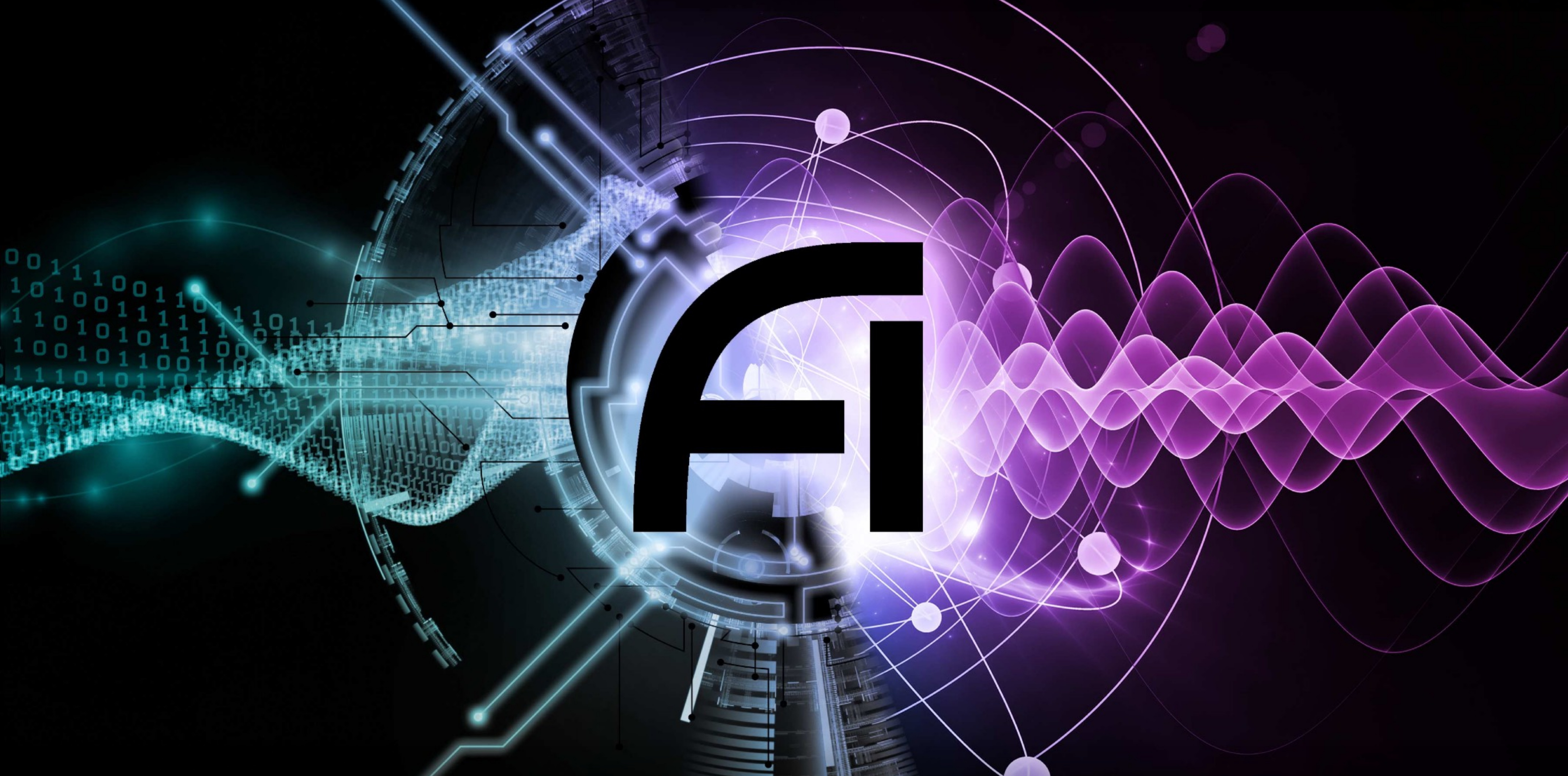
Machine Learning for the Theory Frontier

Jesse Thaler

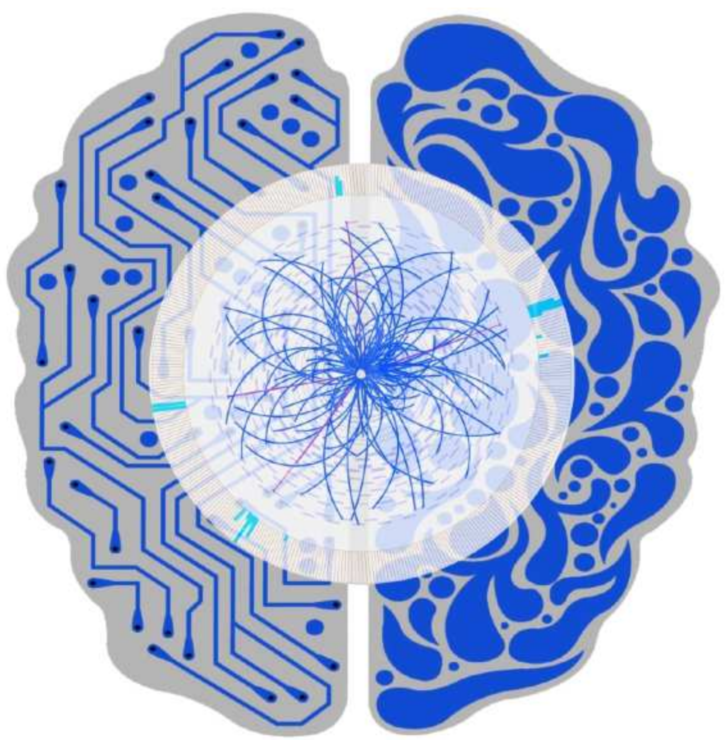
Snowmass Theory Frontier Conference, KITP, Santa Barbara — February 24, 2022

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



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

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



Infuse physics intelligence into artificial intelligence

Symmetries, conservation laws, scaling relations, limiting behaviors, locality, causality, unitarity, gauge invariance, entropy, least action, factorization, unit tests, exactness, systematic uncertainties, reproducibility, verifiability, …

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


The New York Times

By Dennis Overbye

Nov. 23, 2020

Can a Computer Devise a Theory of Everything?

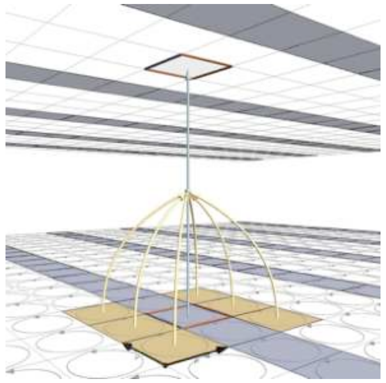
I loathe this question

Overhyping deep learning, which is just one of many computational strategies relevant for the physical sciences

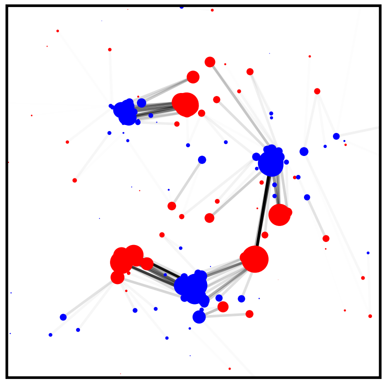
I love this question

Reframes the scientific process and raises questions about what aspects of reductionist logic could be automated

My (Evolving) Perspective



Theoretical high-energy physics has been irreversibly impacted by the rise of deep learning



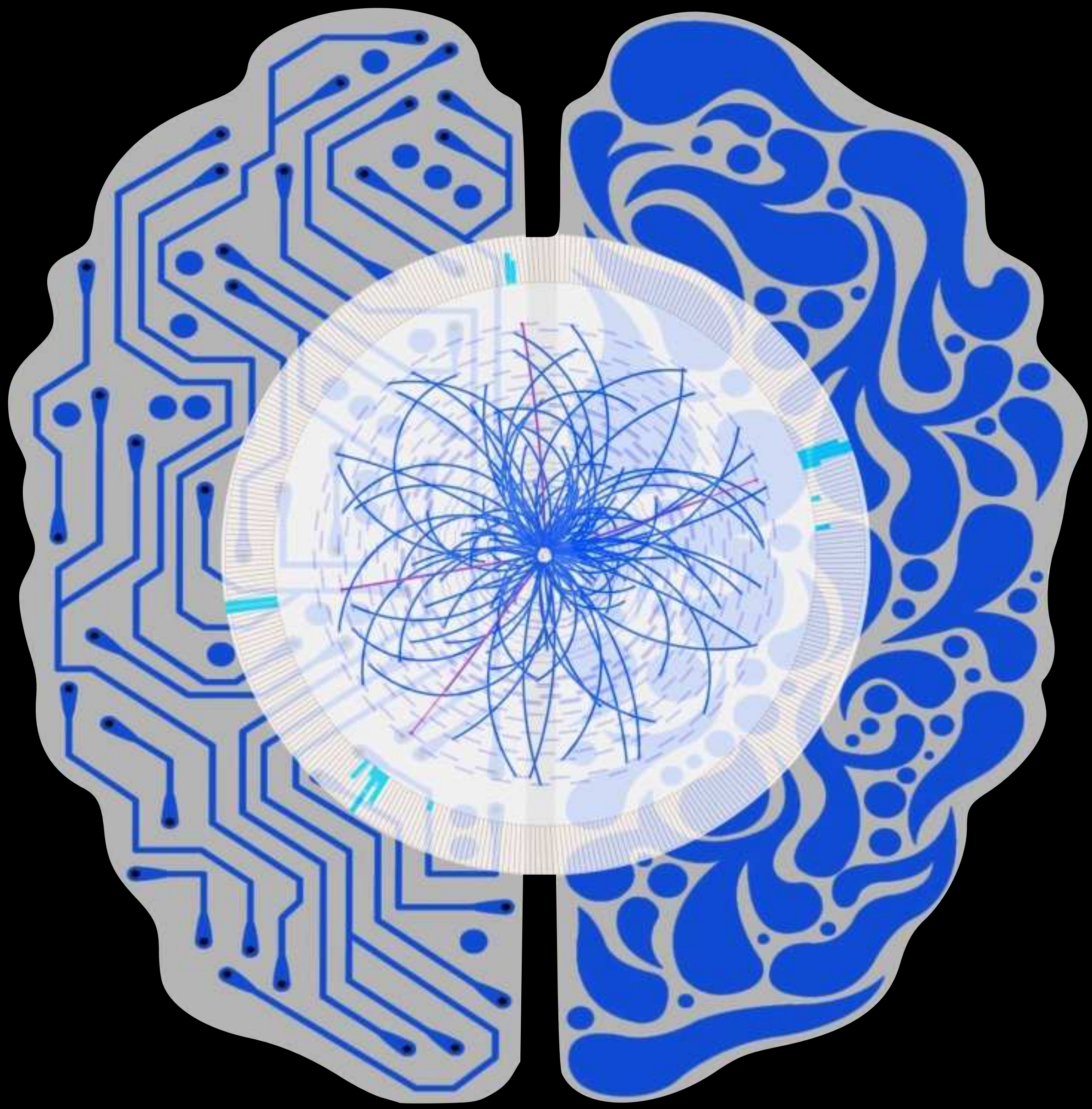
The buzz is around "AI", but we should leverage analysis strategies from various areas of mathematics, statistics, and computer science

We have an opportunity to translate aspects of HEP theory into a computational language

Apologies: citations in this talk are representative, not exhaustive!

[see HEPML-LivingReview for extensive bibliography; see March 2021 KITP talk with Dreyer for related discussion]

The Lens of Machine Learning



What formalisms are needed to leverage ML for HEP theory?

E.g. Likelihood Ratio Trick

Key example of simulation-based inference

Goal: Estimate p(x) / q(x)

Training Data: Finite samples P and Q

Learnable Function: f(x) parametrized by, e.g., neural networks

Loss Function(al): $$L = -\langle \log f(x) \rangle_P + \langle f(x) - 1 \rangle_Q$$

Asymptotically: $$\underset{f(x)}{\arg\min}\, L = \frac{p(x)}{q(x)}$$ *Likelihood ratio*

$$-\min_{f(x)} L = \int \mathrm{d}x\, p(x) \log \frac{p(x)}{q(x)}$$ *Kullback–Leibler divergence*

[see e.g. D'Agnolo, Wulzer, PRD 2019; simulation-based inference in Cranmer, Brehmer, Louppe, PNAS 2020; relation to f-divergences in Nguyen, Wainwright, Jordan, AoS 2009; Nachman, JDT, PRD 2021]

E.g. Likelihood Ratio Trick

Key example of simulation-based inference

Asymptotically, same structure as Lagrangian mechanics!

Action: $$L = \int \mathrm{d}x\, \mathcal{L}(x)$$

Lagrangian: $$\mathcal{L}(x) = -p(x) \log f(x) + q(x)\big(f(x) - 1\big)$$

Euler-Lagrange: $$\frac{\partial \mathcal{L}}{\partial f} = 0$$

Solution: $$f(x) = \frac{p(x)}{q(x)}$$

Requires shift in theoretical focus from solving problems to specifying problems

[see e.g. D'Agnolo, Wulzer, PRD 2019; simulation-based inference in Cranmer, Brehmer, Louppe, PNAS 2020; relation to f-divergences in Nguyen, Wainwright, Jordan, AoS 2009; Nachman, JDT, PRD 2021]

Machine Learning Ingredients

Many HEP theory tasks can phrased as ML optimization

Well-Specified Loss

E.g. classification, regression, generation, …
With implicit or explicit regularization

Reliable Training Data

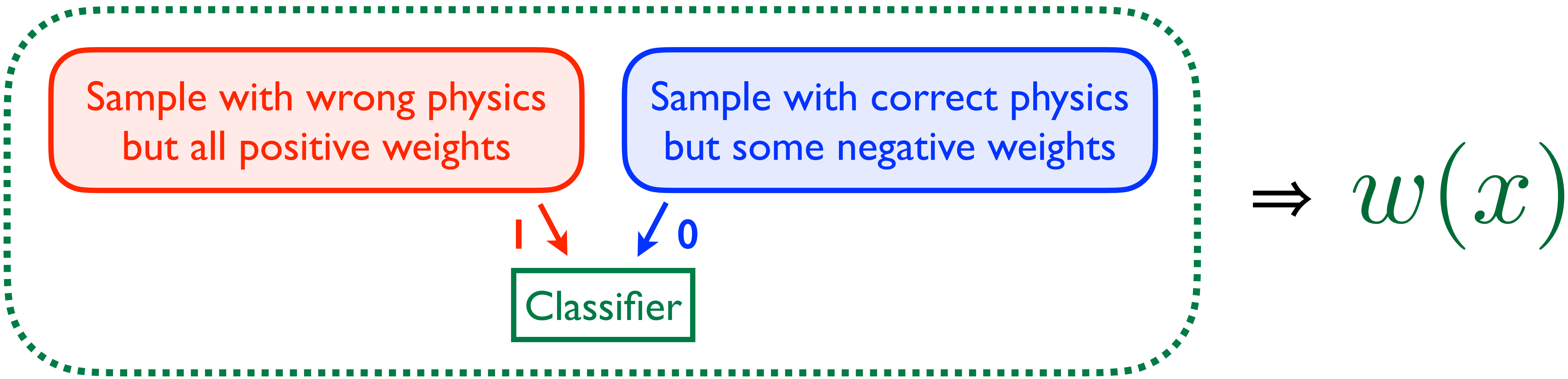
Real or synthetic, fixed or dynamic
Labeled, partially labeled, or unlabeled

Learnable Function

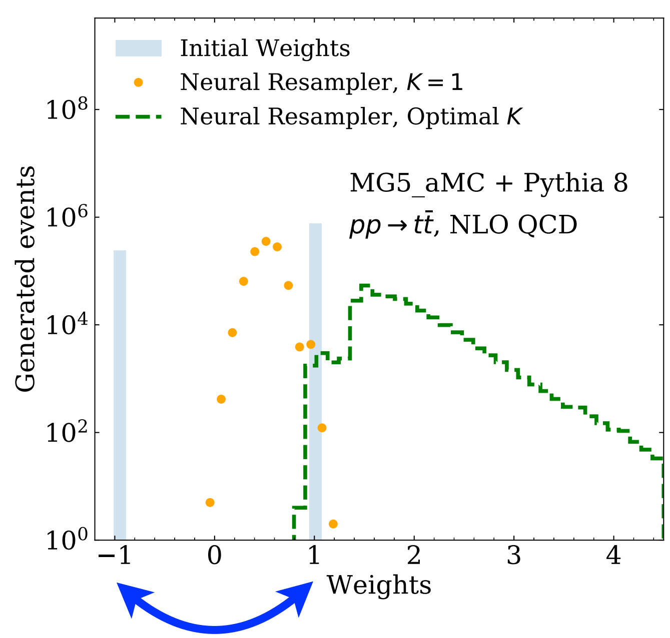
Linear/logit function, neural network,
normalizing flow, other parametrized form, …

Physics input essential for robust usage of these tools,
but interdisciplinary training also valuable

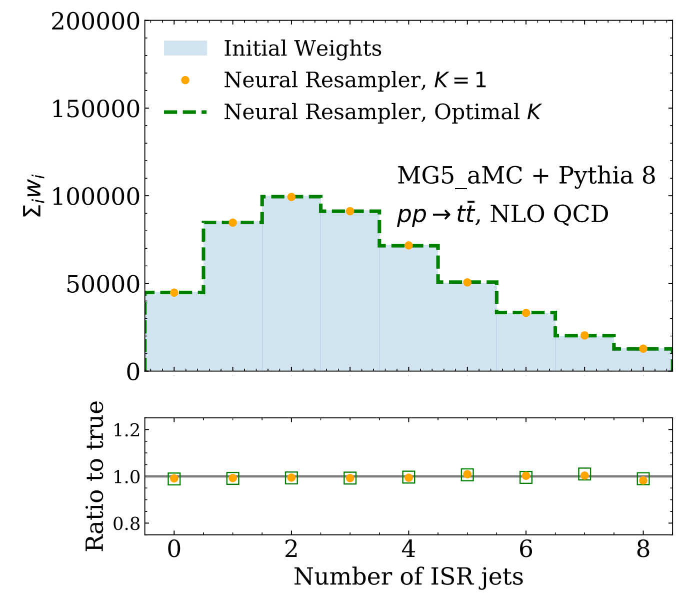
E.g. Neural Resampling for MC Beyond LO



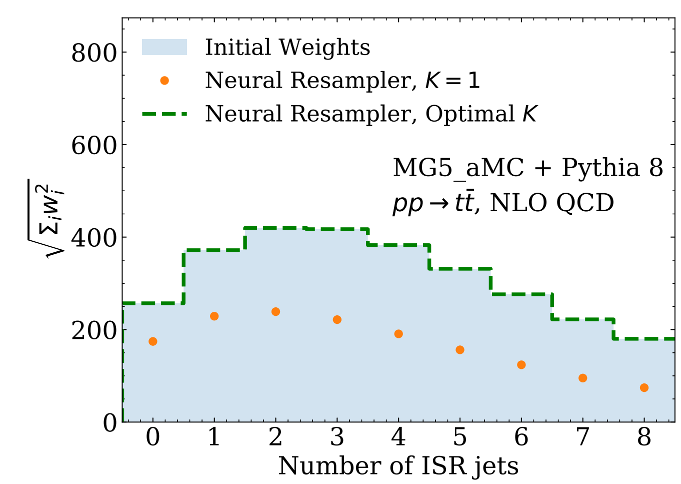
MC@NLO: large weight cancellations



Reweighting recovers desired distribution



Resampling recovers desired uncertainties

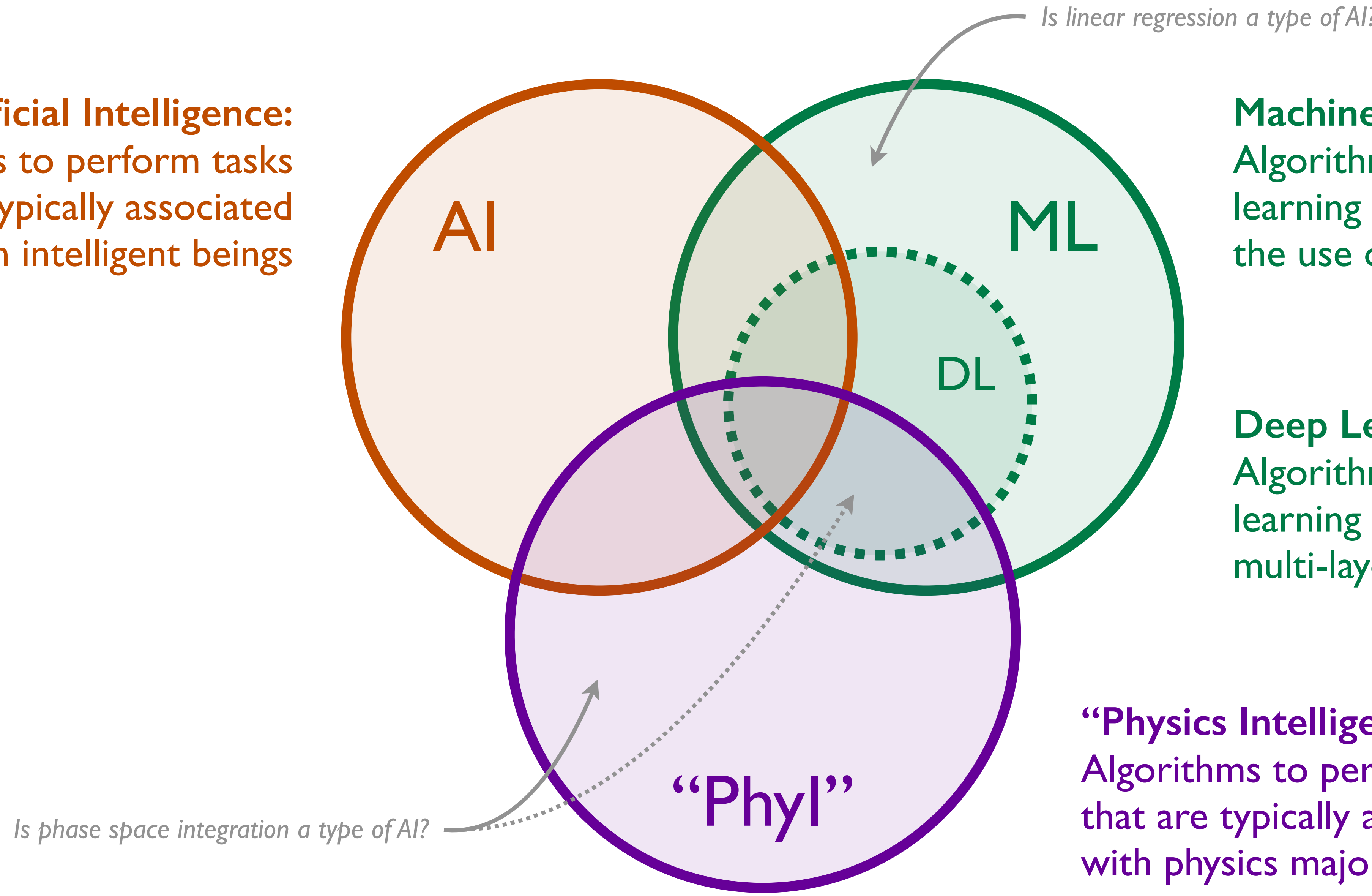


Using custom ML strategy

[Nachman, JDT, PRD 2020; inspired by Andersen, Gutschow, Maier, Prestel, EPJC 2020]

“What is the Machine Learning?”

Artificial Intelligence:
Algorithms to perform tasks that are typically associated with intelligent beings

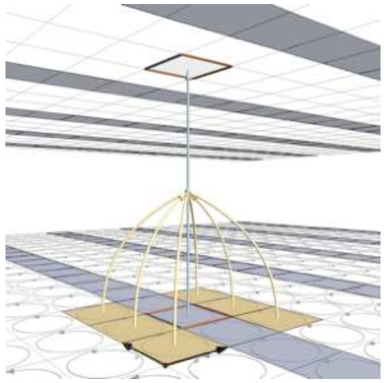


Machine Learning:
Algorithms based on learning solutions through the use of data

Deep Learning:
Algorithms based on learning parameters of multi-layer neural networks

“Physics Intelligence”:
Algorithms to perform tasks that are typically associated with physics majors/PhDs

In most cases, the machine is learning an approximate solution to a well-specified optimization problem



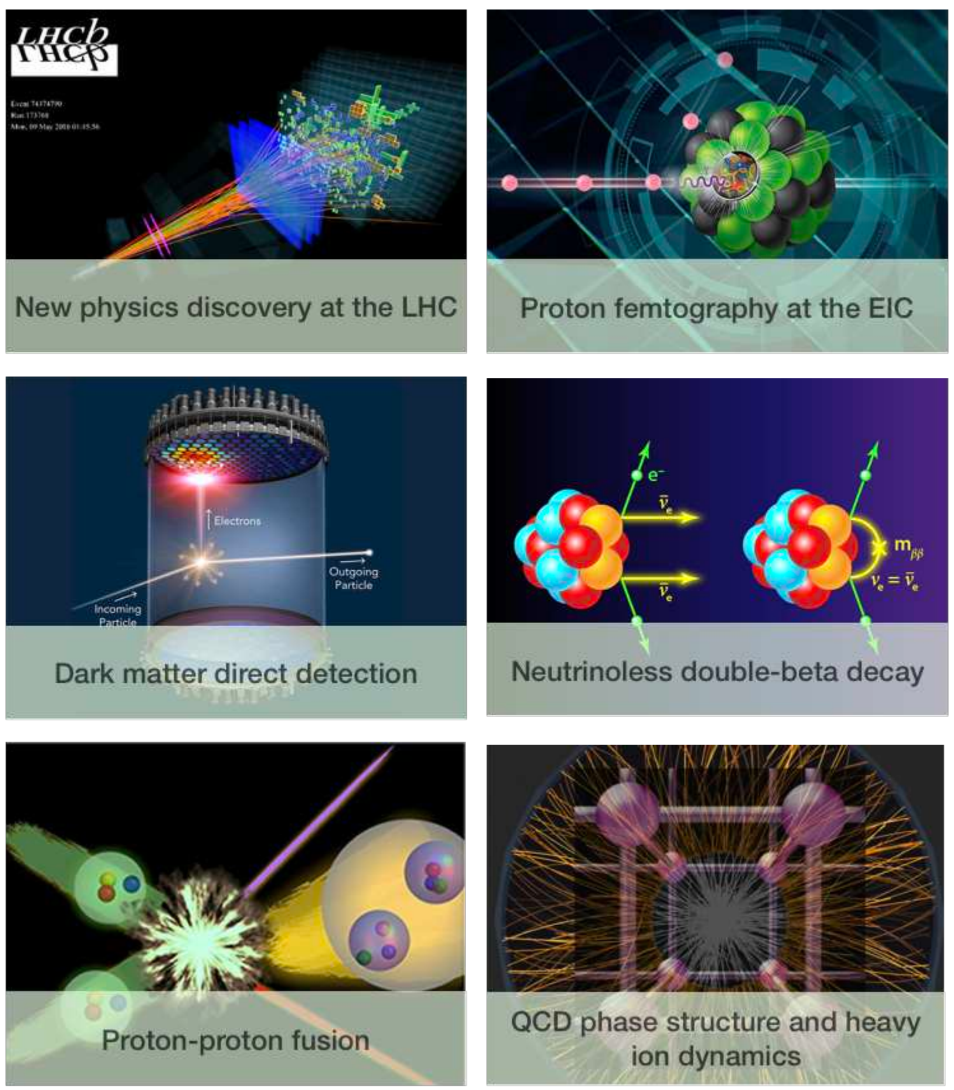
Theoretical high-energy physics has been irreversibly impacted by the rise of deep learning

The buzz is around "AI", but we should leverage analysis strategies from various areas of mathematics, statistics, and computer science

We have an opportunity to translate aspects of HEP theory into a computational language

Deep Learning for Lattice Field Theory (TF05)

Equations governing the strong nuclear force are known, but precision computations are extremely demanding (>10% of open supercomputing in US)



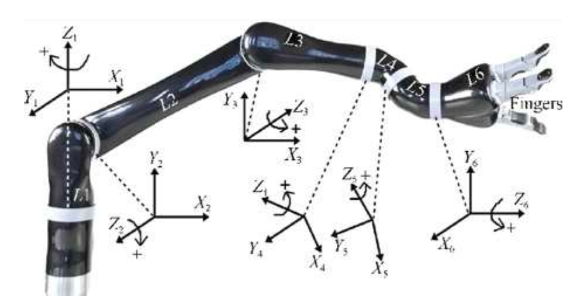
Industry collaboration to develop custom AI tools

Google DeepMind Massachusetts Institute of Technology

Custom generative models based on normalizing flows achieve 1000-fold acceleration while preserving symmetries & guaranteeing exactness

Tools designed for physics find interdisciplinary applications

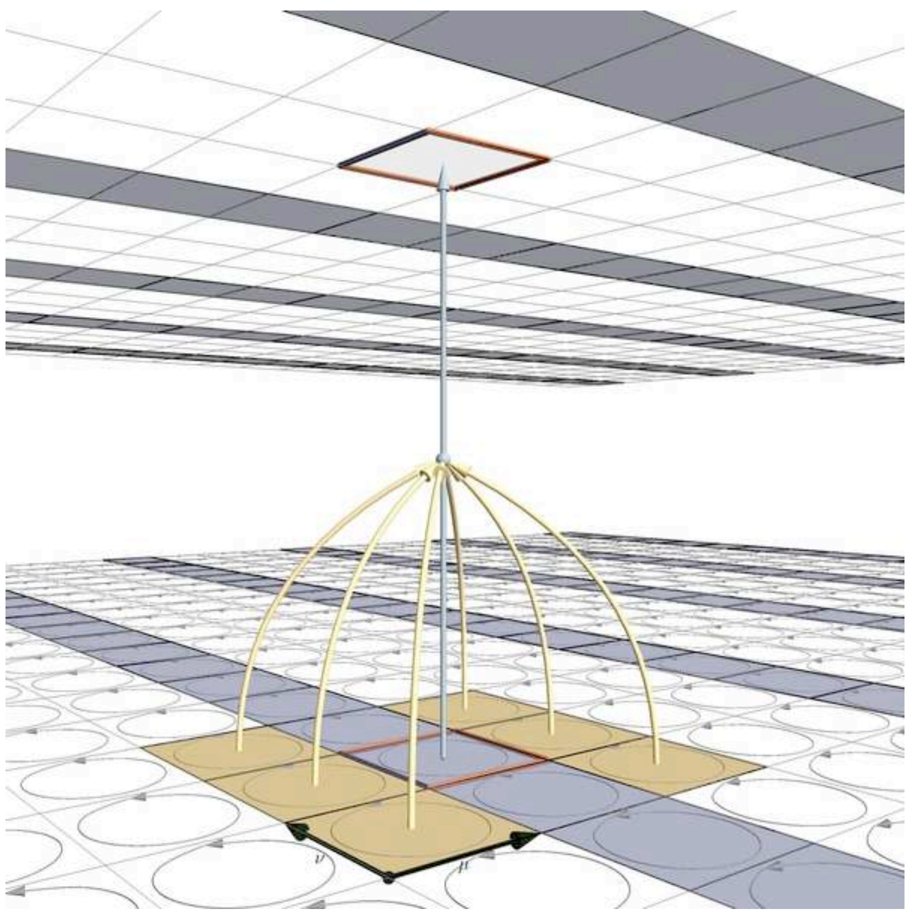
Robotics



[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan, PRL 2020]

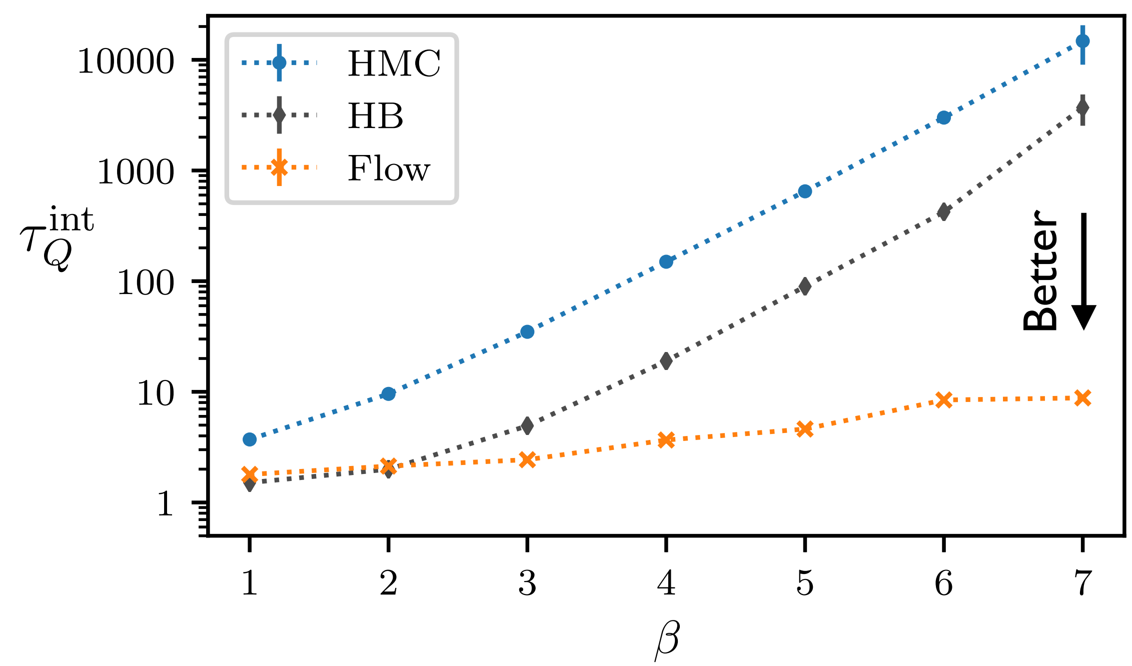
Deep Learning for Lattice Field Theory (TF05)

Equations governing the strong nuclear force are known, but precision computations are extremely demanding (>10% of open supercomputing in US)



Deep Learning ⇔ "Deep Thinking"

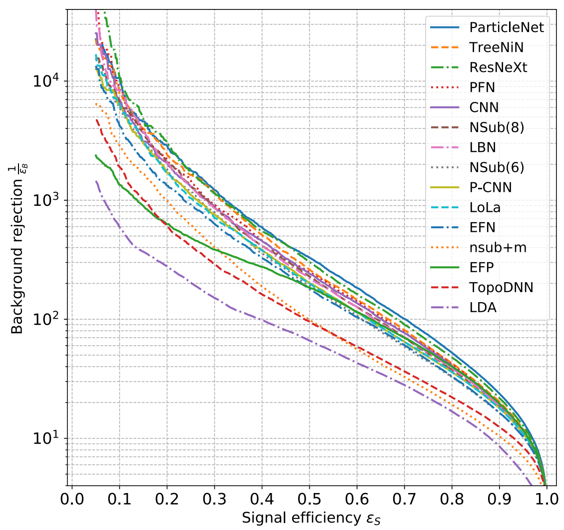
Normalizing Flows   Compact Domains   Gauge Equivariance



[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan, PRL 2020]

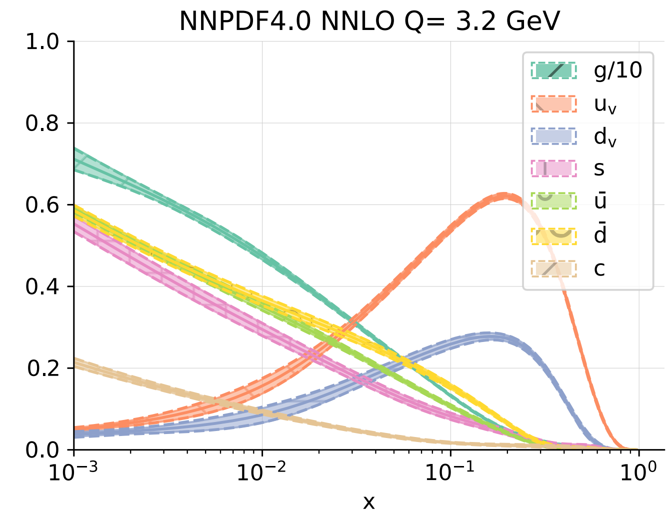
Deep Learning for Colliders (TF07)

Jet Classification



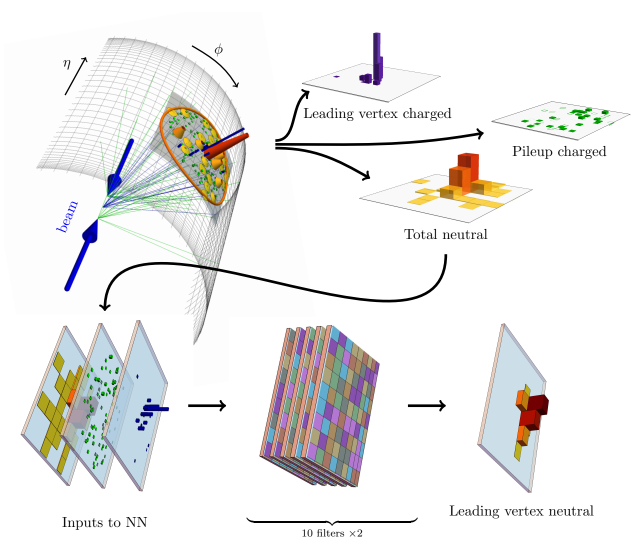
[e.g. Kasieczka, Plehn, et al., SciPost 2019]

Parton Distribution Functions



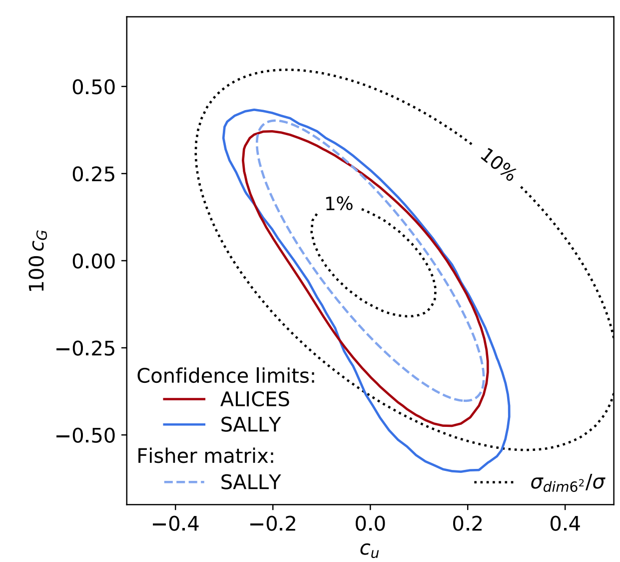
[e.g. NNPDF Collaboration, JHEP 2022]

Pileup Mitigation



[e.g. Komiske, Metodiev, Nachman, Schwartz, JHEP 2017]

Parameter Inference

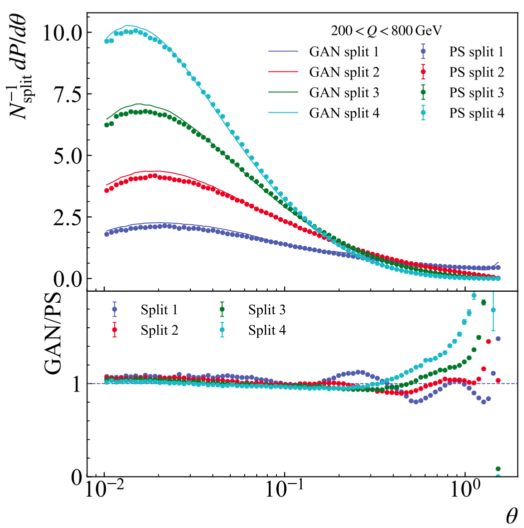


[e.g. Brehmer, Kling, Espejo, Cranmer, CSBS 2020]

. . .

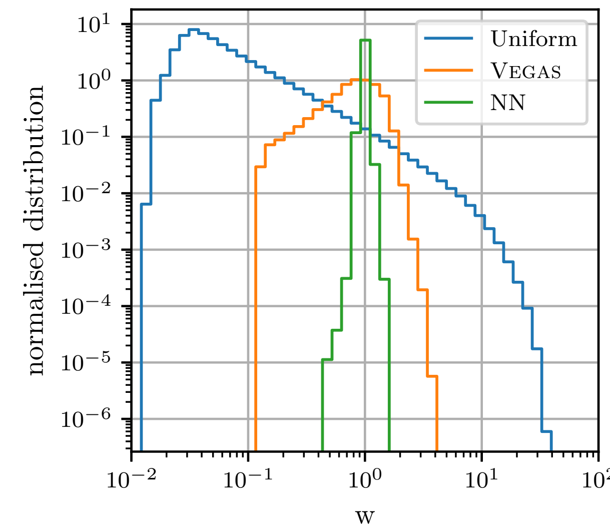
More Deep Learning for Colliders (TF07)

Parton Shower Modeling/Tuning



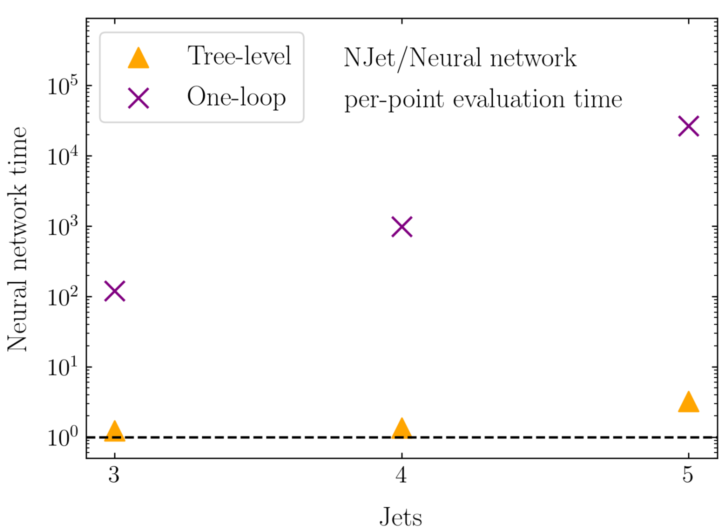
[e.g. Lai, Neill, Płoskoń, Ringer, arXiv 2020;
see also Andreassen, Feige, Frye, Schwartz, EPJC 2019]

Phase Space Integration



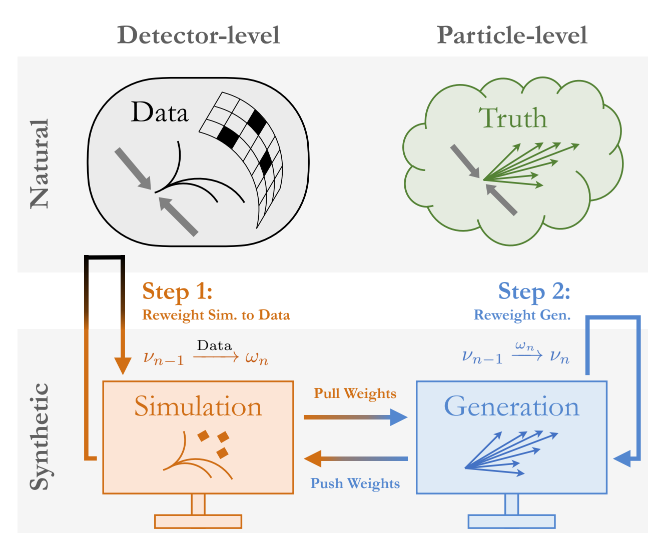
[e.g. Bothmann, Janßen, Knobbe, Schmale, Schumann, SciPost 2020;
see also Gao, Höche, Isaacson, Krause, Schulz, PRD 2020]

Amplitude Approximations



[e.g. Badger, Bullock, JHEP 2020]

Deconvolution/Unfolding



. . .

[e.g. Andreassen, Komiske, Metodiev, Nachman, JDT, PRL 2020;
see also Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Köthe, SciPost 2020]

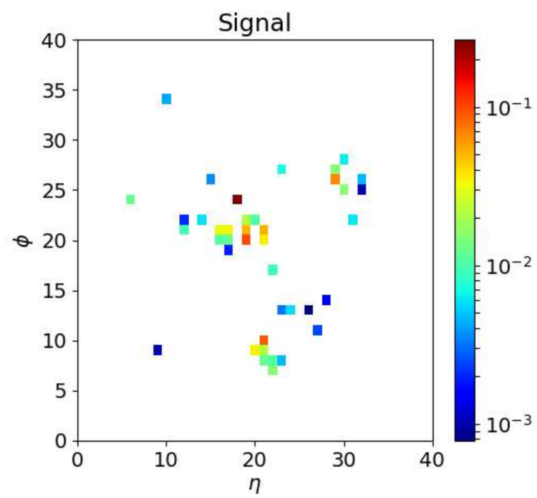
Progress made not just because of
increased computational power and large datasets…

…but also because we have understood the
structure of the underlying problems

(and the structure of HEP theory problems are often optimization tasks,
cf. yesterday's talks by Silverstein & Simmons-Duffin and discussion after Cordova & Bah)

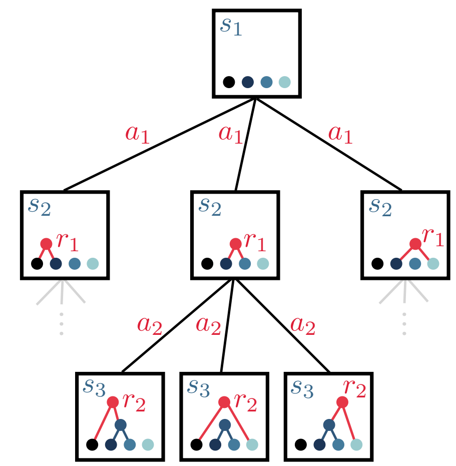
Theoretical Priors ⇒ Network Architectures

Pixelized Calorimetry



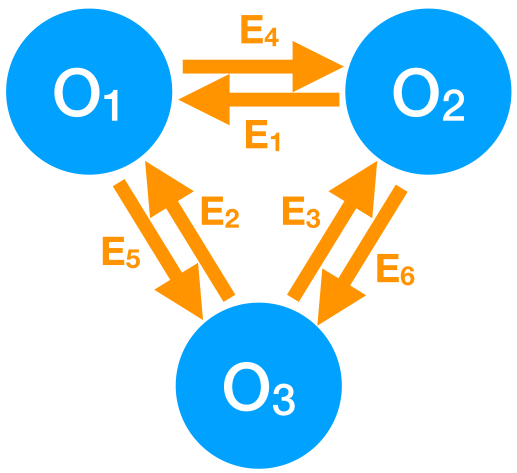
[review in Kagan, arXiv 2020]

Hierarchical Showers



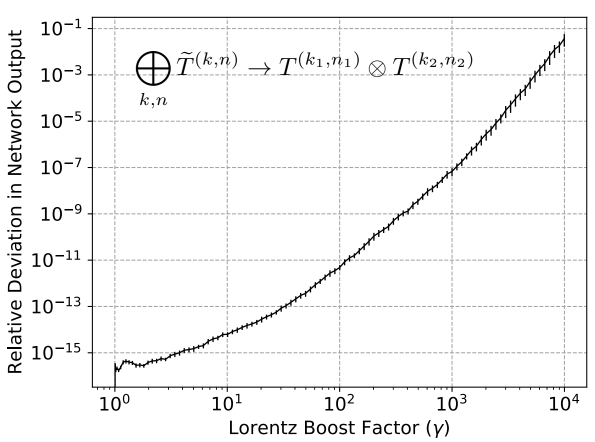
[e.g. Brehmer, Macaluso, Pappadopulo, Cranmer, NeurIPS 2020]

Pairwise Interactions



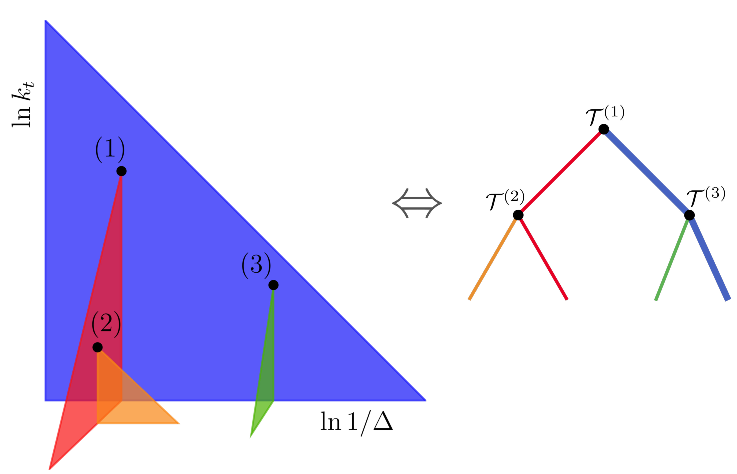
[e.g. Moreno, Cerri, Duarte, Newman, Nguyen, Periwal, Pierini, Serikova, Spiropulu, Vlimant, EPJC 2020]

Lorentz Equivariance



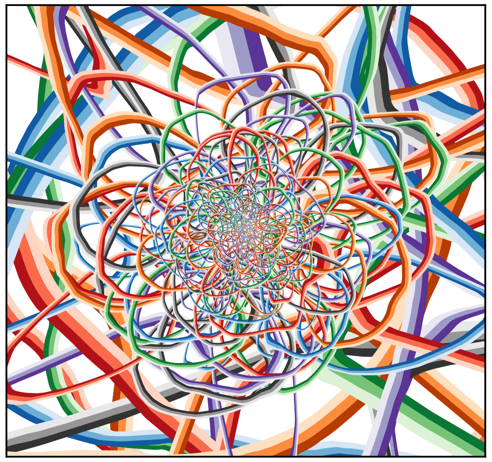
[e.g. Bogatskiy, Anderson, Offermann, Roussi, Miller, Kondor, arXiv 2020]

Lund Plane Emissions



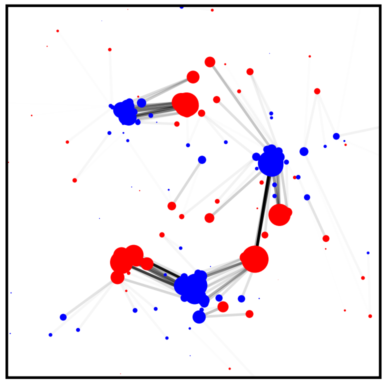
[e.g. Dreyer, Qu, JHEP 2021]

Infrared and Collinear Safety



[e.g. Komiske, Metodiev, JDT, JHEP 2019; see also Dolan, Ore, PRD 2021; Konar, Ngairangbam, Spannowsky, JHEP 2022]

Theoretical high-energy physics has been irreversibly impacted by the rise of deep learning

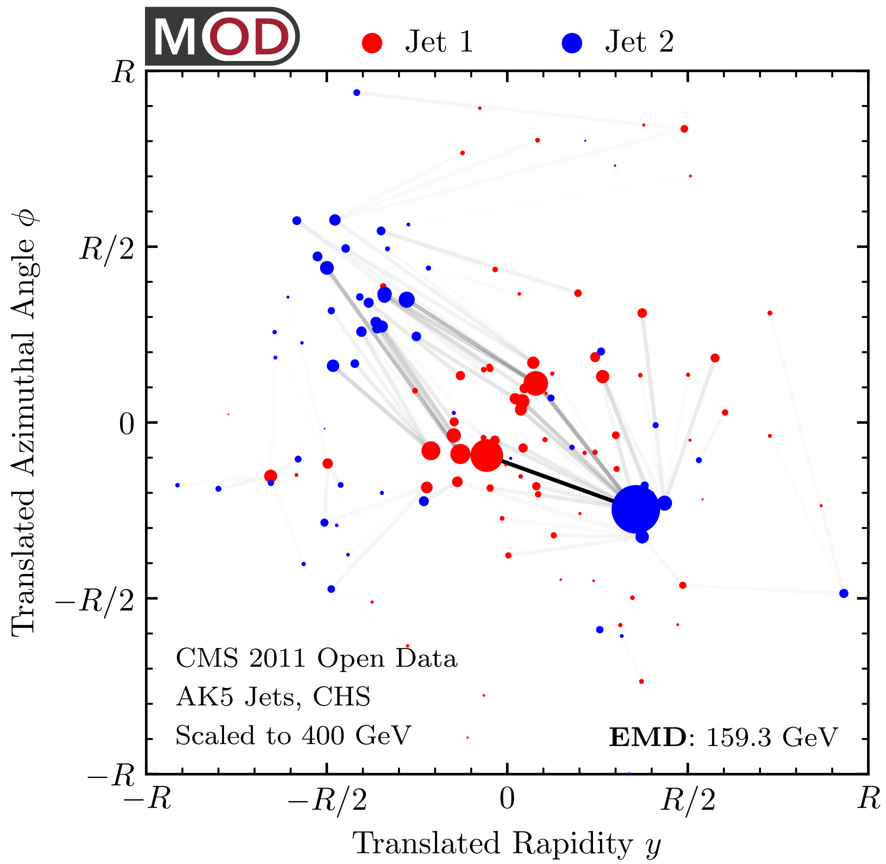


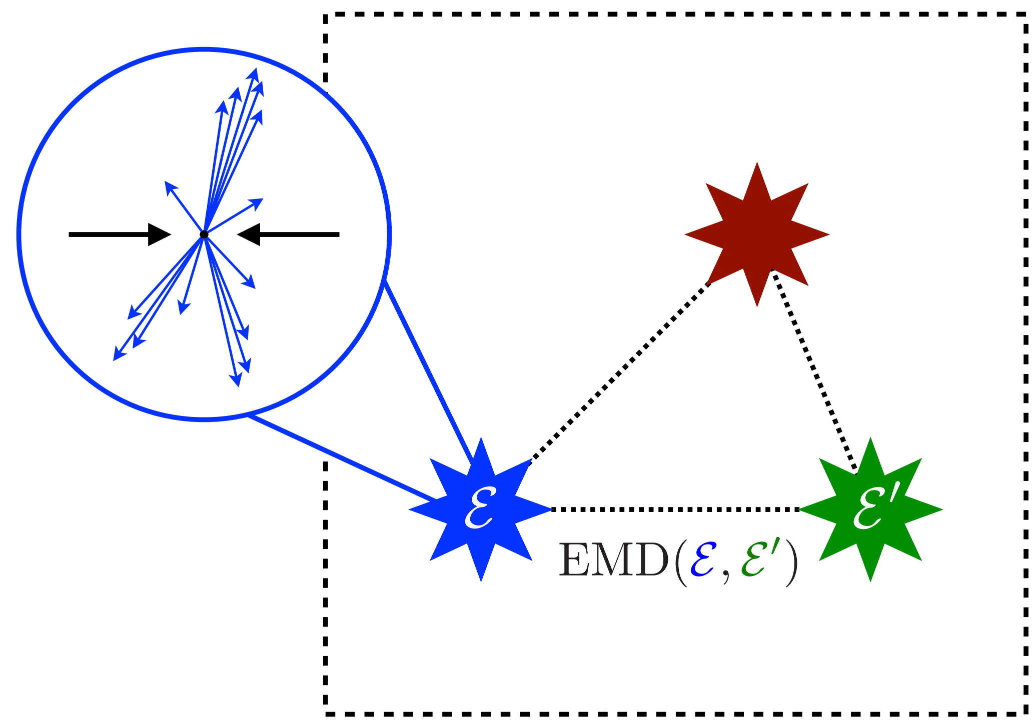
The buzz is around "AI", but we should leverage analysis strategies from various areas of mathematics, statistics, and computer science

We have an opportunity to translate aspects of HEP theory into a computational language

Optimal Transport for Collider Geometry

Energy Mover's Distance ⇒ Metric Space of Collider Events





New insights into high-energy physics facilitated by advances in mathematics, statistics, and computer science

[Komiske, Metodiev, JDT, PRL 2019; code at Komiske, Metodiev, JDT, energyflow.network; open data study in Komiske, Mastandrea, Metodiev, Naik, JDT, PRD 2020]
[based on Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele Taskar, GSI 2013]
[flavored variant in Crispim Romão, Castro, Milhano, Pedro, Vale, EPJC 2021]

Optimal Transport for Collider Geometry

Translating Six Decades of Collider Physics

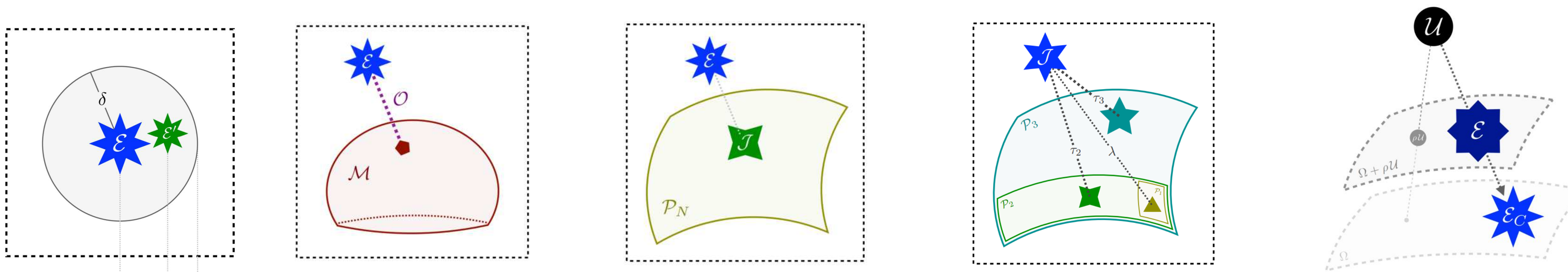
IRC Safety is smoothness in the space of events

Event shapes are distances from events to manifolds.

Jets are projections to few-particle manifolds.

Substructure resolves emissions within the jet.

Pileup mitigation moves away from uniform radiation.



$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$

$\mathcal{J} = \operatorname{argmin}_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$

$\tau(\mathcal{J}) = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_{\beta}(\mathcal{J},\mathcal{E}')$.

$\mathcal{E}_C = \arg\min_{\mathcal{E}'} \text{EMD}(\mathcal{E},\mathcal{E}' + \rho\,\mathcal{U})$.

Taming infinities | Event Shapes | Jet Algorithms | Jet Substructure | Pileup

1960 — 2020

1962-1964
Infrared Safety
[Kinoshita, JMP 1962]
[Lee, Nauenberg, PR 1964]

1977
Thrust, Sphericity
[Farhi, PRL 1977]
[Georgi, Machacek, PRL 1977]

1993
k_T jet clustering
[Ellis, Soper, PRD 1993]
[Catani, Dokshitzer, Seymour, Webber, NPB 1993]

1997-1998
C/A jet clustering
[Wobisch, Wengler, 1998]
[Doskhitzer, Leder, Moretti, Webber, JHEP 1997]

2010-2015
N-(sub)jettiness, XCone
[Stewart, Tackmann, Waalewijn, PRL 2010]
[Thaler, Van Tilburg, JHEP 2011]
[Stewart, Tackmann, Thaler, Vermilion, Wilkason, JHEP 2015]

2014-2019
Constituent Subtraction
[Berta, Spousta, Miller, Leitner, JHEP 2014]
[Berta, Masetti, Miller, Spousta, JHEP 2019]

And many more!

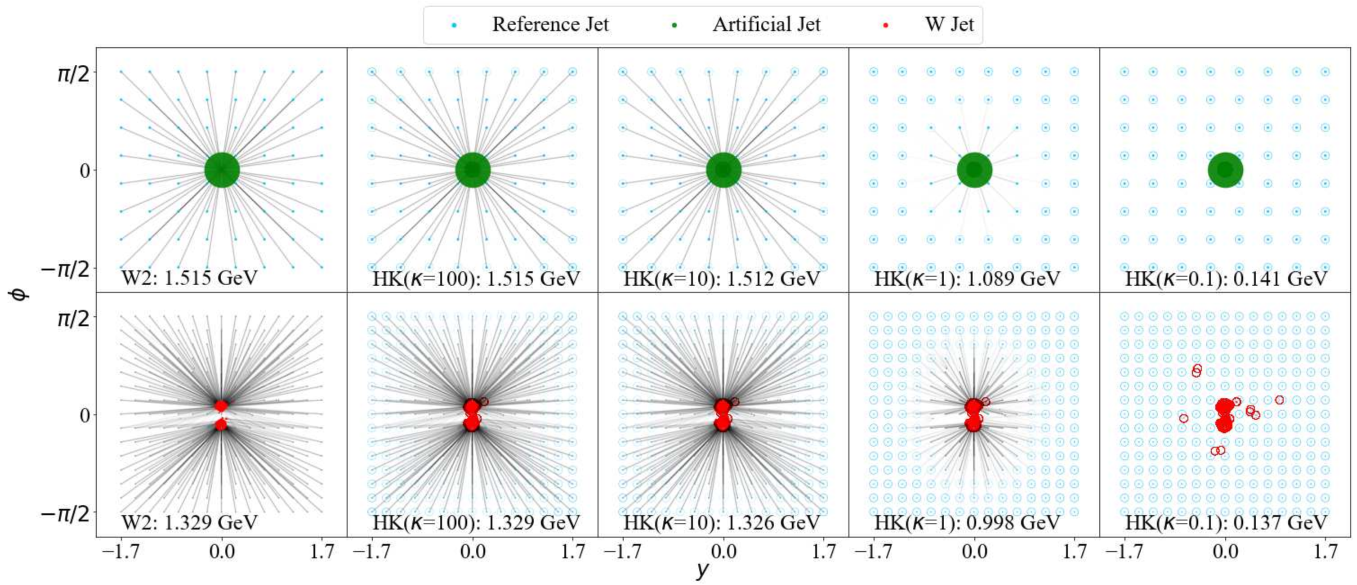
[Komiske, Metodiev, JDT, JHEP 2020; timeline by Metodiev]

[Komiske, Metodiev, JDT, PRL 2019; code at Komiske, Metodiev, JDT, energyflow.network; open data study in Komiske, Mastandrea, Metodiev, Naik, JDT, PRD 2020]
[based on Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele Taskar, GSI 2013]
[flavored variant in Crispim Romão, Castro, Milhano, Pedro, Vale, EPJC 2021]

Opening a Dialogue Between Communities

HEP domain knowledge ⇔ interdisciplinary insights

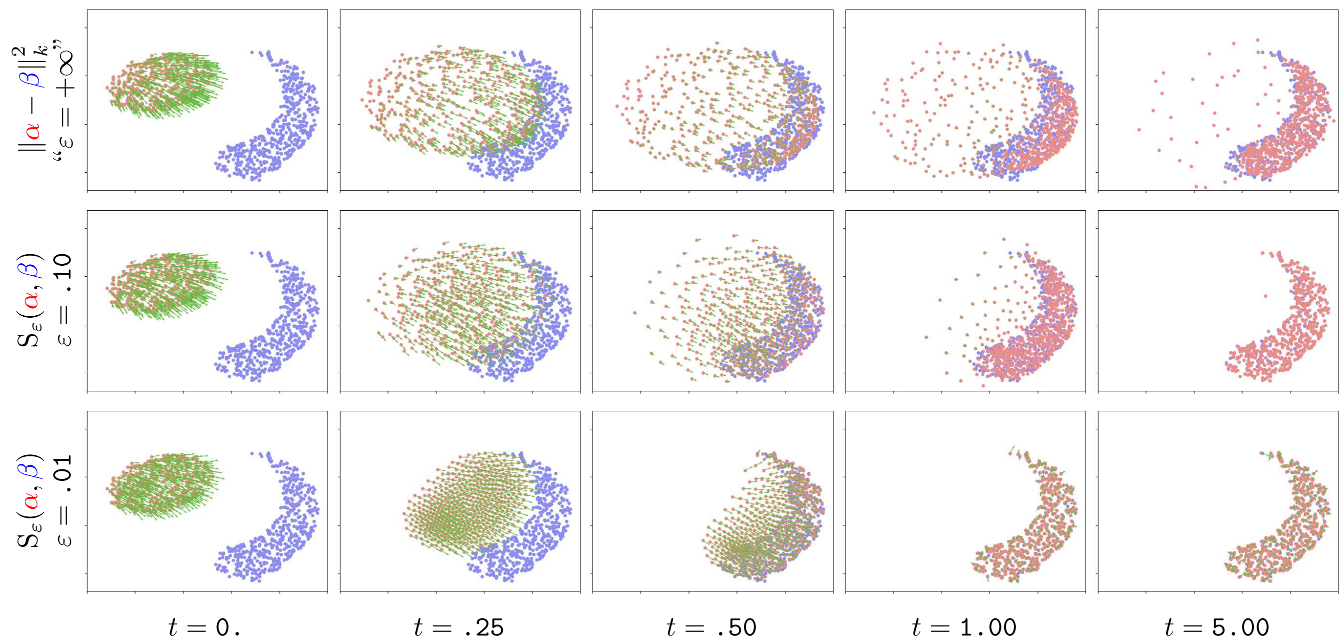
Analyzing Jets with Linearized Transport & Partial Transport



[Cai, Cheng, Craig, Craig, PRD 2020, arXiv 2021]

Interpolating between Optimal Transport & Kernel Methods

(see backup to justify color coding)



[Feydy, Séjourné, Vialard, Amari, Trouvé, Peyré, arXiv 2018]

Theoretical high-energy physics has been irreversibly impacted by the rise of deep learning

The buzz is around “AI”, but we should leverage analysis strategies from various areas of mathematics, statistics, and computer science

We have an opportunity to translate aspects of HEP theory into a computational language

ML for Formal Theory?

E.g. NN-QFT Correspondence

[image from Halverson; see Halverson, Maiti, Stoner, MLST 2021; Roberts, Yaida, Hanin, CUP 2022]

What aspects of formal theory could be rephrased as a data science problem (albeit with theoretical data)?

ML for BSM Theory?

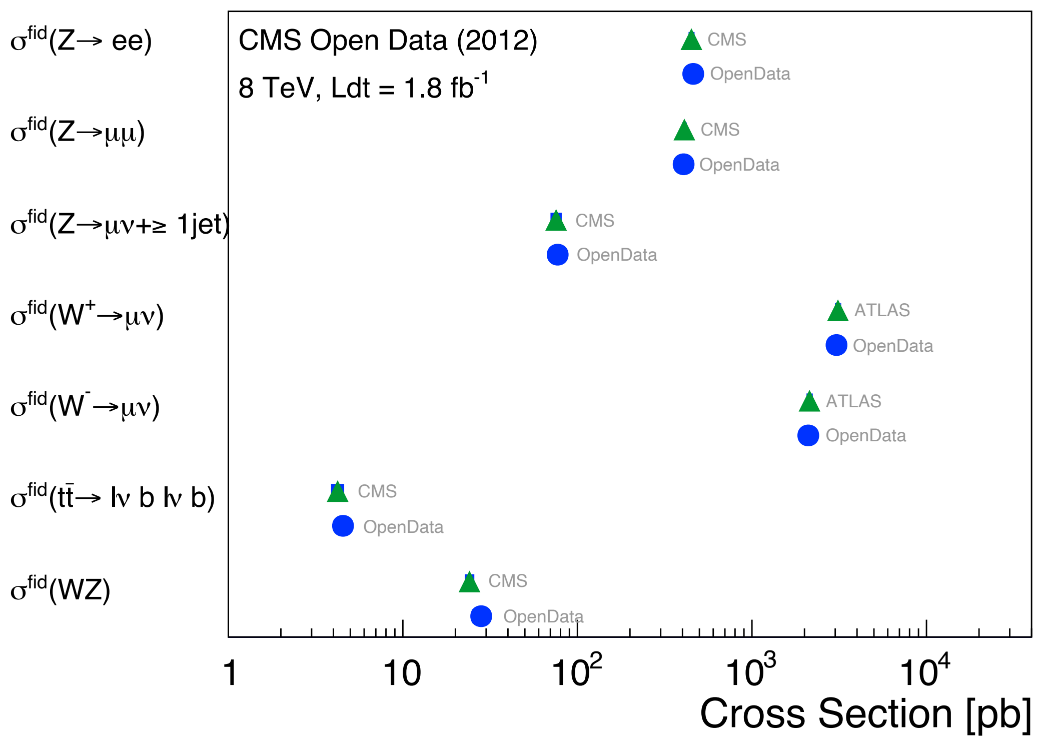
E.g. Anomaly Detection

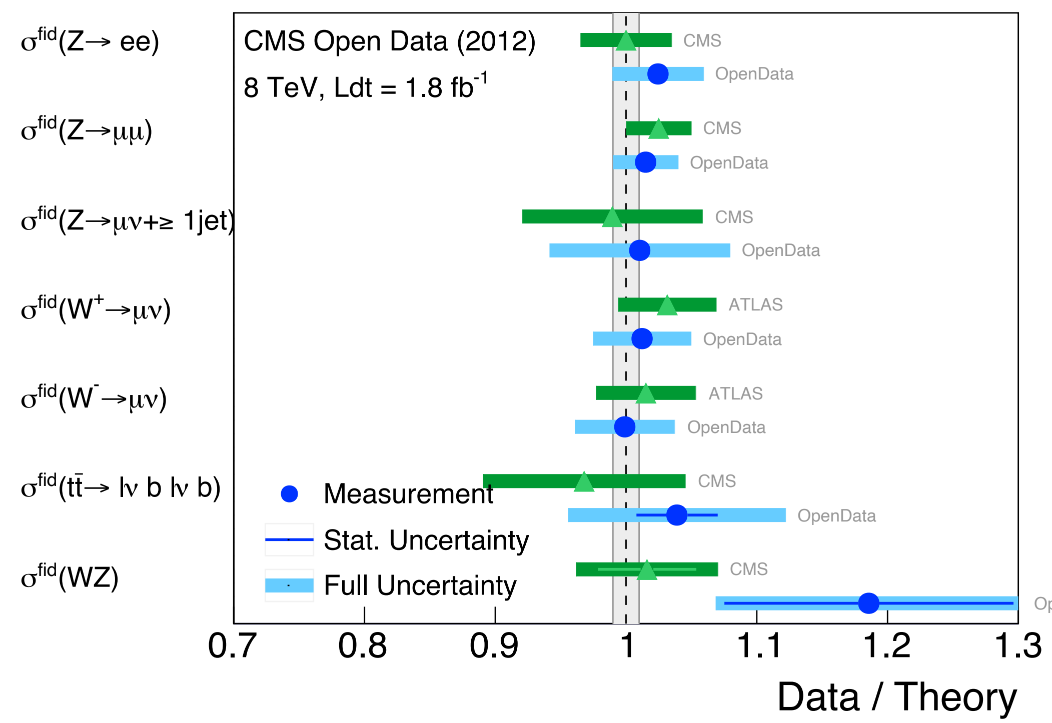
[image from LHC Olympics 2022; see Kasieczka, Nachman, Shih et al., RPP 2021]

What aspects of BSM phenomenology could be streamlined, systematized, and automated?

ML for SM Theory?

E.g. Open Data / Beyond PDF Extraction

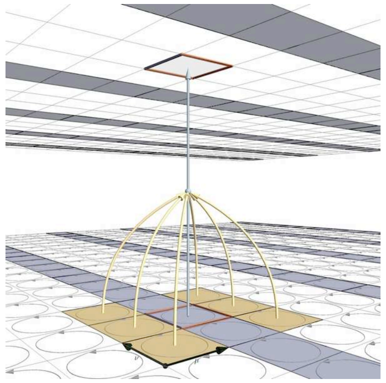




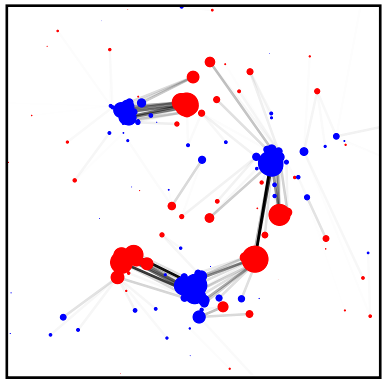
[plots from Apyan, Cuozzo, Klute, Saito, Schott, Sintayehu, JINST 2020]

What aspects of precision SM theory could be more tightly integrated into the analysis pipeline?

Machine Learning for the Theory Frontier



Theoretical high-energy physics has been irreversibly impacted by the rise of deep learning



The buzz is around "AI", but we should leverage analysis strategies from various areas of mathematics, statistics, and computer science

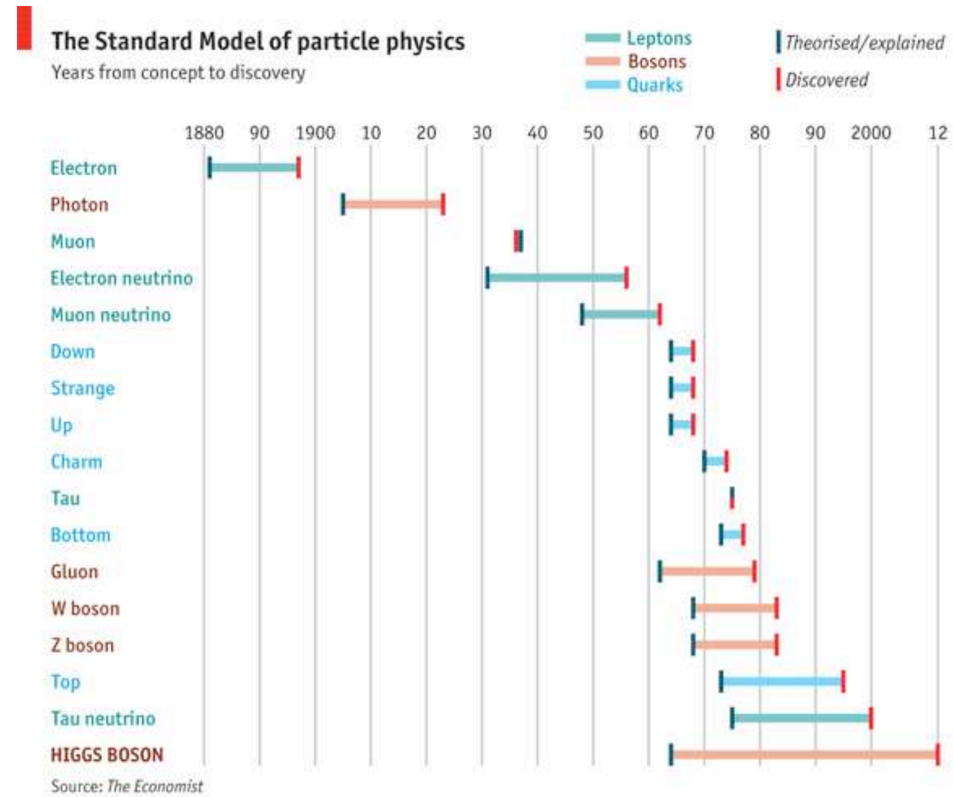
We have an opportunity to translate aspects of HEP theory into a computational language

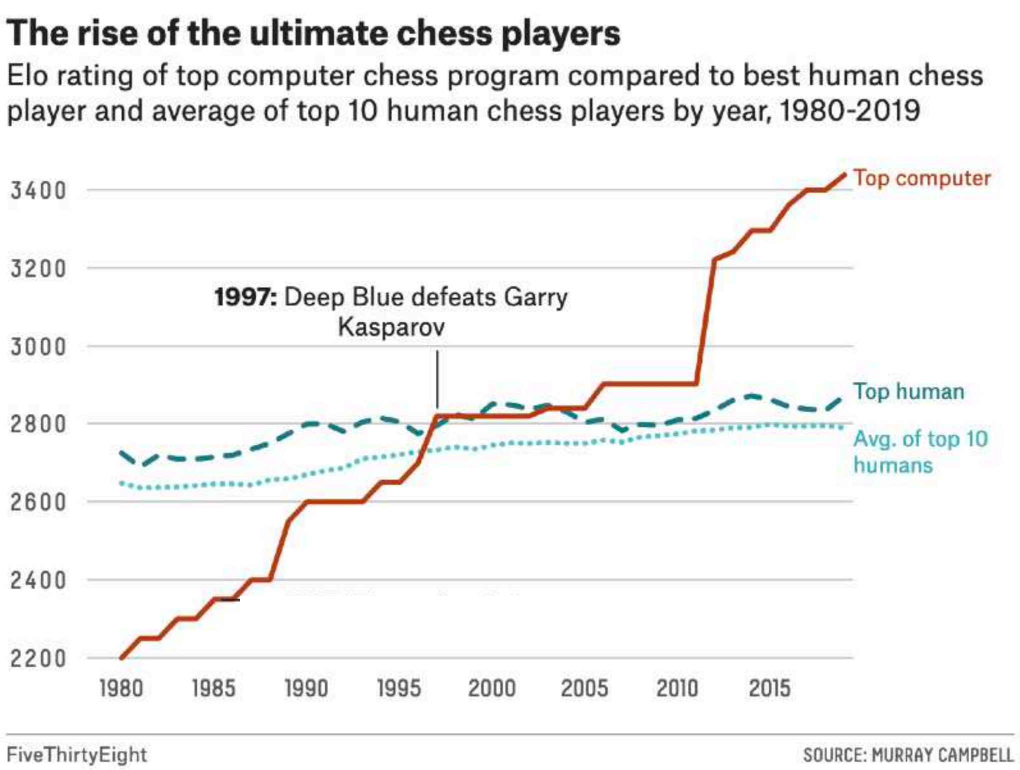
In the spirit of Snowmass, looking forward to your ideas and perspectives!

Backup Slides

Reasons for Skepticism

"How could a machine possibly outcompete a century of triumphs in theoretical physics?"

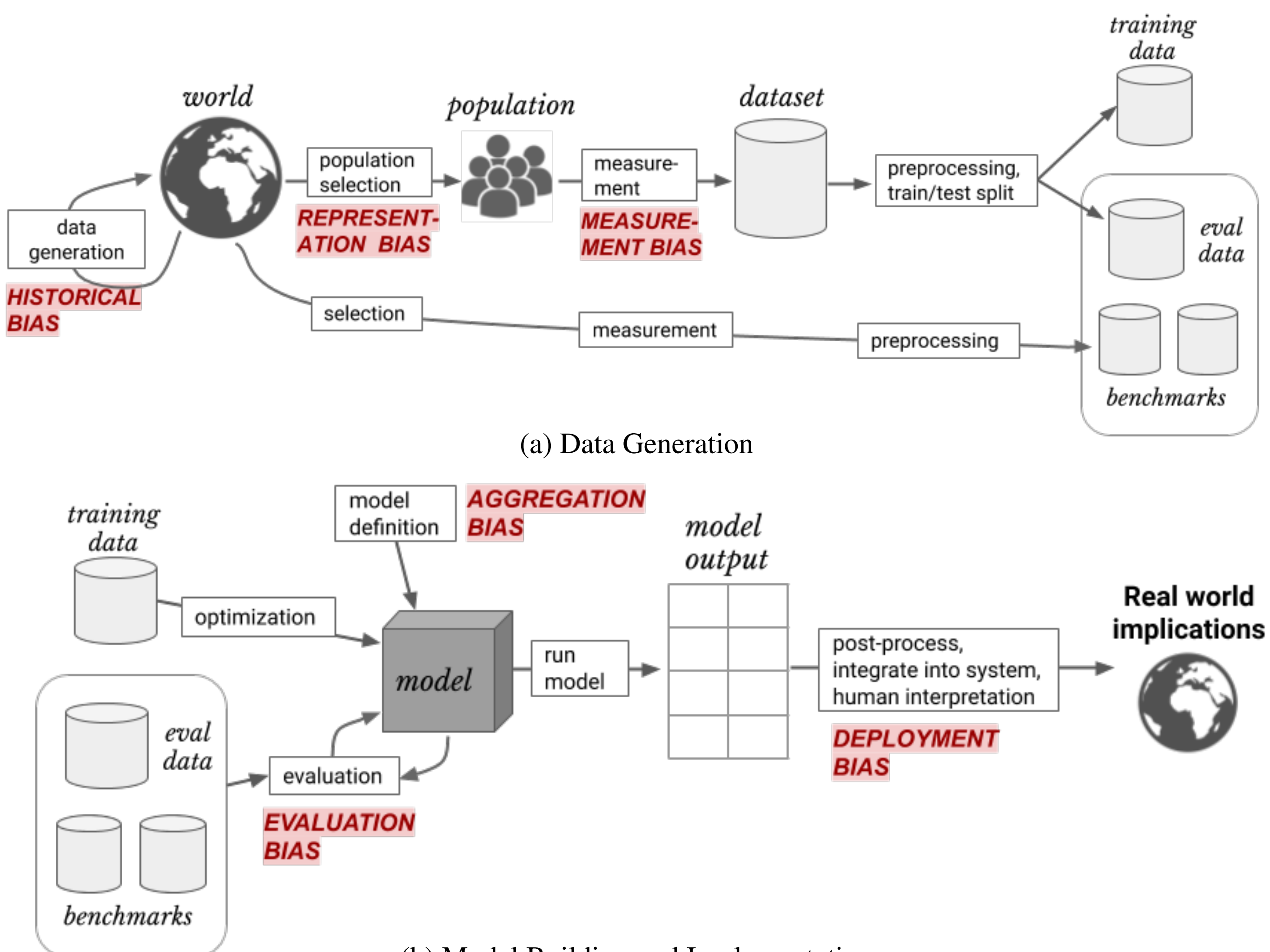




"But these are games with a precise meaning to success and amenable to brute force search"

Reasons to be Wary

"A Framework for Understanding Unintended Consequences of Machine Learning"



(a) Data Generation

(b) Model Building and Implementation

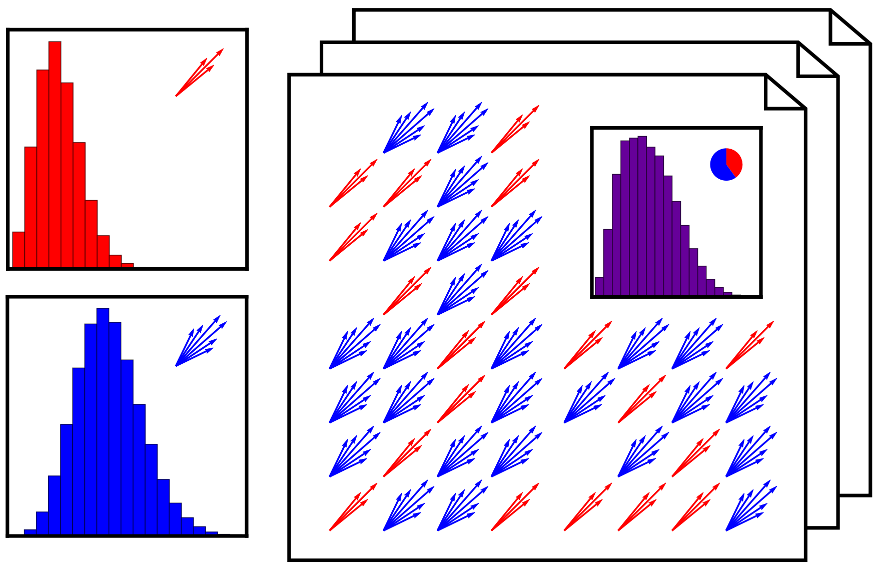
1. **Historical bias** arises when there is a misalignment between world as it is and the values or objectives to be encoded and propagated in a model. It is a normative concern with the state of the world, and exists even given perfect sampling and feature selection.
2. **Representation bias** arises while defining and sampling a development population. It occurs when the development population under-represents, and subsequently fails to generalize well, for some part of the use population.
3. **Measurement Bias** arises when choosing and measuring features and labels to use; these are often proxies for the desired quantities. The chosen set of features and labels may leave out important factors or introduce group- or input-dependent noise that leads to differential performance.
4. **Aggregation bias** arises during model construction, when distinct populations are inappropriately combined. In many applications, the population of interest is heterogeneous and a single model is unlikely to suit all subgroups.
5. **Evaluation bias** occurs during model iteration and evaluation. It can arise when the testing or external benchmark populations do not equally represent the various parts of the use population. Evaluation bias can also arise from the use of performance metrics that are not appropriate for the way in which the model will be used.
6. **Deployment Bias** occurs after model deployment, when a system is used or interpreted in inappproprate ways.

For HEP, "bias" ≈ "systematic uncertainty"

[h/t David Kaiser, MIT SERC; Suresh, Guttag, arXiv 2019]

Other Examples from My Group's Research

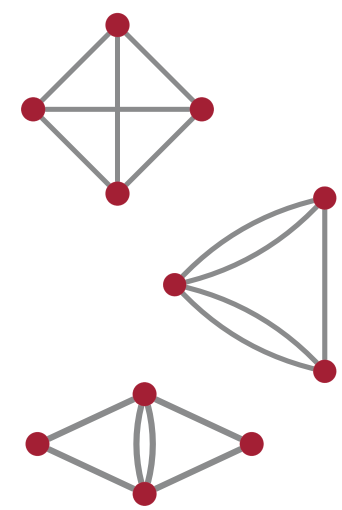
Quark/Gluon Definitions
via Blind Source Separation



[Komiske, Metodiev, JDT, JHEP 2018;
Brewer, JDT, Turner; PRD 2021]

Kinematic Decomposition
via Graph Theory

	Leafless Multigraphs	
	Connected	All
Edges d	A307317	A307316
1	0	0
2	1	1
3	2	2
4	4	5
5	9	11
6	26	34
7	68	87
8	217	279
9	**718**	**897**
10	**2 553**	**3 129**
11	**9 574**	**11 458**
12	**38 005**	**44 576**
13	**157 306**	**181 071**
14	**679 682**	**770 237**
15	**3 047 699**	**3 407 332**
16	**14 150 278**	**15 641 159**



[Komiske, Metodiev, JDT,
JHEP 2018, PRD 2020]

New insights into high-energy physics facilitated by advances in mathematics, statistics, and computer science

(and vice versa!)

Siloing in the Scientific Community

$$\mathrm{Kernel}_k(\alpha,\beta) \;=\; \tfrac{1}{2}\langle\alpha,\,k\star\alpha\rangle - \langle\alpha,\,k\star\beta\rangle + \tfrac{1}{2}\langle\beta,\,k\star\beta\rangle$$

$$= \tfrac{1}{2}\sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{N}} \alpha_i\alpha_j\,k(x_i,x_j) \;-\; \sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{M}} \alpha_i\beta_j\,k(x_i,y_j) \;+\; \tfrac{1}{2}\sum_{i=1}^{\mathrm{M}}\sum_{j=1}^{\mathrm{M}} \beta_i\beta_j\,k(y_i,y_j)$$

Kernel methods. Formulas in the mould of Eqs. (3.99-3.101) are **ubiquitous in applied sciences**: from physics to machine learning, applying a convolution is the simplest way of modelling spatial correlations and pair-wise interactions. Unfortunately though, few papers and textbooks take the time to draw explicit links between fields that have, at first glance, very little in common. Before going any further, we devote a few pages to a short panorama around the six major interpretations of Eq. (3.99). As we identify with each other the theories of:

1. **Newtonian gravitation and electrostatics** in physics,
2. **blurred squared distances** in imaging sciences,
3. **Sobolev norms** in functional analysis,
4. **maximum mean discrepancies** in statistics,
5. **reproducing kernel Hilbert spaces** in machine learning and
6. **Kriging**, **splines** or **Gaussian processes** in geostatistics, imaging and probabilities,

we will hopefully help the reader to get a deeper understanding of a theory that is central to modern data sciences.

[Feydy, Geometric data analysis, beyond convolutions]