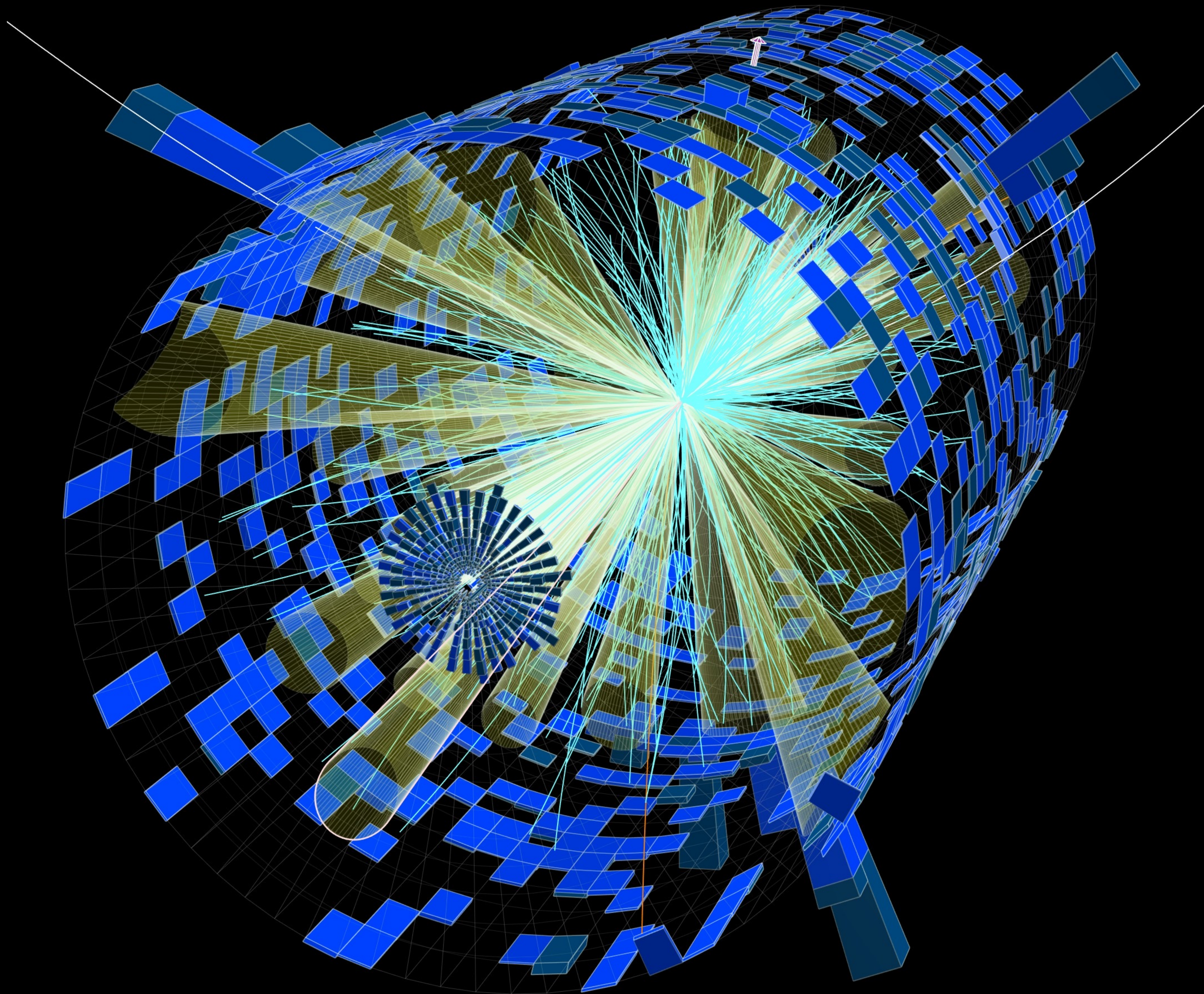
NYU Center for Data Science

Meta AI

Center for Cosmology and Particle Physics

VIGNETTES IN PHYSICS-INSPIRED AI RESEARCH



@KyleCranmer
New York University
Department of Physics
Center for Data Science
CILVR Lab

Collaborators + Many More


Gilles Louppe
U. Liège


Kyunghyun Cho


Joan Bruna


Felix Kling
Sebastian Macaluso


Meghan Frate


Sid Mishra-Sharma


Miles Cranmer


@HServiansky


Nimrod Segol


Johann Brehmer


Craig Greenberg


Nicholas Monath


Peter Battaglia


Alvaro Sanchez Gonzalez

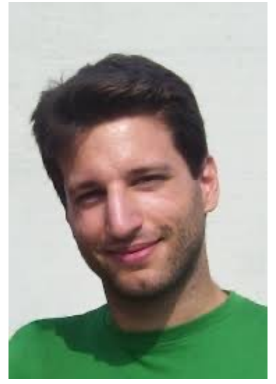

Shirley Ho

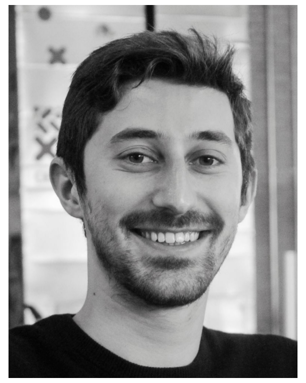

Irina Espejo


@HaggaiMaron


Yaron Lipman


Julian Urban


George Papamakarios


Michael Albergo


Danilo Rezende


Sébastien Racanière


Atılım Güneş Baydin
University of Oxford


Wahid Bhimji
NERSC, Berkeley Lab


Frank Wood
U. Victoria


Jonathan Shlomi


Eilam Gross


Phiala Shanahan


William Detmold


Gurtej Kanwar


Dan Hackett


Denis Boyda


Fernando
Romero-Lopez


Matthew Drnevich


Savannah Thais
Yale University

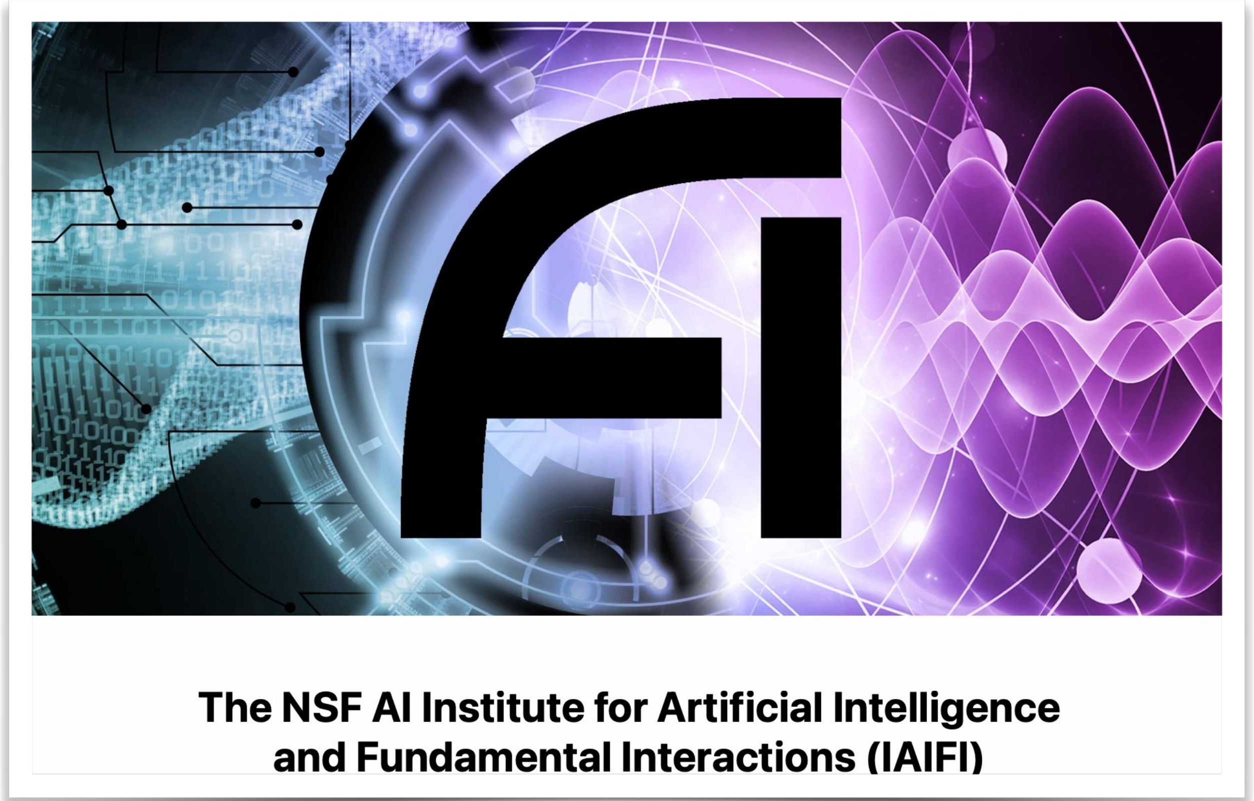
NSF

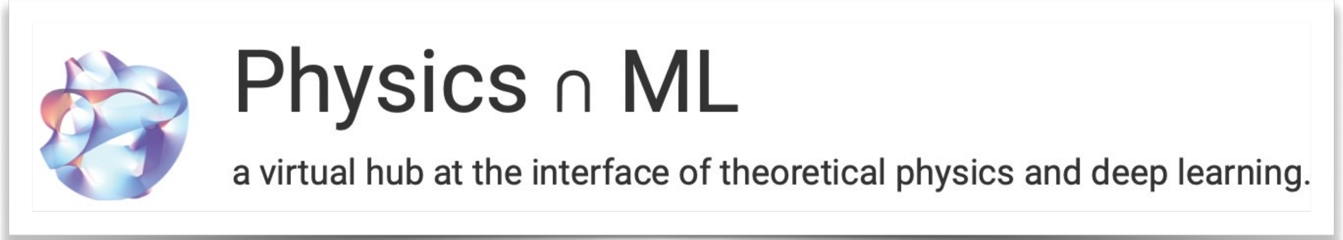
MSDSE
Moore - Sloan Data Science Environments

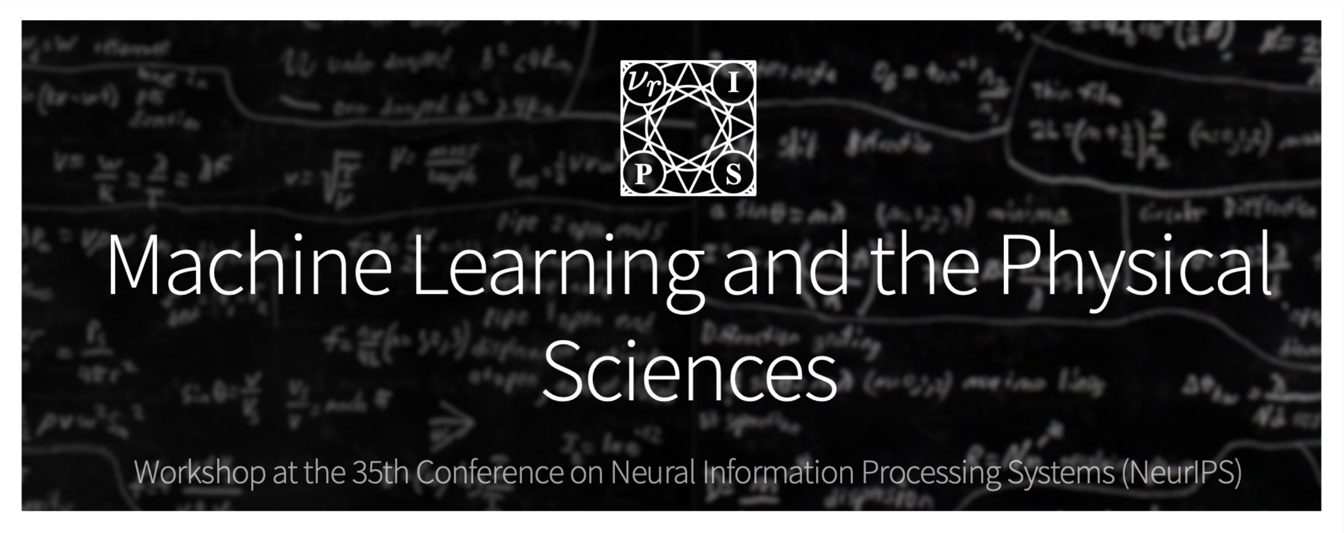
Physics and AI/ML is a hot area

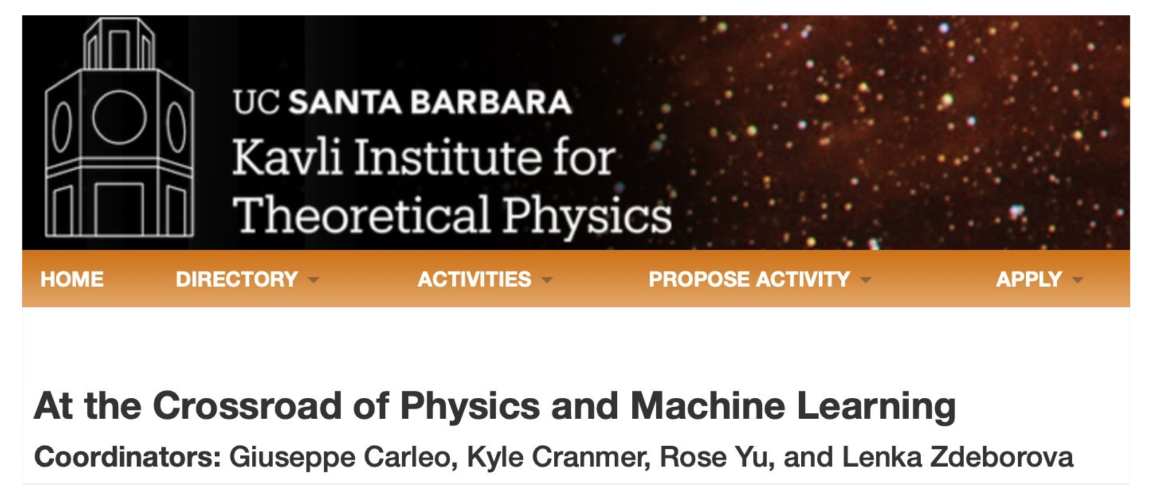
Interaction is often framed in 1 of 2 ways:

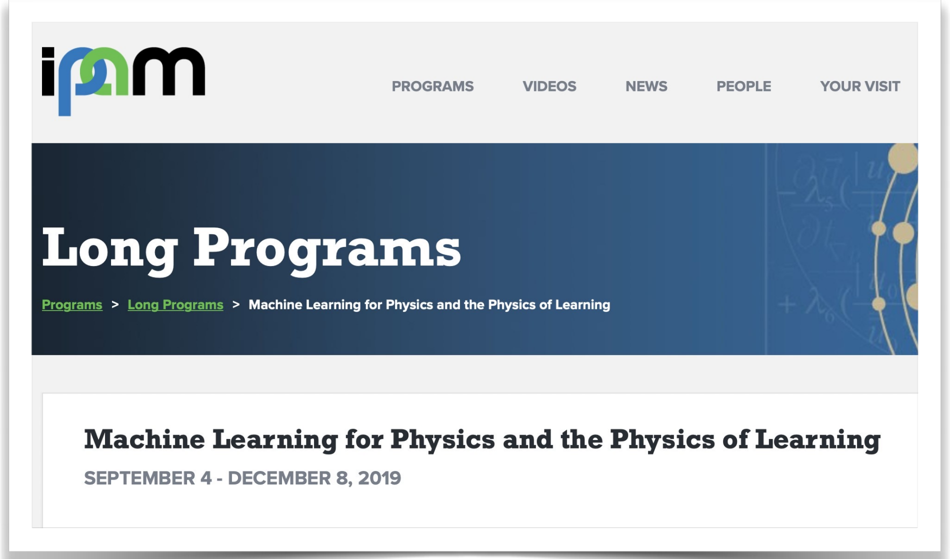
- "ML for physics"
- "Physics for ML"

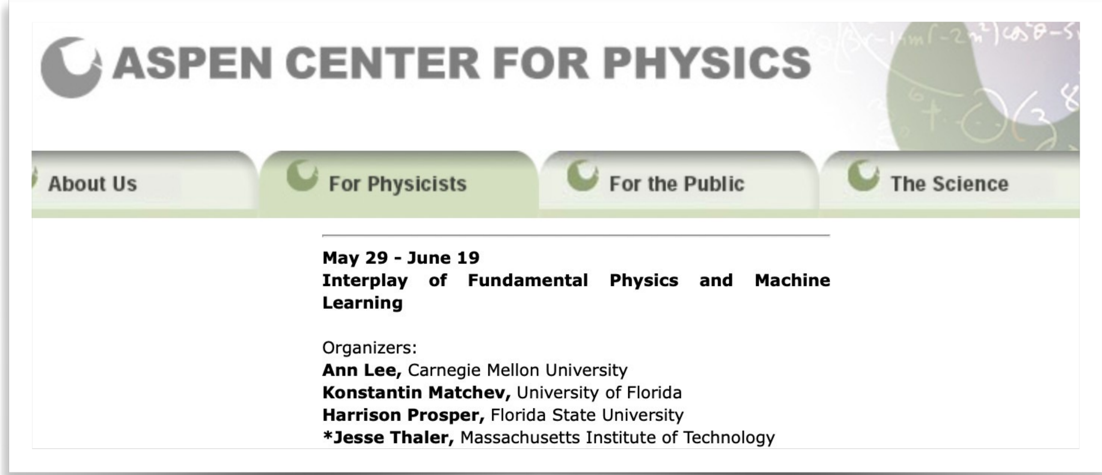












ML for physics

A multitude of examples of ML techniques being used to solve physics problems

- off-the-shelf <——————- spectrum ——————————> custom solutions

- **Kazuhiro Terao, Staff Scientist, SLAC National Accelerator Laboratory, Stanford University**
  - **Friday, December 10, 2:00-3:00pm**
  - *"Machines to find ghost particles in big data"*
  - YouTube Recording
  - Talk Slides
  - Abstract: The neutrino is the most mysterious of the standard model particles. They only interact weakly and can pass through light years of matter without leaving a trace. Since the discovery of neutrino oscillations, experiments provided the first positive evidence for physics beyond the standard model. Yet, there are remaining questions which answers may reveal new physics and possible explanation of the asymmetry in the presence of matters and anti-matters in the current universe. A new generation of neutrino experiments are coming online to address those questions in the era of high precision

- **Giuseppe Carleo, Assistant Professor, Computational Quantum Science Laboratory, École Polytechnique**
  - **Friday, March 4, 2022, 2:00-3:00pm**
  - *"Neural-Network Quantum States: new computational possibilities at the boundaries of the many-body problem"*
  - Abstract: Machine-learning-based approaches, routinely adopted in cutting-edge industrial applications, are being increasingly adopted to study fundamental problems in science. Many-body physics is very much at the forefront of these exciting developments, given its intrinsic "big-data" nature. In this talk, I will present selected applications to the quantum realm. First, I will discuss how a systematic and controlled machine learning of the many-body wave-function can be realized. This goal is achieved by a variational representation of quantum states based on artificial neural networks [1]. I will then discuss recent applications in diverse domains, including prototypical open problems in many-body quantum physics - interacting fermions [2,3,4] and frustrated spins [5,6] — where these approaches typically outperform existing state of the art methods. Finally, I will discuss applications in the context of quantum computing [7,8].

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], M. Fairbairn[5], W. Fedorko[6], C. Gay[6], L. Gouskos[7], P. T. Komiske[8], S. Leiss[1], A. Lister[6], S. Macaluso[3,4], E. M. Metodiev[8], L. Moore[9], B. Nachman,[10,11], K. Nordström[12,13], J. Pearkes[6], H. Qu[7], Y. Rath[14], M. Rieger[14], D. Shih[4], J. M. Thompson[2], and S. Varma[5]

Physics for ML

Primarily using tools developed for statistical mechanics (e.g. replica trick) and quantum field theory (e.g. EFT) to understand the dynamics of deep learning

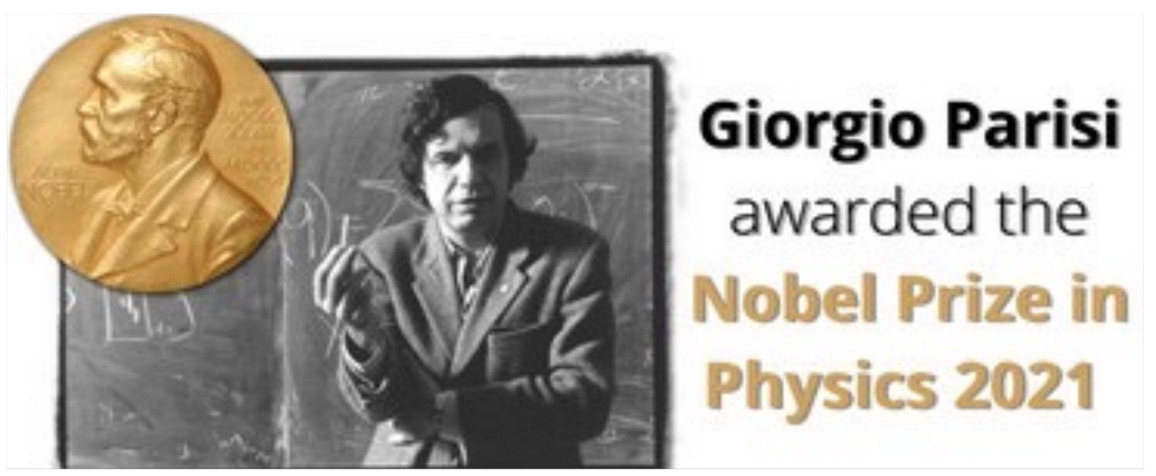
- **Yasaman Bahri, Research Scientist, Google Research, Brain Team**
  - **Friday, November 12, 2:00-3:00pm**
  - *"Understanding deep learning"*
  - YouTube Recording
  - Talk Slides
  - Abstract: Deep neural networks are a rich class of function approximators that are now ubiquitous in many domains and enable new frontiers in physics and other sciences, but their function, limitations, and governing principles are not yet well-understood. I will overview a few results from a research program seeking to understand deep learning by proceeding scientifically. These investigations draw ideas, tools, and hysics, with close guidance from computational experiments, and integrate omputer science and statistics. I will discuss some past highlights from the study networks — such as exact connections to Gaussian processes and linear models, gation in deep networks — and then focus on emerging questions surrounding the ng laws") and predictability.

Dan Roberts

effective theory of deep learning, quantum field theory, black holes & quantum chaos, word play

Fall 2021

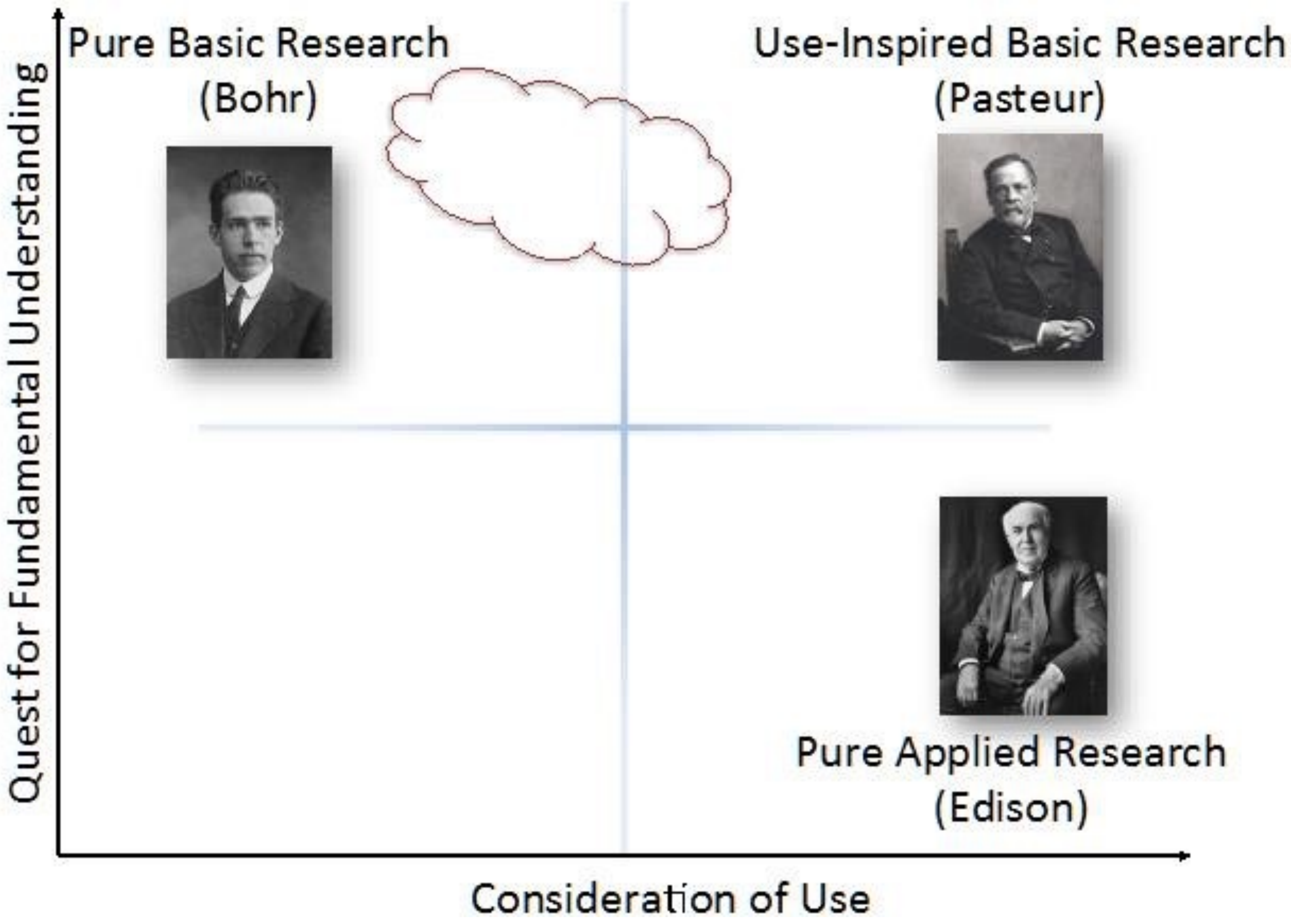
- **Surya Ganguli, Associate Professor, Applied Physics, Stanford University**
  - **Friday, September 17, 2:00-3:00pm**
  - *"Understanding computation using physics and exploiting physics for computation"*
  - YouTube Recording
  - Talk Slides
  - Abstract: We are witnessing an exciting interplay between physics, computation and neurobiology that spans in multiple directions. In one direction we can use the power of complex systems analysis, developed in theoretical physics and applied mathematics, to elucidate design principles



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

Pasteur's Quadrant & Use-inspired research

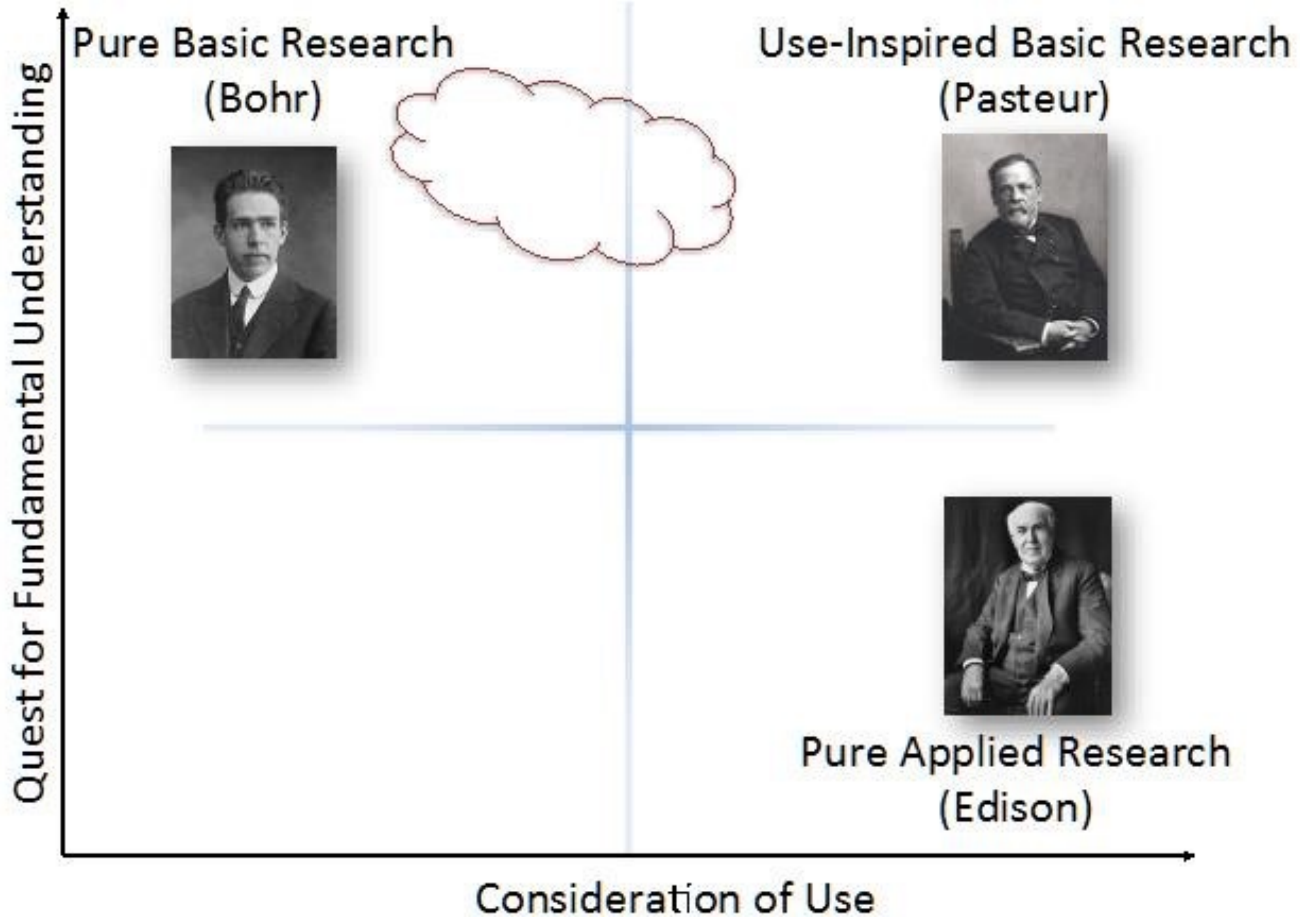
Distinct from pure basic research and pure applied research is the concept of **use-inspired research.**



Donald Stokes, Pasteur's Quadrant
Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

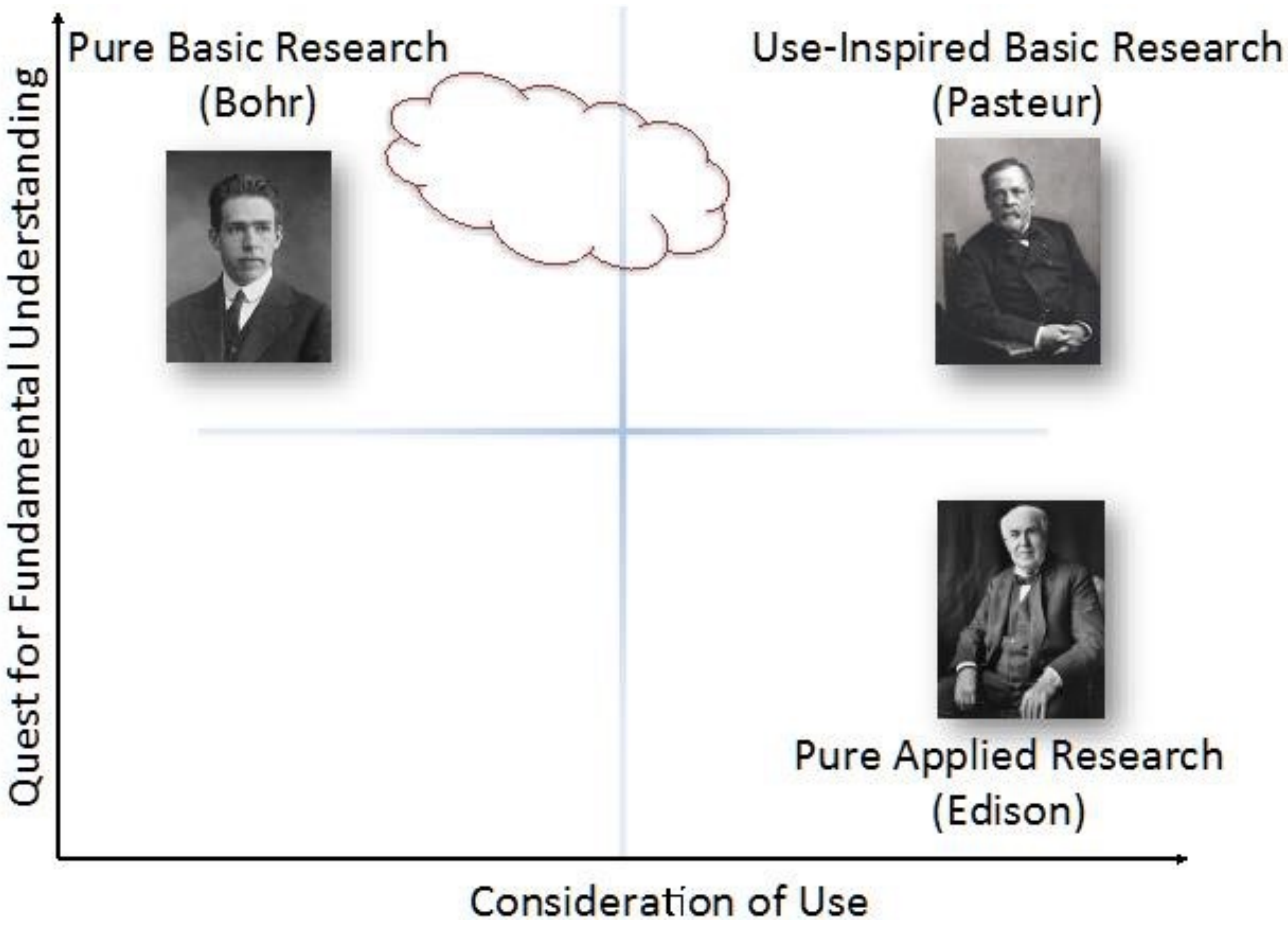
Pasteur's Quadrant & Use-inspired research

The claim is that **foundational advances** are often **inspired by the context** and particularities of a specific applied problem setting: “reality is stranger than fiction”



Donald Stokes, Pasteur's Quadrant
Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

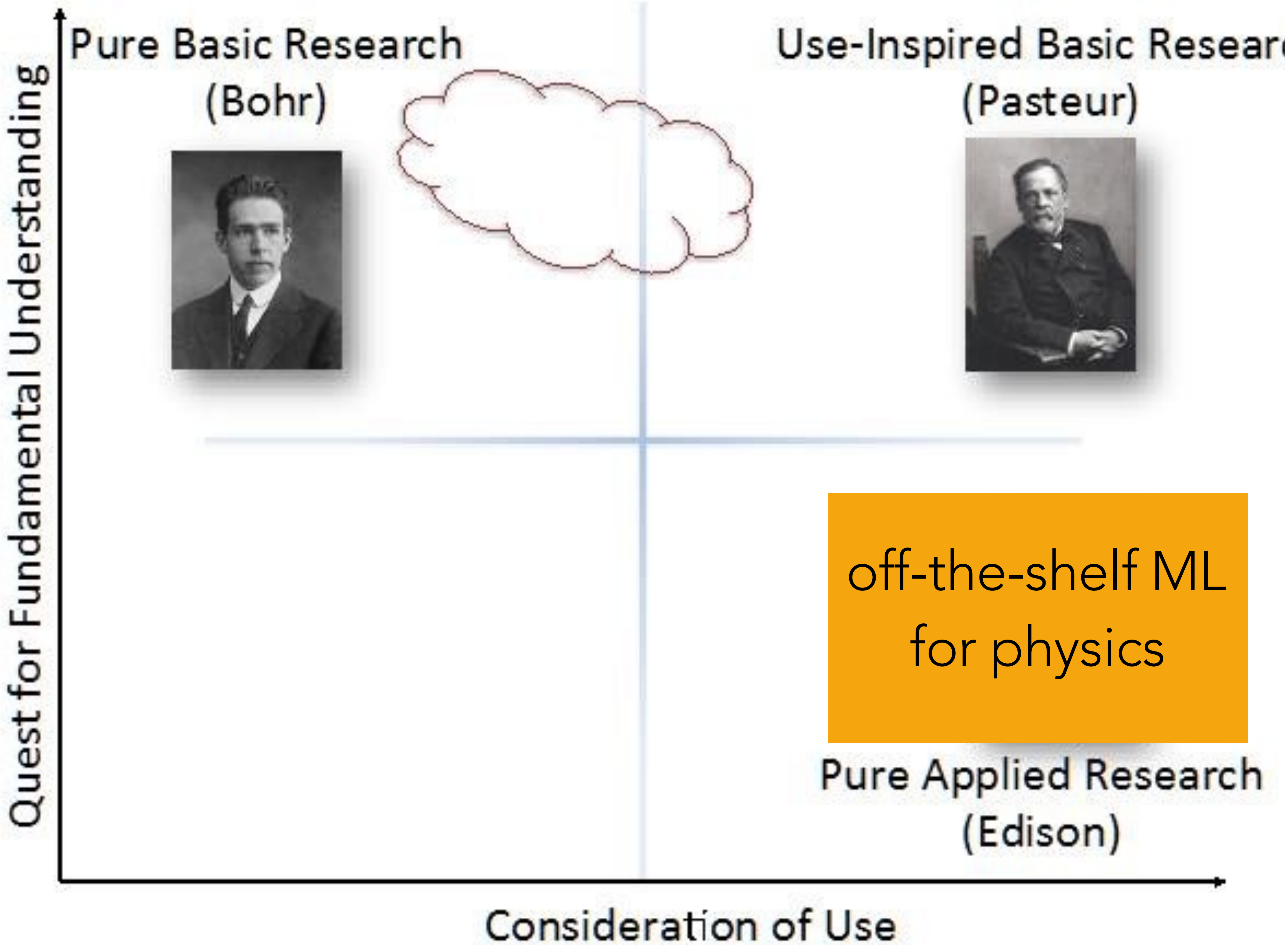
Pasteur's Quadrant & Use-inspired research



Donald Stokes, Pasteur's Quadrant

Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

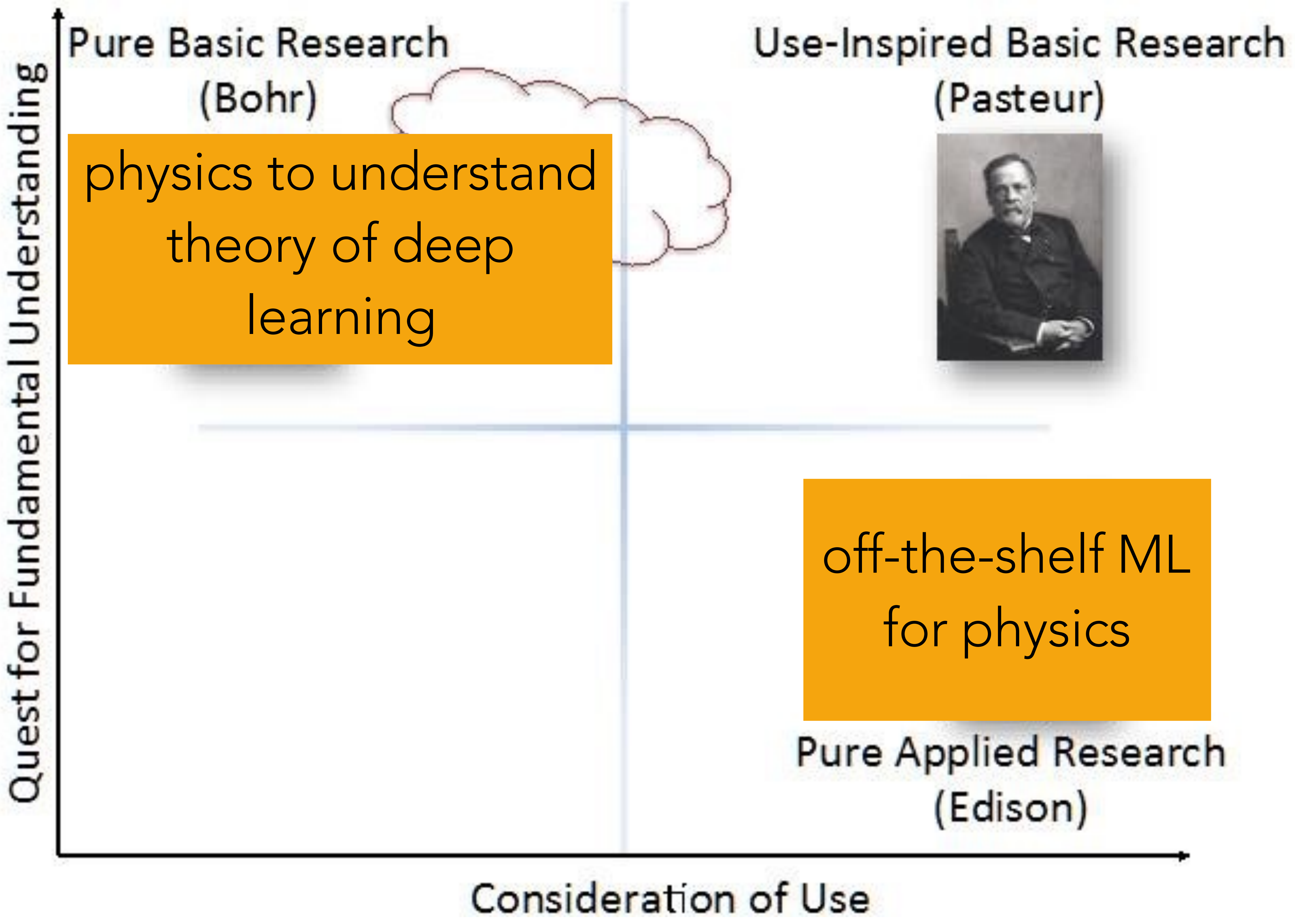
Pasteur's Quadrant & Use-inspired research



Donald Stokes, Pasteur's Quadrant

Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

Pasteur's Quadrant & Use-inspired research

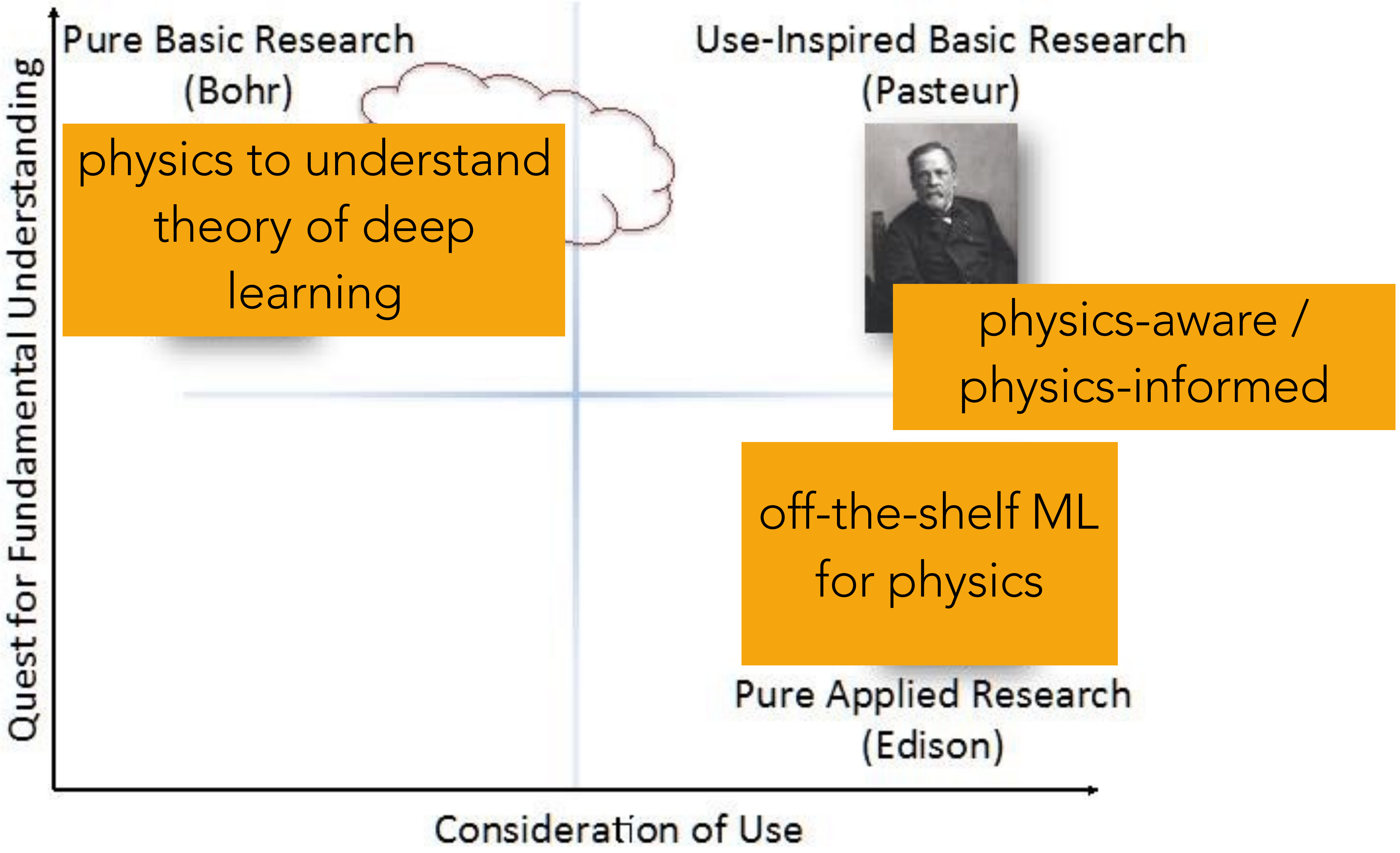


Donald Stokes, Pasteur's Quadrant

Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

Pasteur's Quadrant & Use-inspired research

Physics-aware / physics-informed incorporate physics concepts into ML model



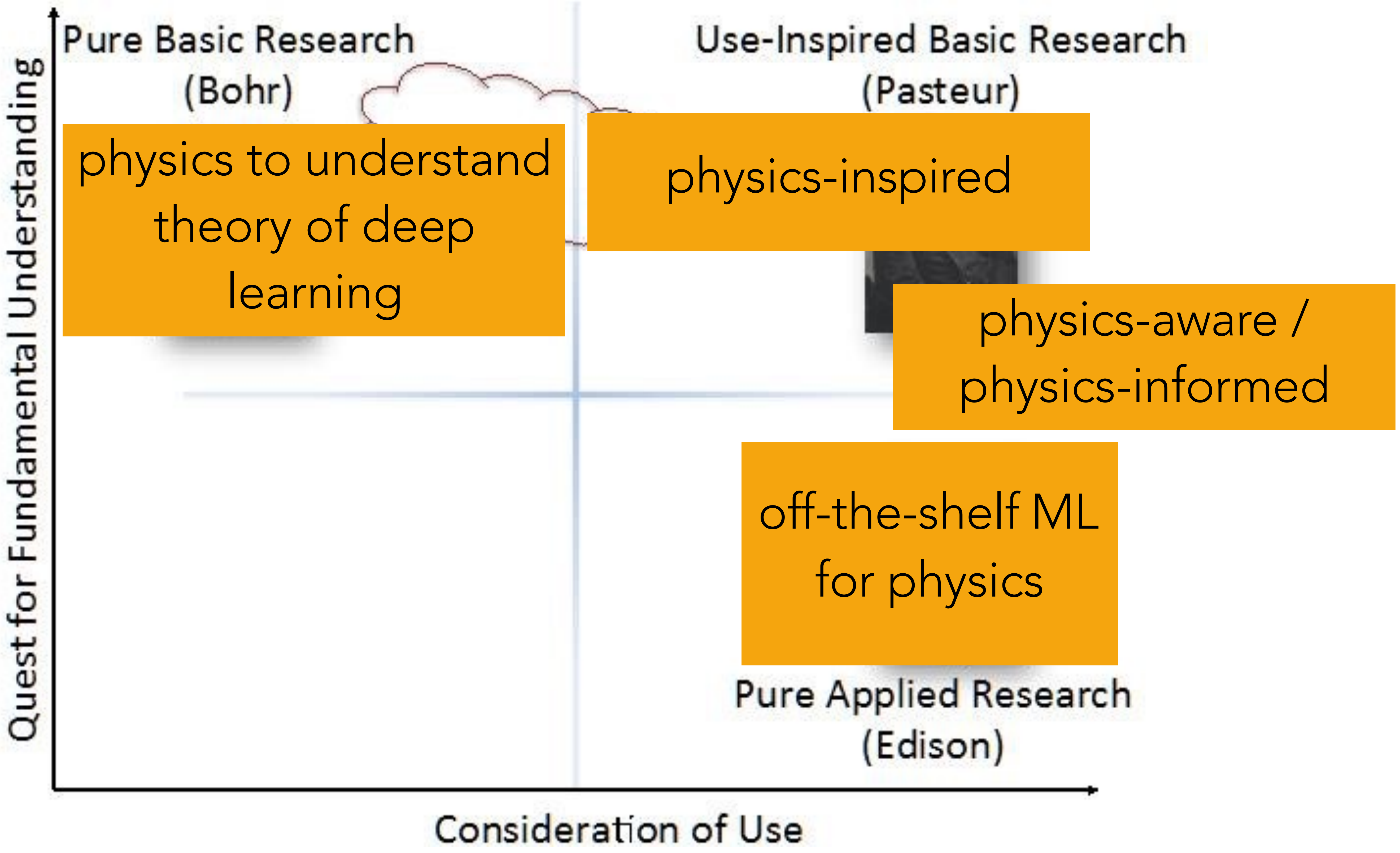
Donald Stokes, Pasteur's Quadrant

Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

Pasteur's Quadrant & Use-inspired research

Physics-aware / physics-informed incorporate physics concepts into ML model

Physics-inspired refers to more general techniques inspired by physics use-case

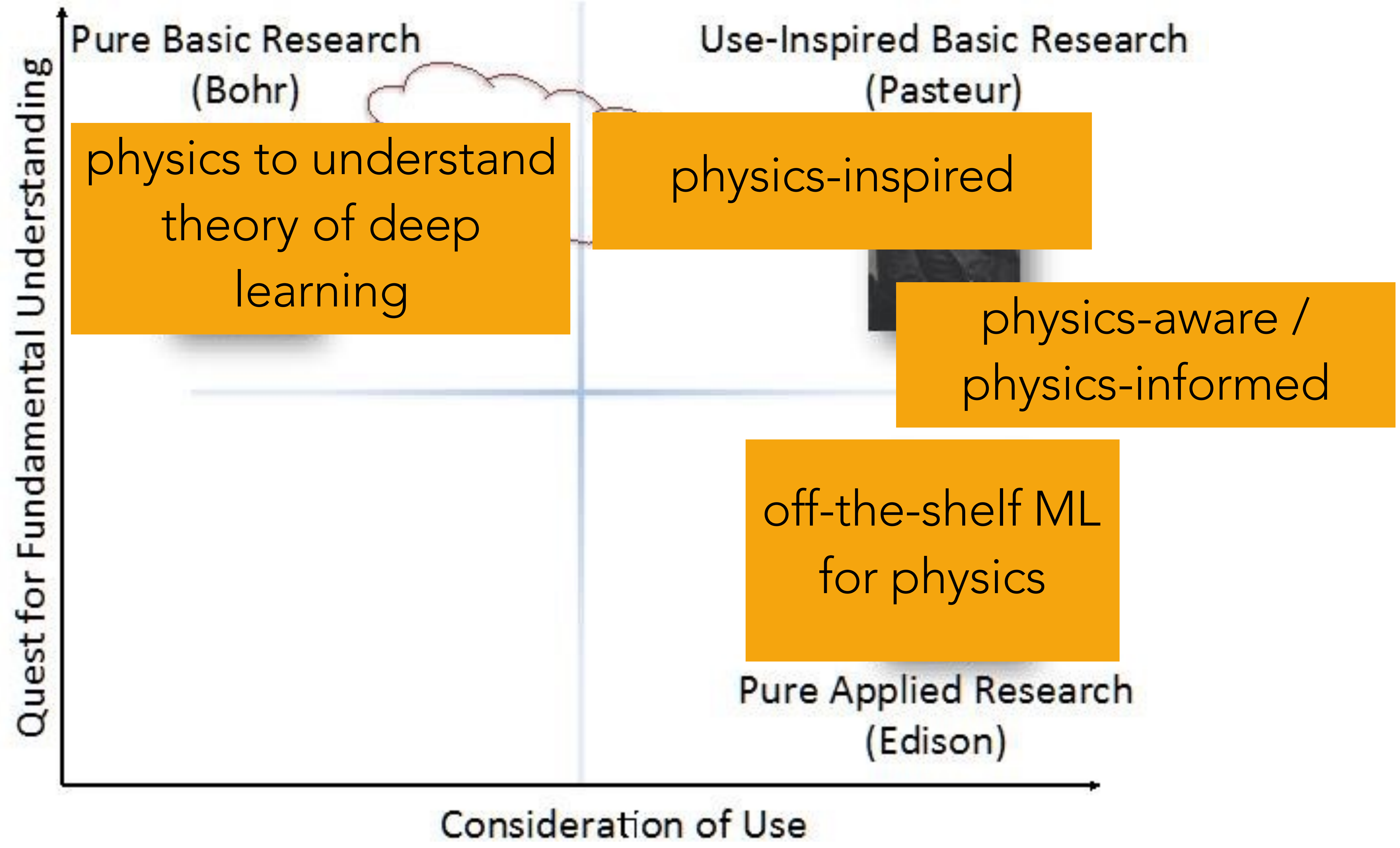


Donald Stokes, Pasteur's Quadrant

Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

Pasteur's Quadrant & Use-inspired research

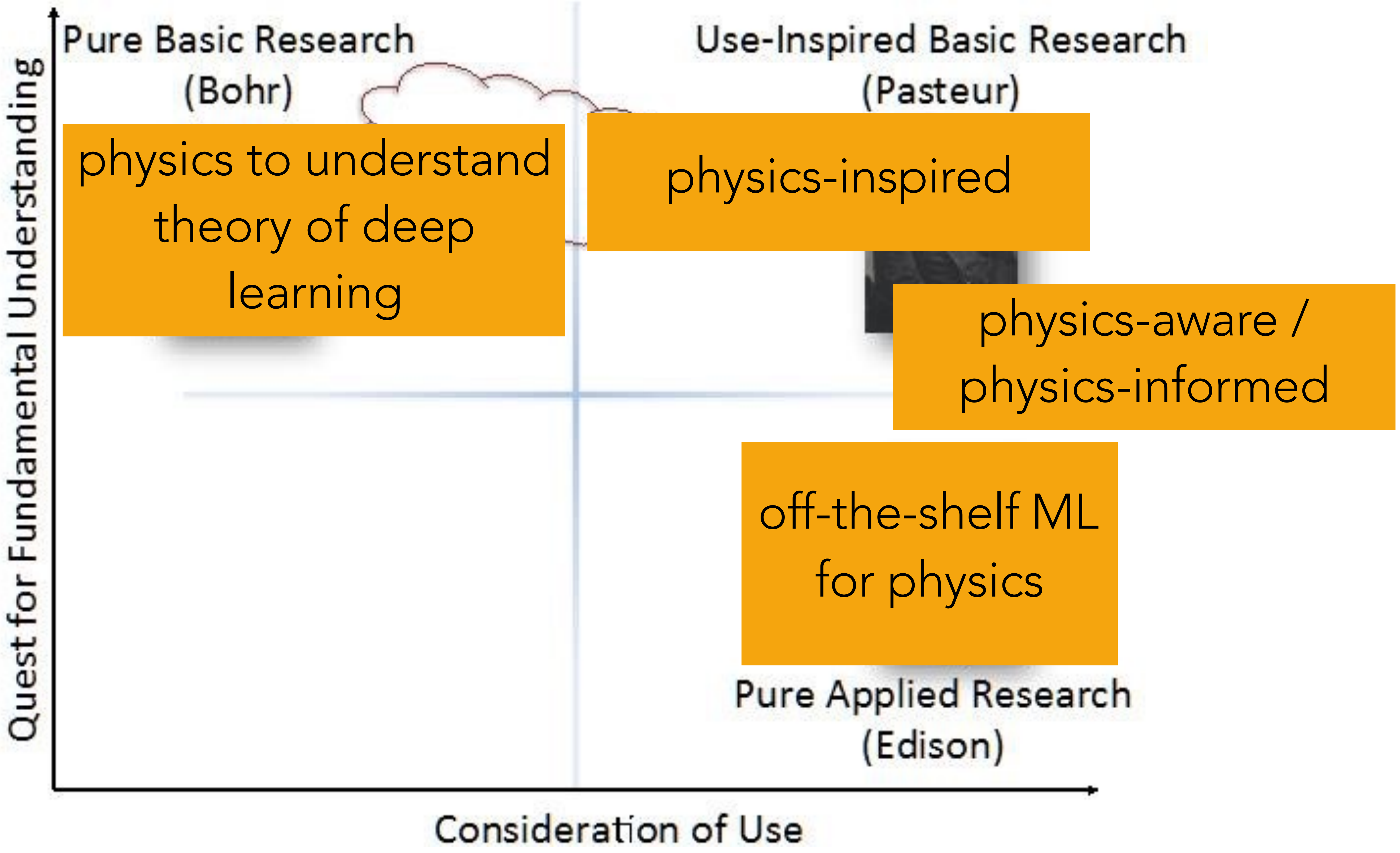
The process of developing **physics-aware / physics-informed** models often reveals more general **physics-inspired** abstractions, observations, and techniques.



Donald Stokes, Pasteur's Quadrant
Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

Pasteur's Quadrant & Use-inspired research

Though sometimes **physics-inspired** techniques come directly from the problem setting or formulation and are identified and abstracted early in the process



Donald Stokes, Pasteur's Quadrant

Source: NSF https://www.nsf.gov/mps/ad_blogs/quadrant_august2014.jsp

Take away

In addition to the personal joy and benefits to society of both basic and applied research, there is great value in exporting general purpose techniques and insight

- An intellectual form of “technology transfer”
- It behooves the field of physics to do this deliberately
  - May require adopting unfamiliar language, avoiding physics jargon, etc.
  - May require publishing in unfamiliar venues & recognizing value of those publications
  - Effective abstractions may not capture all aspects of the initial physics problem
- Aids in building collaborations with methodological researchers
- Avoids siloing & stagnation
- Facilitates importing good ideas from other fields
- Bolsters the real and perceived value of funding physics research

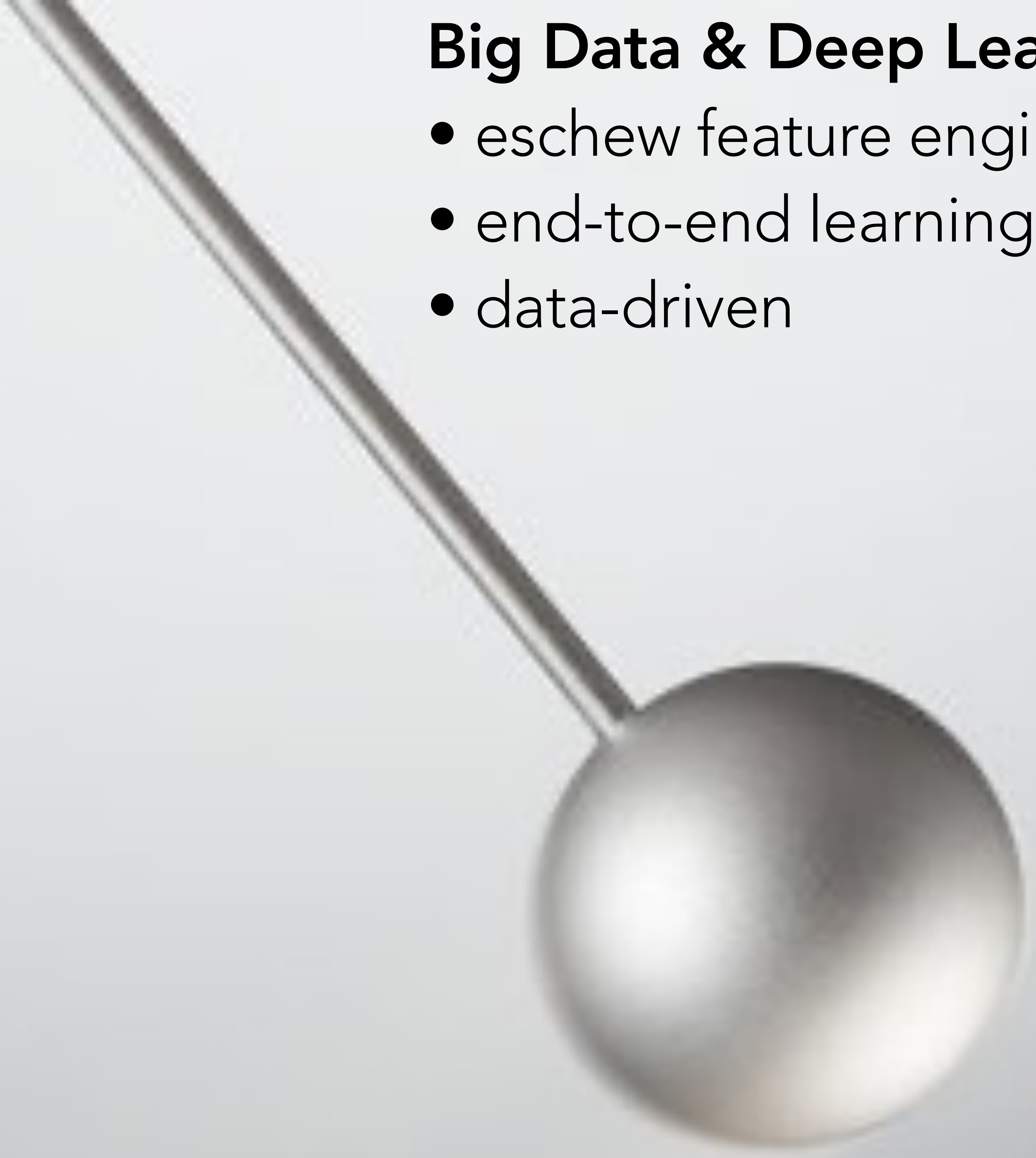
Impedance matching
with the ML community

Traditional approaches in physics

- hand-crafted data analysis
- largely guided by expert knowledge and theoretical insights

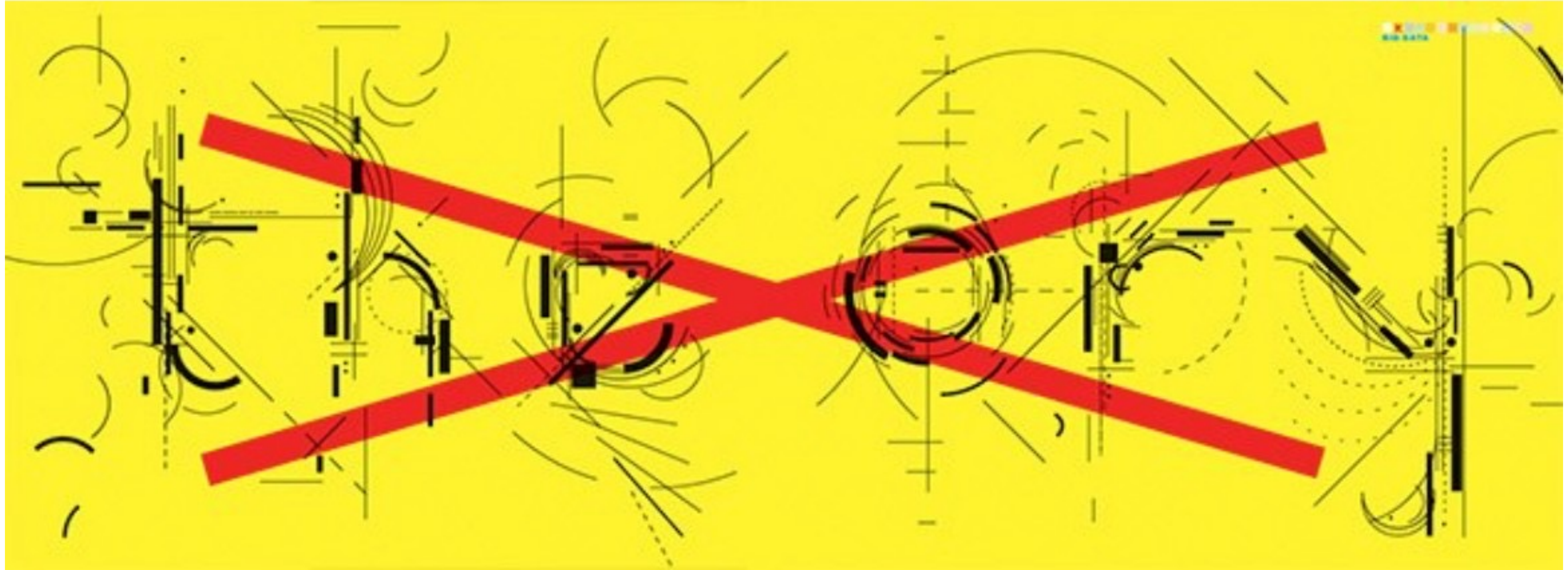
Big Data & Deep Learning

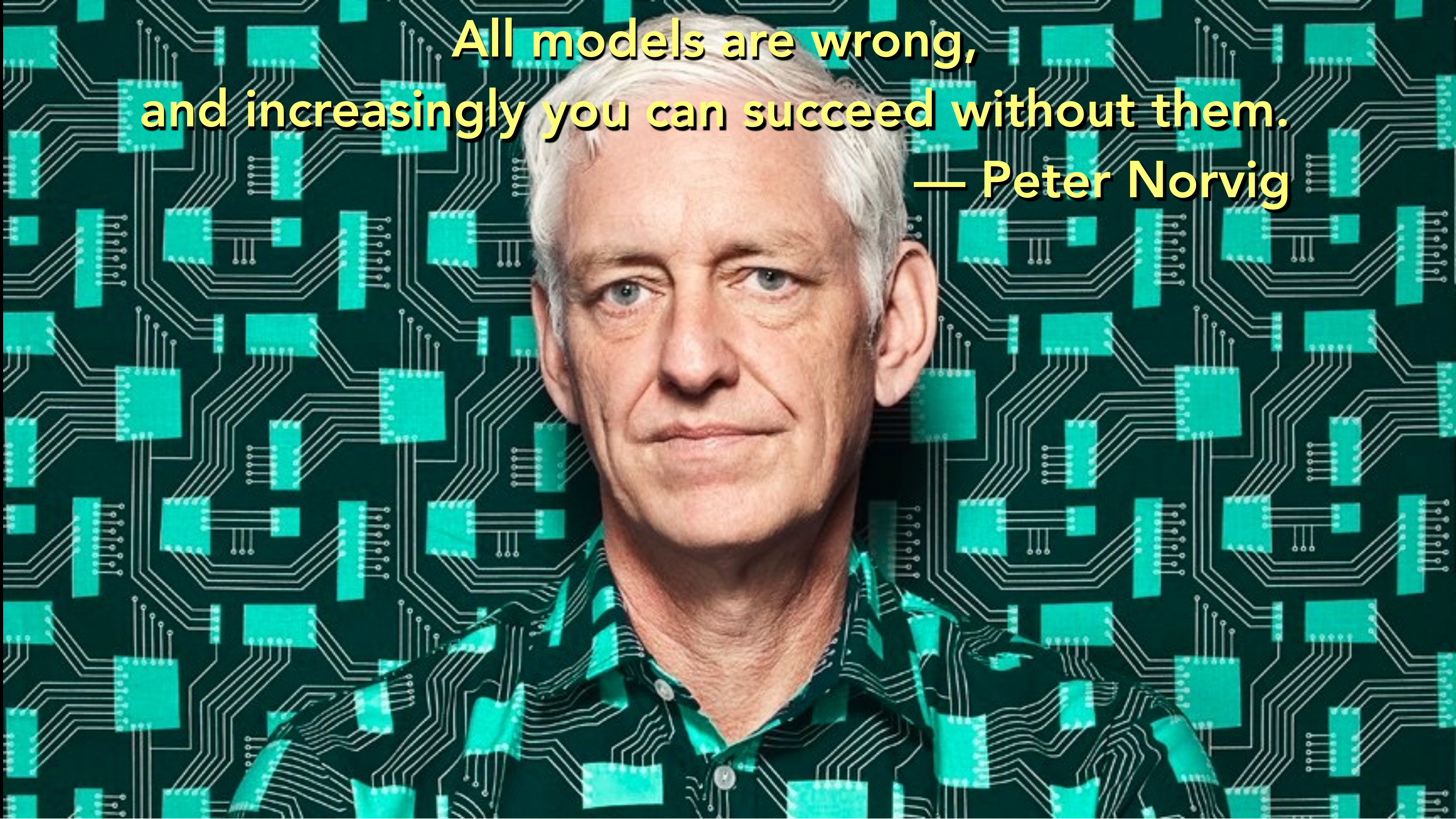
- eschew feature engineering
- end-to-end learning
- data-driven



CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE




All models are wrong,
and increasingly you can succeed without them.
— Peter Norvig

Hybrid approaches

- Some push back in the AI / ML community on data hungry end-to-end approaches

- Increased appreciation in the value of domain knowledge and "inductive bias"

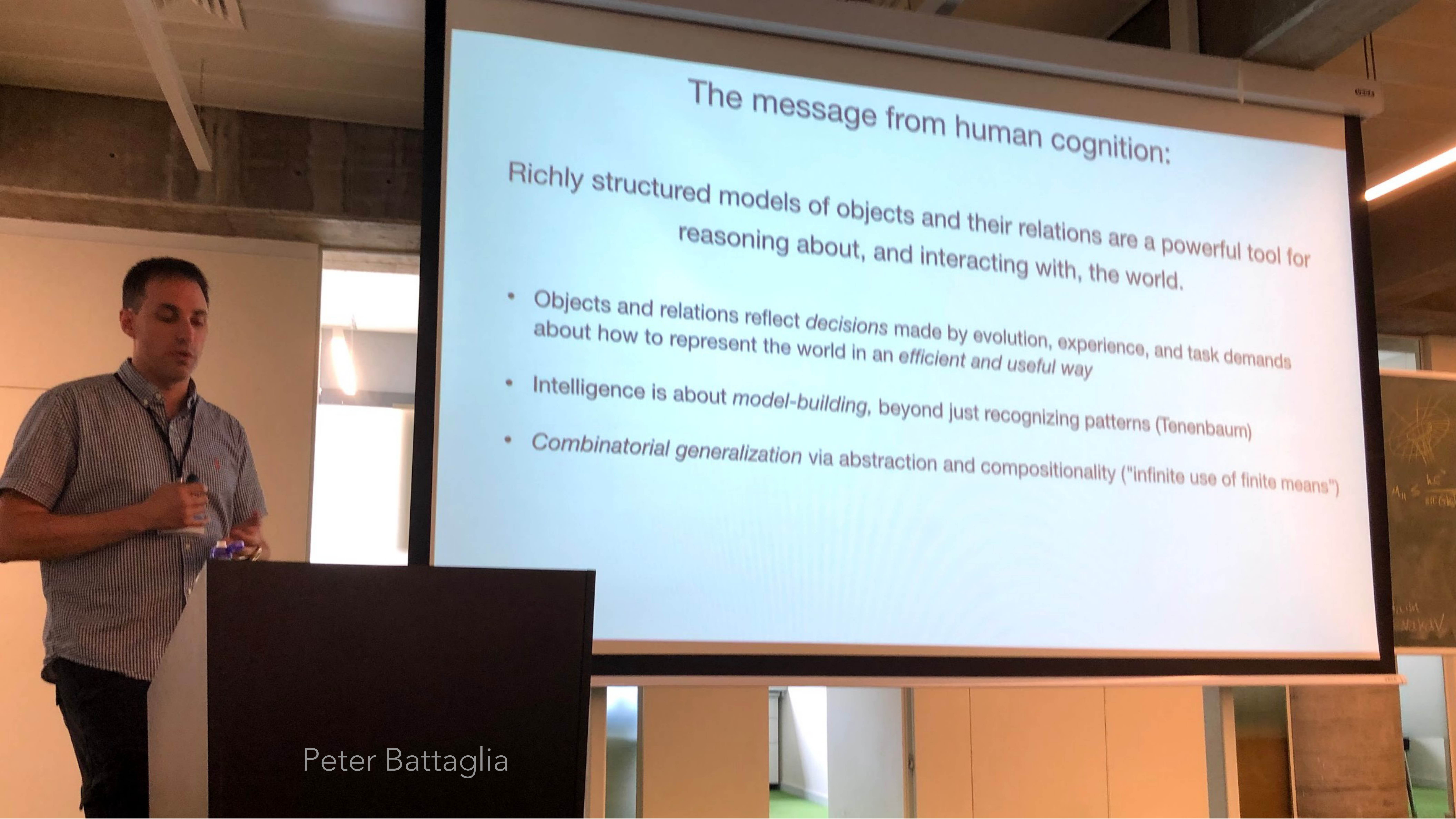
Physical Reasoning Workshop description

Much progress has been made on end-to-end learning for physical understanding and reasoning. If successful, **understanding and reasoning about the physical world** promises far-reaching applications in robotics, machine vision, and the **physical sciences**. Despite this recent progress, **our best artificial systems pale in comparison to the flexibility and generalization of human physical reasoning**.

Our workshop aims to investigate this broad questions:

- 1. What forms of **inductive biases** best enable the development of physical understanding techniques that are applicable to real-world problems?
- 2. How do we ensure that the outputs of a physical reasoning module are reasonable and **physically plausible**?
- 3. Is **interpretability** a necessity for physical understanding and reasoning techniques to be suitable to real-world problems?

Unlike end-to-end neural architectures that distribute bias across a large set of parameters, modern **structured physical reasoning modules** (differentiable physics, relational learning, probabilistic programming) maintain modularity and **physical interpretability**. We will discuss how these **inductive biases** might aid in generalization and interpretability, and how these techniques impact real-world problems.


The message from human cognition:
Richly structured models of objects and their relations are a powerful tool for reasoning about, and interacting with, the world.
• Objects and relations reflect *decisions* made by evolution, experience, and task demands about how to represent the world in an *efficient and useful way*
• Intelligence is about *model-building*, beyond just recognizing patterns (Tenenbaum)
• *Combinatorial generalization* via abstraction and compositionality ("infinite use of finite means")
Peter Battaglia

Insight of data generating process informs inductive bias on architecture

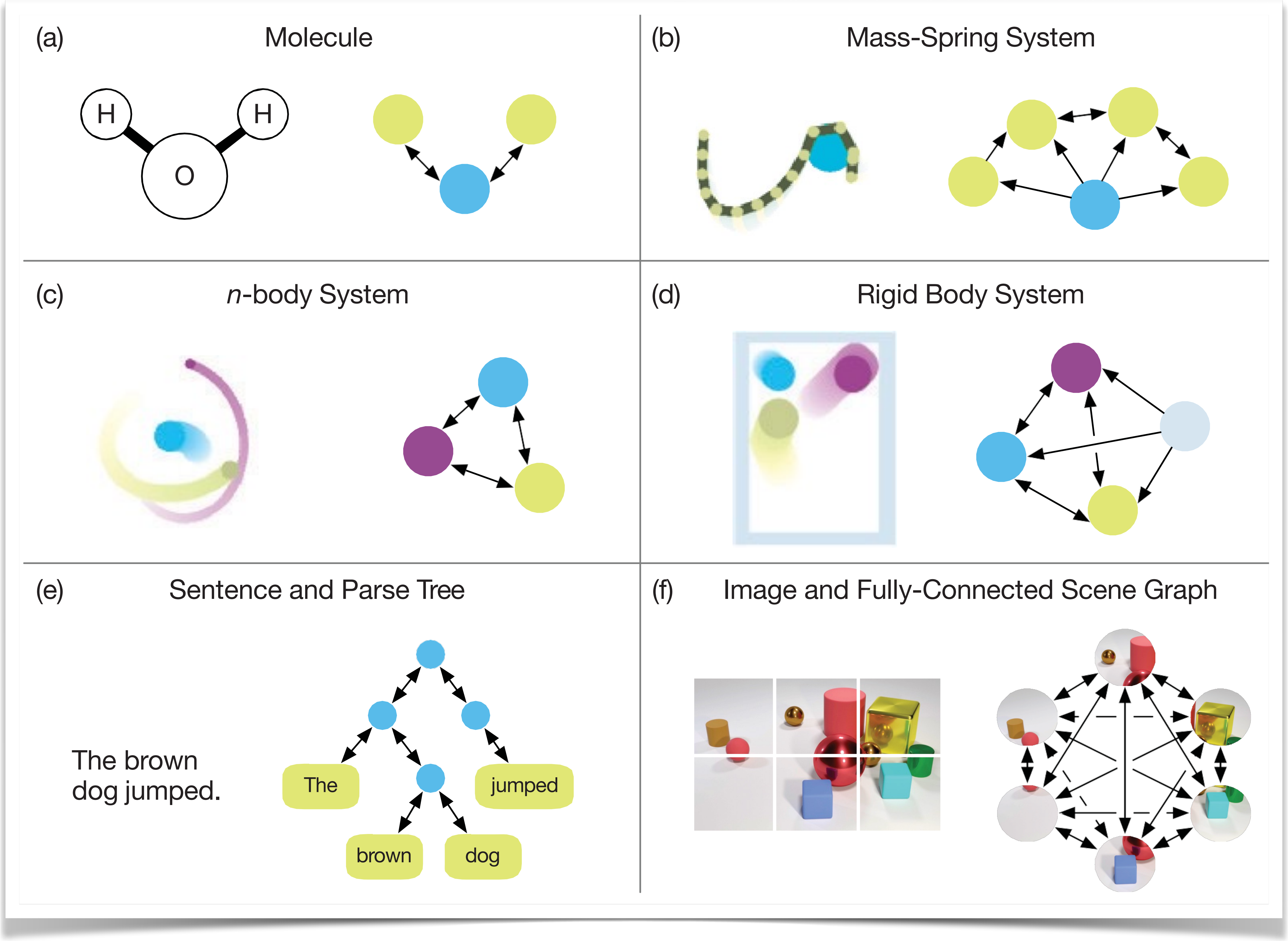
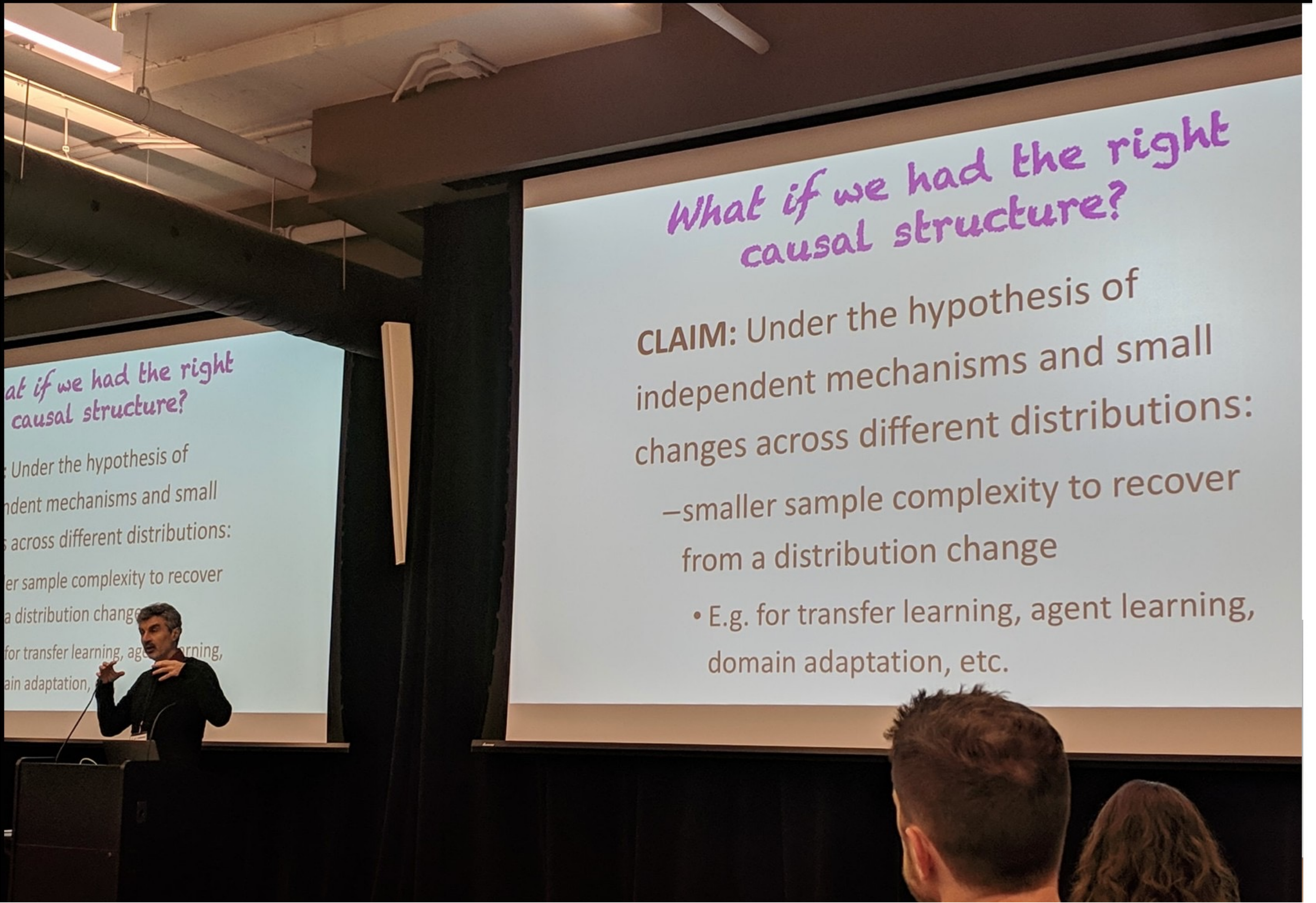


Image credit: Battaglia, et. al. arXiv:1806.01261



Max Welling Isn't this what Bernhard Schoelkopf has been saying for a while?

Like · Reply · 6w

Yann LeCun ...and Leon Bottou ?

Like · Reply · 6w

Leon Bottou Yoshua's paper says: if you observe a distribution change that comes from a causal effect, then you'll adapt faster if your generative model matches the causal model.

Another way of seeing it is : the right causal graph suggests a particular factorization of the joint distribution (a directed bayesian network). A causal intervention means that you only change one of these factors (or a few factors) while leaving the other ones unchanged. Therefore if your generative model is the right causal model, meaning that it factorizes the joint in the same way, it will be easy to adapt it to the change because only a few parameters need changing (those associated with the factors that actually changed).

Max Welling Dan Roy I am, and I think most of us, are keenly aware that Josh has been the big proponent of this view. And I think most people agree with him on this view. Integrating this view with deep learning for more narrowly defined tasks seems to me an interesting intellectual pursuit though. I think that's what's happening here but I was not at the talk 😉

World Models

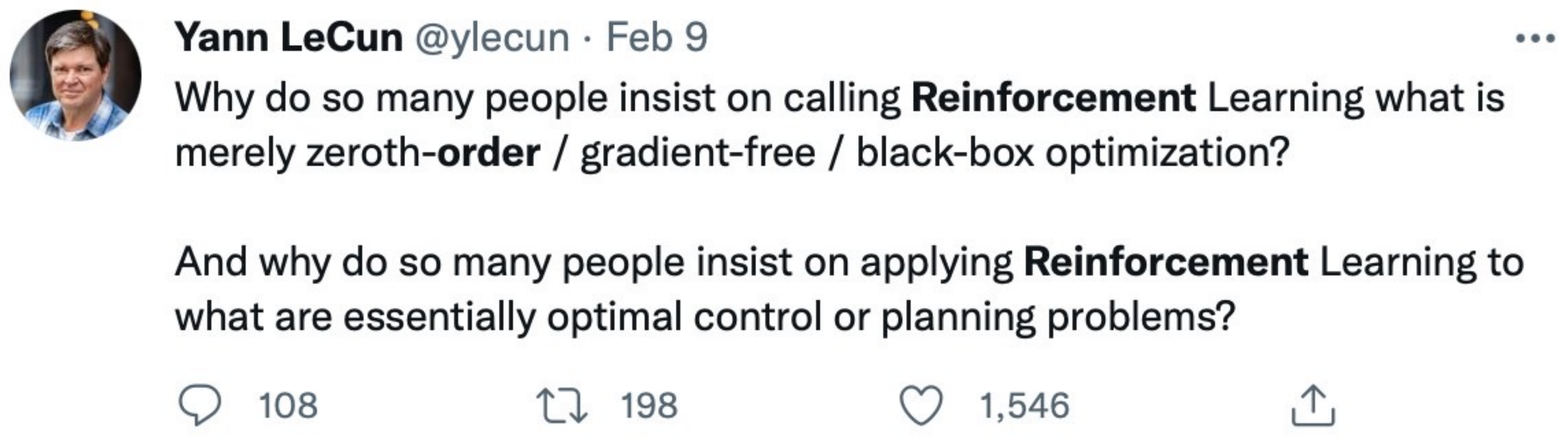
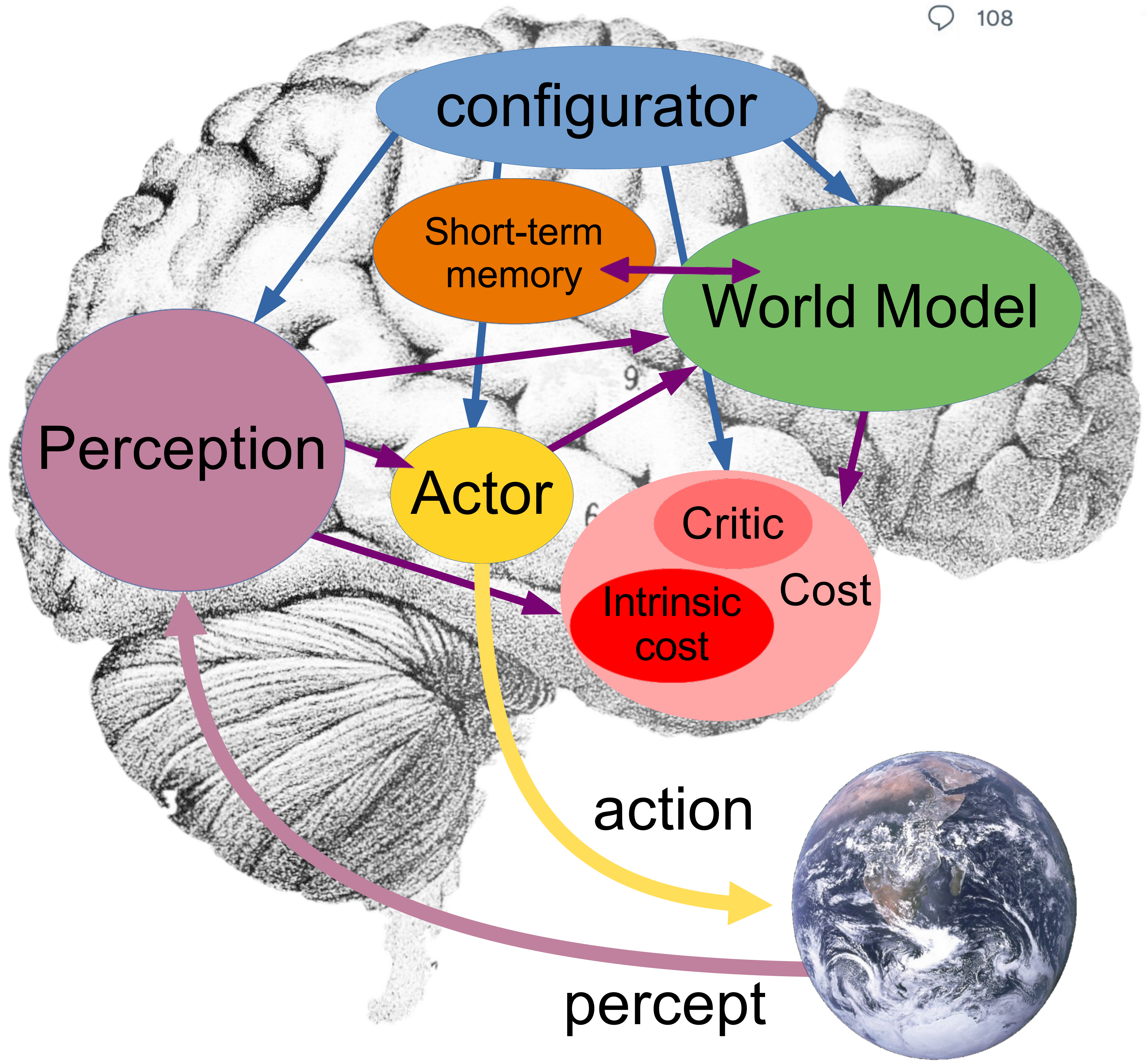


Figure from Yann LeCun



Model for animal-like intelligence

- Somewhat like RL with sequential decision making in real world, but …
- Importance of "world model" to predict potential outcomes of candidate actions
- World model used for planning & control

Inductive Bias

Compositionality

Relationships

Symmetry

Causality

Hybrid "physics-aware" approaches to AI / ML

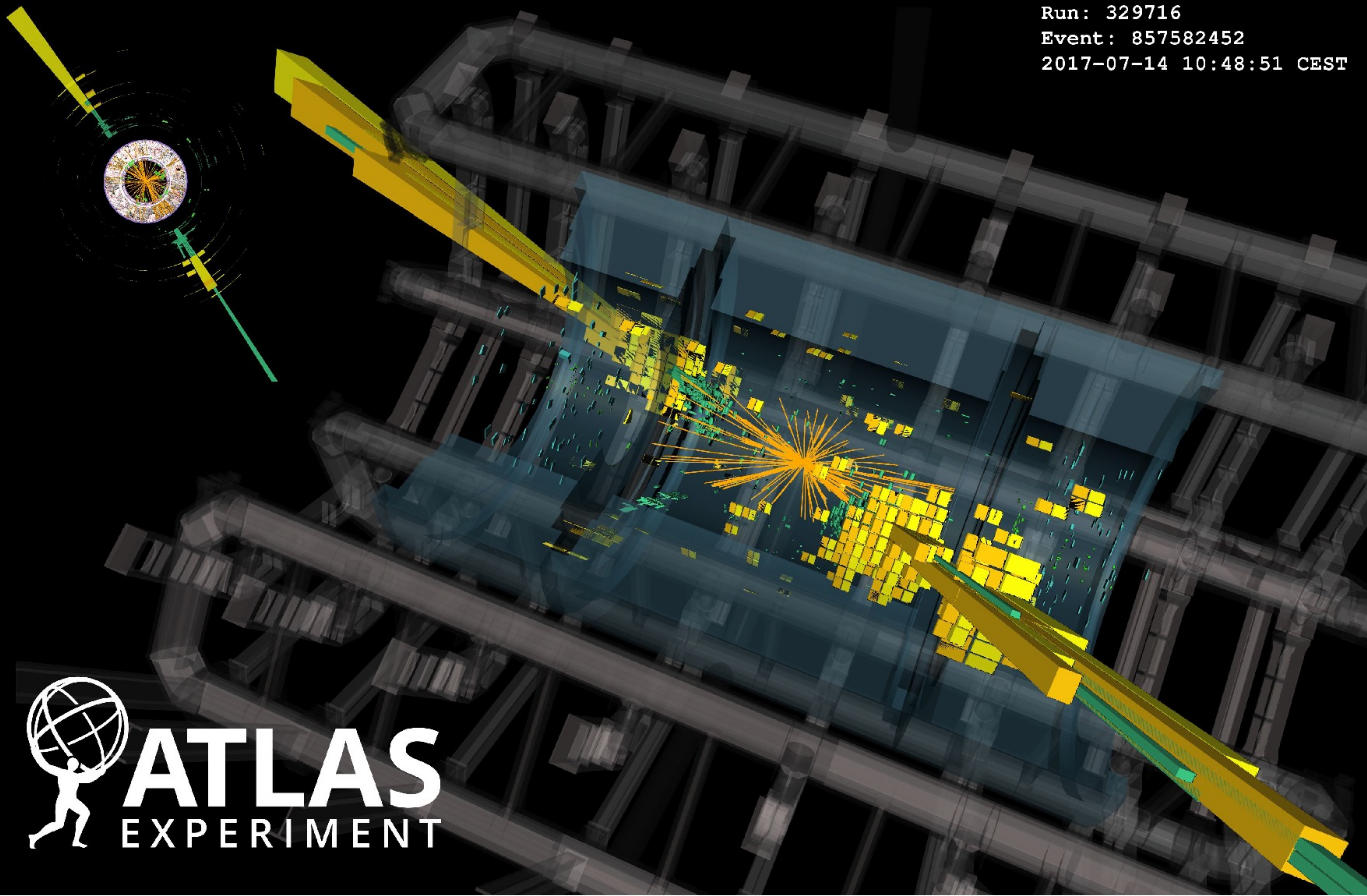
- inject knowledge of data generating process into inductive bias of ML models

- use "black box" style ML components to model the most uncertain aspects of problem where traditional approaches make overly restrictive assumptions

- align architectural components of ML models with causal mechanism

An example of physics-aware ML

JETS

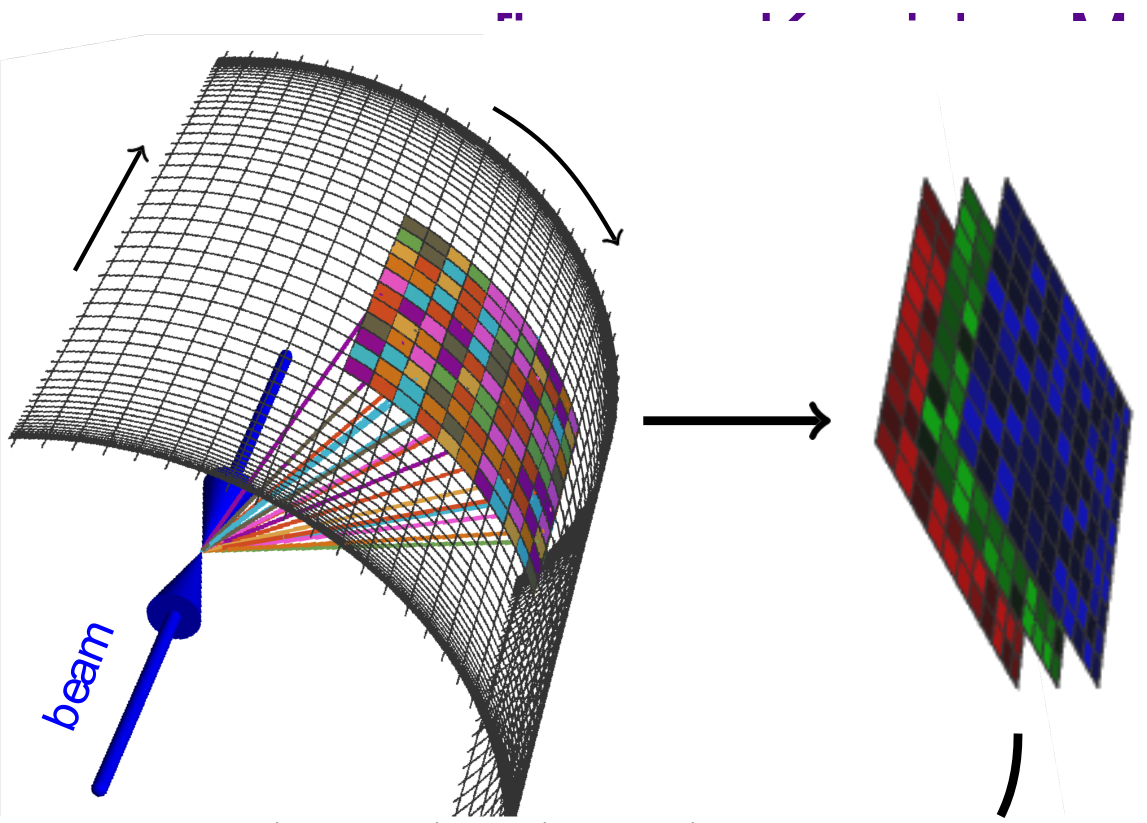

Run: 329716
Event: 857582452
2017-07-14 10:48:51 CEST
ATLAS
EXPERIMENT

Jet Images

Oliveira, et. al arXiv:1511.05190
Whiteson, et al arXiv:1603.09349
Barnard, et al arXiv:1609.00607

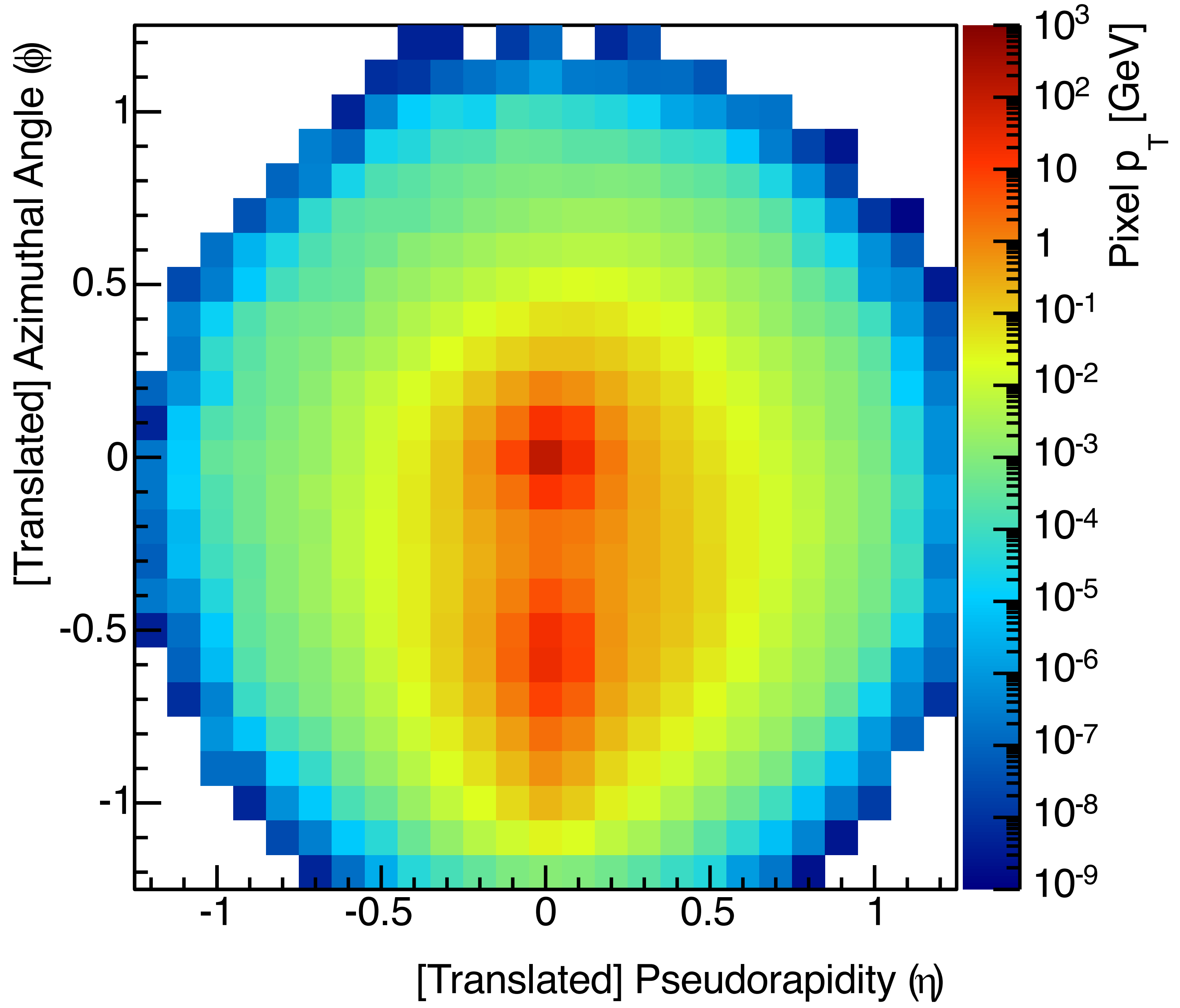
Deep learning algorithms first applied (~2015) to “jet images” for a binary classification task

- based on idealized uniform detector
- ~off-the-shelf use of convolutional neural networks at first

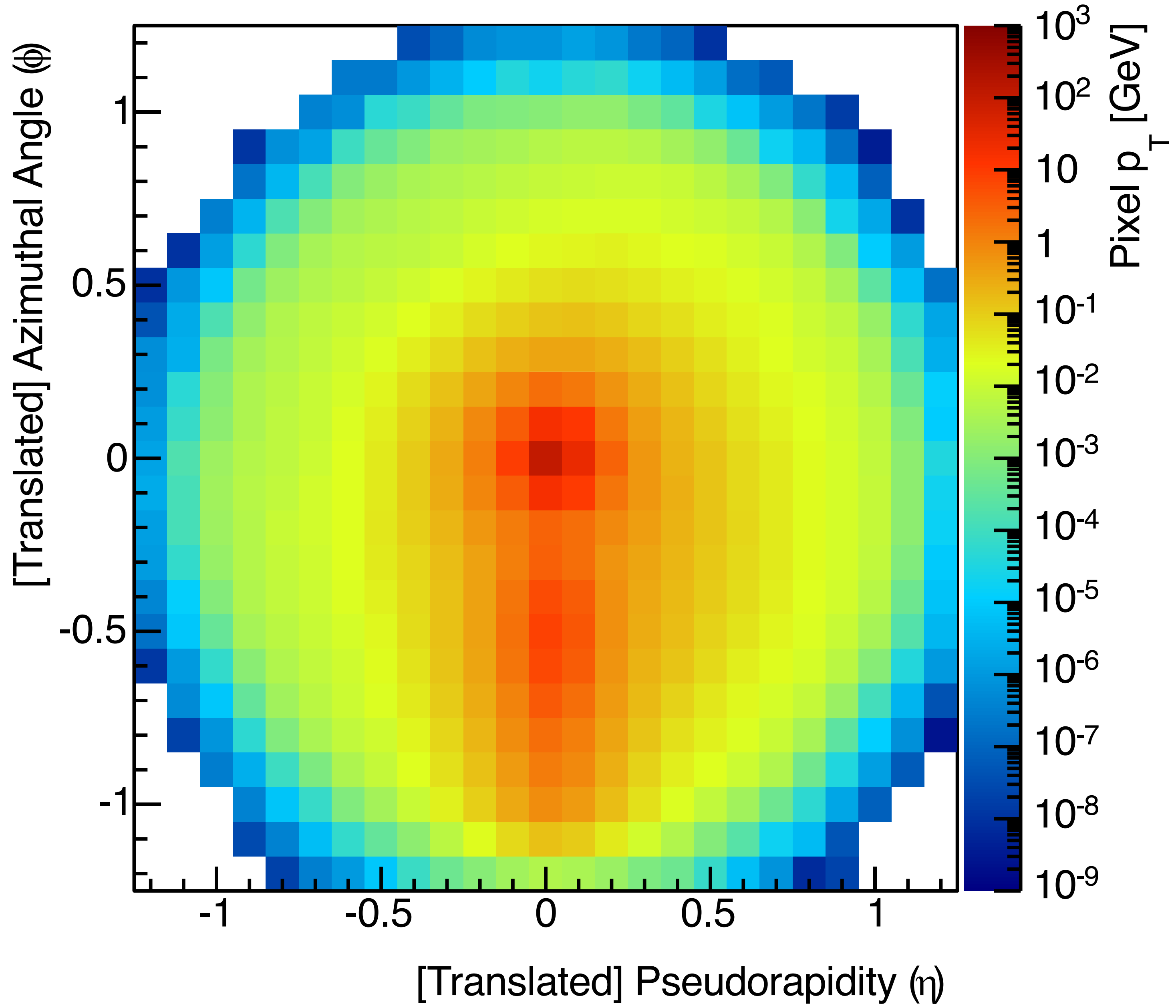


[image: Komiske, Metodiev, Schwartz arkiv:1612.01551]

Average Boosted W Jet

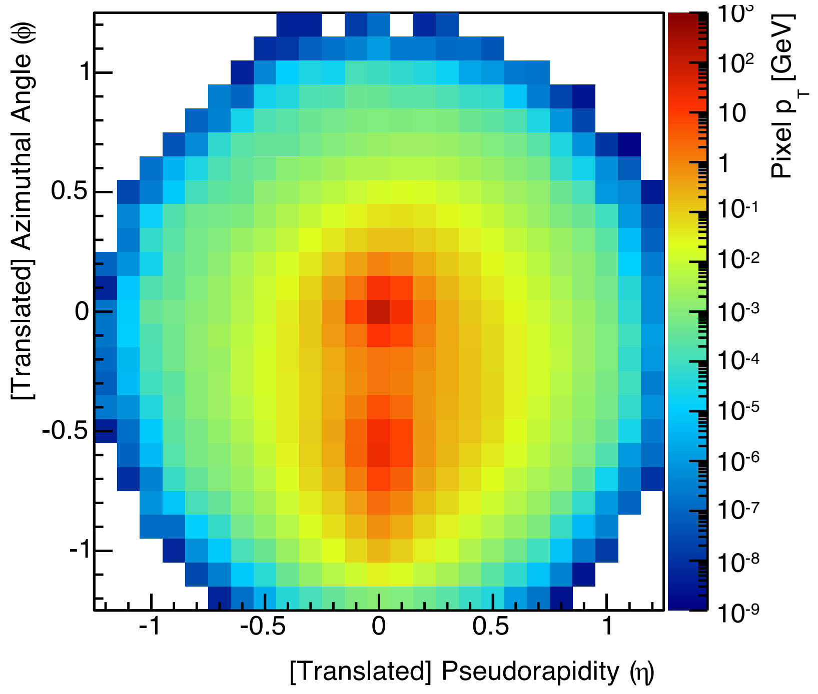


Average QCD Jet

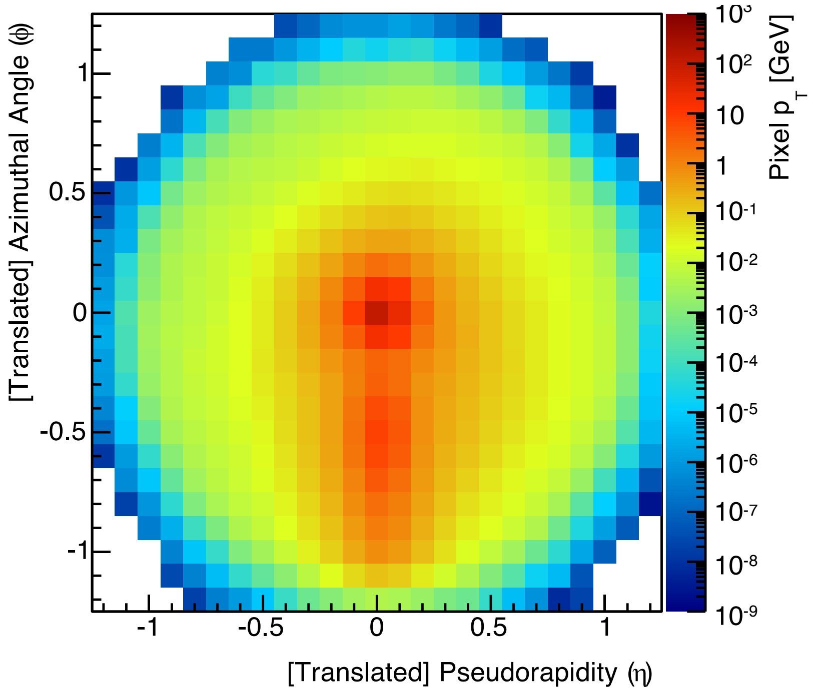


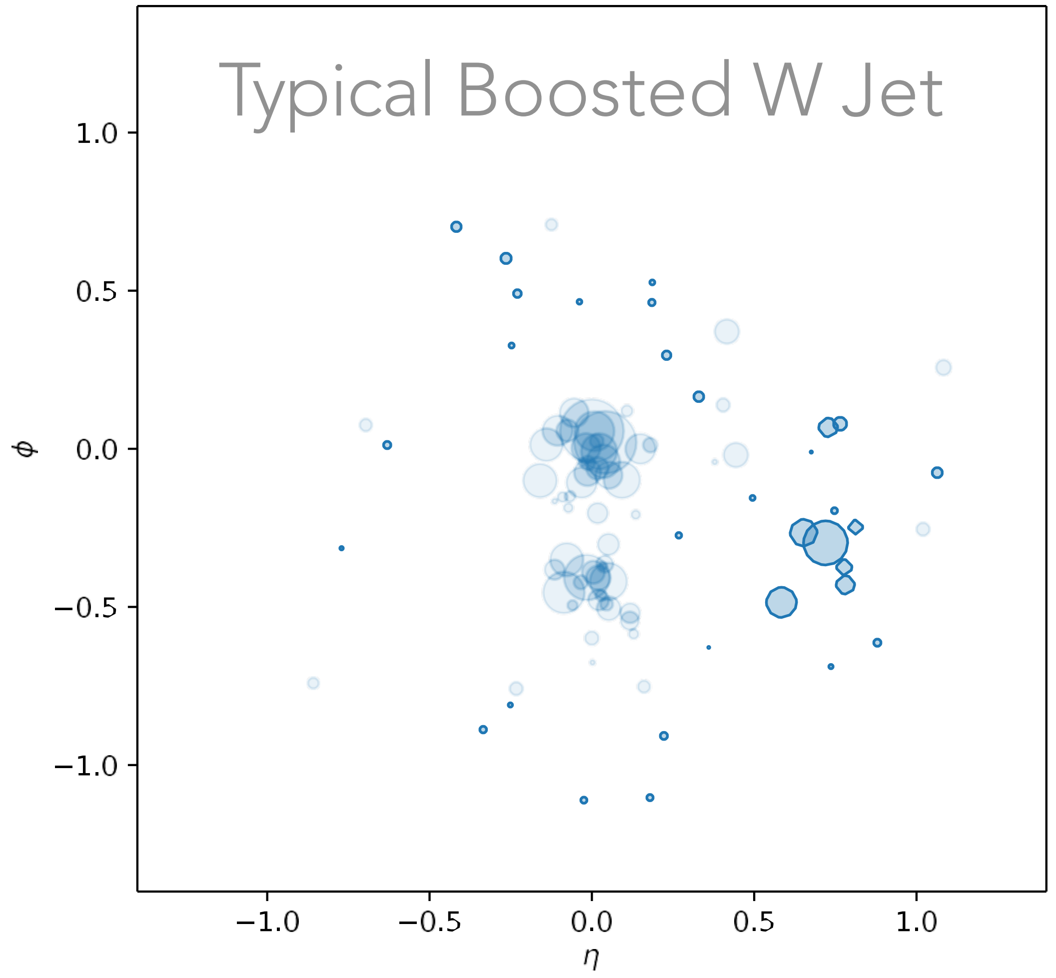
The image representation is sparse

Average Boosted W Jet

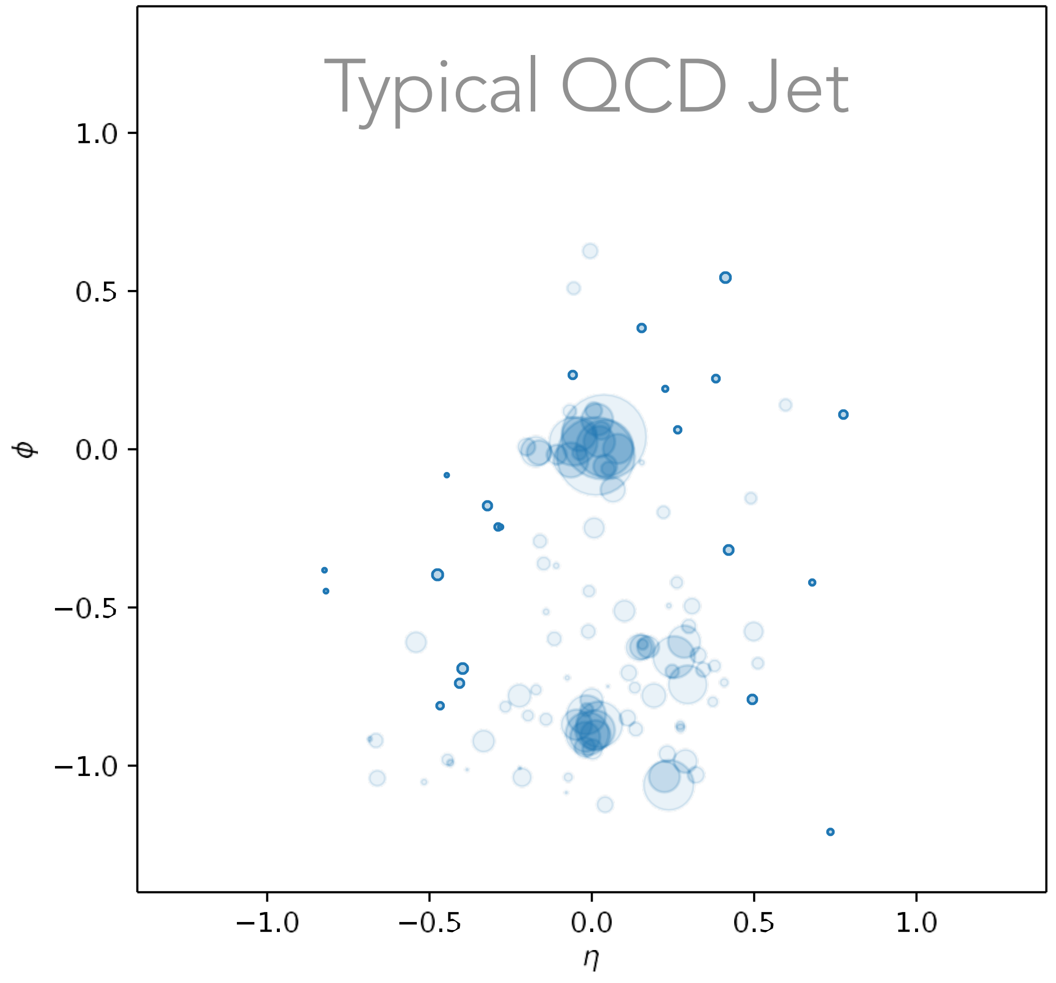


Average QCD Jet

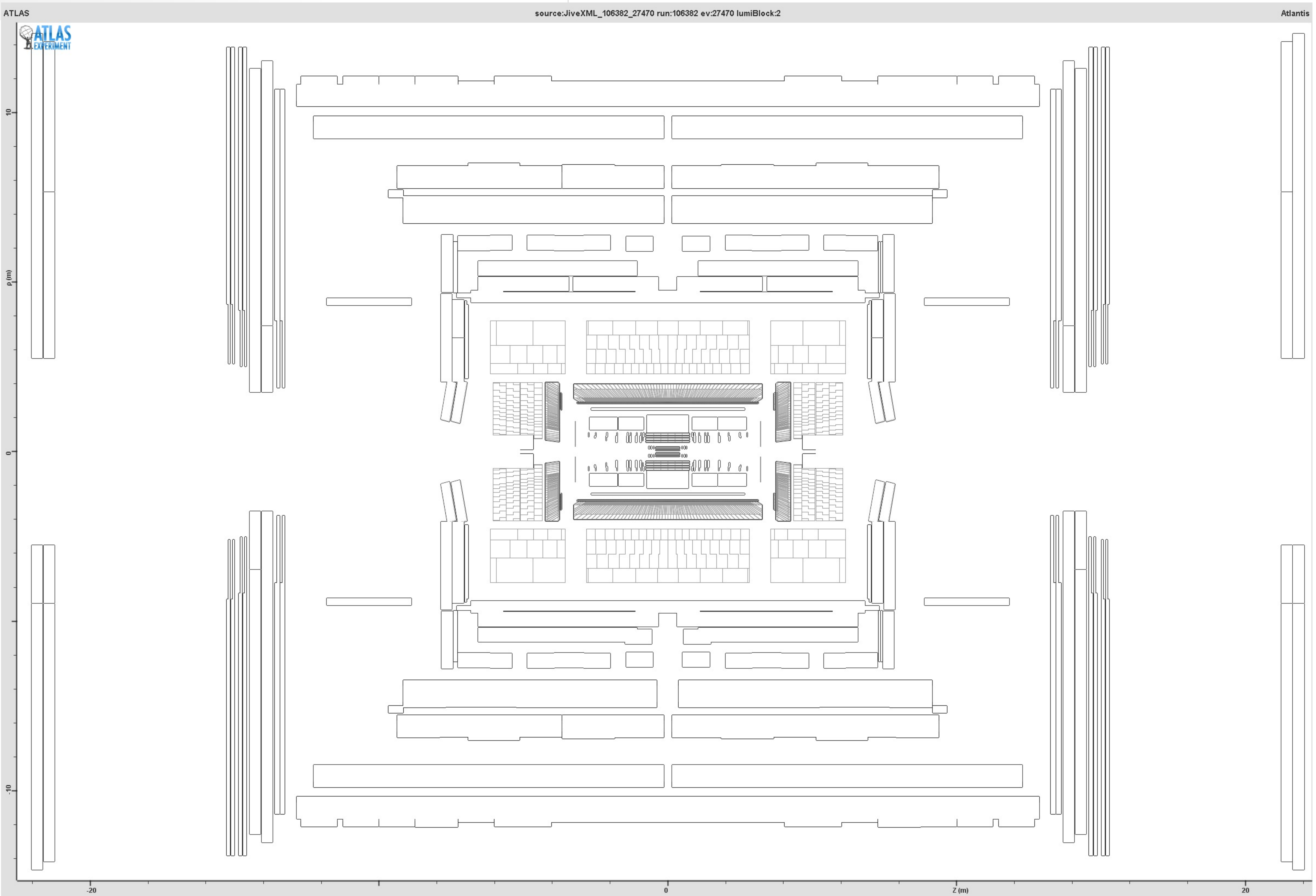




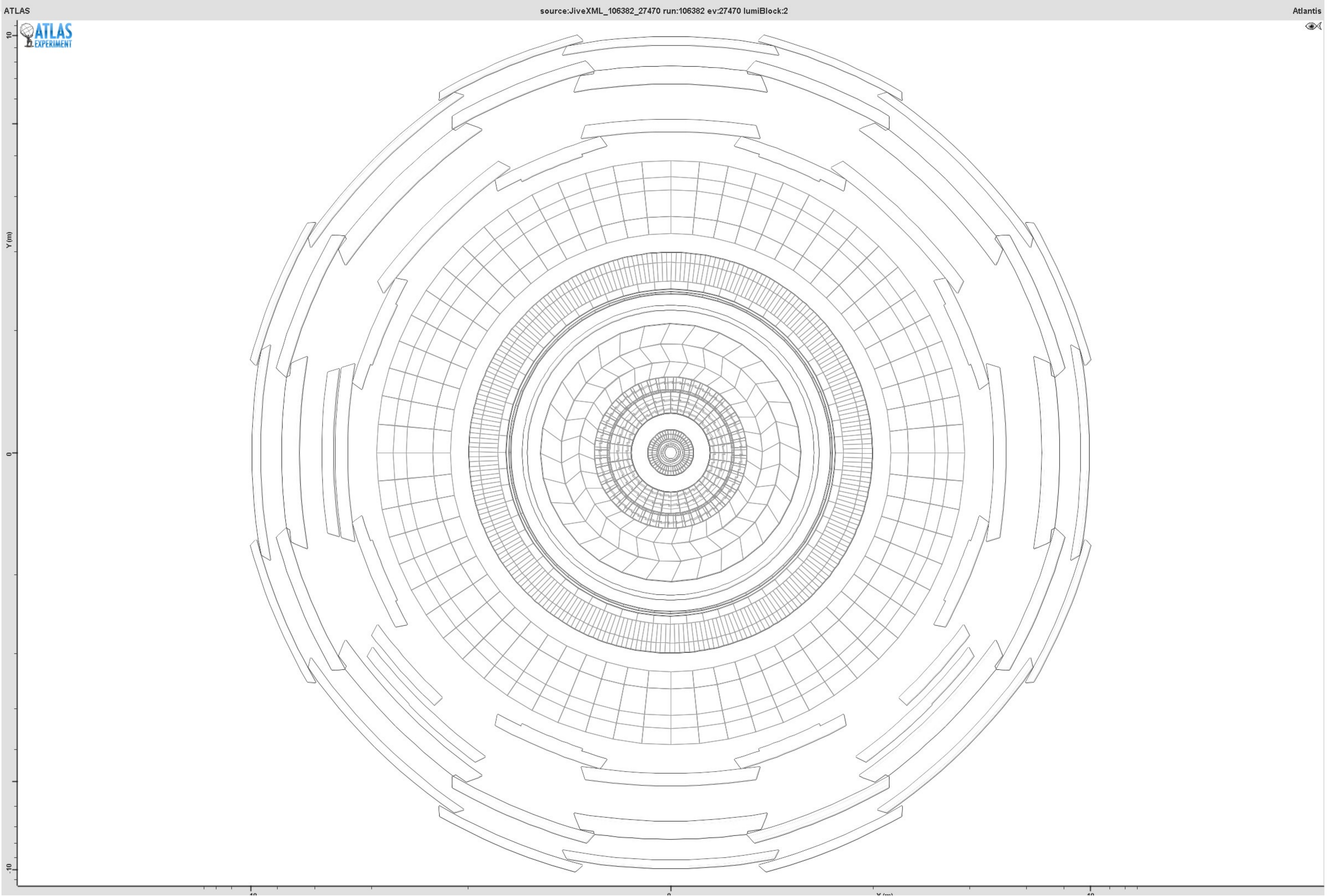
← point clouds →



Non-uniform geometry



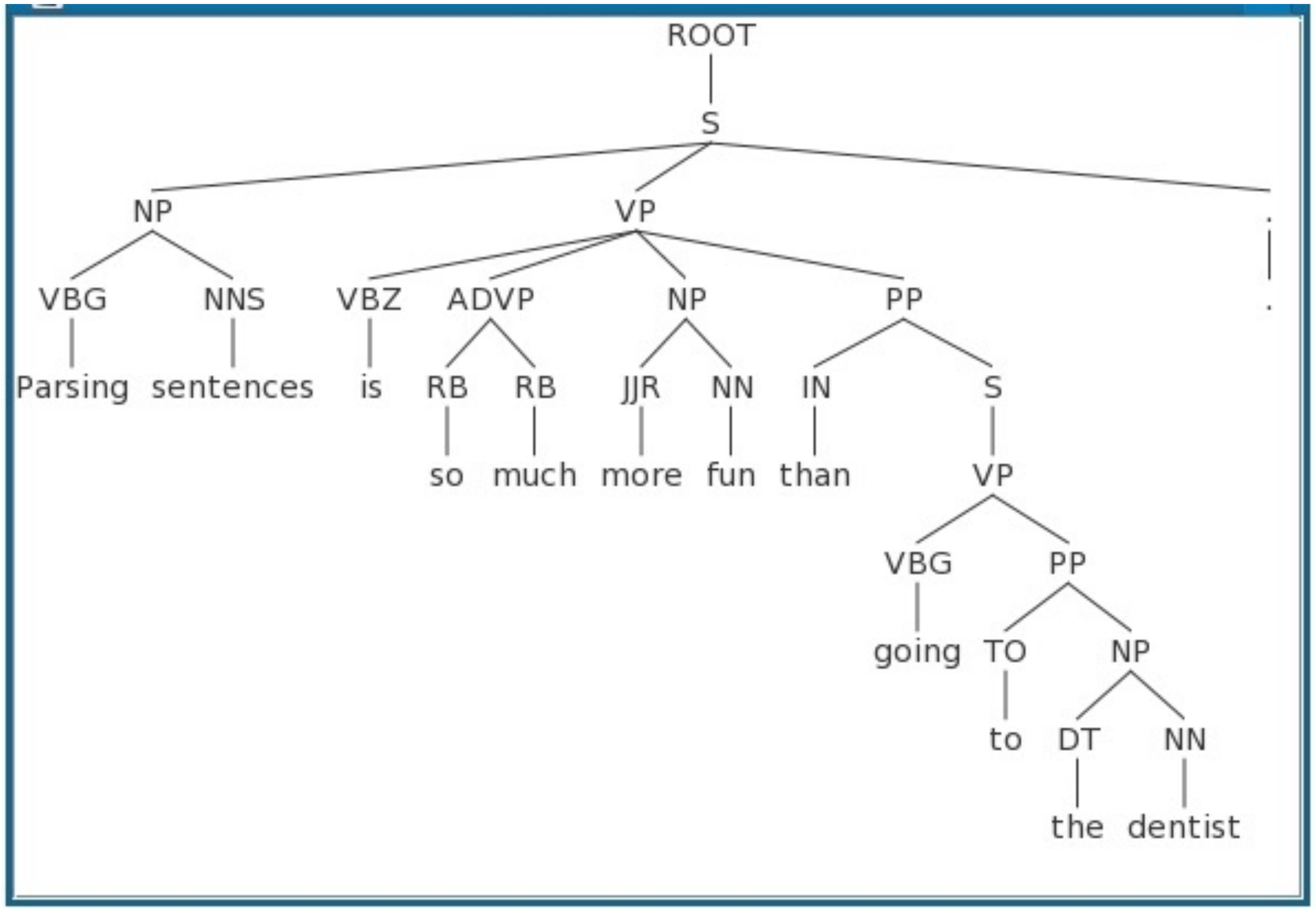
Non-uniform geometry



From images to Sentences (~2016)

Recursive Neural Networks showing great performance for Natural Language Processing tasks

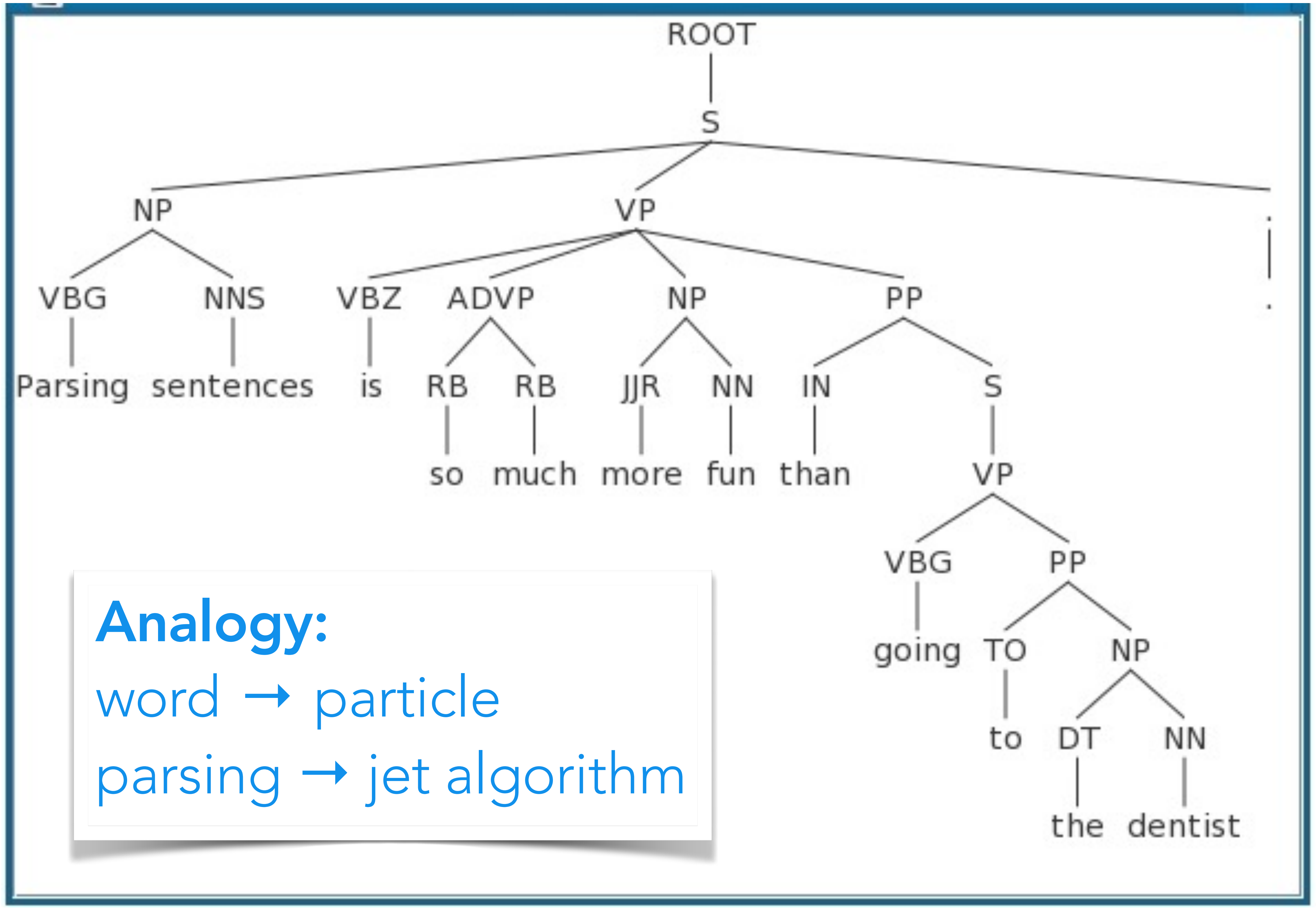
- neural network's topology given by parsing of sentence!



From images to Sentences (~2016)

Recursive Neural Networks showing great performance for Natural Language Processing tasks

- neural network's topology given by parsing of sentence!

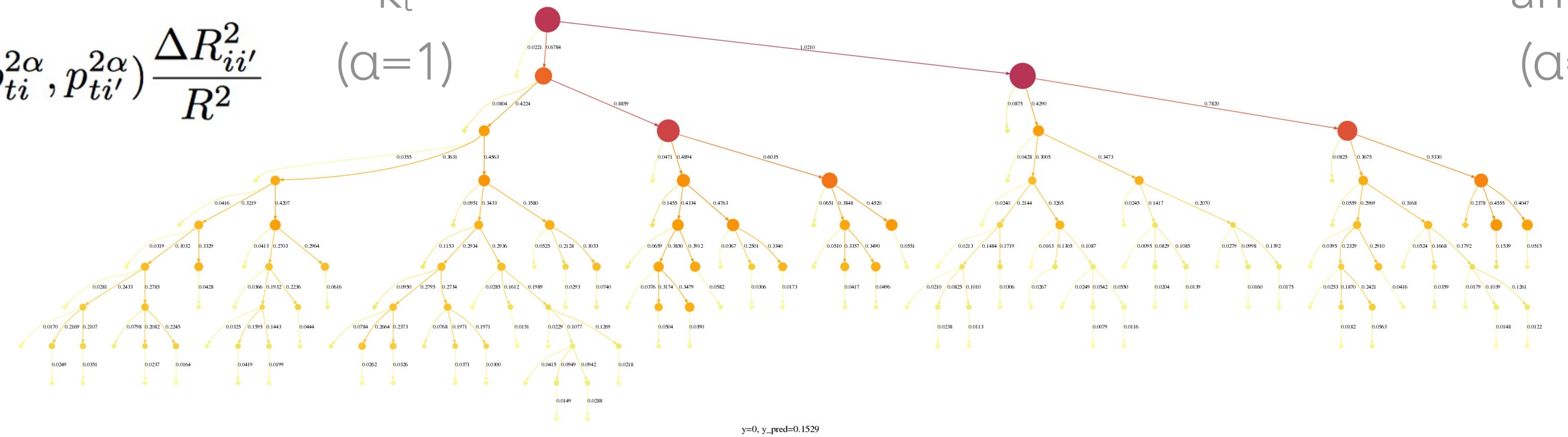


Analogy:
word → particle
parsing → jet algorithm

QCD-inspired Recursive Neural Networks (~2016)

$$d_{ii'}^{\alpha} = \min(p_{ti}^{2\alpha}, p_{ti'}^{2\alpha})\frac{\Delta R_{ii'}^2}{R^2}$$

k_t (α=1)



anti-k_t (α=1)

Work with Gilles Louppe, Kyunghyun Cho, Cyril Becot in 2017

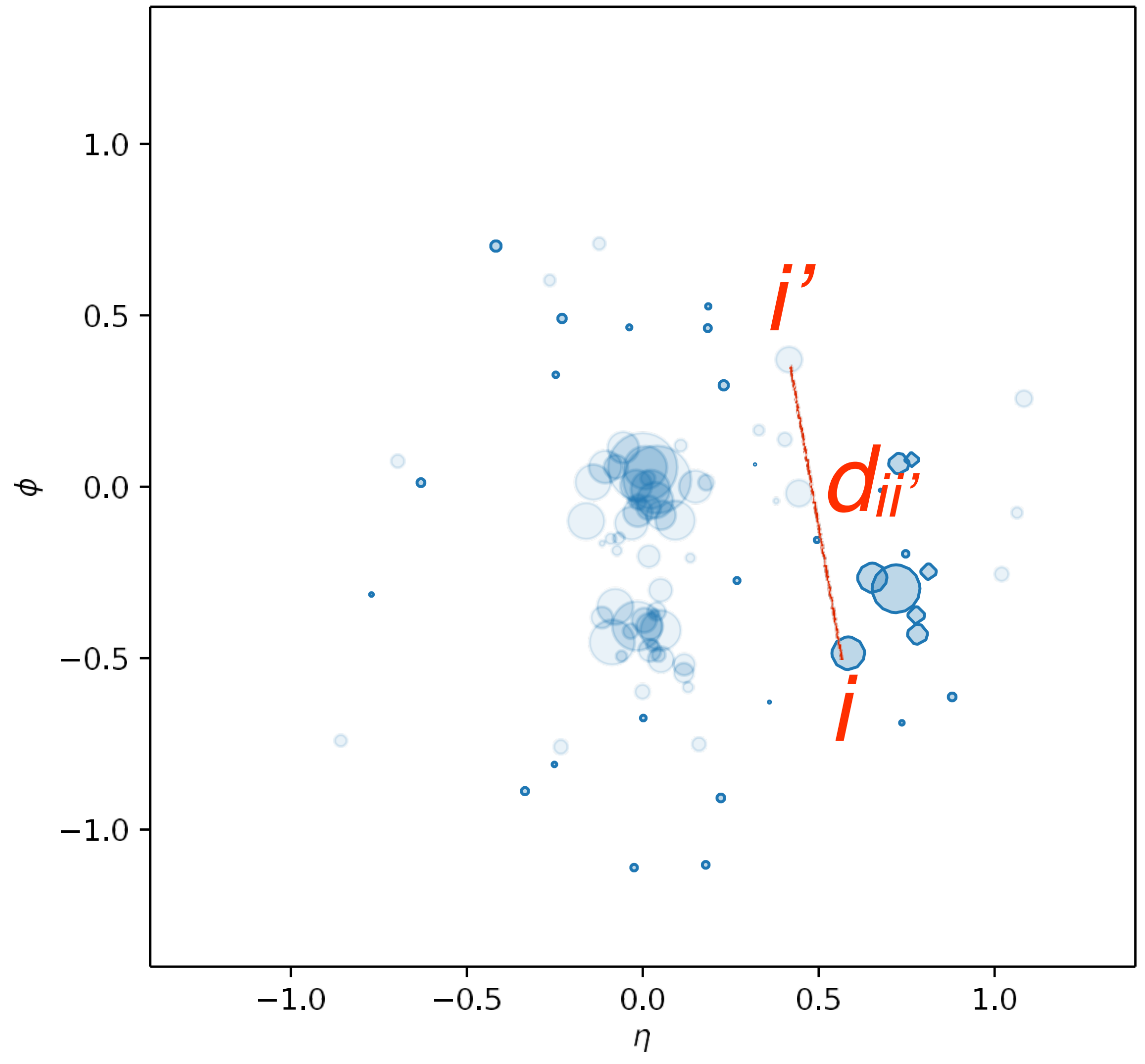
- Use jet clustering algorithms to dynamically provide network connectivity (on a **per-jet basis**)
- Early example of physics-aware ML

arXiv:1702.00748 & follow up work with Joan Bruna using graph conv nets

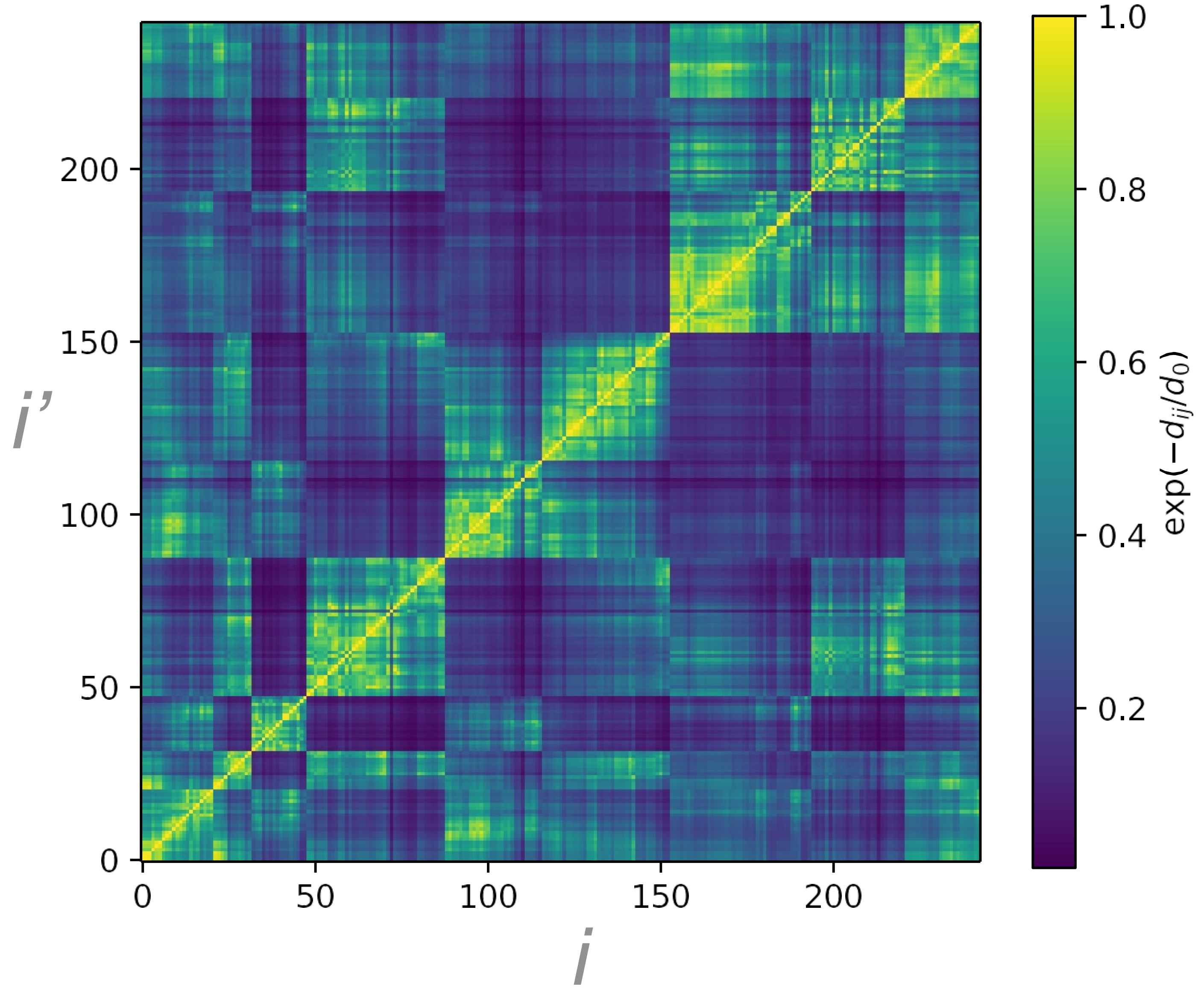
Jets as a graph

Each node of the graph is a particle

- edge length motivated by physics, same as used for clustering algorithms: $d^{\alpha}_{ii'} = \min(p_{ti}^{2\alpha}, p_{ti'}^{2\alpha})\frac{\Delta R^2_{ii'}}{R^2}$
- we can visualize one jet via the $d_{ii'}$ **adjacency matrix**
- adjacency matrix changes **continuously** with and features (momenta of particles)



C/A algorithm with α=0



Deep Sets (and PointNets)

With permutation invariance as a motivation, physicists picked up on Deep Sets architecture. With different input features these are known as "Energy Flow Networks" and "Particle Flow Networks"

- Choice of input features enforces other properties of resulting network important for physicists

Energy Flow Networks: Deep Sets for Particle Jets

Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler

Center for Theoretical Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.

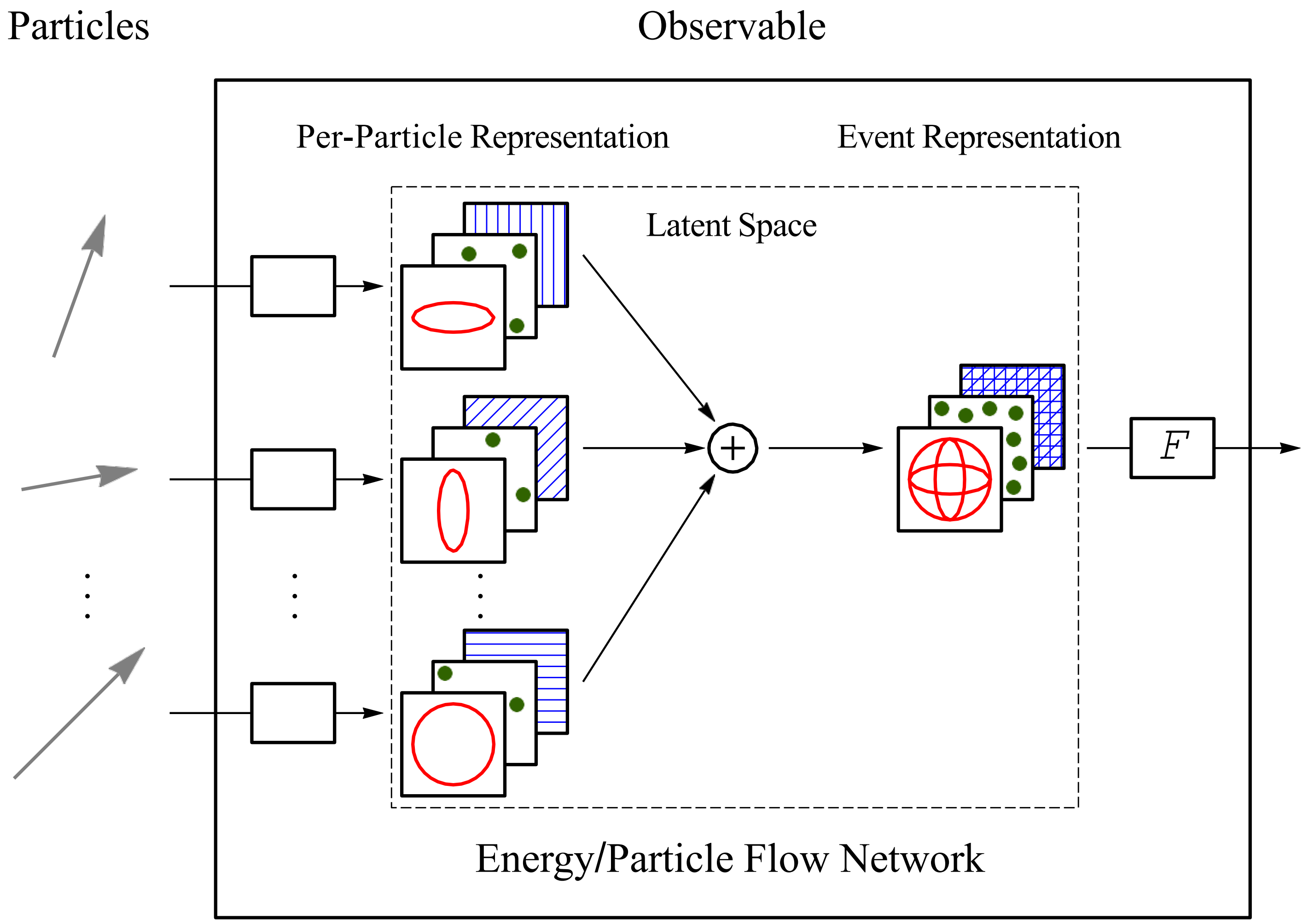
Department of Physics, Harvard University, 17 Oxford Street, Cambridge, MA 02138, U.S.A.

Deep Sets

Manzil Zaheer[1,2], Satwik Kottur[1], Siamak Ravanbhakhsh[1], Barnabás Póczos[1], Ruslan Salakhutdinov[1], Alexander J Smola[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services
{manzilz,skottur,mravanba,bapoczos,rsalakhu,smola}@cs.cmu.edu

PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

Charles R. Qi* Hao Su* Kaichun Mo Leonidas J. Guibas
Stanford University



Physics-aware ML for jets

In the last few years a lot of work on physics-aware ML for jets

Machine Learning for Jet Physics Workshops, January 2020, shortly before the pandemic

Graph Neural Networks in Particle Physics

Jonathan Shlomi[1], **Peter Battaglia**[2], **Jean-Roch Vlimant**[3]

1 Weizmann Institute of Science, Rehovot, Israel
2 DeepMind, London UK
3 California Institute of Technology, PMA, Pasadena, CA, USA 91125-0002

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], M. Fairbairn[5], W. Fedorko[6], C. Gay[6], L. Gouskos[7], P. T. Komiske[8], S. Leiss[1], A. Lister[6], S. Macaluso[3,4], E. M. Metodiev[8], L. Moore[9], B. Nachman,[10,11], K. Nordström[12,13], J. Pearkes[6], H. Qu[7], Y. Rath[14], M. Rieger[14], D. Shih[4], J. M. Thompson[2], and S. Varma[5]

Comparison for binary classification benchmark

arXiv:1902.09914

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], M. Fairbairn[5], W. Fedorko[6], C. Gay[6], L. Gouskos[7], P. T. Komiske[8], S. Leiss[1], A. Lister[6], S. Macaluso[3,4], E. M. Metodiev[8], L. Moore[9], B. Nachman,[10,11], K. Nordström[12,13], J. Pearkes[6], H. Qu[7], Y. Rath[14], M. Rieger[14], D. Shih[4], J. M. Thompson[2], and S. Varma[5]

- Huge range in number of parameters
- Dynamic Graph CNN best performing
- Noticeable performance gap for DeepSets
- TreeRNN doing very well, tiny network in comparison (also train with little data)
- Image-based CNNs do well, but are huge
- Message Passing Networks or Graph Attention Networks not evaluated in this comparison

1/False Positive Rate

		AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	# Param
	CNN [16]	0.981	0.930	914± 14	995± 15	966± 18	610k
"Images" CNN	ResNeXt [30]	0.984	0.936	1122± 47	1246± 28	1286± 31	1.46M
	TopoDNN [18]	0.972	0.916	295± 5	378± 5	391 ± 8	59k
	Multi-body N-subjettiness 6 [24]	0.979	0.922	792± 18	802± 12	783± 13	57k
	Multi-body N-subjettiness 8 [24]	0.981	0.929	867± 15	926± 20	886± 18	58k
Tree RNN	TreeNiN [43]	0.982	0.933	1025± 11	1209± 23	1167± 24	34k
	P-CNN	0.980	0.930	732± 24	838± 13	841± 14	348k
Dynamic Graph CNN	ParticleNet [47]	0.985	0.938	1298± 46	1383± 45	1374± 41	498k
	LBN [19]	0.981	0.931	836± 17	852± 67	971± 20	705k
	LoLa [22]	0.980	0.929	722± 17	768± 11	751± 11	127k
	Energy Flow Polynomials [21]	0.980	0.932	384			1k
DeepSets	Energy Flow Network [23]	0.979	0.927	633± 31	734± 13	729± 11	82k
DeepSets	Particle Flow Network [23]	0.982	0.932	891± 18	1005± 21	1005± 29	82k
	GoaT	0.985	0.939	1368± 140		1549± 208	35k

An example of physics-inspired ML

Use inspired

Peculiarities of the particle physics context:

- Point cloud with permutation symmetry and underlying geometric structure

Maybe there are alternate universes in the multiverse where deep set architecture and graph neural networks emerged from physics-inspired use case

- It didn't quite play out like that

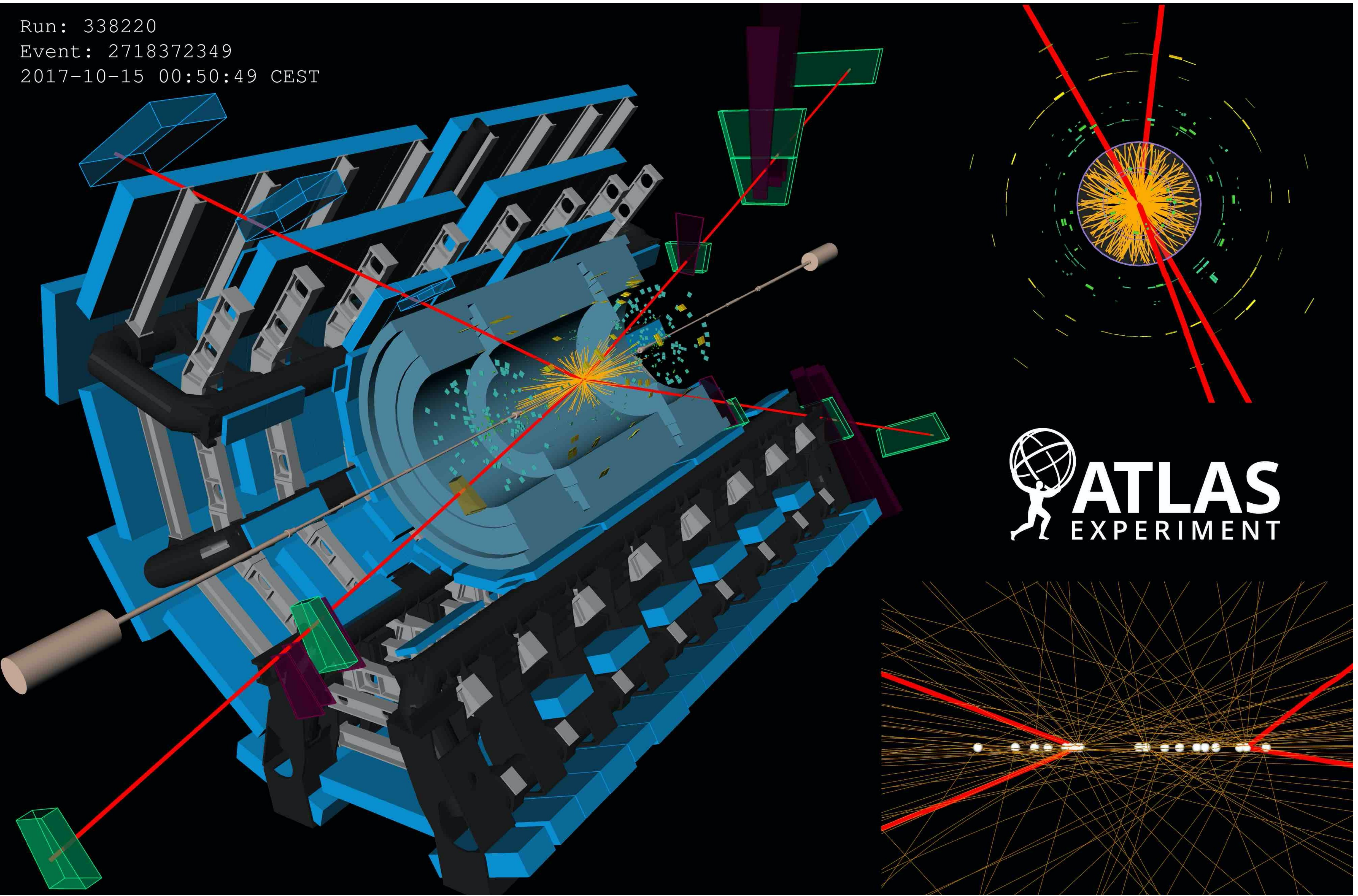
But this next one did…

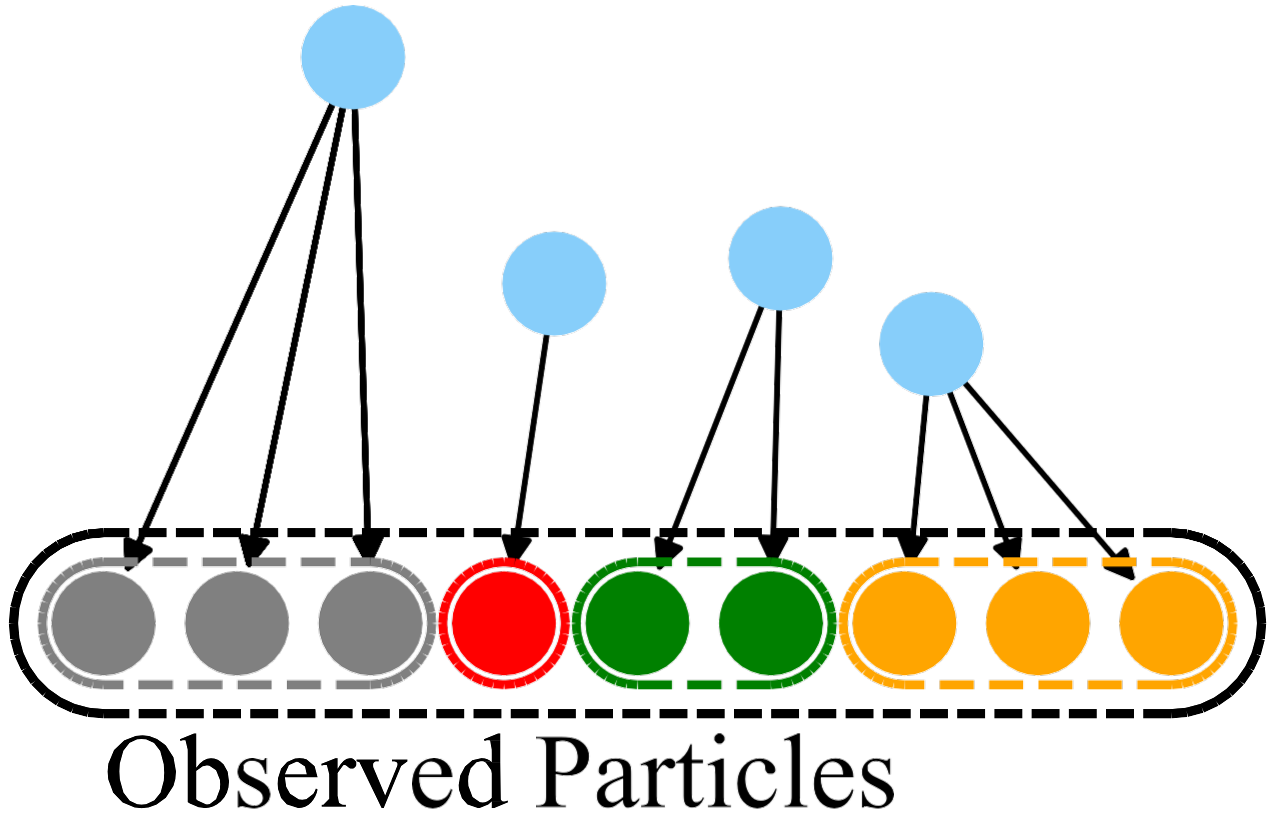
Application to Particle physics

Serviansky, et. al. [arXiv:2002.08772]

Task: group together particles according to common point of origin

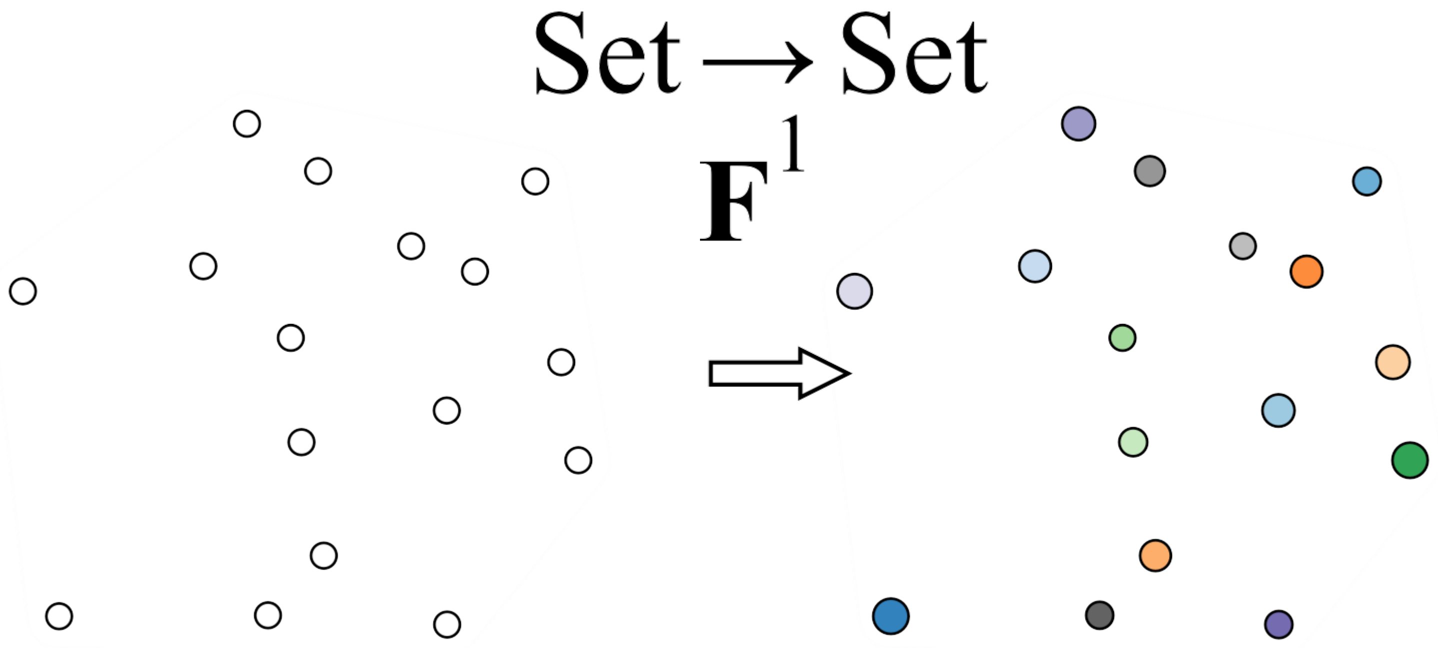
- Accuracy is a bottle neck in several searches for new physics





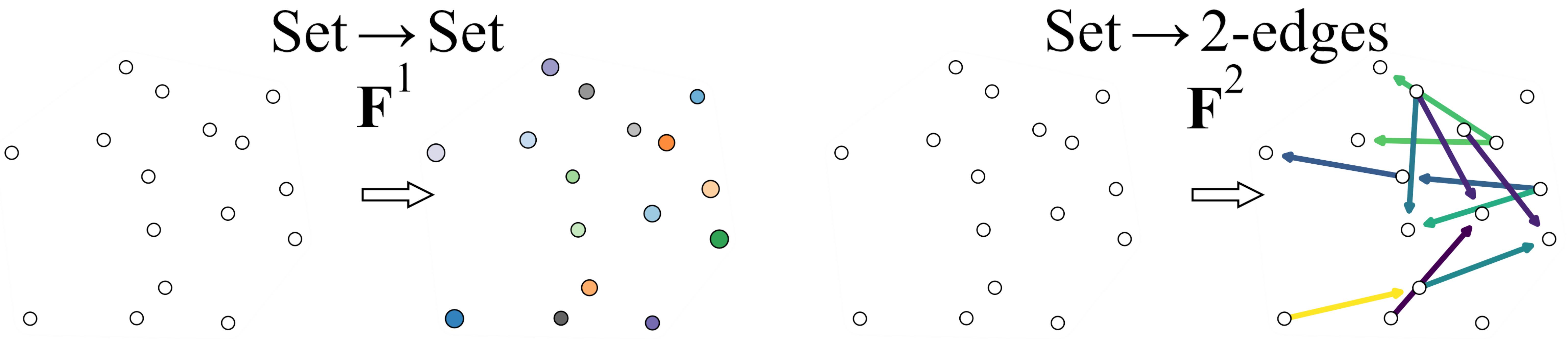
Set to (Hyper-)Graph Functions

Serviansky, et. al. [arXiv:2002.08772]



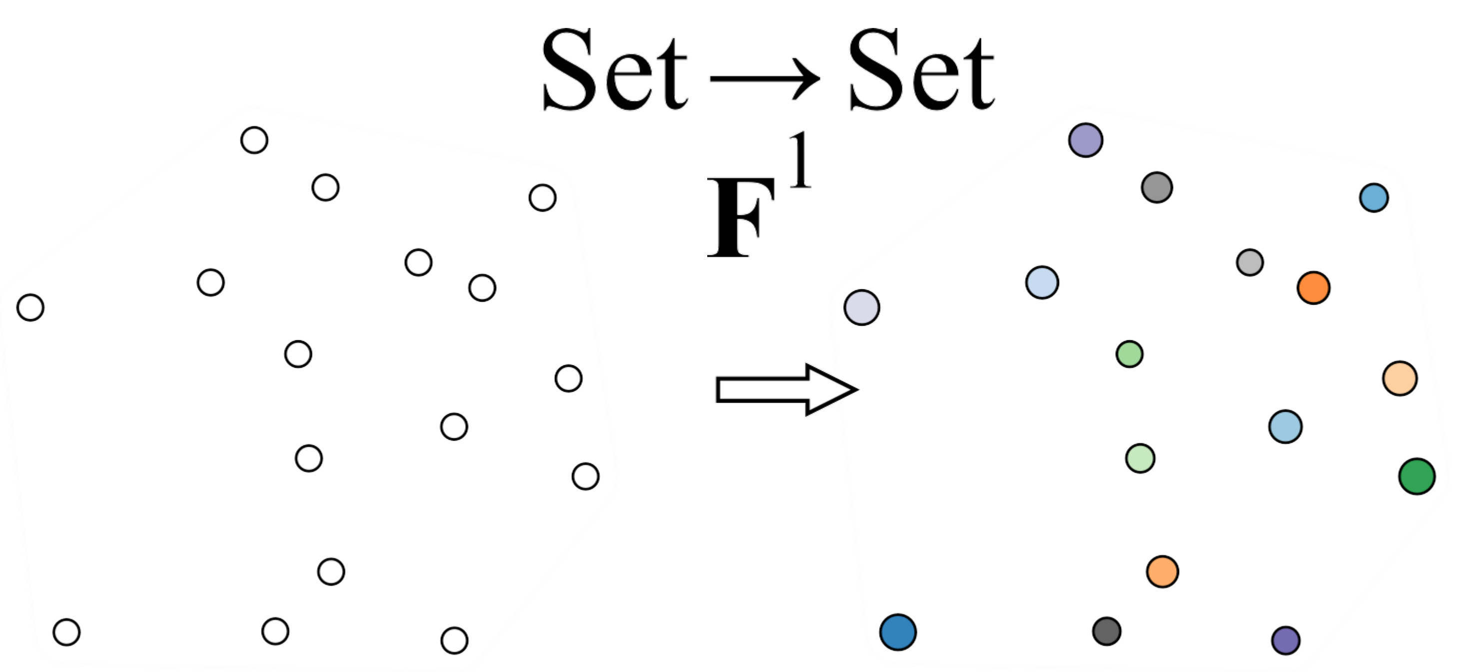
Set to (Hyper-)Graph Functions

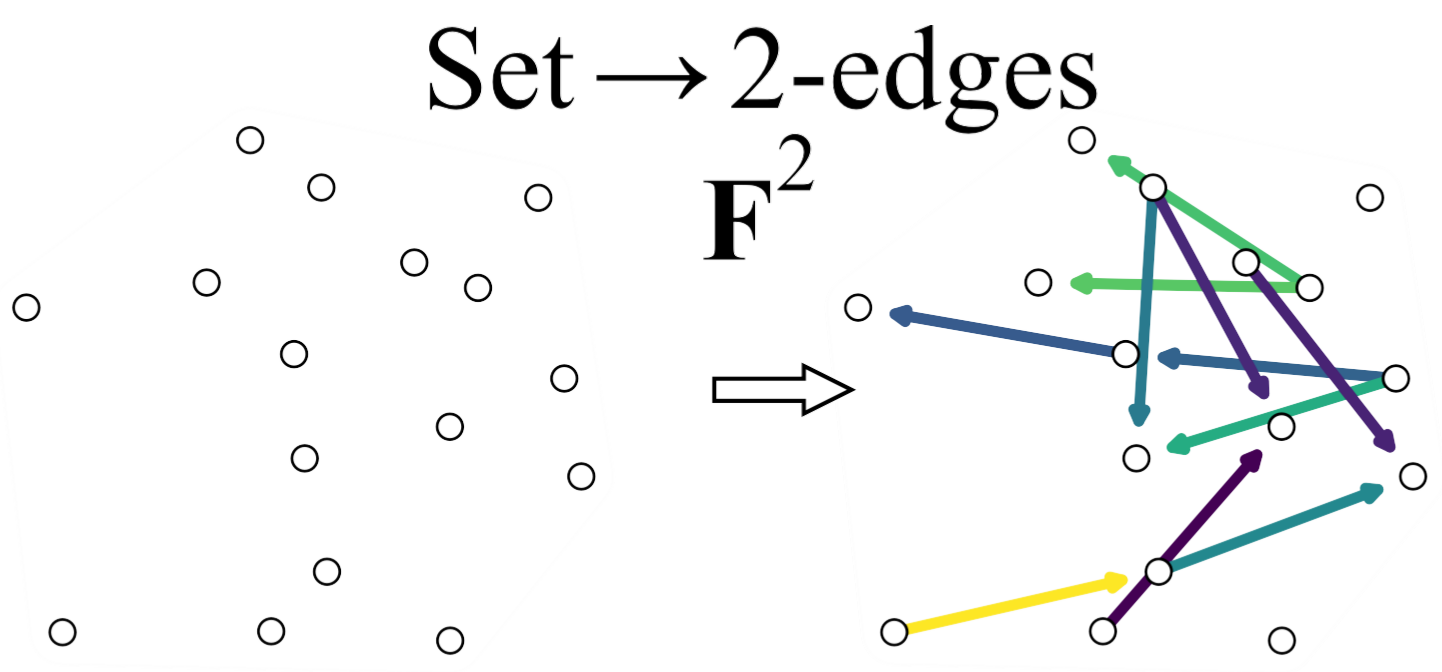
Serviansky, et. al. [arXiv:2002.08772]

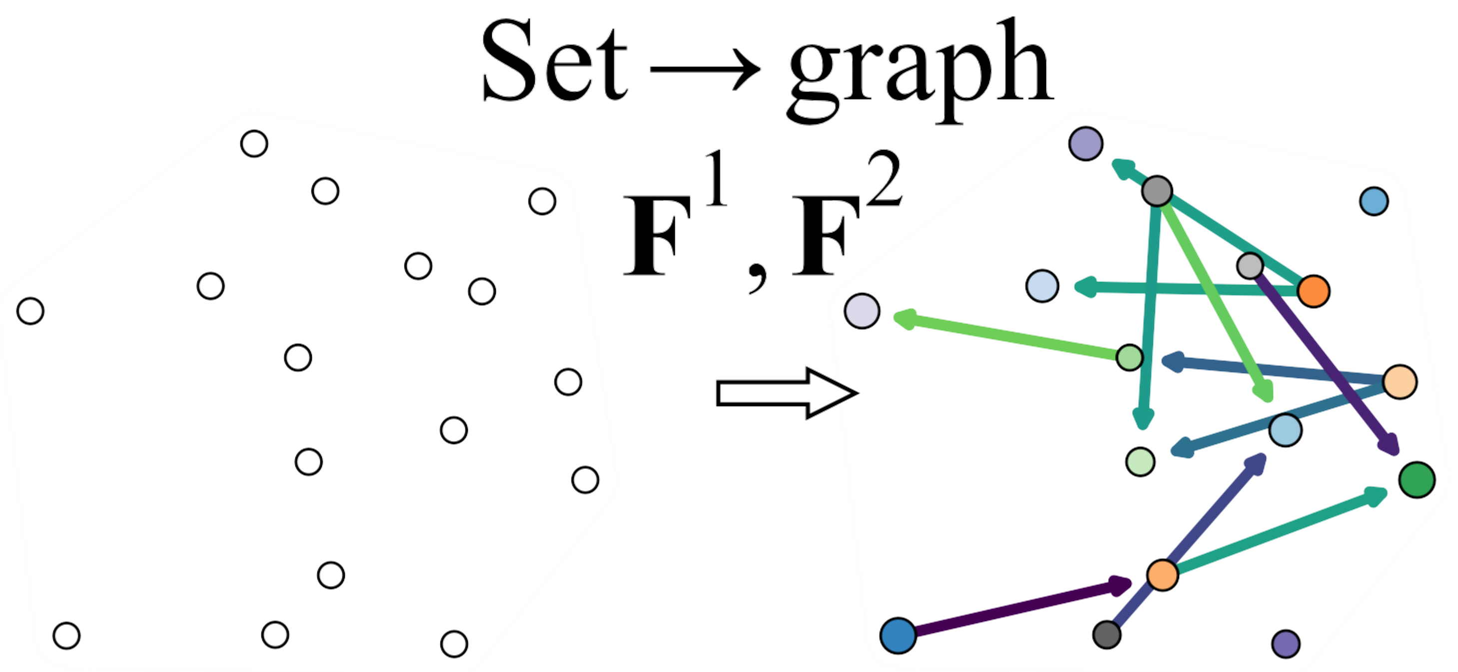


Set to (Hyper-)Graph Functions

Serviansky, et. al. [arXiv:2002.08772]

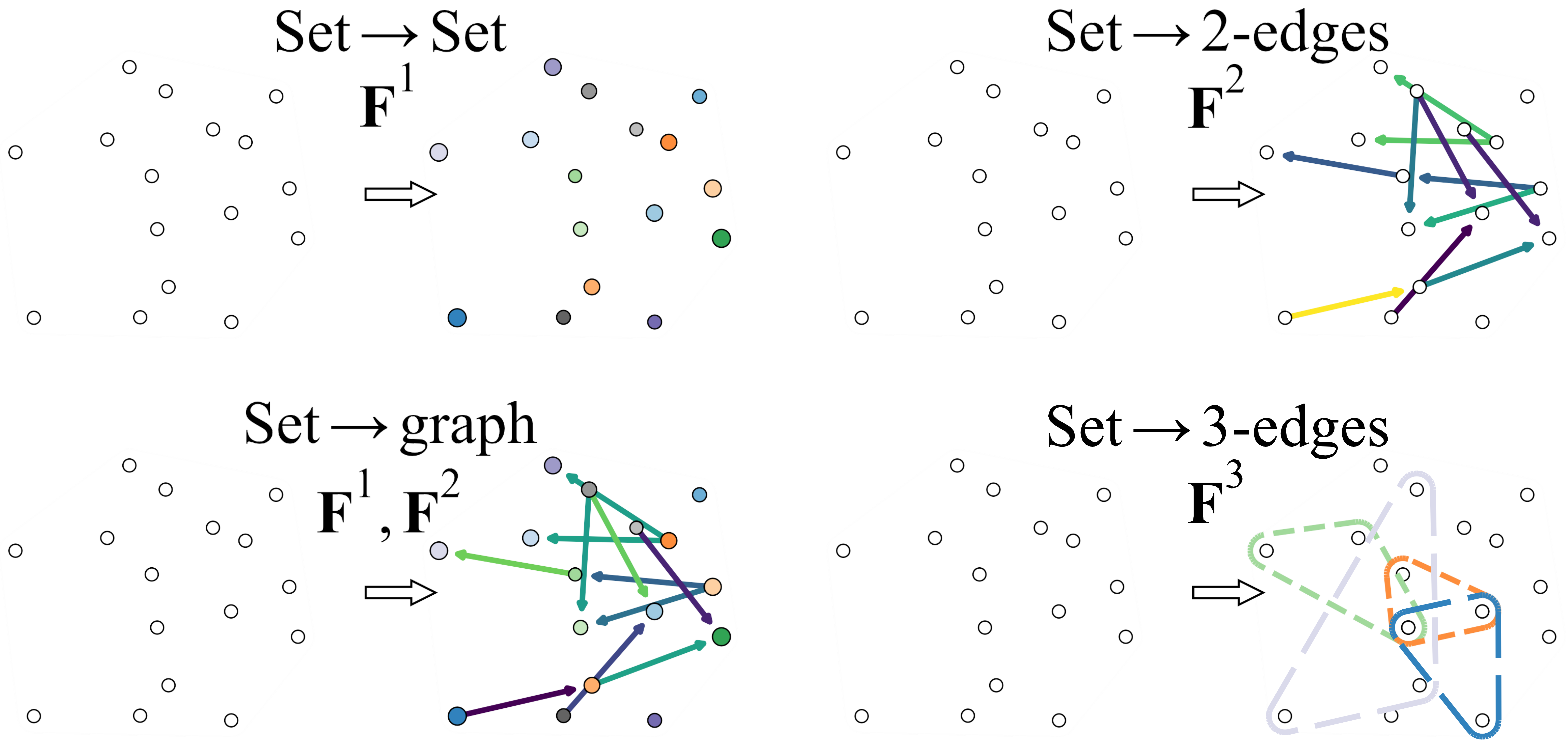






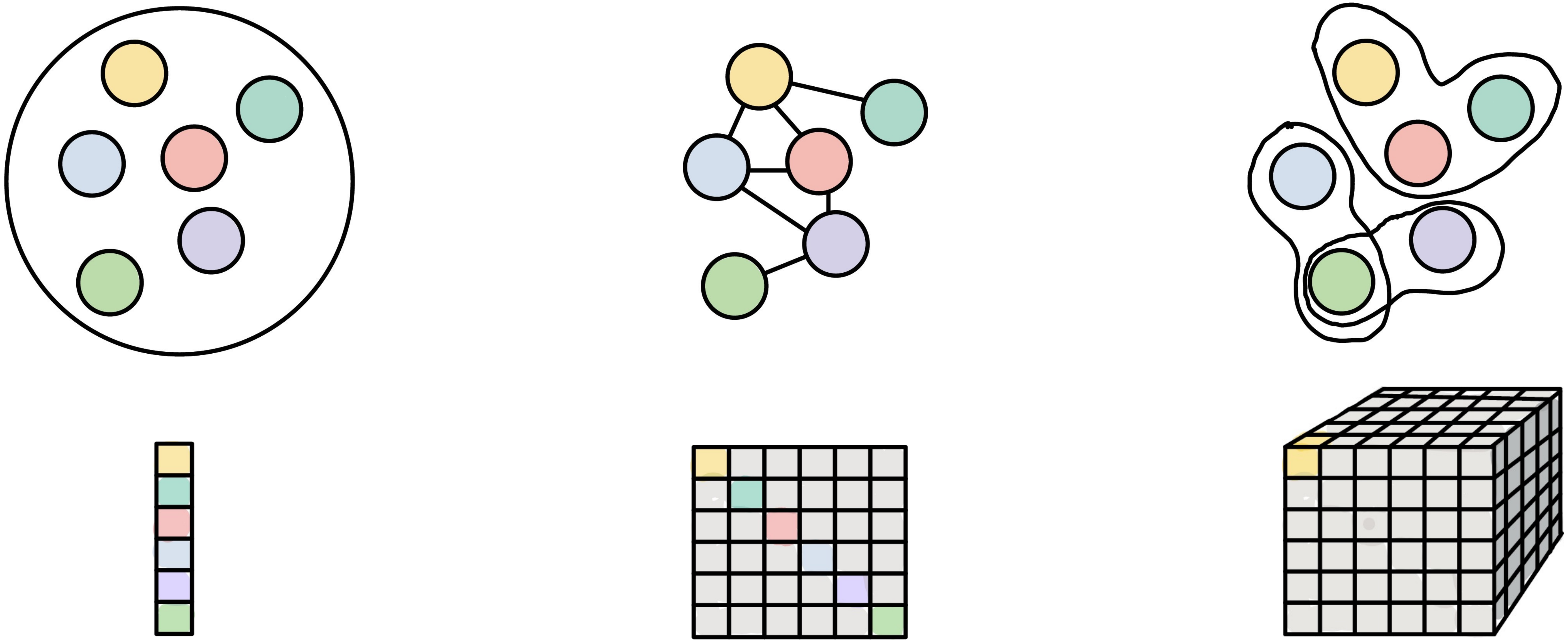
Set to (Hyper-)Graph Functions

Serviansky, et. al. [arXiv:2002.08772]



Hypergraphs as tensors

Serviansky, et. al. [arXiv:2002.08772]

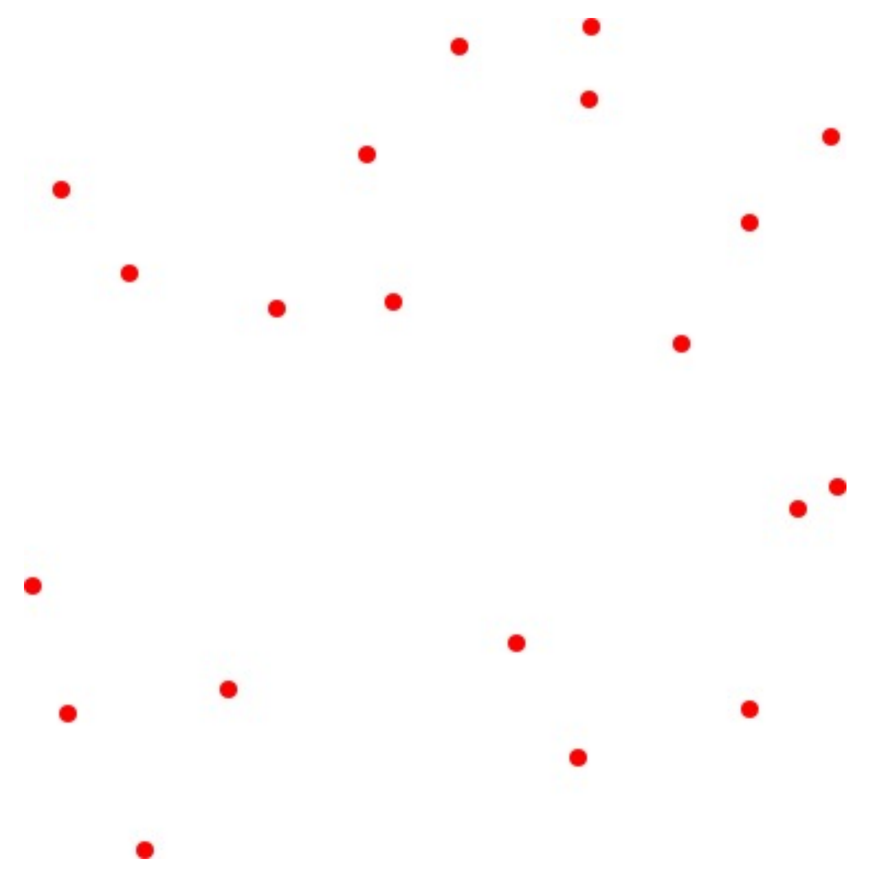


$X \in \mathbb{R}^{n \times d}$ $X \in \mathbb{R}^{n \times n \times d}$ $X \in \mathbb{R}^{n^3 \times d}$

Example

Delaunay triangulation:

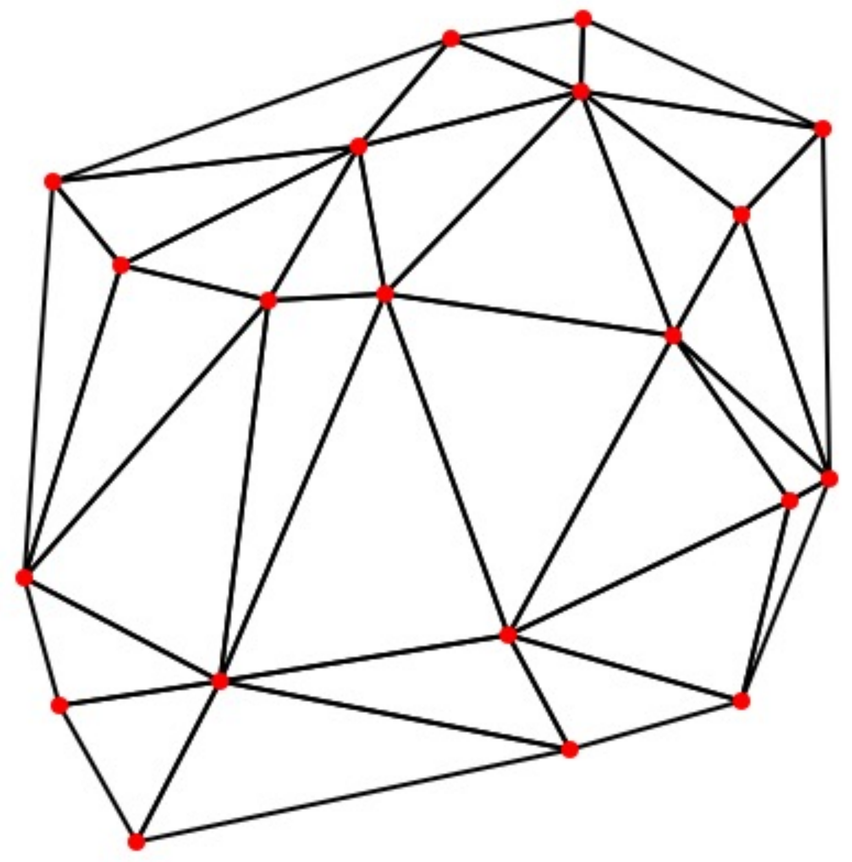
- Input: 2d points as a set



Example

Delaunay triangulation:

- Input: 2d points as a set
- Output: the Delaunay triangulation (graph)



Set2Graph: Learning Graphs from Sets

Hadar Serviansky[1] Nimrod Segol[1] Jonathan Shlomi[1] Kyle Cranmer[2]
Eilam Gross[1] Haggai Maron[3] Yaron Lipman[1]

https://arxiv.org/abs/2002.08772 ICML2020

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE
1

NYU
2

NVIDIA
3

Serviansky, et. al. [arXiv:2002.08772]

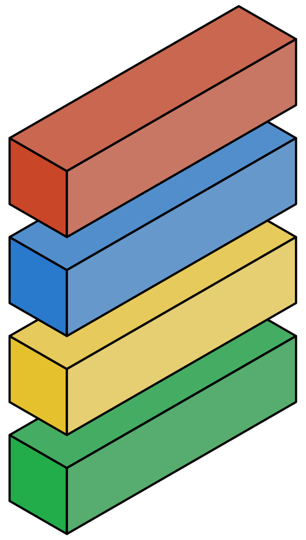
$$F(X; \quad) = \qquad\qquad\qquad\qquad (X)$$

Theorem: Set2Graph model is set-to-graph universal.

Model: Set2Graph

Serviansky, et. al. [arXiv:2002.08772]

$$F(X; \quad) = \qquad\qquad\qquad (X)$$

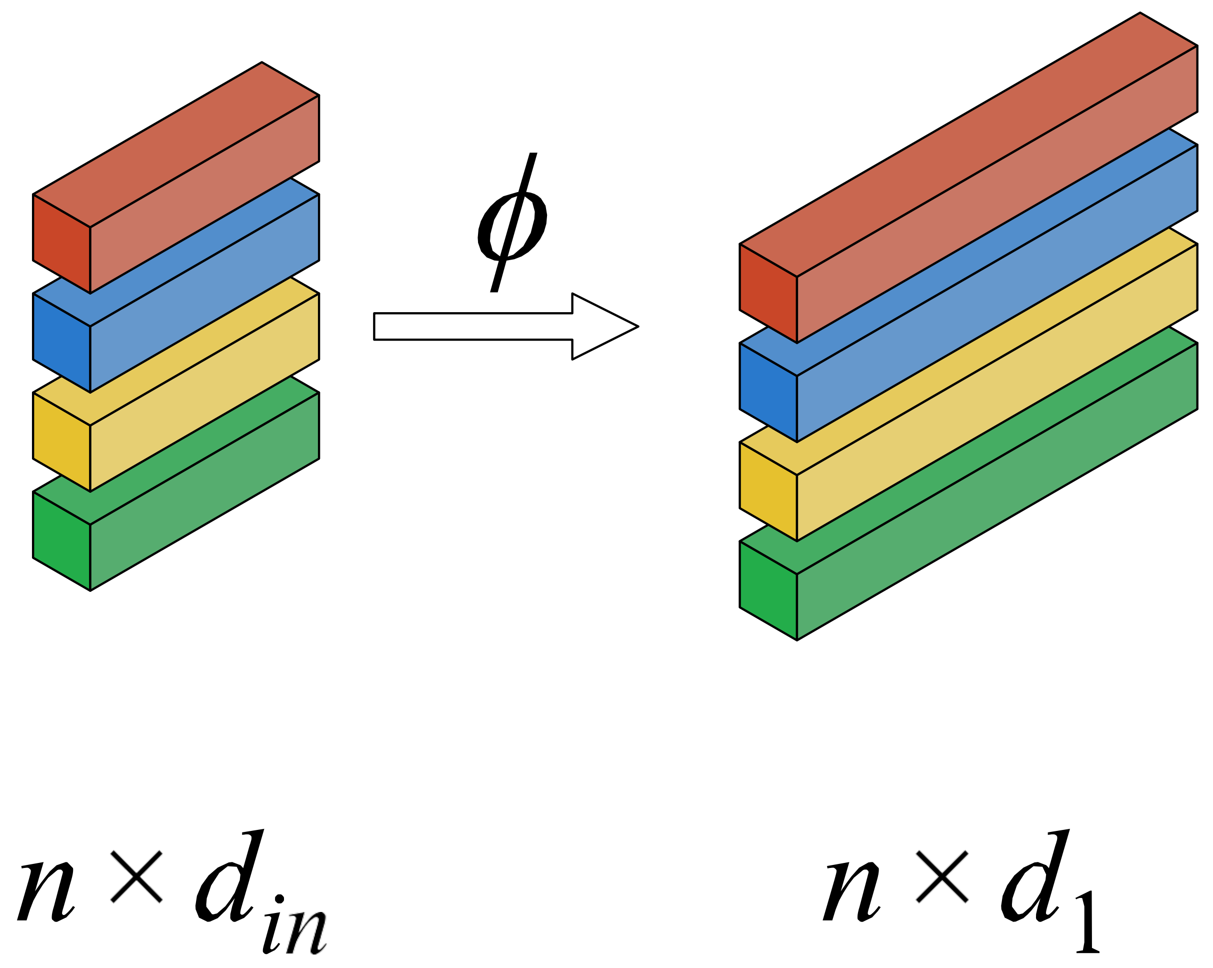


$n \times d_{in}$

Theorem: Set2Graph model is set-to-graph universal.

Model: Set2Graph

$$F(X; \quad) = \qquad\qquad (X)$$

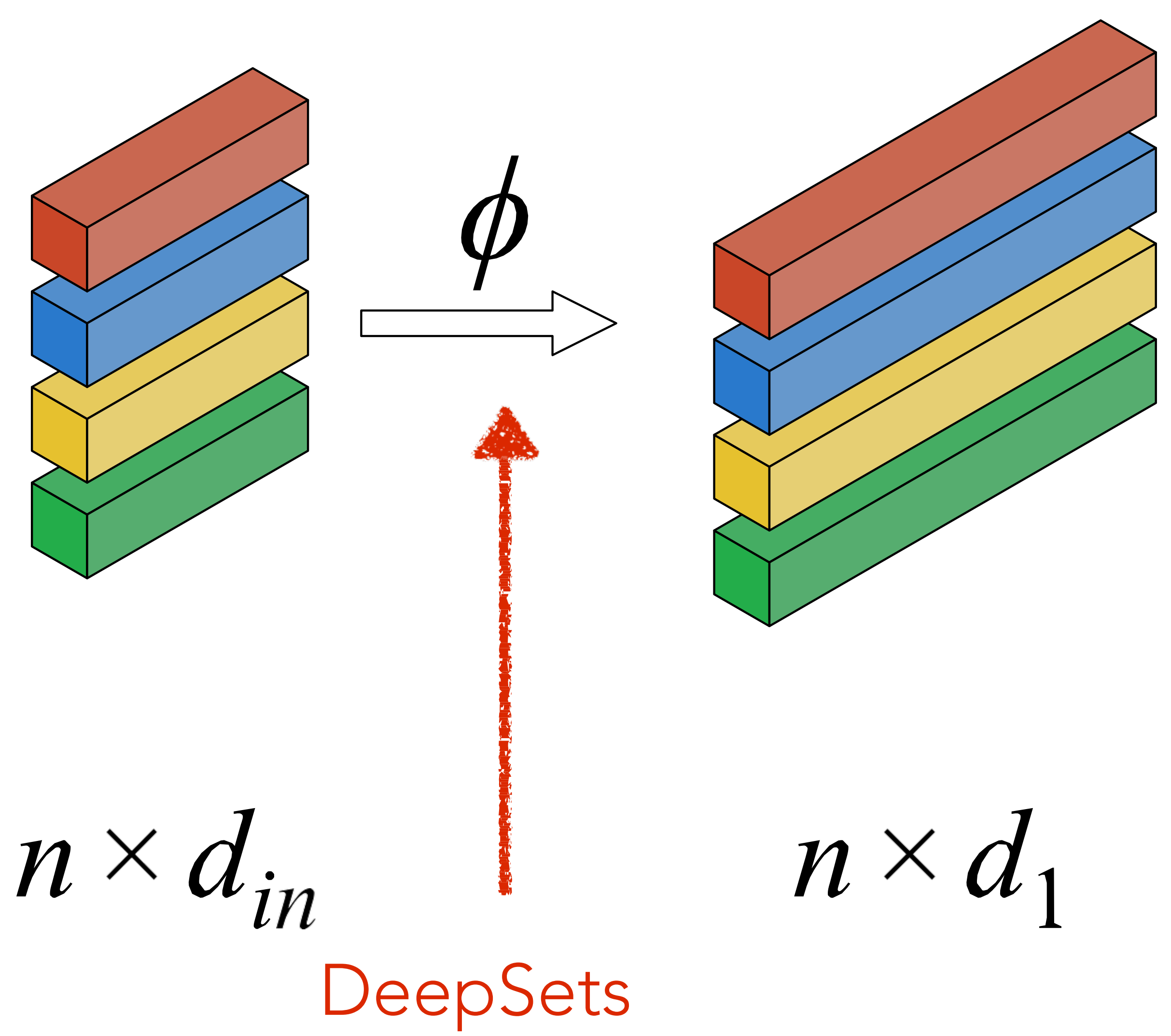


Theorem: Set2Graph model is set-to-graph universal.

Model: Set2Graph

Serviansky, et. al. [arXiv:2002.08772]

$$F(X;\quad) = \qquad\qquad\qquad (X)$$

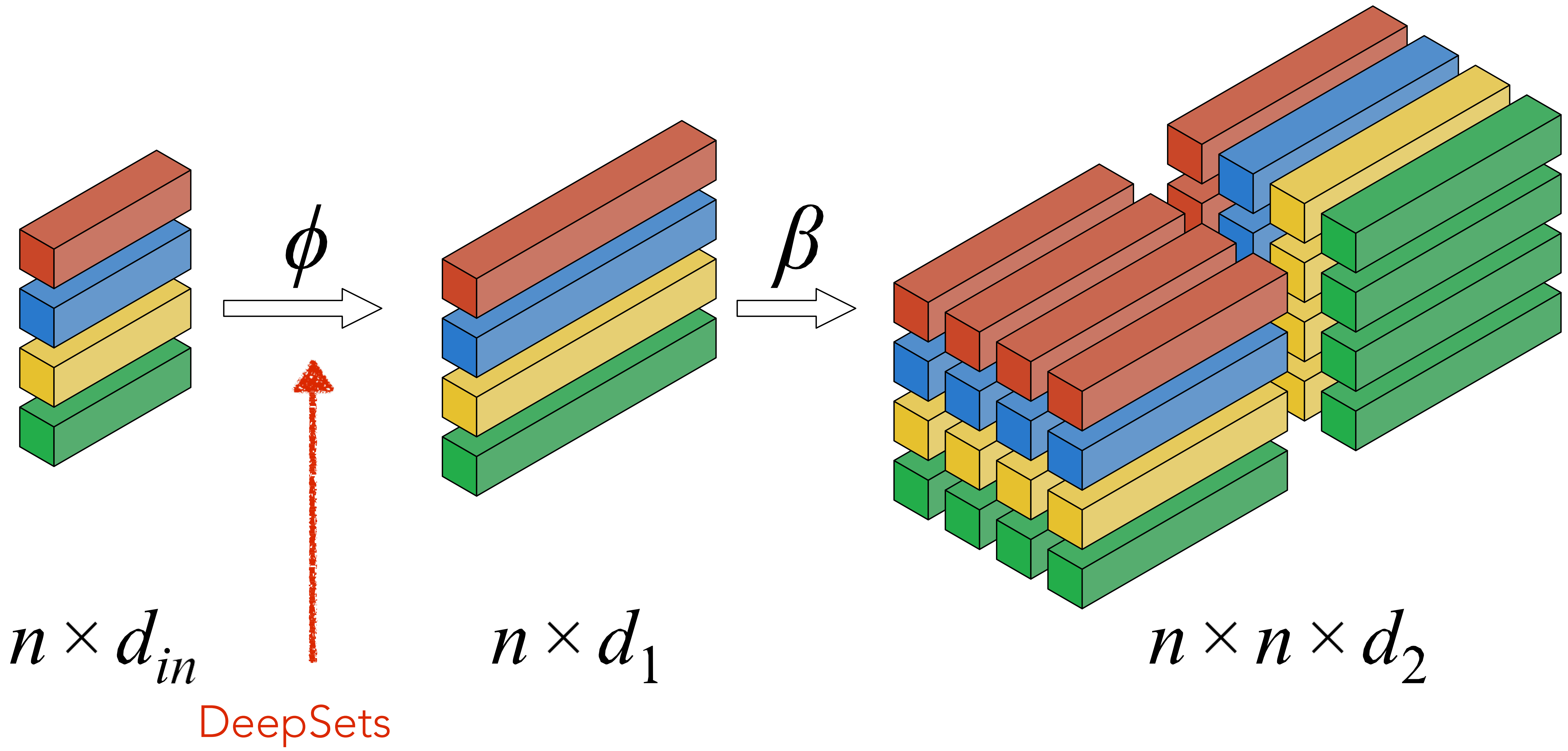


Theorem: Set2Graph model is set-to-graph universal.

Model: Set2Graph

Serviansky, et. al. [arXiv:2002.08772]

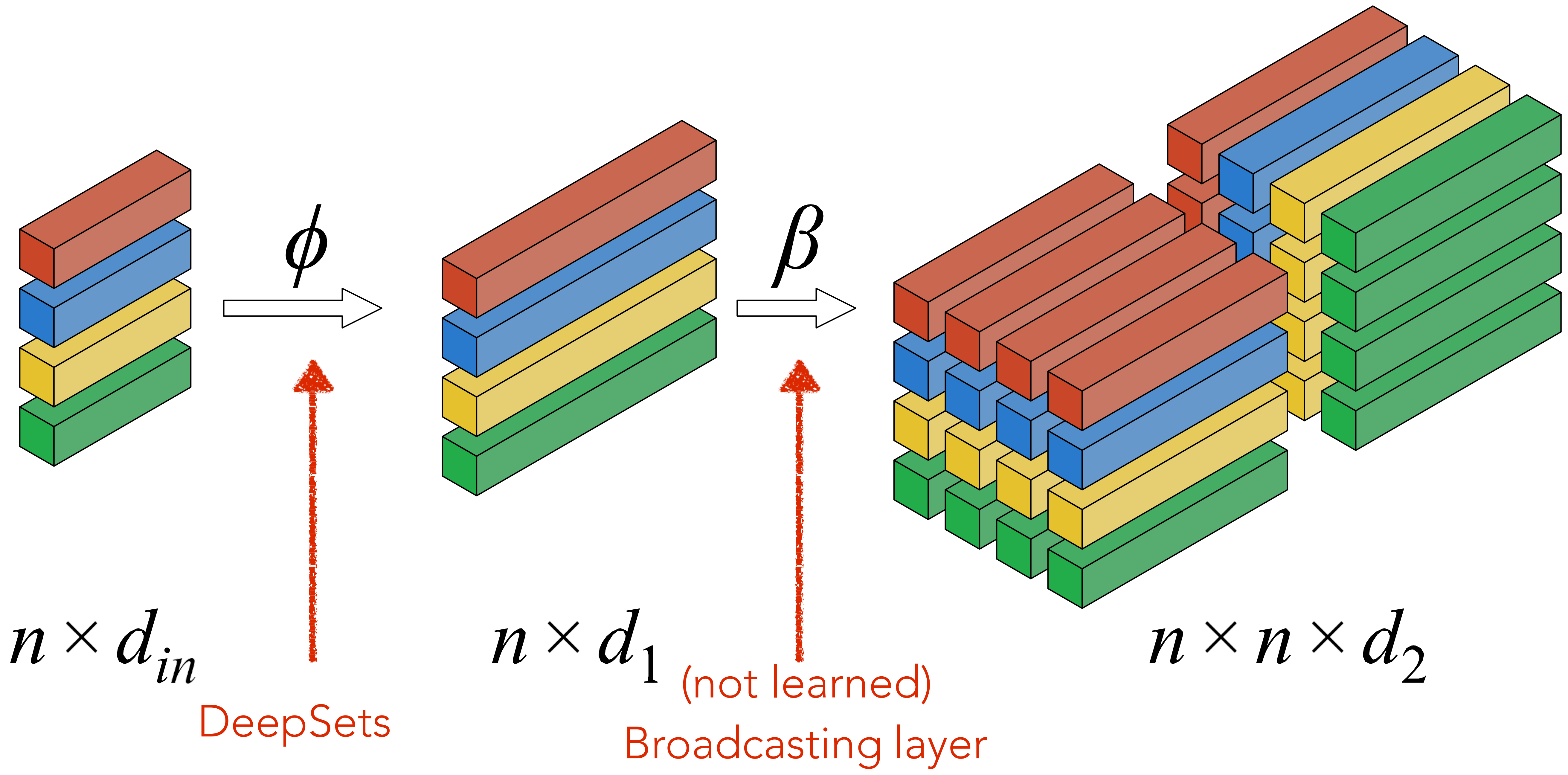
$$F(X; \quad) = \qquad\qquad (X)$$



Theorem: Set2Graph model is set-to-graph universal.

Serviansky, et. al. [arXiv:2002.08772]

$$F(X; \quad) = \qquad\qquad (X)$$

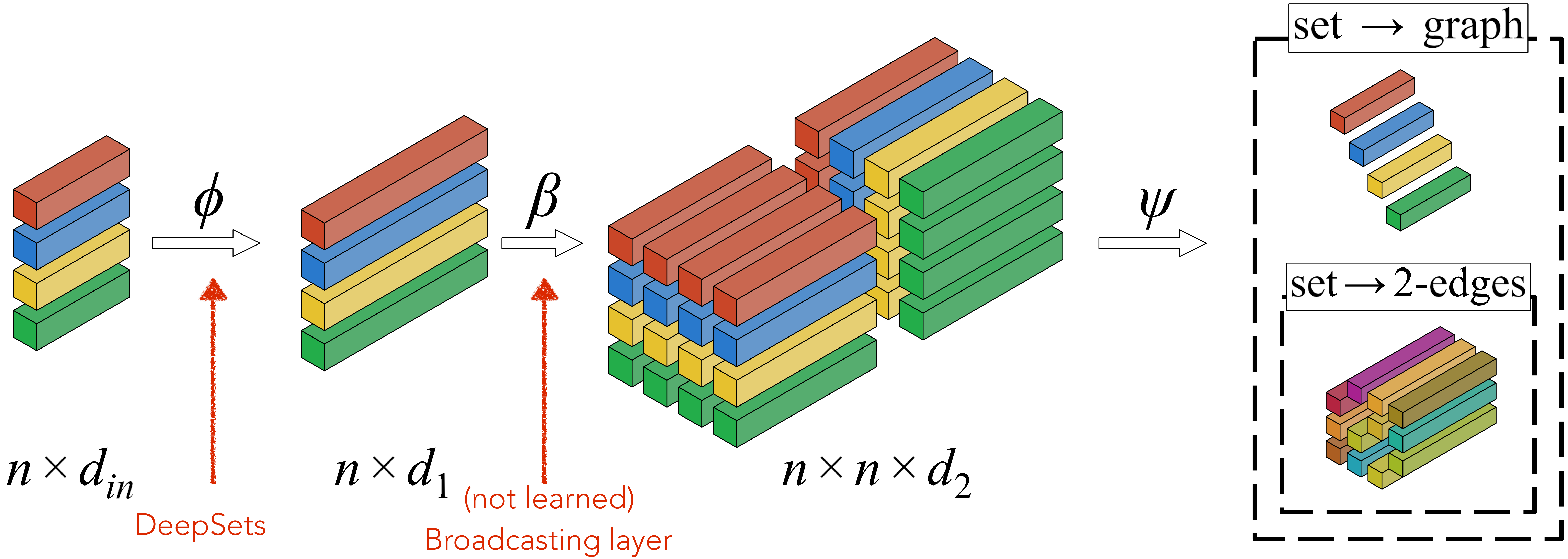


Theorem: Set2Graph model is set-to-graph universal.

Model: Set2Graph

Serviansky, et. al. [arXiv:2002.08772]

$$F(X; \quad) = \qquad\qquad (X)$$

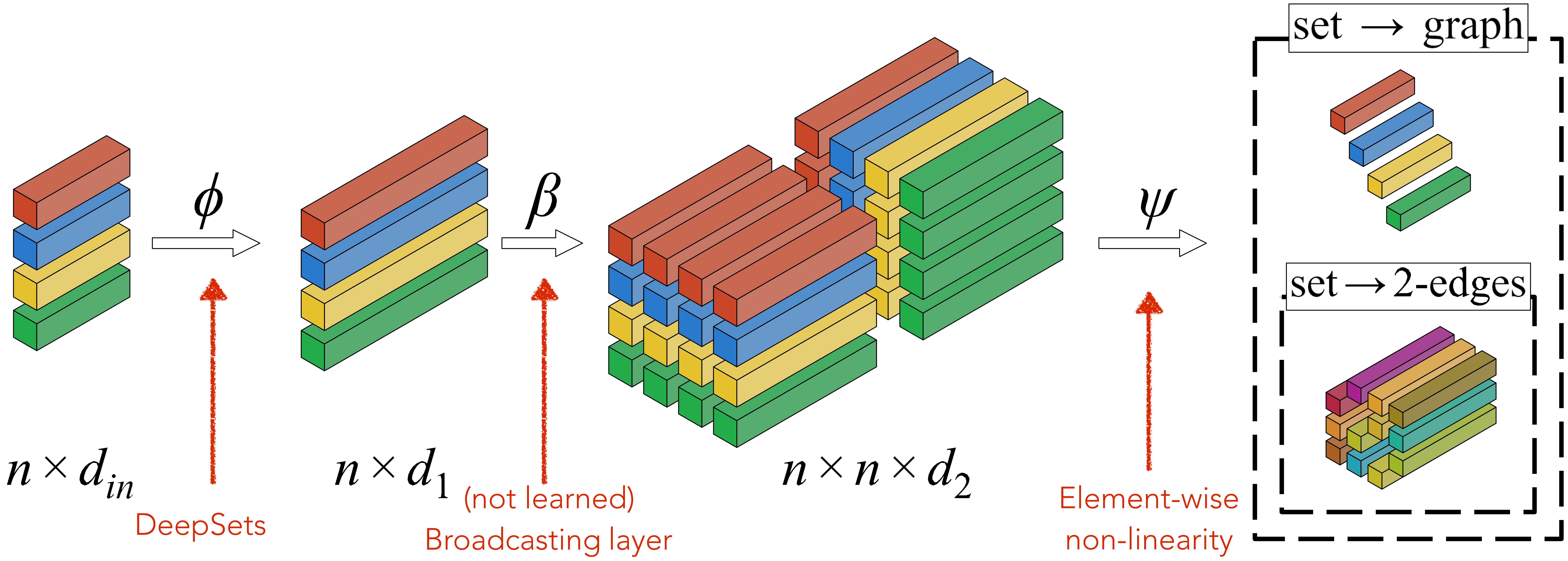


Theorem: Set2Graph model is set-to-graph universal.

Model: Set2Graph

Serviansky, et. al. [arXiv:2002.08772]

$$F(X; \quad) = \qquad\qquad (X)$$



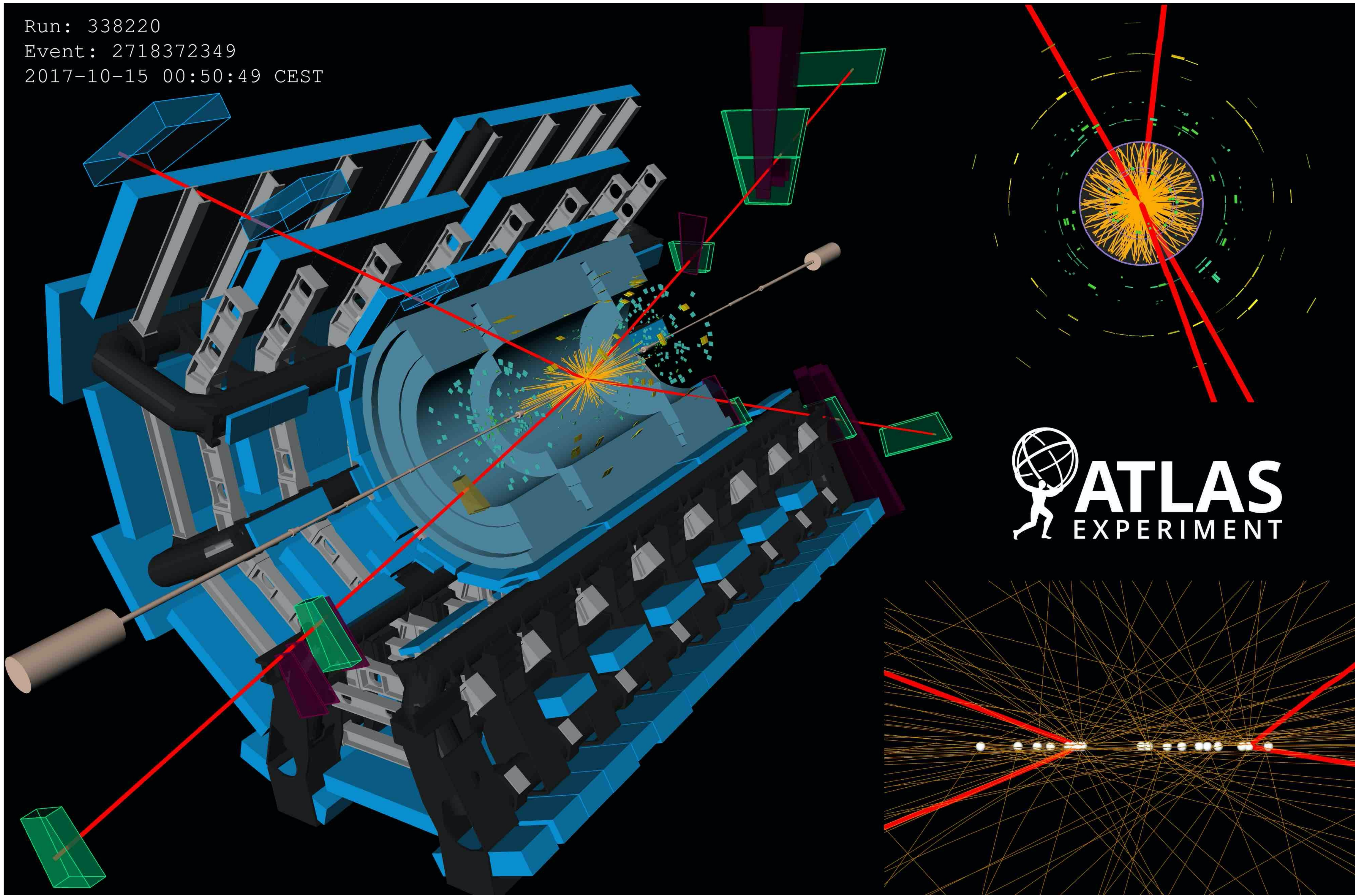
Theorem: Set2Graph model is set-to-graph universal.

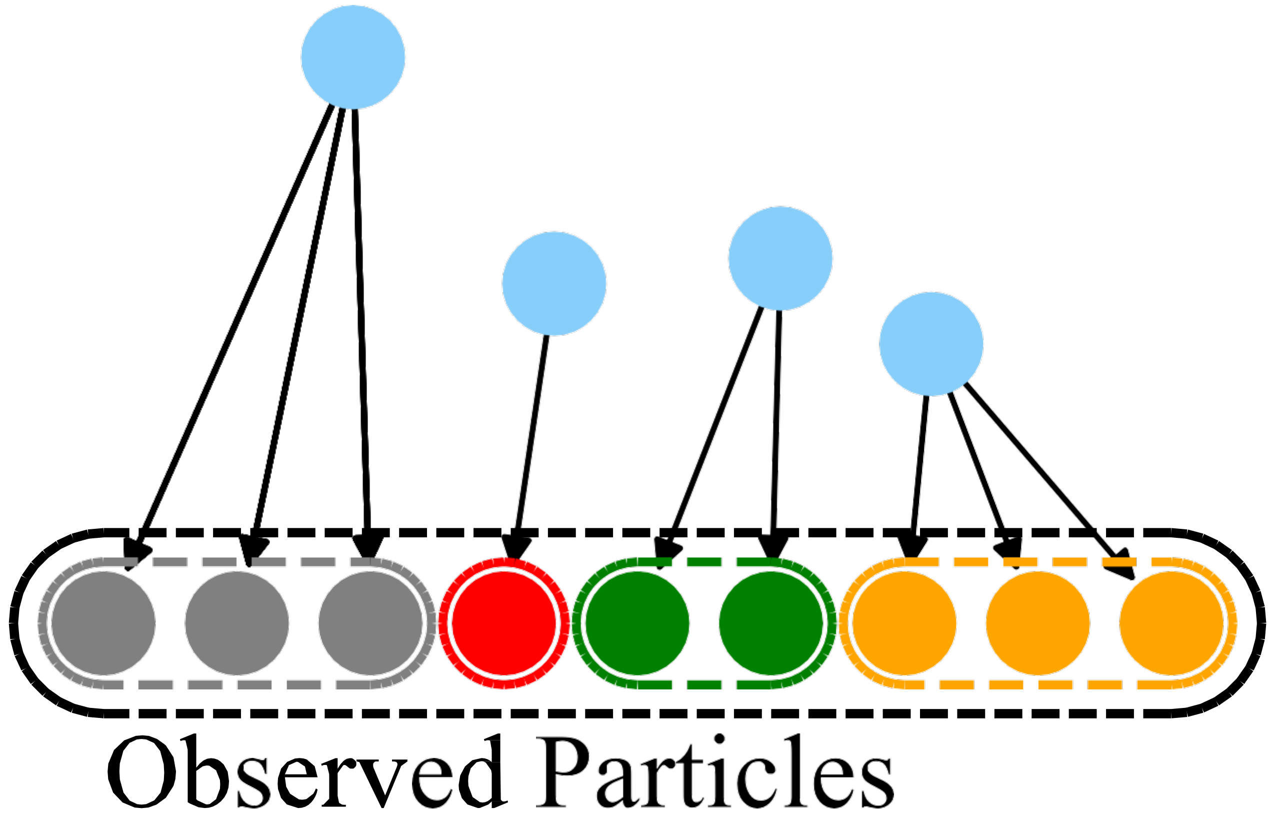
Application to Particle physics

Serviansky, et. al. [arXiv:2002.08772]

Task: group together particles according to common point of origin

- Accuracy is a bottle neck in several searches for new physics



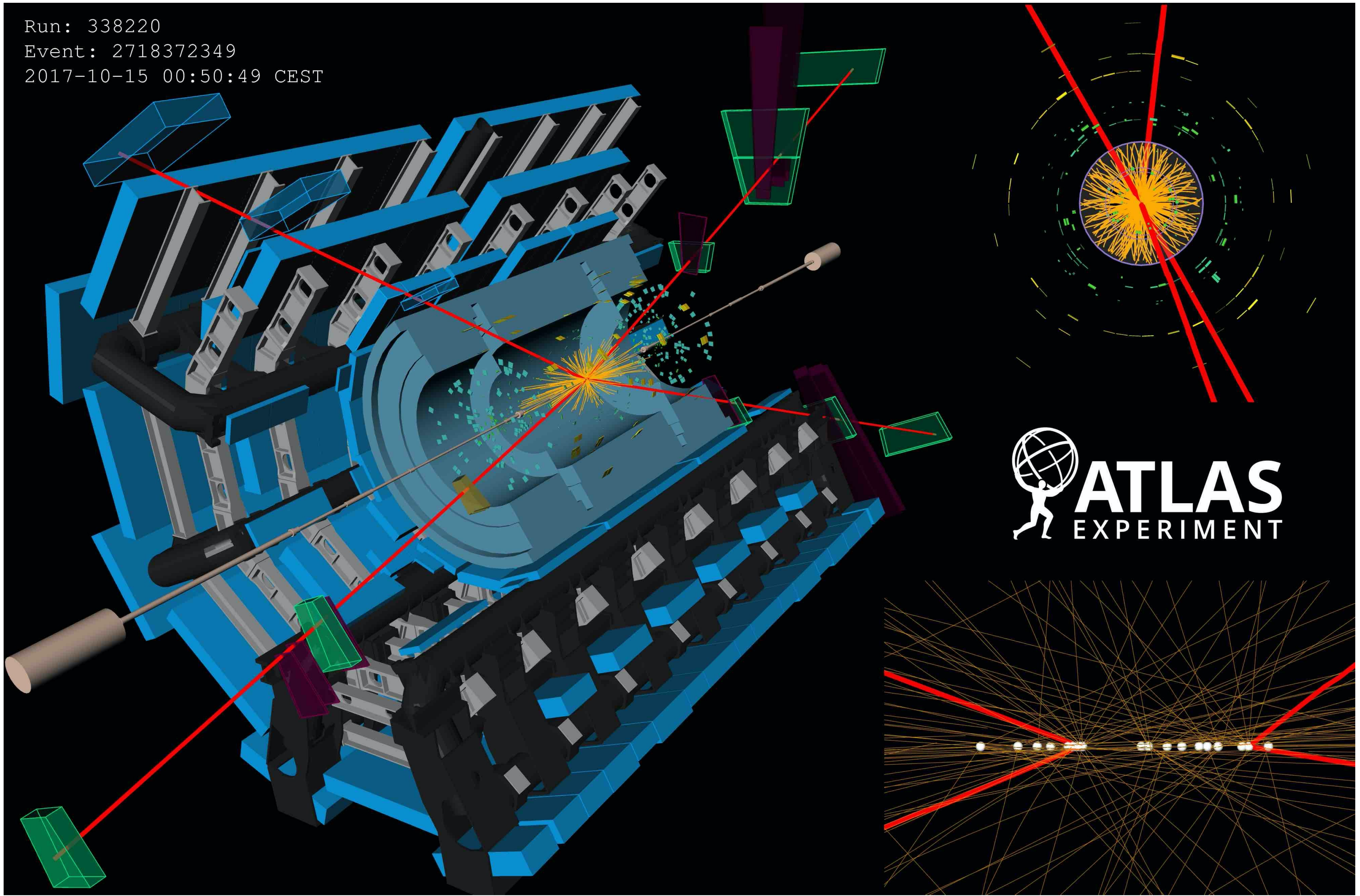


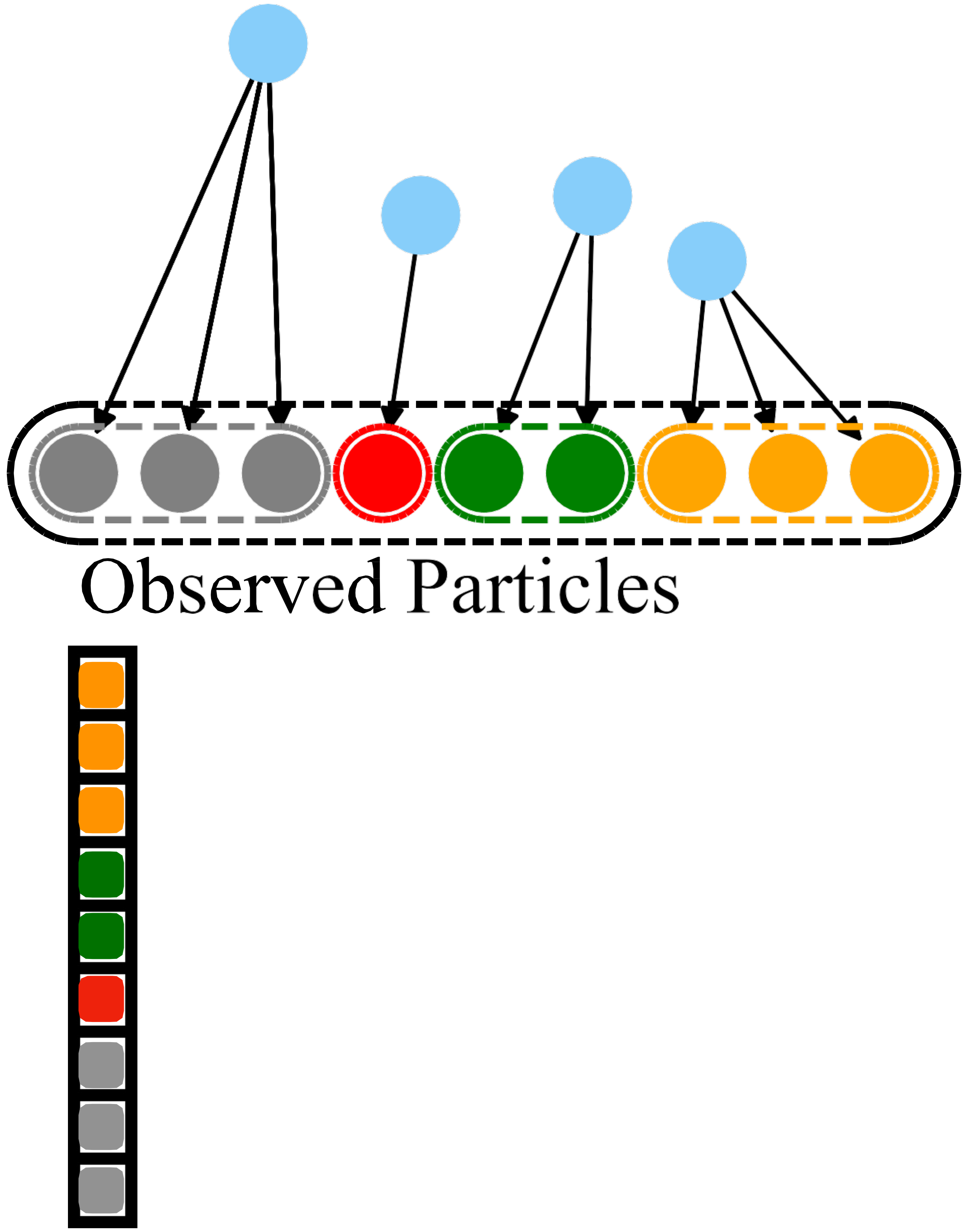
Application to Particle physics

Serviansky, et. al. [arXiv:2002.08772]

Task: group together particles according to common point of origin

- Accuracy is a bottle neck in several searches for new physics



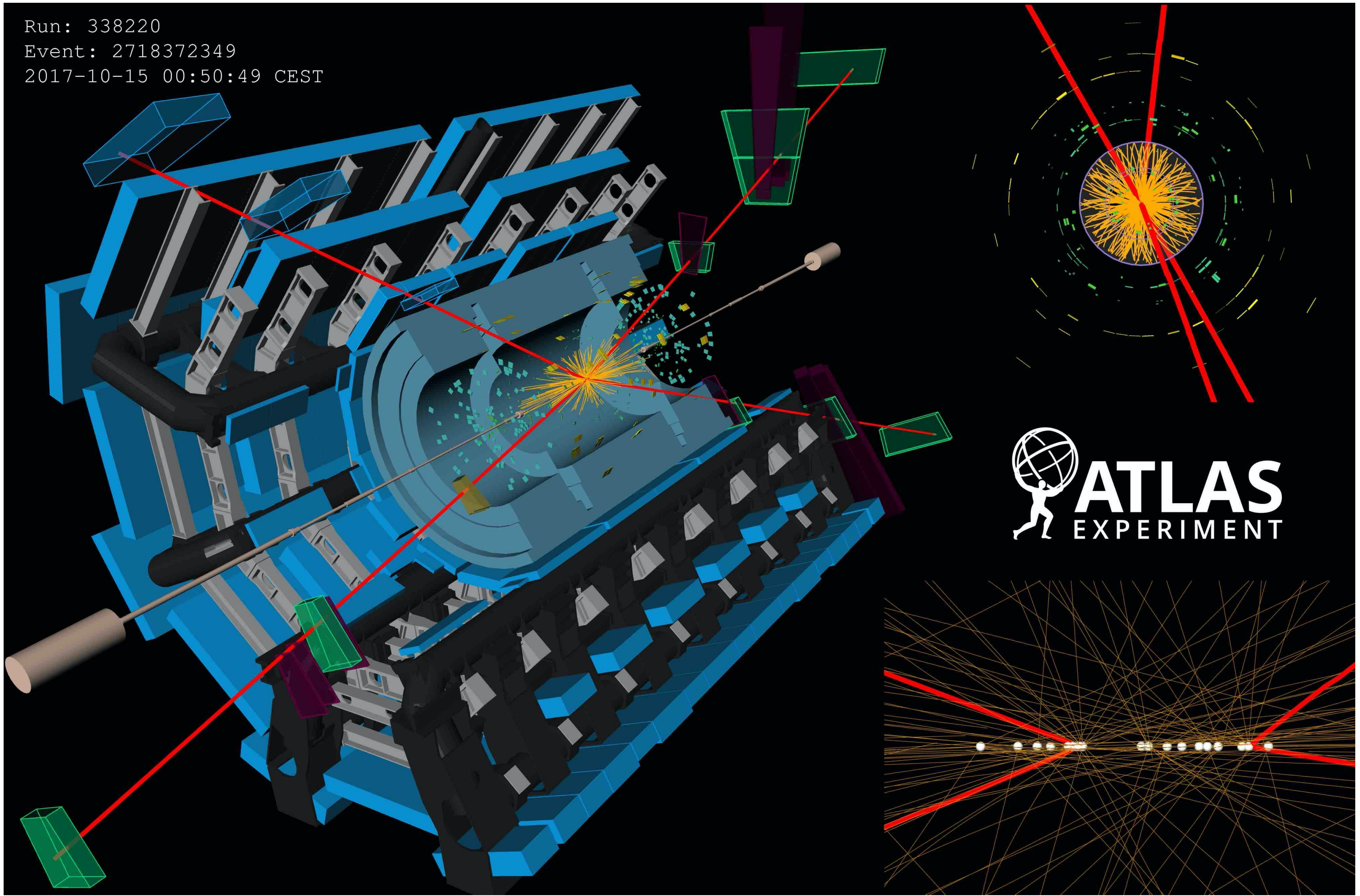


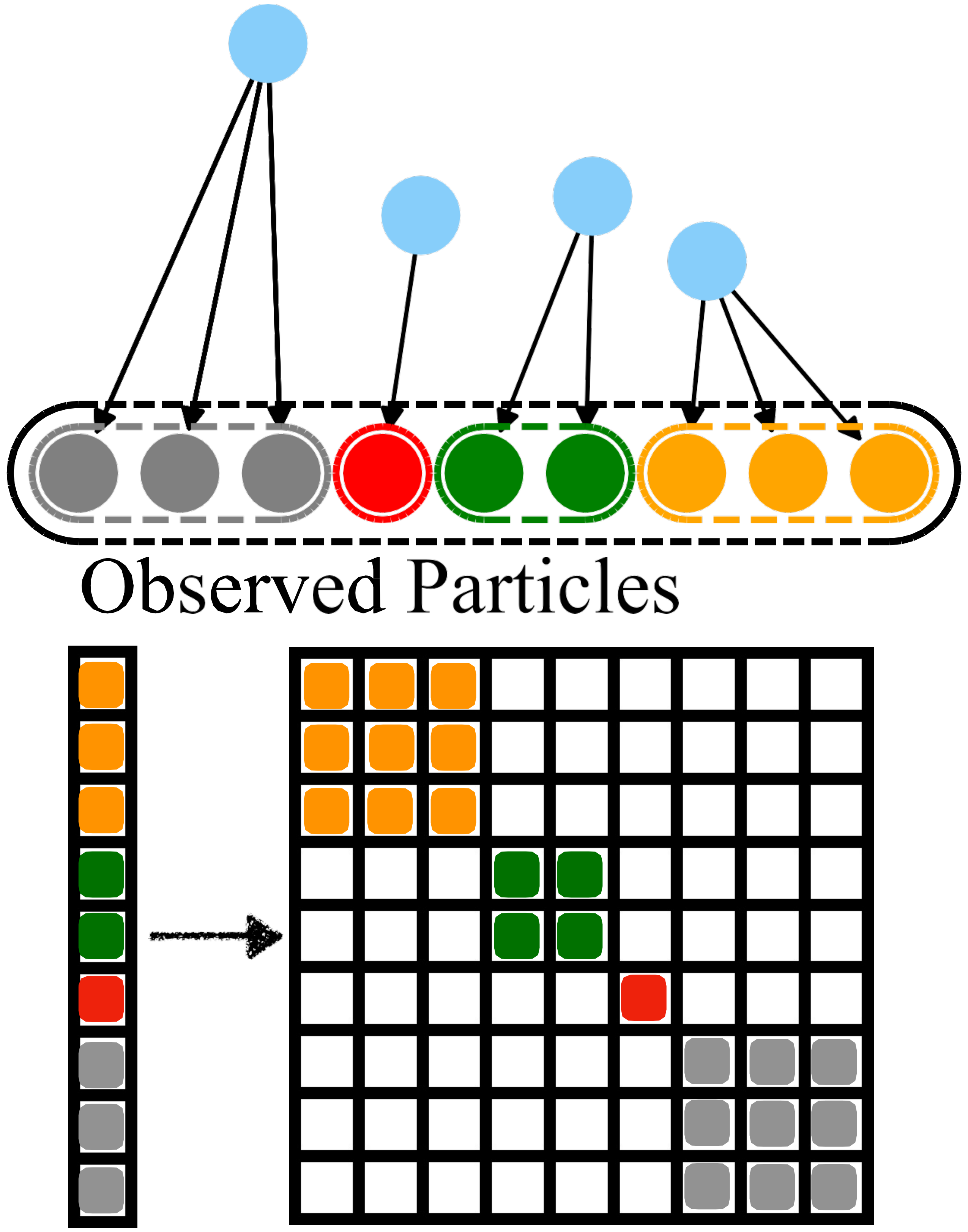
Application to Particle physics

Serviansky, et. al. [arXiv:2002.08772]

Task: group together particles according to common point of origin

- Accuracy is a bottle neck in several searches for new physics





Application to Particle physics

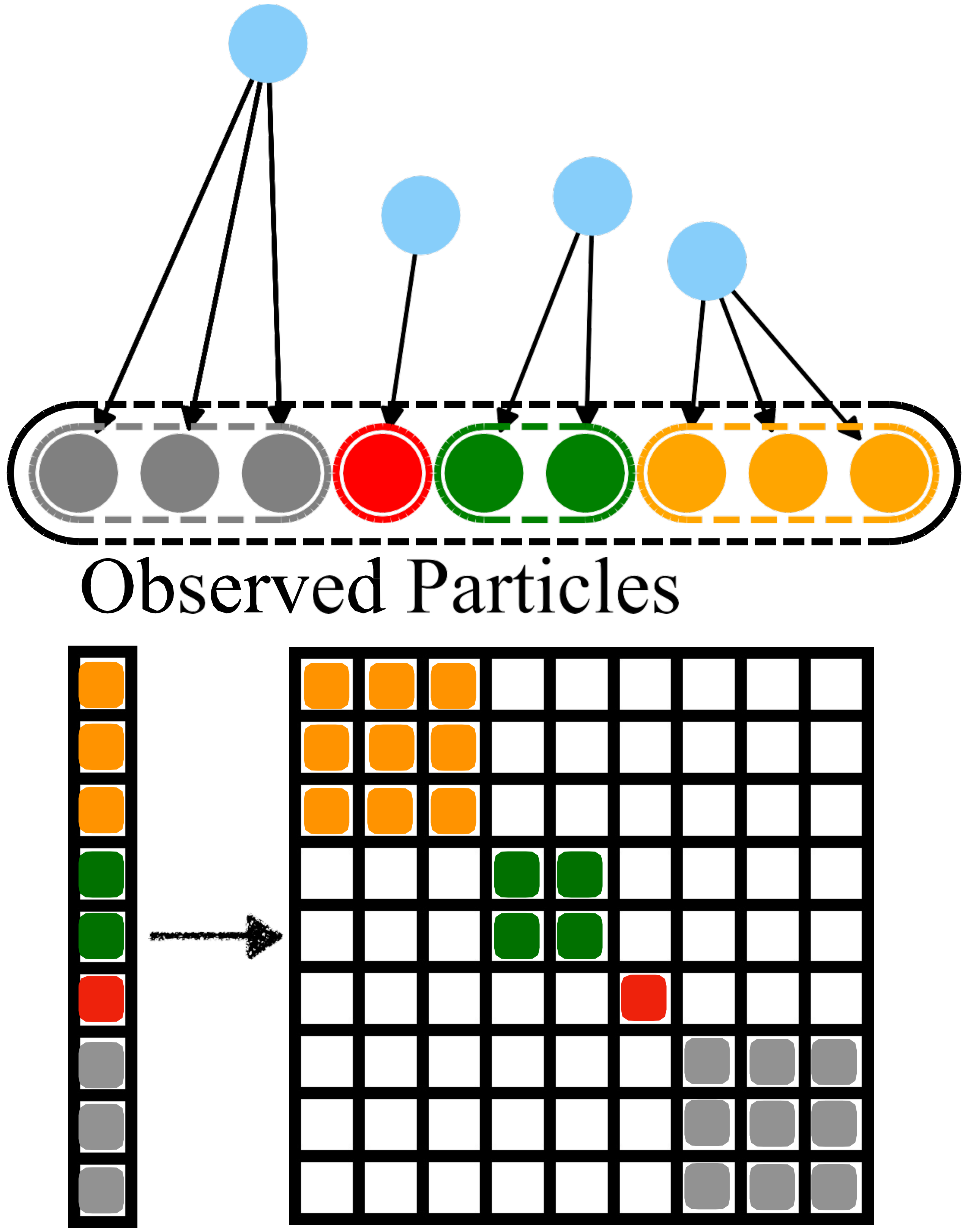
Serviansky, et. al. [arXiv:2002.08772]

Task: group together particles according to common point of origin

- Accuracy is a bottle neck in several searches for new physics

	Model	F1	RI	ARI
B	S2G	0.646±0.003	0.736±0.004	0.491±0.006
	S2G+	**0.655±0.004**	**0.747±0.006**	**0.508±0.007**
	GNN	0.586±0.003	0.661±0.004	0.381±0.005
	SIAM	0.606±0.002	0.675±0.005	0.411±0.004
	SIAM-3	0.597±0.002	0.673±0.005	0.396±0.005
	MLP	0.533±0.000	0.643±0.000	0.315±0.000
	AVR	0.565	0.612	0.318
	trivial	0.438	0.303	0.026
C	S2G	0.747±0.001	0.727±0.003	0.457±0.004
	S2G+	**0.751±0.002**	**0.733±0.003**	**0.467±0.005**
	GNN	0.720±0.002	0.689±0.003	0.390±0.005
	SIAM	0.729±0.001	0.695±0.002	0.406±0.004
	SIAM-3	0.719±0.001	0.710±0.003	0.421±0.005
	MLP	0.686±0.000	0.658±0.000	0.319±0.000
	trivial	0.610	0.472	0.078
	AVR	0.695	0.650	0.326
L	S2G	0.972±0.001	**0.970±0.001**	**0.931±0.003**
	S2G+	0.971±0.002	0.969±0.002	0.929±0.003
	GNN	0.972±0.001	**0.970±0.001**	0.929±0.003
	SIAM	**0.973±0.001**	**0.970±0.001**	0.925±0.003
	SIAM-3	0.895±0.006	0.876±0.008	0.729±0.015
	MLP	0.960±0.000	0.957±0.000	0.894±0.000
	trivial	0.910	0.867	0.675
	AVR	0.970	0.965	0.922

Table 1. Results: partitioning for particle physics.



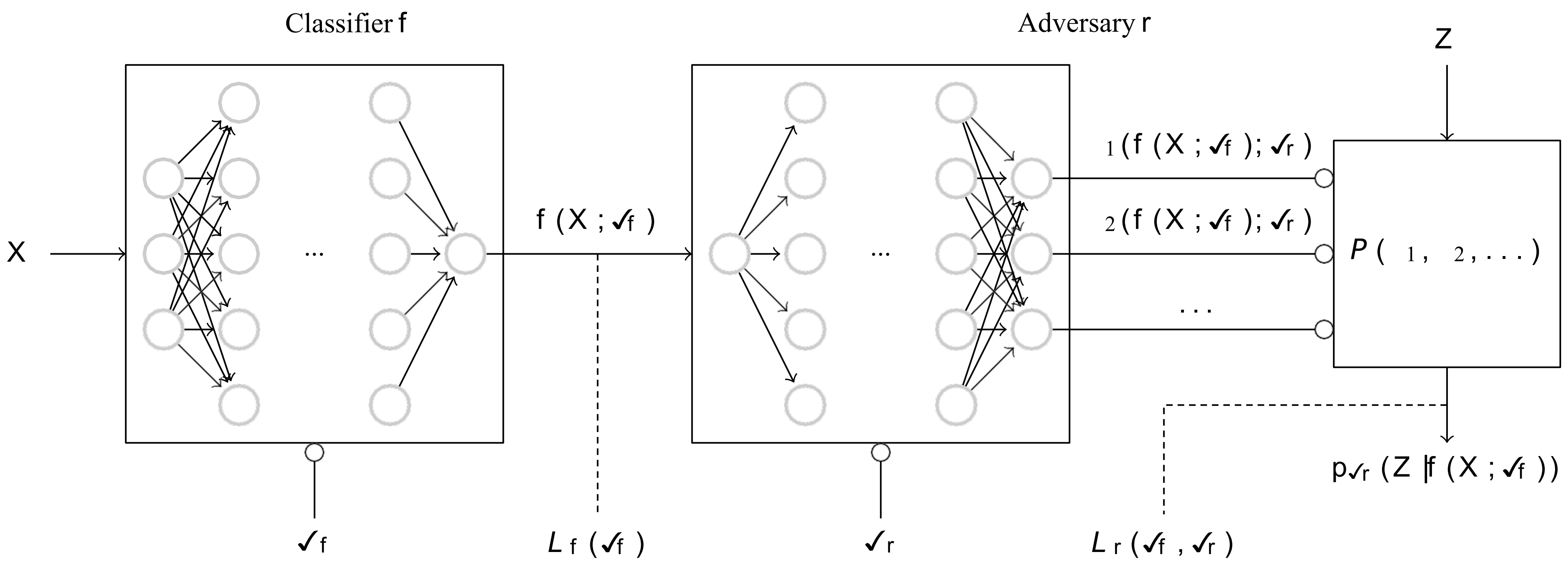
Examples of Physics-inspired ML

From Physics to Fairness

Physics motivation: desire to be robust to systematic uncertainty in experimental particle physics measurements that use machine learning

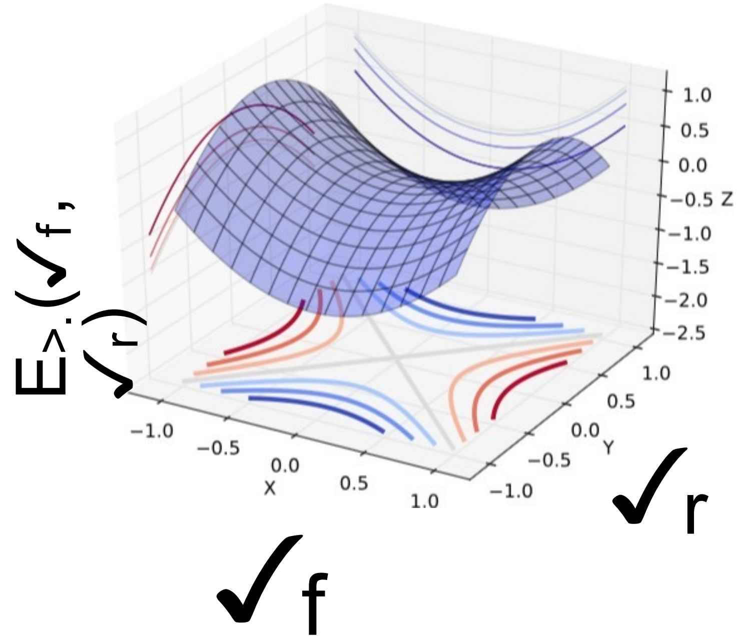
Physics-inspired technique: "learning to pivot with adversarial networks"

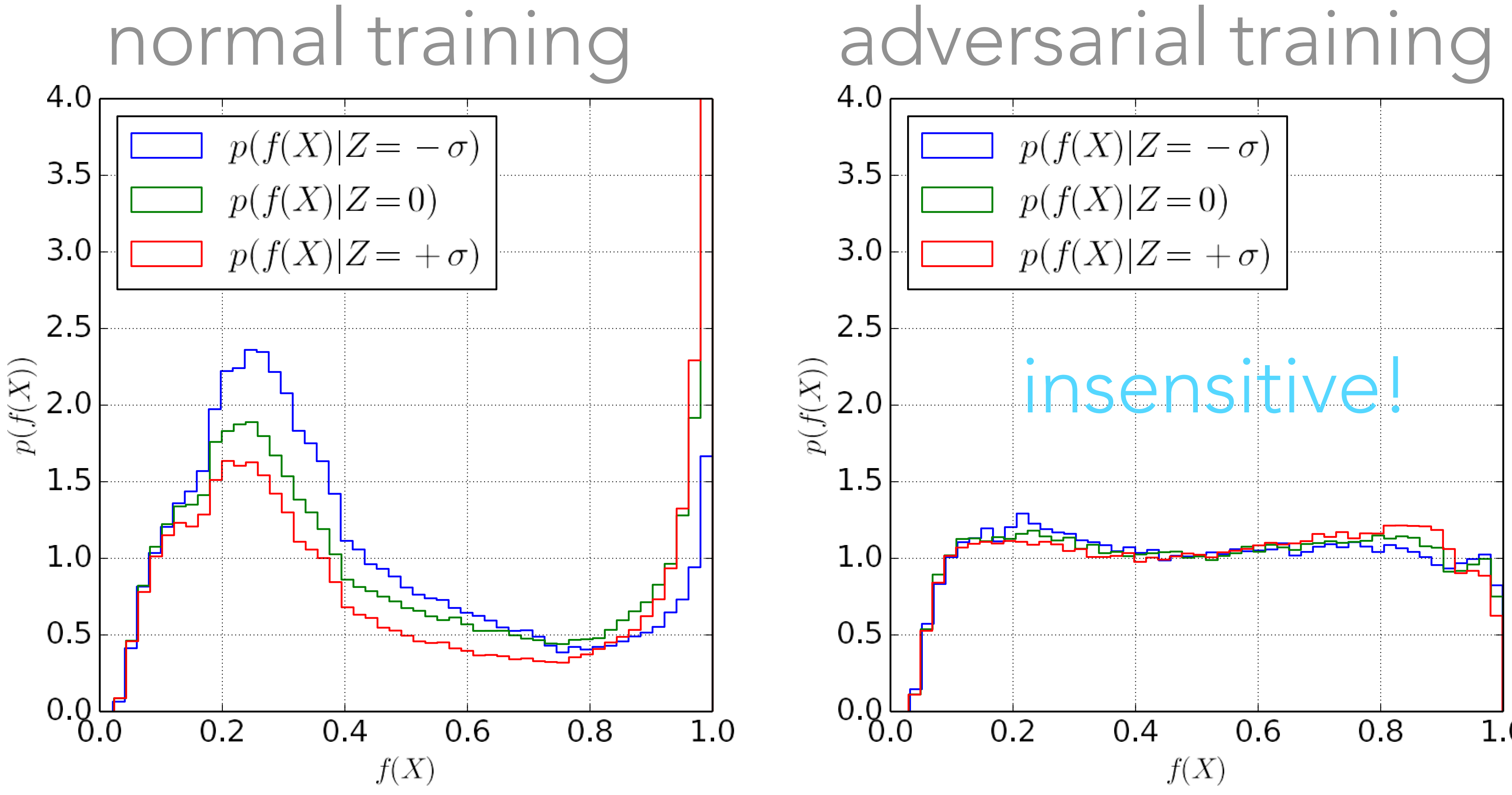
- introduce an **adversary r** that tries to predict nuisance parameter based on classifier output **f**



$$\hat{\theta}_f, \hat{\theta}_r = \arg\min_{\theta_f} \max_{\theta_r} E(\theta_f, \theta_r).$$

$$E_\lambda(\theta_f, \theta_r) = L_f(\theta_f) \qquad L_r(\theta_f, \theta_r)$$





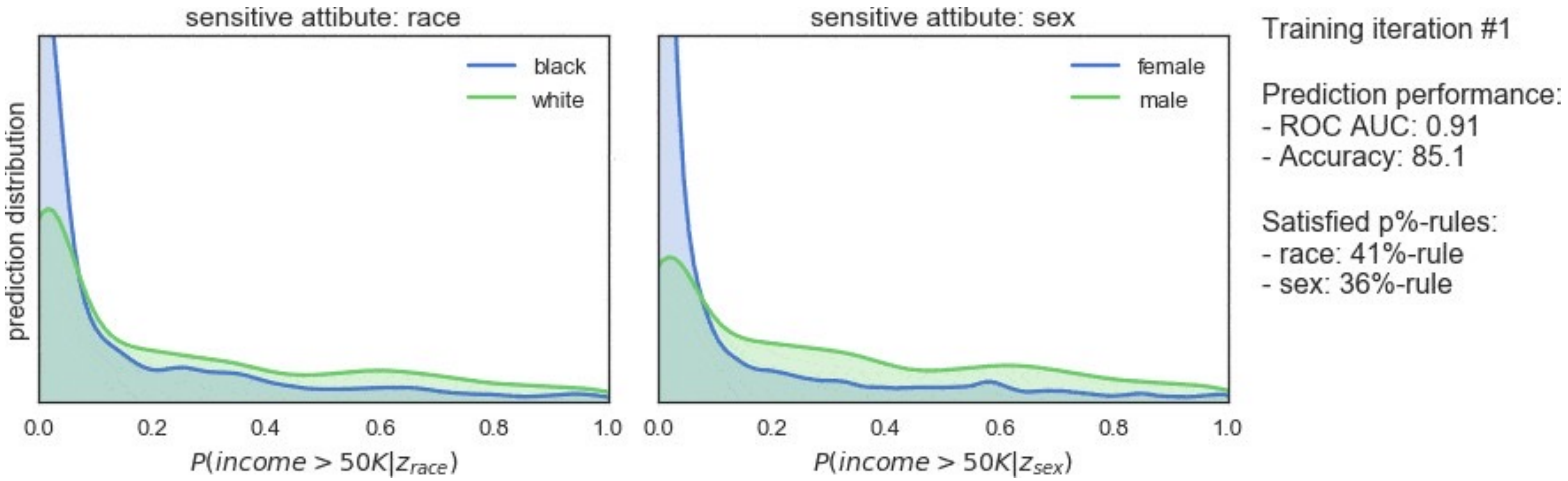
G. Louppe, M. Kagan, K. Cranmer, arXiv:1611.01046, NeurIPS2017

From Physics to Fairness

Physics motivation: desire to be robust to systematic uncertainty in experimental particle physics measurements that use machine learning

Physics-inspired technique: "learning to pivot

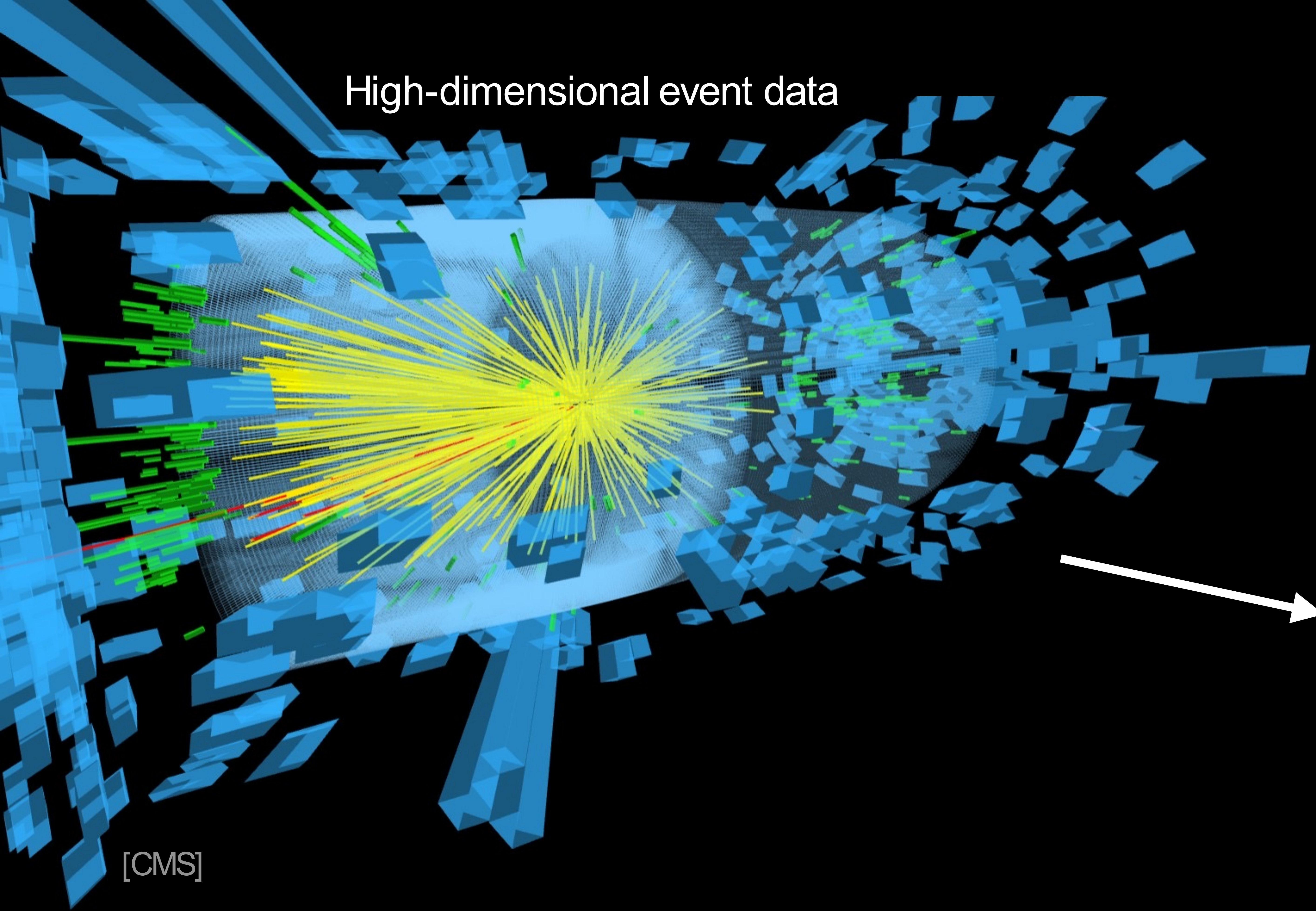
Impact: connection to domain adaptation, algorithmic fairness, privacy, and encryption
Mentioned in Leon Bottou's 2019 ICML keynote on Invariant Risk Minimization



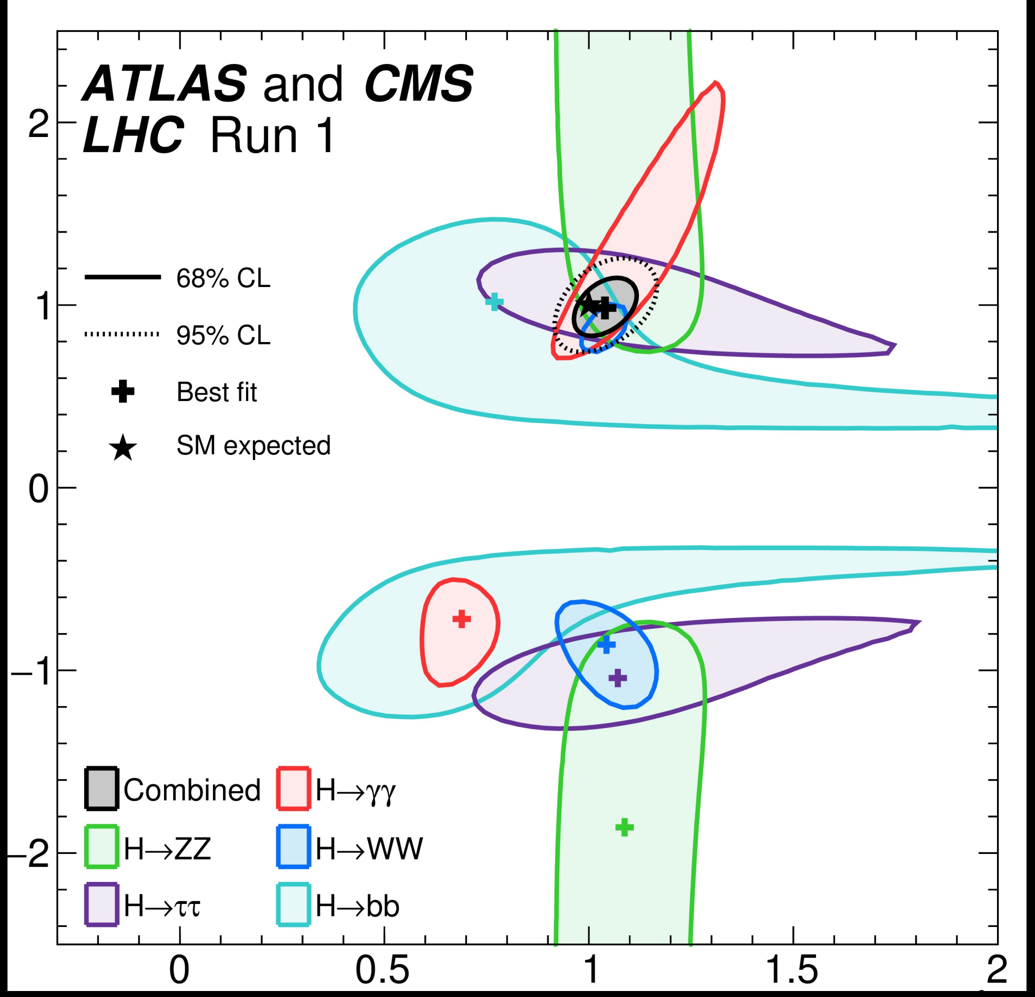
Stijn Tonk, 27 April, 2018
https://godatadriven.com/blog/towards-fairness-in-ml-with-adversarial-networks/

Examples of Physics-inspired ML

High-dimensional event data



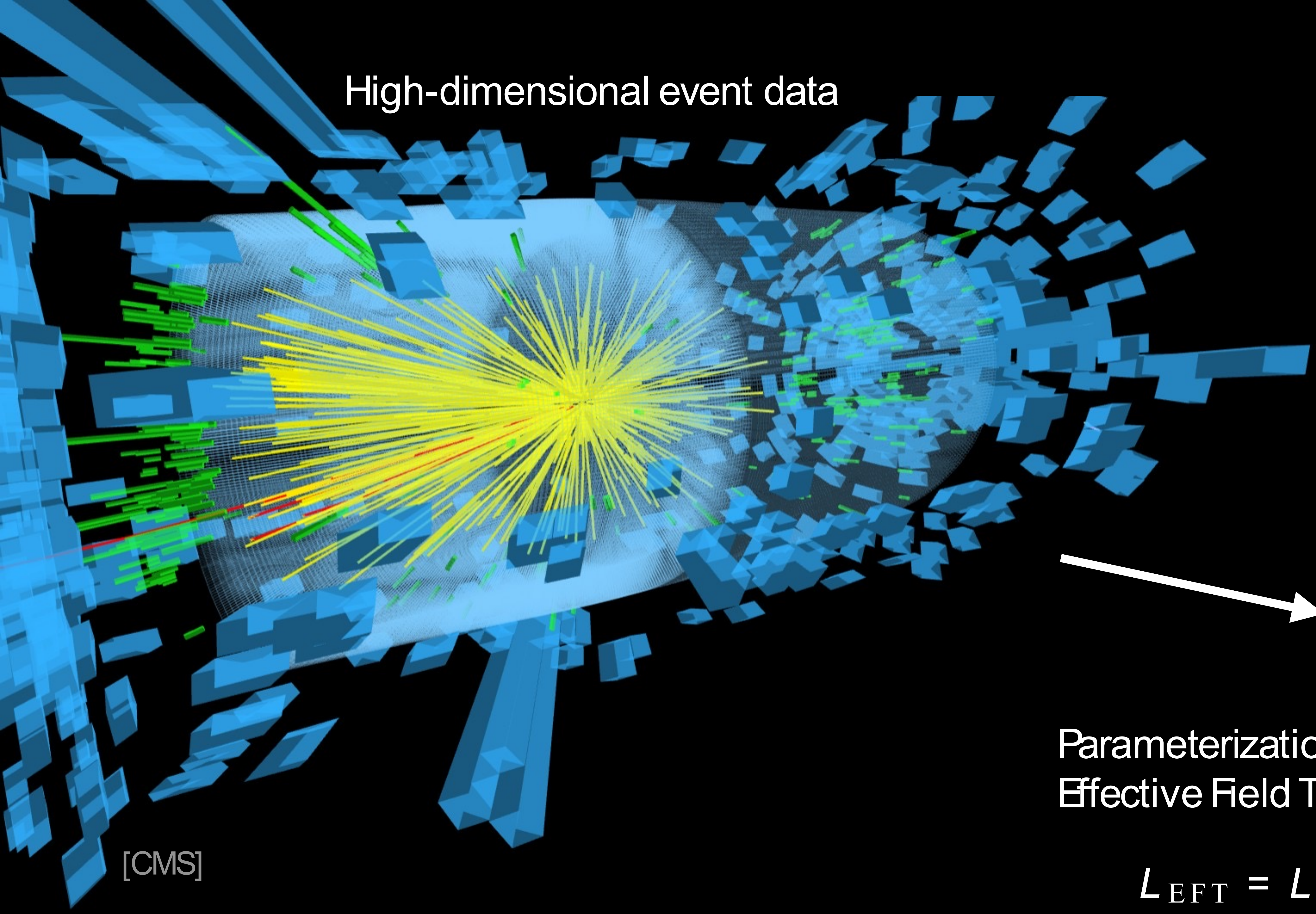
[CMS]



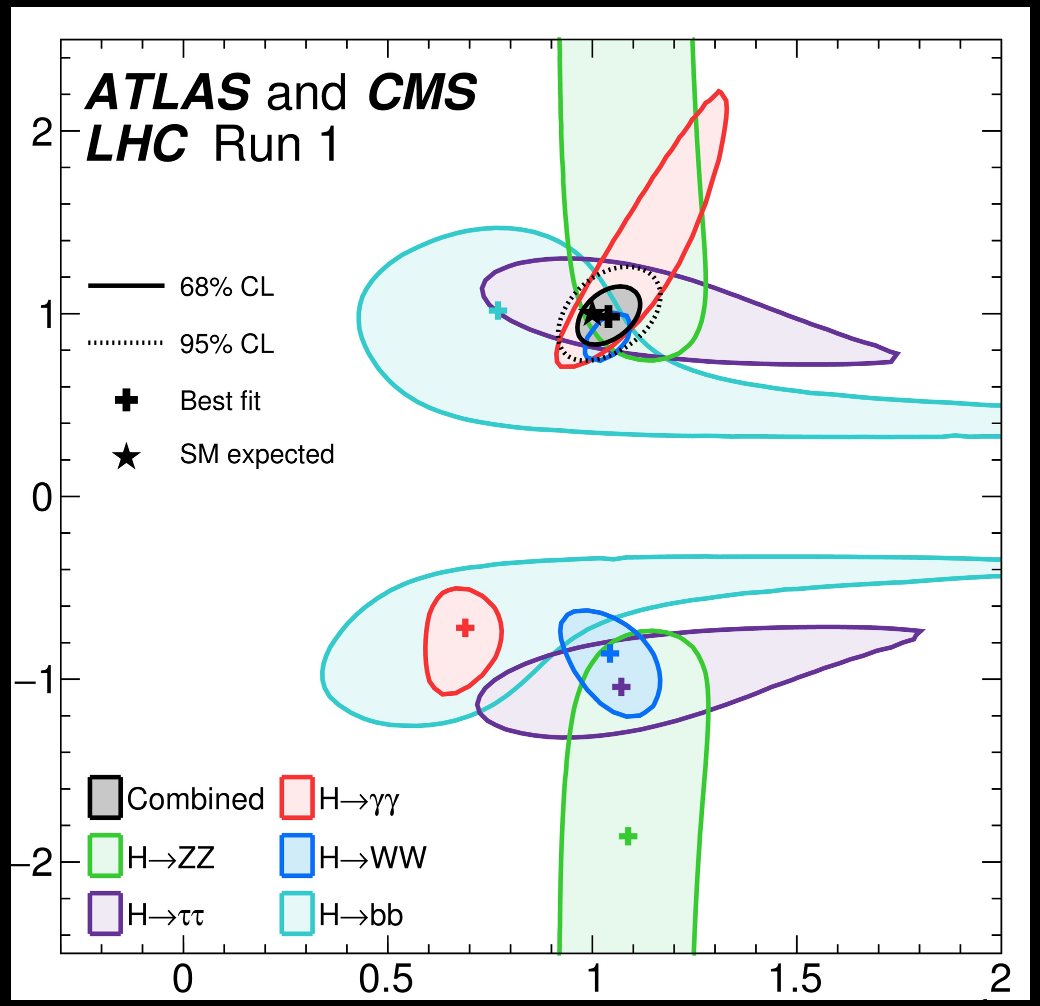
[ATLAS, CMS 1606.02266]

Precision constraints on new physics

High-dimensional event data



[CMS]



[ATLAS, CMS 1606.02266]

Precision constraints on new physics

Parameterization e.g. in Effective Field Theory:

systematic expansion of new physics around Standard Model

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i + \ldots$$

10s to 100s "universal" parameters to measure

Simulating particle physics processes

Theory
parameters
✓

Evolution

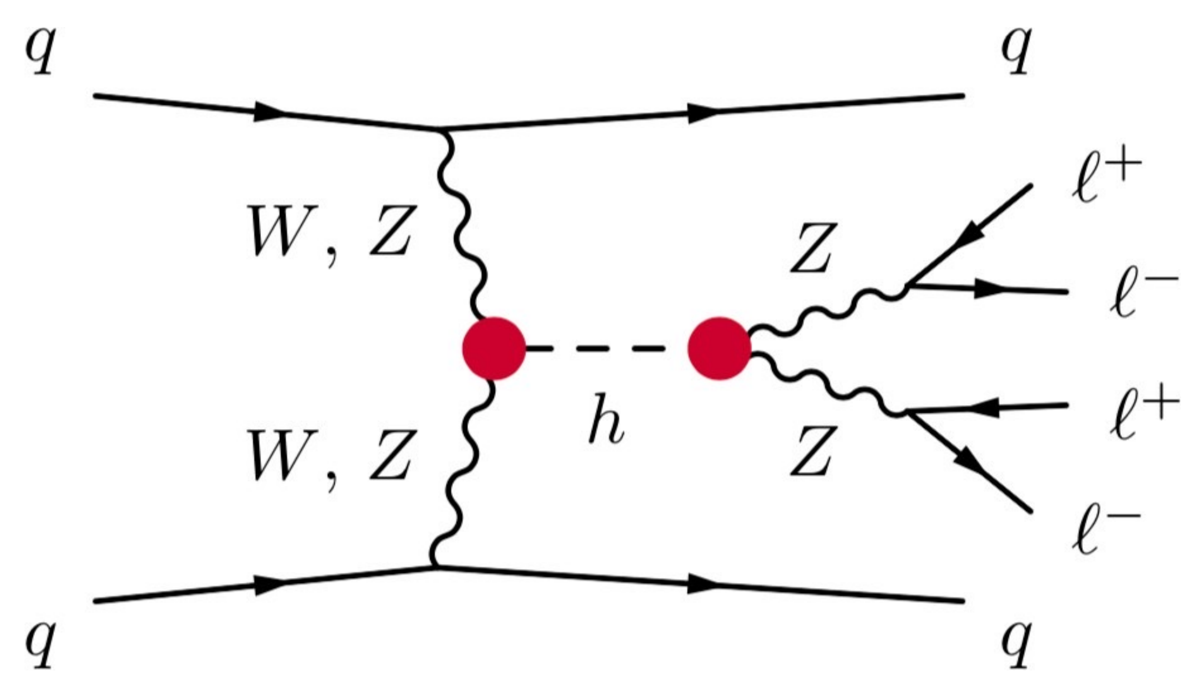
Simulating particle physics processes

Latent variables

Parton-level momenta

Theory parameters

$z_p \longleftarrow \checkmark$

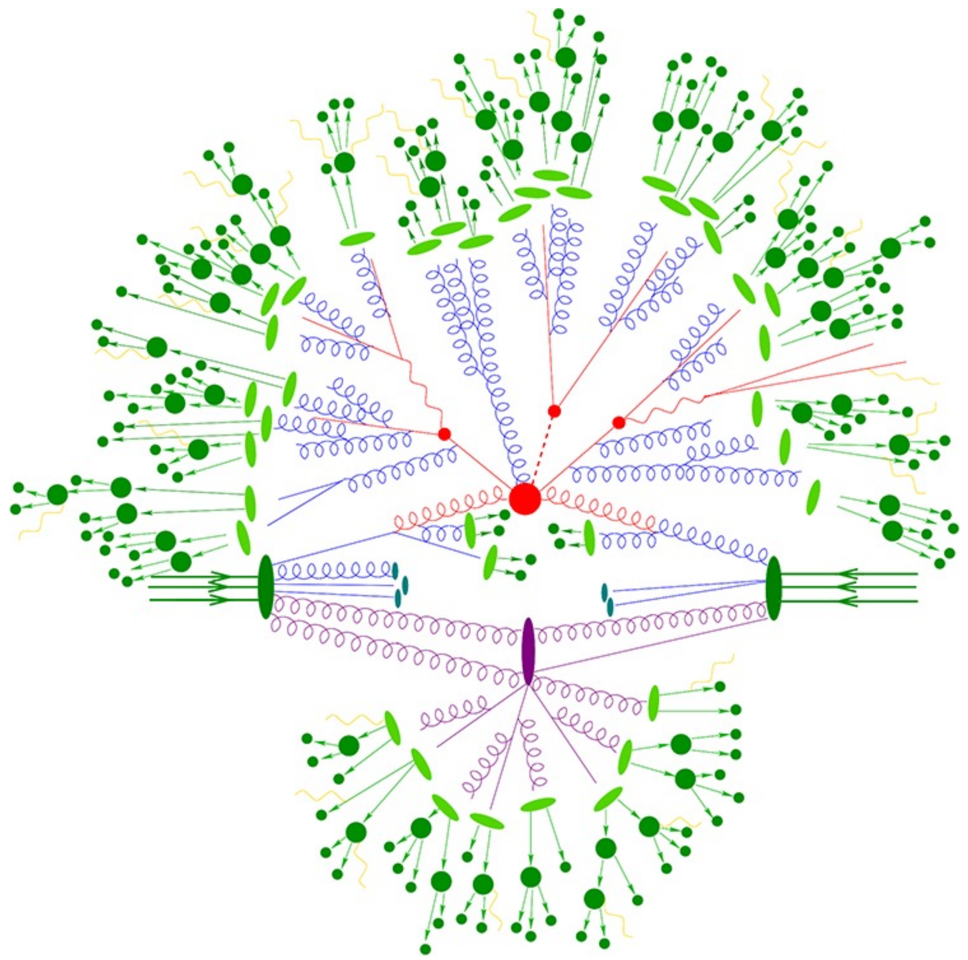


Evolution

Simulating particle physics processes

Latent variables

Shower splittings	Parton-level momenta	Theory parameters
z_s ⟵	z_p ⟵	✓



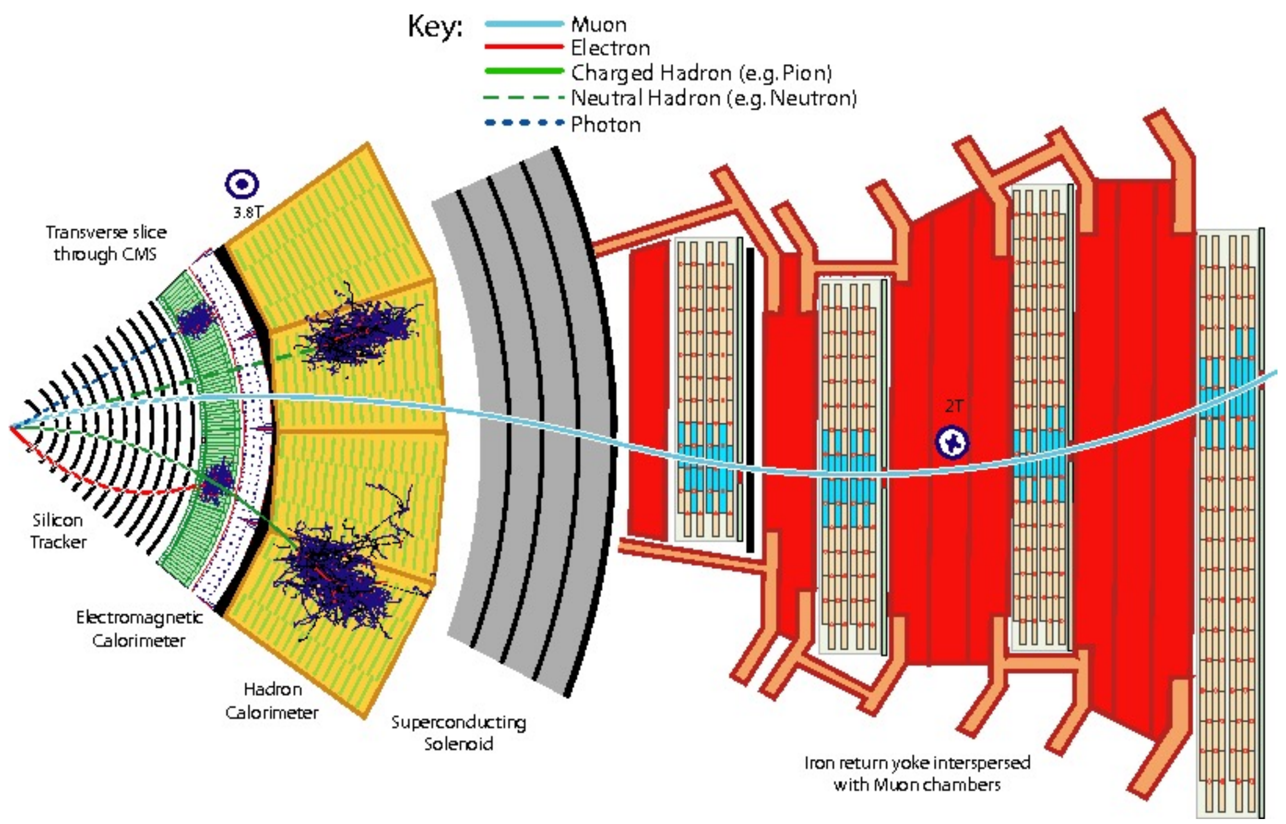
[F. Krauss]

⟵ Evolution

Simulating particle physics processes

Latent variables

Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
z_d ←	z_s ←	z_p ←	✓



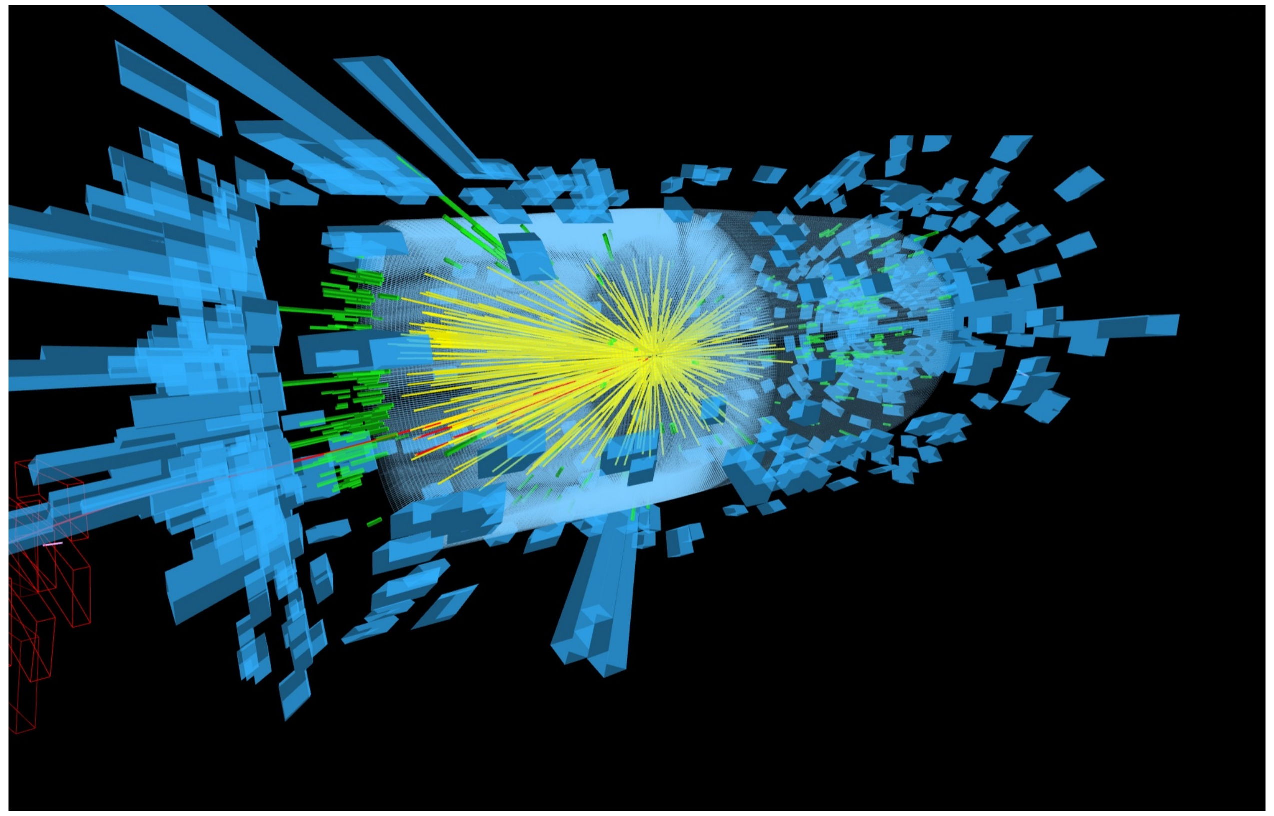
[CMS]

← Evolution

Simulating particle physics processes

Latent variables

Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
←	z_d ←	z_s ←	z_p ←	✓



Evolution ←

Simulating particle physics processes

Latent variables

Observables	Detector interactions	Shower splittings	Parton-level momenta	Theory parameters
←	z_d ←	z_s ←	z_p ←	✓

Sample from $p(x \mid z_d)$ $p(z_d \mid z_s)$ $p(z_s \mid z_p)$ $p(z_p \mid ✓)$

GEANT4 A SIMULATION TOOLKIT

MADGRAPH5_aMC@NLO

Prediction (simulation)

Simulating particle physics processes

Latent variables

Observables ← Detector interactions ← Shower splittings ← Parton-level momenta ← Theory parameters

$z_d \leftarrow z_s \leftarrow z_p \leftarrow$ ✓

$p(x|✓) = Z\ dz_d\ Z\ dz_s\ Z\ dz_p\ p(x|z_d)\ p(z_d|z_s)\ p(z_s|z_p)\ p(z_p|✓)$

Inference

Simulating particle physics processes

Latent variables

Observables		Detector interactions		Shower splittings		Parton-level momenta		Theory parameters
	←	z_d	←	z_s	←	z_p	←	✓

$$p(x|✓) = \boxed{Z\, dz_d\; Z\, dz_s\; Z\, dz_p}\; p(x|z_d)\quad p(z_d|z_s)\quad p(z_s|z_p)\quad p(z_p|✓)$$

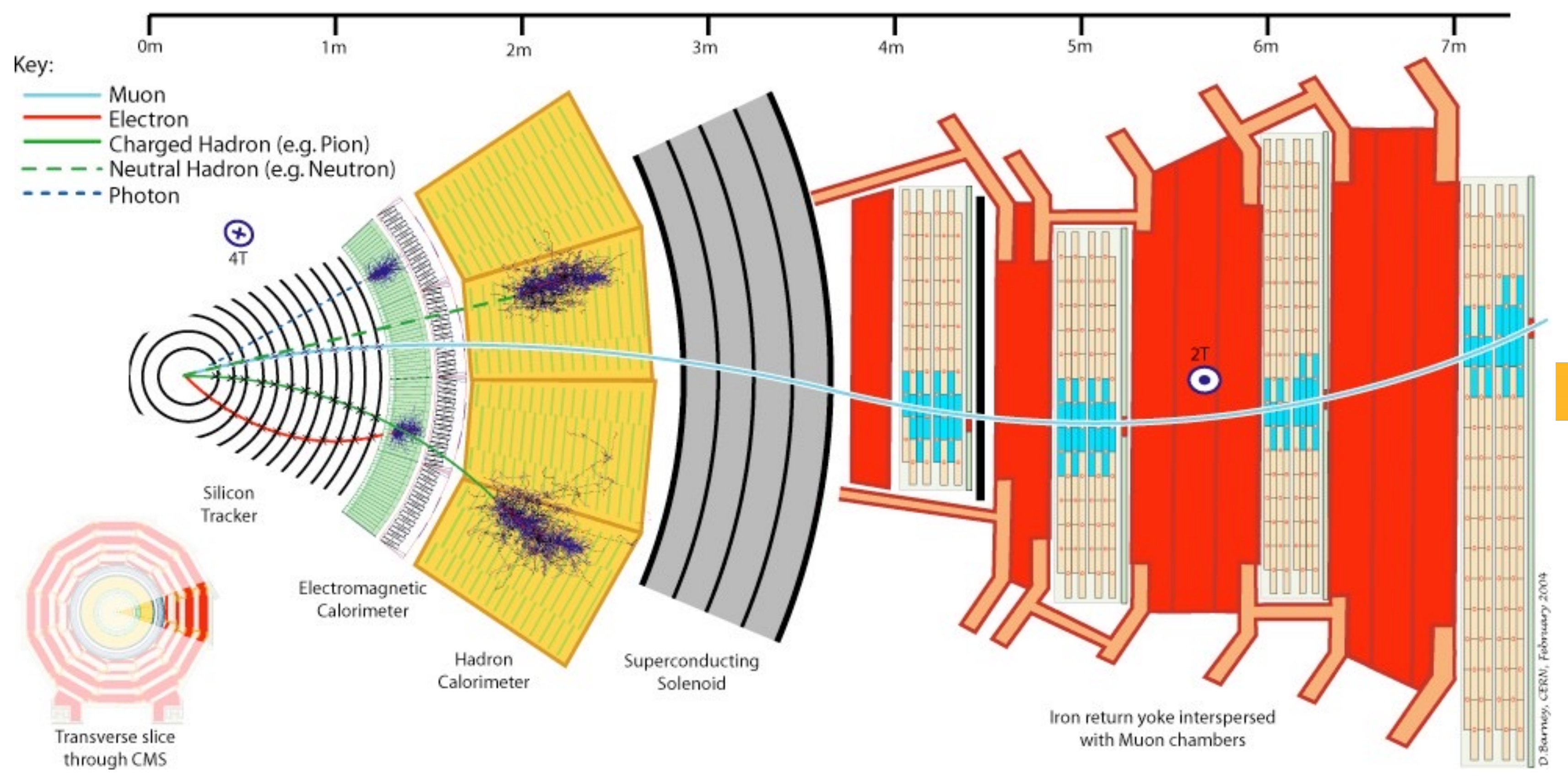
It's infeasible to calculate the integral over this enormous space!

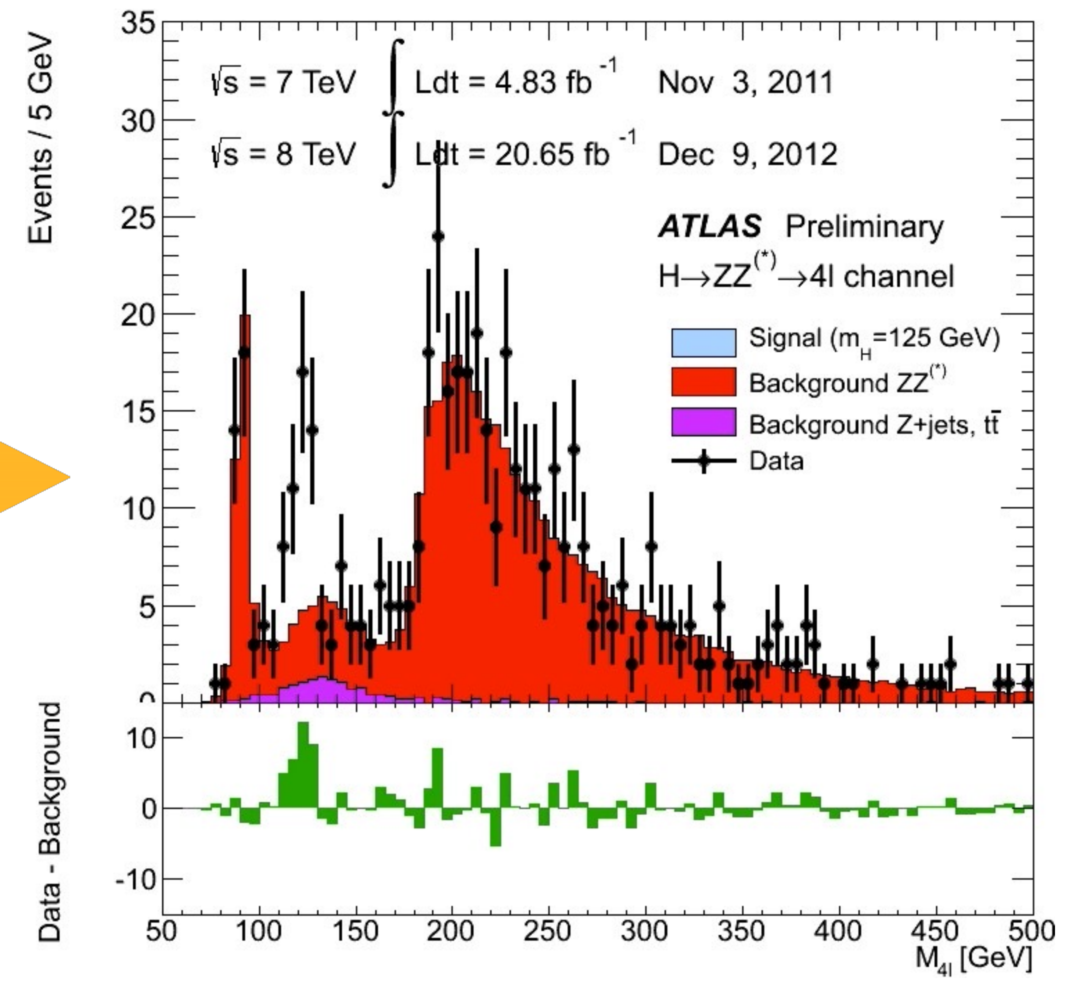
Inference →

10^8 sensors → summary statistic

Most measurements and searches for new particles at the LHC are based on the distribution of a single **summary statistic**

- choosing a good summary statistic $s(x$ (feature engineering) is a task for a skilled physicist and tailored to the goal of measurement or new particle search
- likelihood $p(s$) **approximated** using histograms or kernel density estimation [Similar to Diggle & Gratton (1984)]

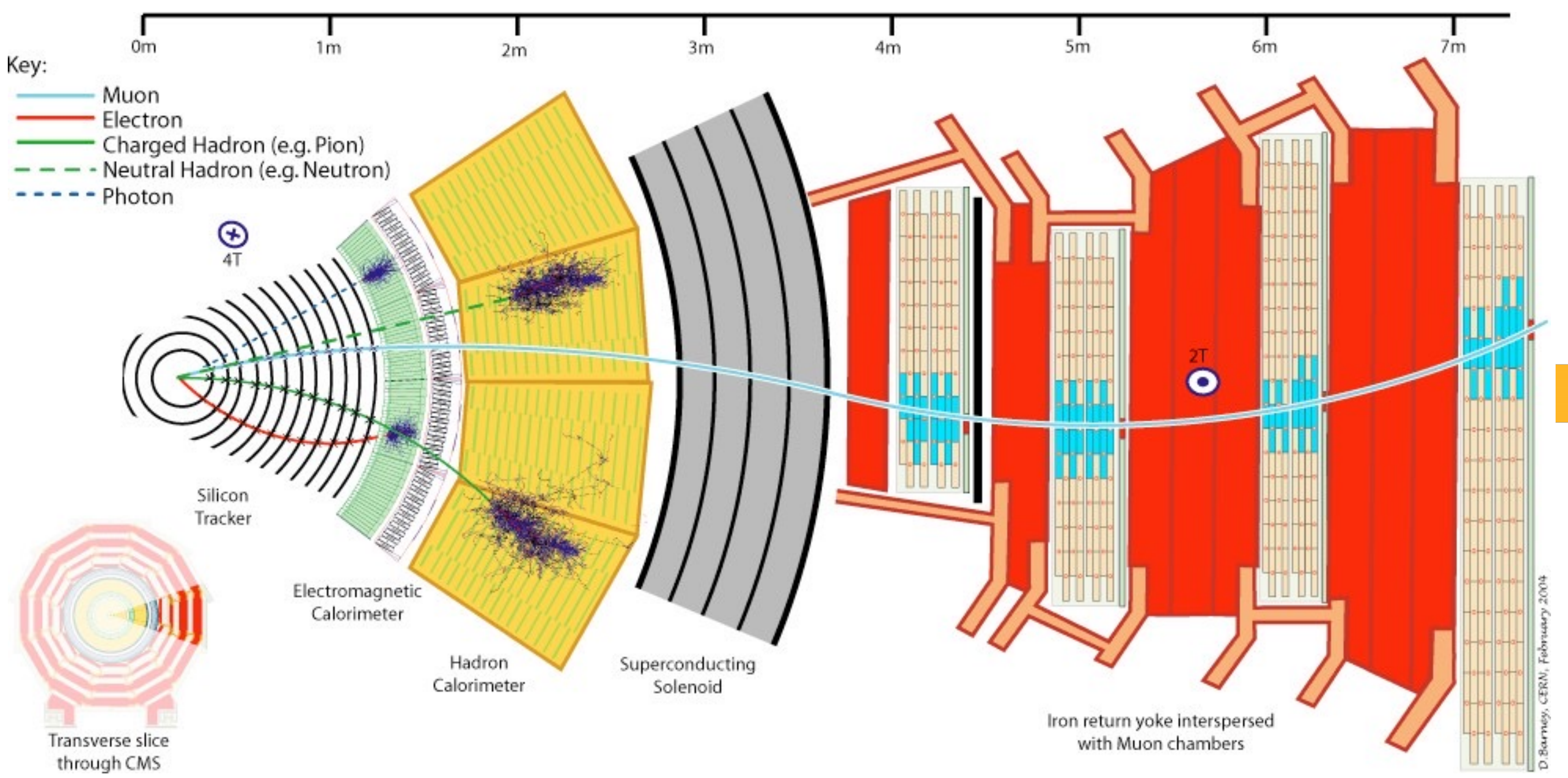


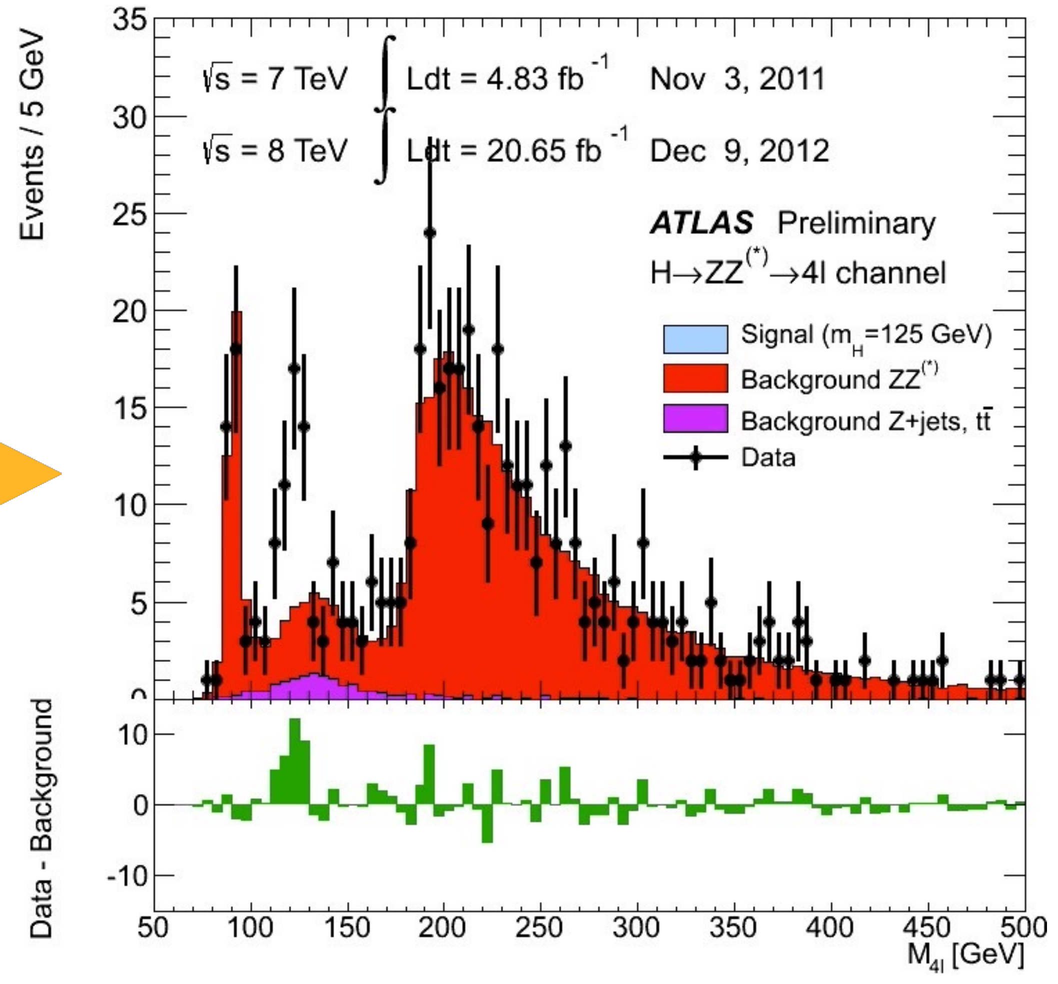


10^8 sensors → summary statistic

Most measurements and searches for new particles at the LHC are based on the distribution of a single **summary statistic**

- choosing a good summary statistic $s(x$ (feature engineering) is a task for a skilled physicist and tailored to the goal of measurement or new particle search
- likelihood $p(s$) **approximated** using histograms or kernel density estimation [Similar to Diggle & Gratton (1984)]

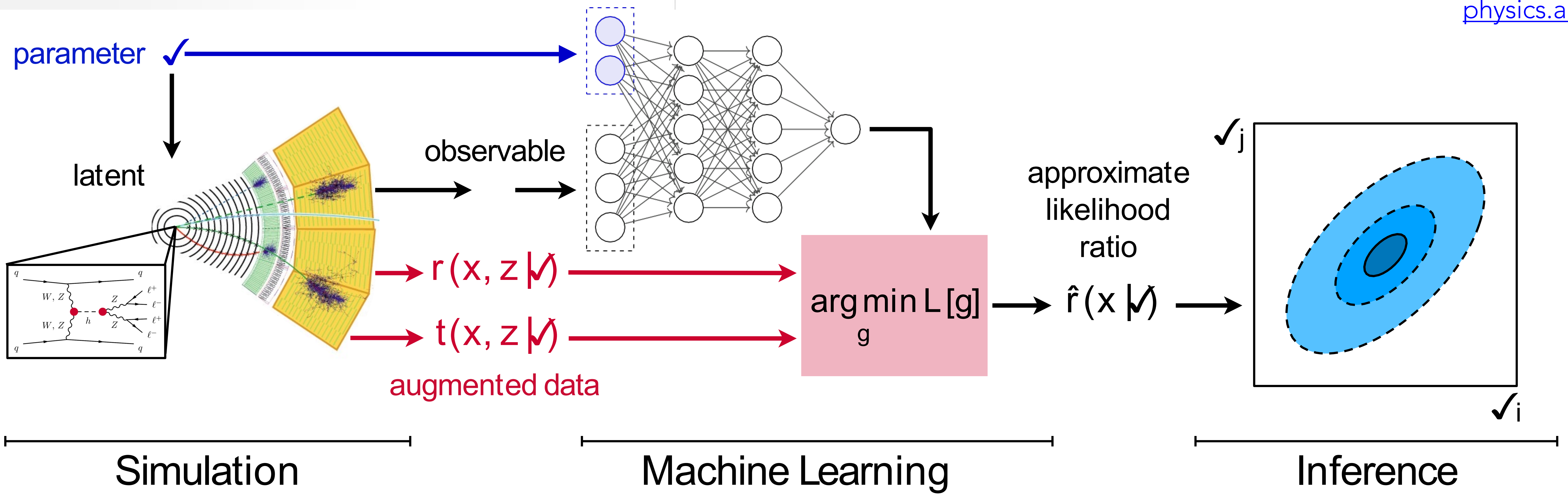


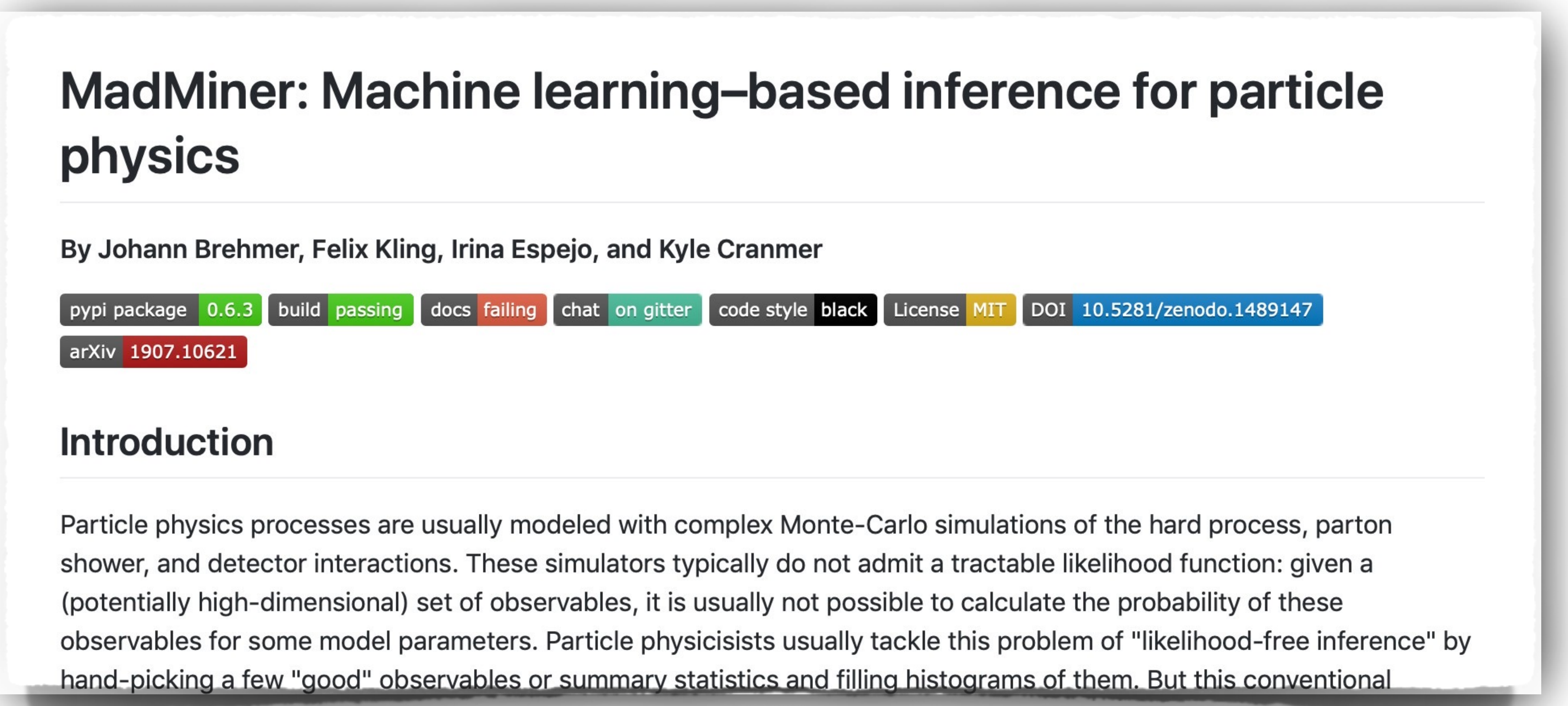


This doesn't scale if summary is high dimensional!

Learning the likelihood ratio

arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
physics.aps.org/articles/v11/90





MadMiner: Machine learning–based inference for particle physics

By Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer

pypi package 0.6.3 | build passing | docs failing | chat on gitter | code style black | License MIT | DOI 10.5281/zenodo.1489147 | arXiv 1907.10621

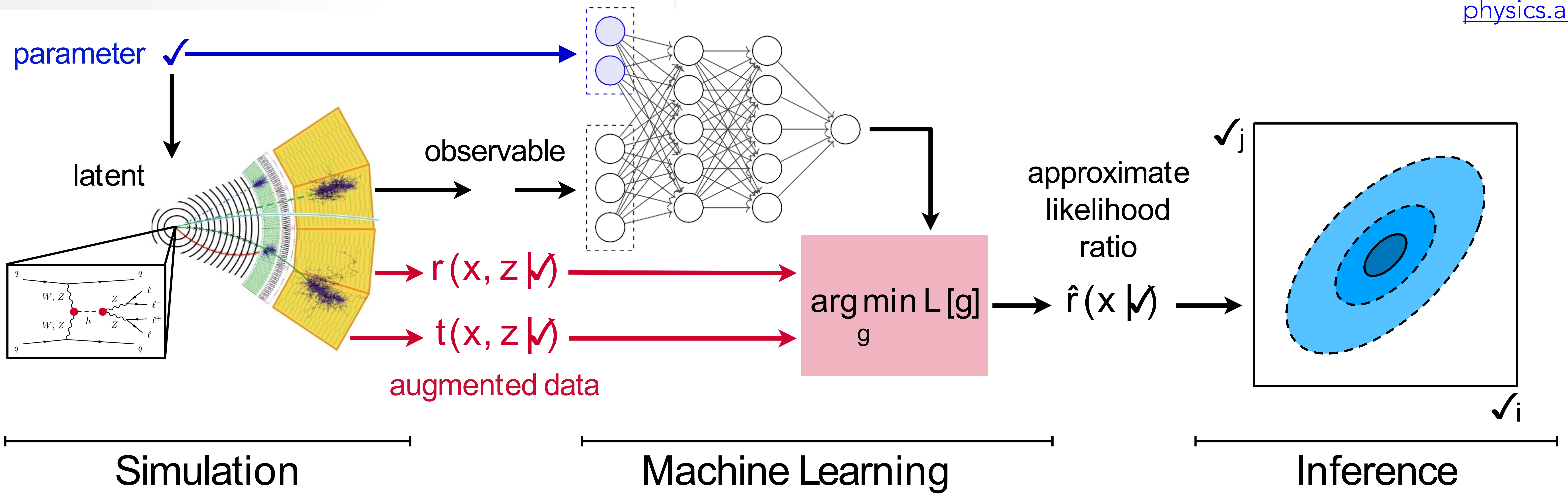
Introduction

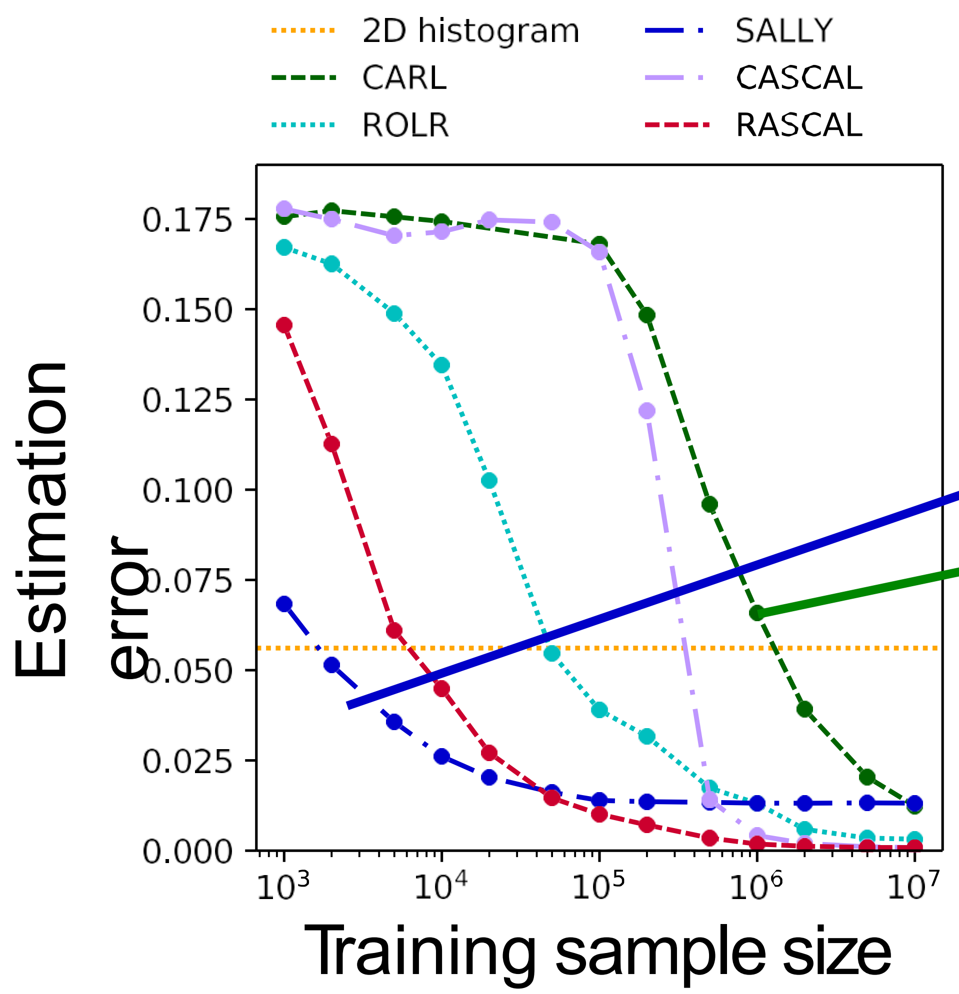
Particle physics processes are usually modeled with complex Monte-Carlo simulations of the hard process, parton shower, and detector interactions. These simulators typically do not admit a tractable likelihood function: given a (potentially high-dimensional) set of observables, it is usually not possible to calculate the probability of these observables for some model parameters. Particle physicisists usually tackle this problem of "likelihood-free inference" by hand-picking a few "good" observables or summary statistics and filling histograms of them. But this conventional

Dedicated software package interfacing with particle physics simulators:
github.com/johannbrehmer/madminer

Learning the likelihood ratio

arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
physics.aps.org/articles/v11/90





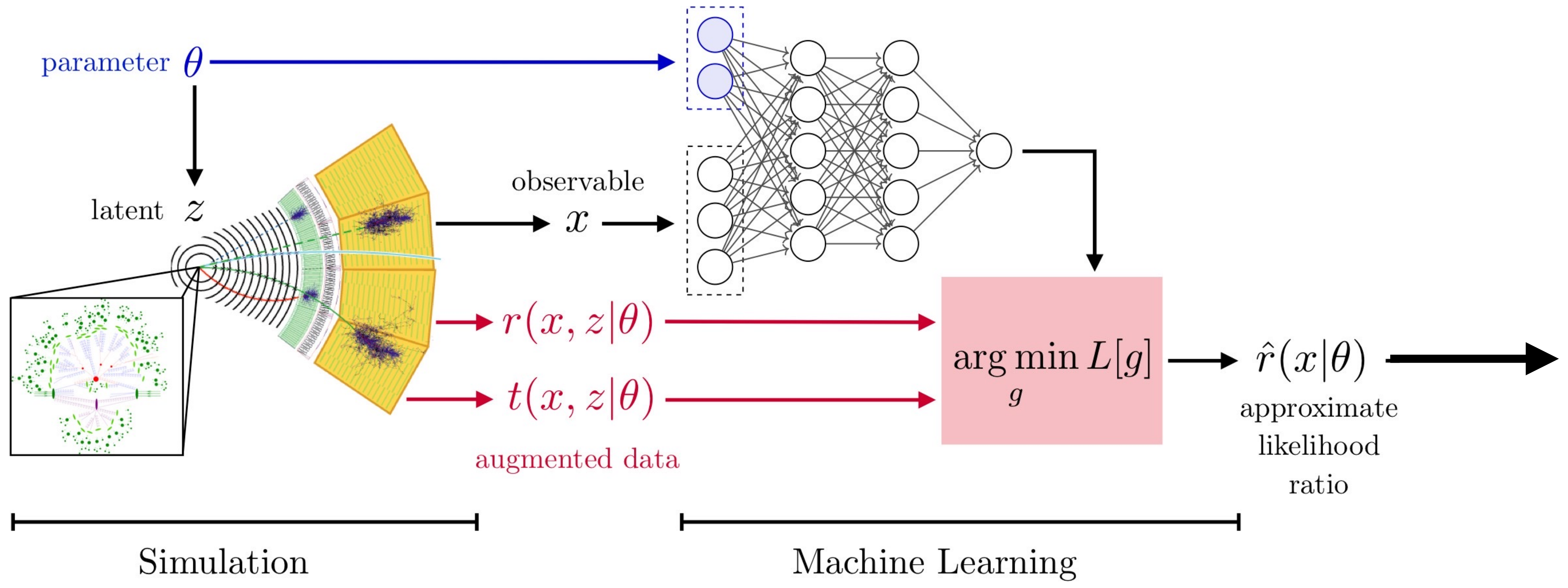
New techniques
require less data than
without augmented data

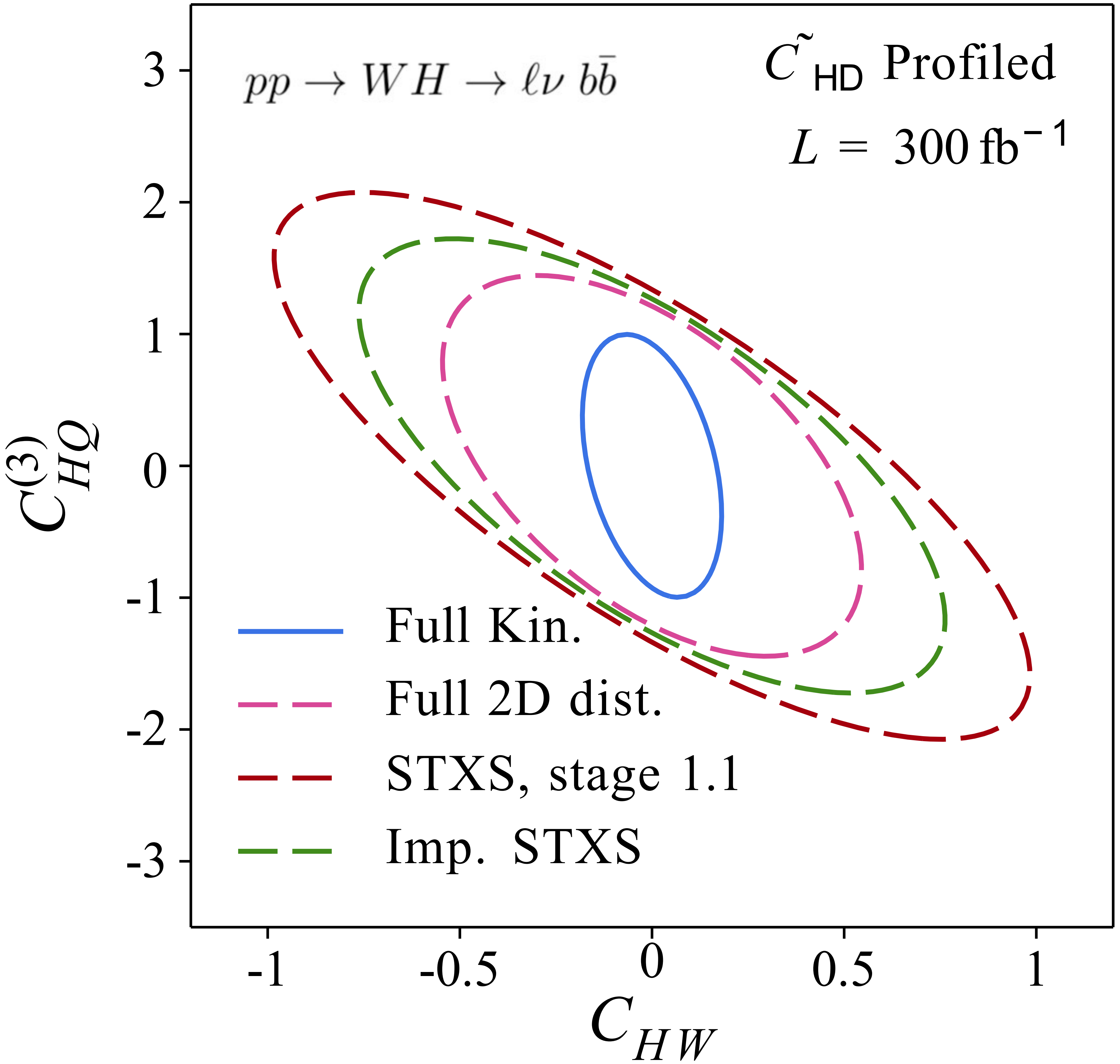
Traditional Approach no NN

Impact on Studies of The Higgs Boson

Massive gains in precision of a flagship measurement at the LHC !

Equivalent increasing data collected by LHC by several factors





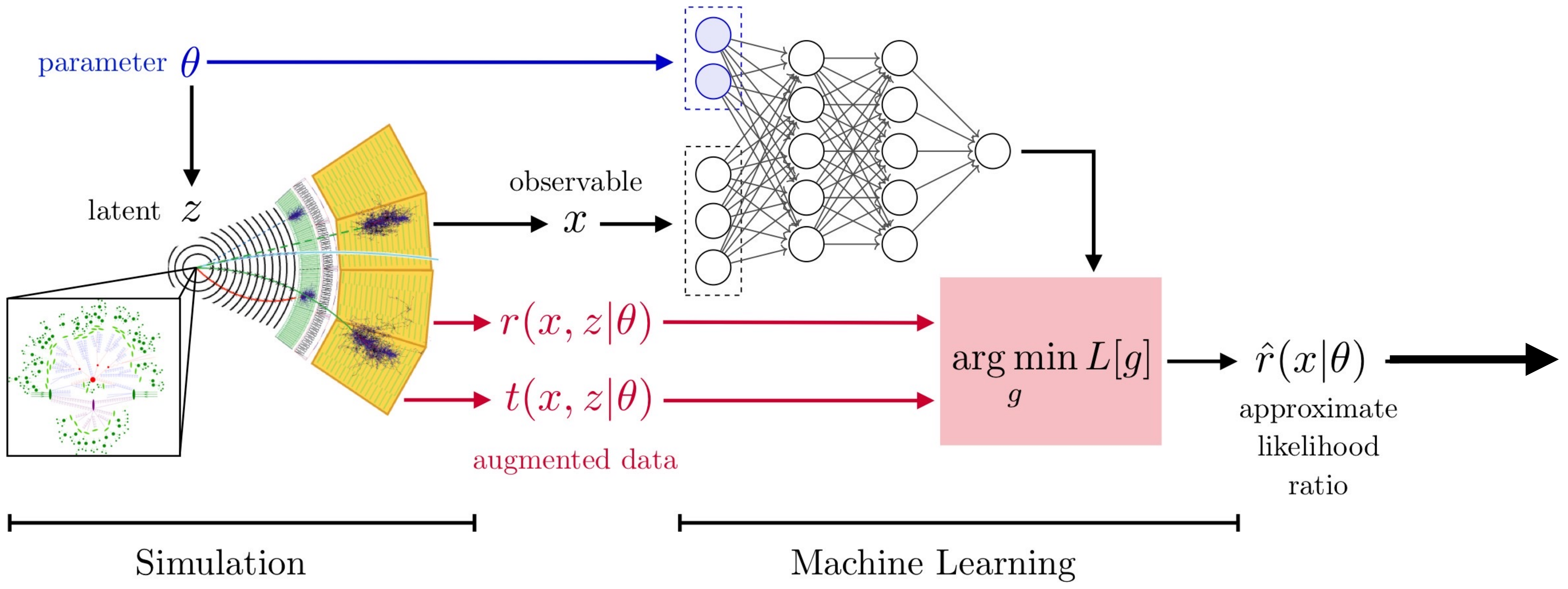
[J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn 1908.06980]

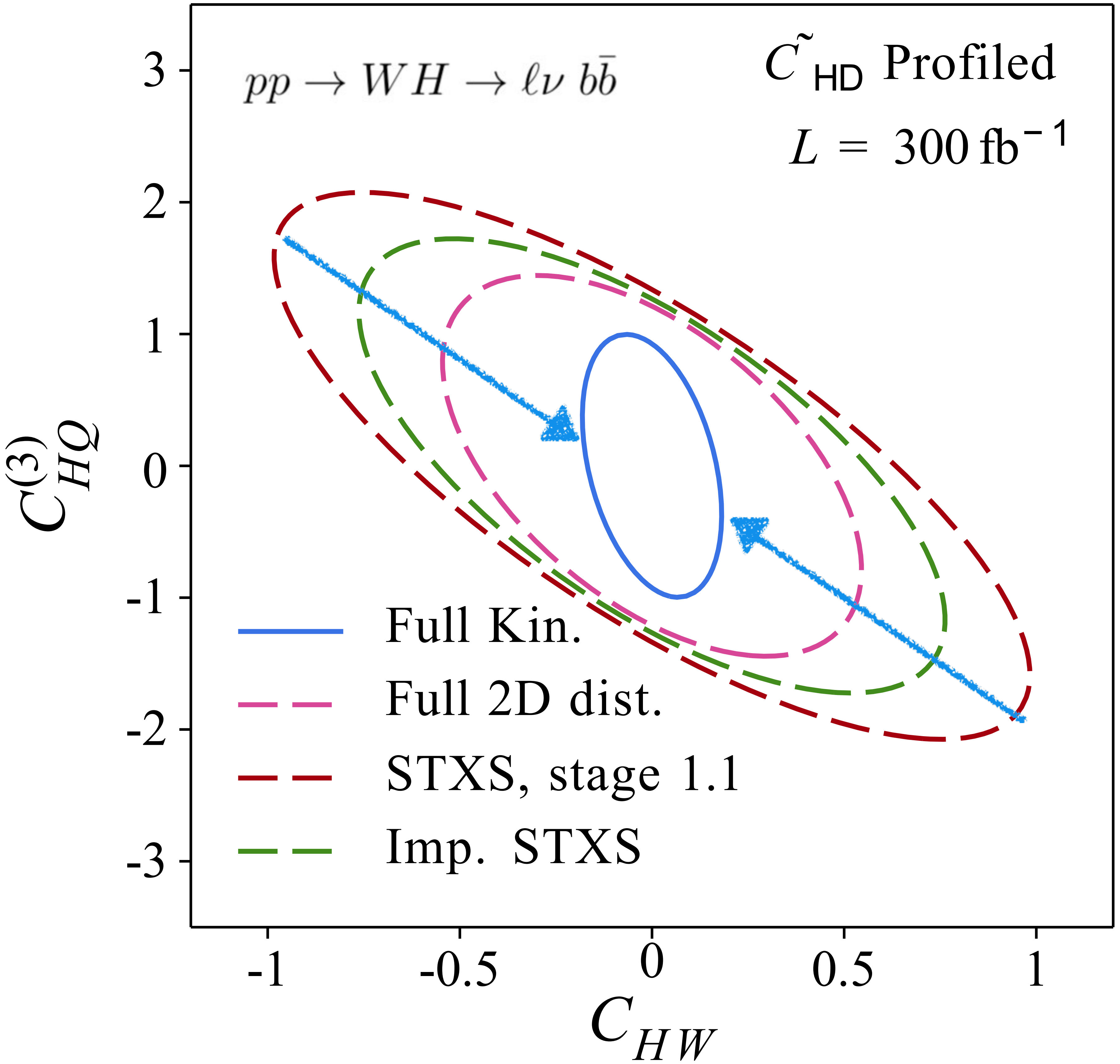
[J. Brehmer, F. Kling, I. Espejo, K. Cranmer 1907.10621]

Impact on Studies of The Higgs Boson

Massive gains in precision of a flagship measurement at the LHC !

Equivalent increasing data collected by LHC by several factors





[J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn 1908.06980]

[J. Brehmer, F. Kling, I. Espejo, K. Cranmer 1907.10621]

A common theme, a common language

ABC

resources on approximate Bayesian computational methods

Search

Home

Home

This website keeps track of developments in approximate Bayesian computation (ABC) (a.k.a. likelihood-free), a class of computational statistical methods for Bayesian inference under intractable likelihoods. The site is meant to be a resource both for biologists and statisticians who want to learn more about ABC and related methods. Recent publications are under Publications 2012. A comprehensive list of publications can be found under Literature. If you are unfamiliar with ABC methods see the Introduction. Navigate using the menu to learn more.

ABC in Montreal | ABC in Montreal (2014)

ABC in Montreal

Approximate Bayesian computation (ABC) or likelihood-free (LF) methods have developed mostly beyond the radar of the machine learning community, but are important tools for a large and diverse segment of the scientific community. This is particularly true for systems and population biology, computational neuroscience, computer vision, healthcare sciences, but also many others.

Interaction between the ABC and machine learning community has recently started and contributed to important advances. In general, however, there is still significant room for more intense interaction and collaboration. Our workshop aims at being a place for this to happen.

ICML 2017 Workshop on Implicit Models

Workshop Aims

Probabilistic models are an important tool in machine learning. They form the basis for models that generate realistic data, uncover hidden structure, and make predictions. Traditionally, probabilistic models in machine learning have focused on prescribed models. Prescribed models specify a joint density over observed and hidden variables that can be easily evaluated. The requirement of a tractable density simplifies their learning but limits their flexibility --- several real world phenomena are better described by simulators that do not admit a tractable density. Probabilistic models defined only via the simulations they produce are called implicit models.

Arguably starting with generative adversarial networks, research on implicit models in machine learning has exploded in recent years. This workshop's aim is to foster a discussion around the recent developments and future directions of implicit models.

Implicit models have many applications. They are used in ecology where models simulate animal populations over time; they are used in phylogeny, where simulations produce hypothetical ancestry trees; they are used in physics to generate particle simulations for high energy processes. Recently, implicit models have been used to improve the state-of-the-art in image and content generation. Part of the workshop's focus is to discuss the commonalities among applications of implicit models.

Of particular interest at this workshop is to unite fields that work on implicit models. For example:

- **Generative adversarial networks** (a NIPS 2016 workshop) are implicit models with an adversarial training scheme.
- Recent advances in **variational inference** (a NIPS 2015 and 2016 workshop) have leveraged implicit models for more accurate approximations.
- **Approximate Bayesian computation** (a NIPS 2015 workshop) focuses on posterior inference for models with implicit likelihoods.
- Learning implicit models is deeply connected to **two sample testing, density ratio and density difference** estimation.

We hope to bring together these different views on implicit models, identifying their core challenges and combining their innovations.

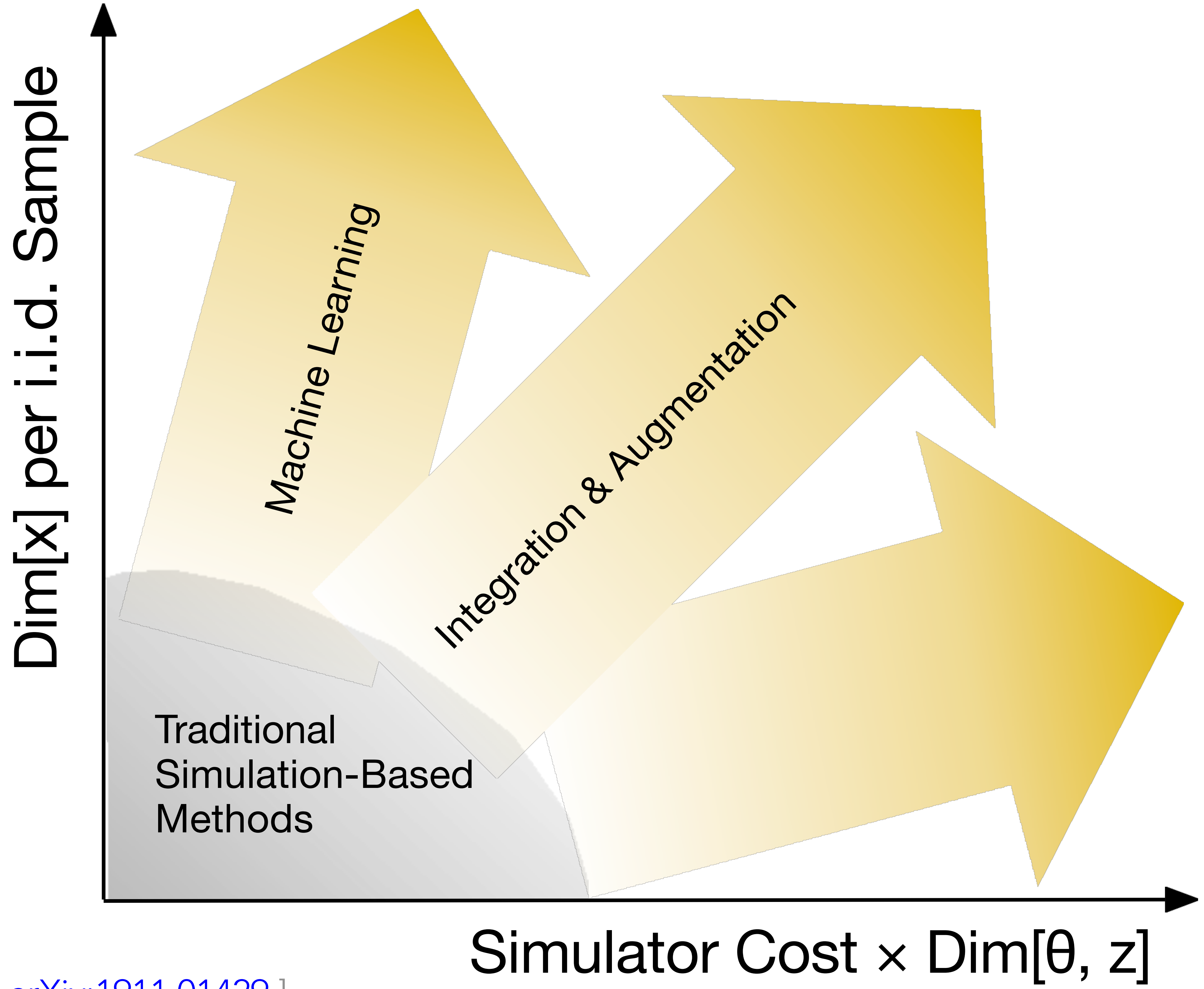
Published in Proceedings of the National Academy of Sciences

The frontier of simulation-based inference

Kyle Cranmer[a,b,1], Johann Brehmer[a,b], and Gilles Louppe[c]

[a]Center for Cosmology and Particle Physics, New York University, USA; [b]Center for Data Science, New York University, USA; [c]Montefiore Institute, University of Liège, Belgium

April 3, 2020



Gilles Louppe

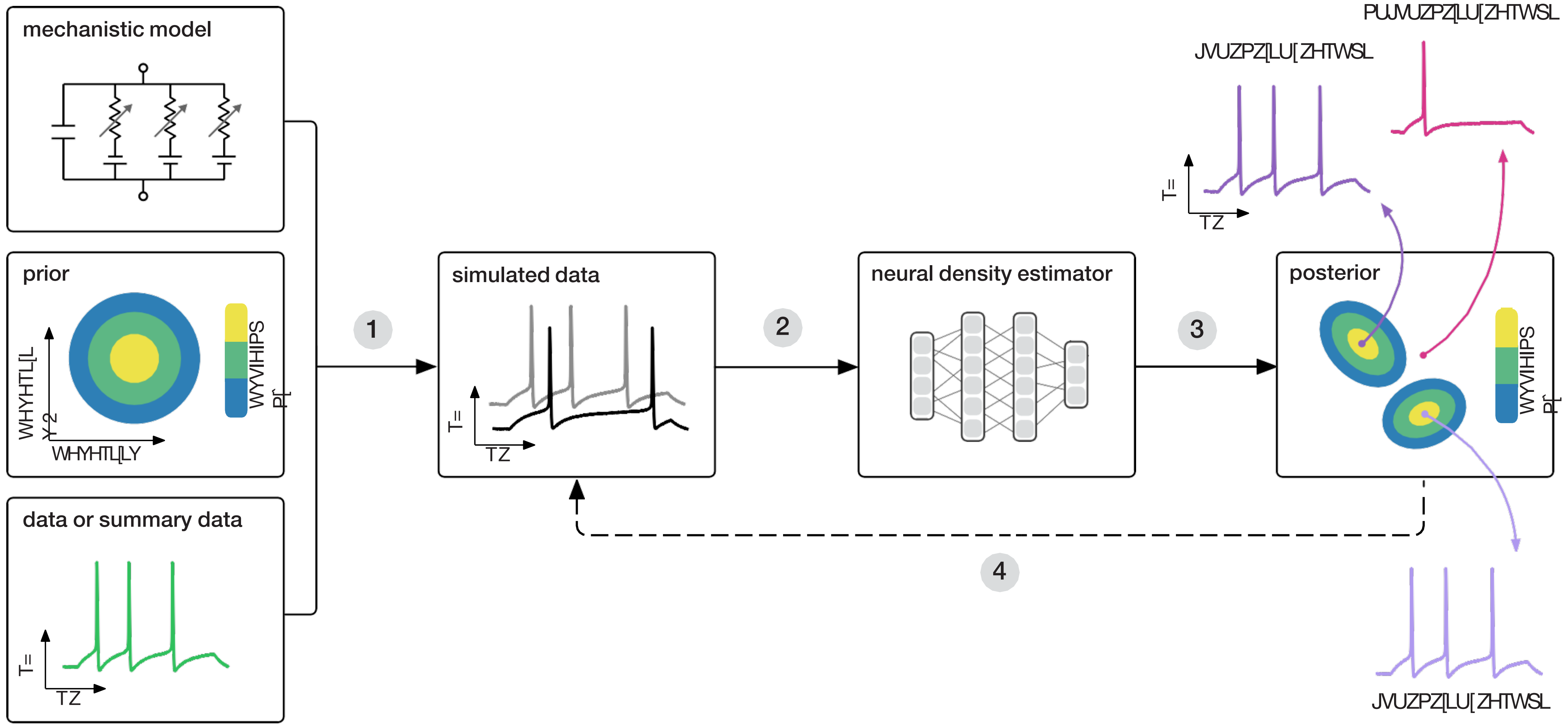
Johann Brehmer

[Cranmer, Brehmer, Louppe PNAS (2020), arXiv:1911.01429]

From Jakob Macke's group that developed the sbi python package

b#B, A iQQHFB 7Q bBKmHaBQM@#ab2/ BM722M+2

AHpa`Q h2D2`Q@*aMi2`Q2- R- CaM "Q2Hib2- R- JB+?a2H .2BbiH2`2- R- CaM@Jaii?Bb Gm2+FKaMM2- R- *QMQ` .m`FaM2- k- S2/`Q CX:QMïaHp2bR- j - .apB/ aX:`22M#2`;R- 9- aM/ CaFQ# >X Ja+F2R- 8- e



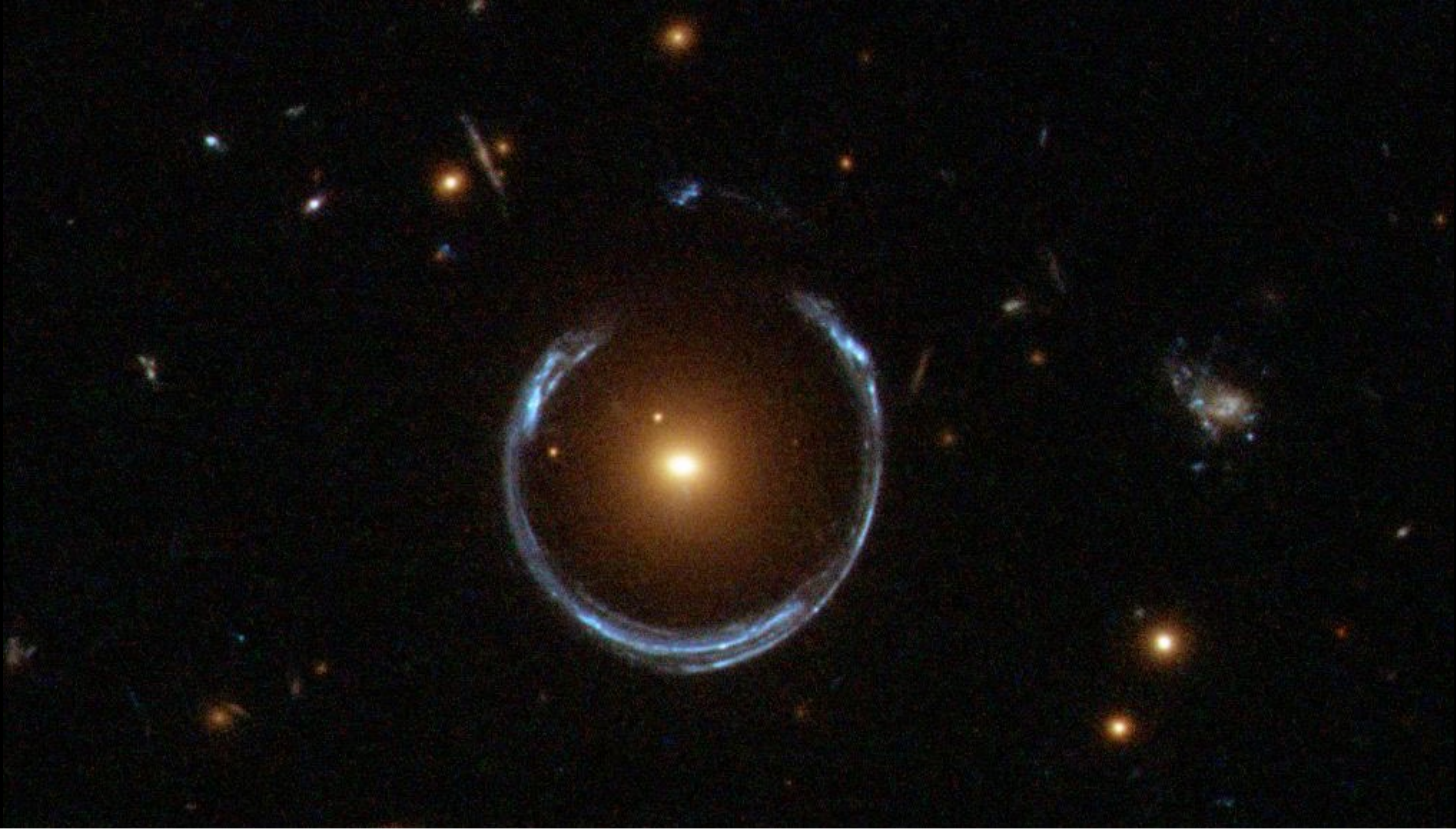
eLife

Training deep neural density estimators to identify mechanistic models of neural dynamics

Pedro J Gonc¸alves[1,2†*], Jan-Matthis Lueckmann[1,2†*], Michael Deistler[1,3†*], Marcel Nonnenmacher[1,2,4], Kaan Öcal[2,5], Giacomo Bassetto[1,2], Chaitanya Chintaluri[6,7], William F Podlaski[6], Sara A Haddad[8], Tim P Vogels[6,7], David S Greenberg[1,4], Jakob H Macke[1,2,3,9*]

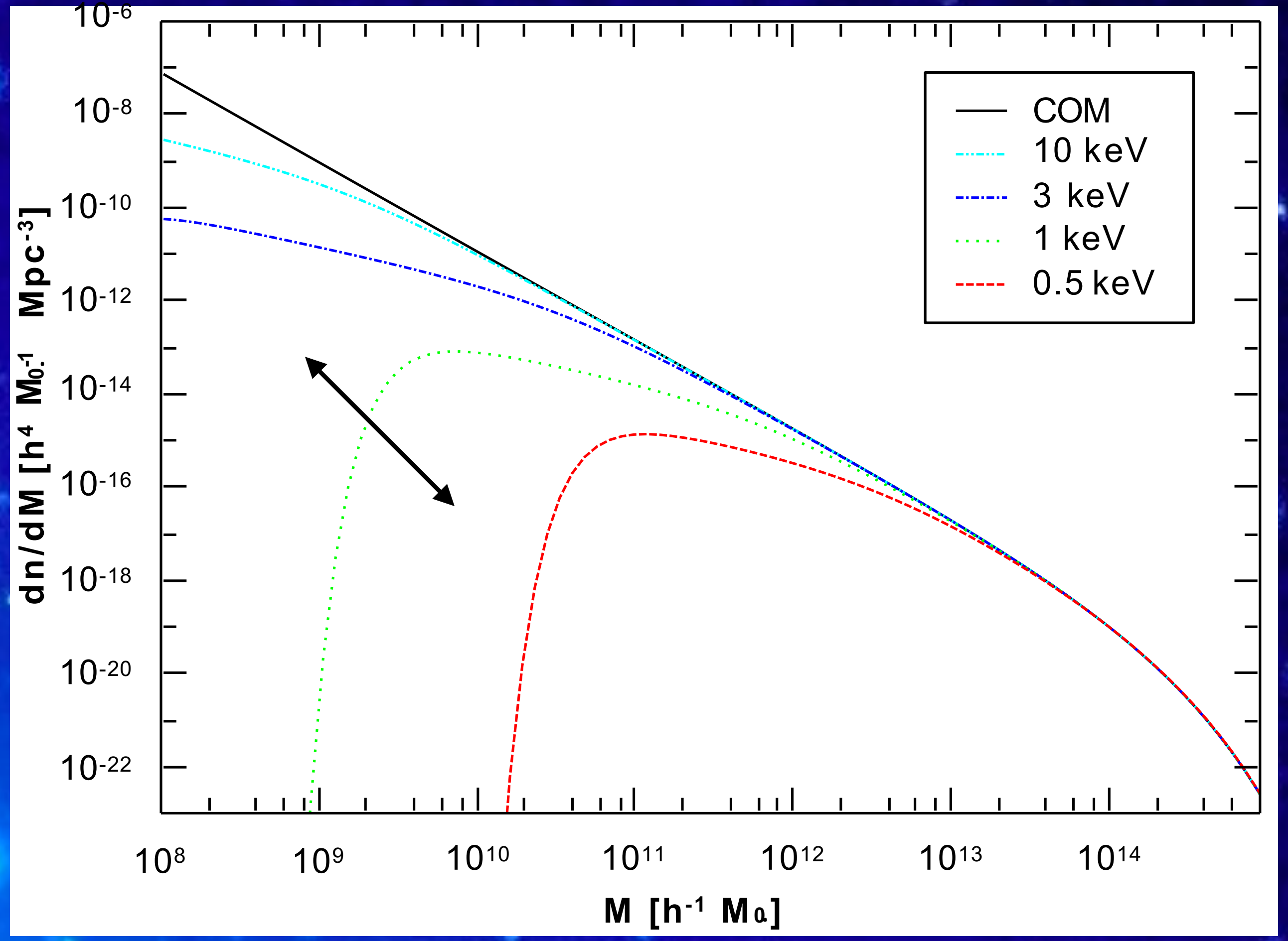
[1]Computational Neuroengineering, Department of Electrical and Computer Engineering, Technical University of Munich, Munich, Germany; [2]Max Planck Research Group Neural Systems Analysis, Center of Advanced European Studies and Research (caesar), Bonn, Germany; [3]Machine Learning in Science, Excellence Cluster Machine Learning, Tübingen University, Tübingen, Germany; [4]Model-Driven Machine Learning, Institute of Coastal Research, Helmholtz Centre Geesthacht, Geesthacht, Germany; [5]Mathematical Institute, University of Bonn, Bonn, Germany; [6]Centre for Neural Circuits and Behaviour, University of Oxford, Oxford, United Kingdom; [7]Institute of Science and Technology Austria, Klosterneuburg, Austria; [8]Max Planck Institute for Brain Research, Frankfurt, Germany; [9]Max Planck Institute for Intelligent Systems, Tübingen, Germany

Abstract Mechanistic modeling in neuroscience aims to explain observed phenomena in terms of underlying causes. However, determining which model parameters agree with complex and stochastic neural data presents a significant challenge. We address this challenge with a machine learning tool which uses deep neural density estimators—trained using model simulations—to carry out Bayesian inference and retrieve the full space of parameters compatible with raw data or selected data features. Our method is scalable in parameters and data features and can rapidly analyze new data after initial training. We demonstrate the power and flexibility of our approach on receptive fields, ion channels, and Hodgkin–Huxley models. We also characterize the space of circuit configurations giving rise to rhythmic activity in the crustacean stomatogastric ganglion, and use these results to derive hypotheses for underlying compensation mechanisms. Our approach will help close the gap between data-driven and theory-driven models of neural dynamics.



Dark Matter Substructure

Abundance of DM subhalos vs mass:

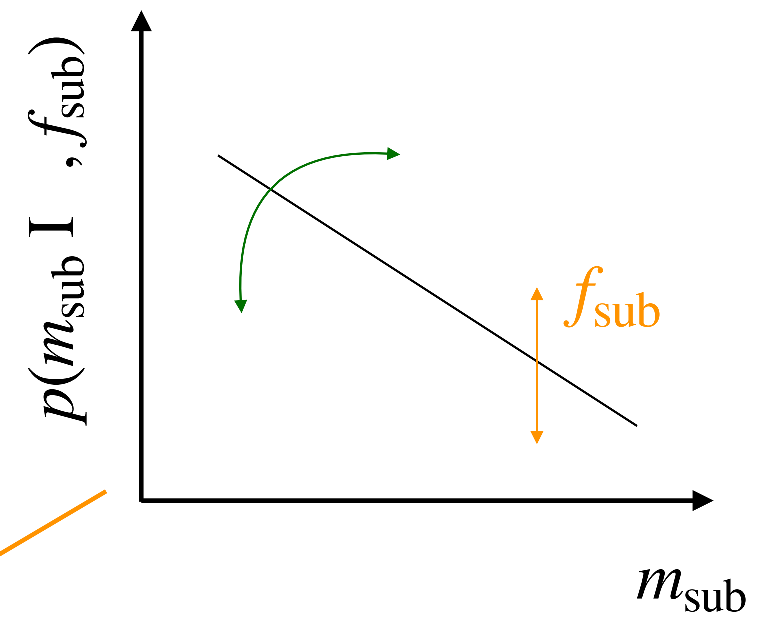


[R. *Dunstan et al* 1109.6291]

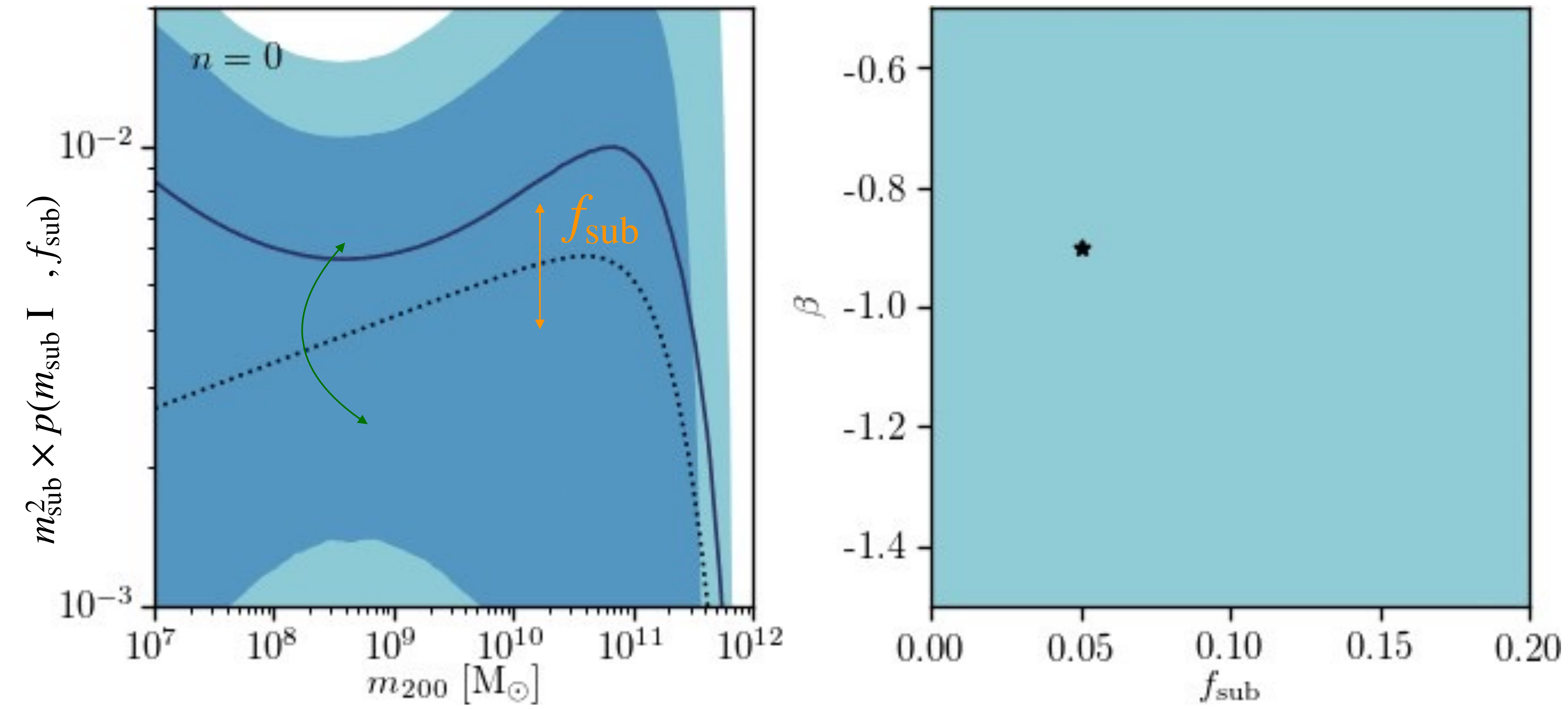
[T. Brown, J. Tumlinson]

Posterior from amortized likelihood ratio

Watch how the posterior for two population parameters concentrate around true value used to generate mock data.



(plotted slightly differently)

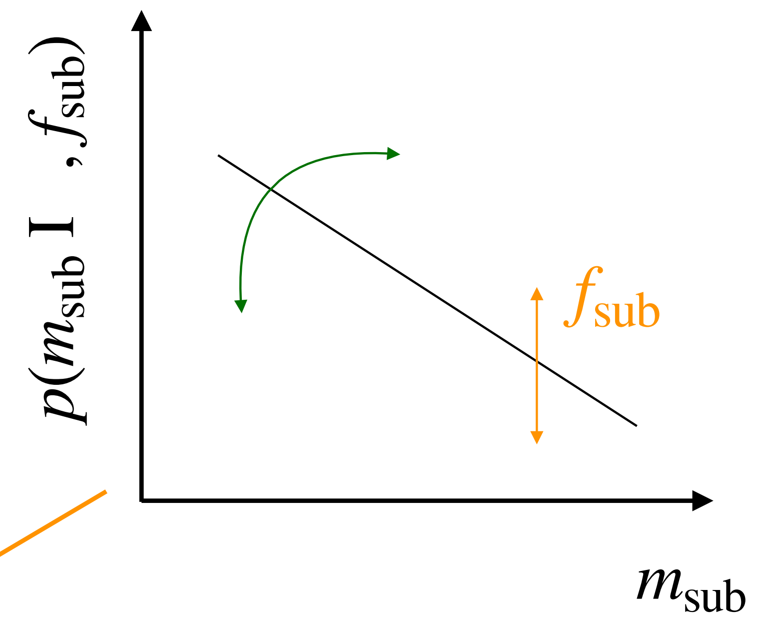




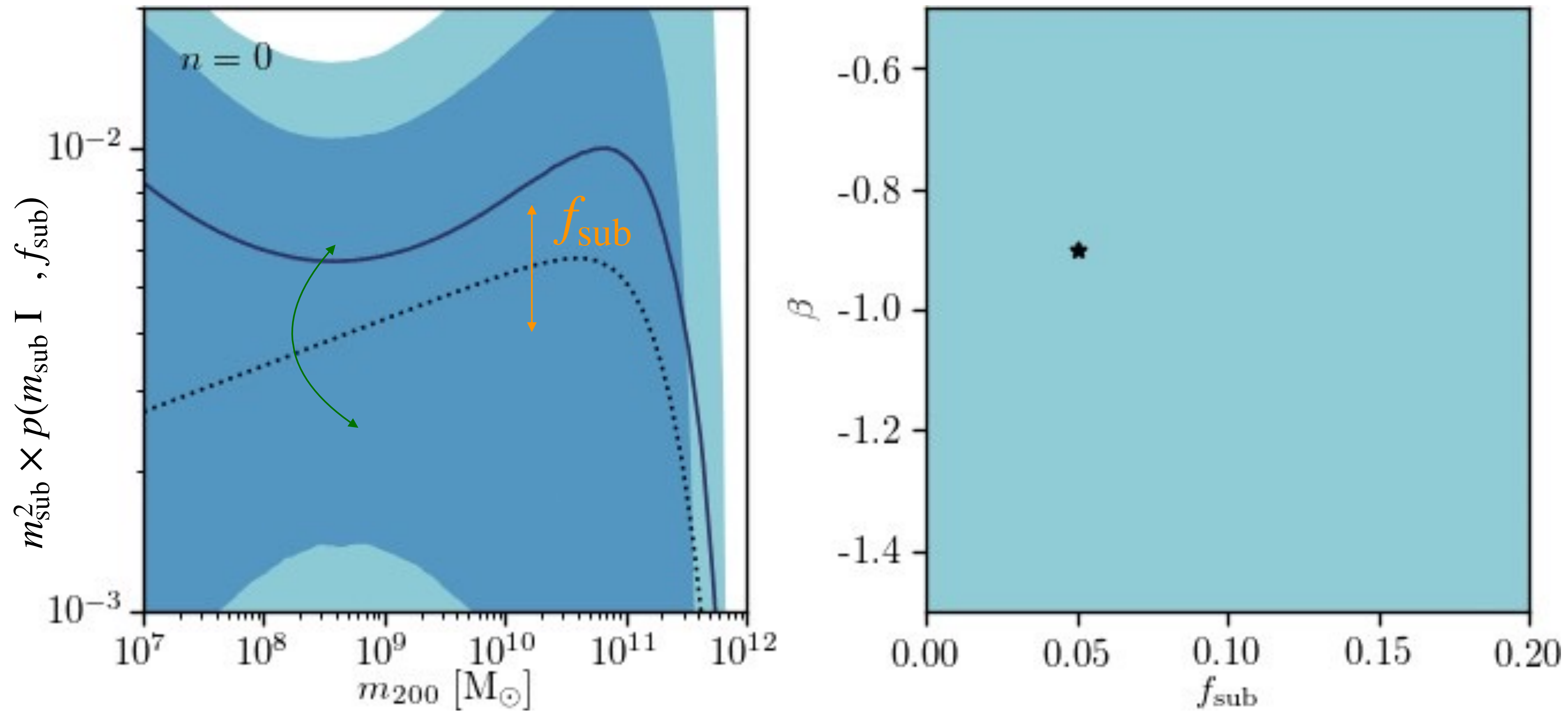

Sid Mishra-Sharma Johann Brehmer Gilles Louppe Joeri Hermans, K. Cranmer. published in ApJ, 886 (2019) arXiv:1909.02005

Posterior from amortized likelihood ratio

Watch how the posterior for two population parameters concentrate around true value used to generate mock data.

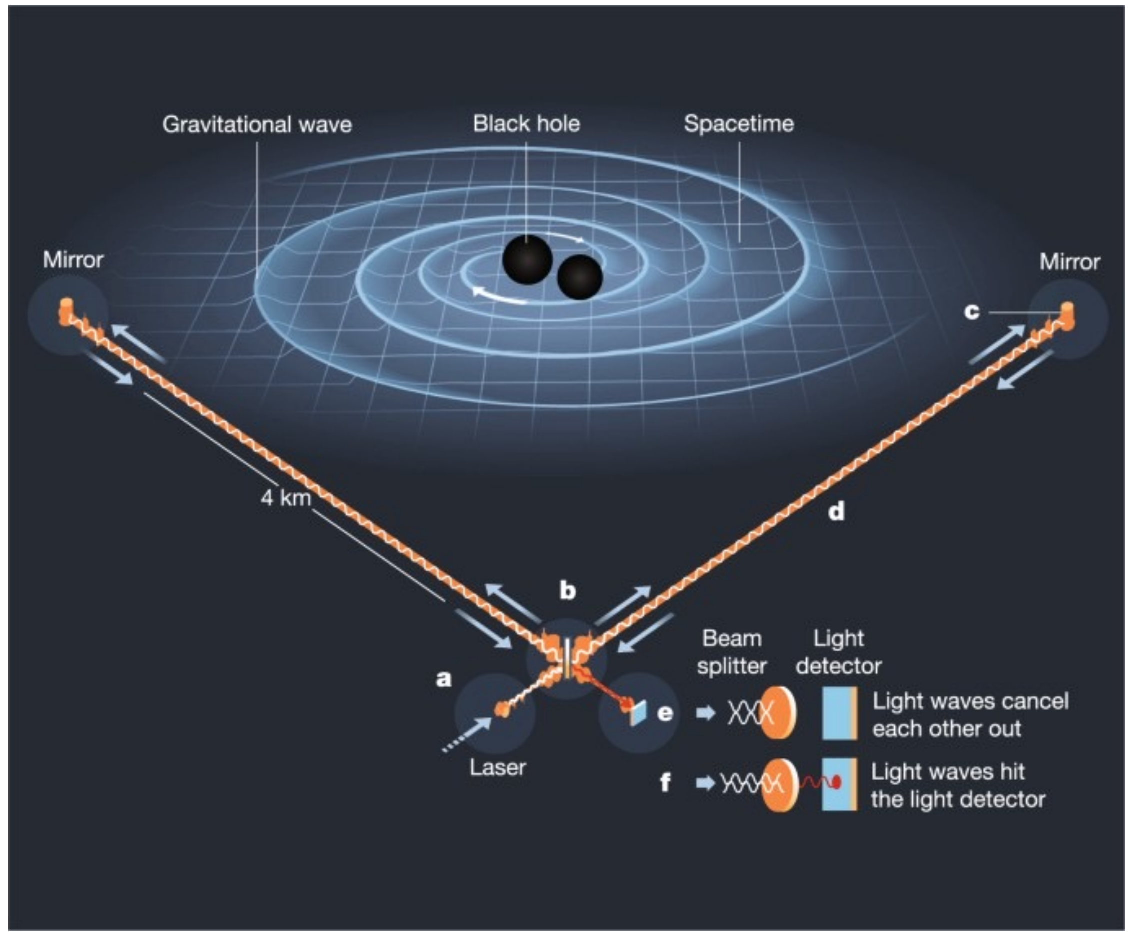


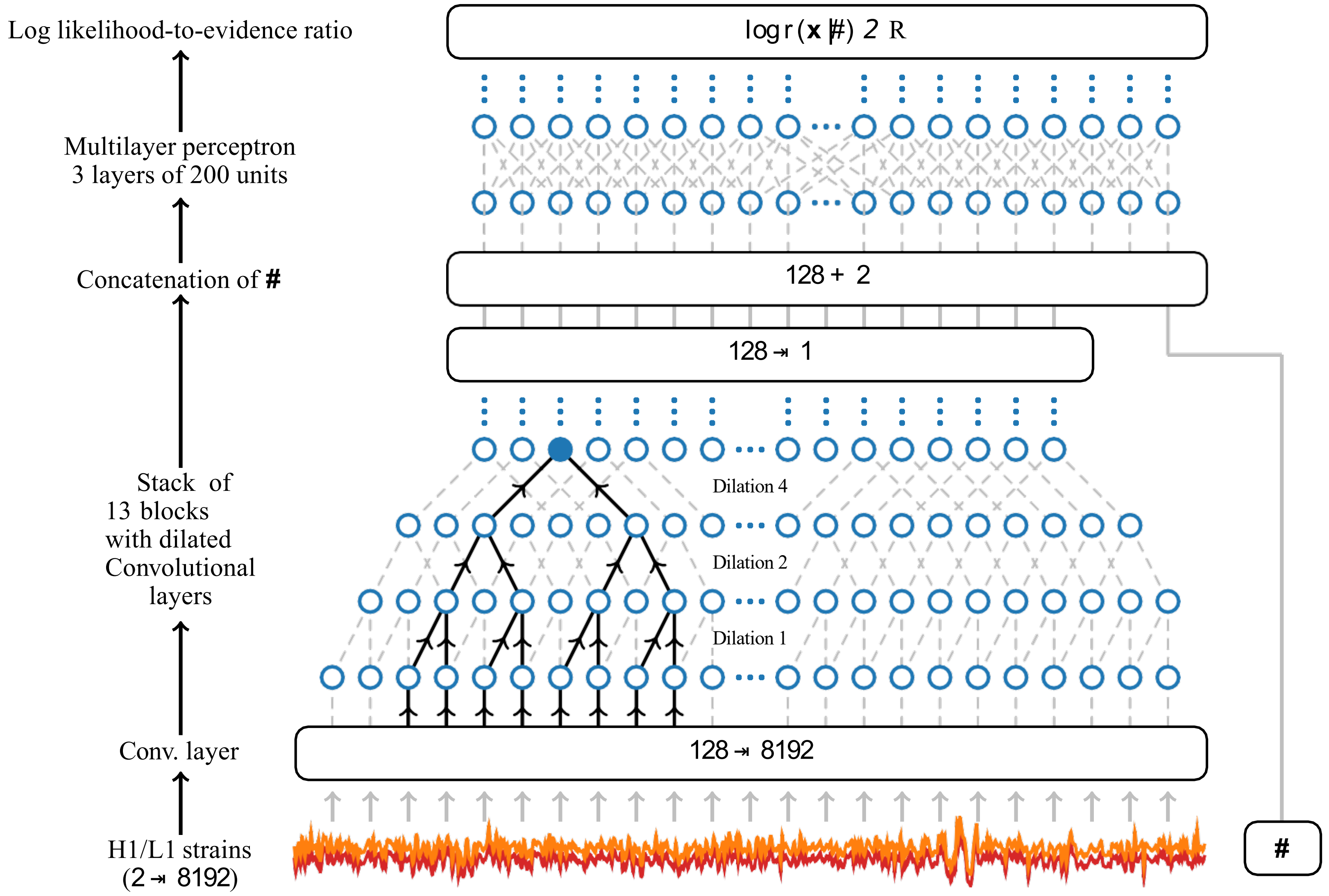
(plotted slightly differently)



Sid Mishra-Sharma Johann Brehmer Gilles Louppe Joeri Hermans, K. Cranmer. published in ApJ, 886 (2019) arXiv:1909.02005

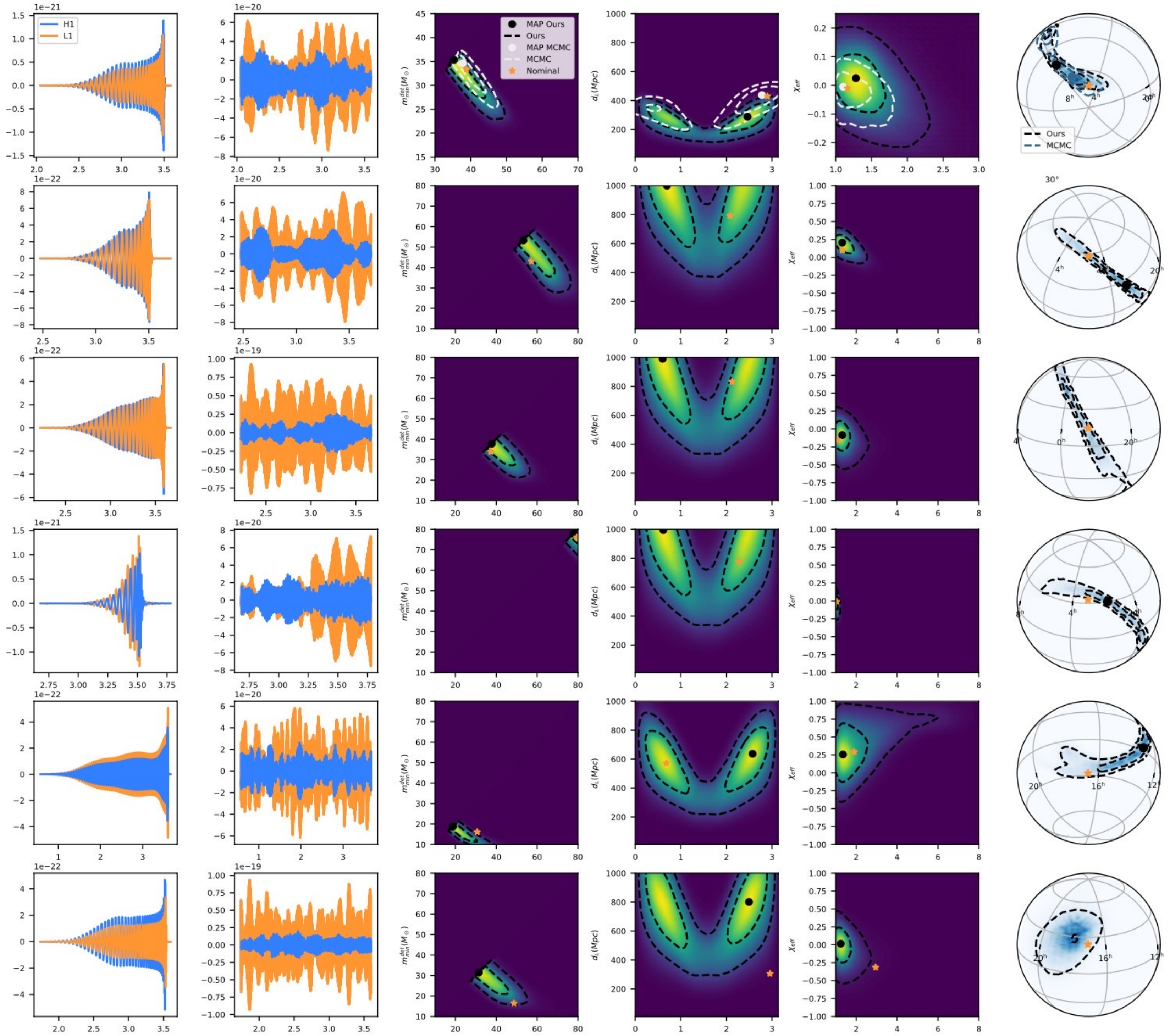
Gravitational Wave Astronomy





Lightning-Fast Gravitational Wave Parameter Inference through Neural Amortization

Delaunoy, Wehenkel, Hinderer, Nissanke, Weniger, Williamson, Louppe
[arXiv:2010.12931]



Speed is a concern if we want to use this to quickly point telescopes to look for optical counterpart

Two other recent examples

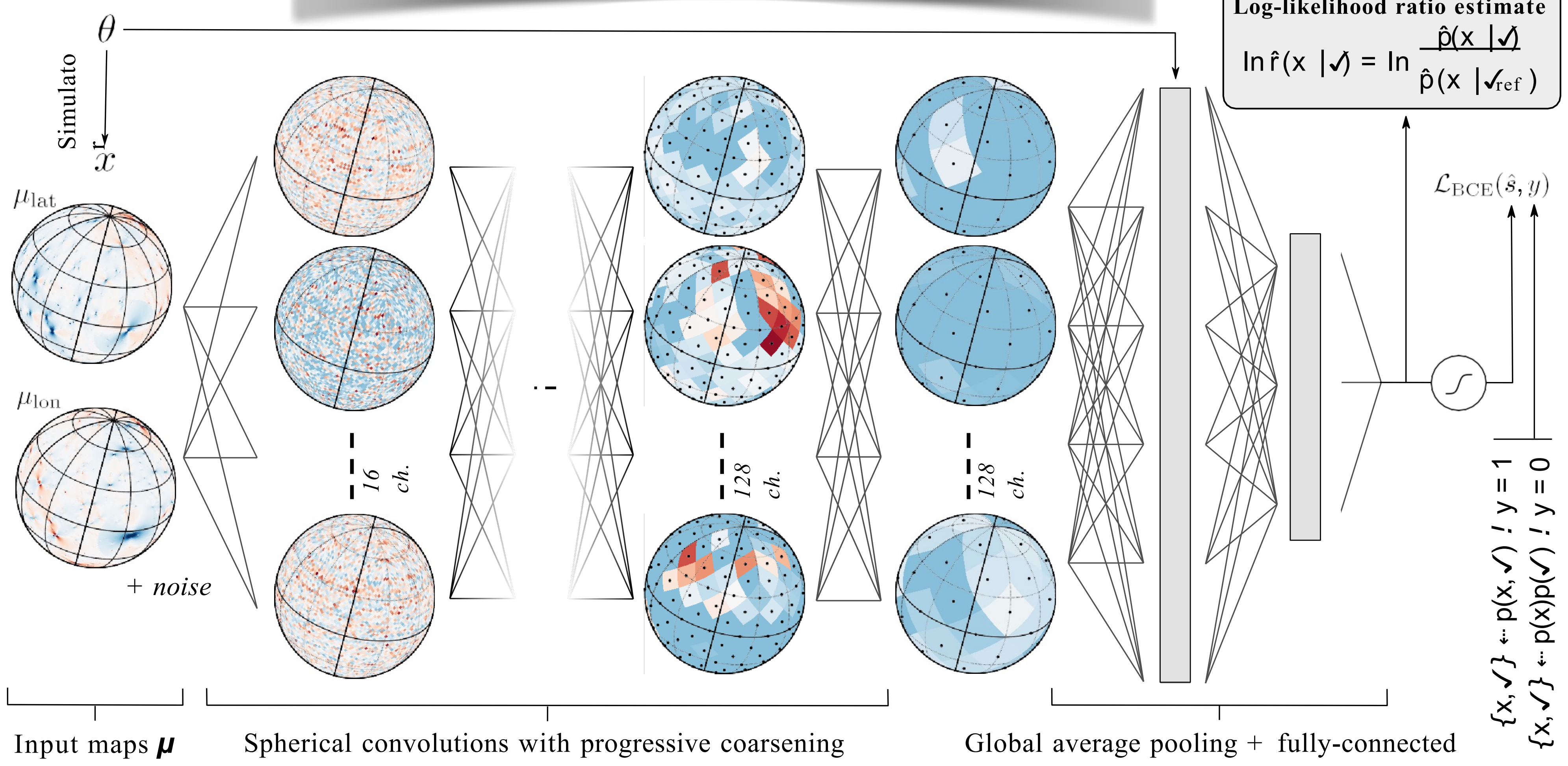
Sid Mishra-Sharma

Both works use spherical CNN to process maps of the sky into relevant summaries

Right: data distribution is a mixture and coefficients are parameters of interest.

Inferring dark matter substructure with astrometric lensing beyond the power spectrum

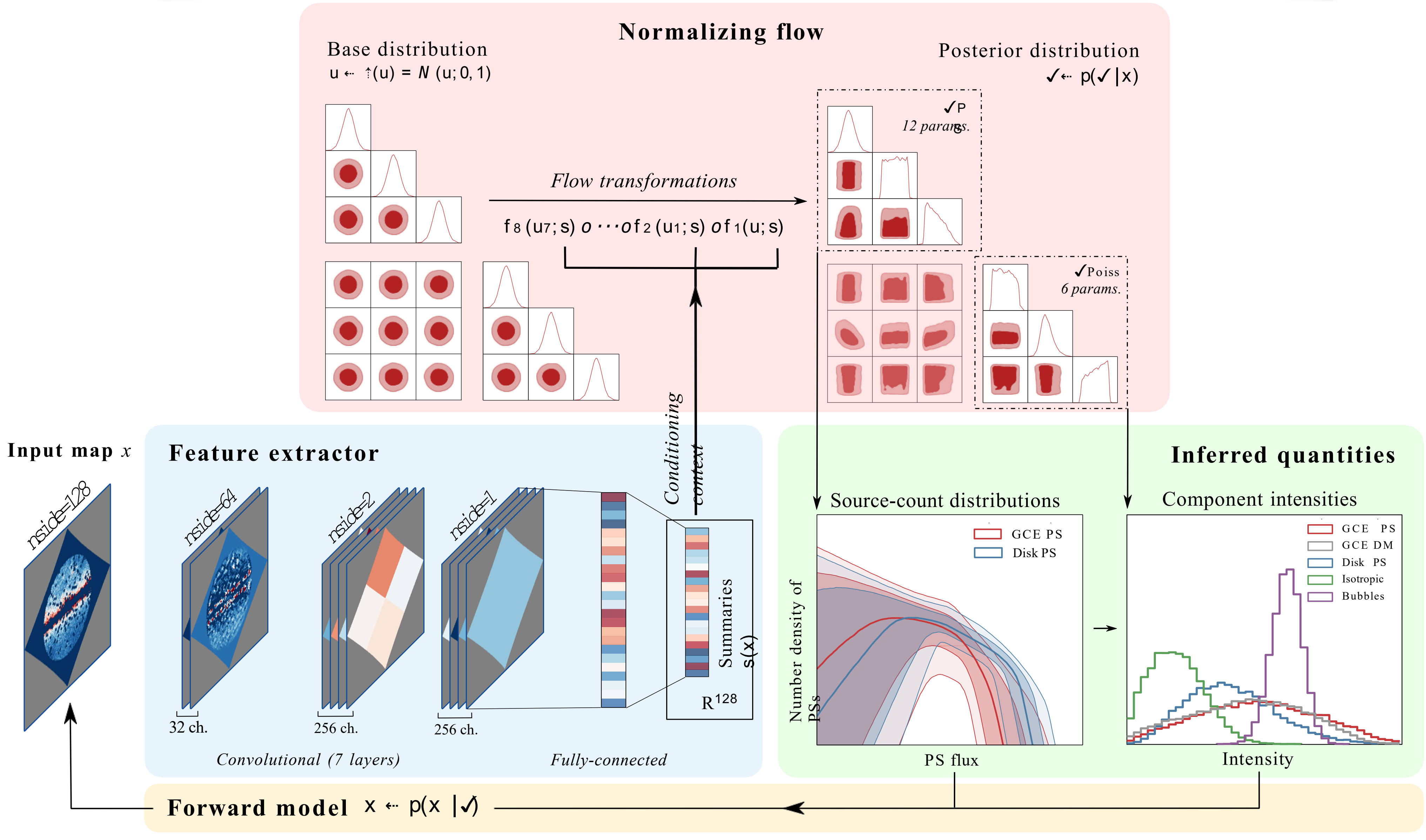
Siddharth Mishra-Sharma
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Massachusetts Institute of Technology
Harvard University
New York University
smsharma@mit.edu



[arXiv:2110.01620]

A neural simulation-based inference approach for characterizing the Galactic Center γ-ray excess

Siddharth Mishra-Sharma[1,2,3,4,5,*] and Kyle Cranmer[5,6,†]

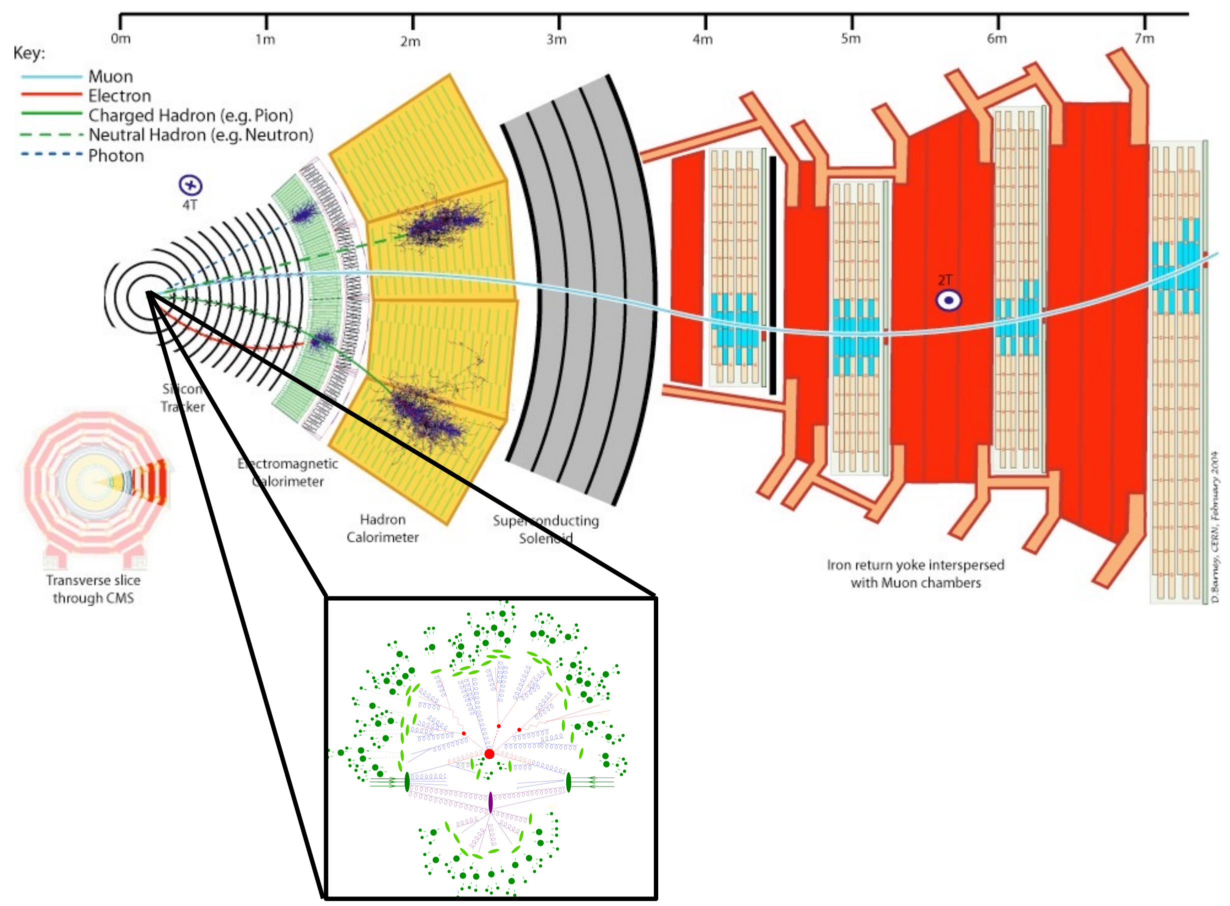


[arXiv:2110.06931]

Two approaches simulation-based inference

Use simulator

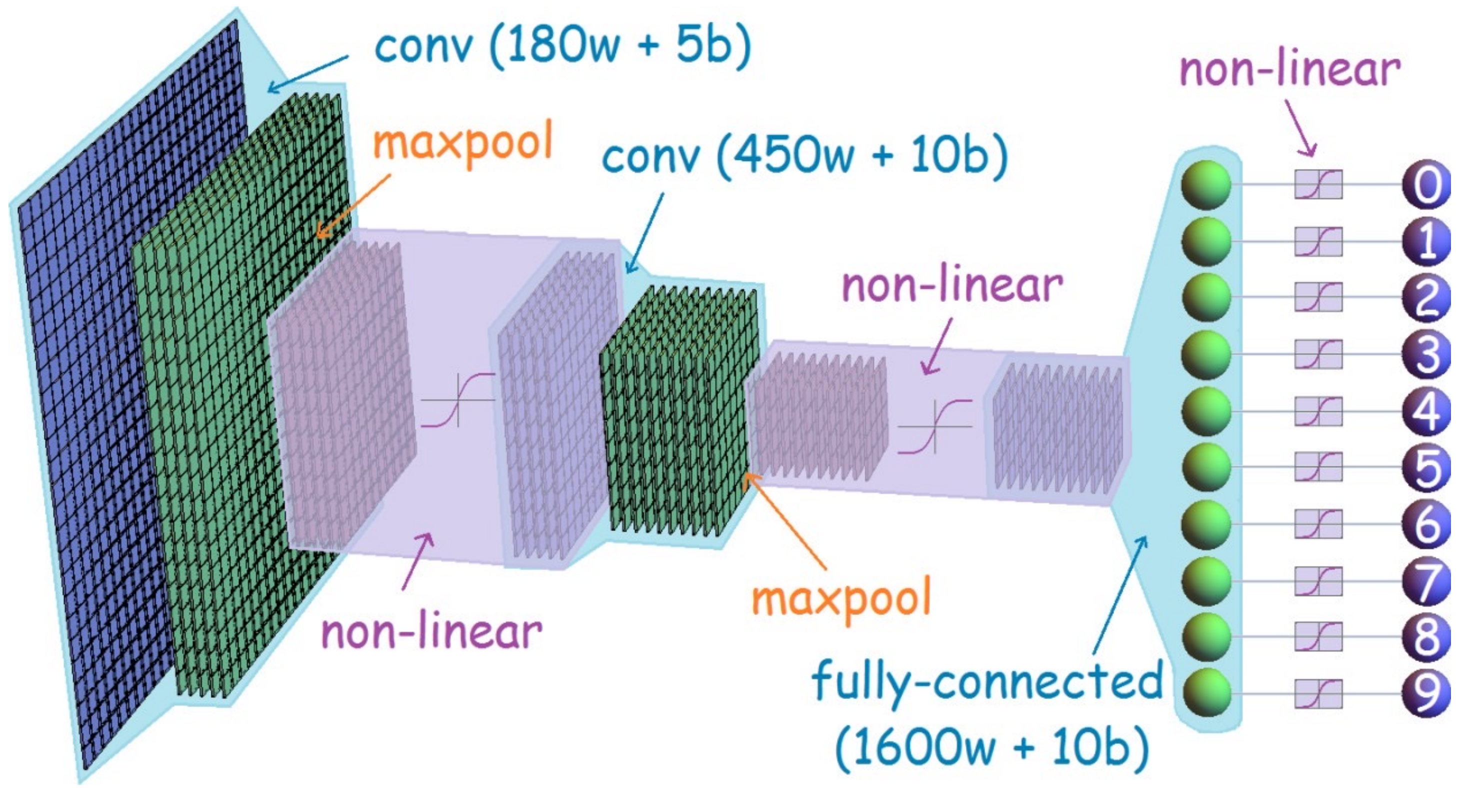
(much more efficiently)



- Approximate Bayesian Computation (ABC)
- Probabilistic Programming
- Adversarial Variational Optimization

Learn simulator

(with deep learning)



- Likelihood ratio trick (with classifiers)
- Conditional density estimate (with normalizing flows)
- Learned summary statistics

[image credit: A.P. Goucher]

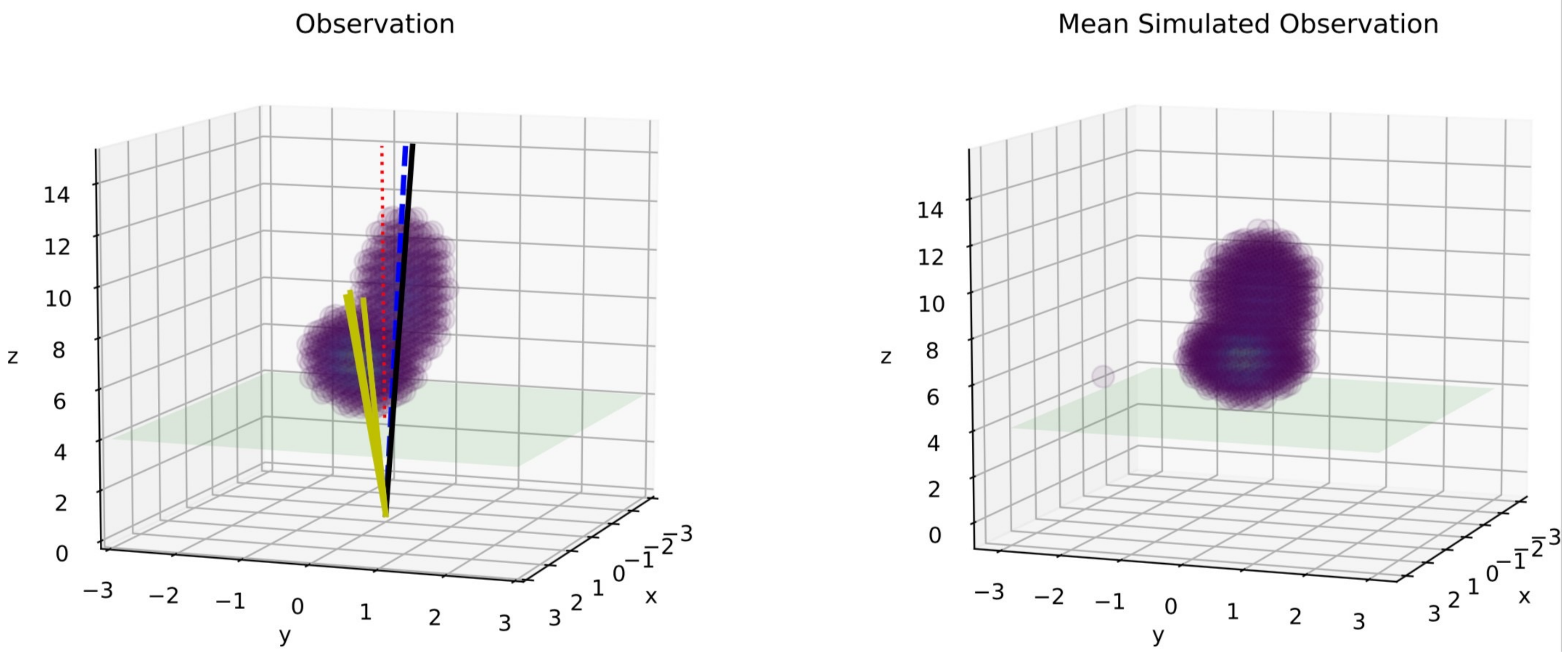
G. Baydin, et al SC19 arXiv:1907.03382
G. Baydin, et al. NeurIPS 2019 arXiv:1807.07706

Previously had to use a special purpose probabilistic programming language. With **ppx** protocol, we decouple inference engine & control existing simulator.

- **Augment** real-world physics simulator
  (C++, 1M lines of code)

- 3DCNN-LSTM architecture for $q_\phi(z|x)$
  (Stack traces with Dim[ ] ranging from 100 — 2,000)

- Inference is embarrassingly parallelizable unlike MCMC. 230x speedup



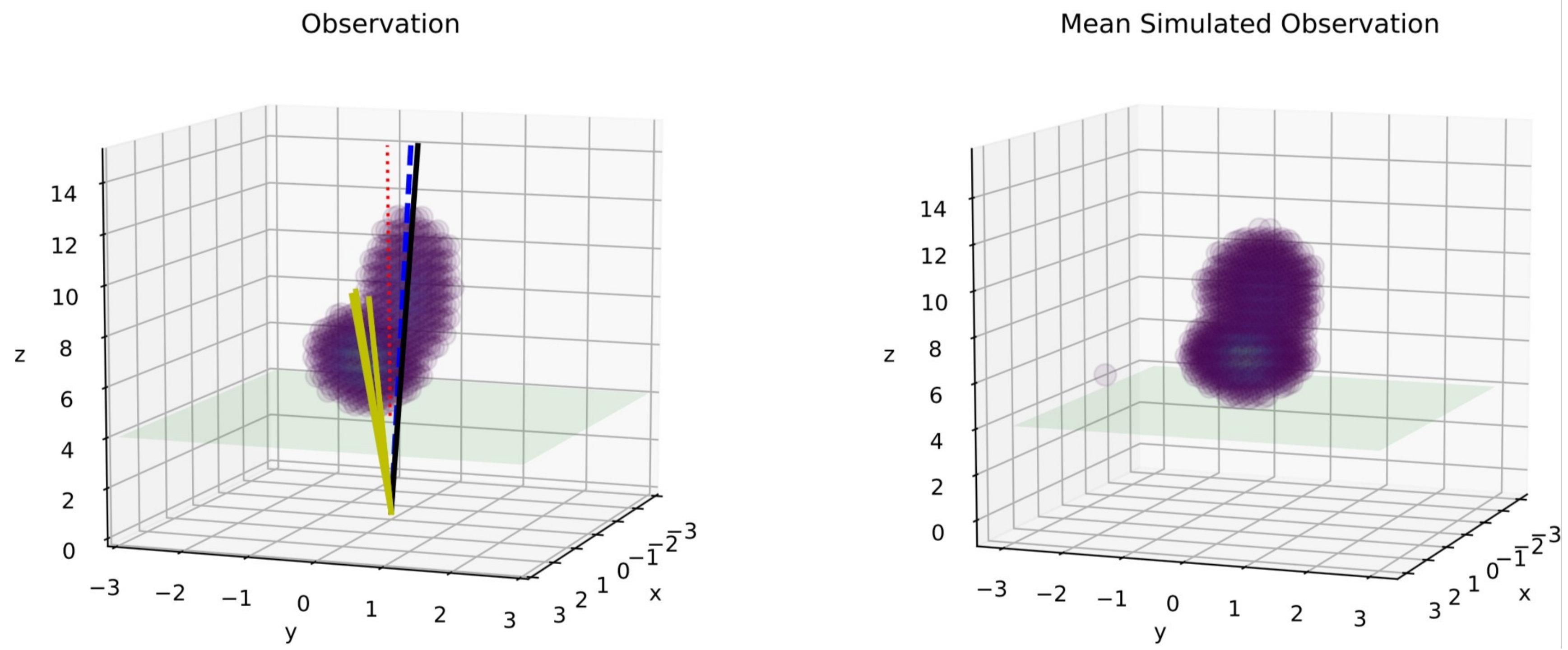
G. Baydin, et al SC19 arXiv:1907.03382
G. Baydin, et al. NeurIPS 2019 arXiv:1807.07706

Previously had to use a special purpose probabilistic programming language. With **ppx** protocol, we decouple inference engine & control existing simulator.

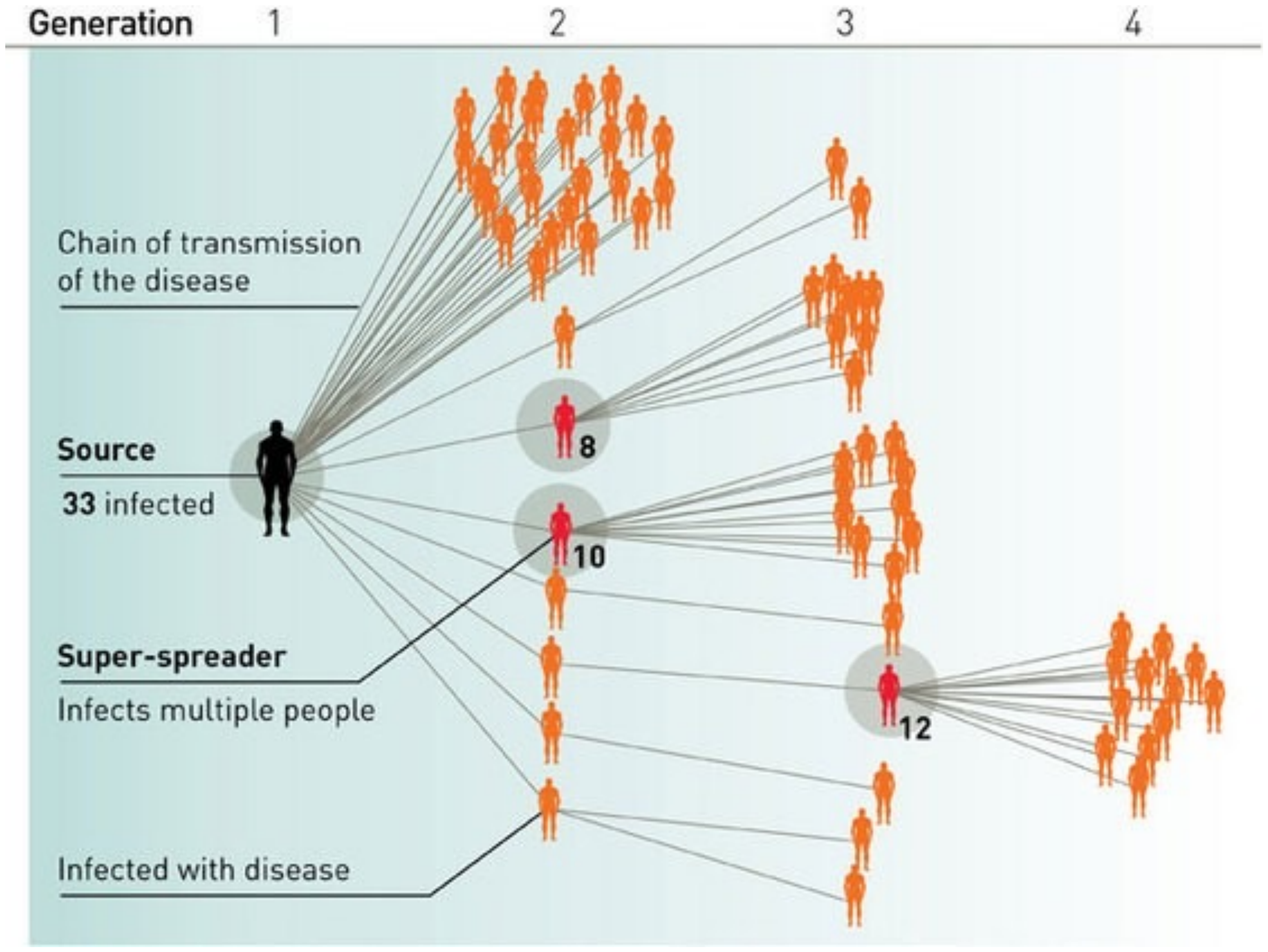
- **Augment** real-world physics simulator
  (C++, 1M lines of code)

- 3DCNN-LSTM architecture for $q_<(z \ |x)$
  (Stack traces with Dim[ ] ranging from 100 — 2,000)

- Inference is embarrassingly parallelizable unlike MCMC. 230x speedup



Epidemiology & population Genetics



Simulation-Based Inference for Global Health Decisions

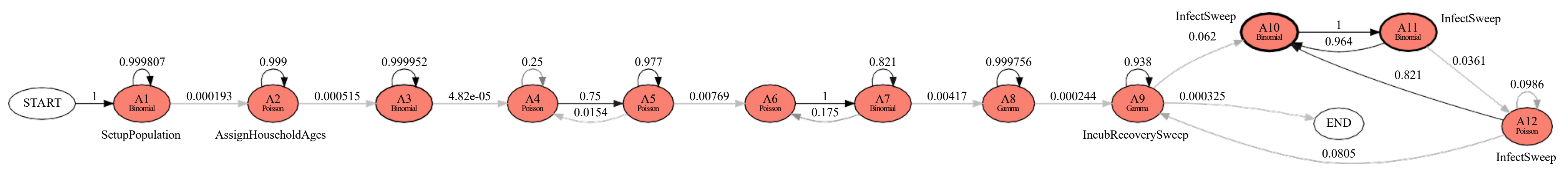


Figure 1: Latent probabilistic structure uncovered using PyProb from the Imperial College CovidSim simulator run on Malta, demonstrating the first step in working with this simulator as a probabilistic program. Uniform distributions are omitted for simplicity.

Simulation-Based Inference for Global Health Decisions

Christian Schroeder de Witt [1] Bradley Gram-Hansen [1] Nantas Nardelli [1]
Andrew Gambardella [1] Rob Zinkov [1] Puneet Dokania [1] N. Siddharth [1]
Ana Belen Espinosa-Gonzalez [2] Ara Darzi [2] Philip Torr [1] Atılım Güneş Baydin [1]

https://arxiv.org/abs/2005.07062

PLANNING AS INFERENCE IN EPIDEMIOLOGICAL DYNAMICS MODELS

A PREPRINT

Frank Wood[1,3,4], Andrew Warrington[2], Saeid Naderiparizi[1], Christian Weilbach[1], Vaden Masrani[1], William Harvey[1], Adam Ścibior[1], Boyan Beronov[1], and Ali Nasseri[1]

[1]Department of Computer Science, University of British Columbia
[2]Department of Engineering Science, University of Oxford
[3]MILA
[4]CIFAR AI Chair
{fwood,awarring,saeidnp,weilbach,vadmas,wsgh,ascibior,beronov}@cs.ubc.ca, ali.nasseri@ubc.ca

https://arxiv.org/abs/2003.13221

Hijacking Malaria Simulators with Probabilistic Programming

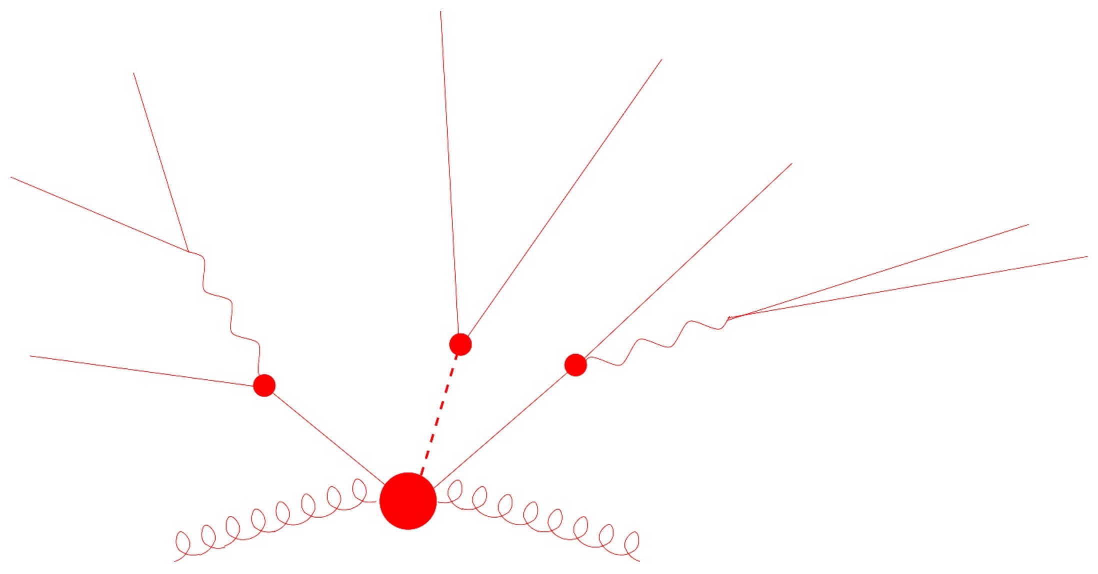
Bradley J. Gram-Hansen *[1] Christian Schröder de Witt *[1]
Tom Rainforth [2] Philip H.S. Torr [1] Yee Whye Teh [2] Atılım Güneş Baydin [1]

https://arxiv.org/abs/1905.12432

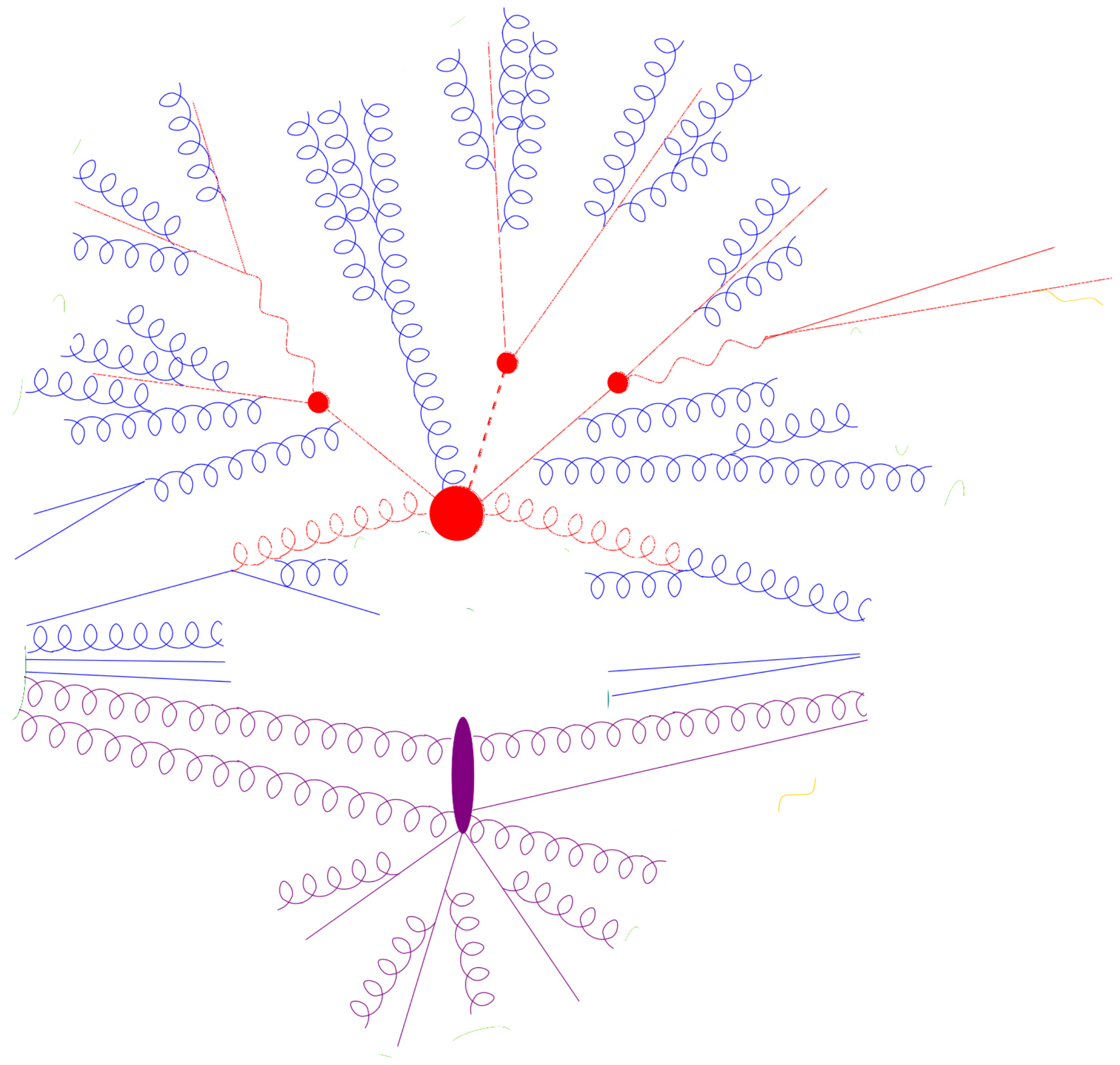
Examples of Physics-inspired ML

Causal, Generative Model For Jets

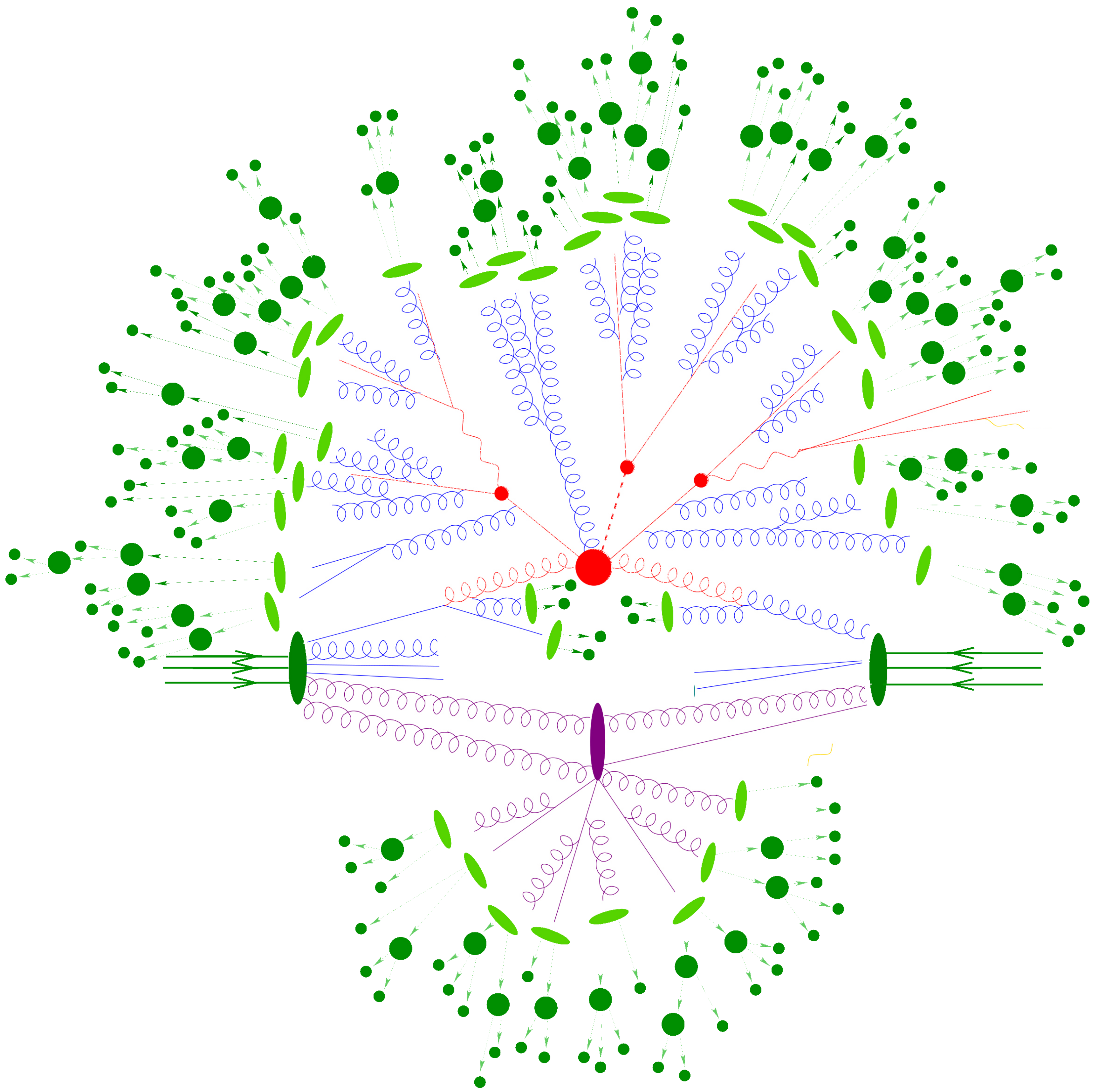
Causal, Generative Model For Jets



Causal, Generative Model For Jets



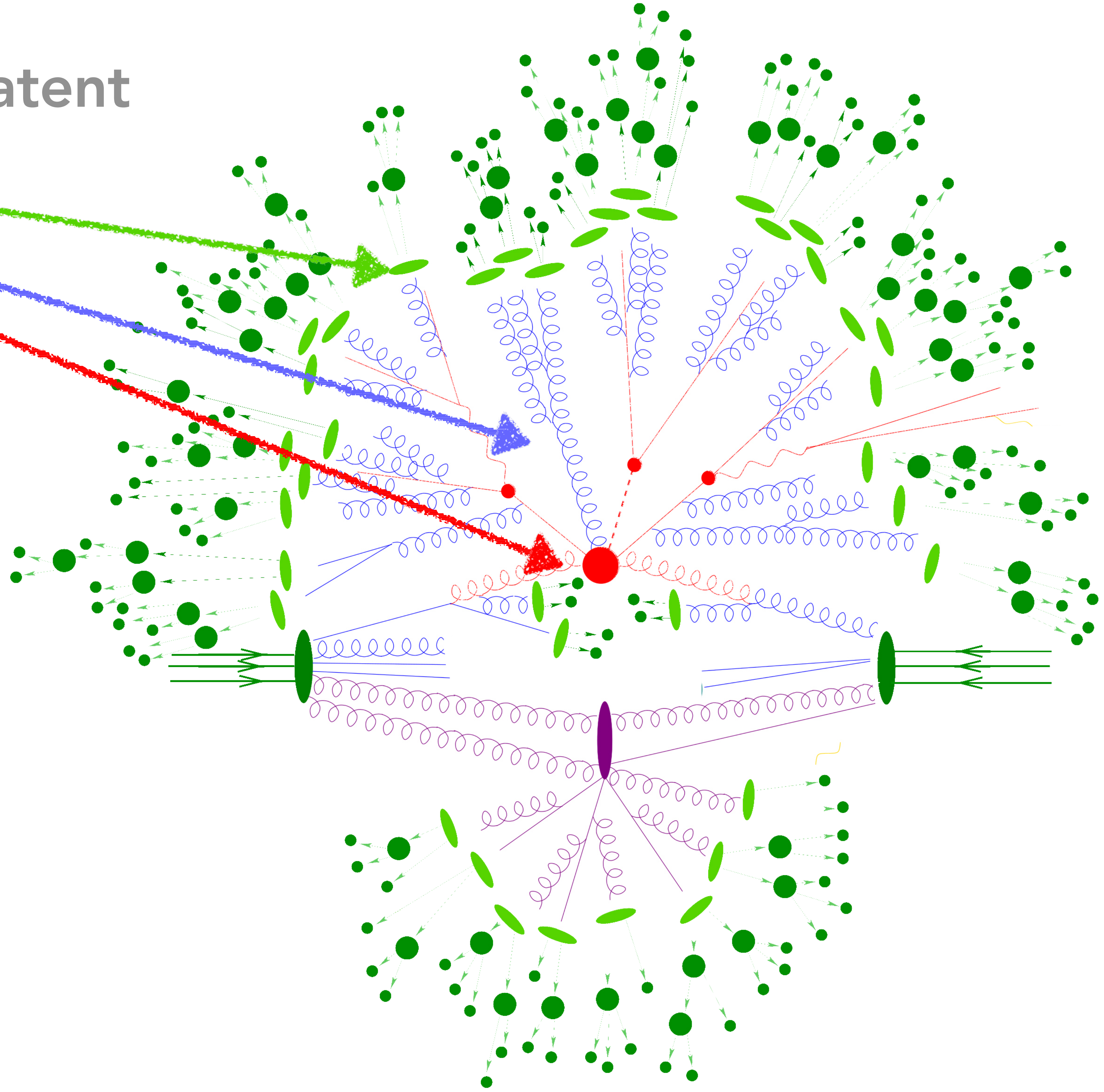
Causal, Generative Model For Jets



Causal, Generative Model For Jets

z

Evolution of the tree is **latent**



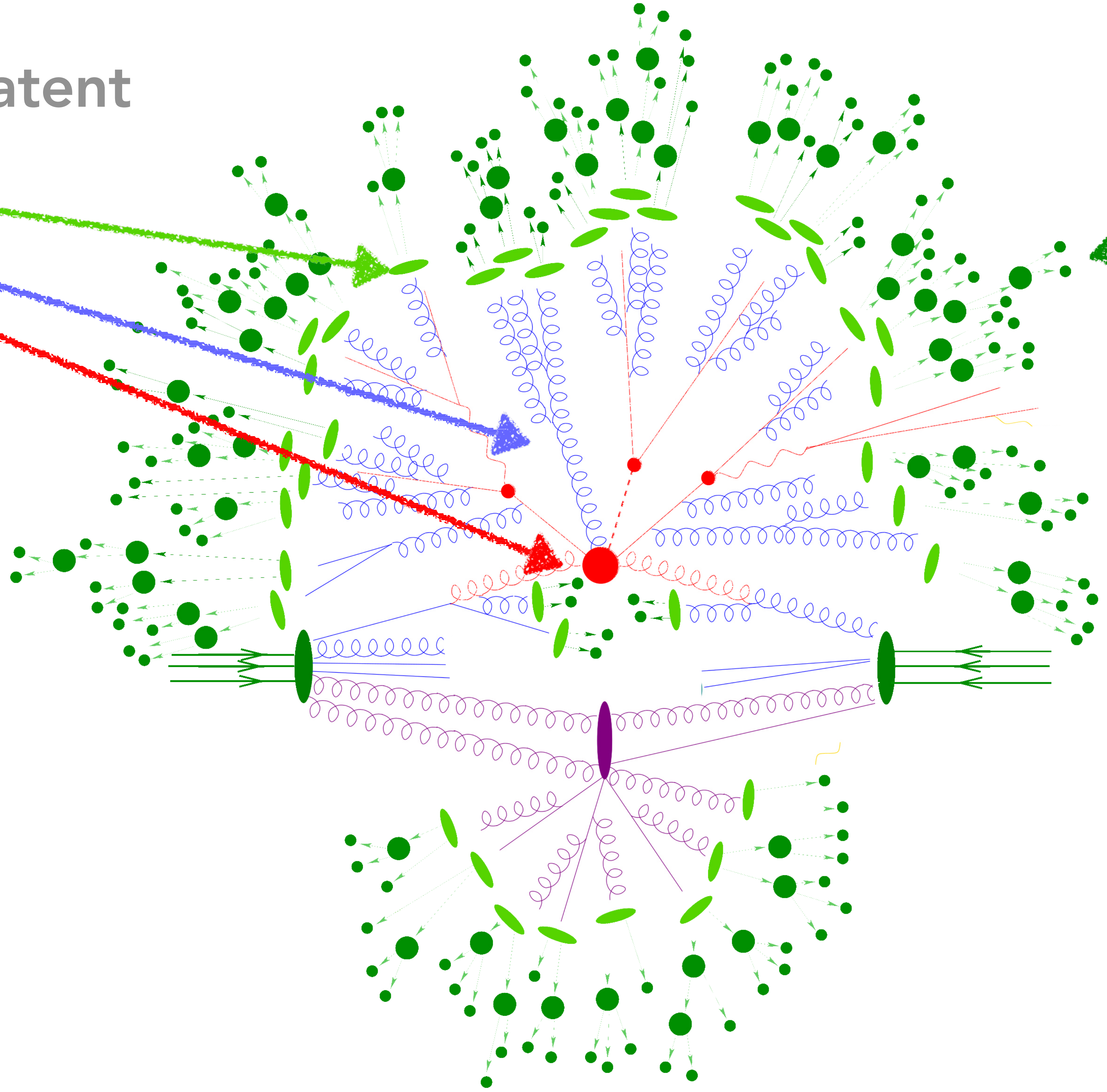
Causal, Generative Model For Jets

z

Evolution of the tree is **latent**

x

We only **observe** the **leaves**



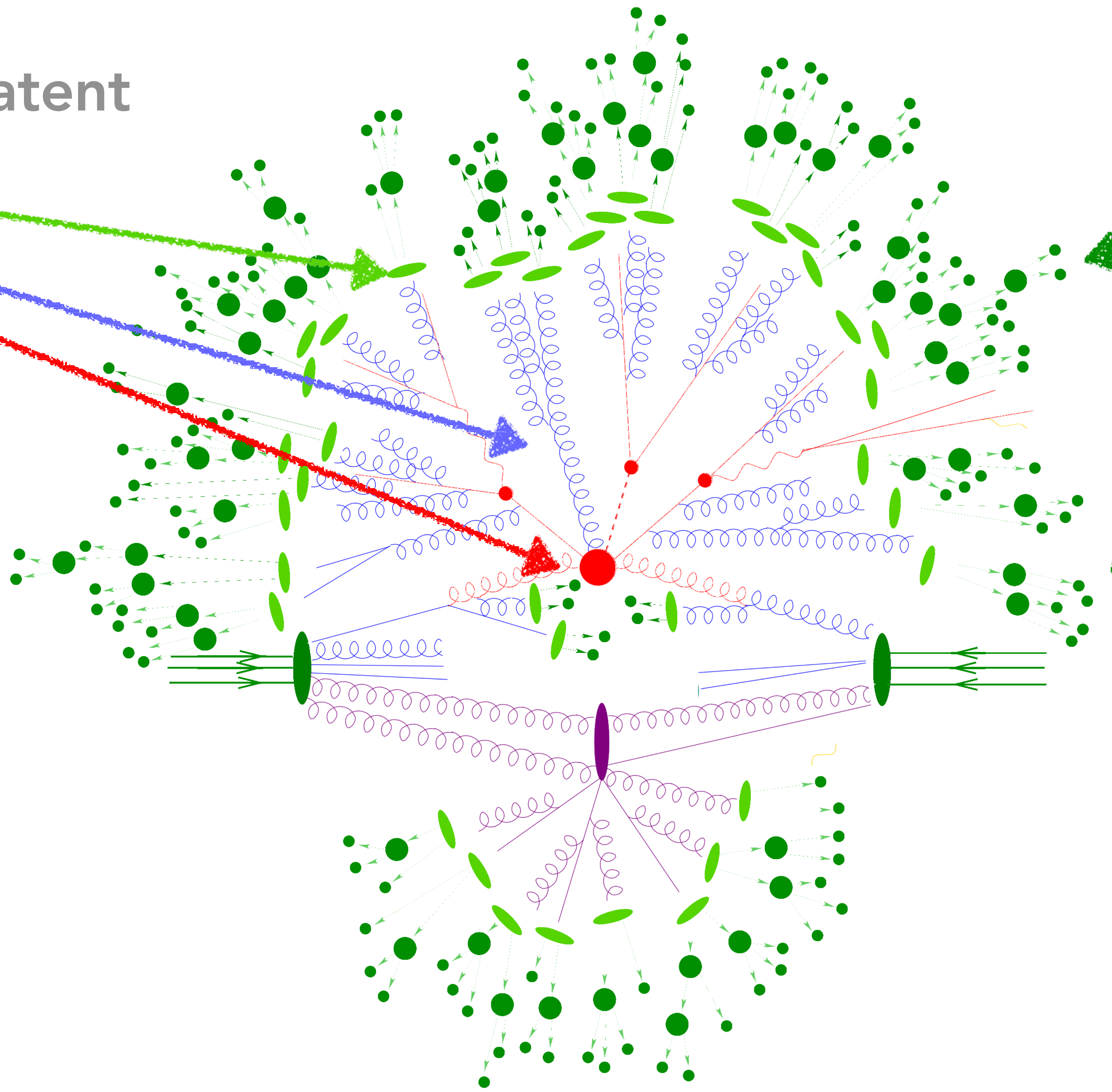
Causal, Generative Model For Jets

z

Evolution of the tree is **latent**

x

We only **observe** the **leaves**



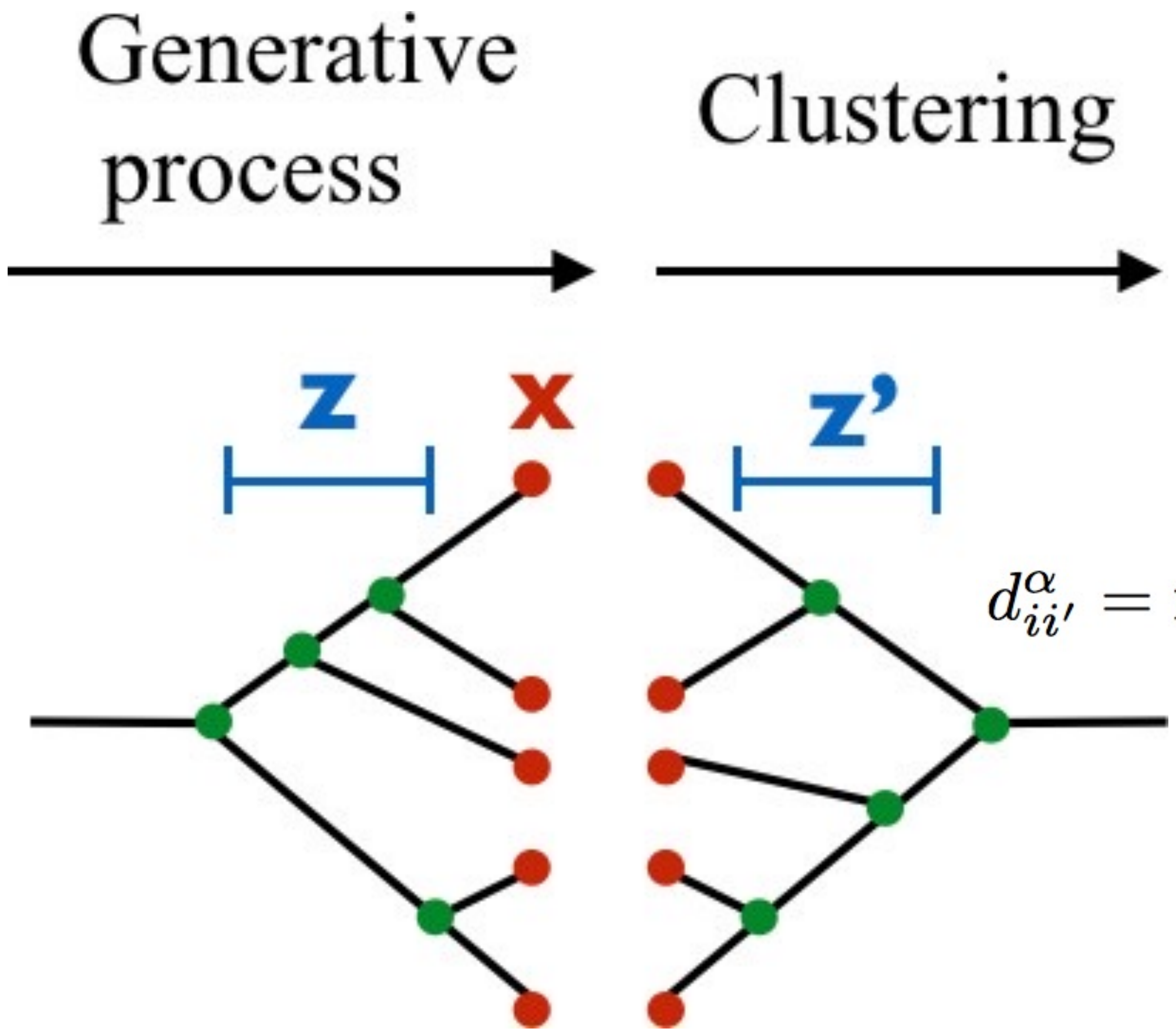
energy & momentum is conserved at each splitting

$$p(\text{tree}\quad) = \prod_{\text{nodes}} p(\text{split, left, right}\ \ \text{parent},\ )$$

Jet Clustering

Traditionally, physicists try to "reconstruct" the latent tree from observed

- Hierarchical clustering can be seen as inverting the generative process
- Standard techniques use bottom-up / greedy / agglomerative clustering HAC
  - Similarity measure is motivated by underlying physics: QCD, relativity, etc.



$$d^{\alpha}_{ii'} = \min(p^{2\alpha}_{ti}, p^{2\alpha}_{ti'})\frac{\Delta R^2_{ii'}}{R^2}$$

The anti-k_t jet clustering algorithm #1

Matteo Cacciari (Paris, LPTHE), Gavin P. Salam (Paris, LPTHE), Gregory Soyez (Brookhaven) (Feb, 2008)

Published in: *JHEP* 04 (2008) 063 • e-Print: 0802.1189 [hep-ph]

pdf DOI cite 7,676 citations

FastJet User Manual #1

Matteo Cacciari (Paris, LPTHE and Diderot U., Paris), Gavin P. Salam (CERN and Princeton U. and Paris, LPTHE), Gregory Soyez (Saclay, SPhT) (Nov, 2011)

Published in: *Eur.Phys.J.C* 72 (2012) 1896 • e-Print: 1111.6097 [hep-ph]

pdf DOI cite 3,945 citations

Dispelling the N^3 myth for the k_t jet-finder #1

Matteo Cacciari (Paris, LPTHE), Gavin P. Salam (Paris, LPTHE) (Dec, 2005)

Published in: *Phys.Lett.B* 641 (2006) 57-61 • e-Print: hep-ph/0512210 [hep-ph]

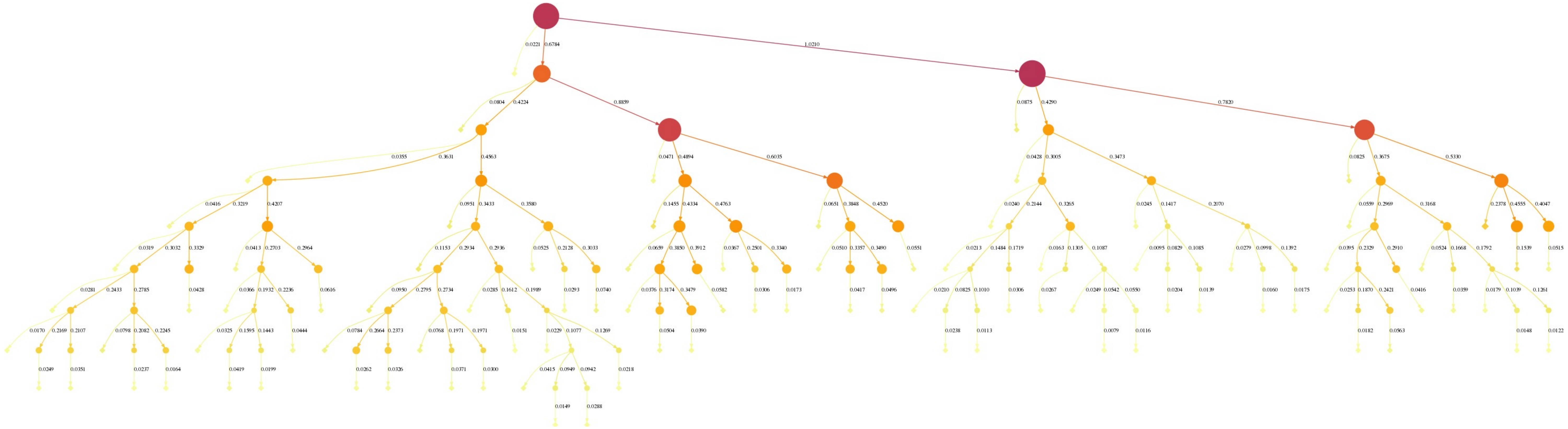
pdf DOI cite 1,988 citations

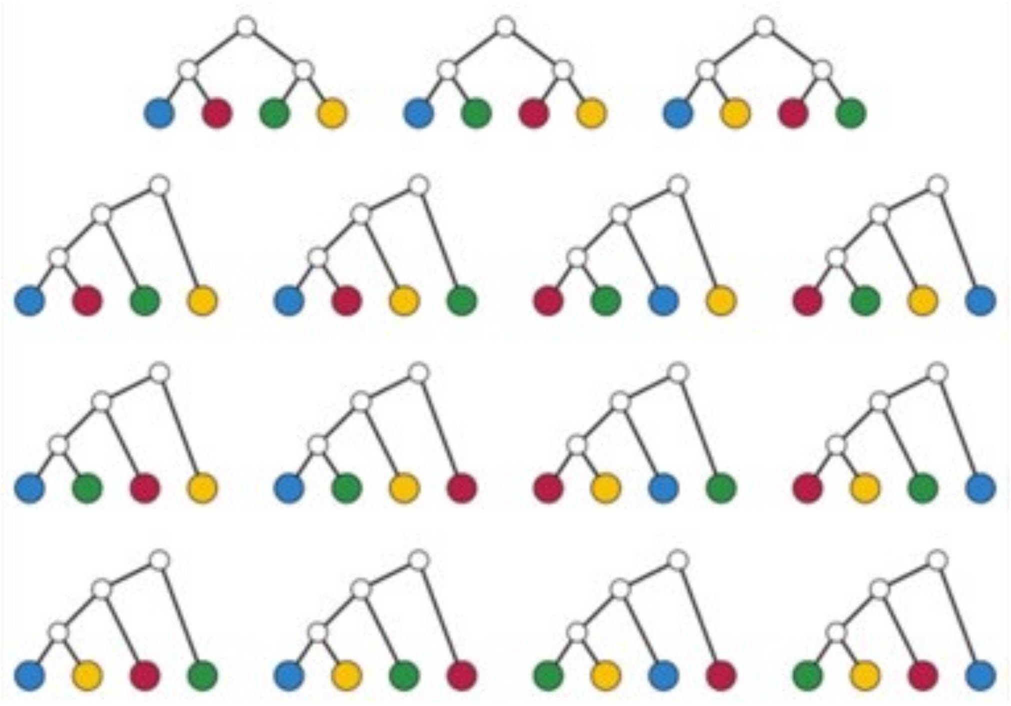
Impact of clustering on downstream tasks

Optimal solutions for many down-stream tasks would be ~trivial if an oracle could give us the correct / ground-truth tree, but

- Several trees could potentially lead to the same set of leaves, so **we need to think probabilistically if we want to work directly with shower model** $p(x, z)$

Alternatively, we can use machine learning to **learn functions** related to the **marginal distribution** $p(x) = p(x, z)dz$.





Trellis

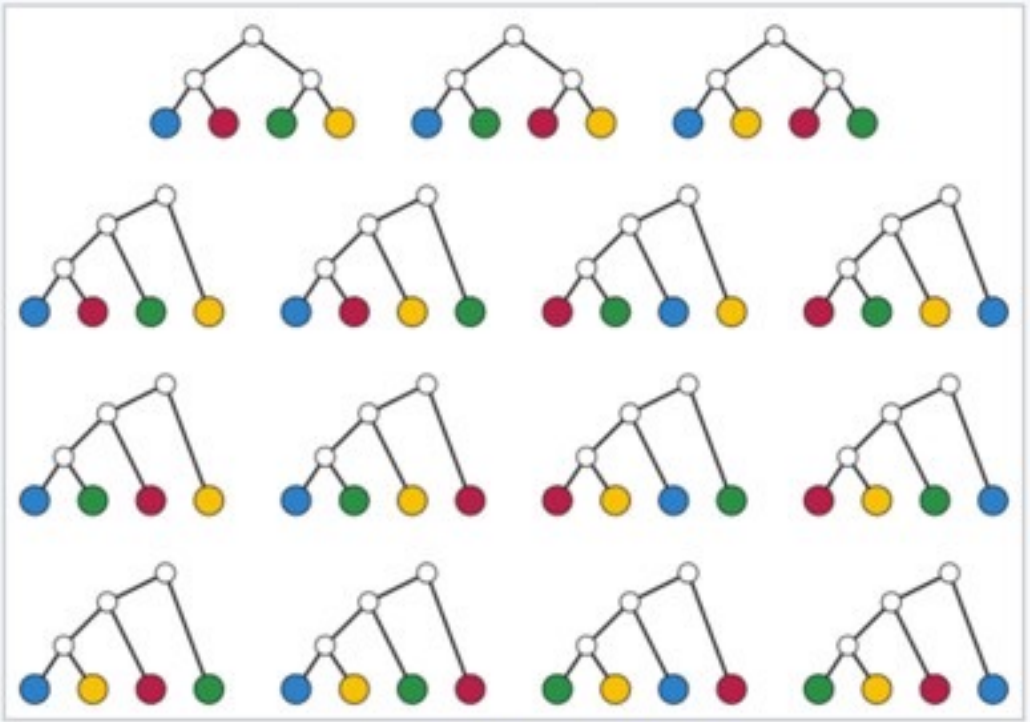
$$p(x|\theta) = \int dz \; p(x, z|\theta)$$

Trellis Ginkgo

The number of trees (hierarchical clusterings) is enormous!

- It grows like $2N-3)!$ where N is the number of jet constituents
- and there are 2^{N-1} permutations of)

# of leaves	Approx. # of trees
4	15
5	100
7	10 K
9	2 M
11	600 M
150	10^{300}



The fifteen different rooted binary trees (with unordered children) on a set of four labeled leaves, illustrating 15 = (2 × 4 − 3)!! (see article text).

https://en.wikipedia.org/wiki/Double_factorial

Trellis

$$p(x|\theta) = \int dz \; p(x, z|\theta)$$

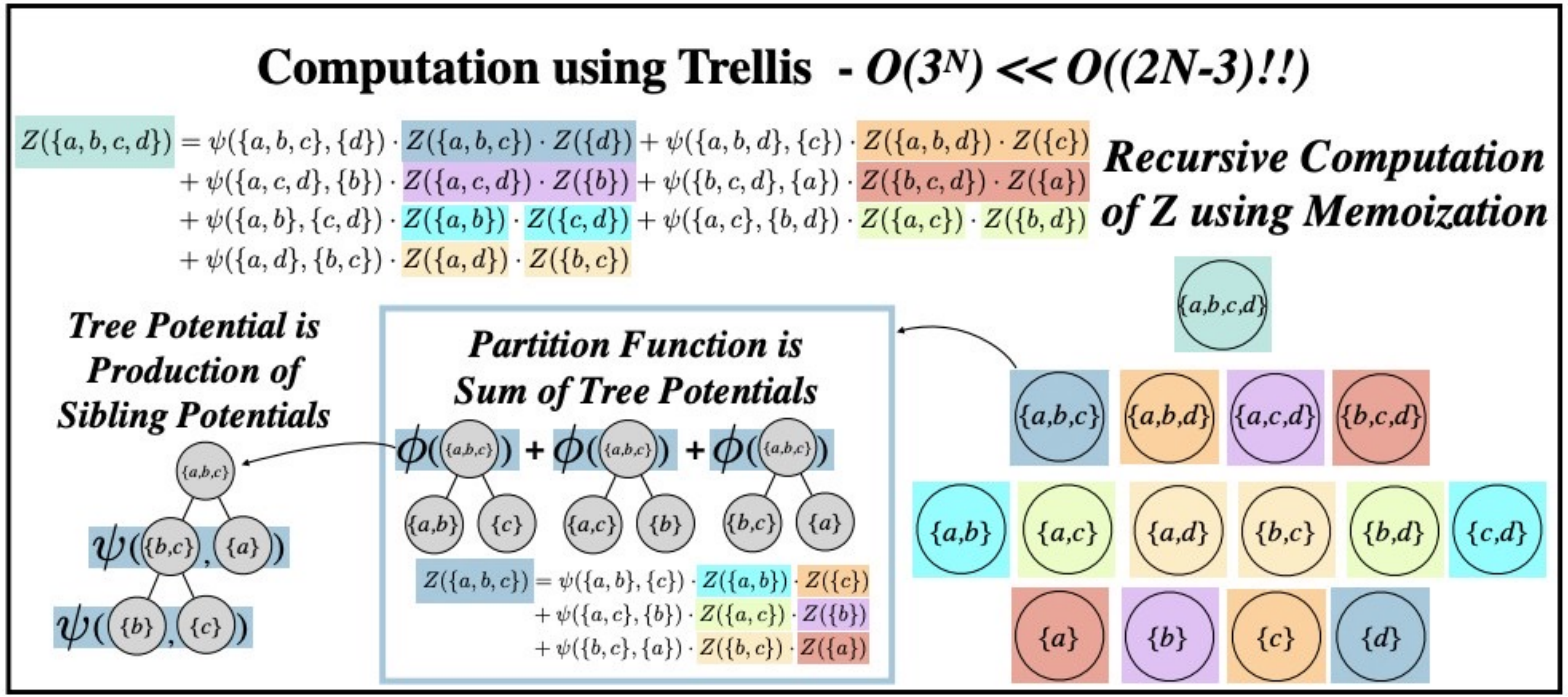
Trellis | Ginkgo

The number of trees (hierarchical clusterings) is enormous!

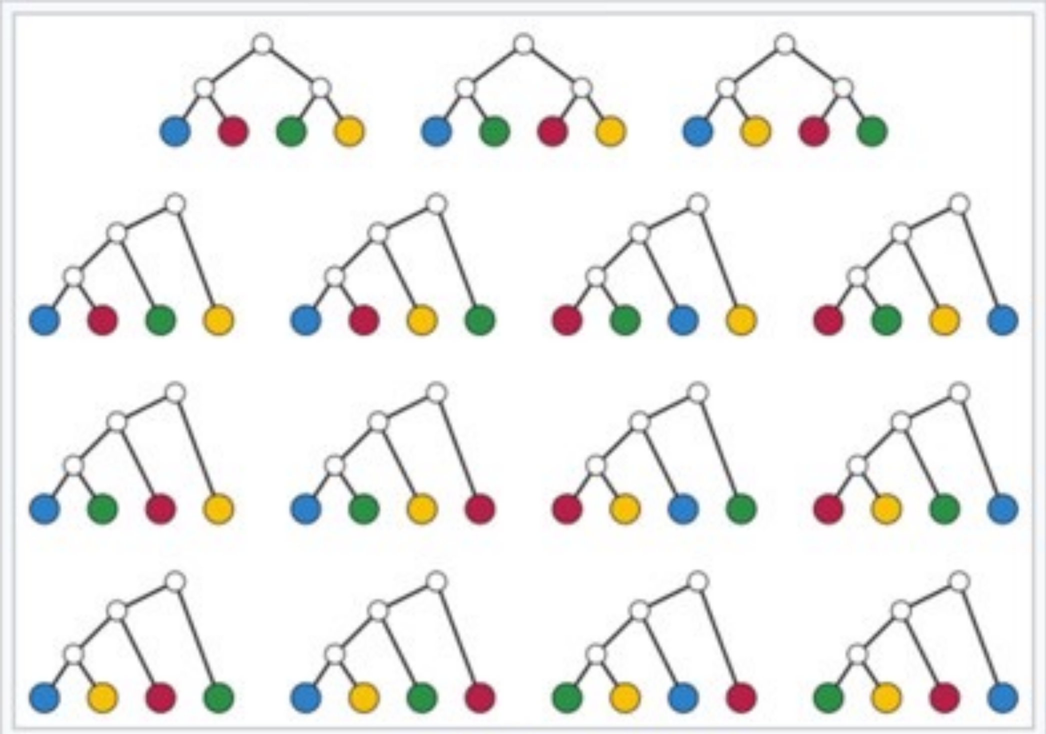
- It grows like $2N - 3)!$ where N is the number of jet constituents
- and there are 2^{N-1} permutations of)

The trellis is a data structure and dynamic programming algorithm that allows us to efficiently carry out this sum or find $z\hat{}= \arg\max_z p(z\;x,\;)$

# of leaves	Approx. # of trees
4	15
5	100
7	10 K
9	2 M
11	600 M
150	10^{300}



AISTATS 2021 [Greenberg, Macaluso, Monath, Cranmer, McCallum, et al '20, arXiv:2002.11661]



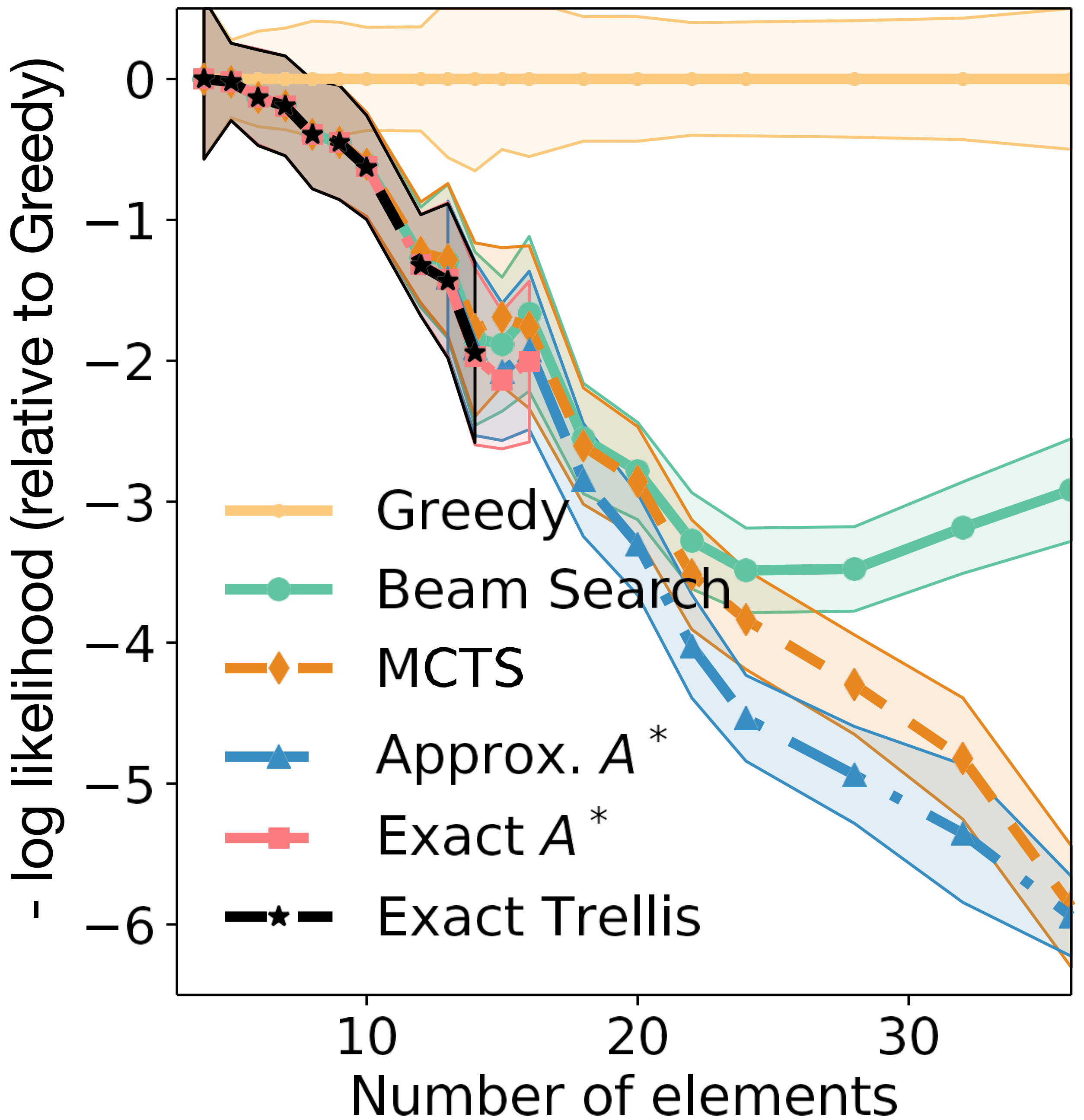
The fifteen different rooted binary trees (with unordered children) on a set of four labeled leaves, illustrating 15 = (2 × 4 − 3)!! (see article text).

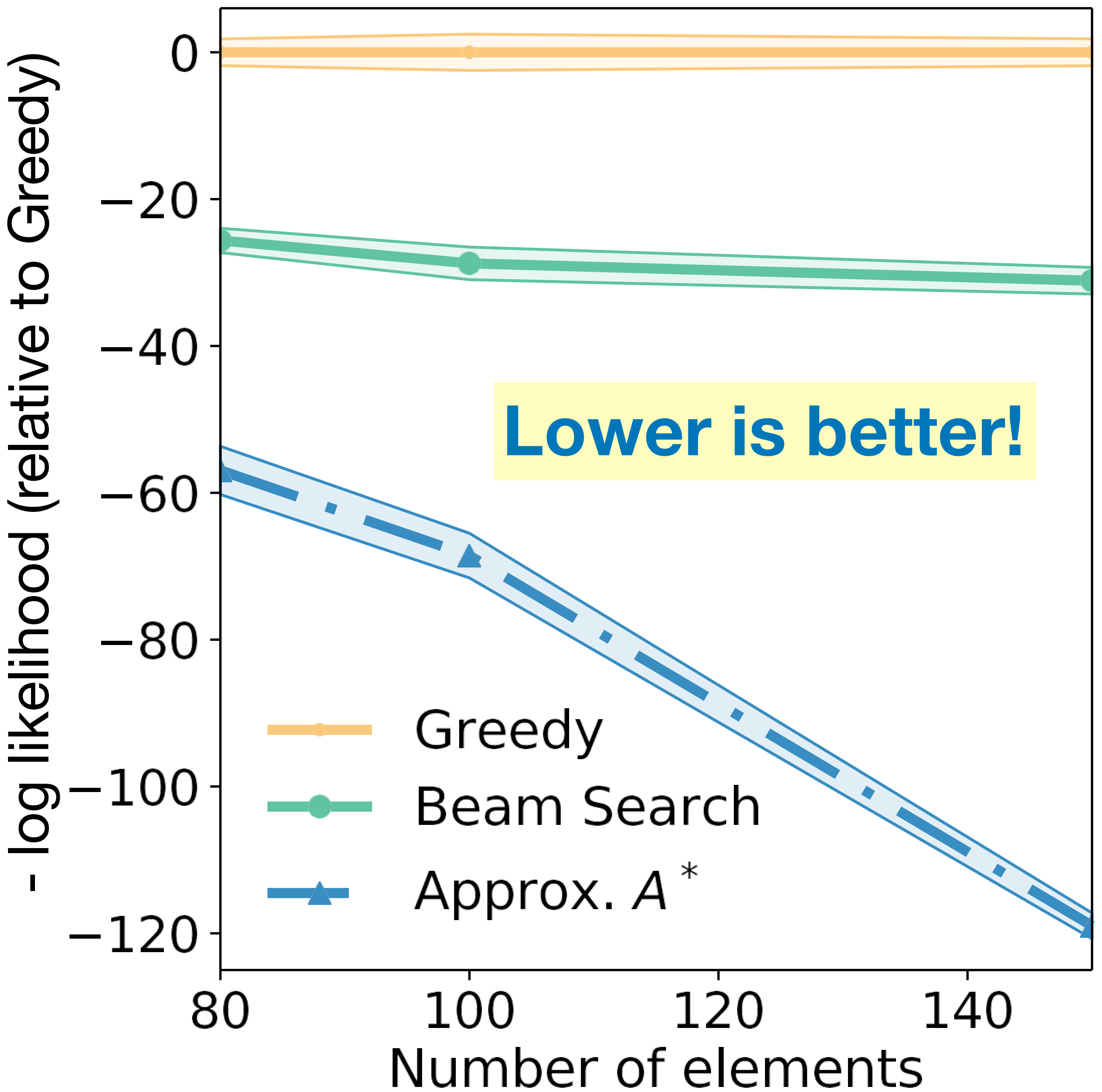
https://en.wikipedia.org/wiki/Double_factorial

Searching for the MAP tree / Combinatorial optimization

We find the exact MAP tree for up to 16 jet constituents: $z_{\text{MAP}} = \arg\max_{z} p(z \ \ x, \ \ )$

- We explored Monte Carlo Tree Search, other RL algs, approximate sparse trellis algorithms, and a novel A* search algorithm that uses the trellis
- Our approx. algorithms greatly improve over greedy and beam search baselines.





# of leaves	Approx. # of trees
4	15
5	100
7	10 K
9	2 M
11	600 M
16	10^{16}
150	10^{300}

Unexpected applications

These physics-inspired algorithms are relevant for genomics, phylogenetic trees and networks

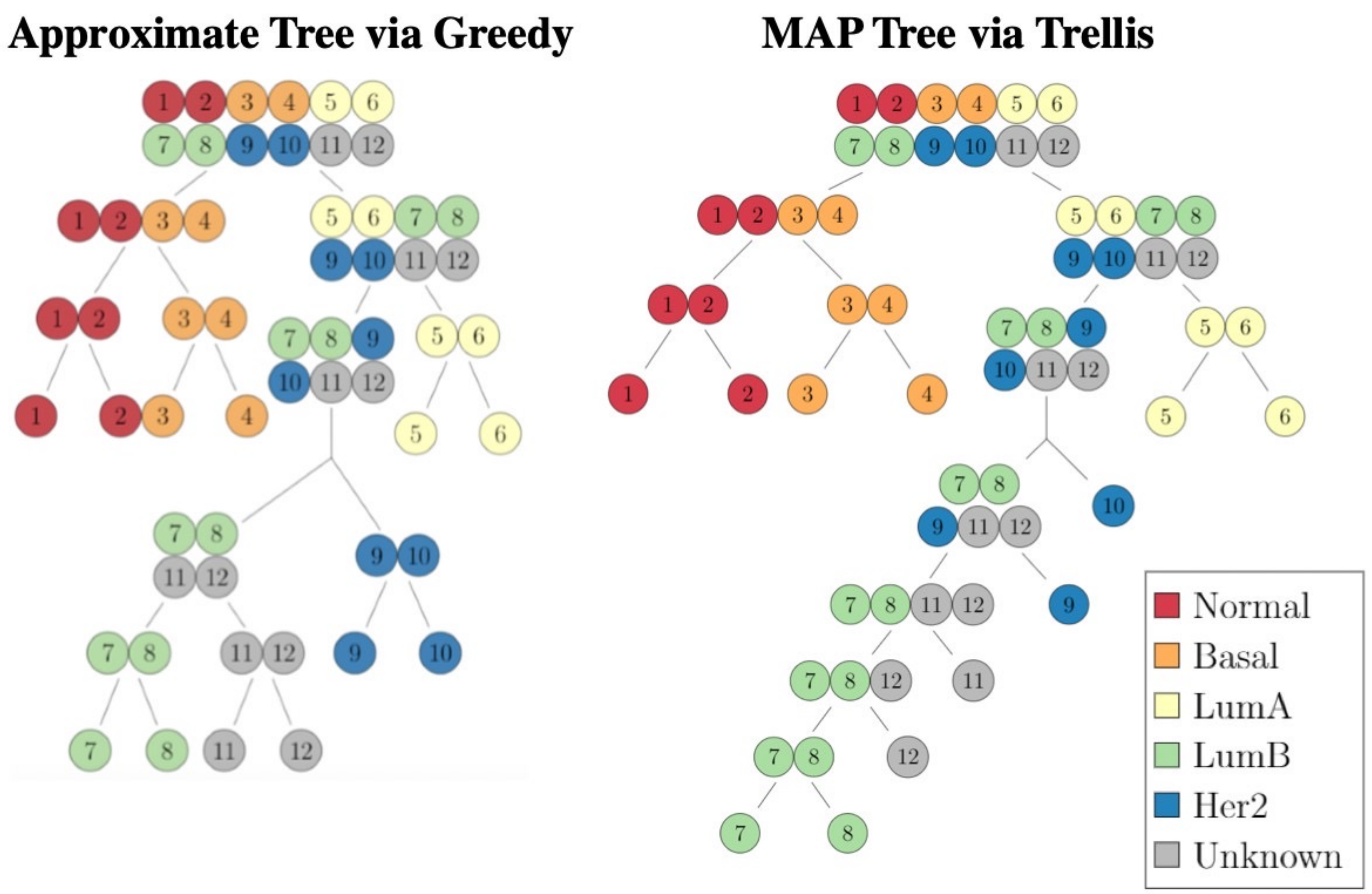
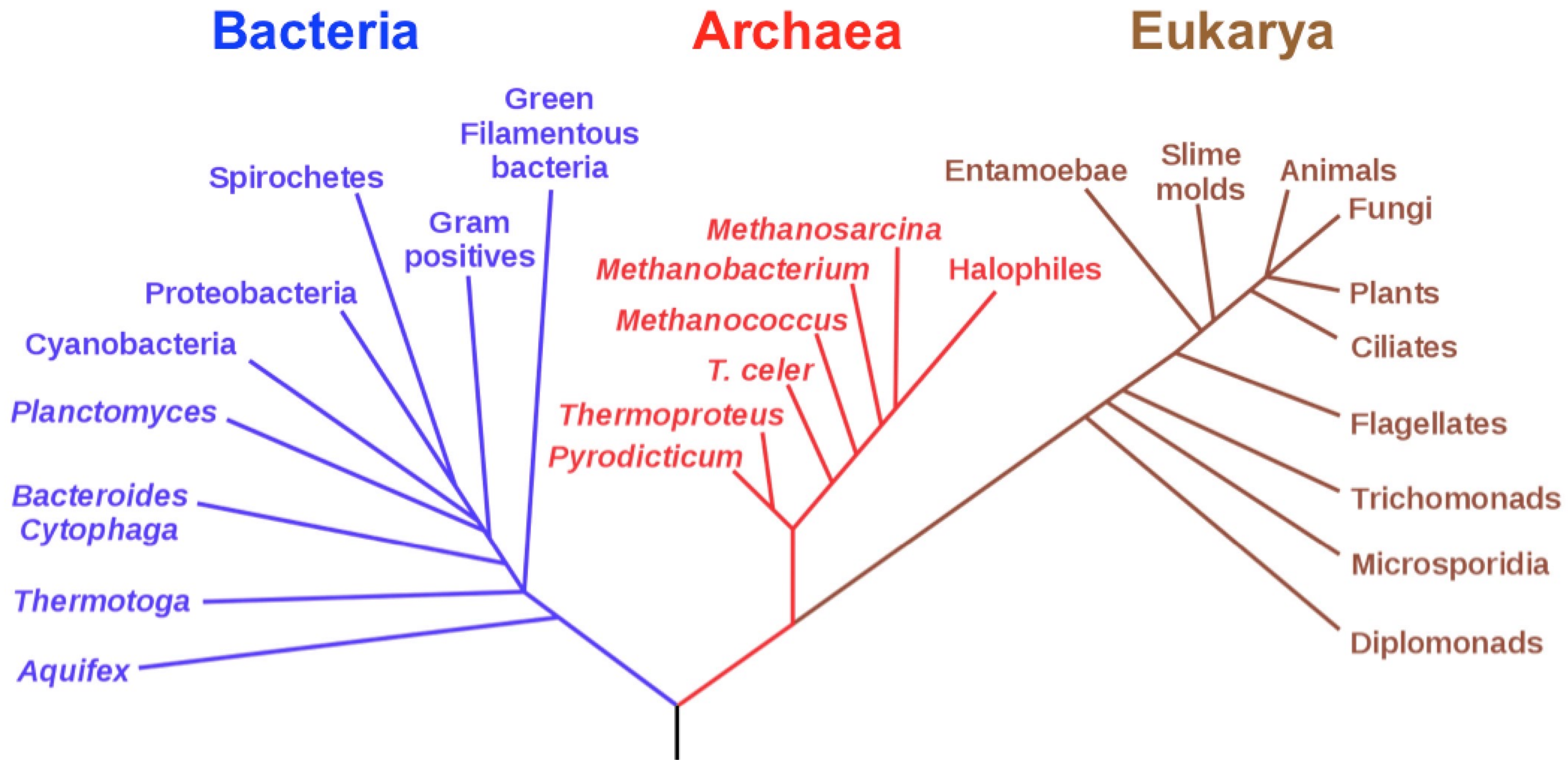


Figure 8: **Cancer Genomics**. Comparison of trees from greedy hierarchical clustering (left) and exact MAP clustering using the trellis (right) on the subsampled `pam50` data set. The colors indicate subtypes of breast cancer (grey if unknown). Though both appear to assign unknown samples to LumB, the right tree positions the unknown samples closer to the Her2 samples.



An IPAM workshop

Highly recommended:

http://www.ipam.ucla.edu/programs/workshops/deep-learning-and-combinatorial-optimization/

Deep Learning and Combinatorial Optimization

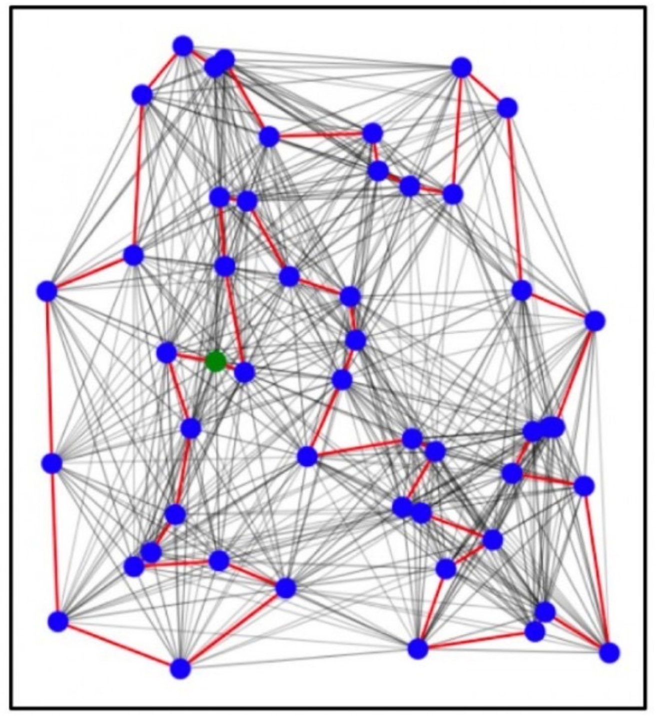
FEBRUARY 22 - 25, 2021

OVERVIEW | SPEAKER LIST | SCHEDULE

Overview

Virtual Workshop: In response to COVID-19, all participants will attend this workshop virtually via Zoom. Workshop registrants will receive the Zoom link a few days prior to the workshop, along with instructions on how to participate. The video of the recorded sessions will be made available on IPAM **website**.

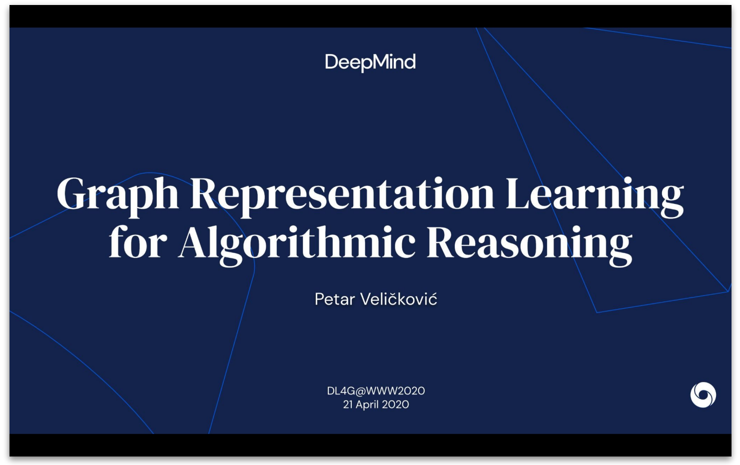
Workshop Overview: In recent years, deep learning has significantly improved the fields of computer vision, natural language processing and speech recognition. Beyond these traditional fields, deep learning has been expended to quantum chemistry, physics, neuroscience, and more recently to combinatorial optimization (CO). Well-known CO problems are Travelling Salesman Problem, assignment problems, routing, planning, Bayesian search, and scheduling. CO is basically used every day in finance and revenue management, transportation, manufacturing, supply chain, public policy, hardware design, computing and information technology.



Most combinatorial problems are difficult to solve, often leading to heuristic solutions which require years of research work and significant specialized knowledge. For example, the famous TSP problem has been studied for more than 80 years, and the best solver leverages 30 years of theoretical developments, data structures and heuristics from computer science. In the last few years, deep learning has developed some preliminary but promising approaches to deal with classical CO problems such as TSP, MaxCut, Minimum Vertex Cover, Knapsack, Quadratic Assignment Problem and Vehicle Routing Problems. DL is particularly attractive to address CO problems given its high flexibility, approximate nature, and self-learning paradigm. In other words, DL has the potential to learn universal

Further insight

If you would like to know more details about constructing good processor networks:



https://www.youtube.com/watch?v=IPQ6CPoluok

https://drive.google.com/file/d/1_EQ9Yu7VEkvrHaVHl_WbT5ABvxrSNY-s/view?usp=sharing

Want to know more?

Combinatorial optimization and reasoning with graph neural networks

Quentin Cappart[1], Didier Chételat[2], Elias Khalil[3], Andrea Lodi[2], Christopher Morris[2], and Petar Veličković[*4]

[1]Department of Computer Engineering and Software Engineering, Polytechnique Montréal
[2]CERC in Data Science for Real-Time Decision-Making, Polytechnique Montréal
[3]Department of Mechanical & Industrial Engineering, University of Toronto
[4]DeepMind

Combinatorial optimization is a well-established area in operations research and computer science. Until recently, its methods have focused on solving problem instances in isolation, ignoring the fact that they often stem from related data distributions in practice. However, recent years have seen a surge of interest in using machine learning, especially graph neural networks (GNNs), as a key building block for combinatorial tasks, either as solvers or as helper functions. GNNs are an inductive bias that effectively encodes combinatorial and relational input due to their permutation-invariance and sparsity awareness. This paper presents a conceptual review of recent key advancements in this emerging field, aiming at both the optimization and machine learning researcher.

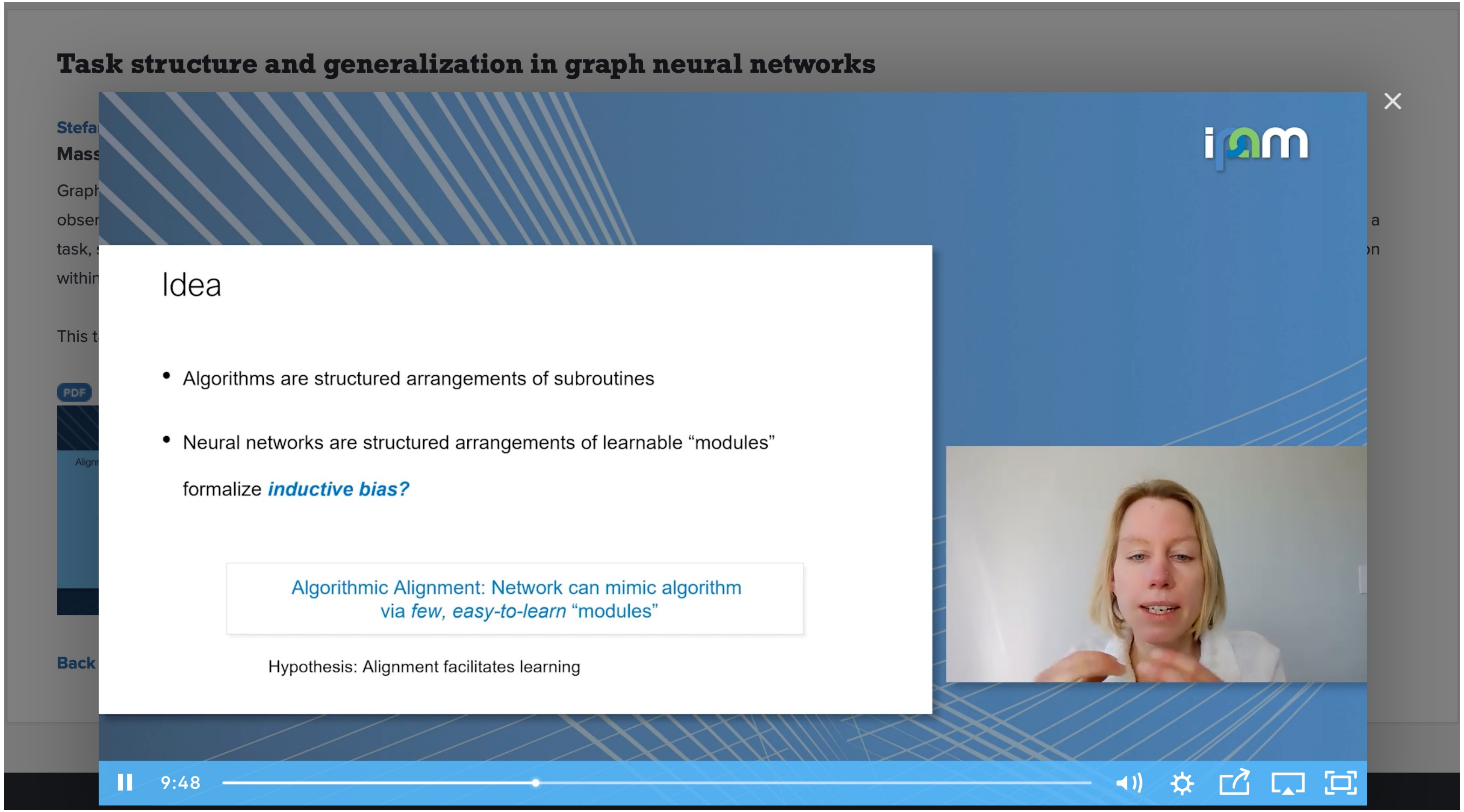
Our *43-page* survey on GNNs for CO!

https://arxiv.org/abs/2102.09544

Section 3.3. details algorithmic reasoning, with comprehensive references.

Algorithmic Alignment

http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021



Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021

Alignment more generally

More generally: GNNs align with Dynamic Programming

$$\text{Answer}[k][i] = \text{DP-Update}(\{\text{Answer}[k-1][j],\ j = 1 \ldots n\})$$

$$h_s^{(k)} = \boxed{\sum_{t \in S} \text{MLP}_1^{(k)}} \left(h_s^{(k-1)}, h_t^{(k-1)}\right)$$

Many algorithmic / physical reasoning tasks are DPs!

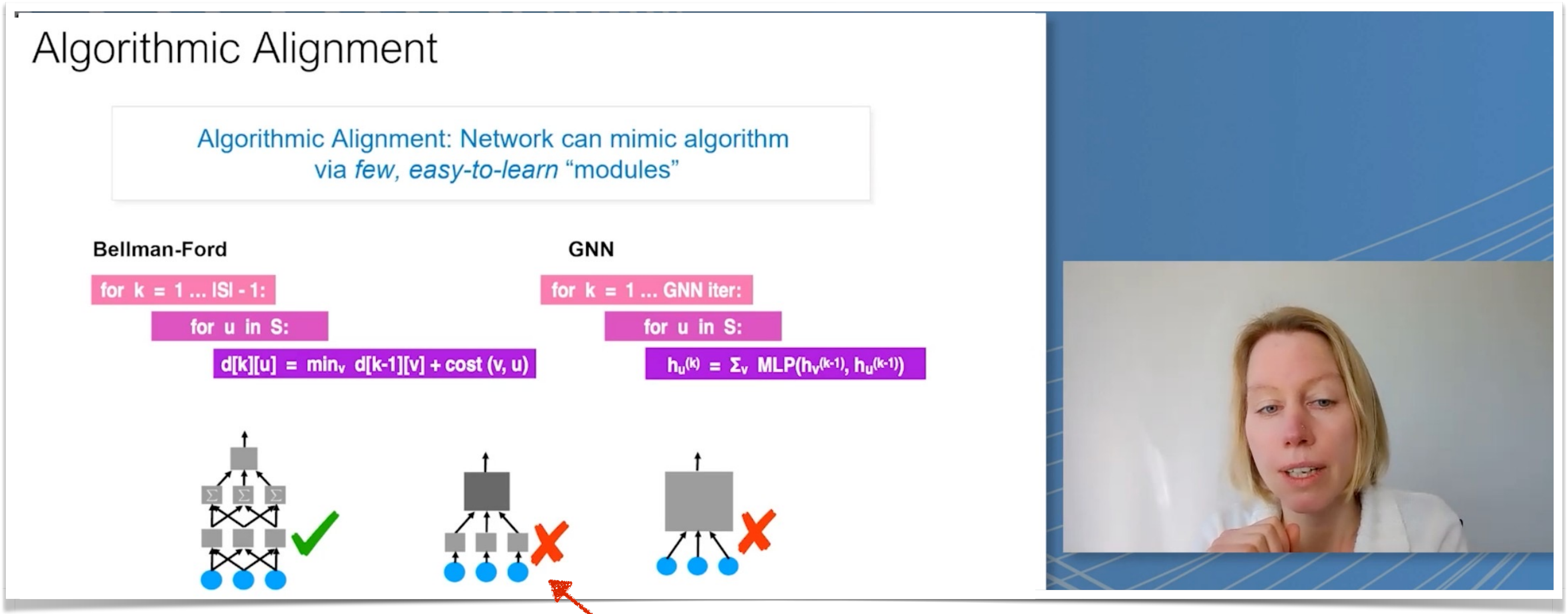
Formalization:

A neural network (M, ϵ, δ)-aligns with an algorithm if it can mimic the algorithm via *n* different (shared) network modules, each of which is PAC-learnable with at most M/n samples.

Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

Algorithmic Alignment

http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021

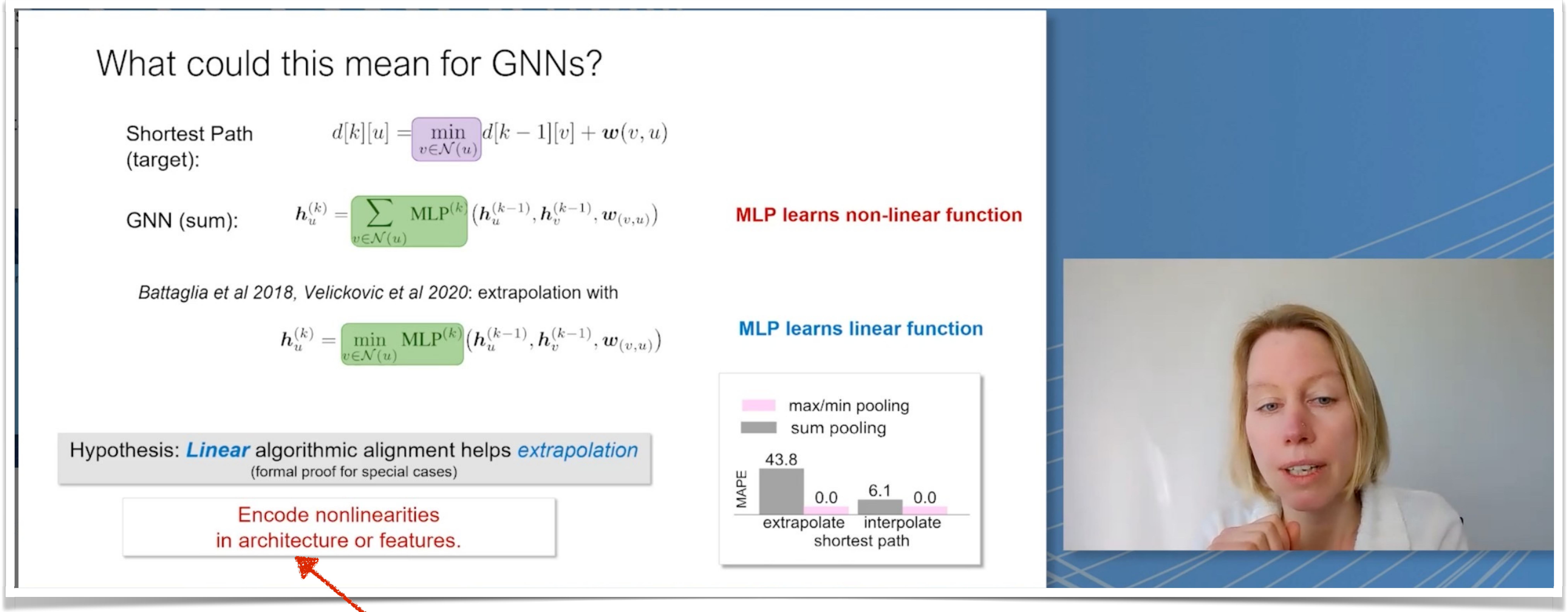


Deep Sets with sum aggregation will have problems

Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

Algorithmic Alignment

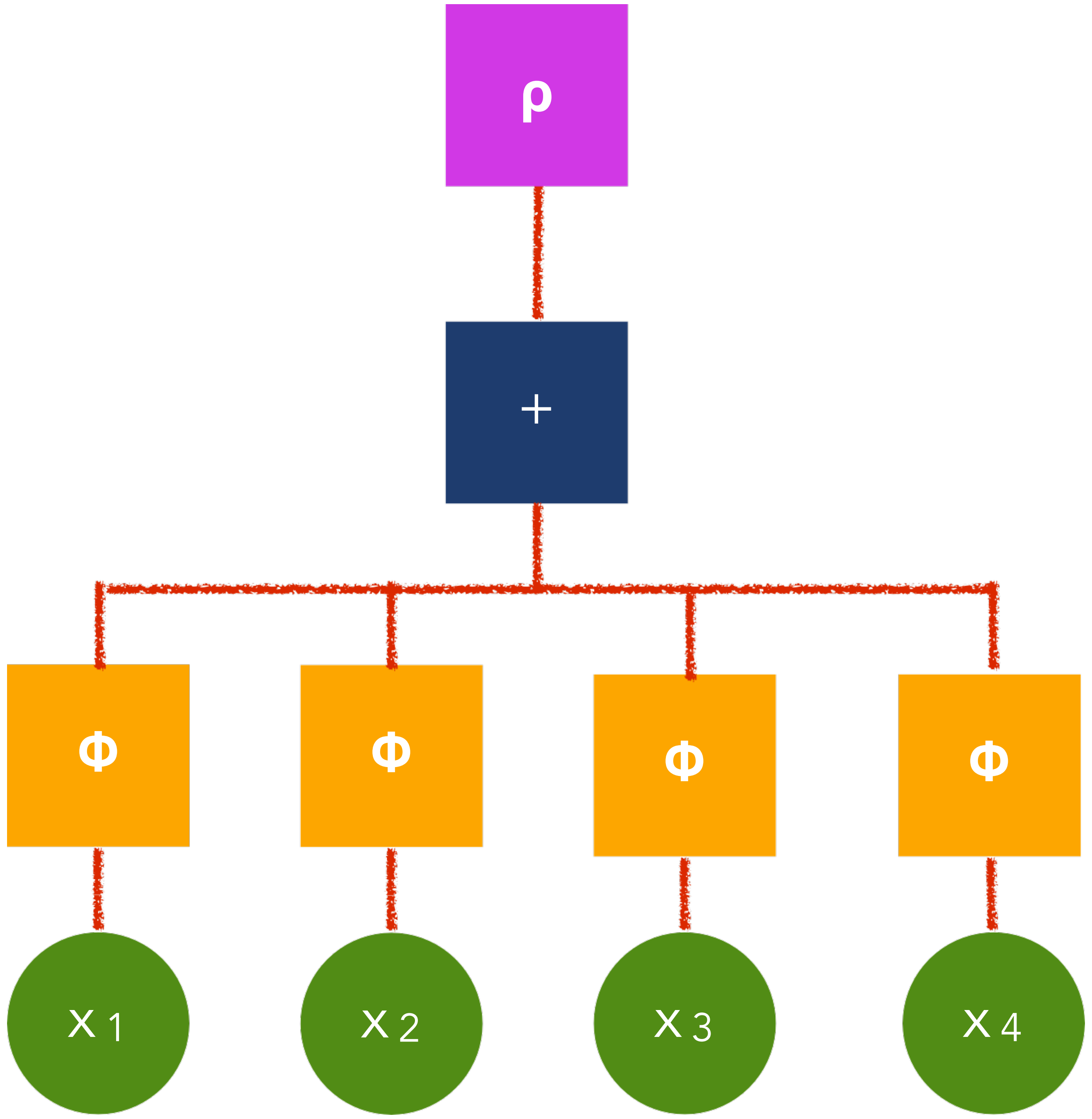
http://www.ipam.ucla.edu/abstract/?tid=16710&pcode=DLC2021

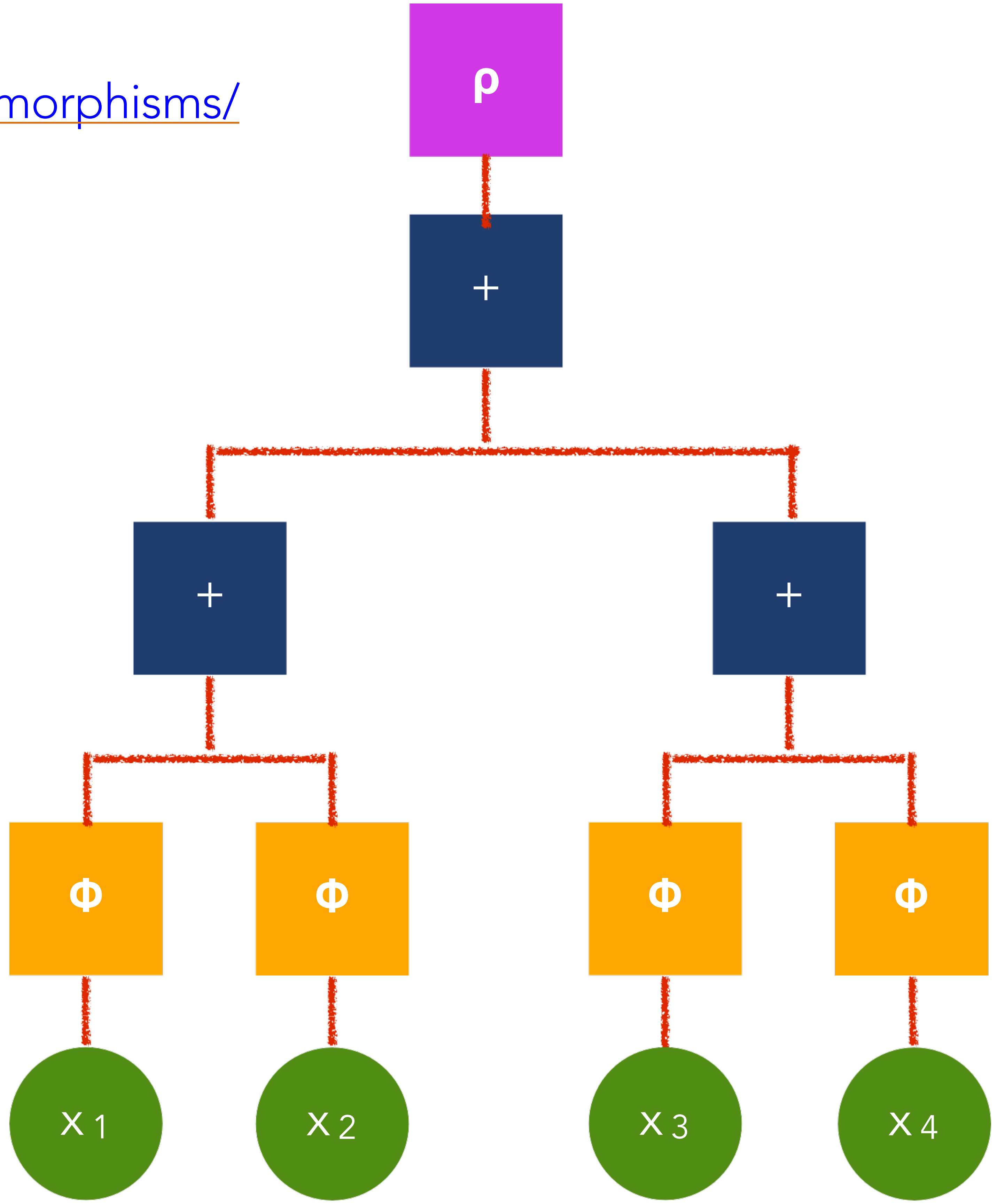


Stefanie Jegelka's talk at IPAM workshop on Deep Learning and Combinatorial Optimization

We can aggregate the sum for DeepSets hierarchically if we want, so DeepSets has a TreeRNN like architecture

- A "tree catamorphism" https://blog.ploeh.dk/2019/04/29/catamorphisms/





Examples of Physics-inspired ML

Monte Carlo

Famous case of physics-inspired "ML"

- Stanislaw Ulam invented the modern version of the **Markov Chain Monte Carlo** method while he was working on nuclear weapons projects at the Los Alamos National Laboratory in the 1940s
- Metropolis–Hastings - Equation of State Calculations by Fast Computing Machines by Nicholas Metropolis, Arianna W. Rosenbluth, Marshall Rosenbluth, Augusta Teller and Edward Teller.
- Hamiltonian (Hybrid) MC - originally proposed by Simon Duane, Anthony Kennedy, Brian Pendleton and Duncan Roweth in 1987 for calculations in lattice quantum chromodynamics.

•… many more approaches developed for molecular dynamics, quantum systems, etc.

Current development: flow-based sampling strategies

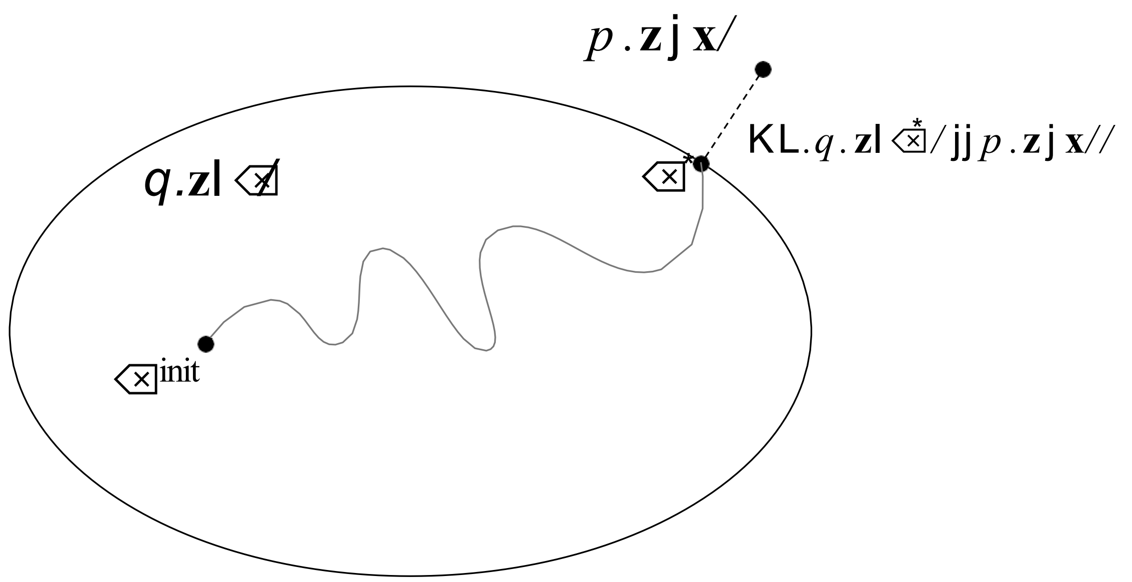
Variational Inference

Variational Inference also has its origin in physics (Free energy, reverse KL divergence, ELBO)

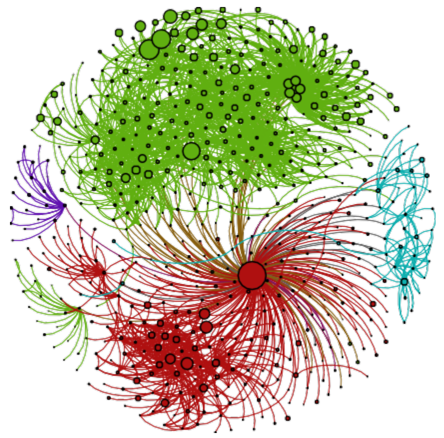
$$\mathcal{L}(\nu) = \mathbb{E}_q[\log p(\mathbf{z})] + \sum_{i=1}^{n} \mathbb{E}_q[\log p(x_i \mid \mathbf{z})] - \mathbb{E}_q[\log q(\mathbf{z}\,;\nu)]$$

- KL is intractable; VI optimizes the **evidence lower bound** (ELBO) instead.
- Two ways to form **noisy gradients** with Monte Carlo:
  - Sample from $q(\cdot)$; no need to take expectations
  - Subsample from the data; no need to process all data
- This leads to **scalable** and **generic** variational inference.

Variational inference

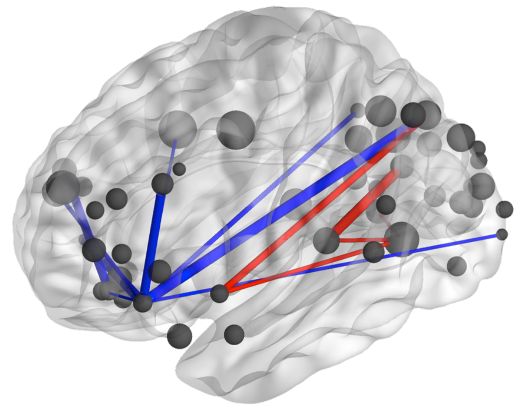


- VI solves **inference** with **optimization**.
- Posit a **variational family** of distributions over the latent variables.
- Fit the **variational parameters** ν to be close (in KL) to the exact posterior.



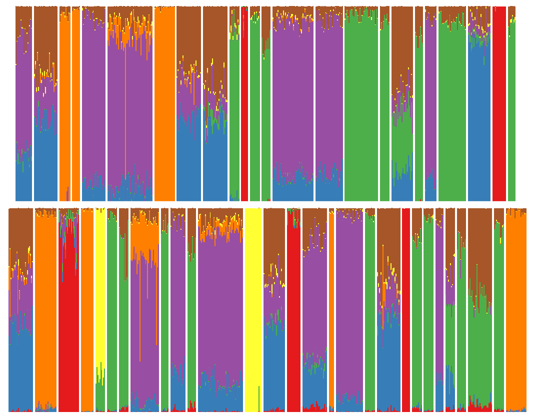
Communities discovered in a 3.7M node network of U.S. Patents

[Gopalan and Blei PNAS 2013]



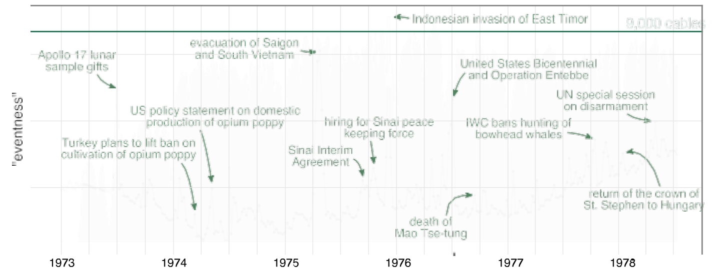
Neuroscience analysis of 220 million fMRI measurements

[Manning+ PLOS ONE 2014]



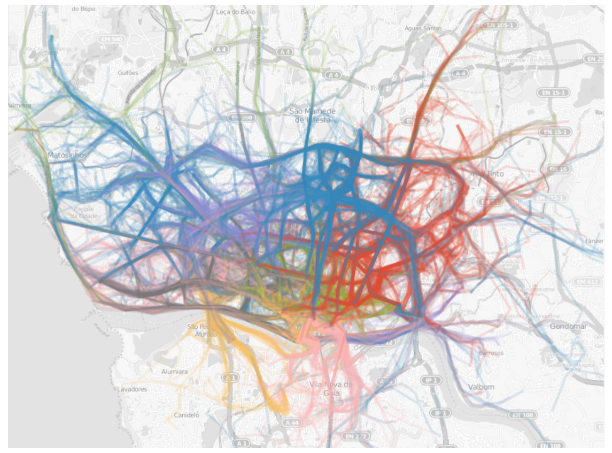
Population analysis of 2 billion genetic measurements

[Gopalan+ Nature Genetics 2016]



Analysis of 2M declassified cables from the State Dept

[Chaney+ EMNLP 2016]

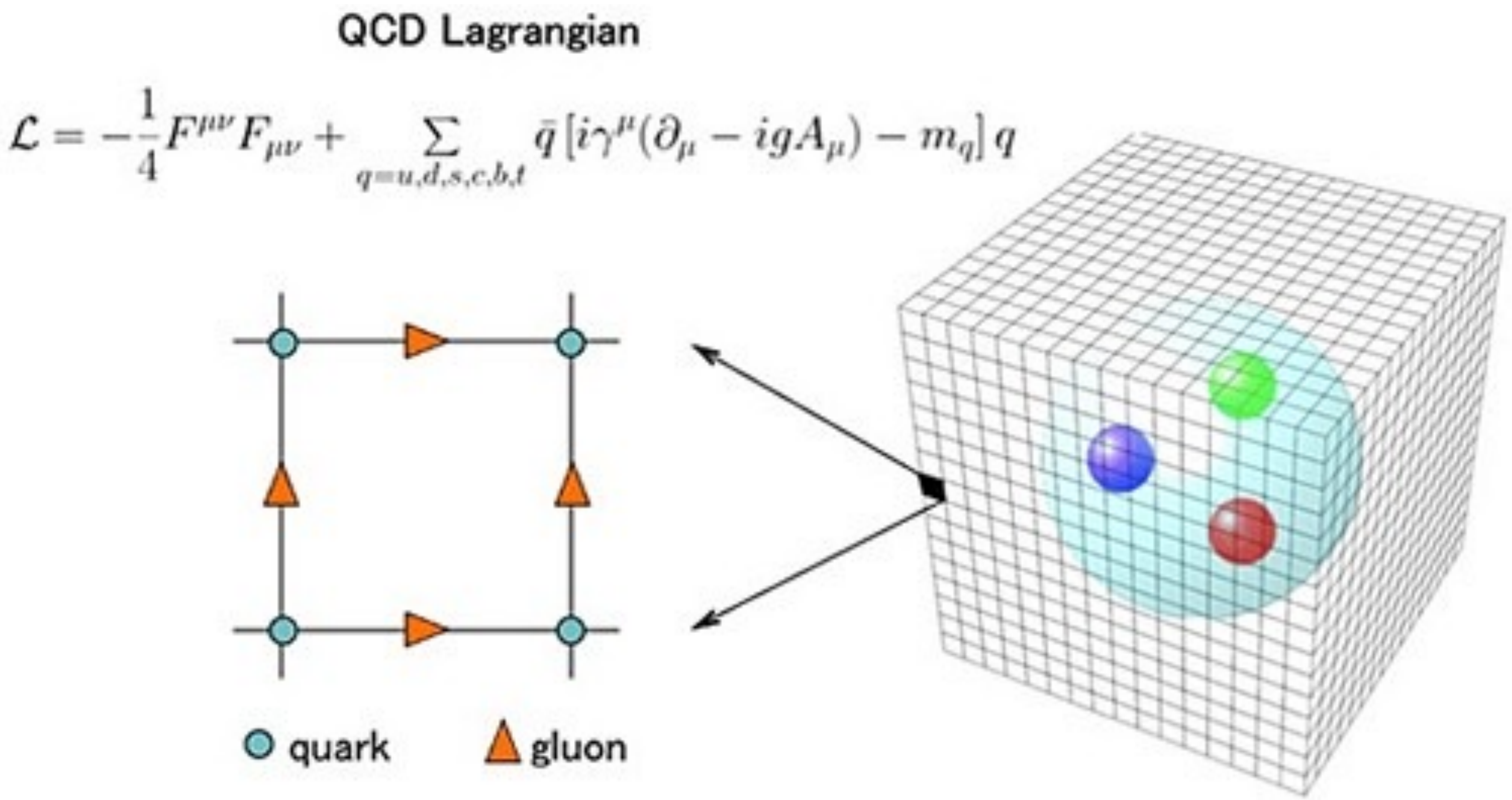


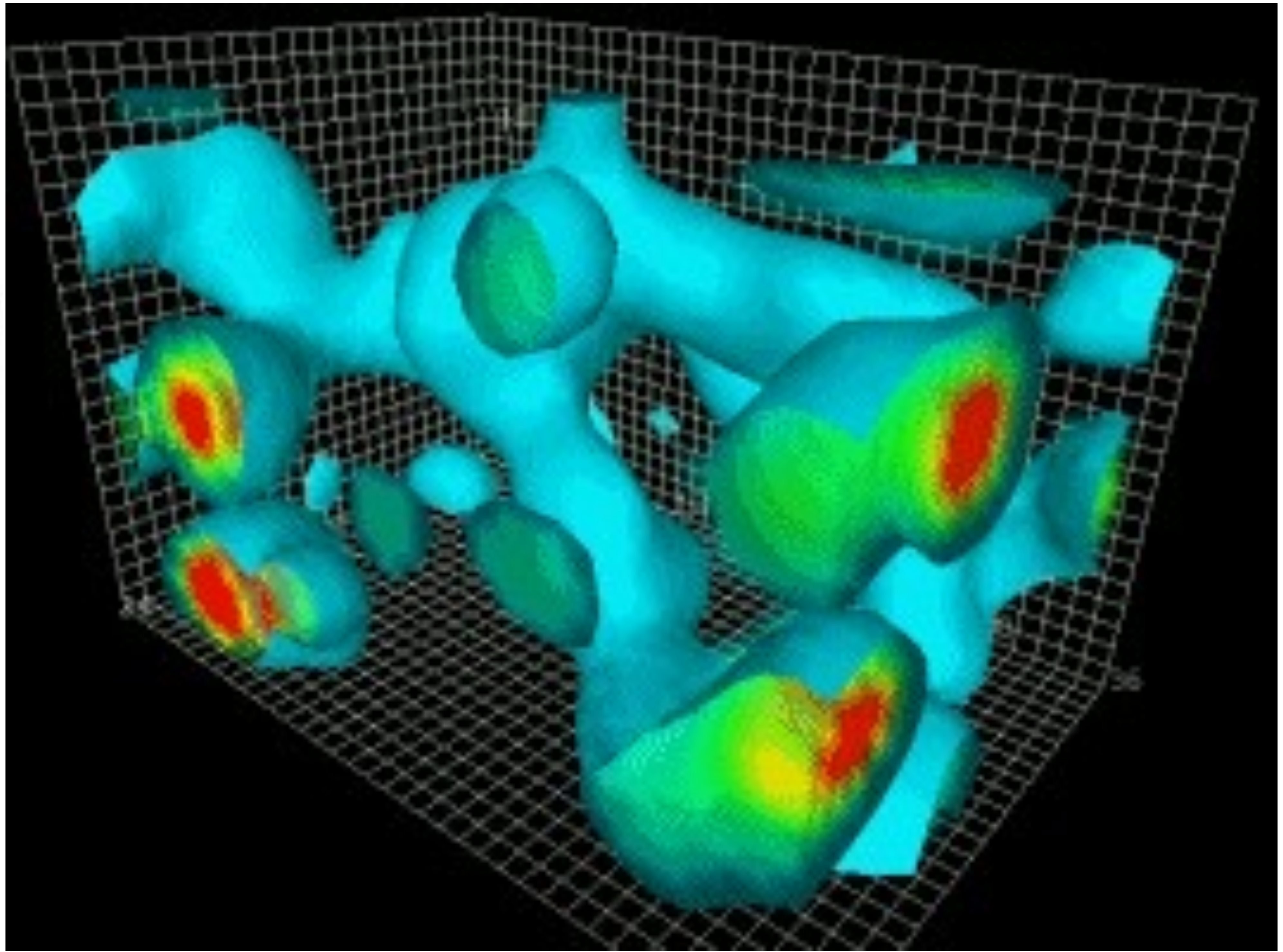
Analysis of 1.7M taxi trajectories, in Stan

[Kucukelbir+ JMLR 2016]

Very expensive simulations with high dimensional data: eg. $x \in \mathbb{R}^{10^9}$

Use normalizing flow to approximate target distribution of configurations that is implied by the action $S(x$ via the Boltzmann Equation $p(x) = e^{-S(x)}/Z$

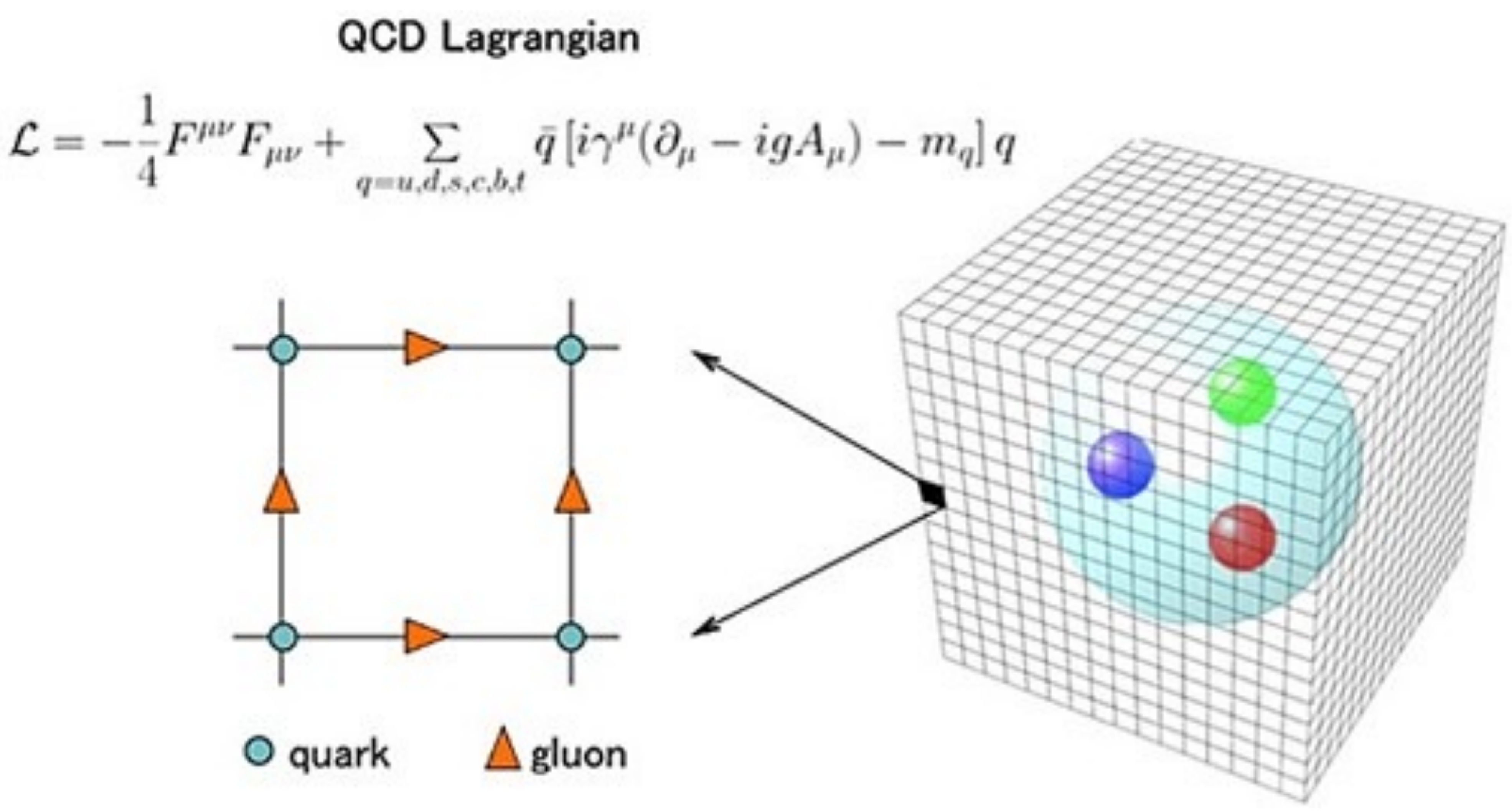


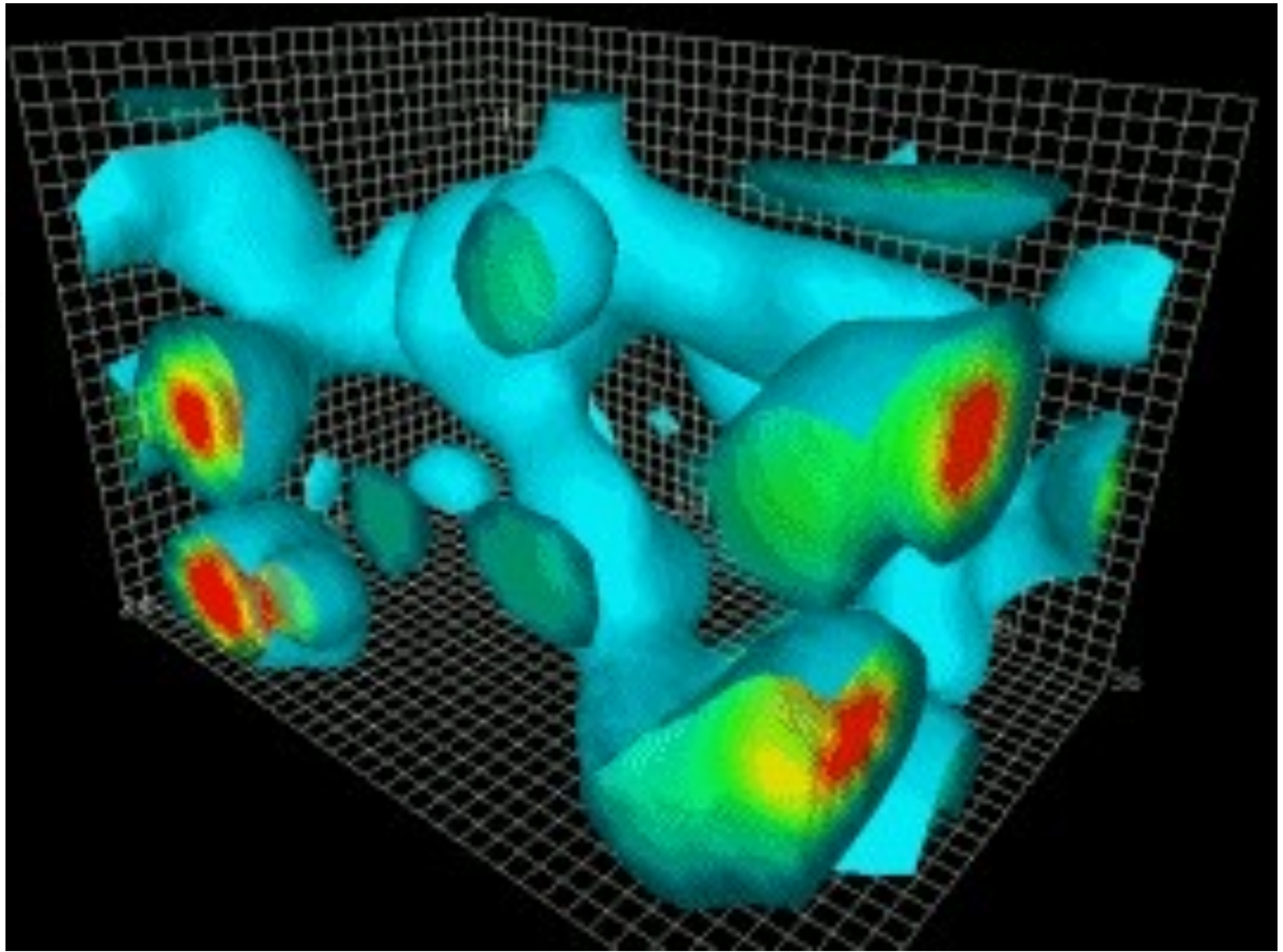


Lattice Field Theory

Very expensive simulations with high dimensional data: eg. $x \in \mathbb{R}^{10^9}$

Use normalizing flow to approximate target distribution of configurations that is implied by the action $S(x$ via the Boltzmann Equation $p(x) = e^{-S(x)}/Z$





Flows for LQCD

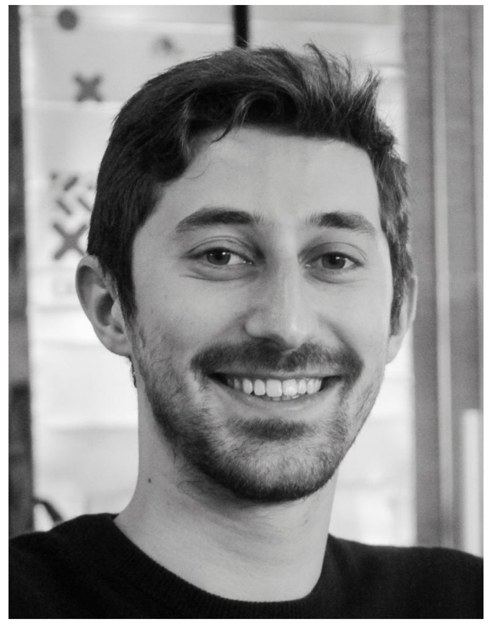
Albergo, Kanwar, Shanahan, PRD (2019) arXiv:1904.12072

Flow-based generative models for Markov chain Monte Carlo in lattice field theory

M. S. Albergo,[1,2,3] G. Kanwar,[4] and P. E. Shanahan[4,1]

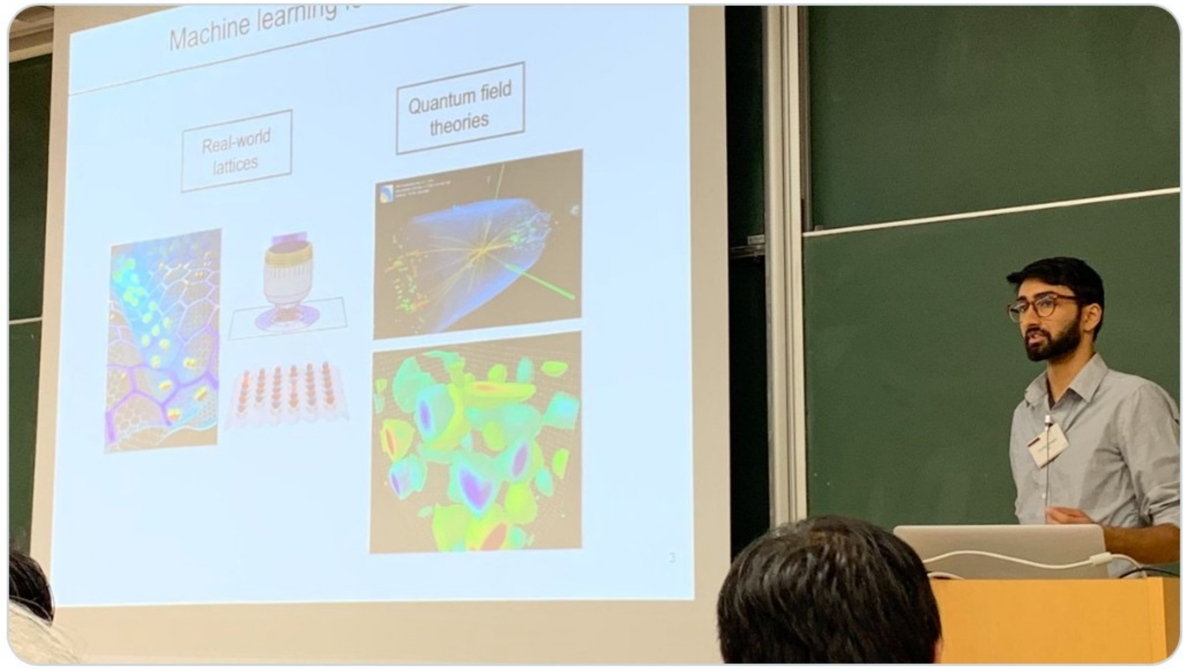
[1]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[2]*Cavendish Laboratories, University of Cambridge, Cambridge CB3 0HE, U.K.*
[3]*University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[4]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

A Markov chain update scheme using a machine-learned *flow-based generative model* is proposed for Monte Carlo sampling in lattice field theories. The generative model may be optimized (trained) to produce samples from a distribution approximating the desired Boltzmann distribution determined by the lattice action of the theory being studied. Training the model systematically improves autocorrelation times in the Markov chain, even in regions of parameter space where standard Markov chain Monte Carlo algorithms exhibit critical slowing down in producing decorrelated updates. Moreover, the model may be trained without existing samples from the desired distribution. The algorithm is compared with HMC and local Metropolis sampling for ϕ^4 theory in two dimensions.



Enrico Rinaldi @enricesena · Nov 1

Yesterday **Gurtej Kanwar** told us about machine learning for lattice field theories and exciting progress in Generative Models for gauge theories (collaboration with @DeepMindAI) at #DLAP2019 Today is the last day of this great conference!



5 15

Flows for molecular dynamics

RESEARCH   Noé *et al.*, *Science* 365, 1001 (2019)   6 September 2019

RESEARCH ARTICLE SUMMARY

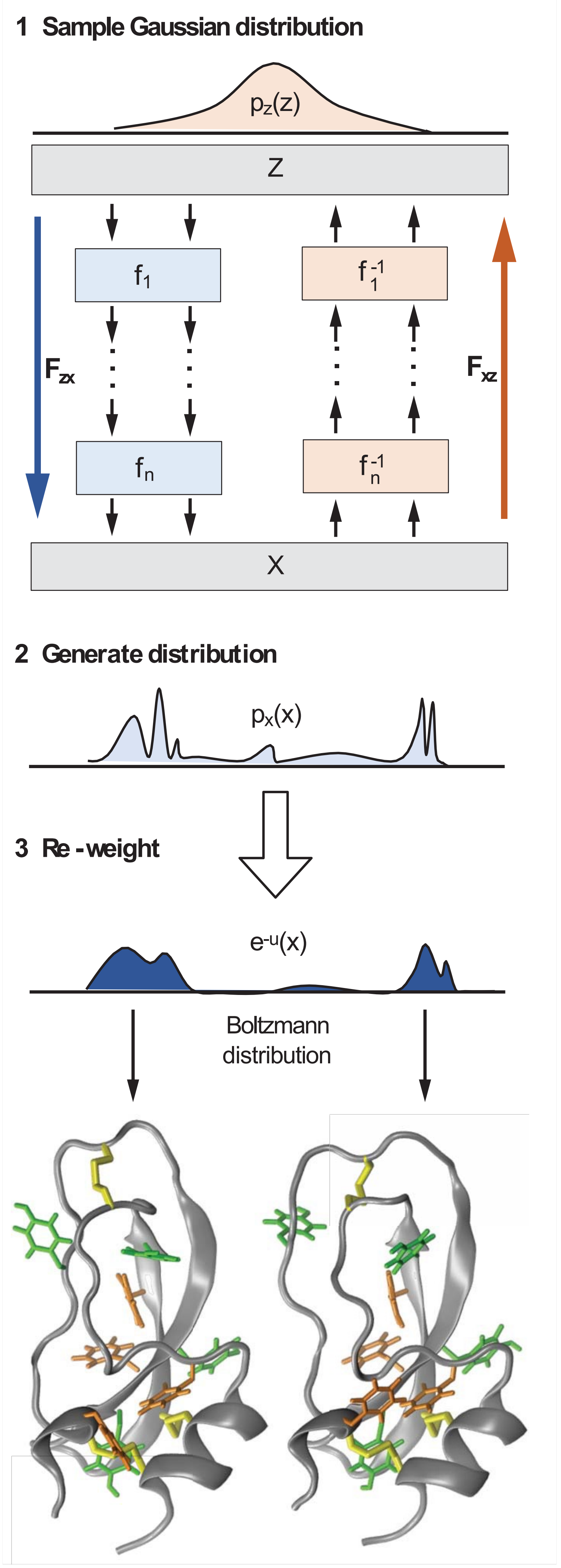
MACHINE LEARNING

Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

Frank Noé*†, Simon Olsson*, Jonas Köhler*, Hao Wu

The main approach is thus to start with one configuration, e.g., the folded protein state, and make tiny changes to it over time, e.g., by using Markov-chain Monte Carlo or molecular dynamics (MD). However, these simulations get trapped in metastable (long-lived) states: For example, sampling a single folding or unfolding event with atomistic MD may take a year on a supercomputer.

Boltzmann generators overcome sampling problems between long-lived states. The

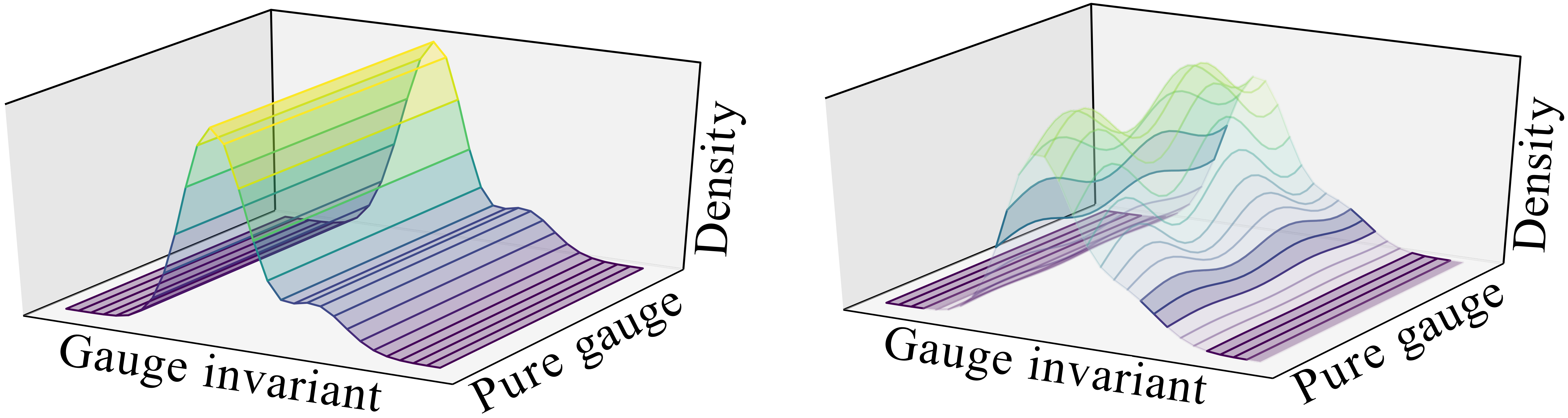


Incorporating symmetries

Kanwar, Albergo, Boyda, Cranmer, Hackett, Racaniere, Rezende, Urban, Shanahan
arXiv:2002.02428, 2003:06413, 2106.05934, 2008.05456

The action is invariant to gauge transformations, so the distribution is constant in those directions

- Architectures that don't bake in symmetry totally fail
- We've developed normalizing flow models that impose symmetry, and it works!



Space-time & Local Gauge Symmetry

Kanwar, Albergo, Boyda, Cranmer, Hackett, Racaniere, Rezende, Shanahan arXiv:2002.02428 & arXiv:2003:06413

Michael Albergo George Papamakarios Danilo Rezende Sébastien Racanière

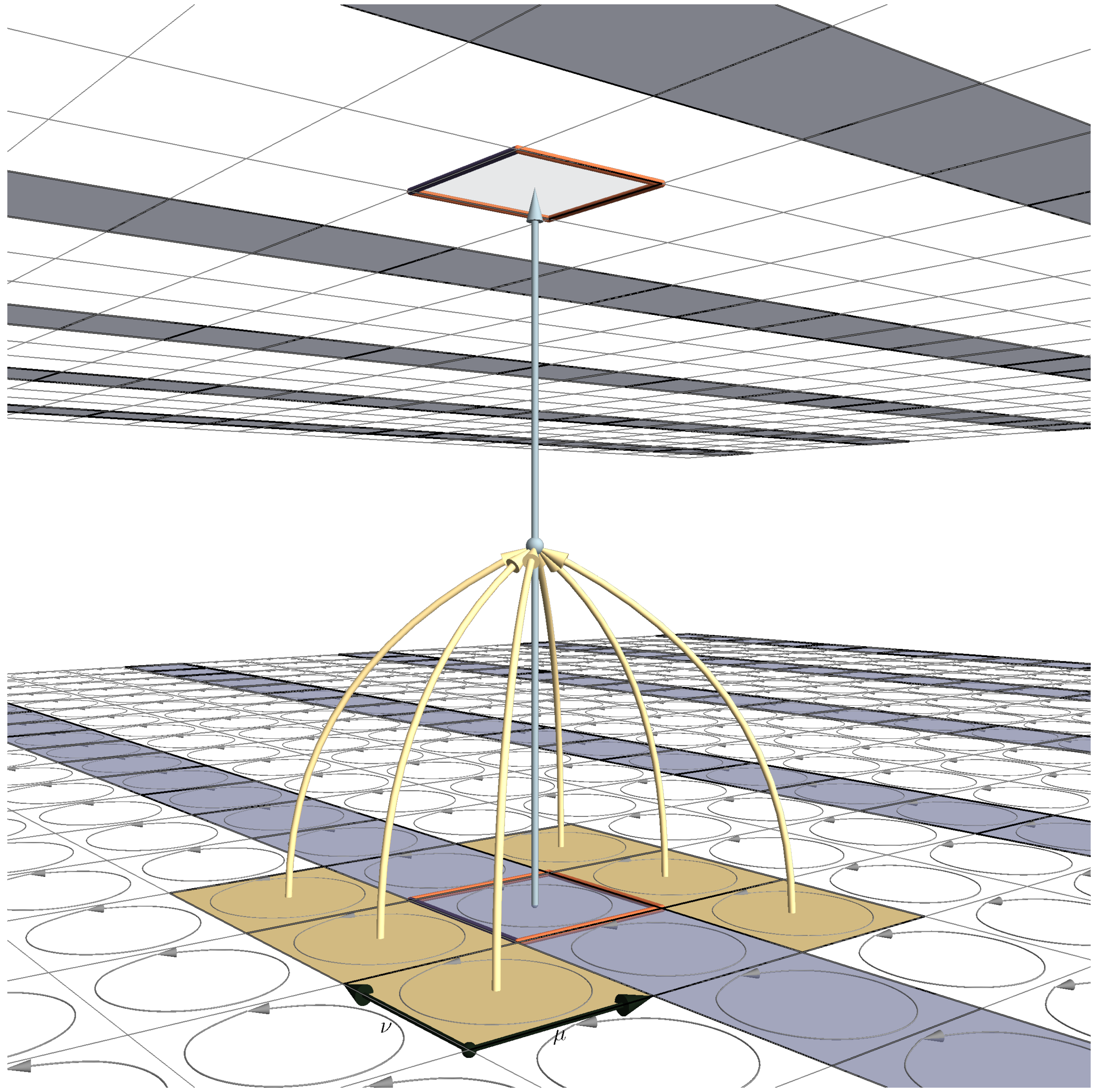
Phiala Shanahan Gurtej Kanwar Dan Hackett Denis Boyda

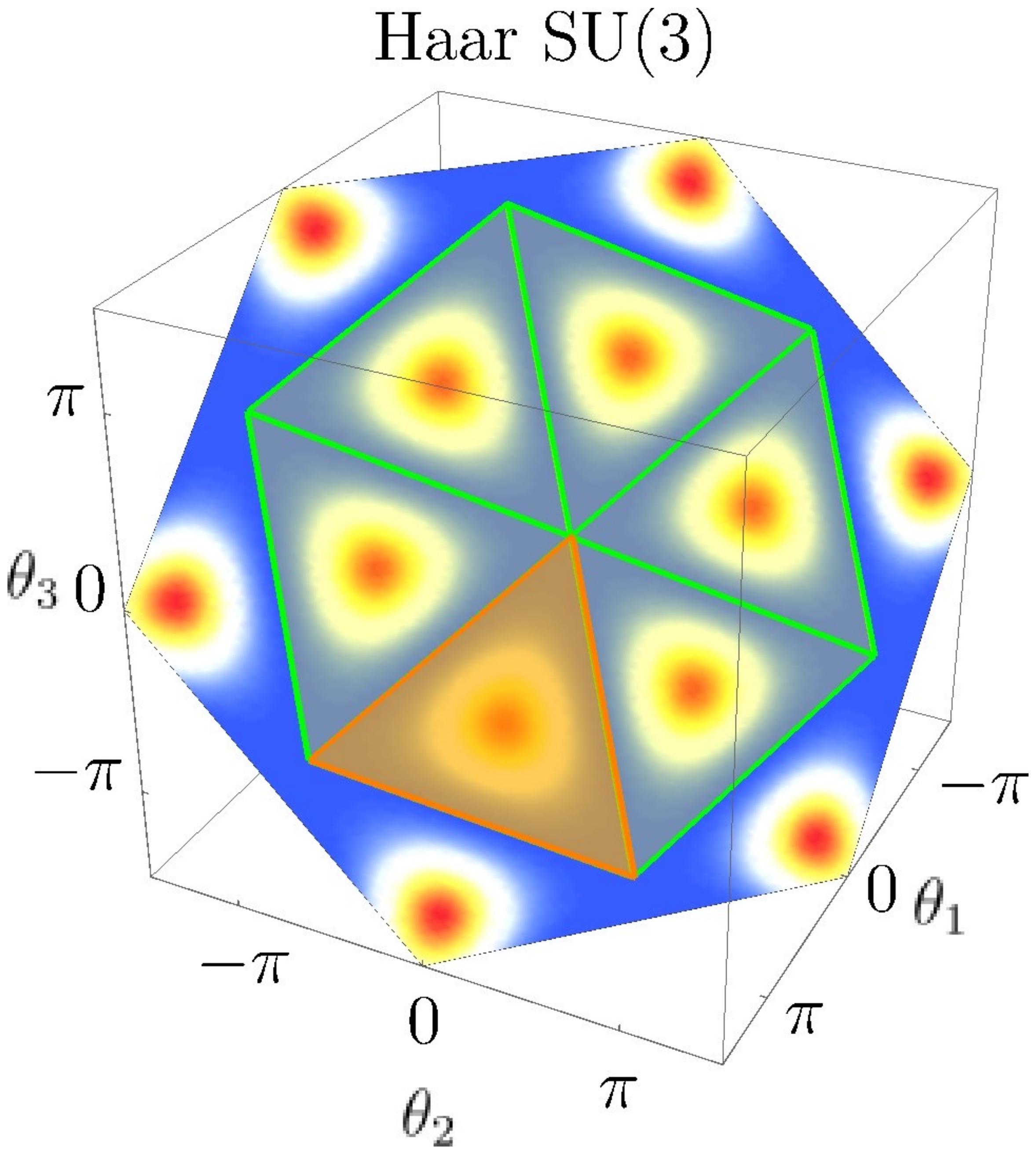
20 May 2020

Building symmetries into generative flow models

Phiala Shanahan, MIT, 12:00 EDT

Abstract: I will discuss recent work to incorporate symmetries, in particular gauge symmetries (local symmetry transformations that form Lie groups), into generative flow models. This work is motivated by the applications of generative models for physics simulation, in particular for lattice field theory.





Flows on Spheres and Tori

[arXiv:2002.02428, ICML2020]

Along the way, we designed flows on compact manifolds like Spheres and Tori that correspond to Lie groups:

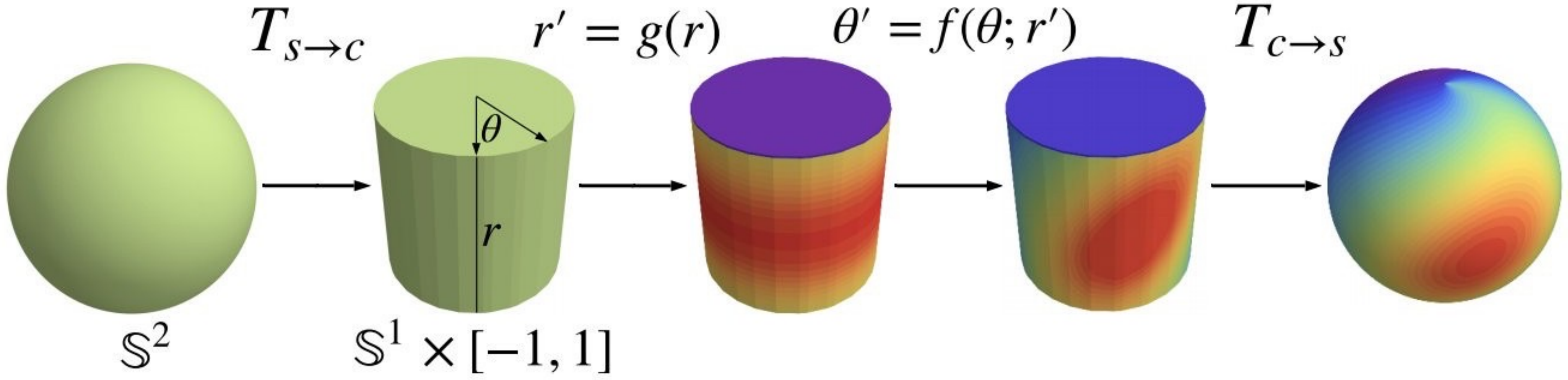
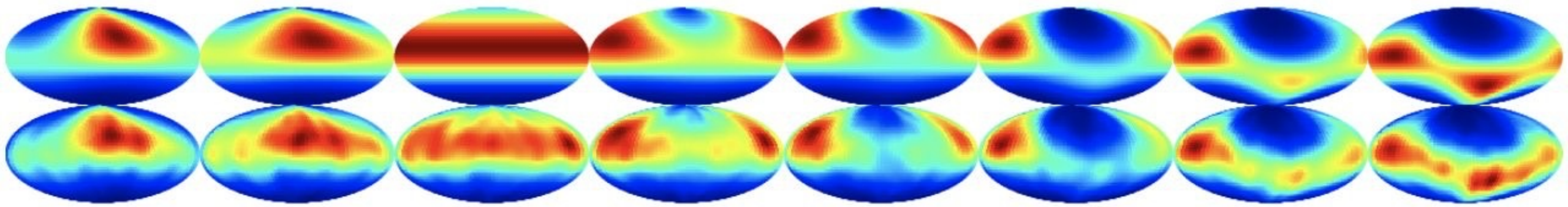


Figure 1. Illustration of the recursive flow on the sphere $\mathbb{S}^2$.



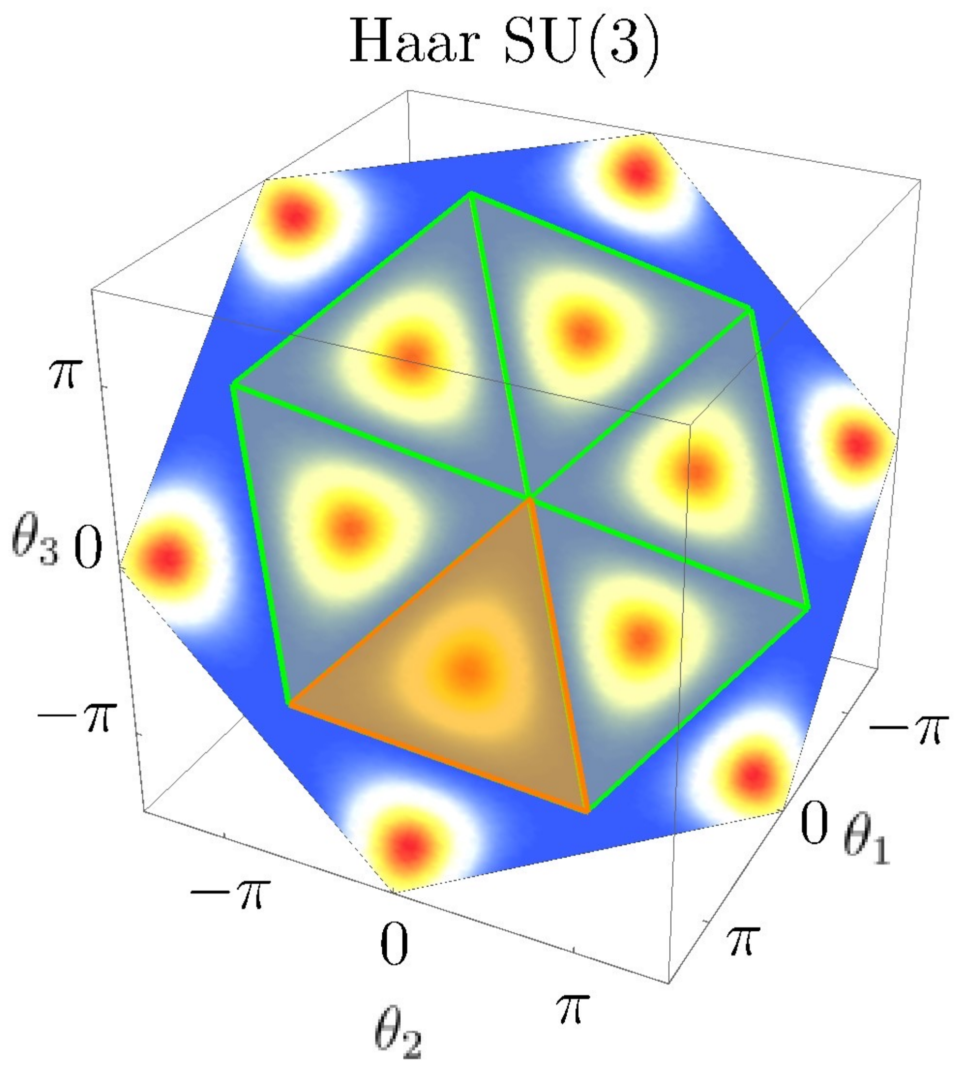


Figure 5. Learned multi-modal density on $SU(2) \equiv \mathbb{S}^3$ using the recursive flow. Each column shows an $\mathbb{S}^2$ slice of the $\mathbb{S}^3$ density

Flows on Spheres and Tori

[arXiv:2002.02428, ICML2020]

Along the way, we designed flows on compact manifolds like Spheres and Tori that correspond to Lie groups:

RELEVANT FOR ROBOTICS WHERE CONFIGURATION SPACE OF ROBOTIC ARM IS AN N-DIM TORUS!

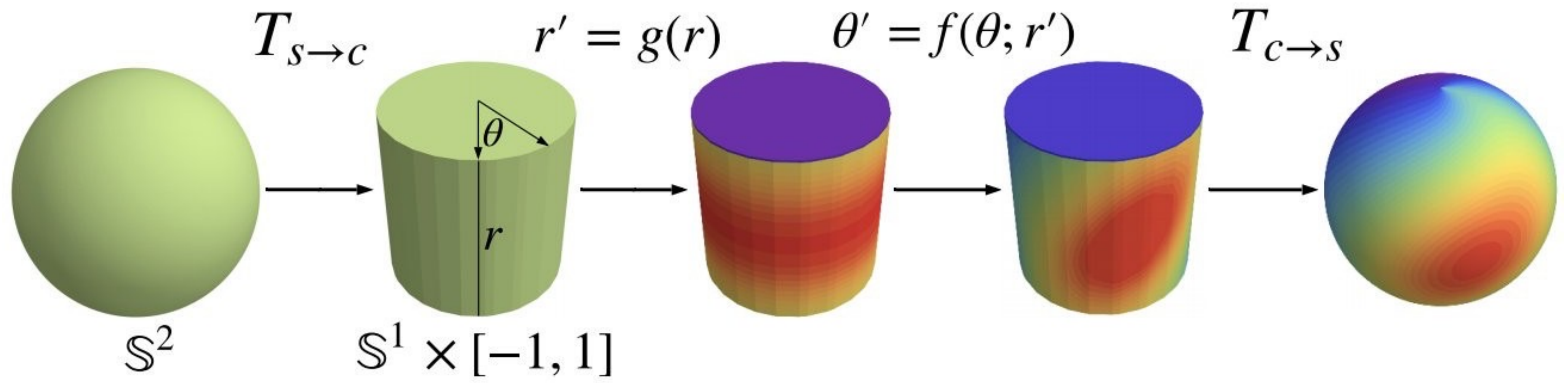


Figure 1. Illustration of the recursive flow on the sphere $\mathbb{S}^2$.

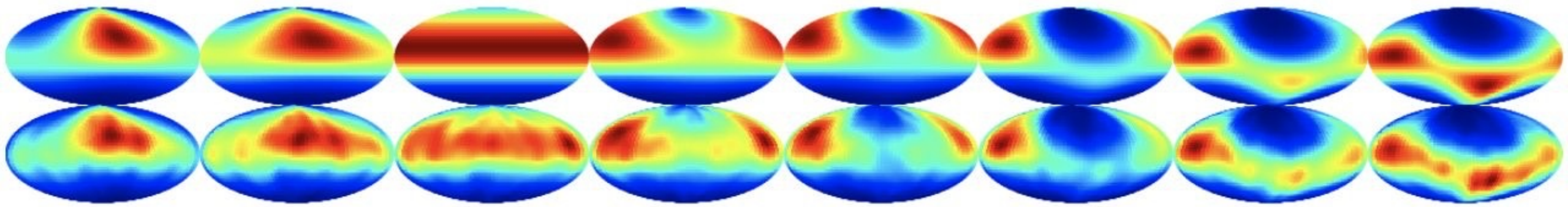
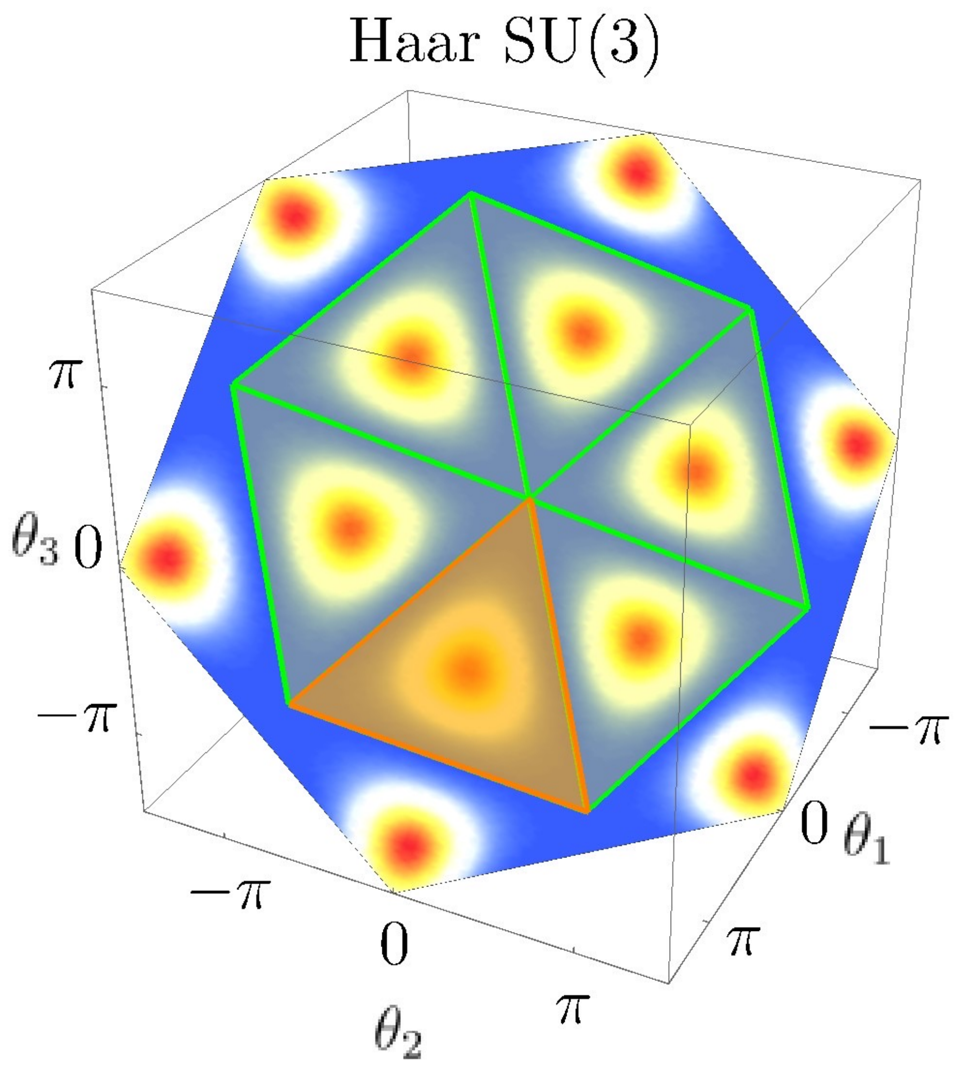


Figure 5. Learned multi-modal density on $SU(2) \equiv \mathbb{S}^3$ using the recursive flow. Each column shows an $\mathbb{S}^2$ slice of the $\mathbb{S}^3$ density



The bitter lesson

– RICH SUTTON

The bitter lesson

Sutton 2019 http://www.incompleteideas.net/IncIdeas/BitterLesson.html

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.

The bitter lesson

Sutton 2019 http://www.incompleteideas.net/IncIdeas/BitterLesson.html

This is a big lesson. As a field, we still have not thoroughly learned it, as we are continuing to make the same kind of mistakes. To see this, and to effectively resist it, we have to understand the appeal of these mistakes. We have to learn **the bitter lesson that building in how we think we think does not work in the long run**.

The bitter lesson is based on the historical observations that

- 1) AI researchers have often tried to build knowledge into their agents,
- 2) this always helps in the short term, and is **personally satisfying** to the researcher, but
- 3) in the long run it plateaus and even inhibits further progress, and
- 4) breakthrough progress eventually arrives by an opposing approach based on scaling computation by *search* and *learning*.

The eventual success is tinged with bitterness, and often incompletely digested, because it is success over a favored, human-centric approach.

The bitter lesson

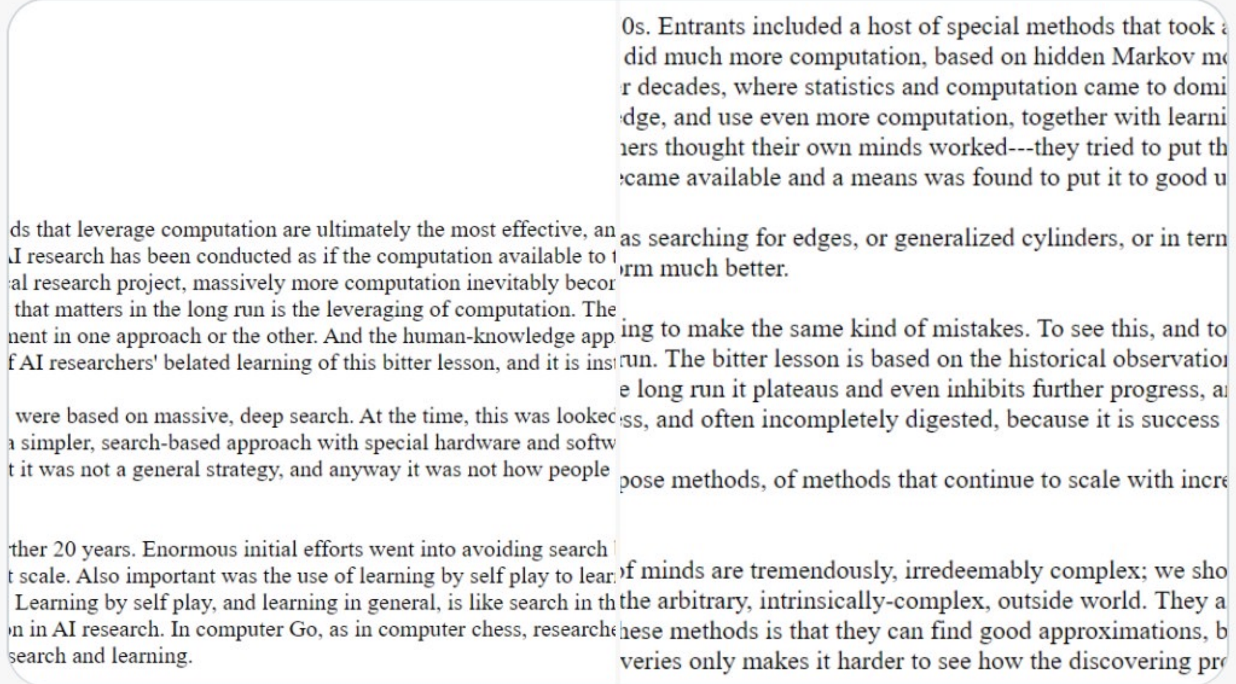
Sutton 2019 http://www.incompleteideas.net/IncIdeas/BitterLesson.html

One thing that should be learned from the bitter lesson is the great power of general purpose methods, of methods that continue to scale with increased computation even as the available computation becomes very great. The two methods that seem to scale arbitrarily in this way are *search* and *learning*.

The second general point to be learned from the bitter lesson is that the actual contents of minds are tremendously, irredeemably complex; **we should stop trying to find simple ways to think about the contents of minds, such as simple ways to think about space, objects, multiple agents, or symmetries.** All these are part of the arbitrary, intrinsically-complex, outside world. They are not what should be built in, as their complexity is endless; instead we should build in only the meta-methods that can find and capture this arbitrary complexity. Essential to these methods is that they can find good approximations, but the search for them should be by our methods, not by us. We want AI agents that can discover like we can, not which contain what we have discovered. **Building in our discoveries only makes it harder to see how the discovering process can be done.**

Physics-inspired work impacting the debate

Michael Nielsen @michael_nielsen · Jul 21, 2021
This is fascinating: Rich Sutton on the "bitter lesson" of AI research:
incompleteideas.net/IncIdeas/Bitte...


0s. Entrants included a host of special methods that took
did much more computation, based on hidden Markov m
r decades, where statistics and computation came to domi
dge, and use even more computation, together with learni
ers thought their own minds worked---they tried to put th
came available and a means was found to put it to good u
ds that leverage computation are ultimately the most effective, an
I research has been conducted as if the computation available to
al research project, massively more computation inevitably beco
that matters in the long run is the leveraging of computation. The
nent in one approach or the other. And the human-knowledge app
f AI researchers' belated learning of this bitter lesson, and it is ins
as searching for edges, or generalized cylinders, or in tern
rm much better.
ing to make the same kind of mistakes. To see this, and to
run. The bitter lesson is based on the historical observatio
e long run it plateaus and even inhibits further progress, a
ss, and often incompletely digested, because it is success
were based on massive, deep search. At the time, this was looke
a simpler, search-based approach with special hardware and softw
t it was not a general strategy, and anyway it was not how people
pose methods, of methods that continue to scale with incr
ther 20 years. Enormous initial efforts went into avoiding search
t scale. Also important was the use of learning by self play to lear
Learning by self play, and learning in general, is like search in th
n in AI research. In computer Go, as in computer chess, research
search and learning.
f minds are tremendously, irredeemably complex; we sho
the arbitrary, intrinsically-complex, outside world. They a
hese methods is that they can find good approximations, b
veries only makes it harder to see how the discovering pr

28 116 583

Kyle Cranmer @KyleCranmer · Jul 21, 2021
We have a few examples of problems (Eg lattice field theory) that are ~hopeless with traditional deep learning, but work when you bake in / enforce symmetries. It seems to take much (exponentially?) more data and compute to learn without that inductive bias. @DaniloJRezende

3 6 54

Danilo J. Rezende
@DaniloJRezende

Replying to @KyleCranmer and @michael_nielsen

Agree!The rapid progress of ML applied to LQCD, mol. dyn., protein folding and computer graphics is the result of the combining domain knowledge (e.g. symmetries) with ML
The "bitter lesson" applies more to domains where domain knowledge is weak or hard to express mathematically.

Danilo J. Rezende @DaniloJRezende · Jun 18, 2020
With all due respect, this conclusion is absolutely preposterous. It comes from asking the wrong question. It is irrelevant to ask "Can we learn X?". The answer is always YES.The important question is "How much data we need? And What kind of data (observational, interventional)?"

> **Jeff Clune** @jeffclune · Jun 18, 2020
> In line with AI-Generating Algorithms, we should stop trying to hand-design the solutions to ML problems. Instead, with the right problems and enough data/compute, ML will do the heavy lifting and solve the problems for us (and do a better job). arxiv.org/abs/1905.10985

15 34 259

Jeff Clune
@jeffclune

Replying to @DaniloJRezende

I am interested to hear that the default assumption is now that we can learn anything if we have enough data/compute. I do not think we would have concluded that a few years back, especially prior to 2012. My point is basically the same as Rich Sutton's The Bitter Lesson. 1/

1:58 PM · Jun 19, 2020 · Twitter Web App

Take away

In addition to the personal joy and benefits to society of both basic and applied research, there is great value in exporting general purpose techniques and insight

- An intellectual form of "technology transfer"
- It behooves the field of physics to do this deliberately
  - May require adopting unfamiliar language, avoiding physics jargon, etc.
  - May require publishing in unfamiliar venues & recognizing value of those publications
  - Effective abstractions may not capture all aspects of the initial physics problem
- Aids in building collaborations with methodological researchers
- Avoids siloing & stagnation
- Facilitates importing good ideas from other fields
- Bolsters the real and perceived value of funding physics research