Inductive bias

Inductive Bias

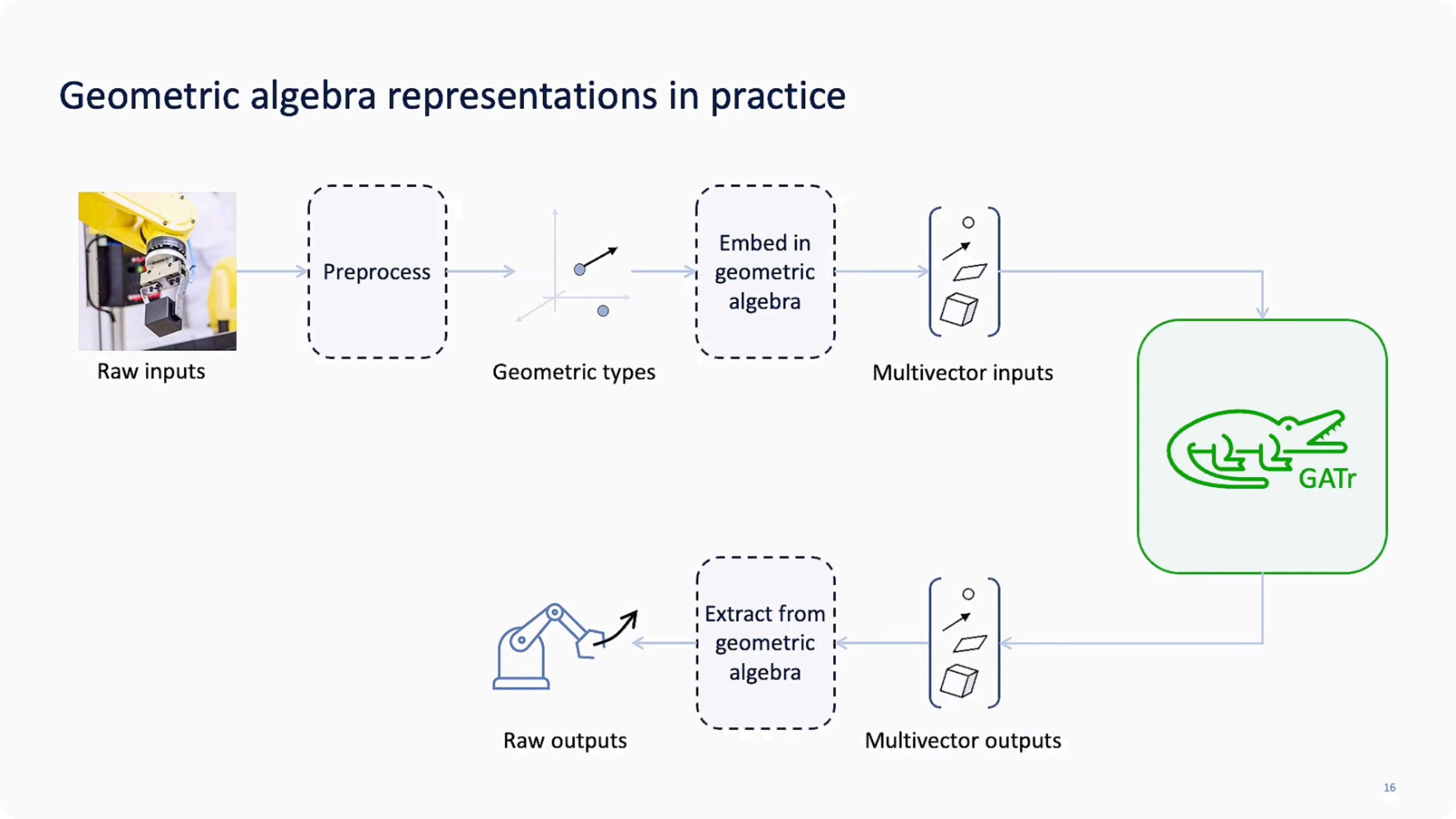
Compositionality

Relationships

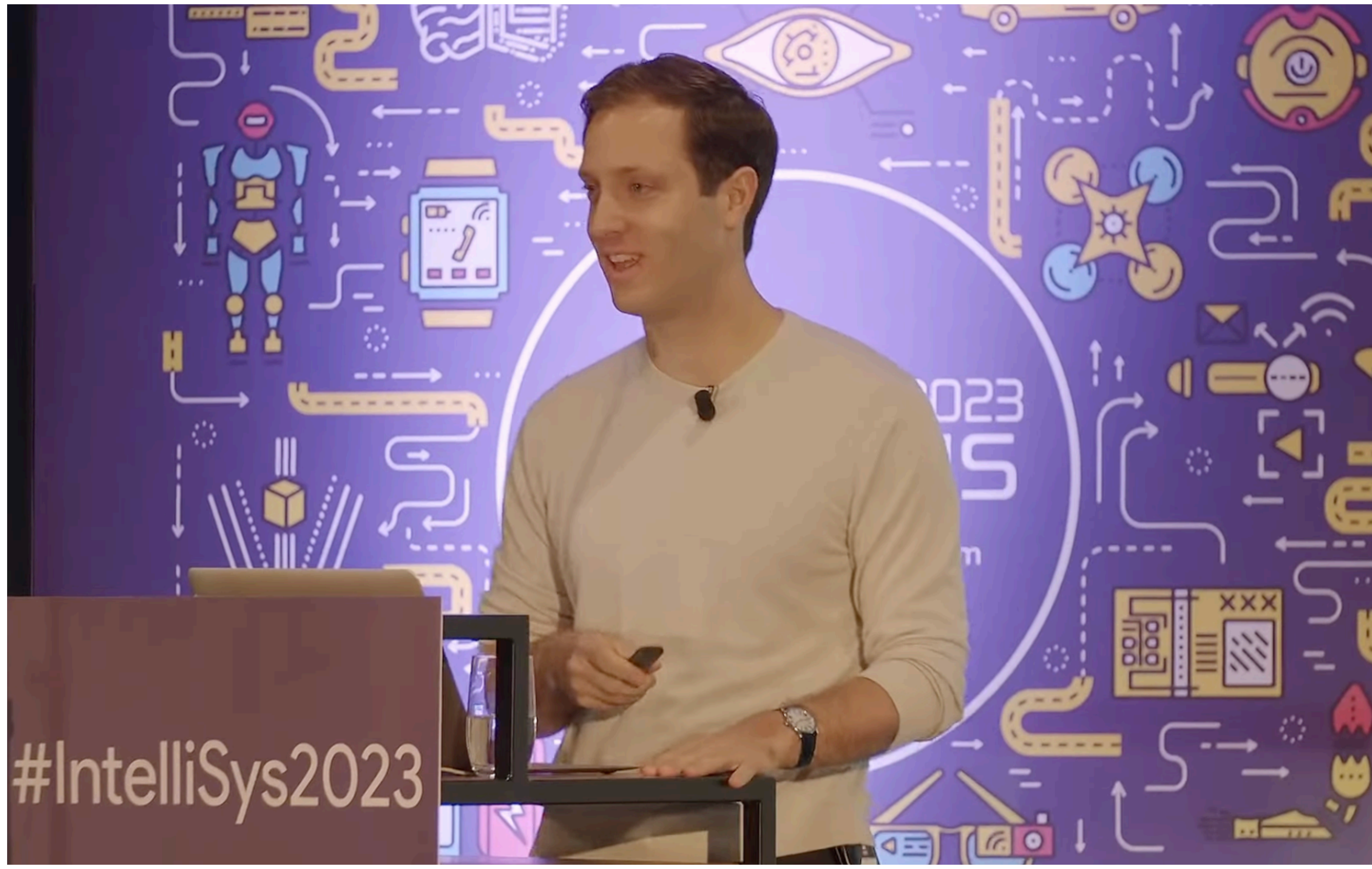
Symmetry

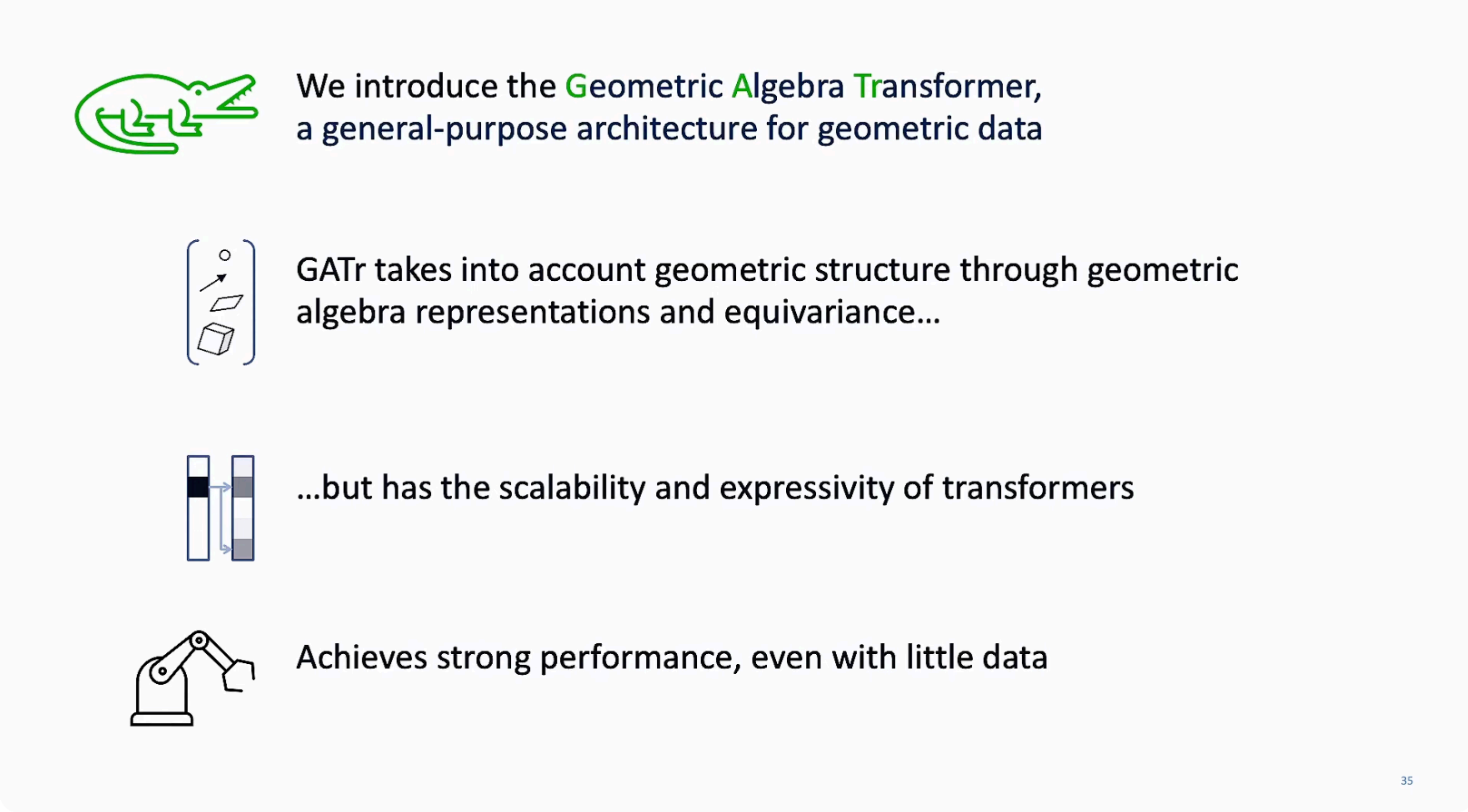
Causality

Separation



Geometric Algebra Transformers: Revolutionizing Geometric Data with Taco Cohen, Qualcomm AI Research

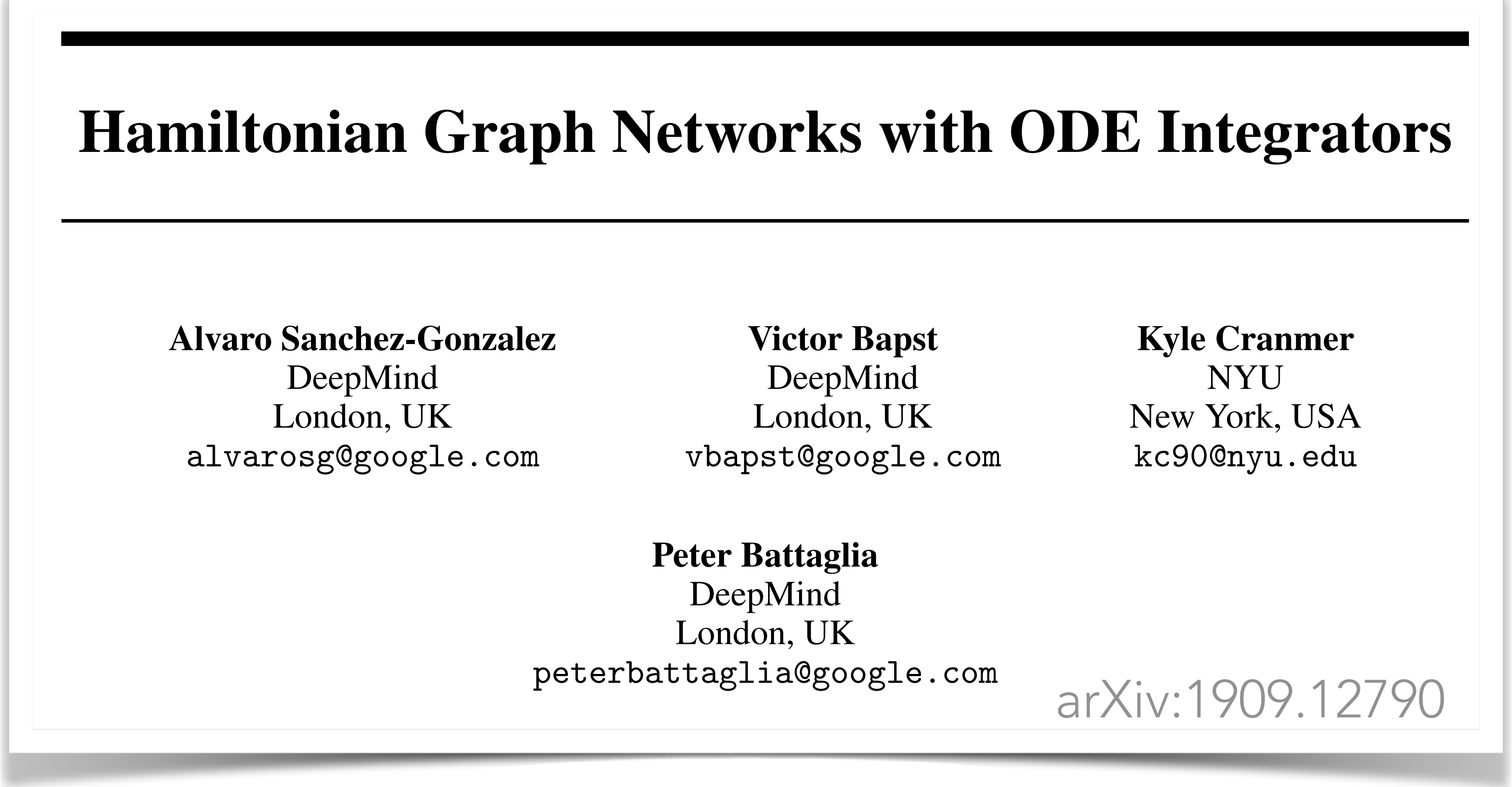




Geometric Algebra Transformers: Revolutionizing Geometric Data with Taco Cohen, Qualcomm AI Research

Hamiltonian graph Networks

Alvaro Sanchez Gonzalez



Hamiltonian Graph Networks with ODE Integrators

Alvaro Sanchez-Gonzalez
DeepMind
London, UK
alvarosg@google.com

Victor Bapst
DeepMind
London, UK
vbapst@google.com

Kyle Cranmer
NYU
New York, USA
kc90@nyu.edu

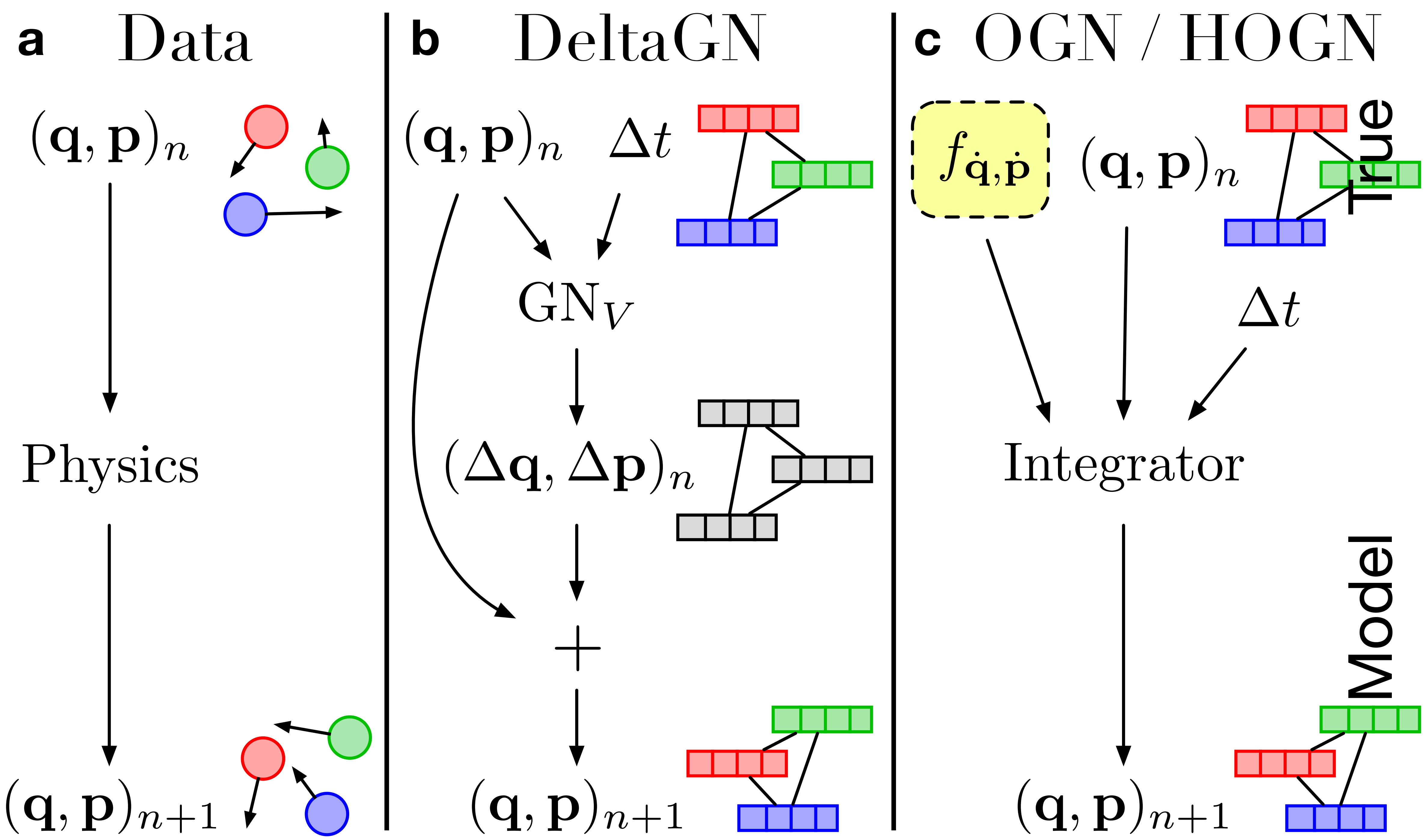
Peter Battaglia
DeepMind
London, UK
peterbattaglia@google.com

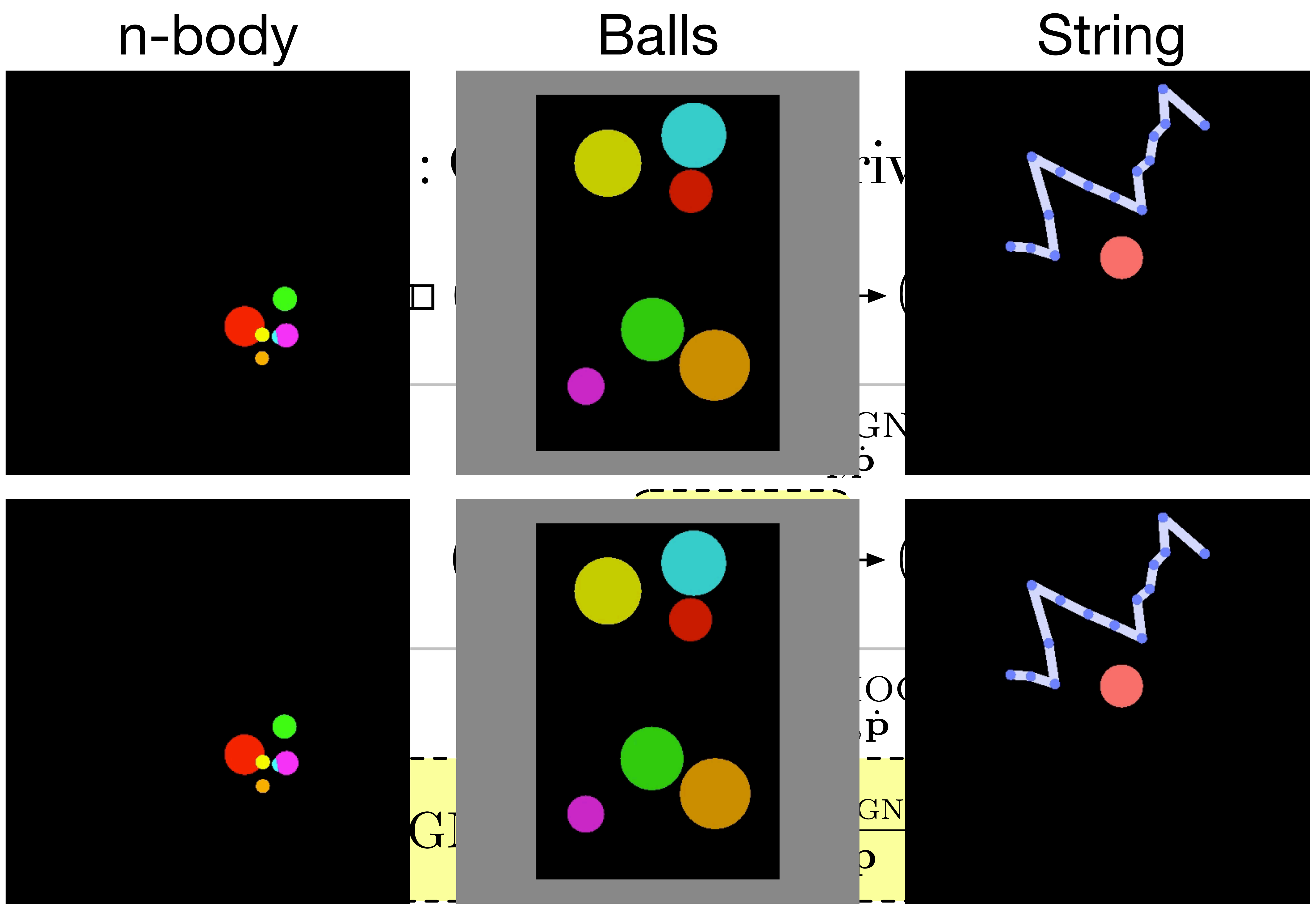
arXiv:1909.12790

We incorporated two physically-informed inductive biases

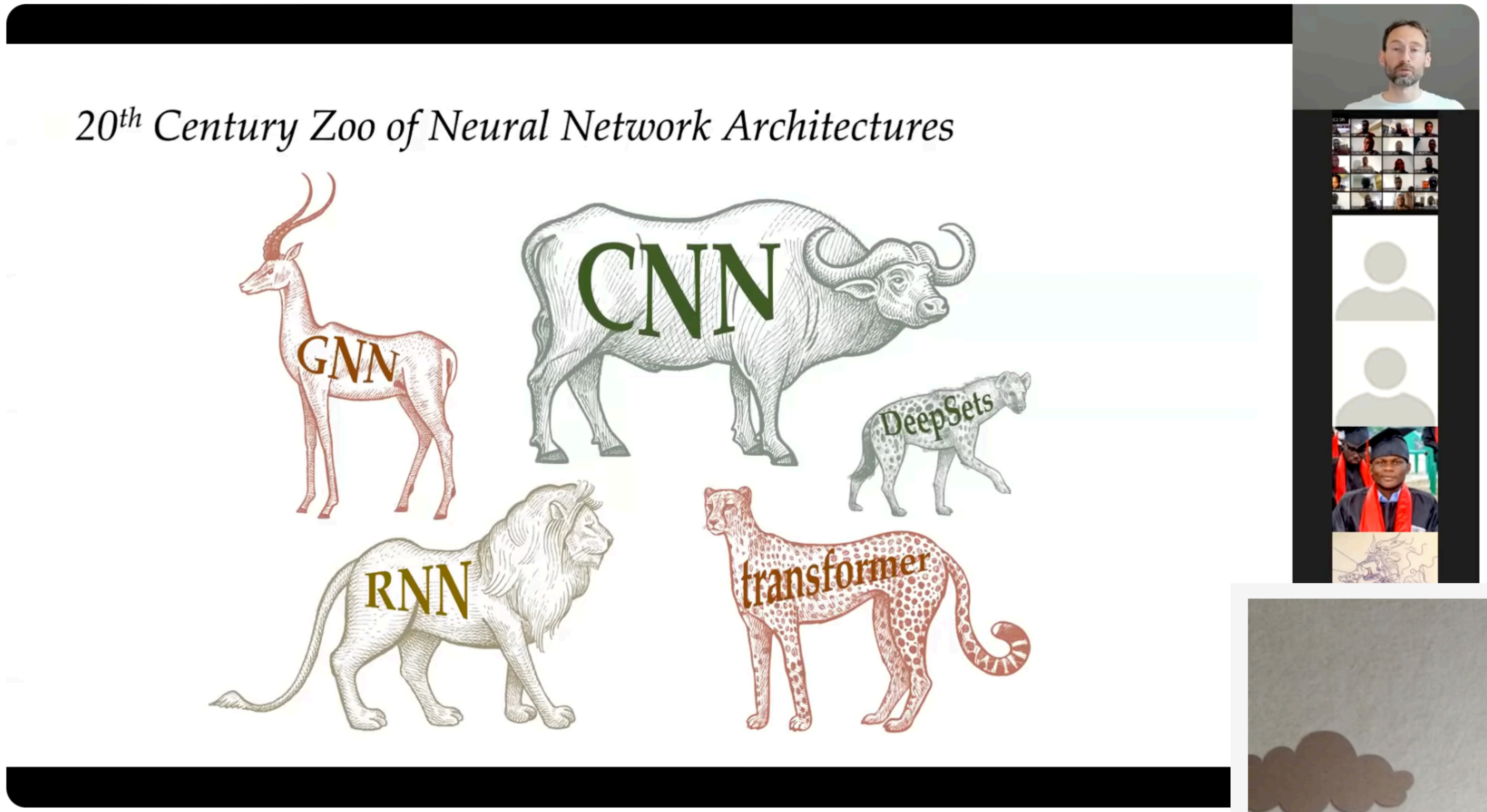
- ODE integrators
- Hamiltonian mechanics

COMMENT HERE IS THAT CAN POOL DOWN TO A SCALAR

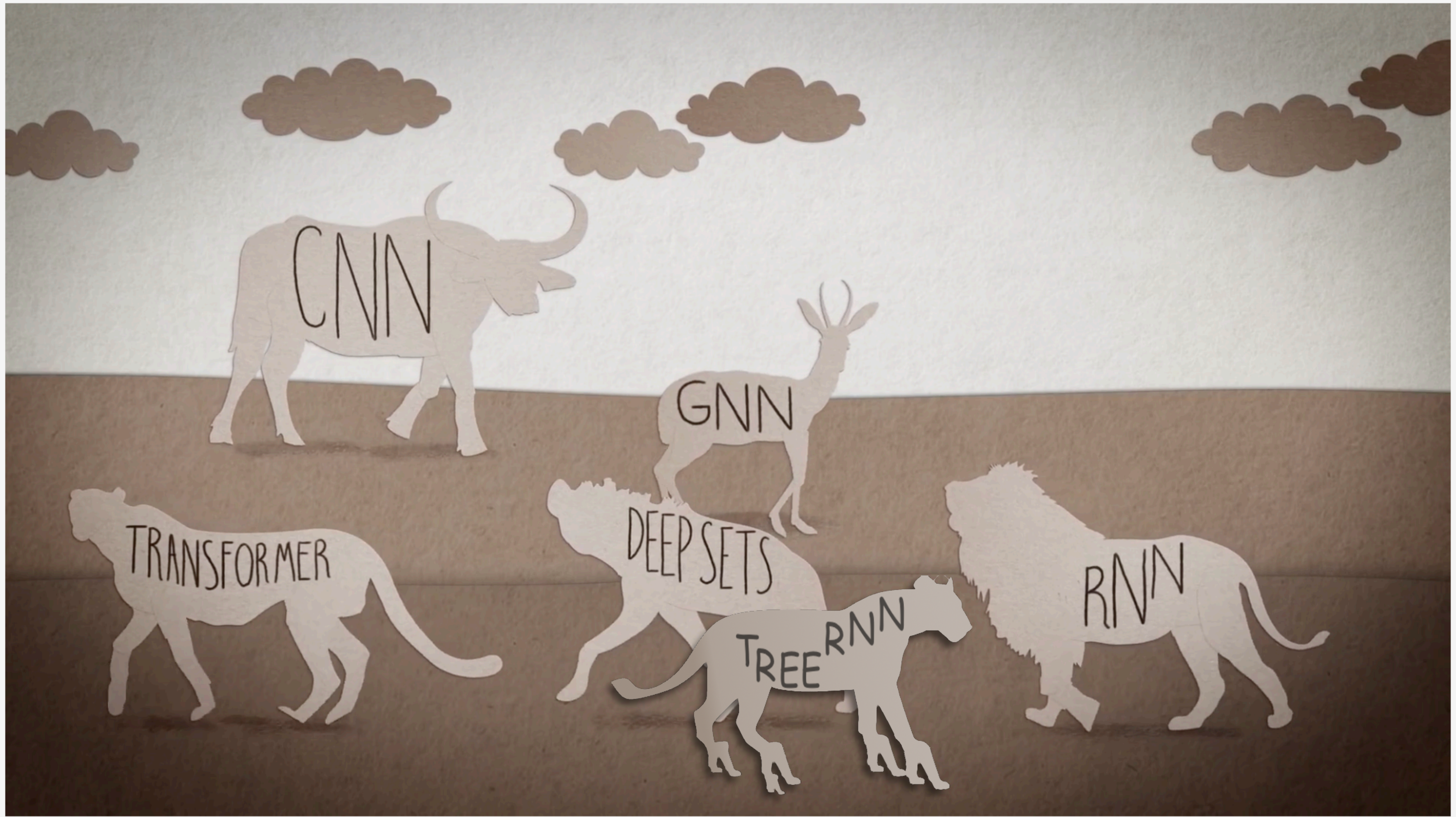




Don't forget TreeRNNs

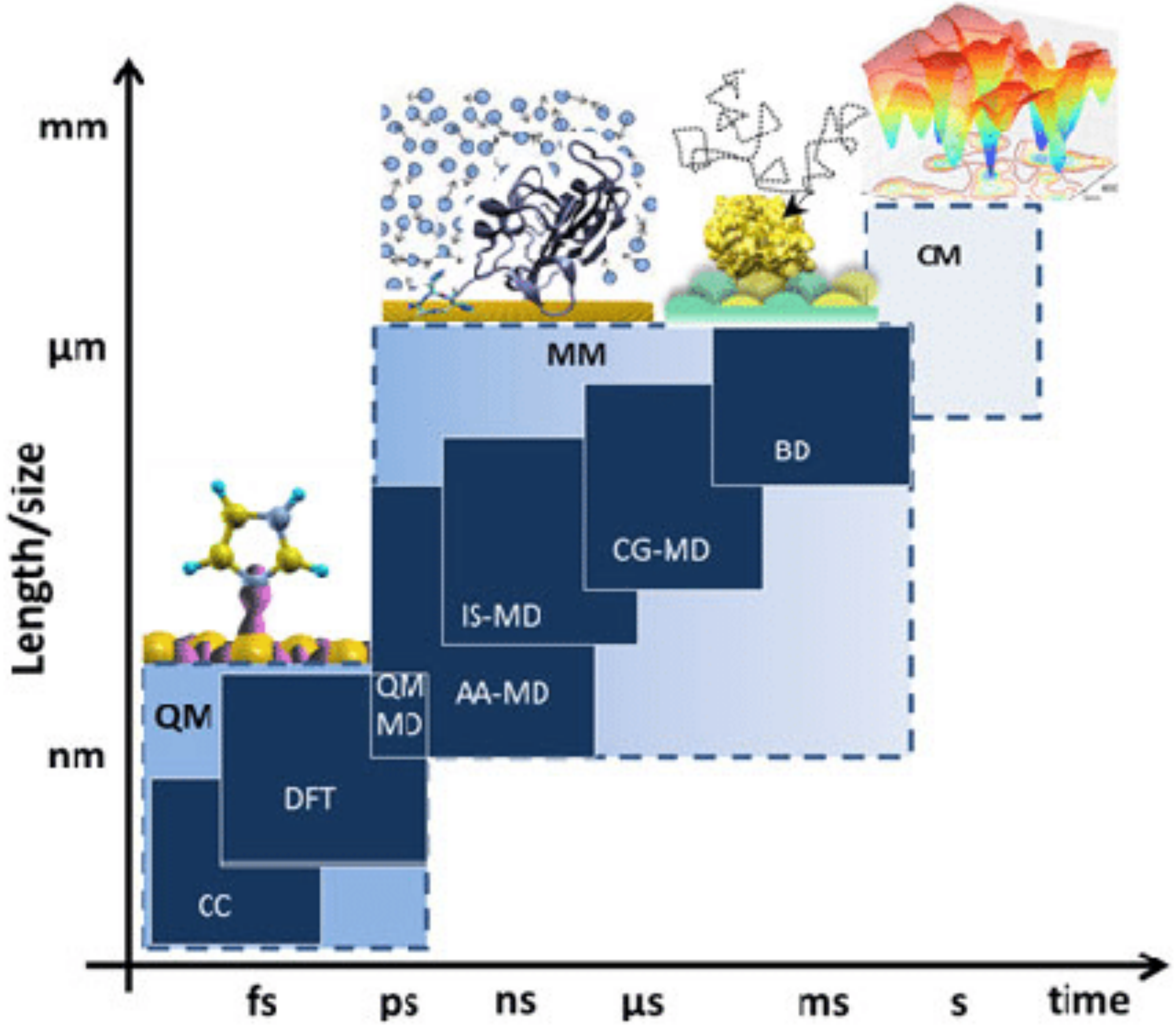


AMMI Course "Geometric Deep Learning" - Lecture 1 (Introduction) - Michael Bronstein



Bigger picture

Time and length scales of different simulation techniques: quantum mechanics (QM), including coupled cluster (CC) and DFT methods, molecular mechanics (MM), and the Brownian dynamics (BD) technique; and continuum mechanics (CM). **Demand for surrogate multiscale modelling.**



Reduction and closure of dynamical systems using deep learning

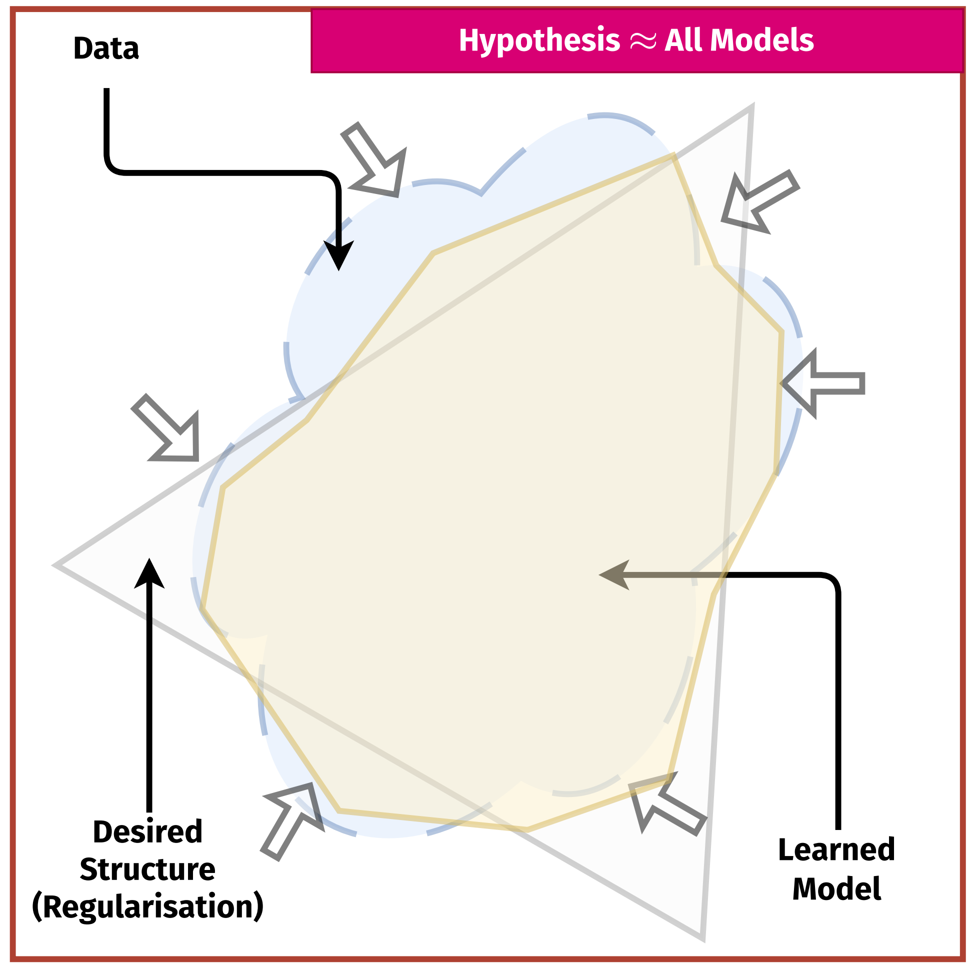
Qianxiao Li
Department of Mathematics
Institute for Functional Intelligent Materials
blog.nus.edu.sg/qianxiaoli

Hammer & Nails 2023, Swiss Edition
Congressi Stefano Franscini, Ascona, Switzerland
31 Oct 2023

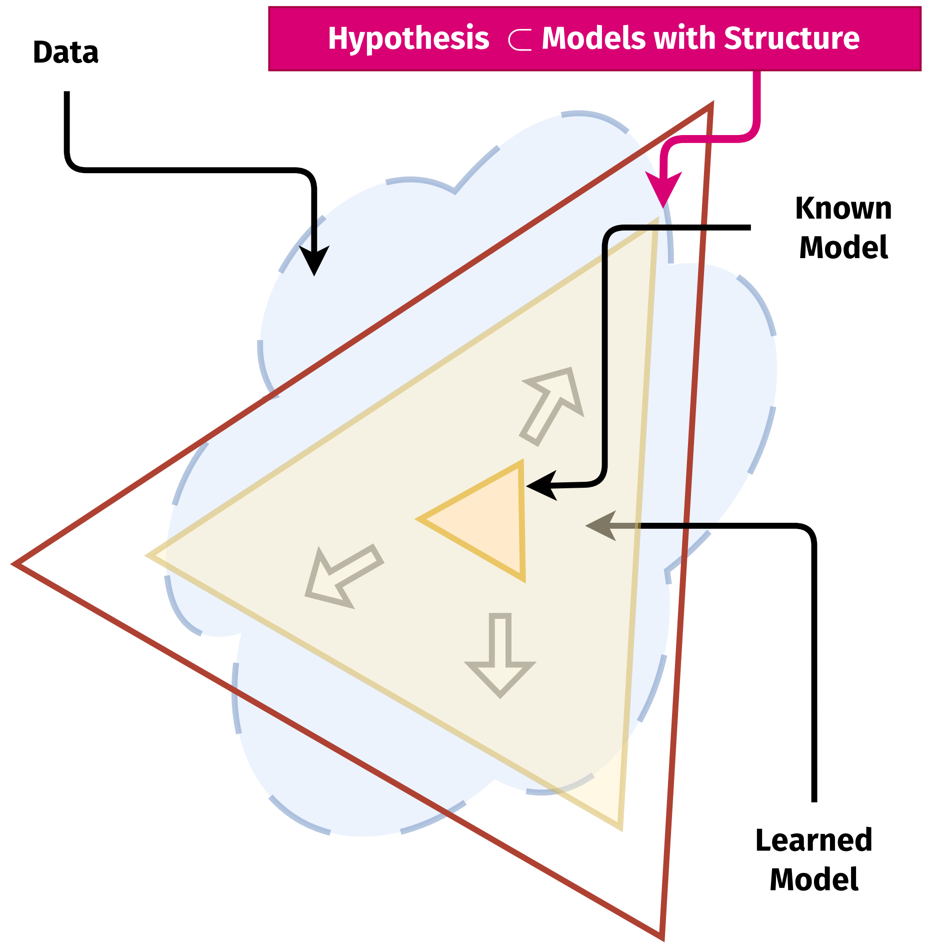
NUS National University of Singapore

Two viewpoints for learning dynamics from data

Data-based approach



Model-based approach



Invited speakers

09:00 **AI & material science**
Speaker: Kostya Novoselov

10:00 **Coffee break**

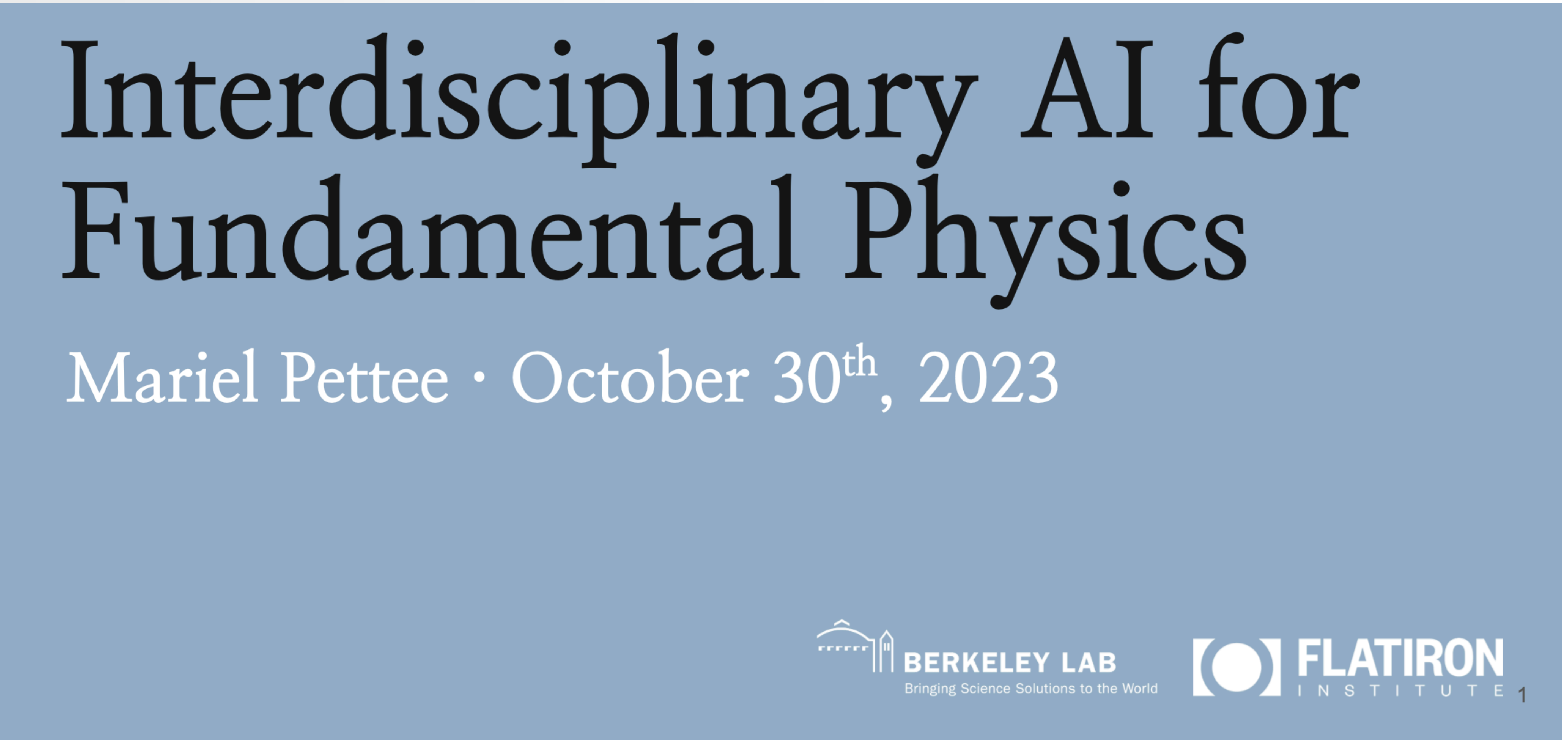
10:30 **Machine learning: bridging scale gap between the worlds of materials and particles**
Speaker: Andrey Ustyuzhanin
Machine Intelligence... Machine Intelligence...

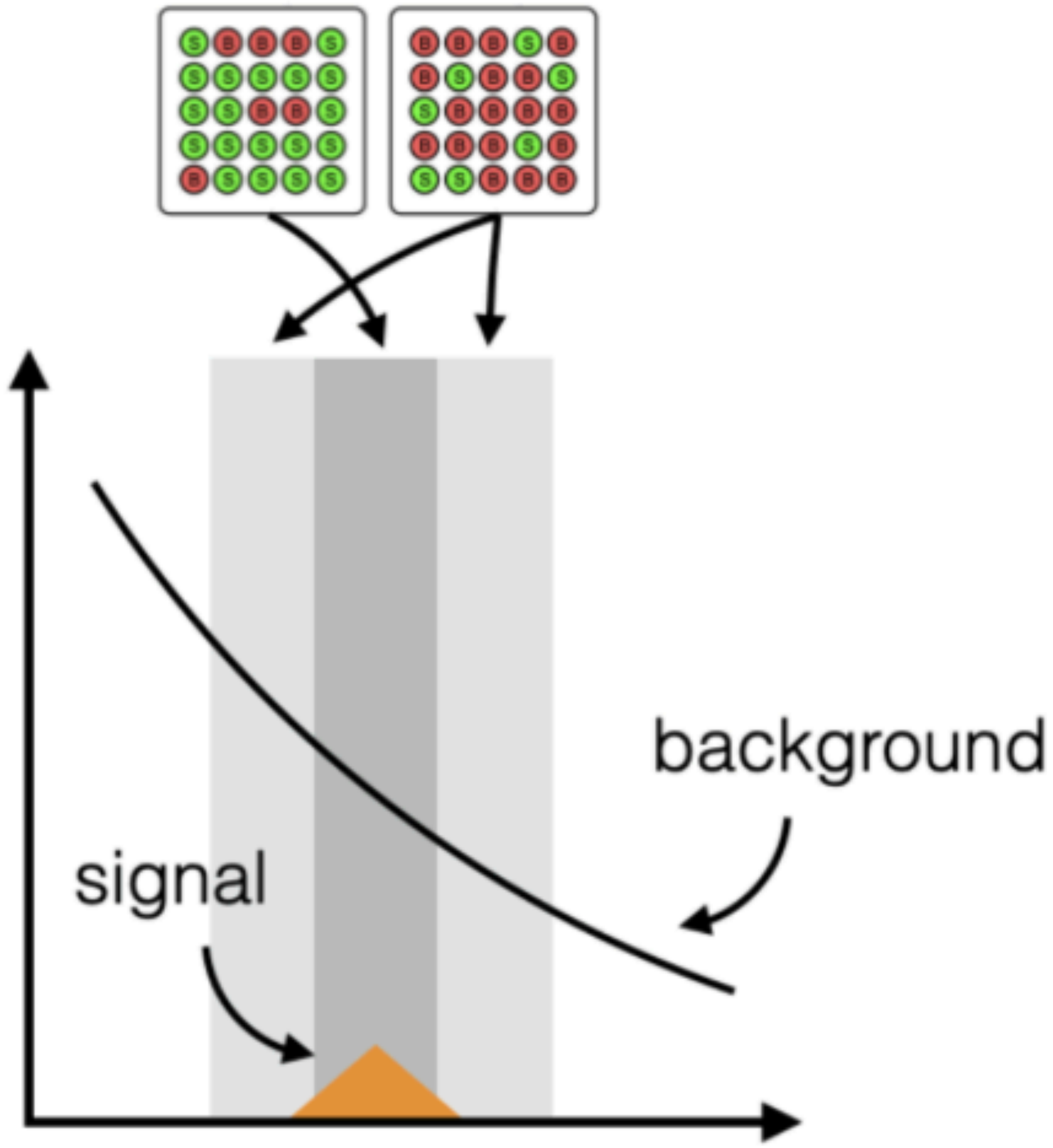
11:15 **Reduction and Closure of Dynamical Systems using Deep Learning**
Speaker: Qianxiao Li
slides_QianxiaoLi.pdf

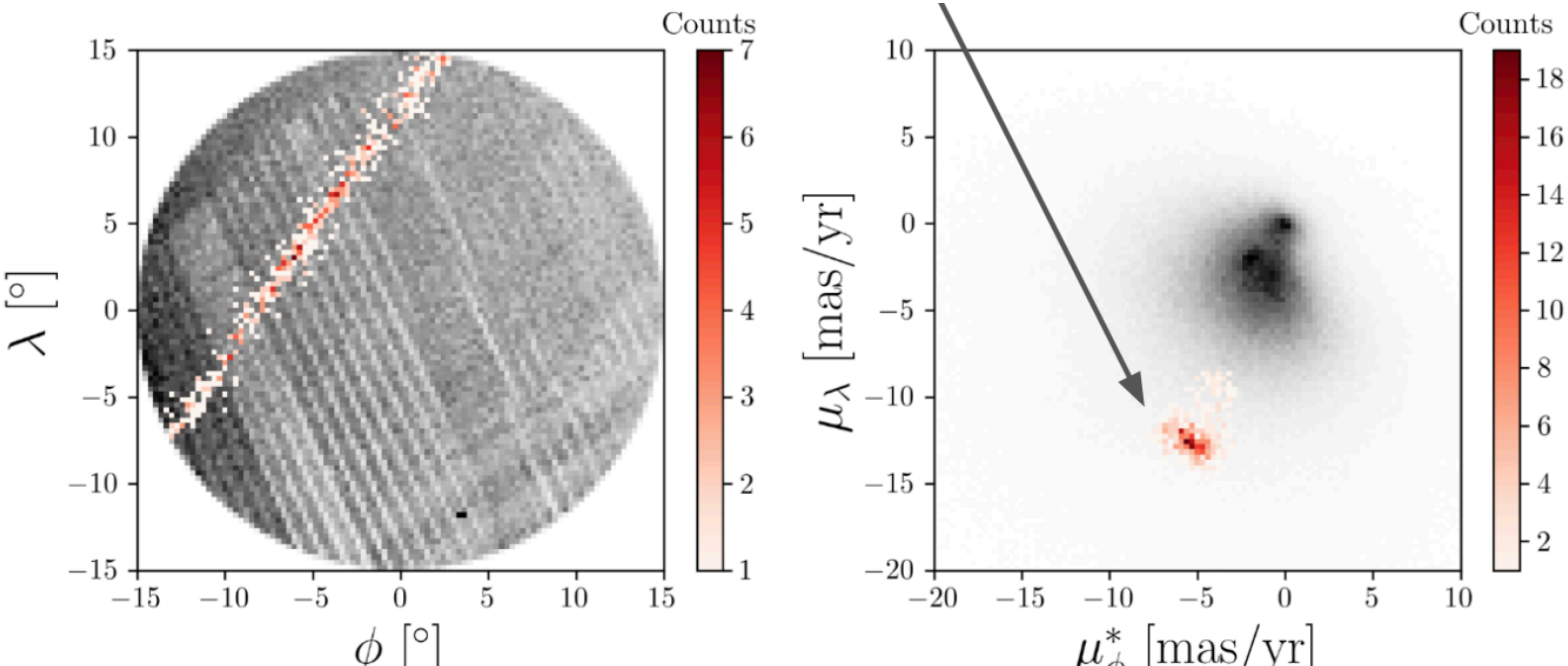
Lunch

Brainstorming - CANCELLED

Multidisciplinary research and collaborations & cross-pollination

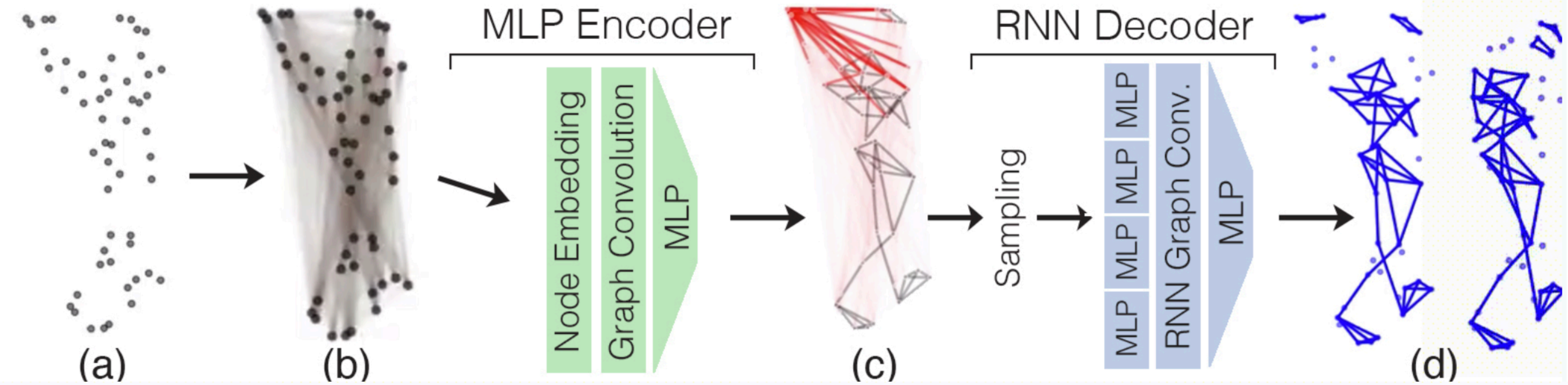

Interdisciplinary AI for Fundamental Physics
Mariel Pettee · October 30th, 2023
BERKELEY LAB
Bringing Science Solutions to the World
FLATIRON
INSTITUTE
1

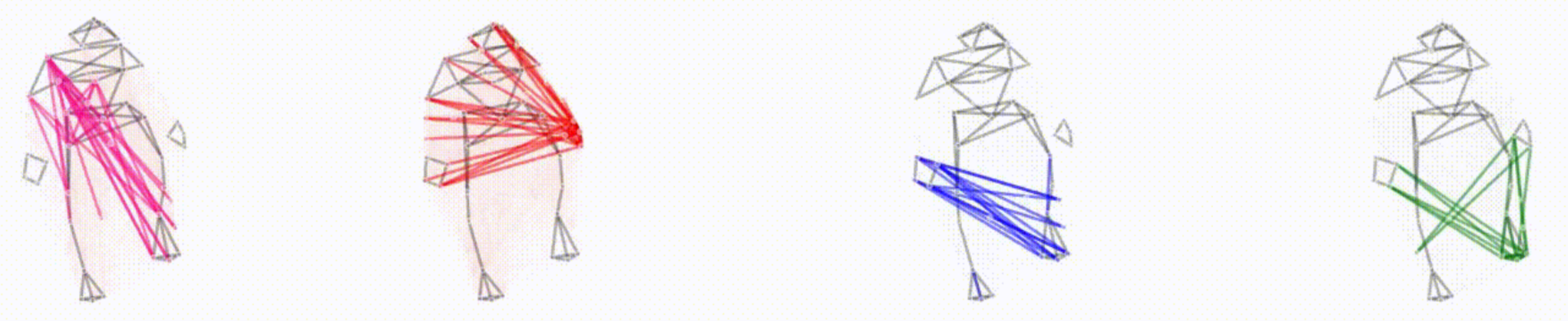

background
signal


Counts
λ [°]
ϕ [°]
Counts
μ_λ [mas/yr]
μ^*_ϕ [mas/yr]

Interdisciplinary AI for Fundamental Physics

Mariel Pettee · October 30th, 2023

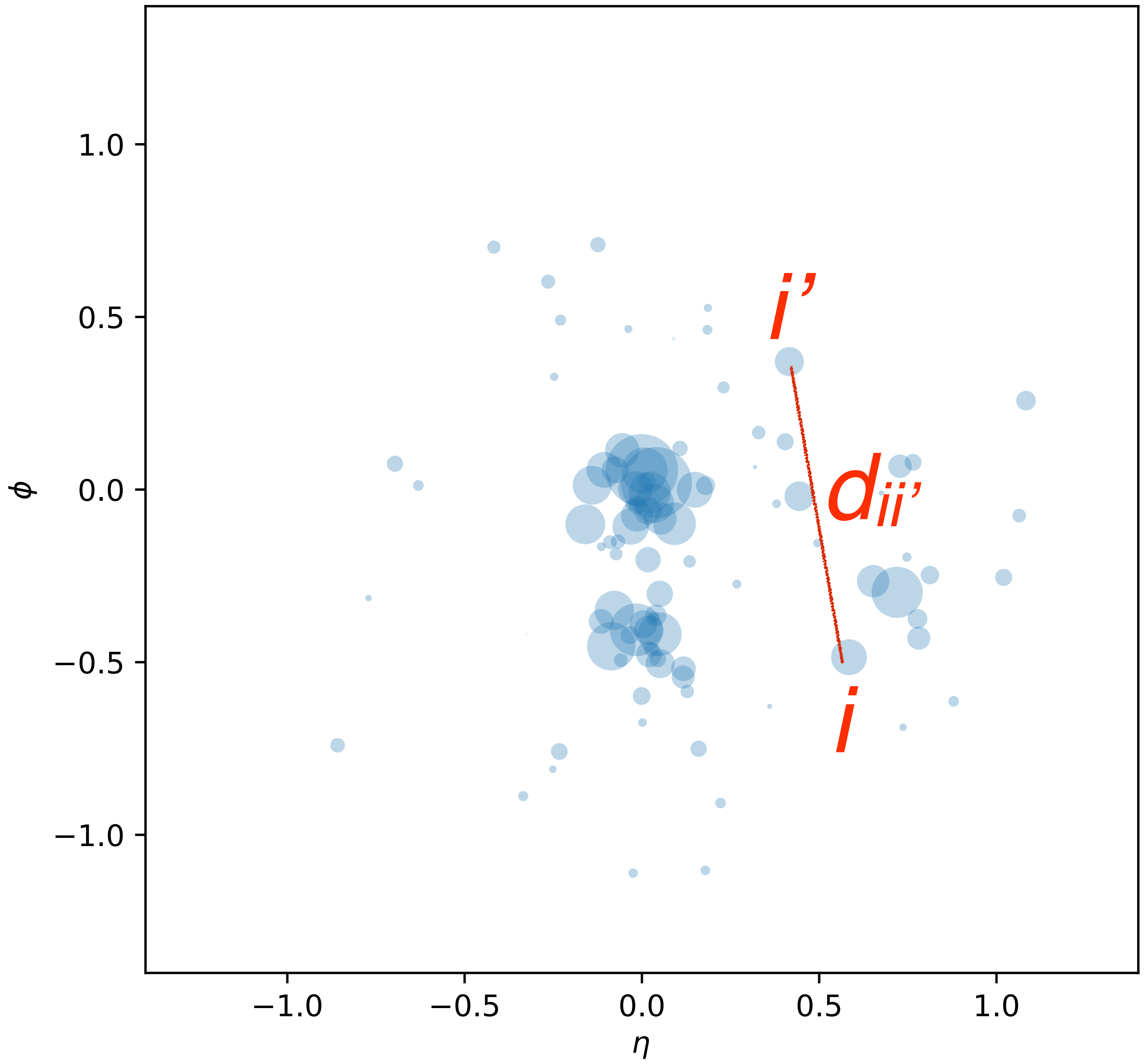


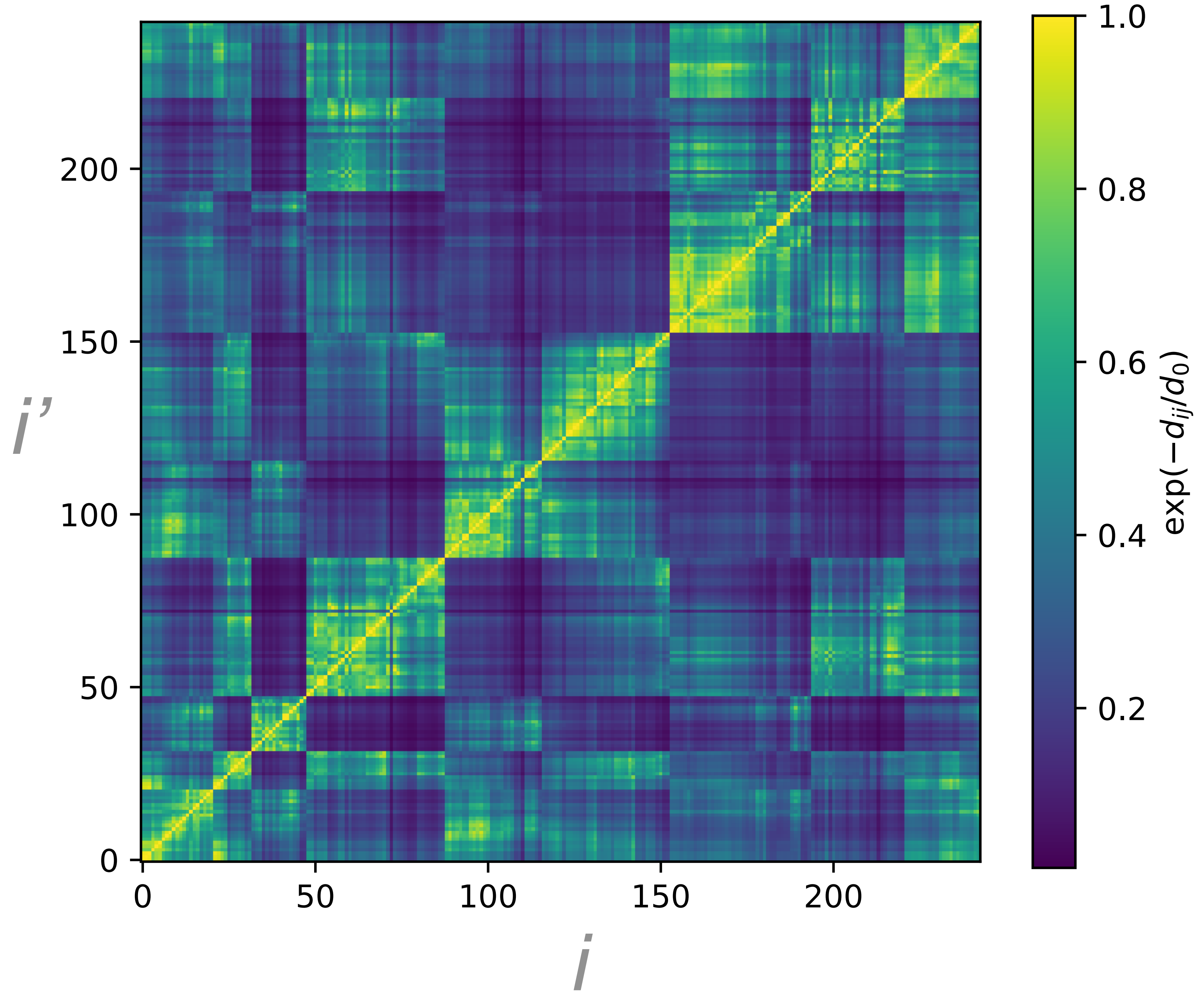


Jets as a graph

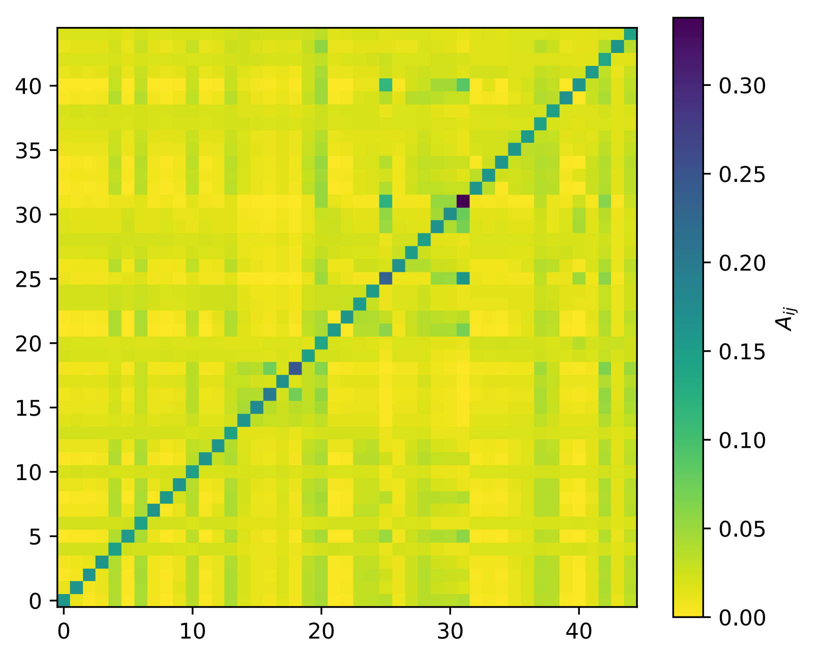
Connecting the semantics of physical closeness with **attention** in graph network provides a **bi-directional interface**. It allows us to either

- **import** physics knowledge by using distance measure of jet physics $d^{\alpha}_{ii'} = \min(p_{ti}^{2\alpha}, p_{ti'}^{2\alpha}) \frac{\Delta R^2_{ii'}}{R^2}$
- learn adjacency matrix and **export** an optimized notion of "distance" that can be used in other contexts (eg. for clustering)

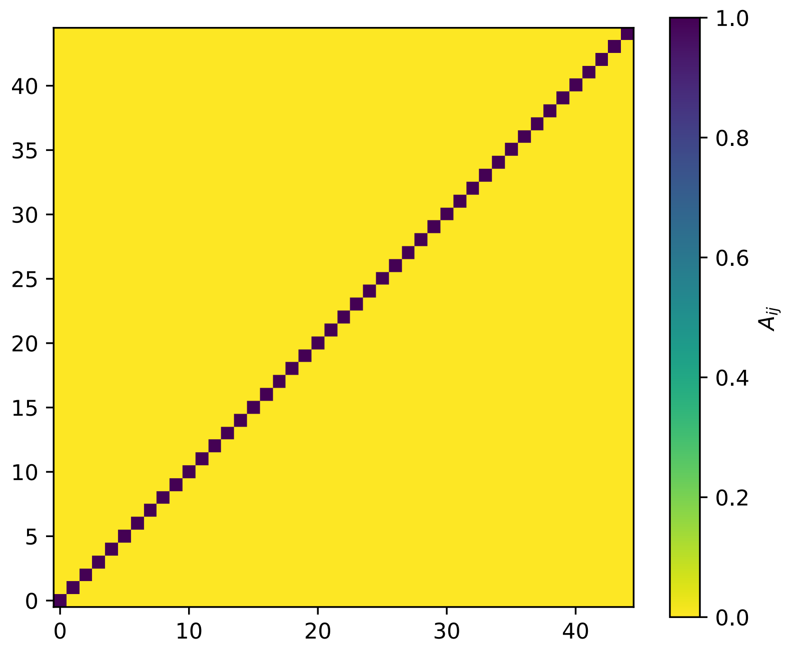




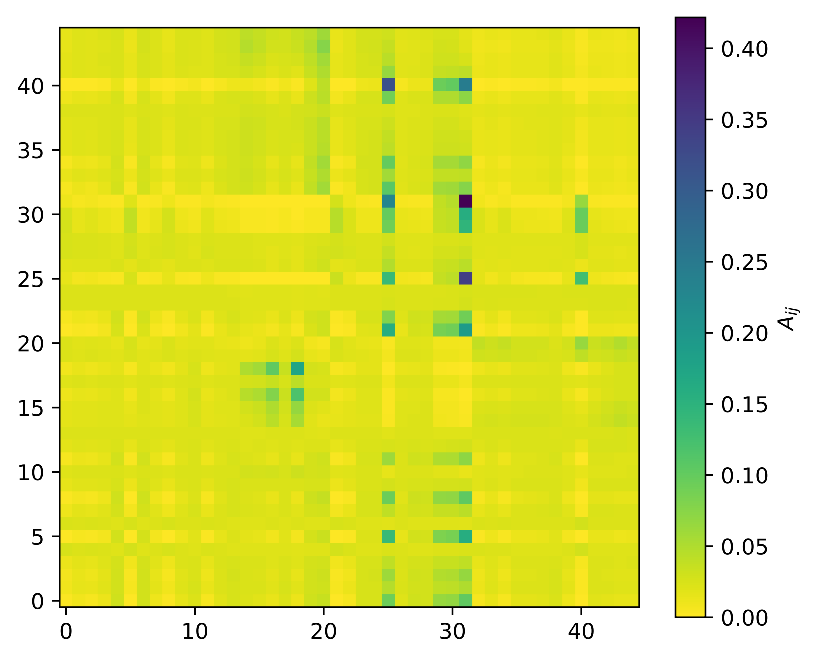
Attention



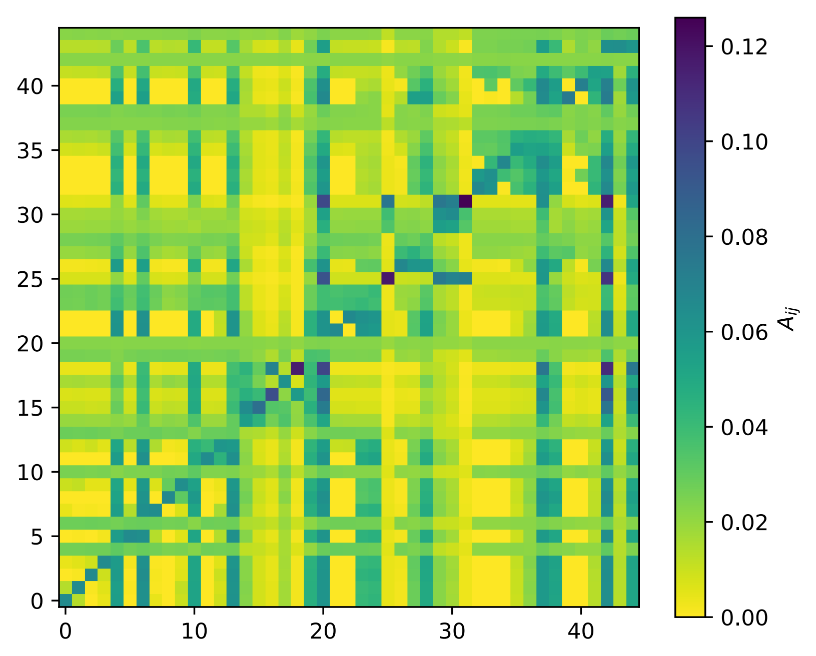
$= W_{id}$



$+ W_{learn}$



$+ W_{phys}$



Total

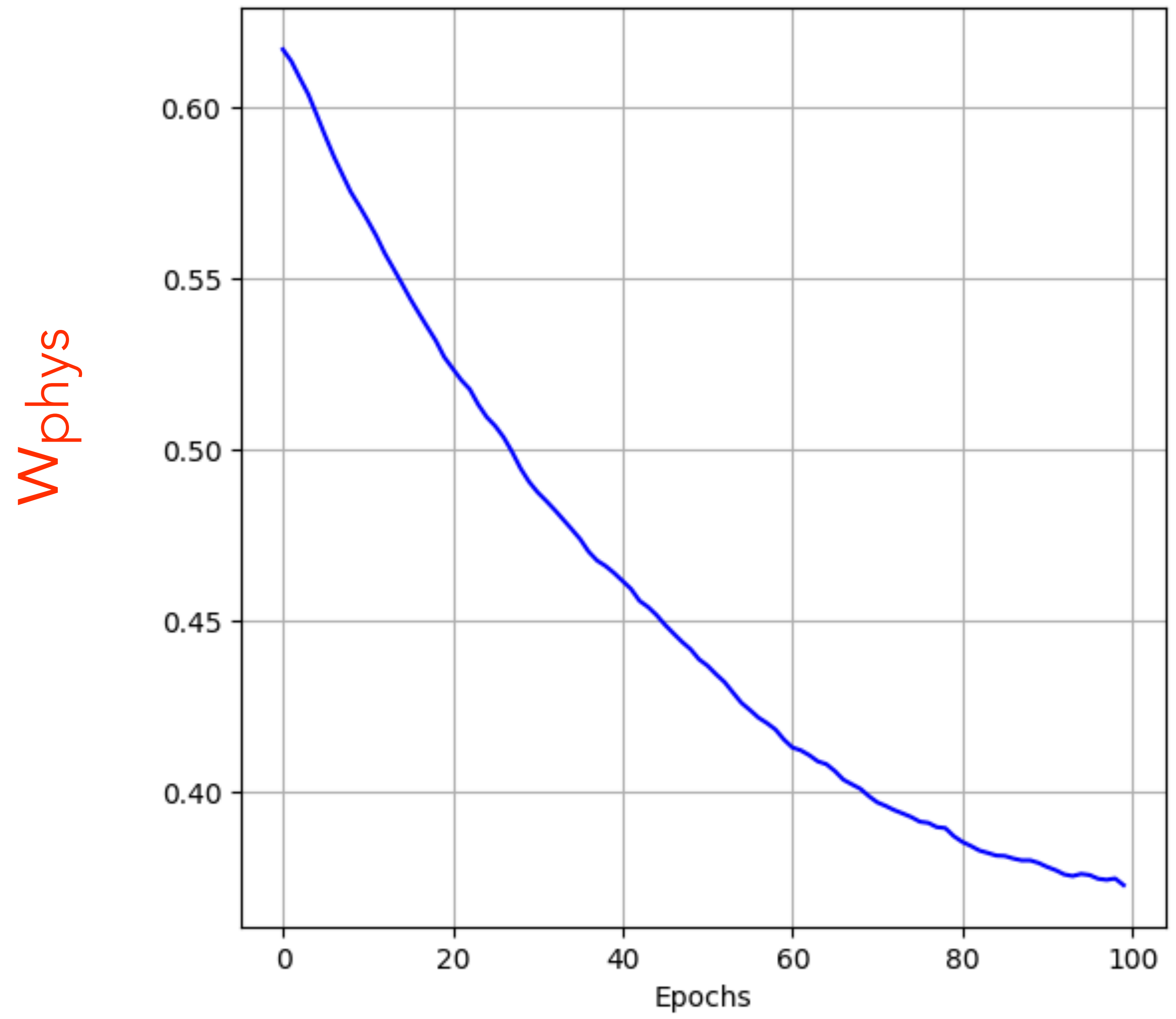
Identity

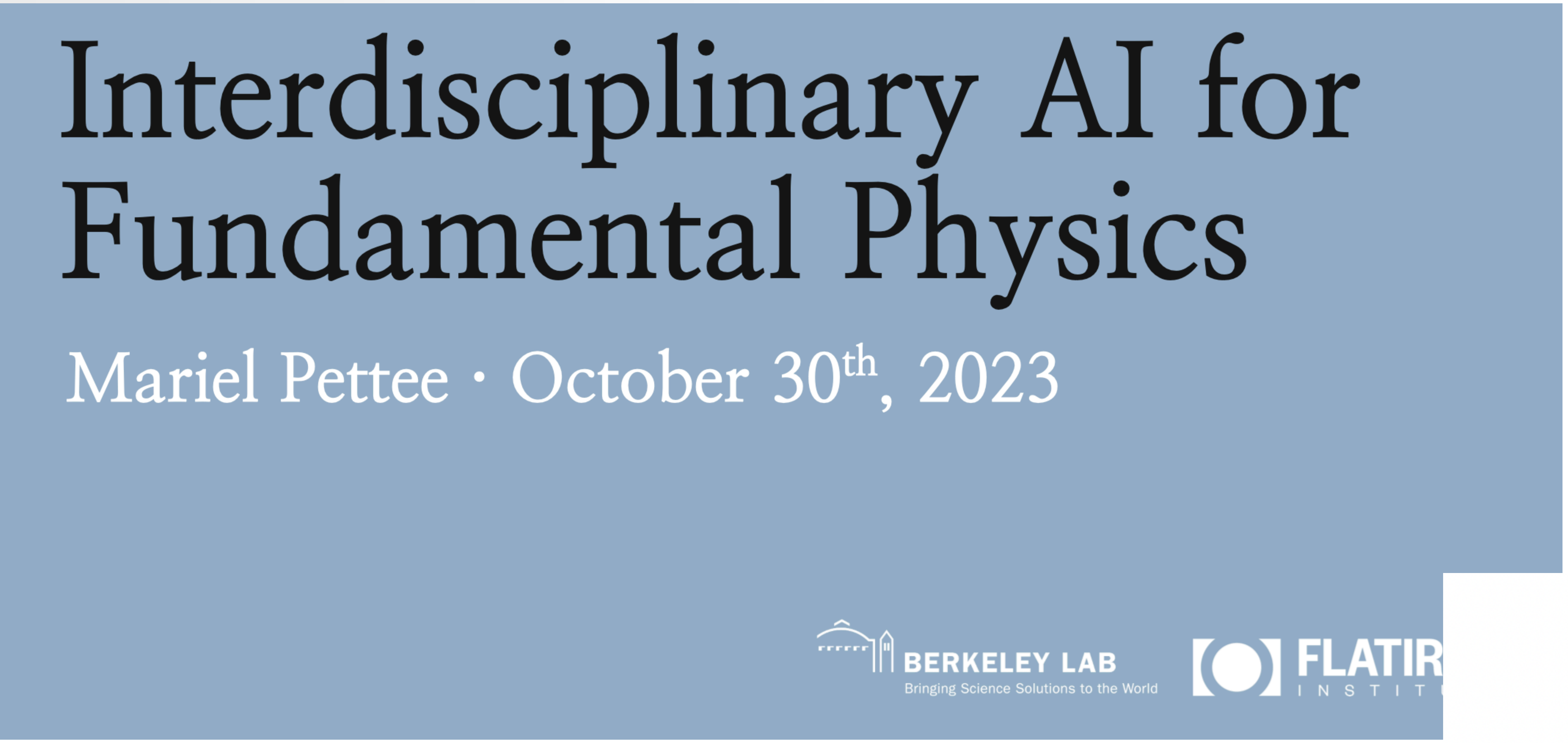
learned

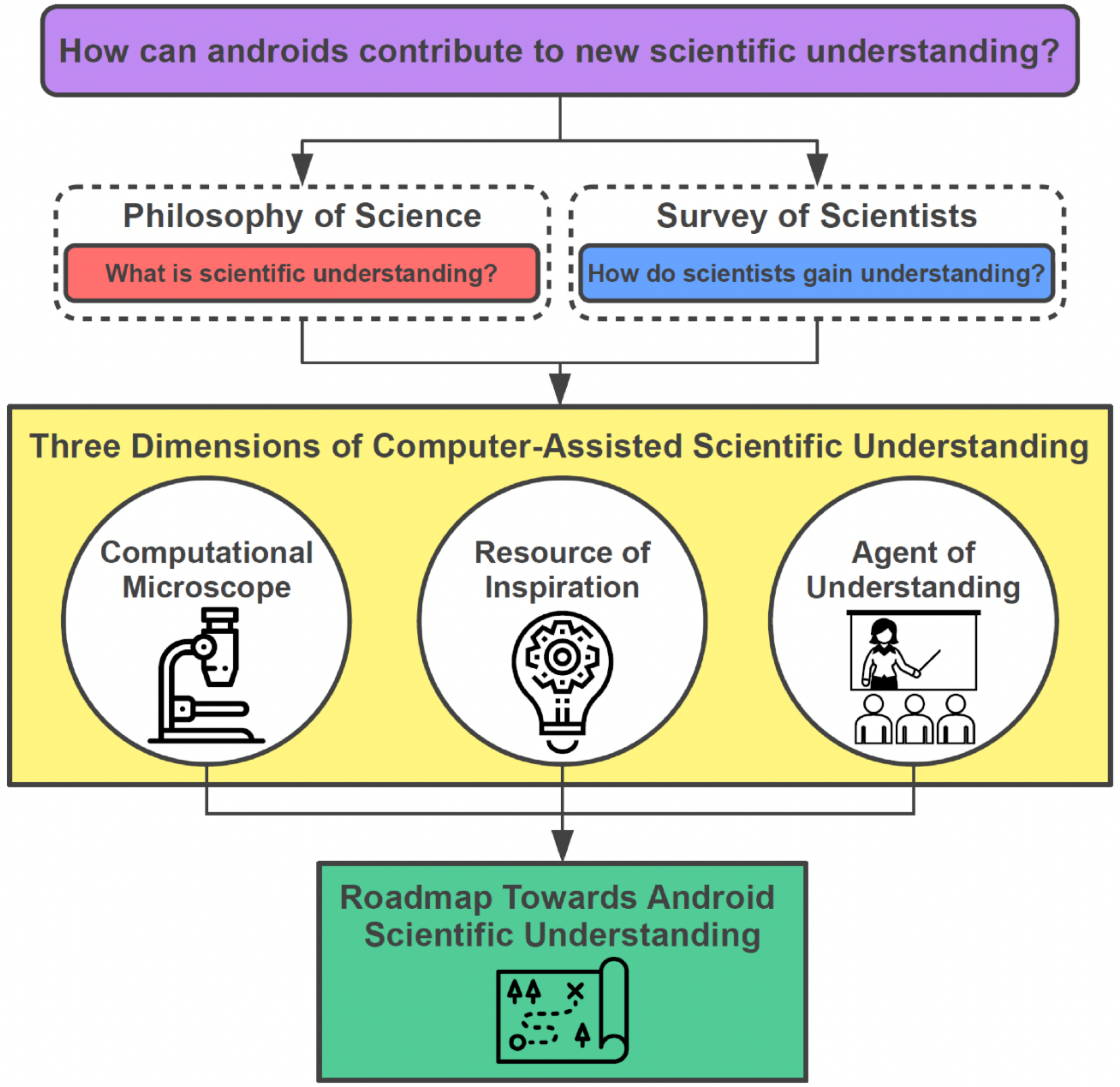
physics

The adjacency matrix starts off as purely physics-inspired (kT) and then learned adjacency matrix becomes more important

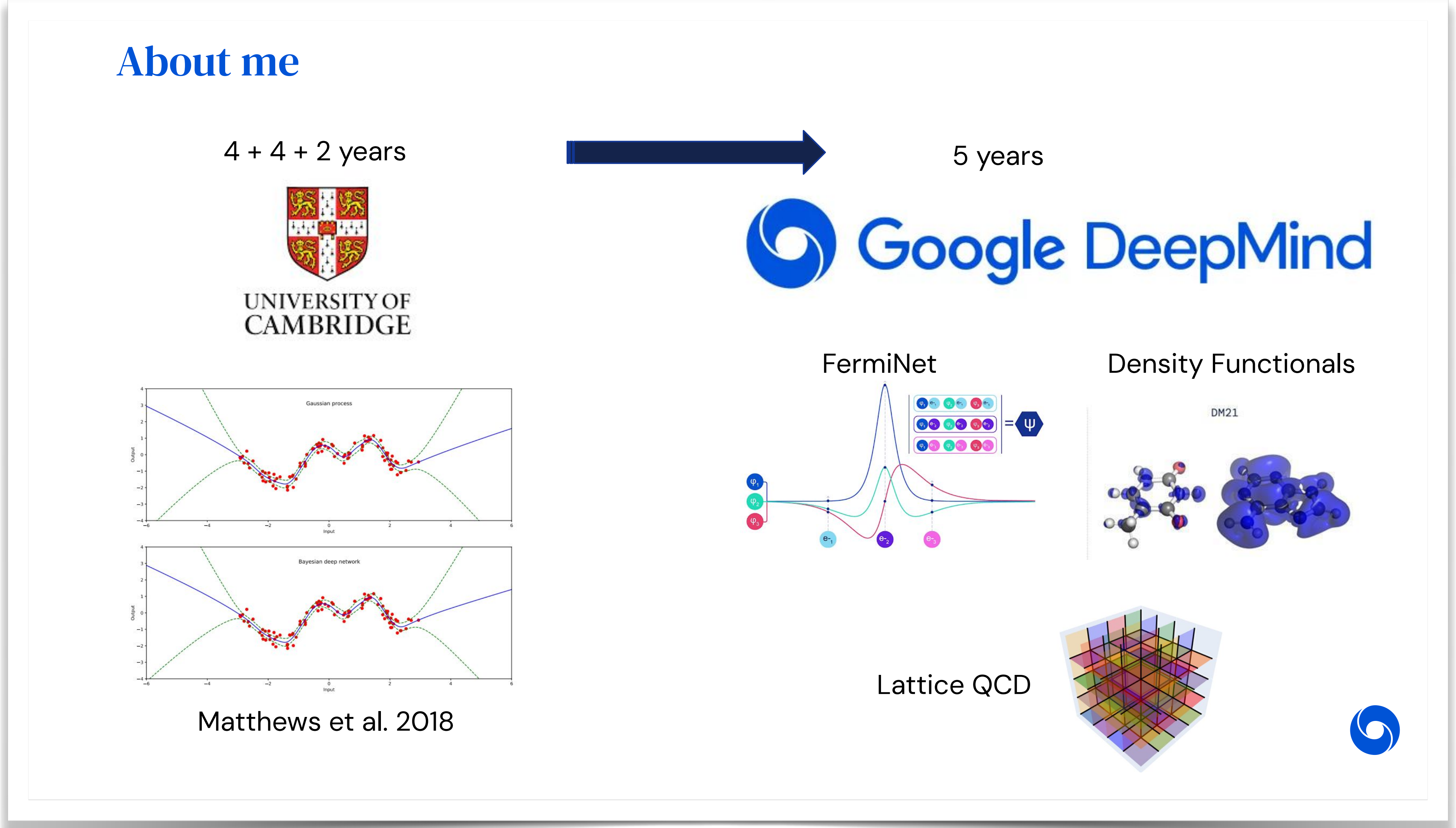
- weight given to physics adjacency matrix slowly decays, but stays relevant →

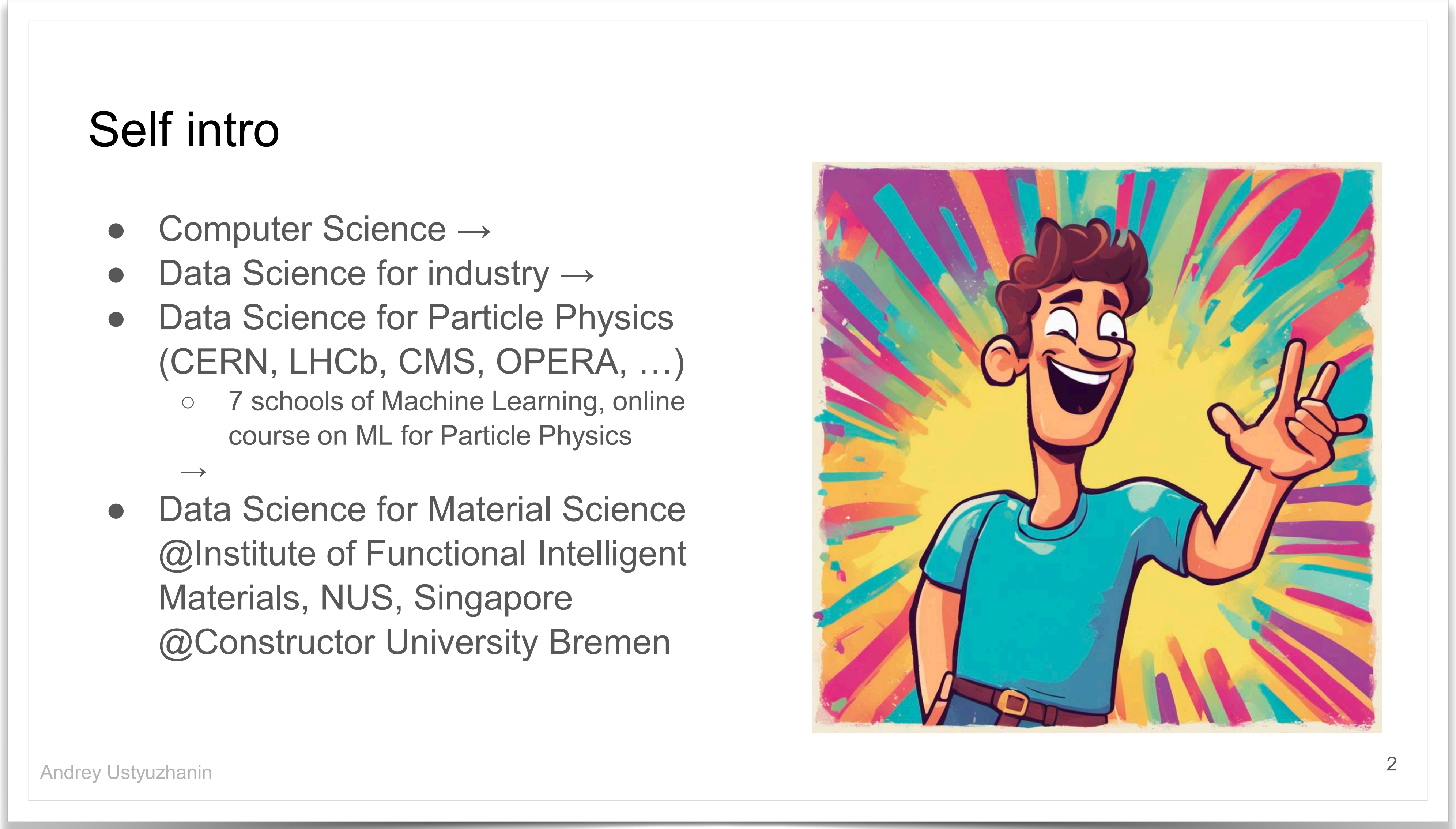


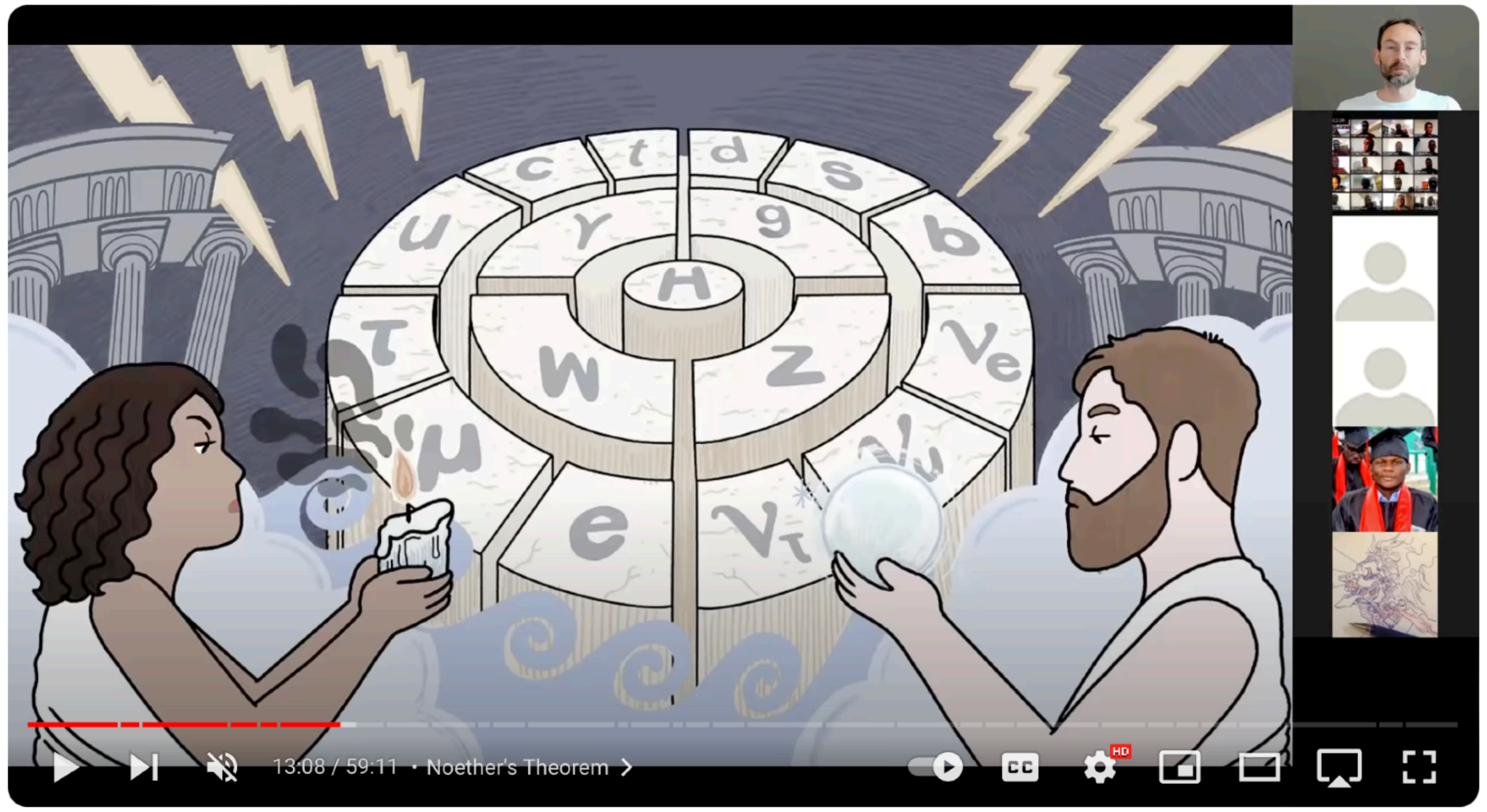




Mario Krenn, et. Al. https://arxiv.org/abs/2204.01467



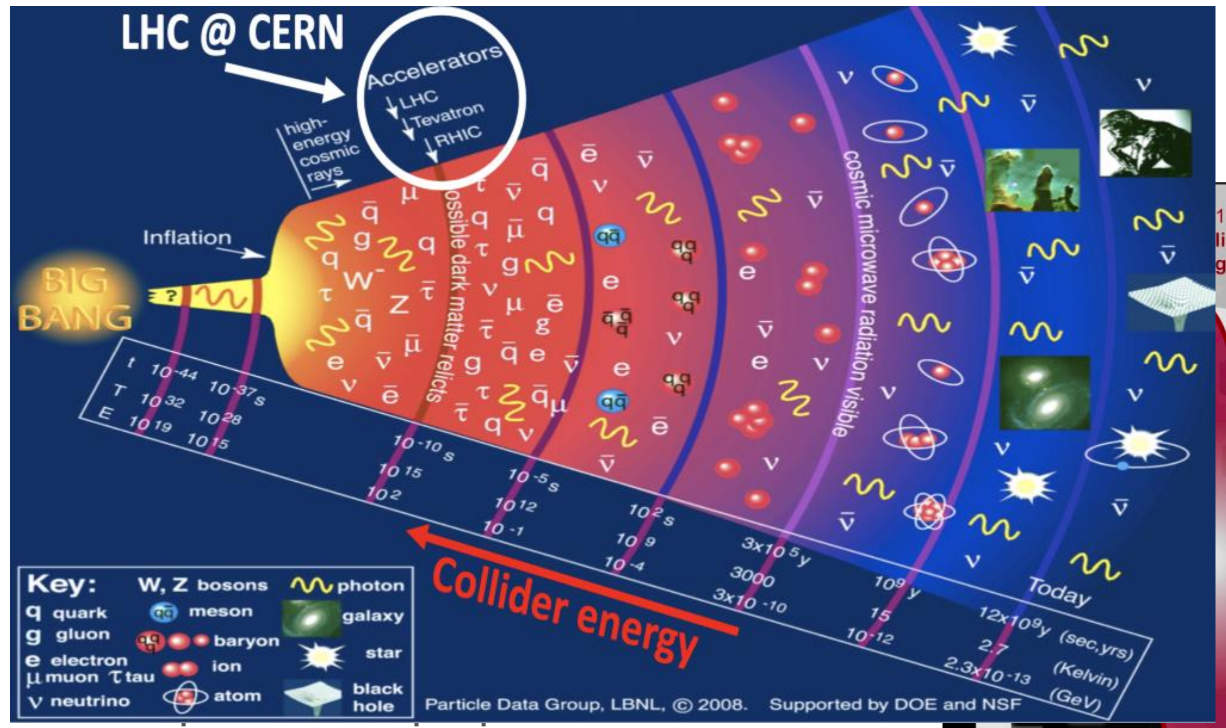


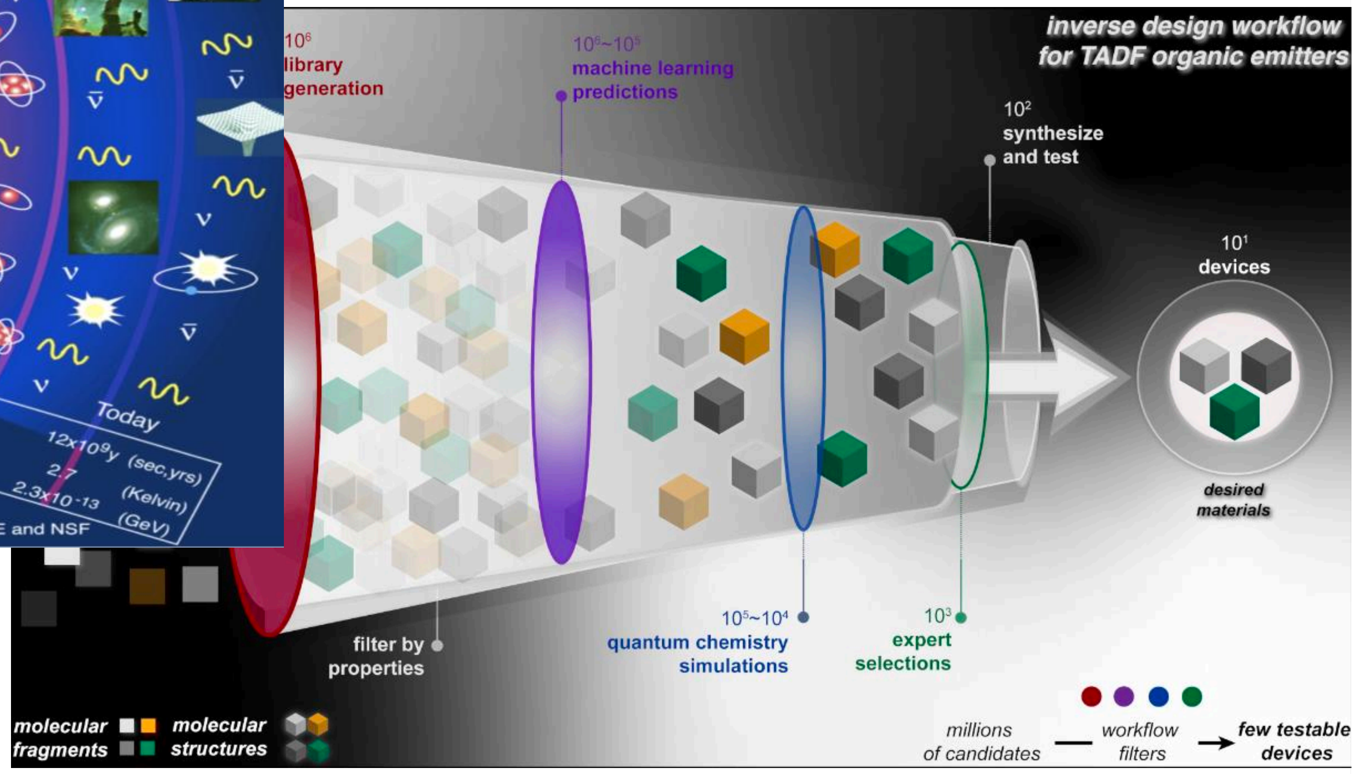


AMMI Course "Geometric Deep Learning" - Lecture 1 (Introduction) - Michael Bronstein

Cross-polination

Inverse problem, high-throughput screening (HTS)





stages.

27

HEP vs Material Science (MS)

Similarities

- fast simulation / generative models
- need for foundation models
- representation learning
- optimal transport methods
- inverse design / design optimization
- ML model uncertainty estimation
- spatial structures representations
- need for differential simulations / simulation-based inference
- denoising / stability estimation methods
- anomaly detection methods

HEP distinct features

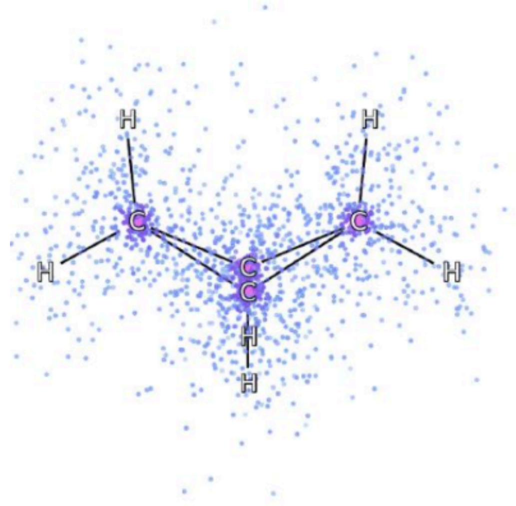
- Centralized data collection
- Bump hunting
- Science of confidence intervals
- Plenty of theoretical models for unknown
- Search for unknown

MS distinct features

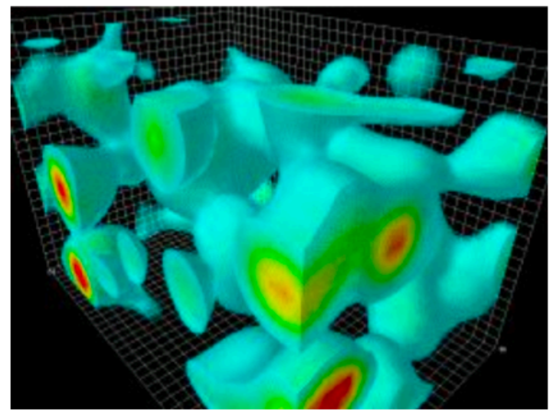
- Multiscale effects / modelling
- Time-dependent modelling
- Data is heavily fragmented

Andrey Ustyuzhanin

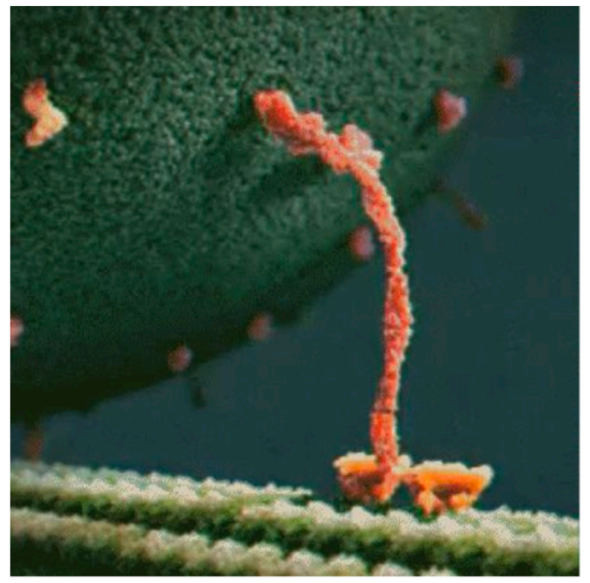
36



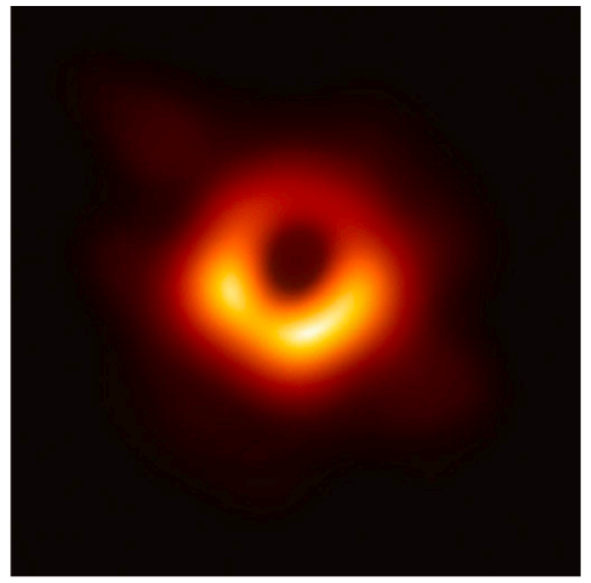
Quantum Monte Carlo



Lattice QCD



Protein physics



Black hole astronomy

Technical advances in deploying AI/ML in experiments

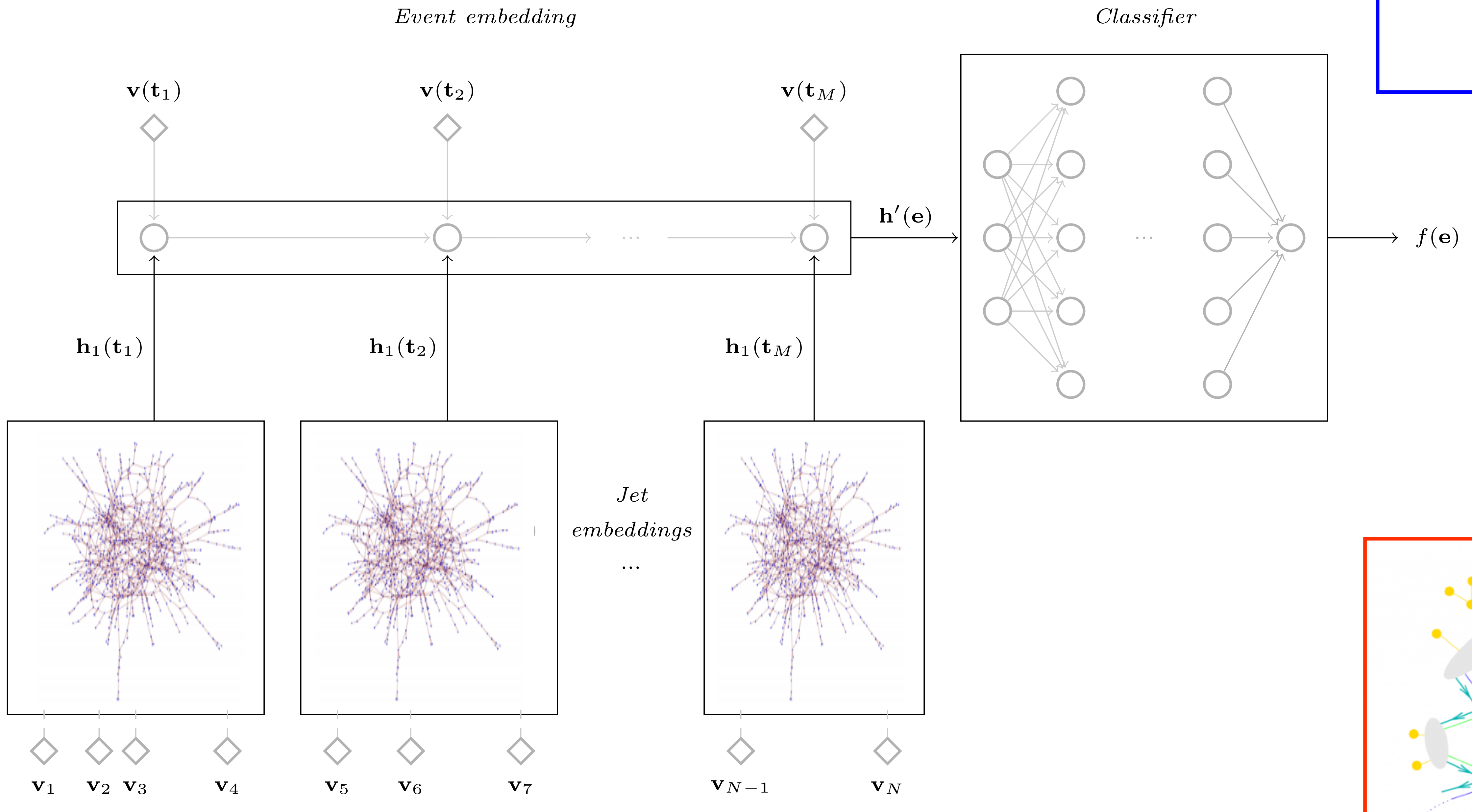
Bigger data, shorter time: Real-time inference for scientific discovery

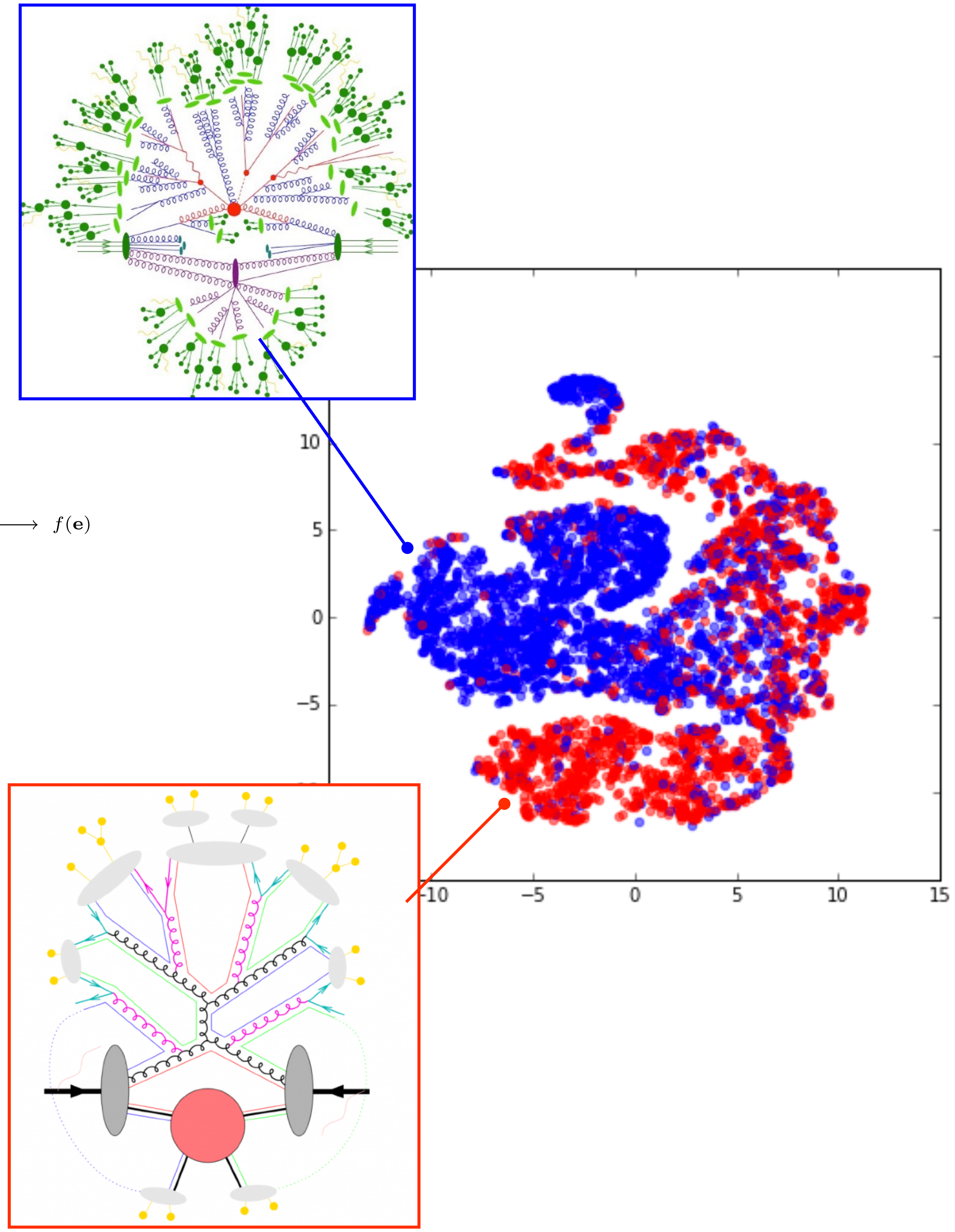
Thea Klæboe Årrestad
(ETH Zürich)

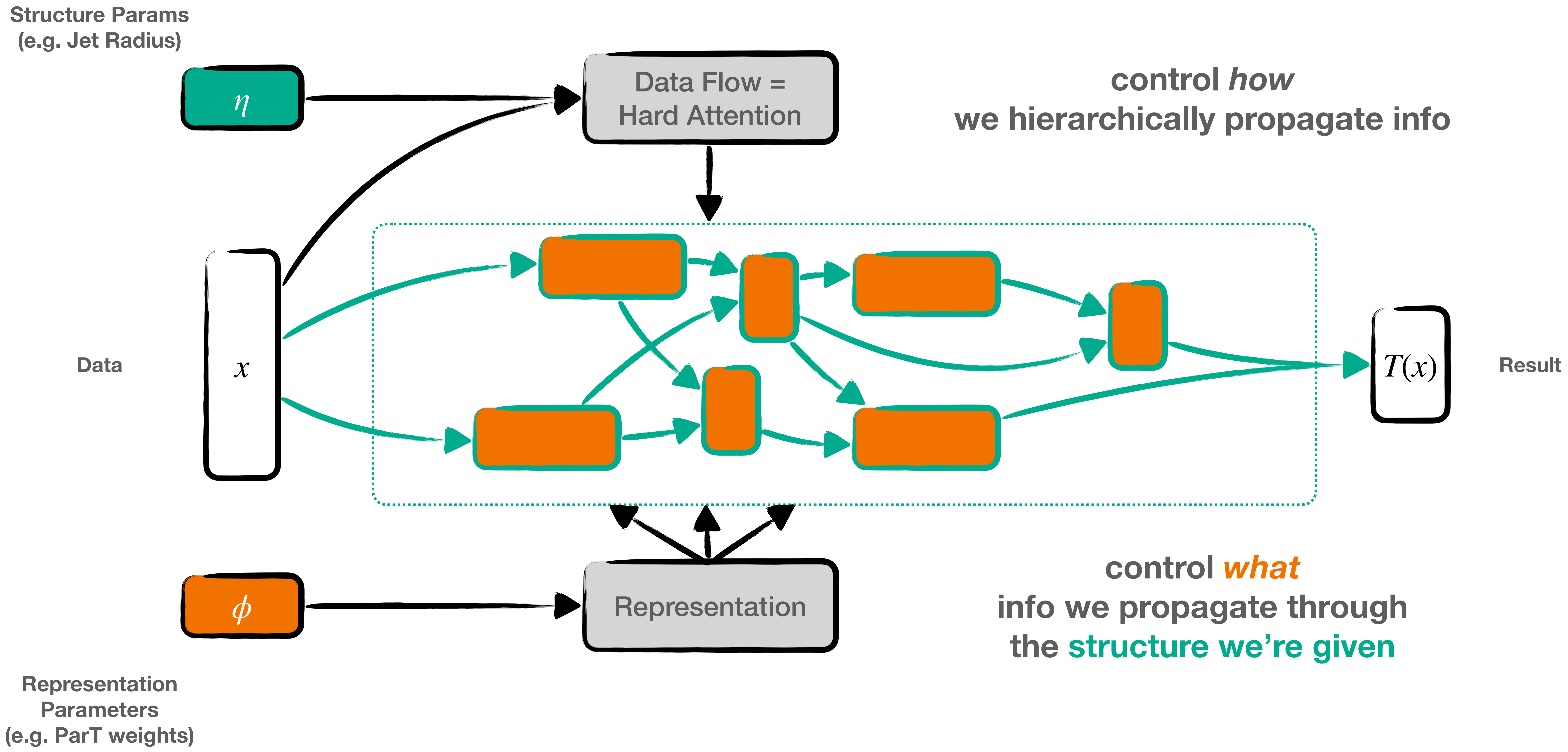
EVENT EMBEDDINGS

Jointly optimize jet embedding → event embedding → classifier

It scales!







control *how*
we hierarchically propagate info

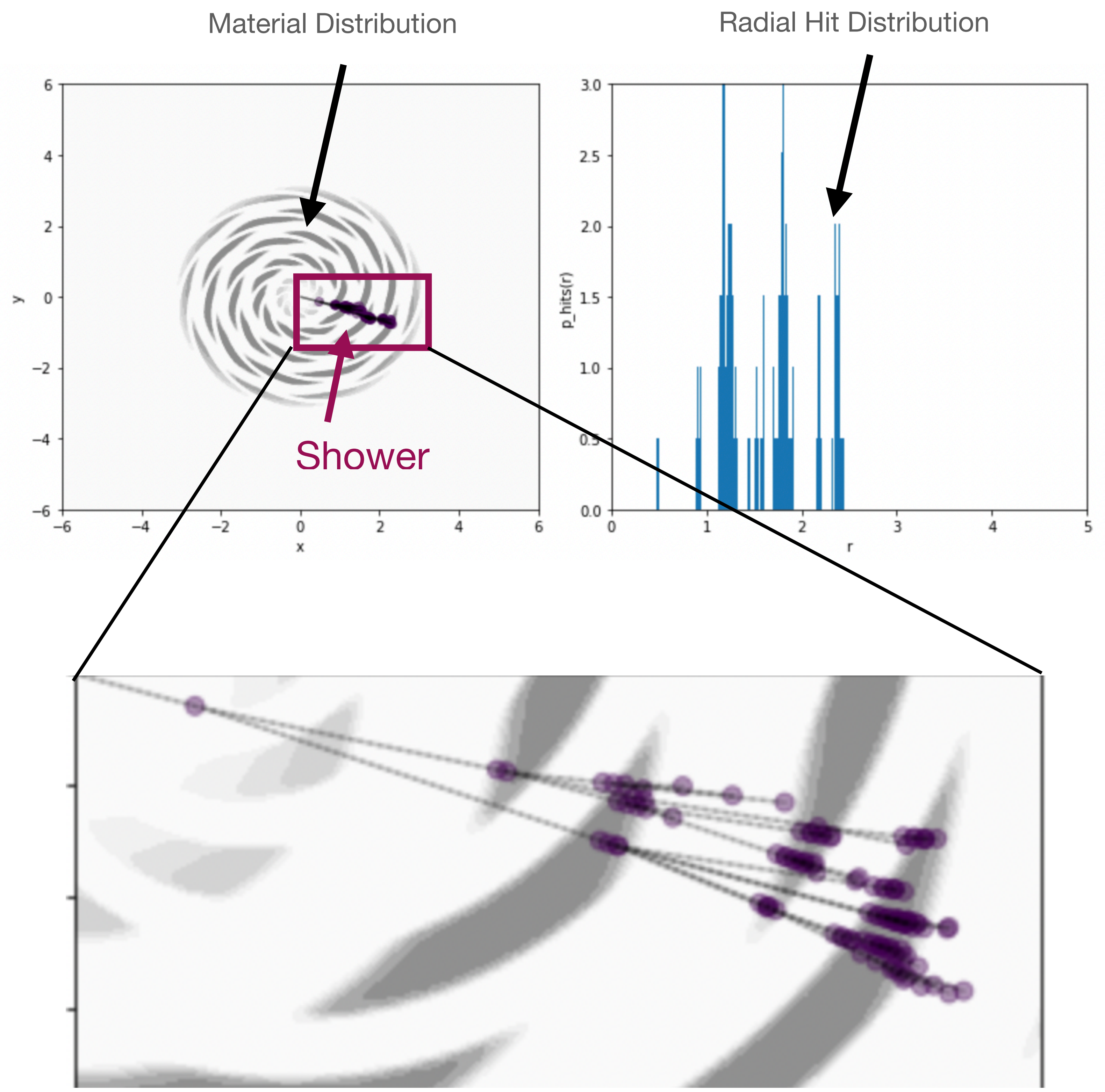
control *what*
info we propagate through
the **structure we're given**

Differentiating through Particle Showers

High Density: E-loss and splitting
Low Density: linear propagation

Design Parameter:
Radial Distance of Material

Design Goal: Shower Depth



Some new Questions for next H&N ?

How do we calibrate high-dim representation?

Will we get a “safe” calibrated fine tuning manifold?

Can we optimize structural pieces (e.g. jet definition) → stochastic reconstruction?

Supervised vs Self-supervised Backbones (JetCLR, ReSim, MPM,…)

Michael’s Talk Next

How will AI enable autonomous particle accelerators?

V. Kain
Data Science for Beam Operation
Beams Department, CERN

Predicting magnetic hysteresis and eddy current effects

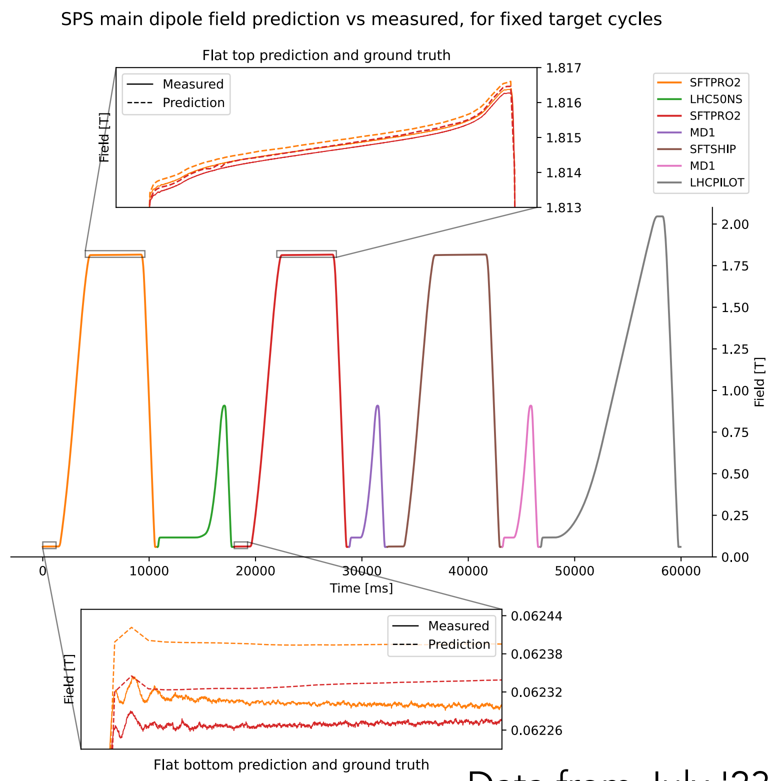
Potentially game-changing!

Time-series forecasting problem: need magnets to be measured on test bench
$[B_t, B_{t+1}, \ldots, B_{t+n-1}], [I_t, I_{t+1}, \ldots, I_{t+n+N}] \rightarrow [B_{t+n}, B_{t+n+1}, \ldots, B_{t+n+N}]$

First operational experience:

- **feedforward correction** triggered before every cycle
- accuracy not sufficient yet

First results PhyLSTM for SPS main dipoles assuming $\ddot{B} + g(B,\dot{B}) = \Gamma I(t)$, next: Transformers



Data from July '23

7 recommendations → Automating exploitation

CERN
Esplanade des Particules 1
P.O. Box
1211 Geneva 23 Switzerland

EMDS NO.	REV.	VALIDITY
2922514	1.0	RELEASED

REFERENCE
2922514

Date: July 28, 2023

PROJECT REPORT

Efficiency Think Tank Report

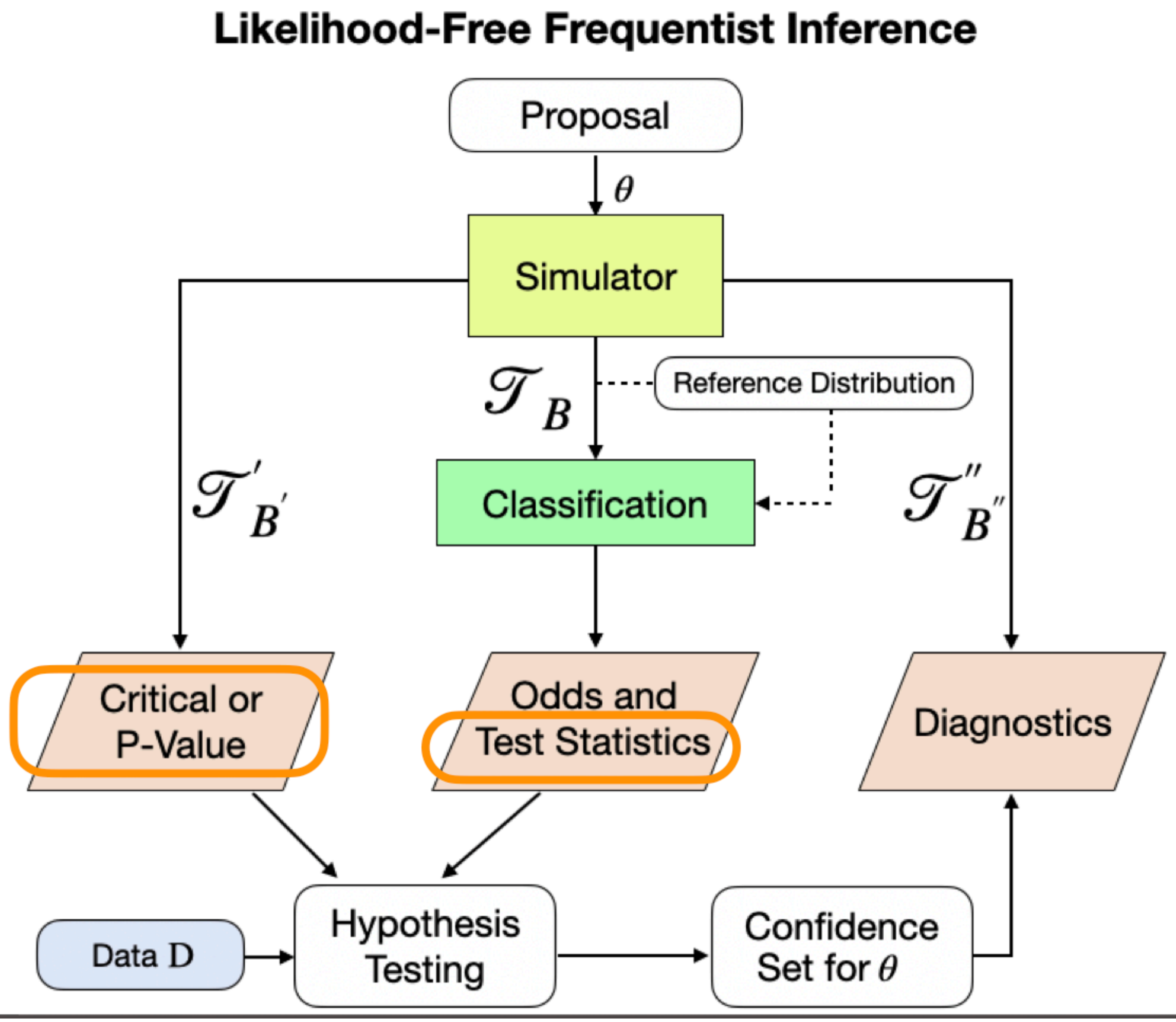
1. **AI** Hysteresis compensation
2. Automatic and dynamic beam scheduling
3. **AI** Automatic LHC filling
4. **AI** Auto-pilots
5. **AI** Automatic fault analysis, recovery and prevention

→ Fully automated standard physics operation

6. Automatic testing and sequencing
7. **AI** Automatic parameter optimisation

→ Goal: reduce commissioning time by 50 %

Our Inference Machinery

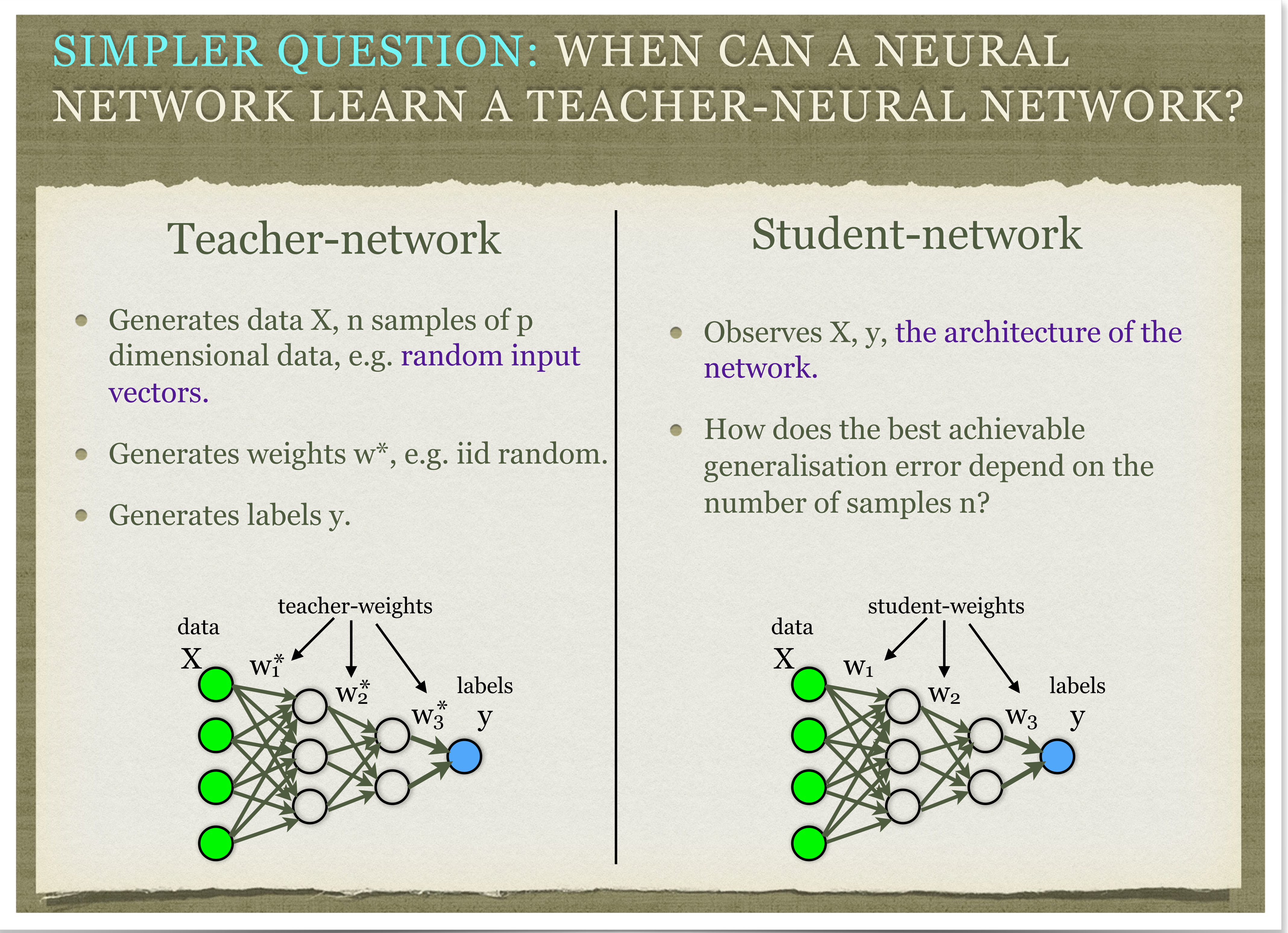


Parting thoughts

Interplay of key ingredients of Deep Learning

Hard to analyze the effect of data structure for real-world data sources.

Toy models are useful!



* Lenka Zdeborova:
Talk: https://ml4physicalsciences.github.io
Position piece: https://rdcu.be/b4p1m

Teacher → Causal, Generative Model (Simulator)

Richer set of problems can be investigated.

TENSION BETWEEN TWO THEMES?

Some tension between

- Imposing inductive bias & domain knowledge
- recent results showing that over-parameterized networks may play better with optimization algorithms

A topic for further study: when does (correct) inductive bias hurt?

Thank you Tobias and Eilam!