KISS
KI zur schnellen Simulation von wissenschaftlichen Daten

Prof. Dr. Gregor Kasieczka
Email: gregor.kasieczka@uni-hamburg.de
@kasieczka.bsky.social / Gregor Kasieczka
BDA Annual Meeting — 13.03.2025

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

KISS

FSP CMS

PUNCH 4NFDI

DASHH

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

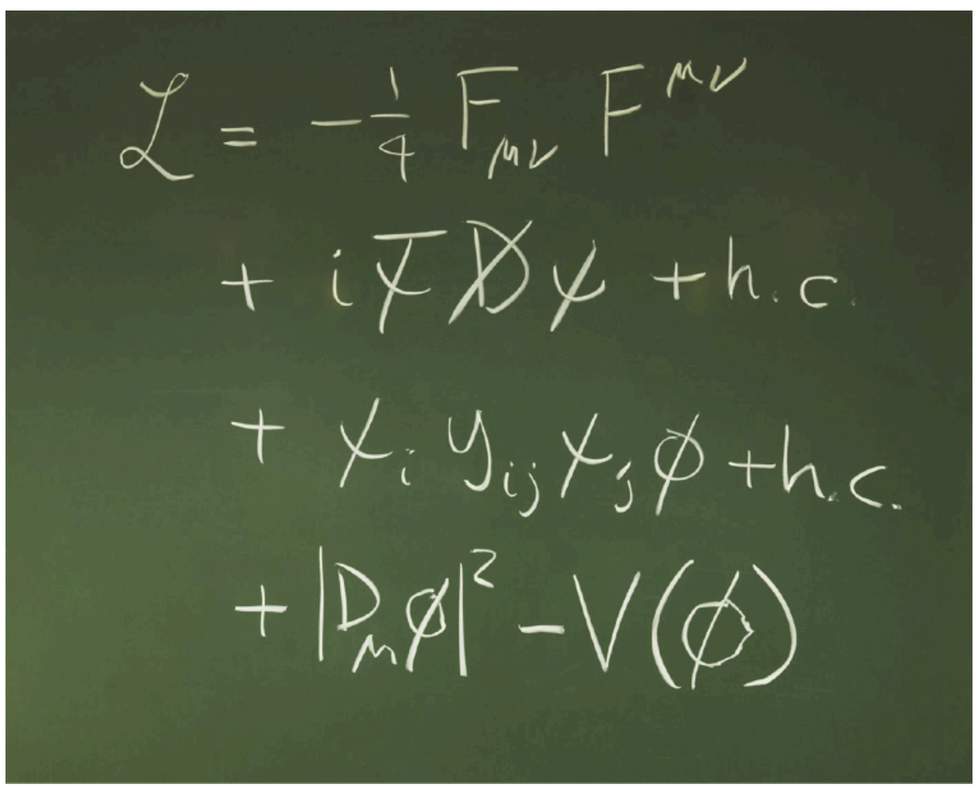
CDCS
CENTER FOR DATA AND COMPUTING IN NATURAL SCIENCES

PIER
Partnership of Universität Hamburg and DESY

Emmy Noether-Programm
Deutsche Forschungsgemeinschaft
DFG

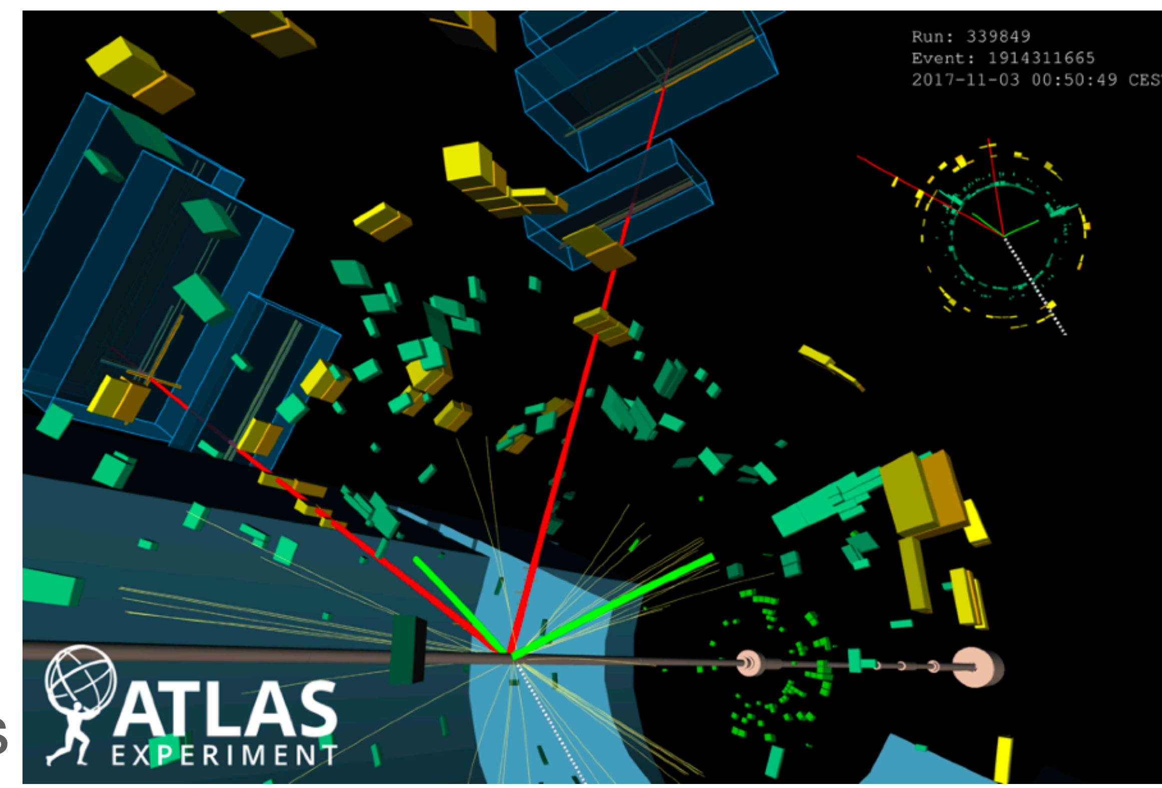
Motivation

Simulations relate fundamental physical structures to observable quantities.



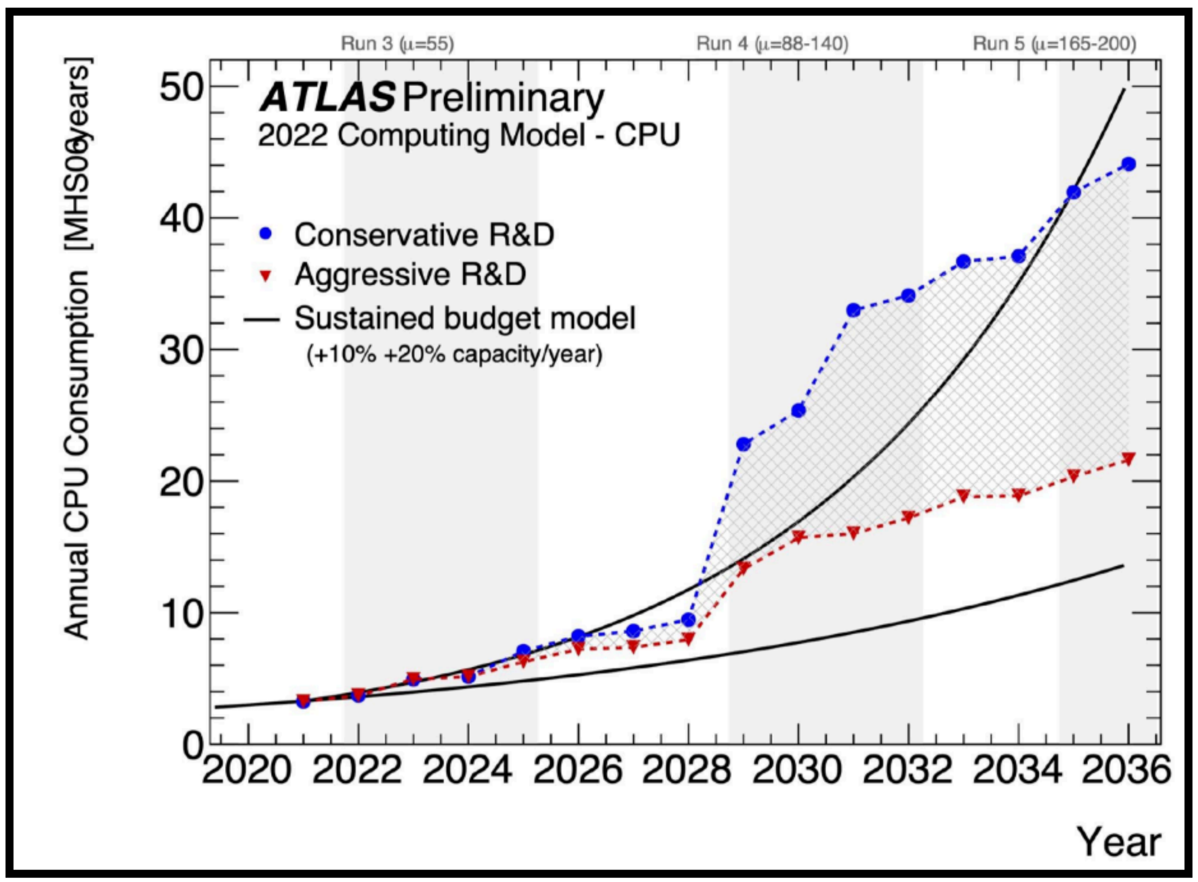
Complex chain of simulations

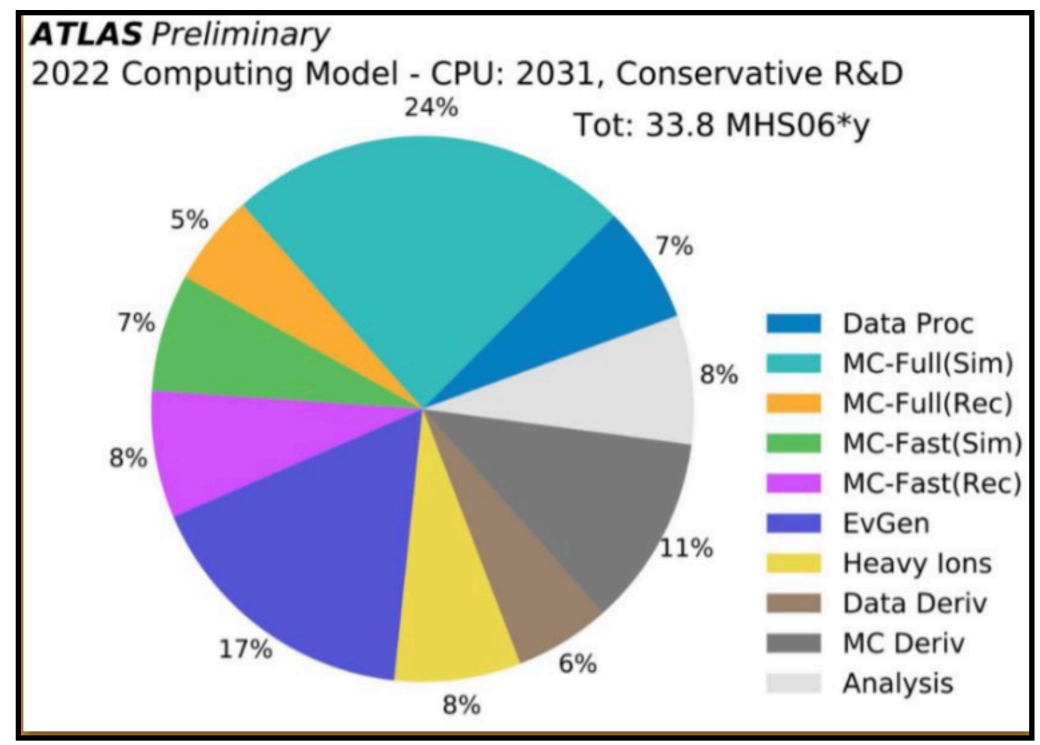
Simulation-aided inverse problems



Motivation

Simulations relate fundamental physical structures to observable quantities, but are computationally very expensive.





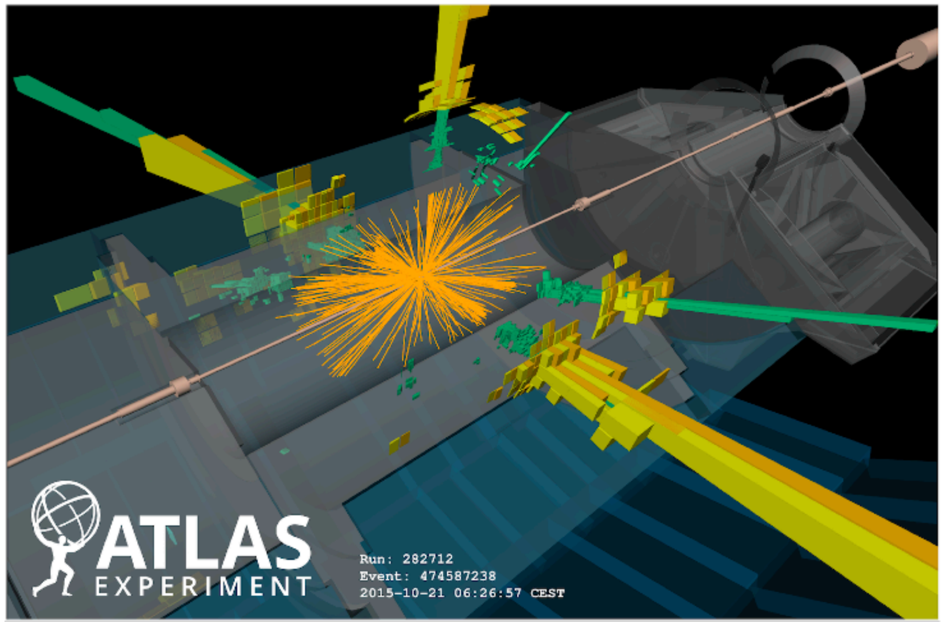
Example for collider physics, similar issues in all domains

https://cds.cern.ch/record/2802918/files/LHCC-G-182.pdf

Motivation

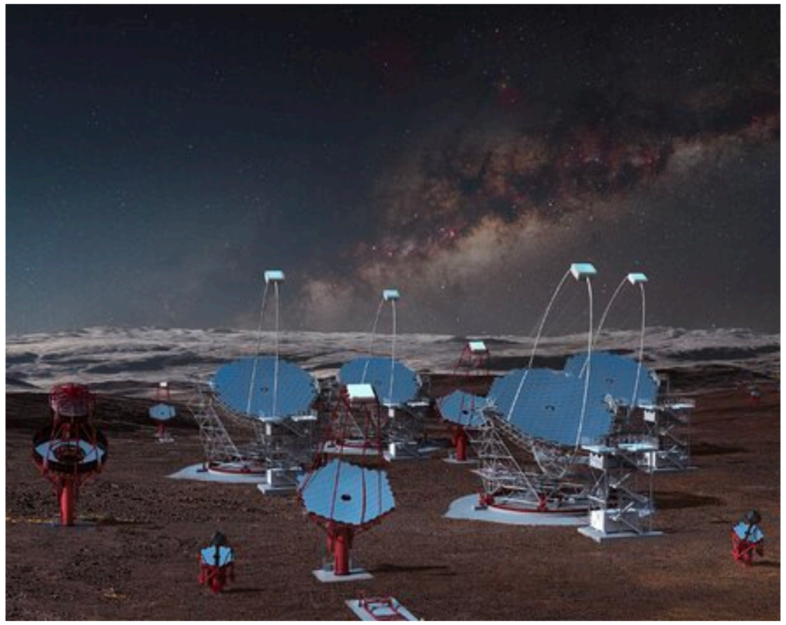
Simulations relate fundamental physical structures to observable quantities, but are computationally very expensive.

They are crucially required for research in particle physics, hadron- and nuclear physics, astro-particle physics, and astronomy.



Experimental particle physics probes nature at length scales of 10^{-18} meters.

Between particle physics and astronomy, KISS covers 45 orders of magnitudes.

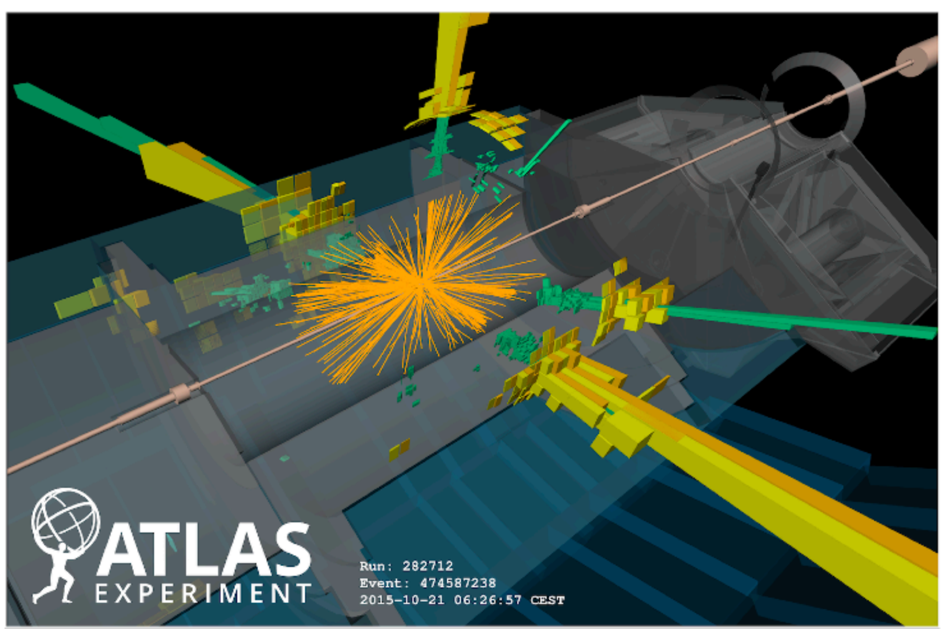


The observable universe has a diameter of 10^{27} meters.

Motivation

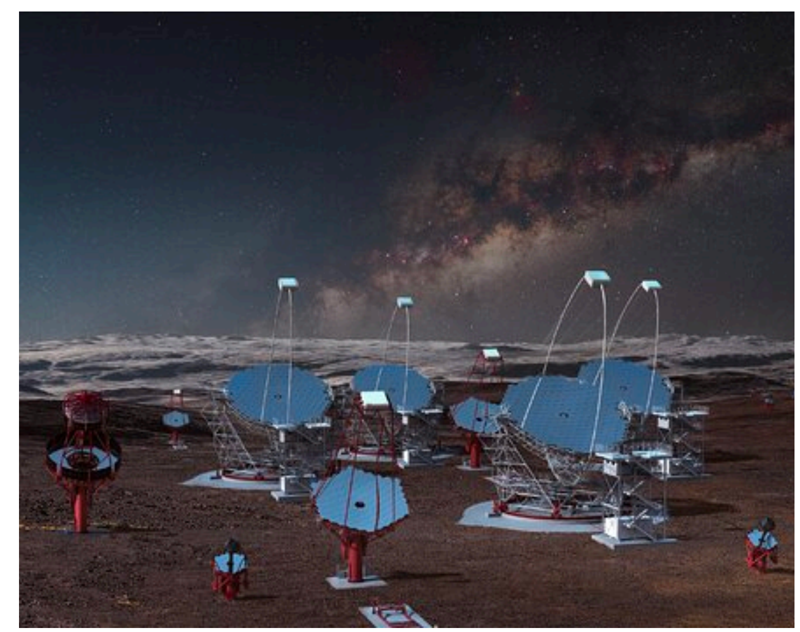
Simulations relate fundamental physical structures to observable quantities, but are computationally very expensive.

They are crucially required for research in particle physics, hadron- and nuclear physics, astro-particle physics, and astronomy.



Experimental particle physics probes nature at length scales of 10^{-18} meters.

United by scientific questions and key methods



The observable universe has a diameter of 10^{27} meters.

Motivation

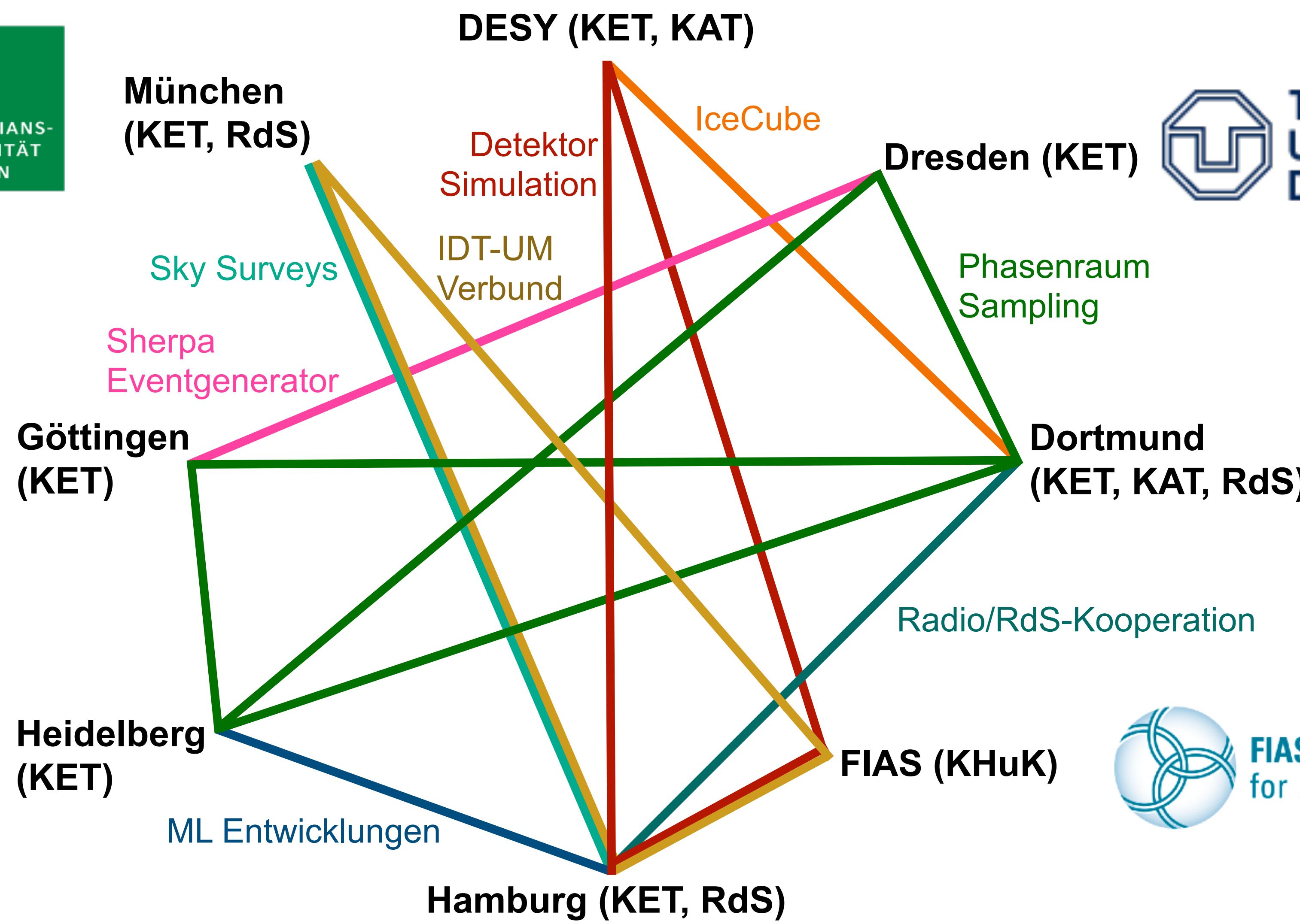
Simulations relate fundamental physical structures to observable quantities.

They are crucially required for research in particle physics, hadron- and nuclear physics, astro-particle physics, and astronomy.

KISS develops and researches generative AI tools to increase the efficiency of scientific simulations

KISS

Composition

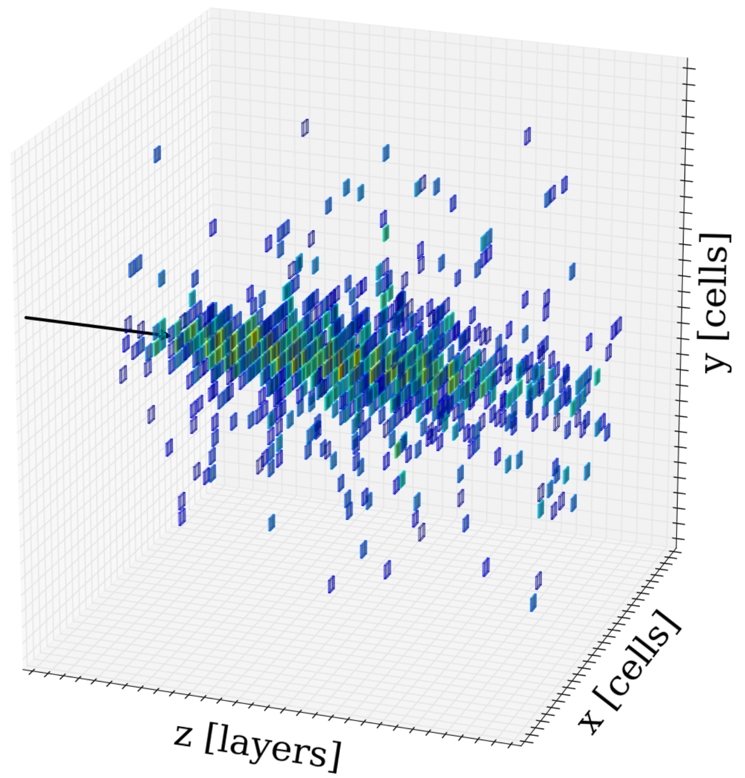


Associated:

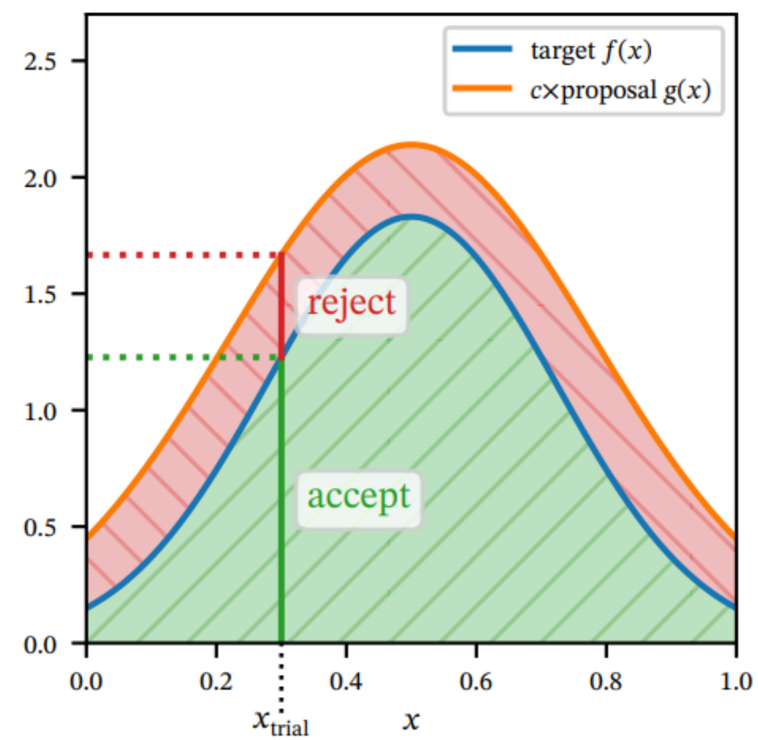
Karlsruher Institut für Technologie

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

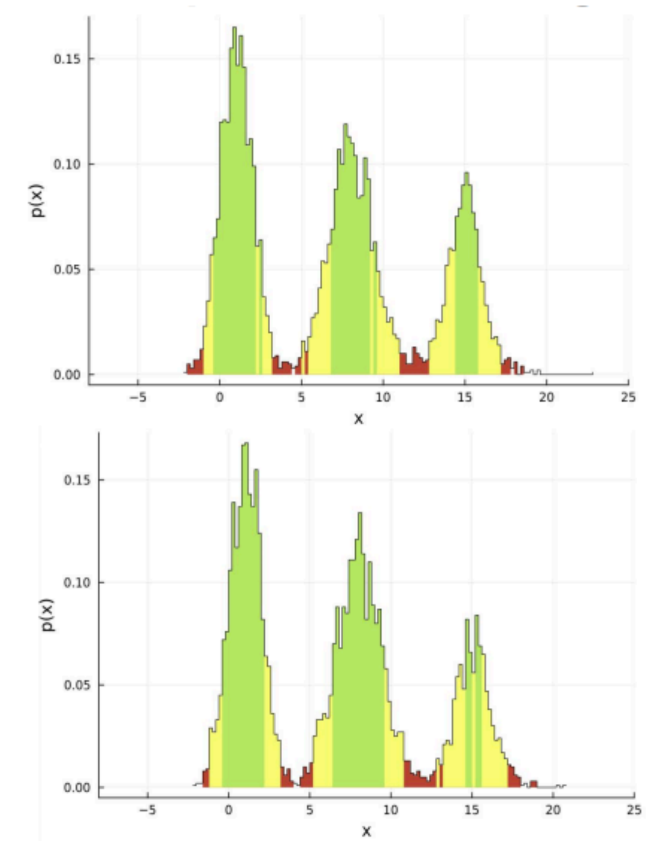
Overview



Surrogates



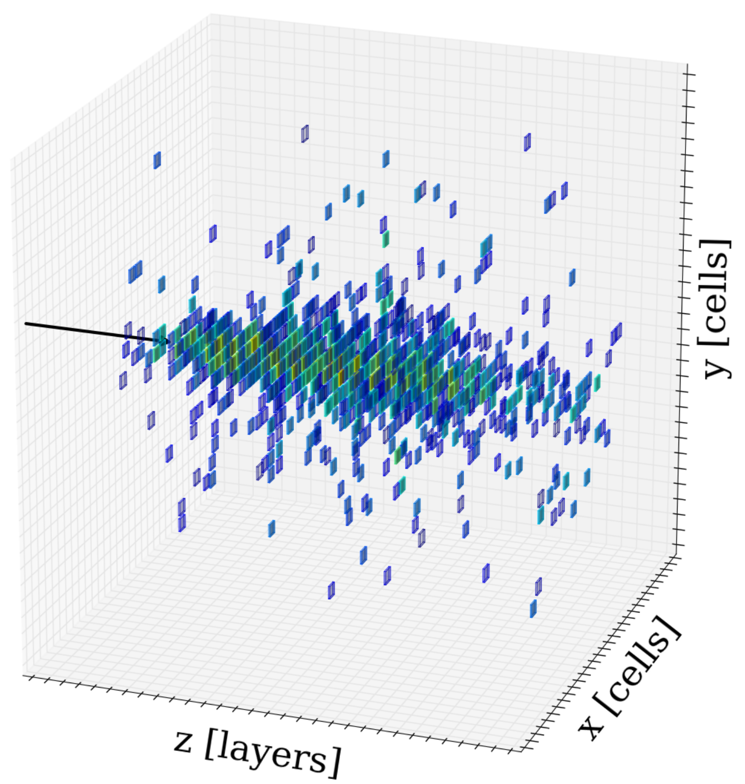
Sampling



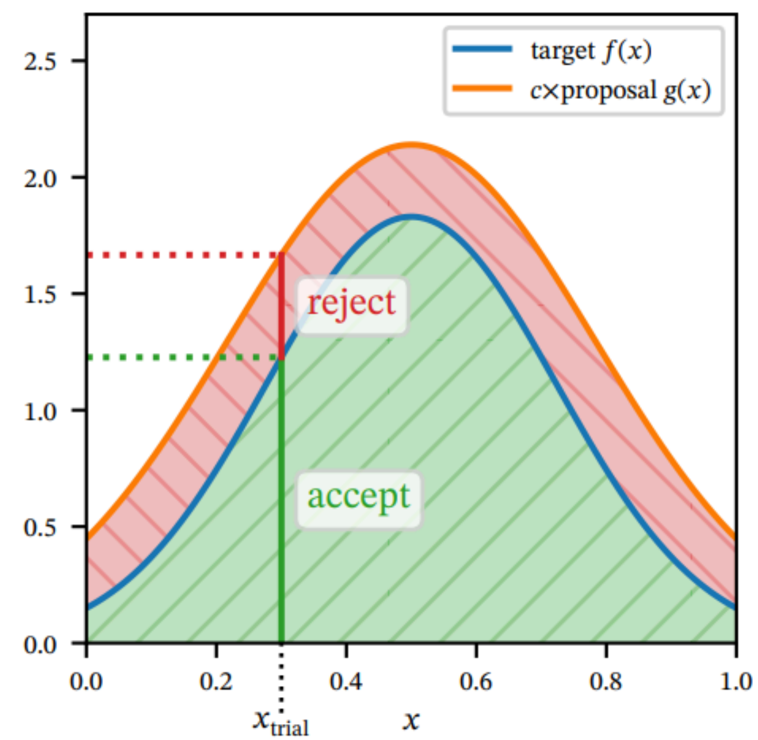
Quality

*apologies for selection bias!

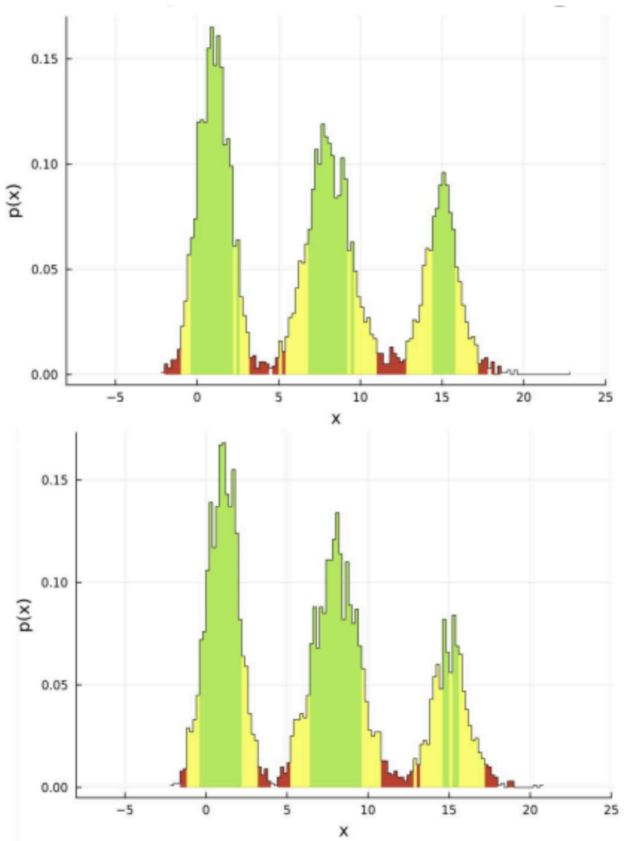
Overview



Surrogates



Sampling



Quality

Generative Image Models

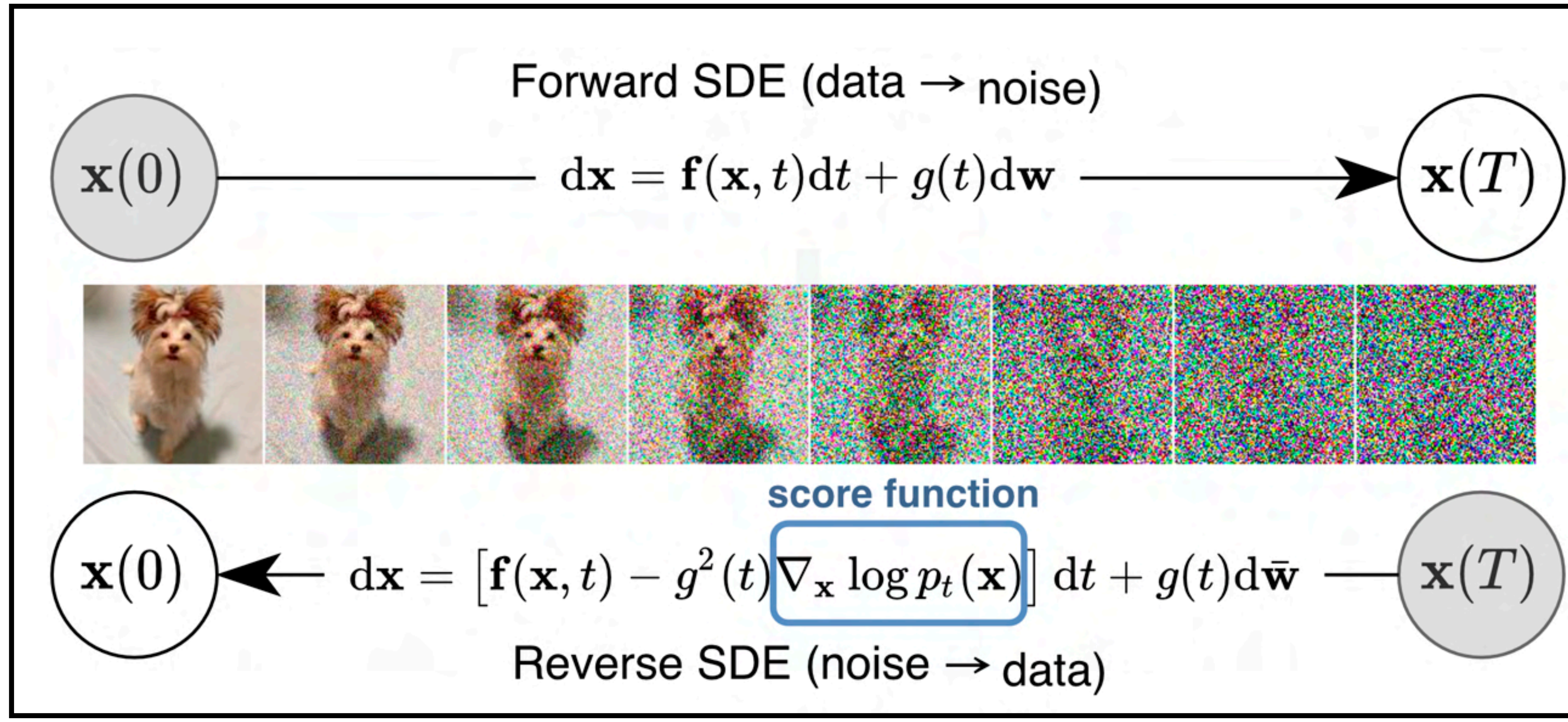
Massive progress in the
generation of artificial images

2024

Generative Image Models

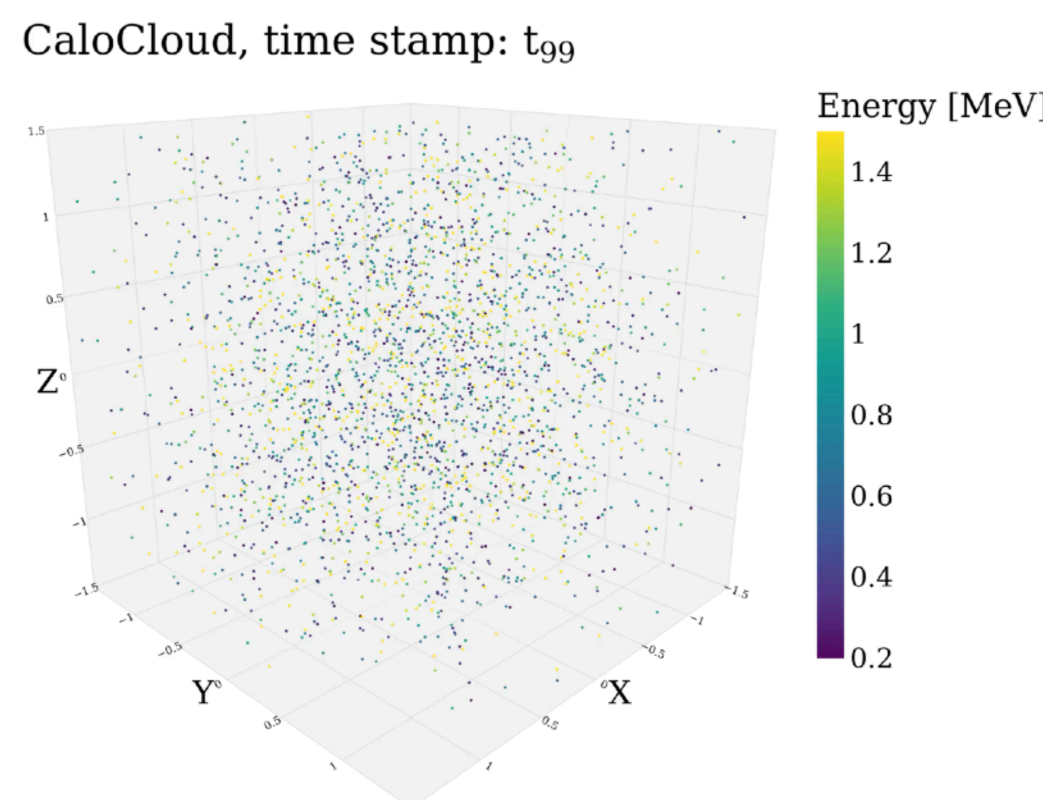
Massive progress in the generation of artificial images

Main driver: Normalising flows and diffusion



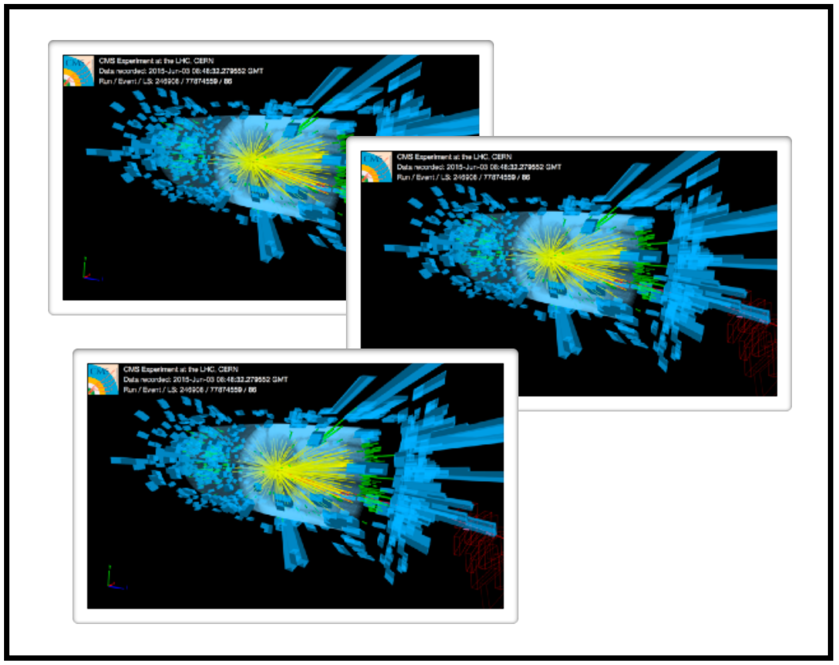
Idea:
Use classical simulation to produce initial training data for generative model

Example: surrogate model for particle interaction in high granularity detector.



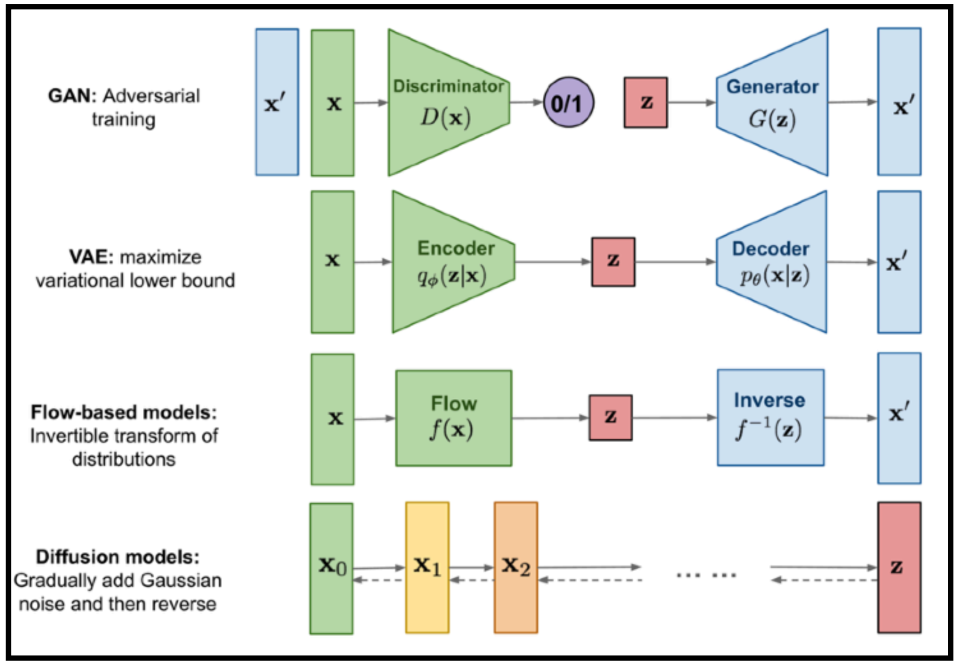
Strategy

1. Use classical simulation or data as input

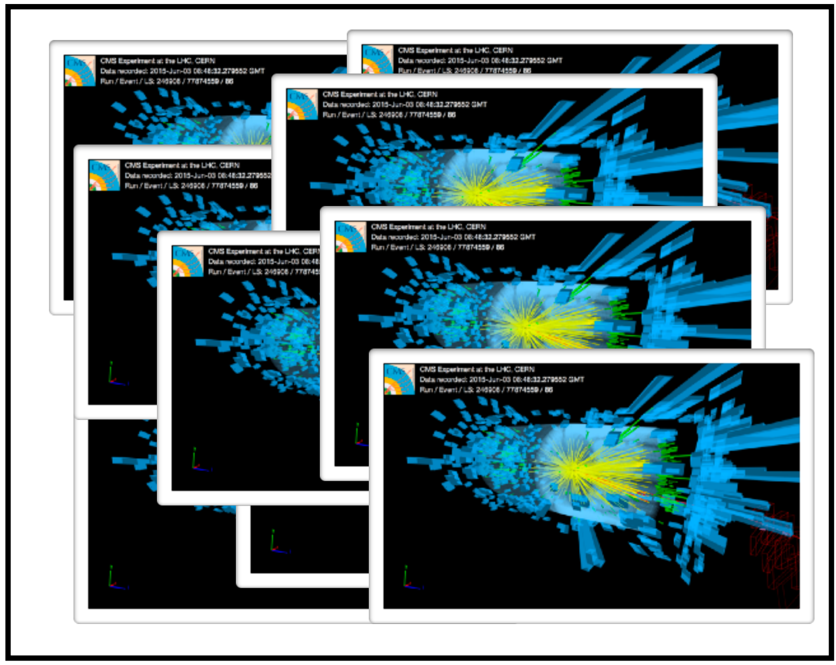


(slow)

2. Train generative surrogate



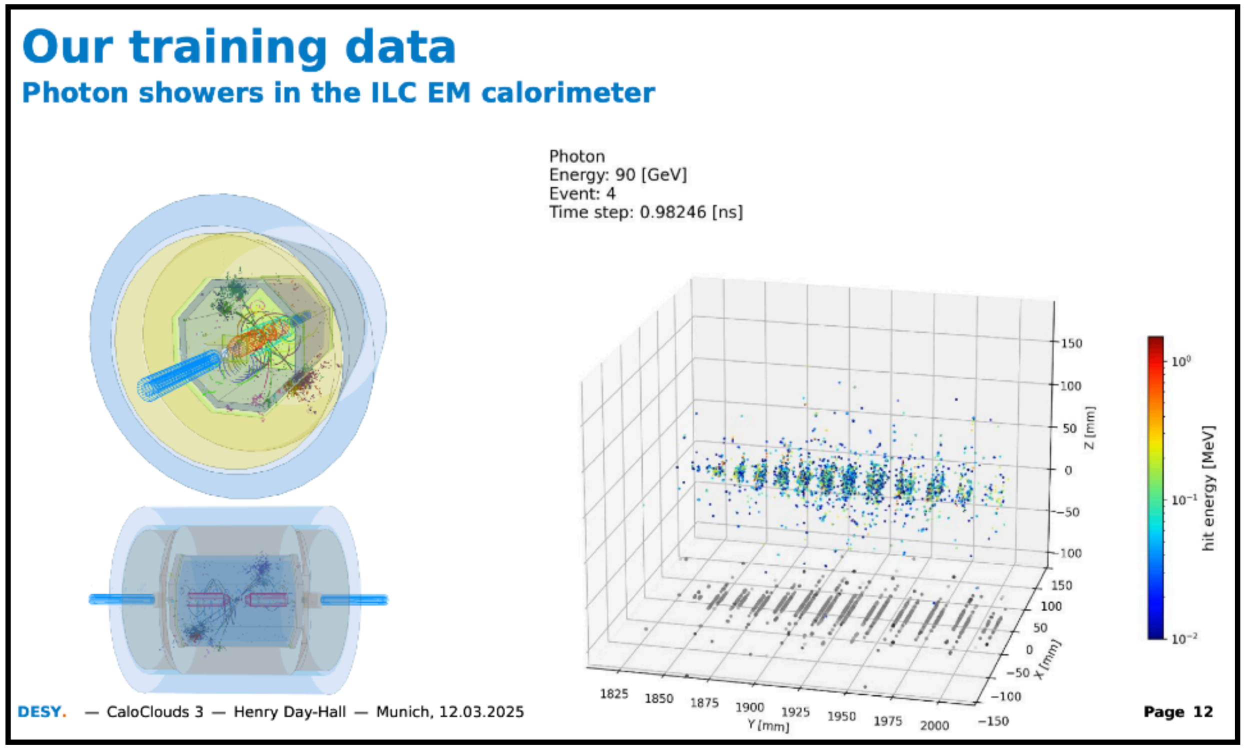
3. Oversample



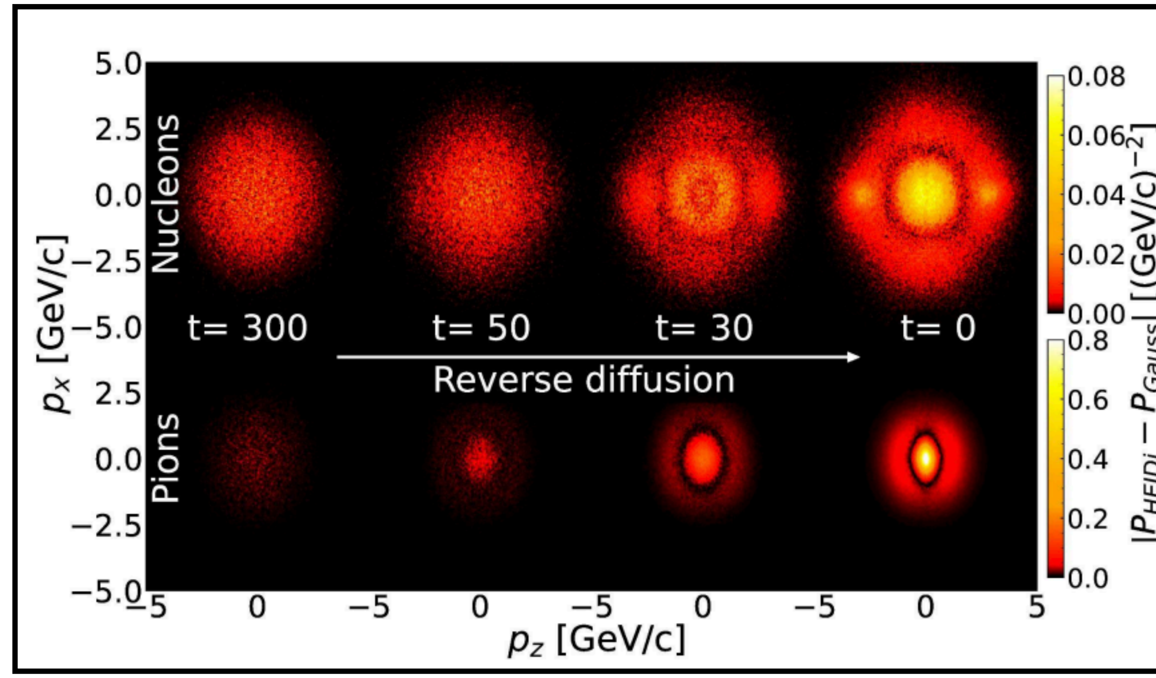
(**fast**)

Paganini, Oliveira, Nachman 1705.02355; Butter, Diefenbacher, **GK**, et al 2008.06545;

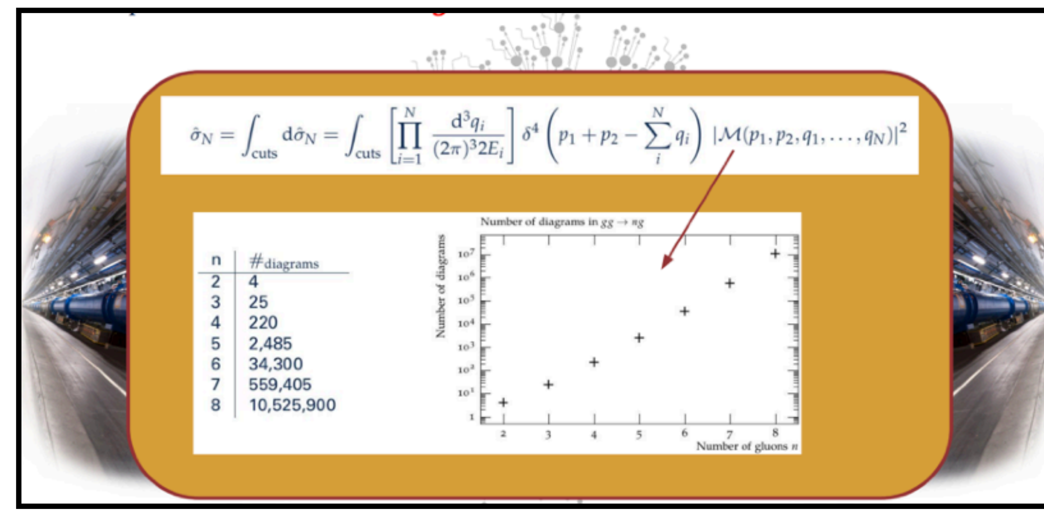
Targets



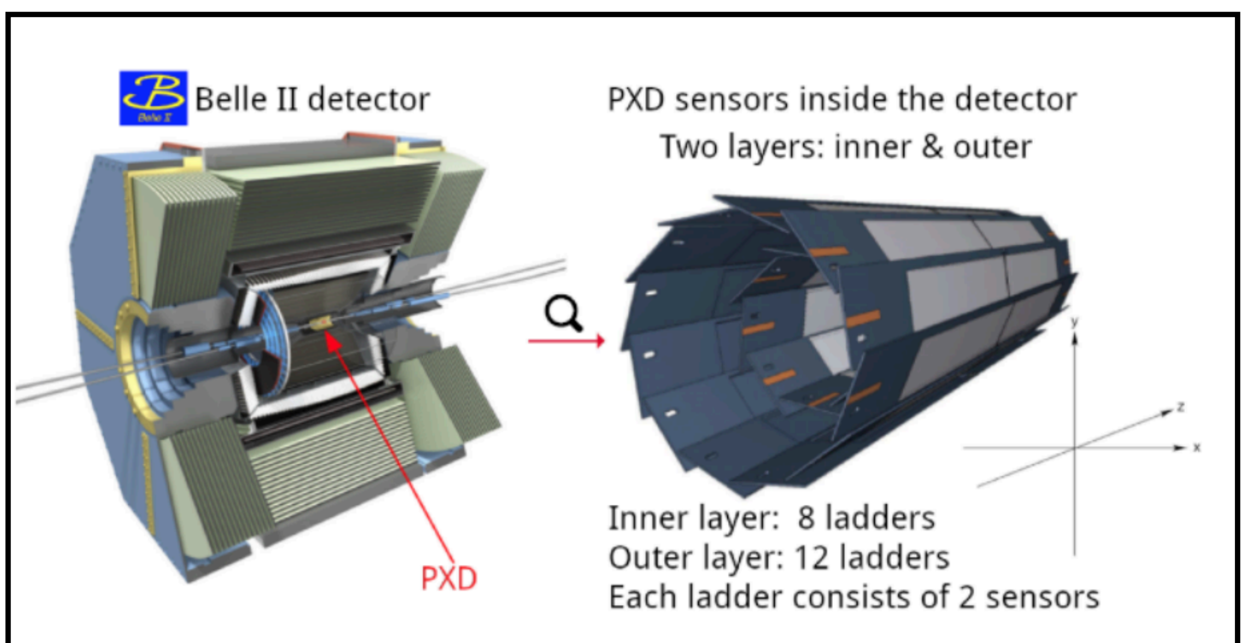
Particle showers in calorimeters



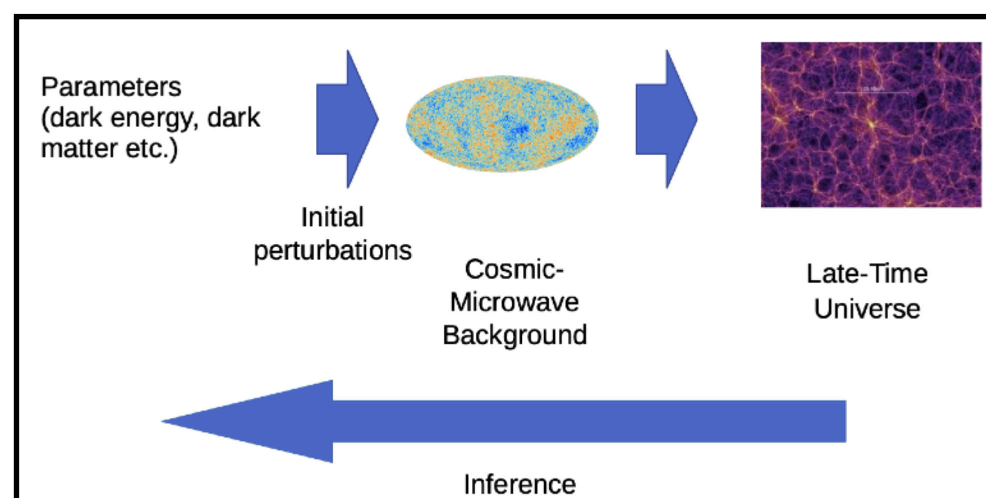
Heavy ion collisions



Hard-scattering matrix elements

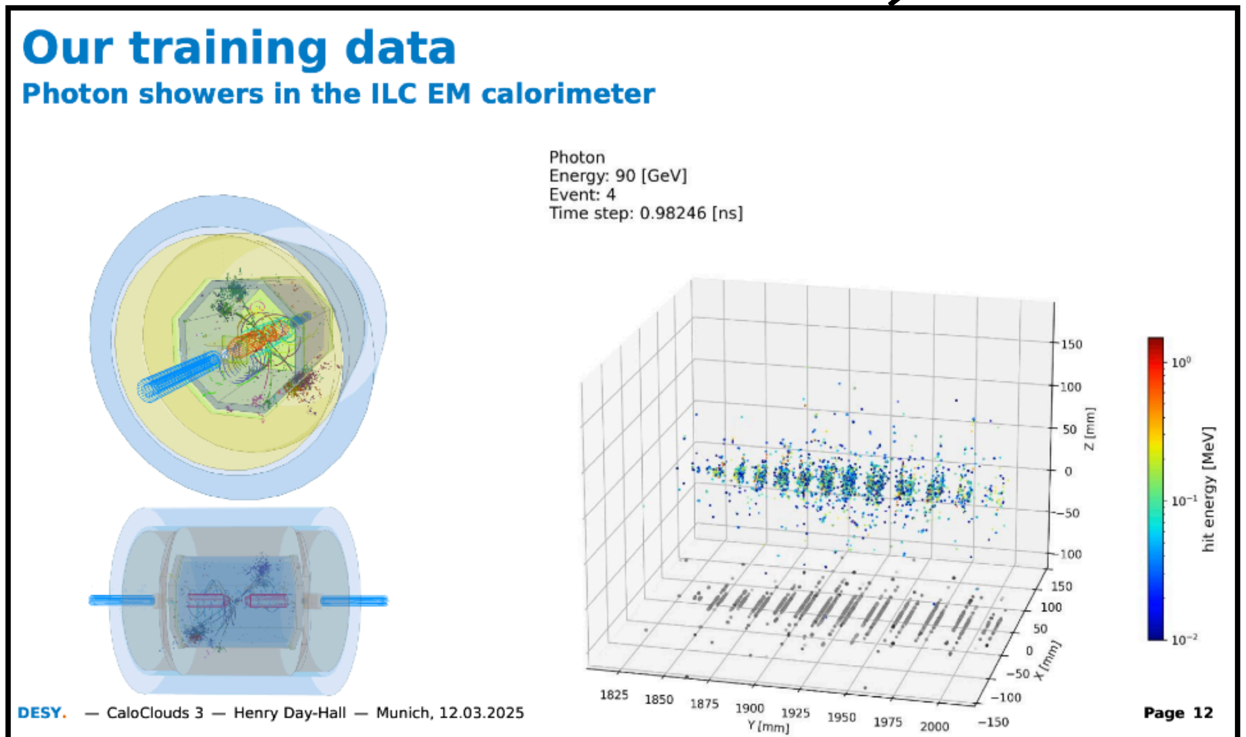


Background hits in pixel sensors

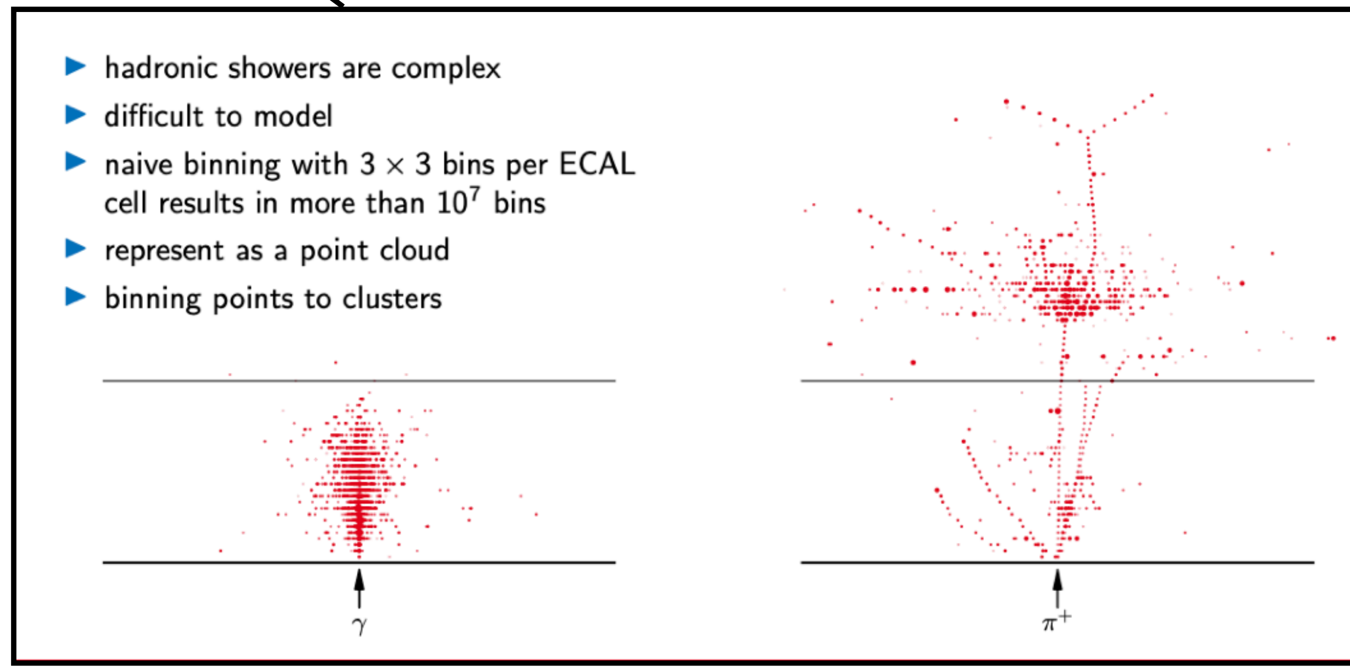


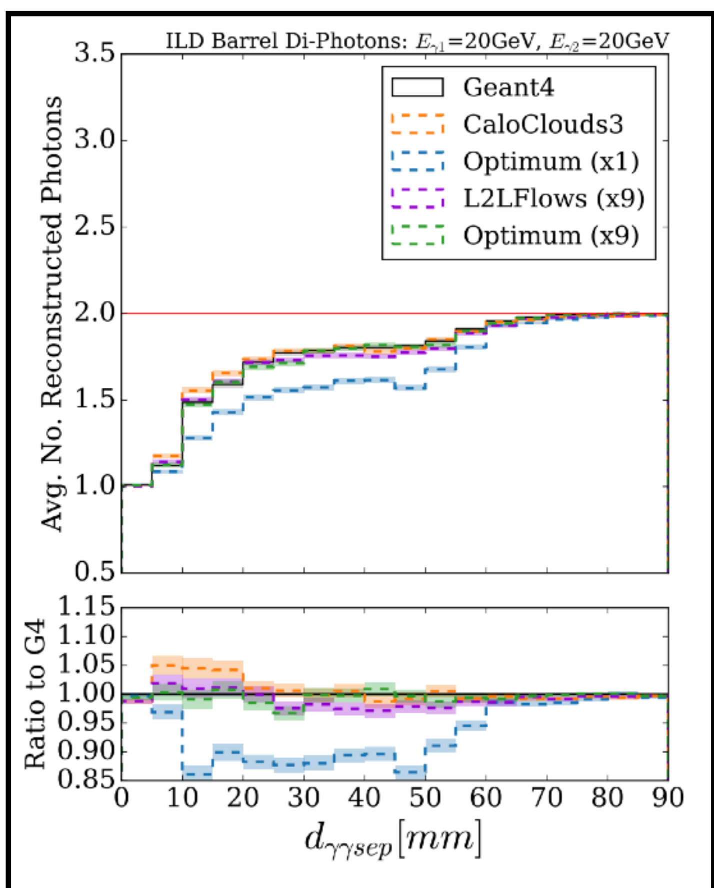
Distributions of galaxies

Targets

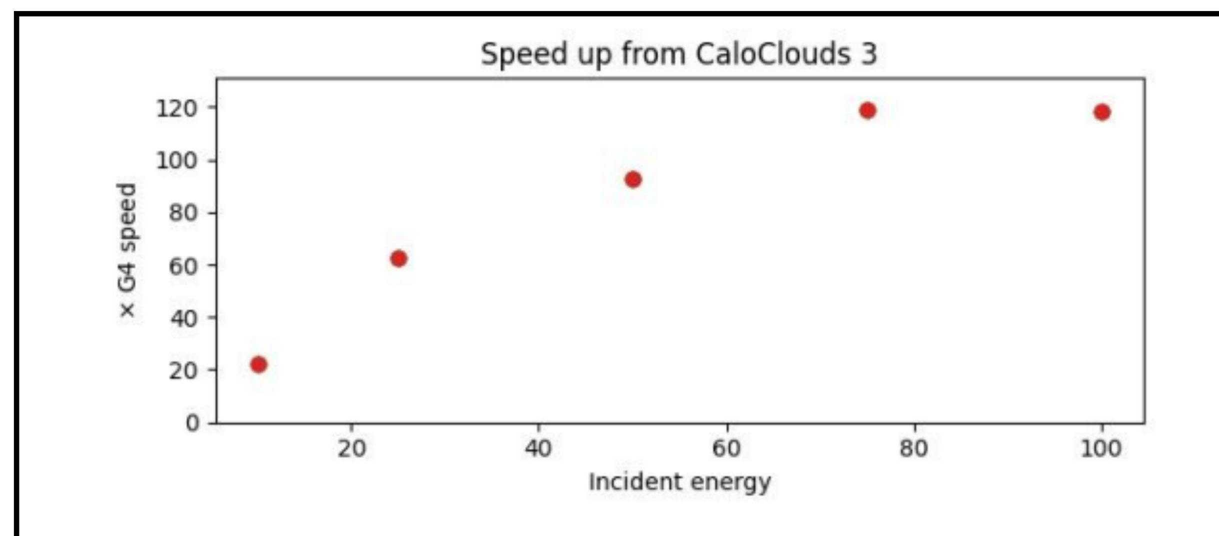


Adressing more complex pion showers



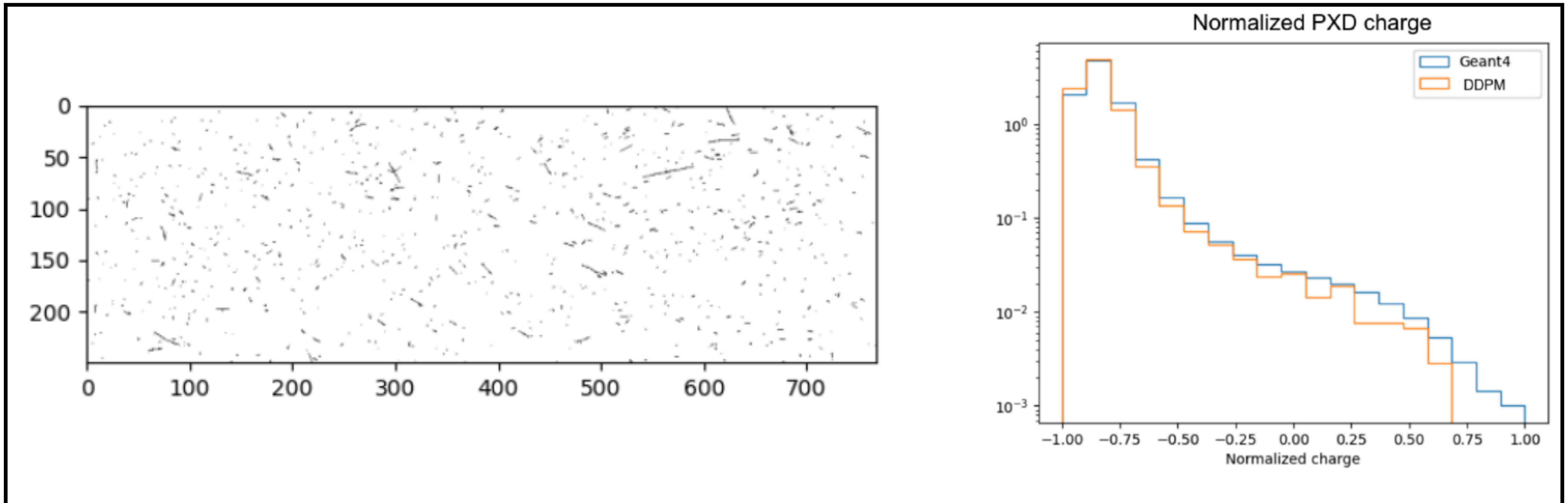


Point-cloud diffusion model (CaloClouds) with very good generative fidelity

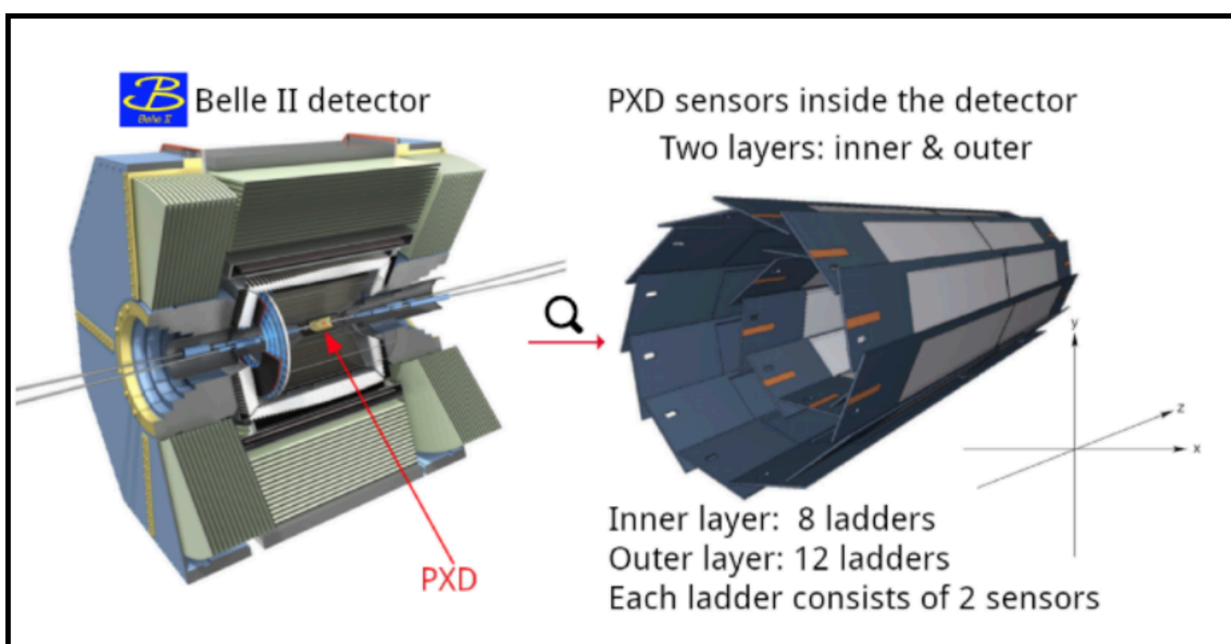


2 orders of magnitude speed-up on same hardware

Targets

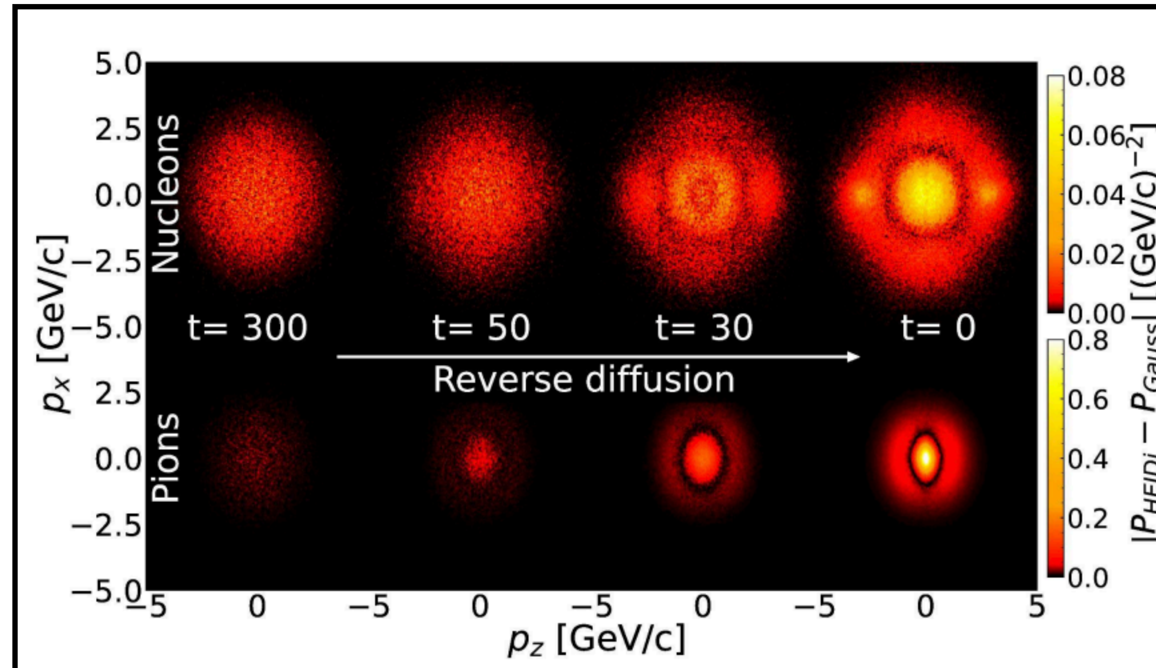


Diffusion model simulates difficult background hits
Good agreement on charge, more tuning needed on correlations

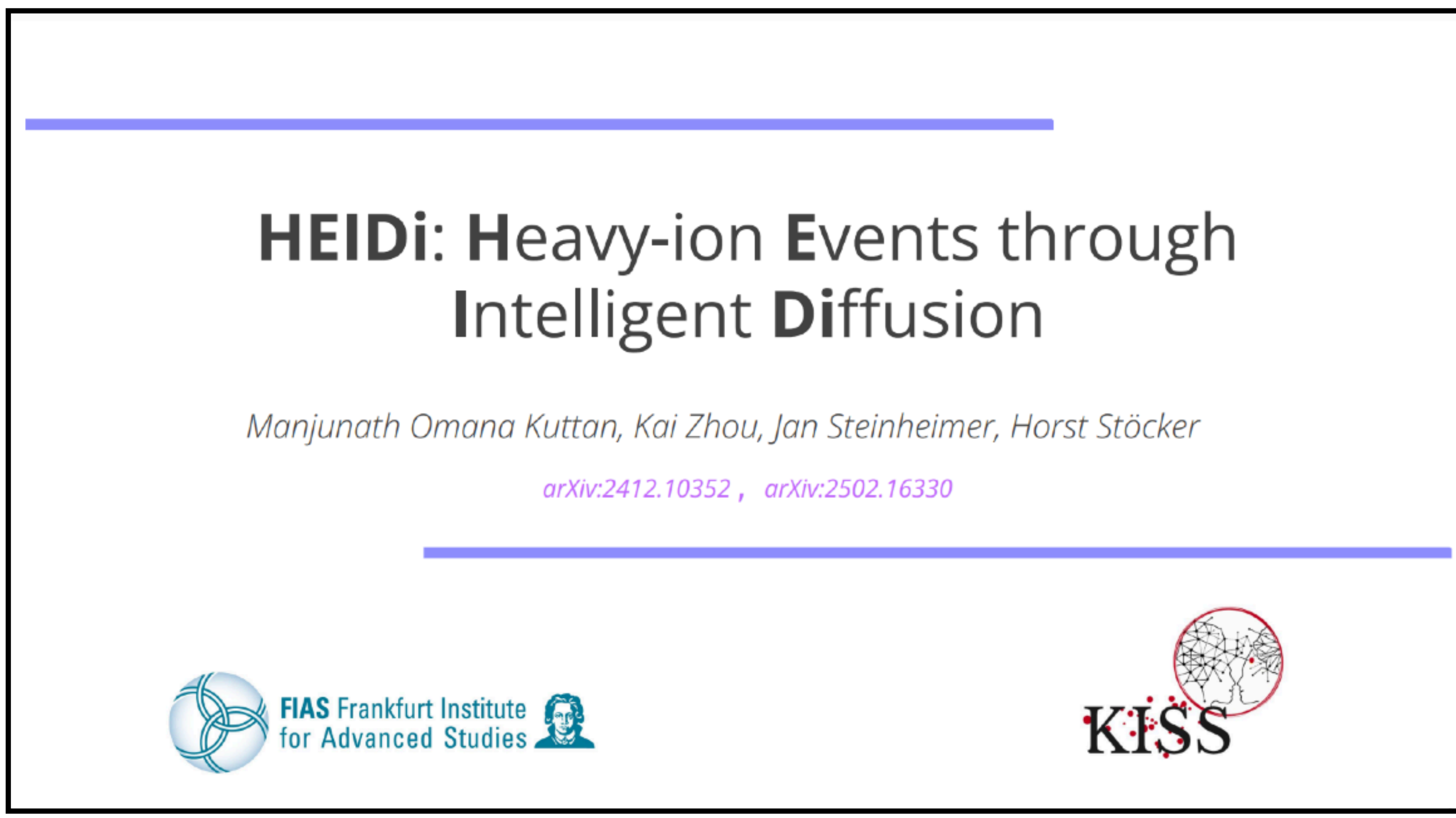


Background hits in pixel sensors

Targets



Heavy ion collisions



Dedicated talk
(Spoiler: substantial speed-up)

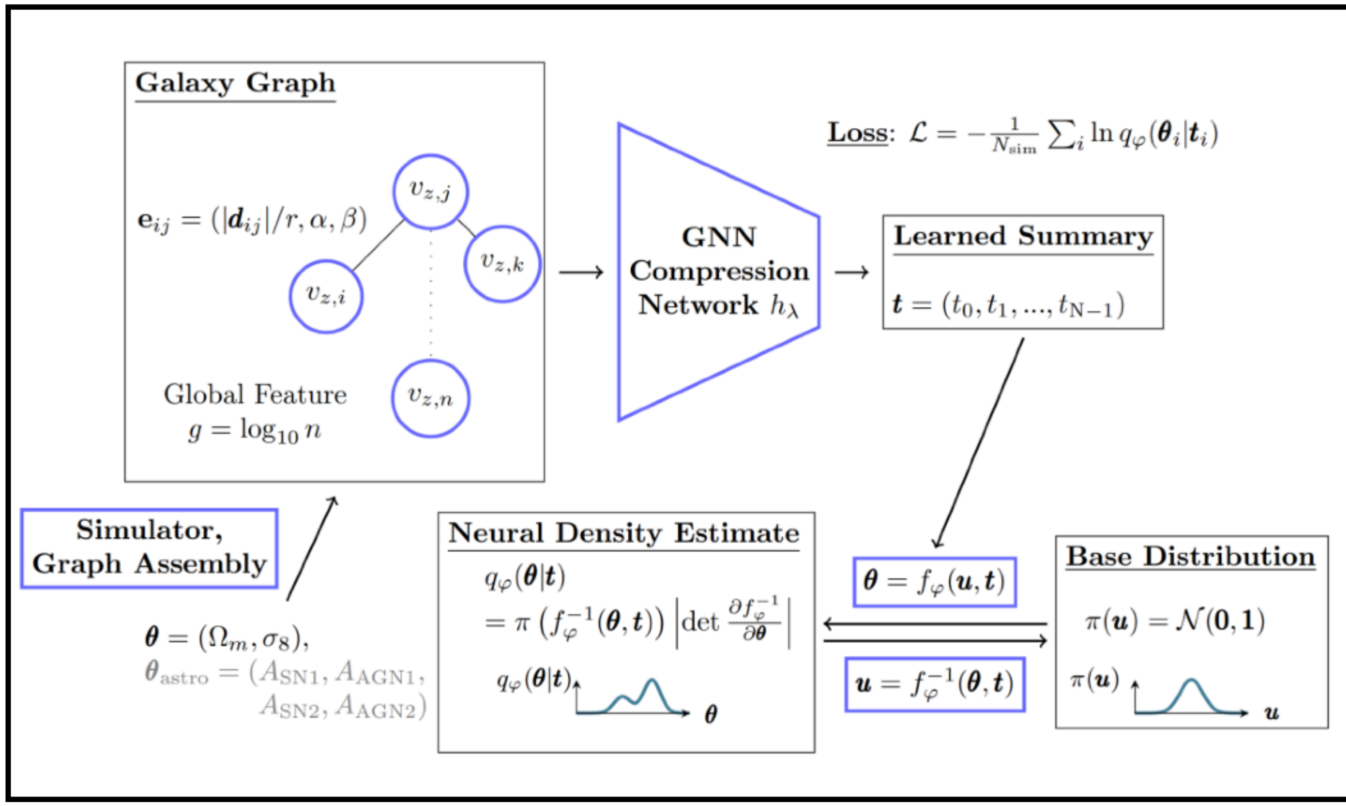
Targets

Ω_m — Matter content

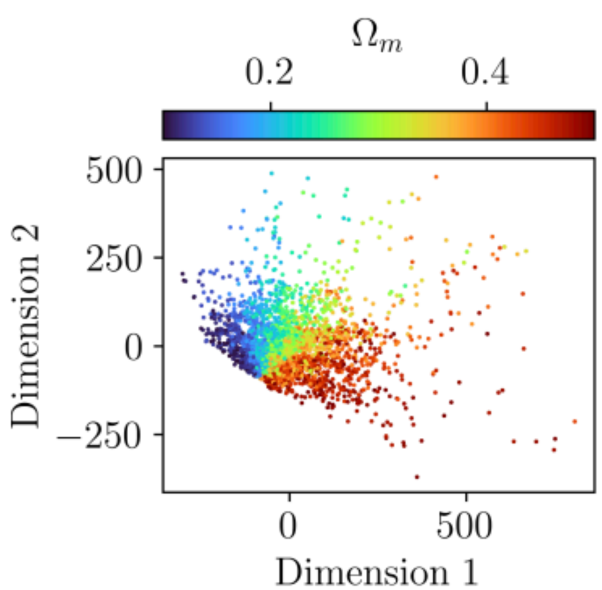
σ_8 — Density fluctuations

$\underbrace{A_{\mathrm{AGN1}} \quad A_{\mathrm{AGN2}}}_{\text{BH feedback}}$

$\underbrace{A_{\mathrm{SN1}} \quad A_{\mathrm{SN2}}}_{\text{Stellar feedback}}$

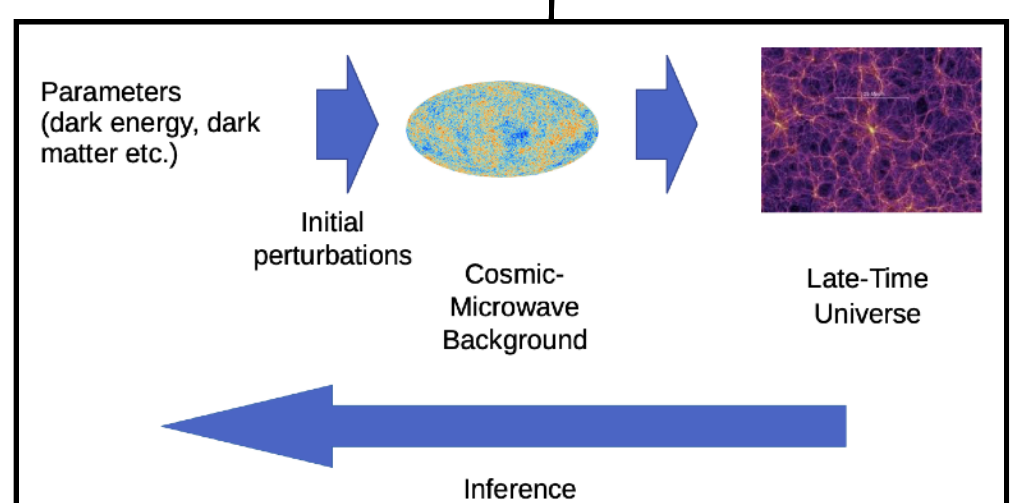


Learn parametrised surrogate of 3000 baryonic simulations, parametrised by 6 physics quantities



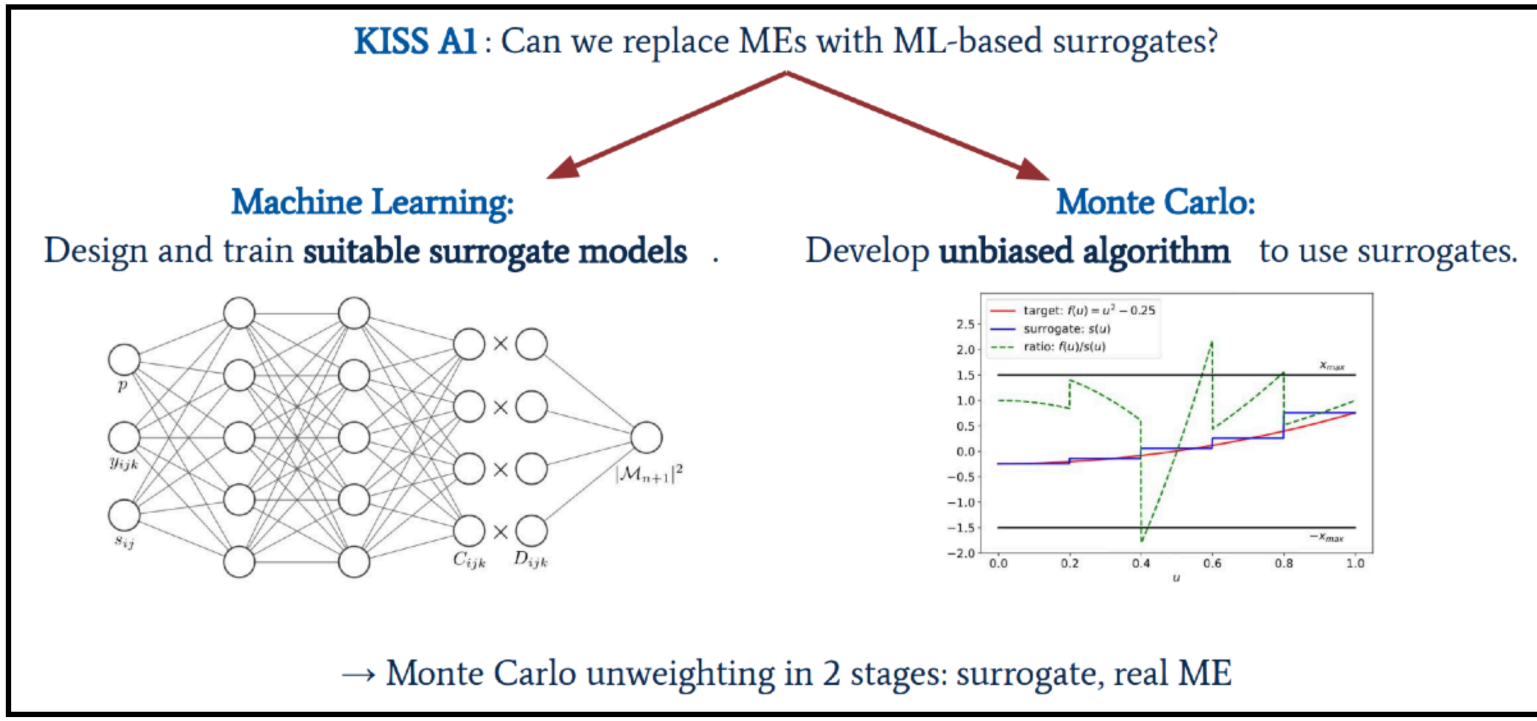
GNN summaries

Identify physics obs. in latent space

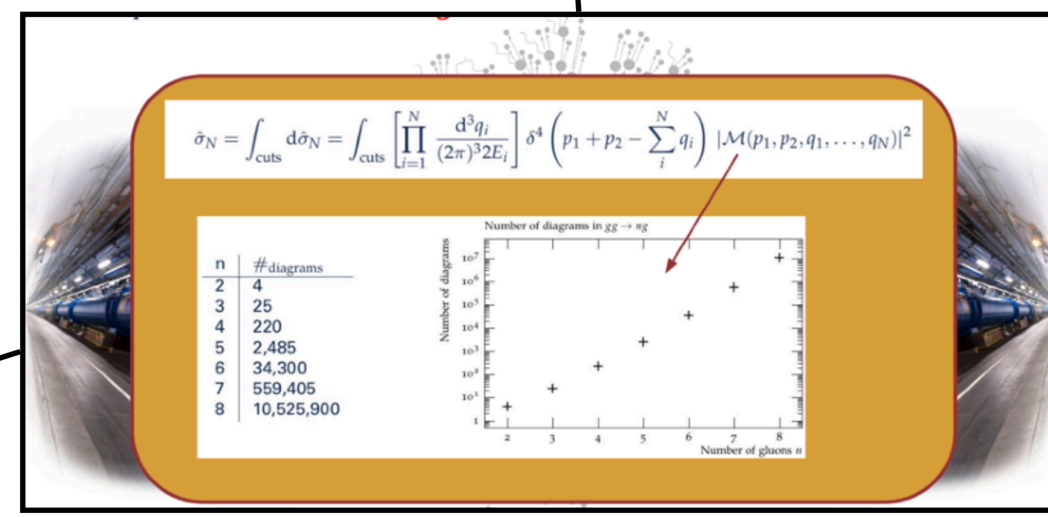


Distributions of galaxies

Targets



Surrogate + Sampling to remove bias



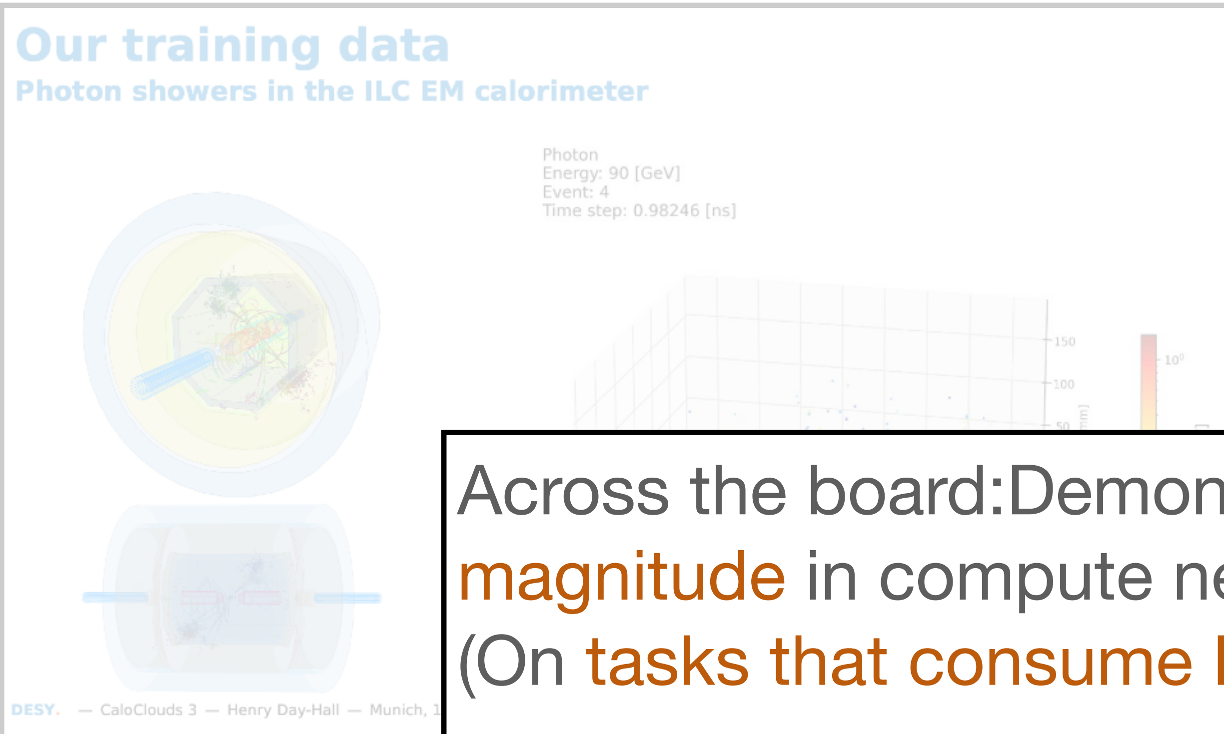
Hard-scattering matrix elements

Expect savings of O(50) Million CPU years!

Process	Events [M]	Summed [MHS23*y]	Sampled [MHS23*y]	Surrogates [MHS23*y]	**Ratio**
Z + ≤0 jets	53 797	0.03	0.03	-	
Z + ≤1 jets	164 479	0.10	0.10	-	
Z + ≤2 jets	456 960	0.29	0.30	-	
Z + ≤3 jets	835 797	0.58	0.74	-	
Z + ≤4 jets	1 164 974	1.16	2.25	-	
Z + ≤5 jets	1 381 719	**8.93**	18.08	**3.10**	**2.8**
Z + ≤6 jets	1 505 067	161.32	**51.68**	**7.57**	**6.7**

Targets

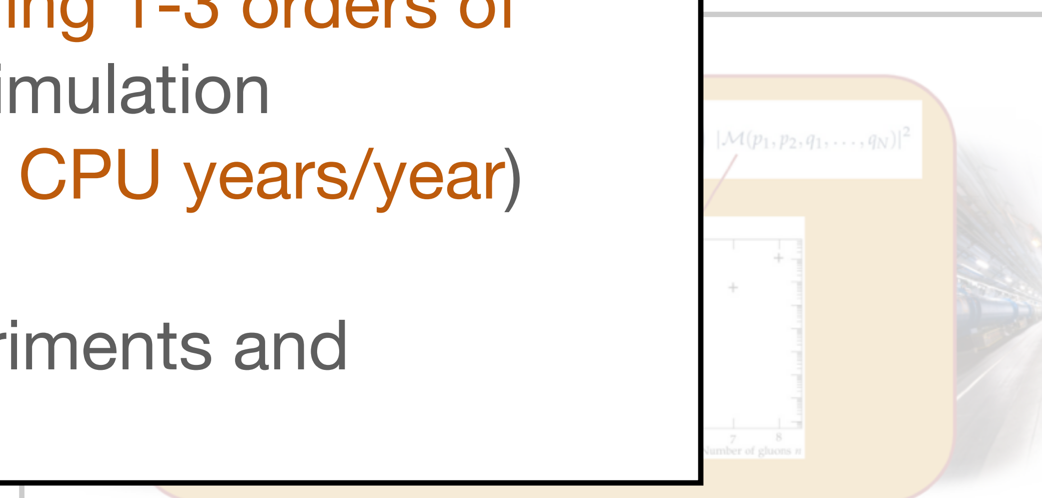
Heavy ion collisions

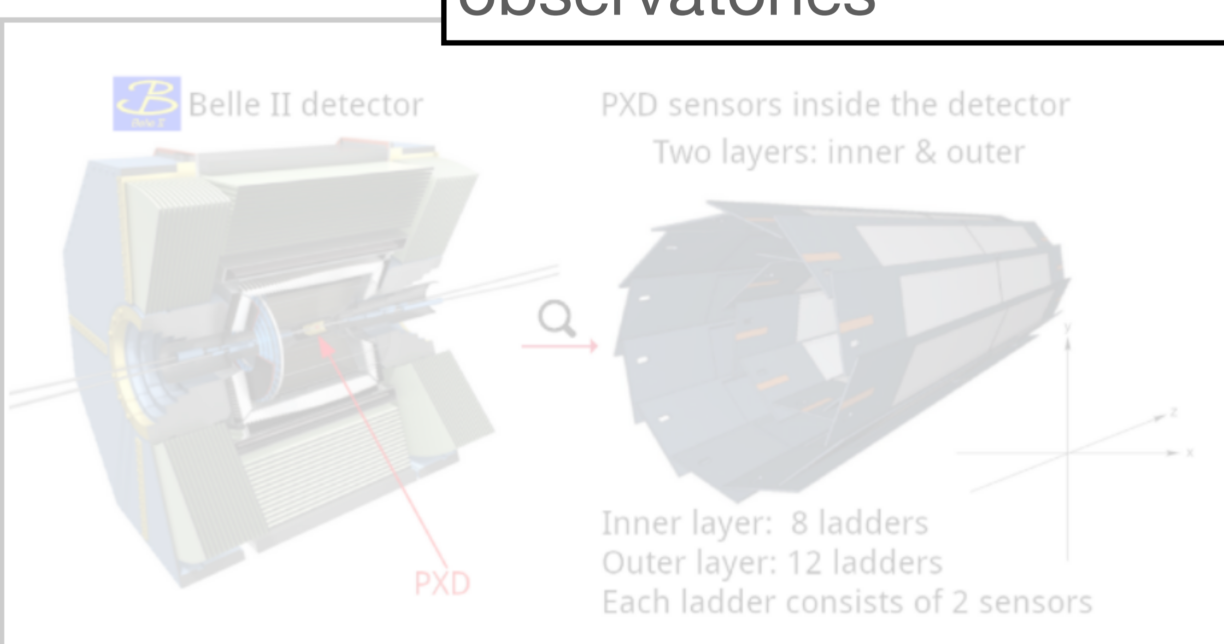


Particle showers

Across the board:Demonstrate saving 1-3 orders of magnitude in compute needs for simulation
(On tasks that consume Millions of CPU years/year)

Required for next generation experiments and observatories

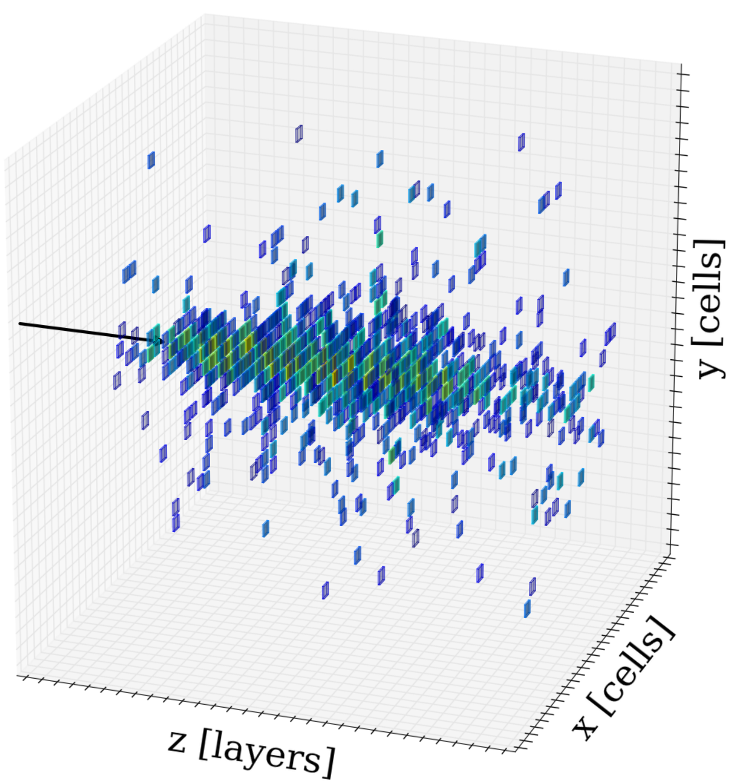

Hard-scattering matrix elements



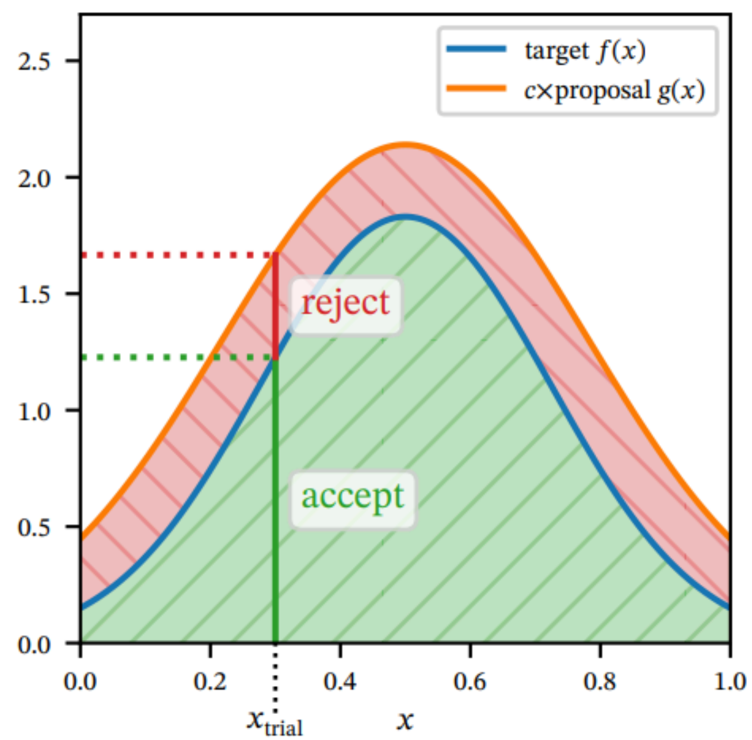
Background hits in pixel sensors

Distributions of galaxies

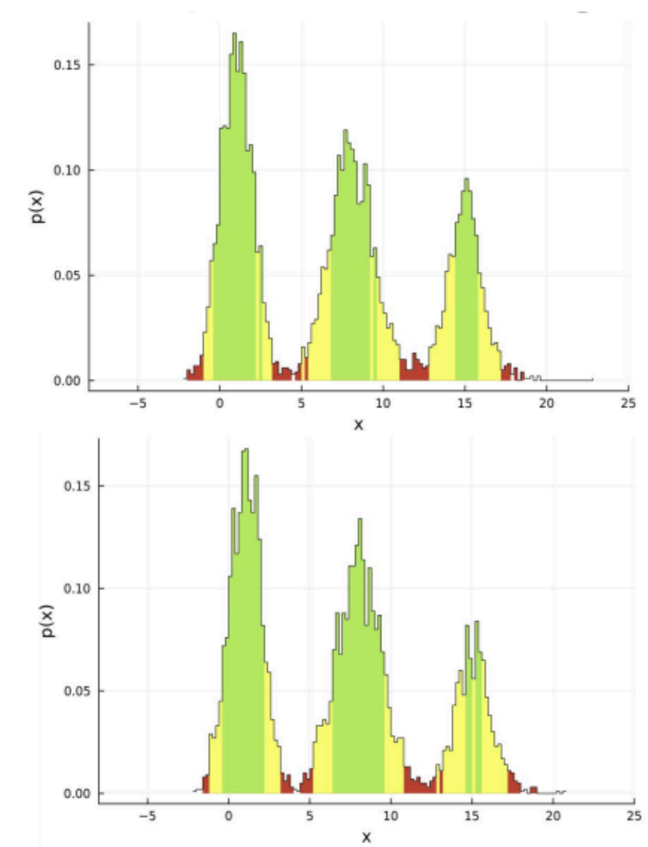
Overview



Surrogates

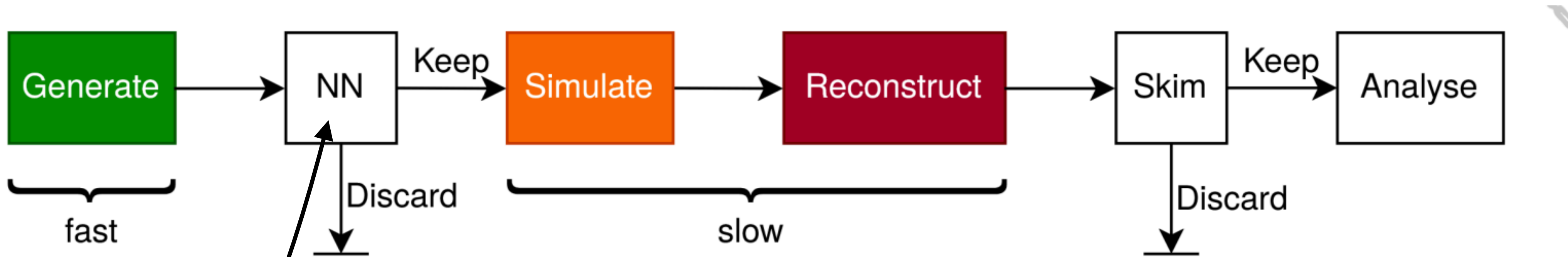


Sampling

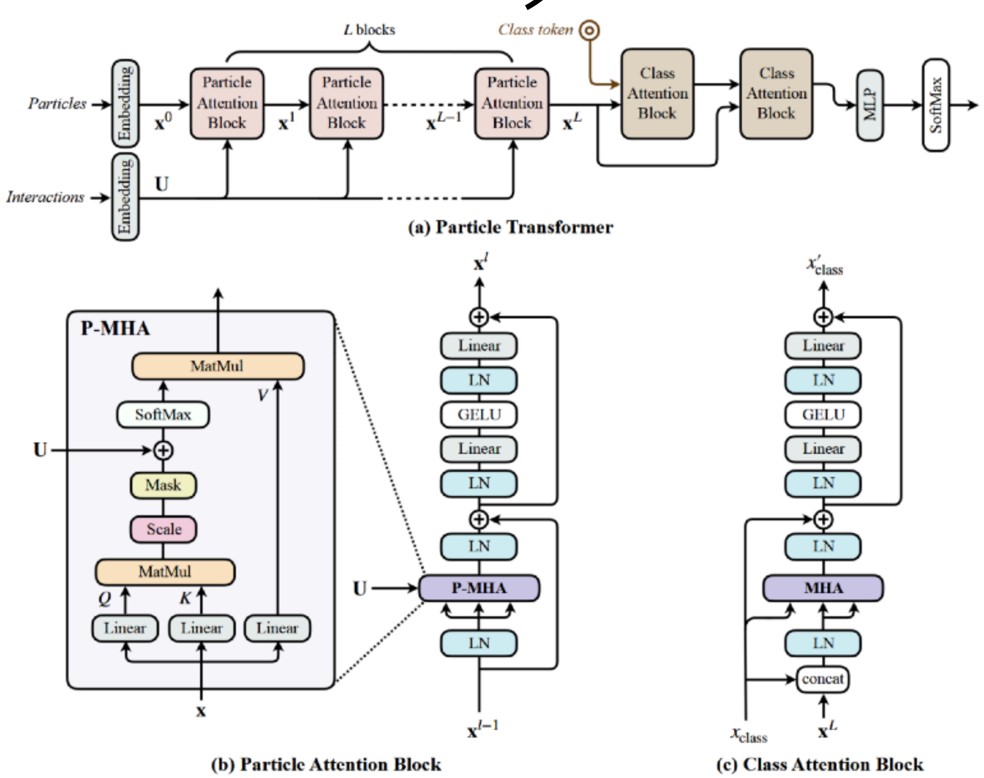


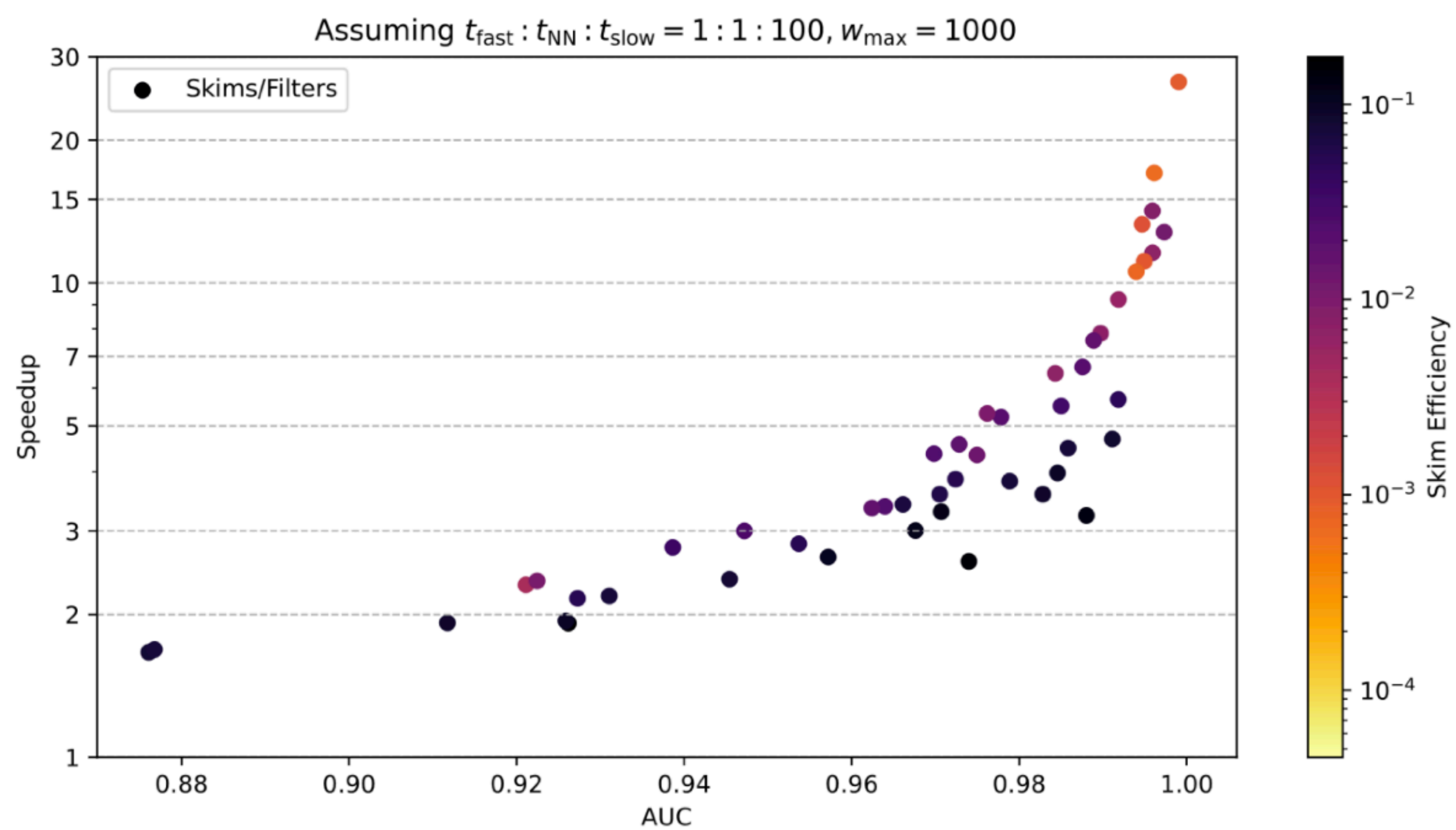
Quality

Smart background simulation



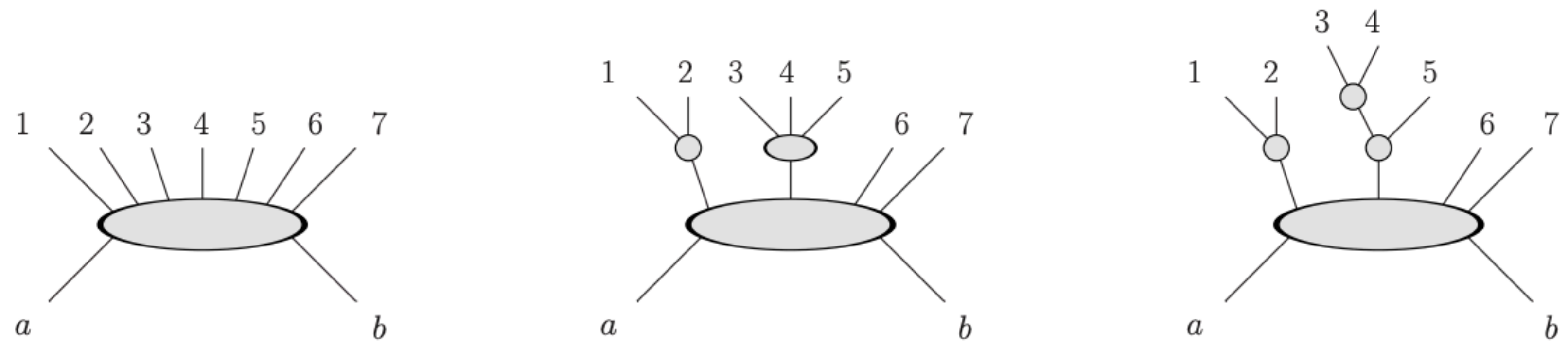
Transformer



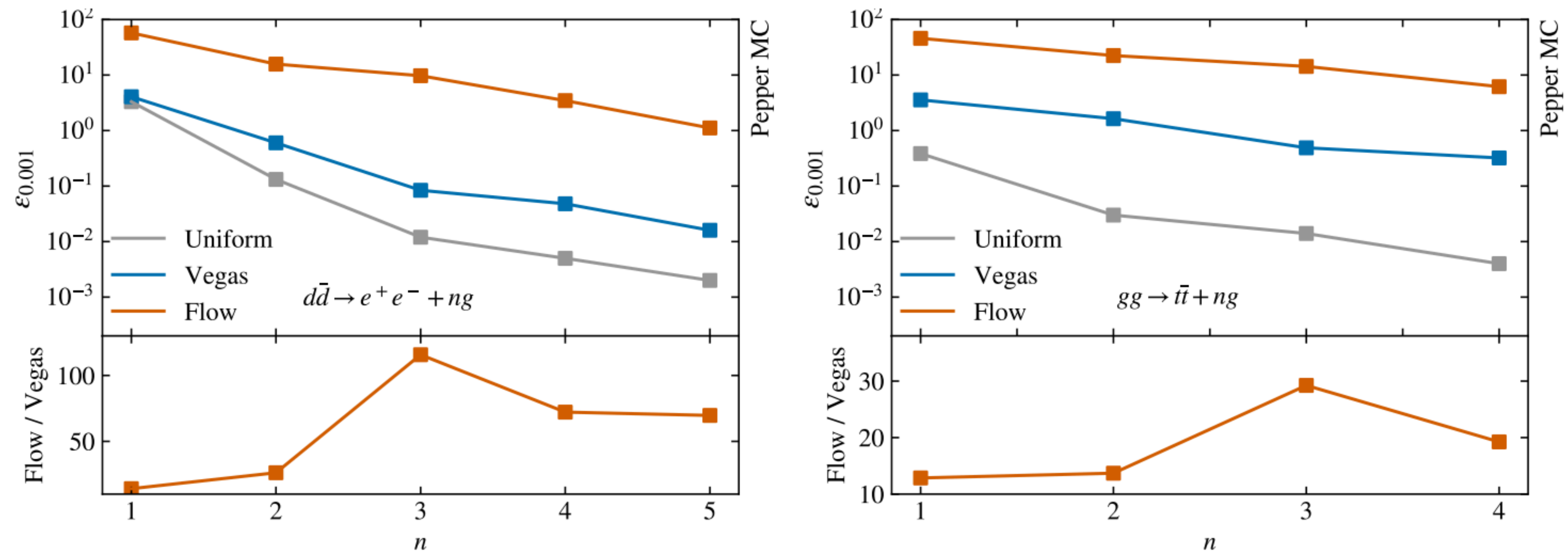


Speed-up depends on physics selection: ~1 order of magnitude

Smart background simulation



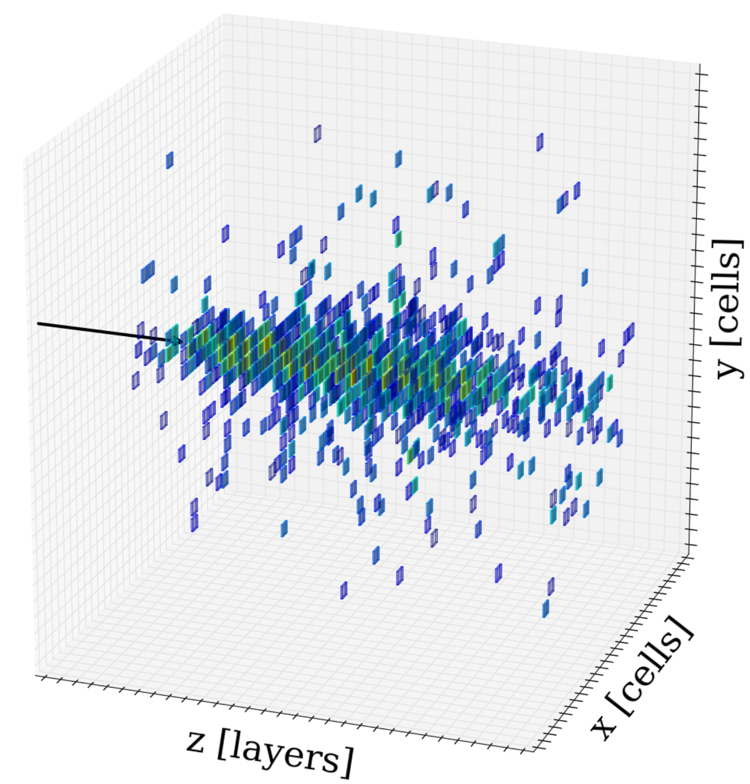
Use normalising flows to predict sampling distributions at LO/NLO



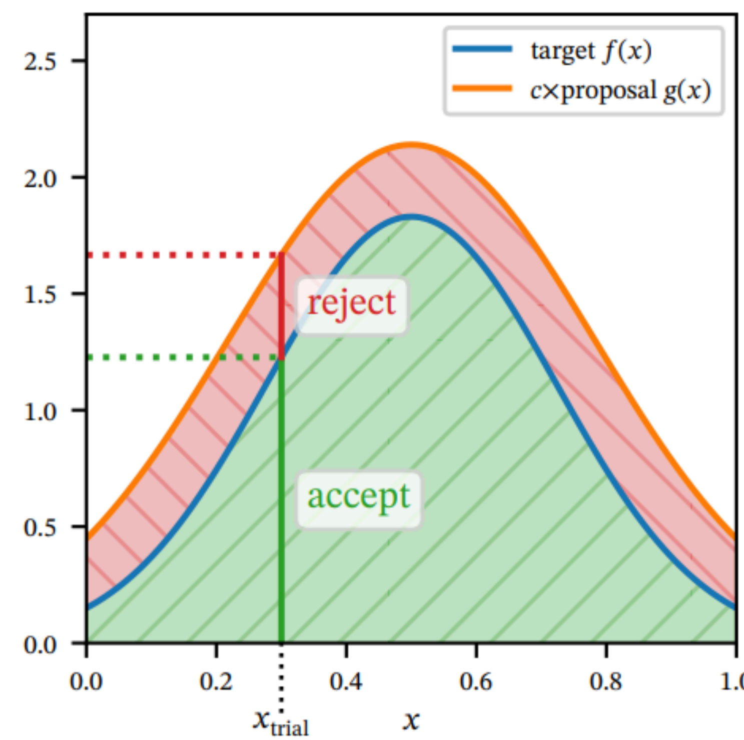
- significantly increased unw. eff. for all multiplicities
- factors up to 115
- seems to scale well

Again, 1-2 order of magnitude

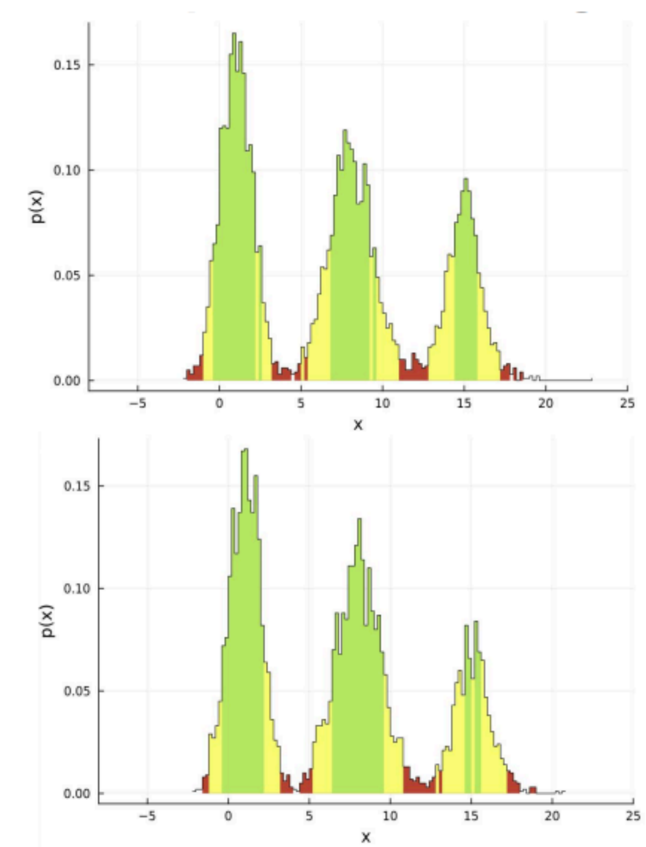
Overview



Surrogates

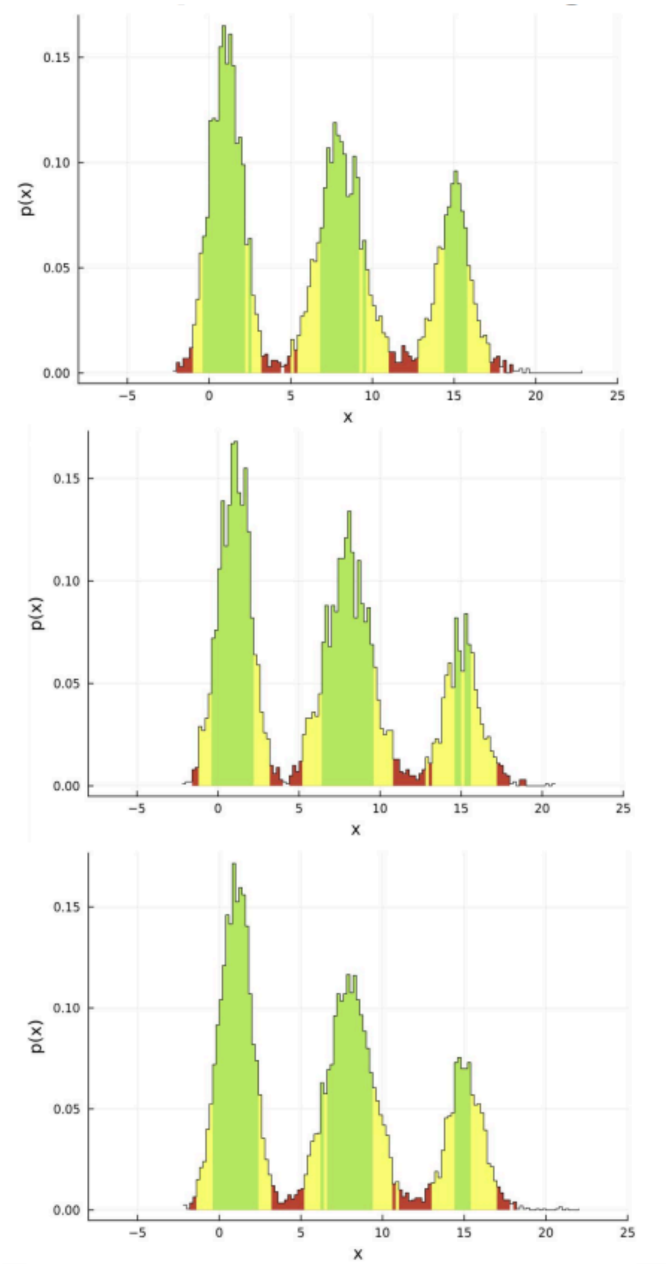


Sampling

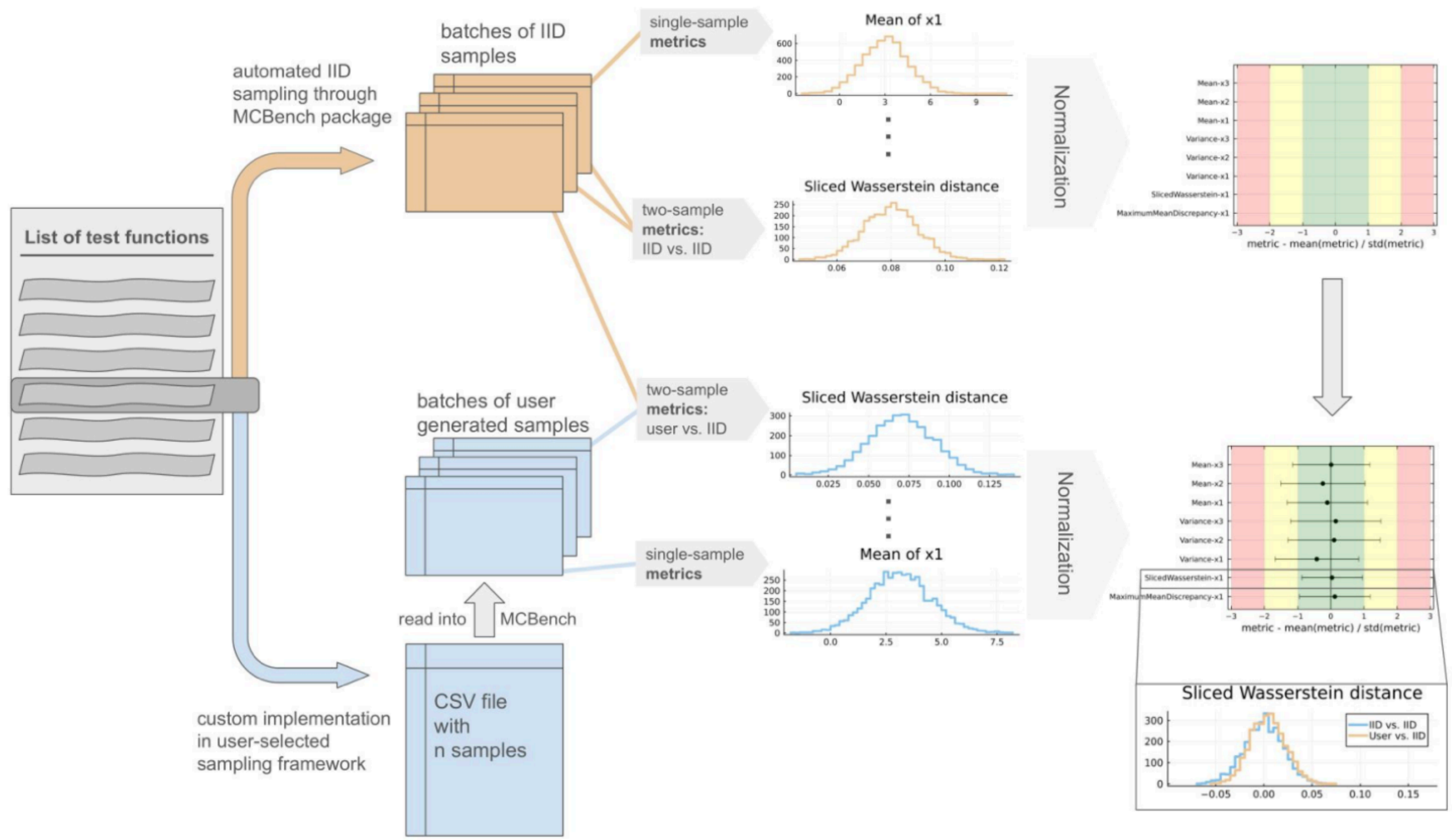


Quality

Foundational Topics



Which samples are drawn from the same underlying distribution?



Build automated tools for comparing and benchmarking distributions

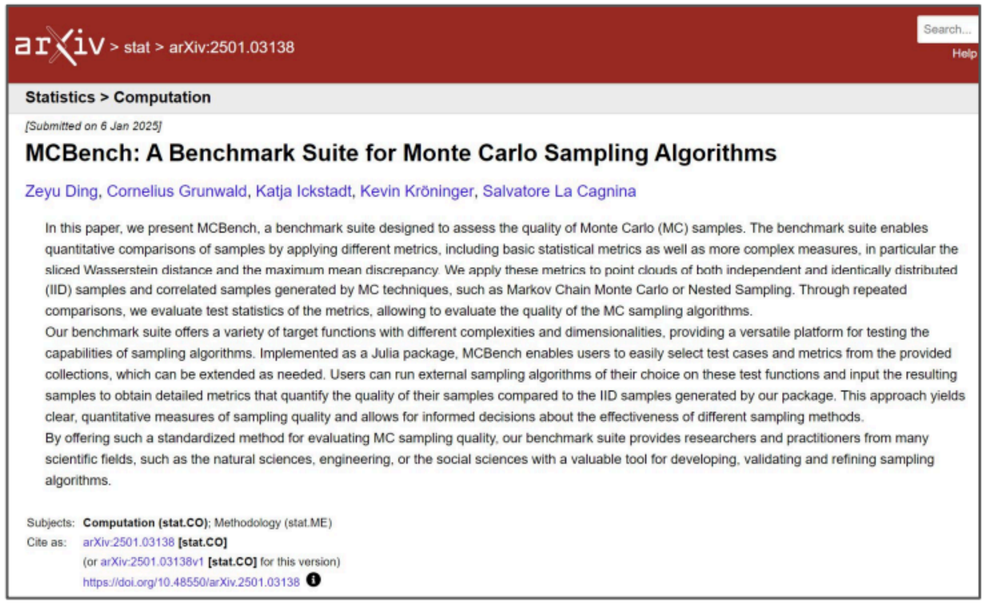
Foundational Topics

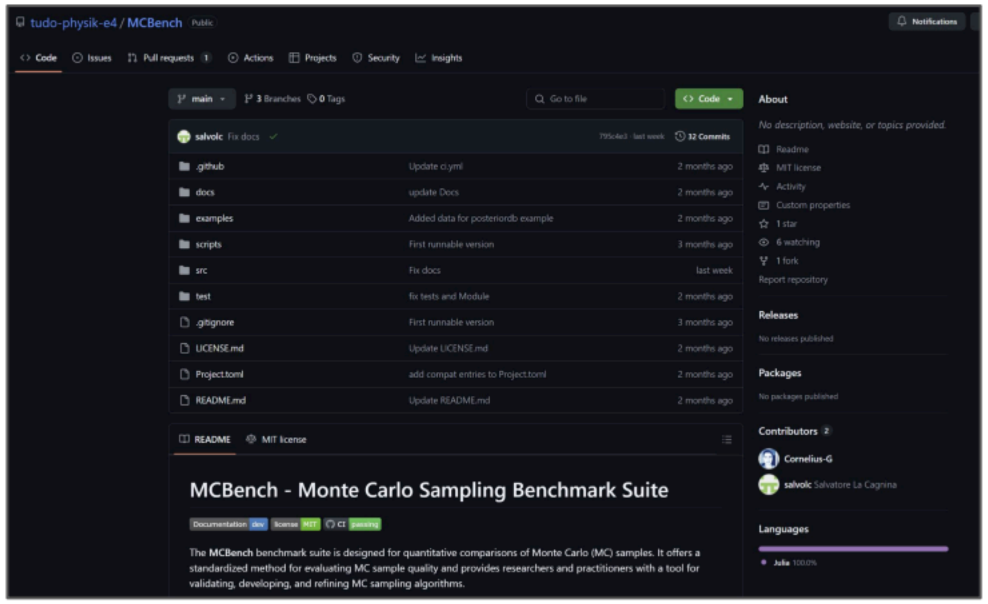
Summary and Conclusions

- Developed a test suite (in the julia programming language
- Compare samplers to IID samples using metrics
- Provide a selection of (IID sampleable) test functions and (one and two-sample) metrics
- Visit our suite on github and paper on arxiv
  - https://github.com/tudo-physik-e4/MCBench
  - https://arxiv.org/abs/2501.03138

Next Steps:

- Add full test case support for different platforms (R, stan, pymc) including testpoints
- Lookout to include more complex test cases and applications



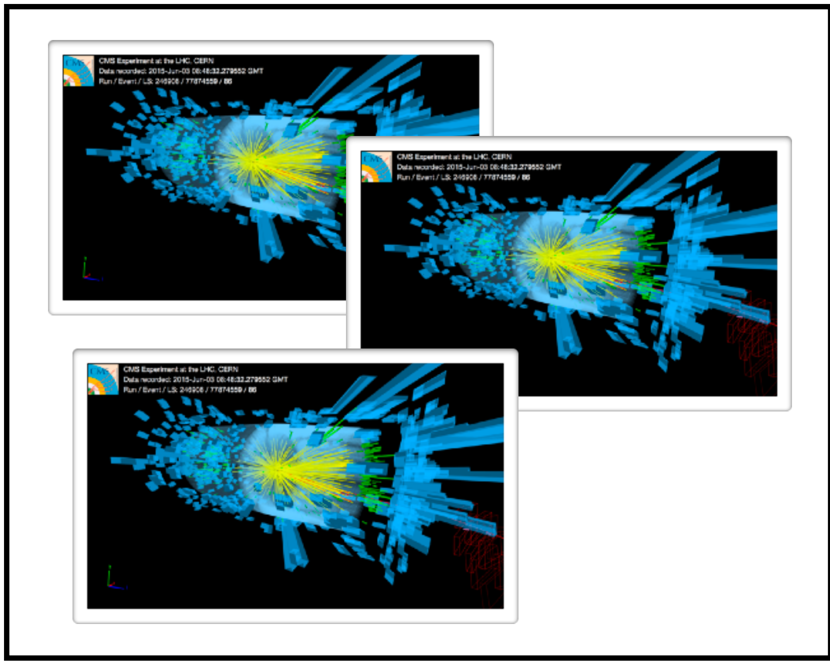


Paper & code now publically available!

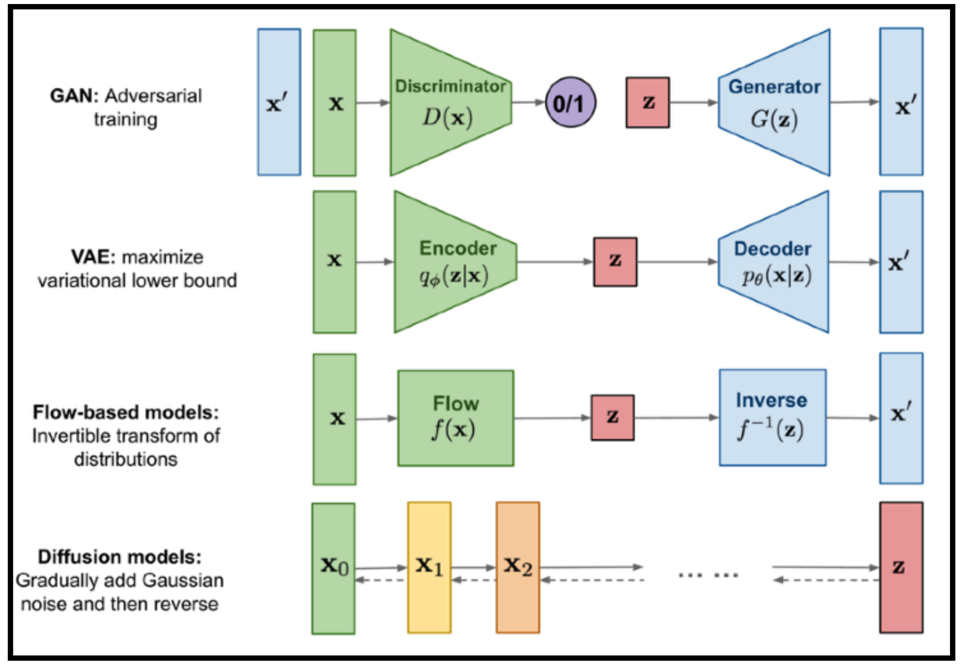
1. Use classical simulation or data as input

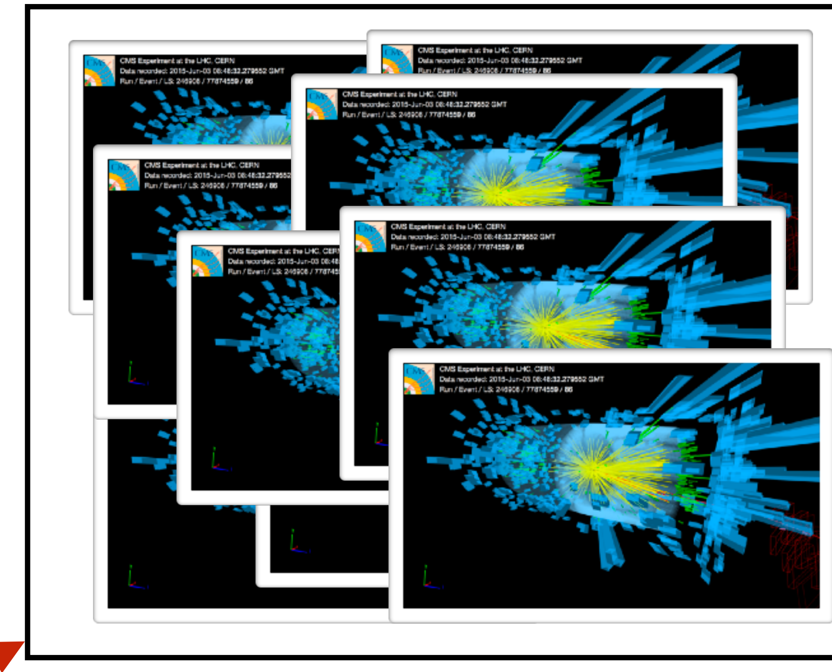
2. Train generative surrogate

3. Oversample



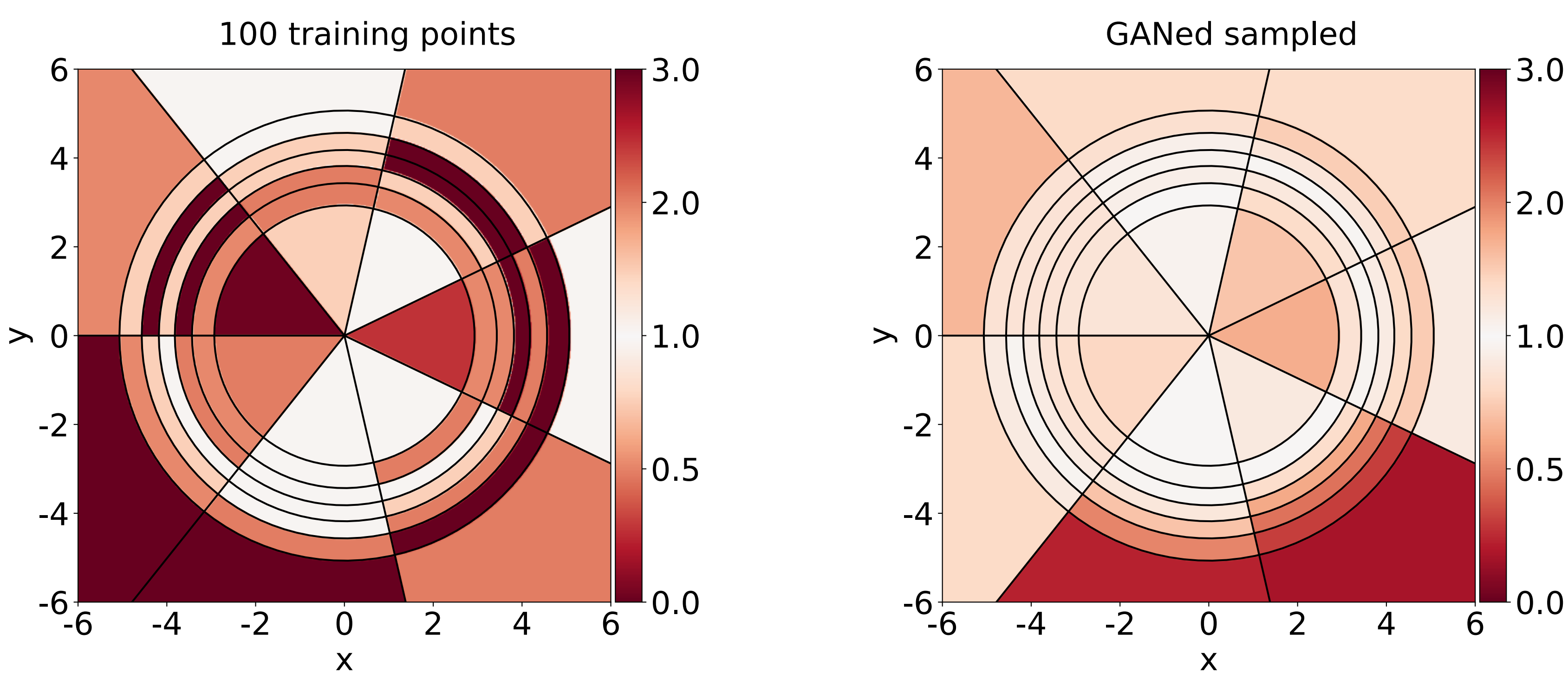
(slow)





(**fast**)

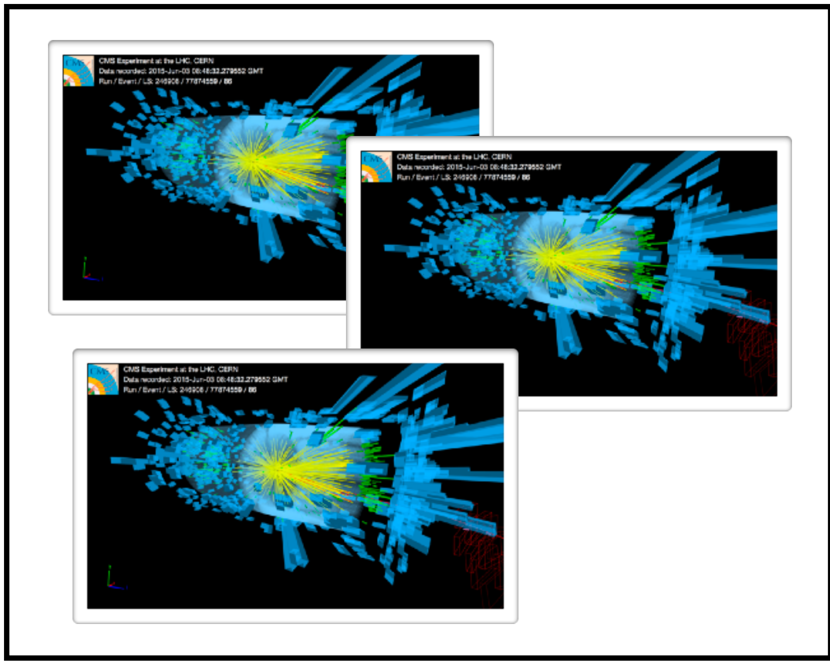
Does this even make sense?



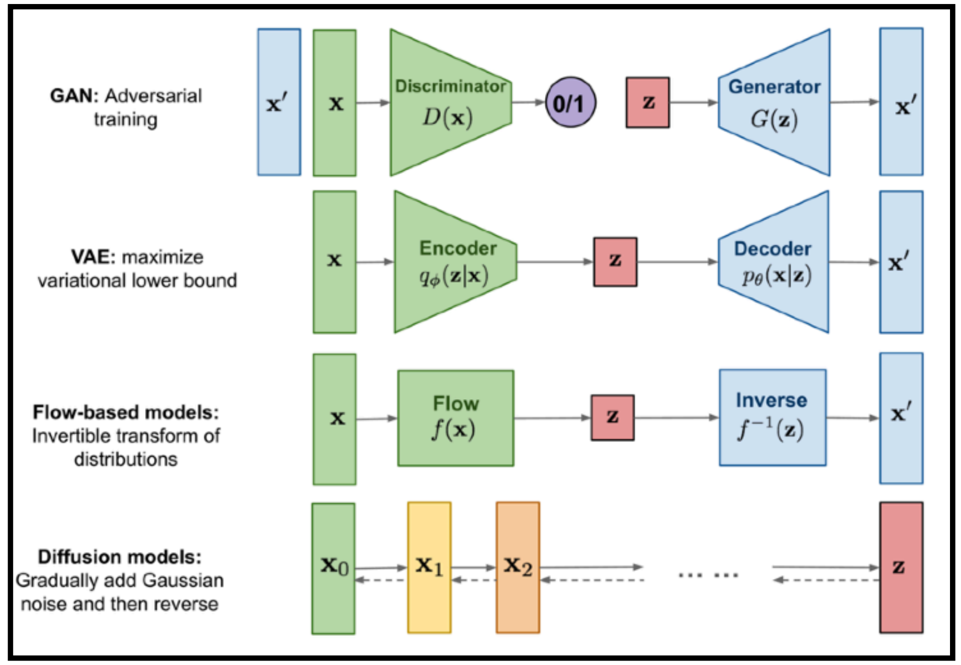
1. Use classical simulation or data as input

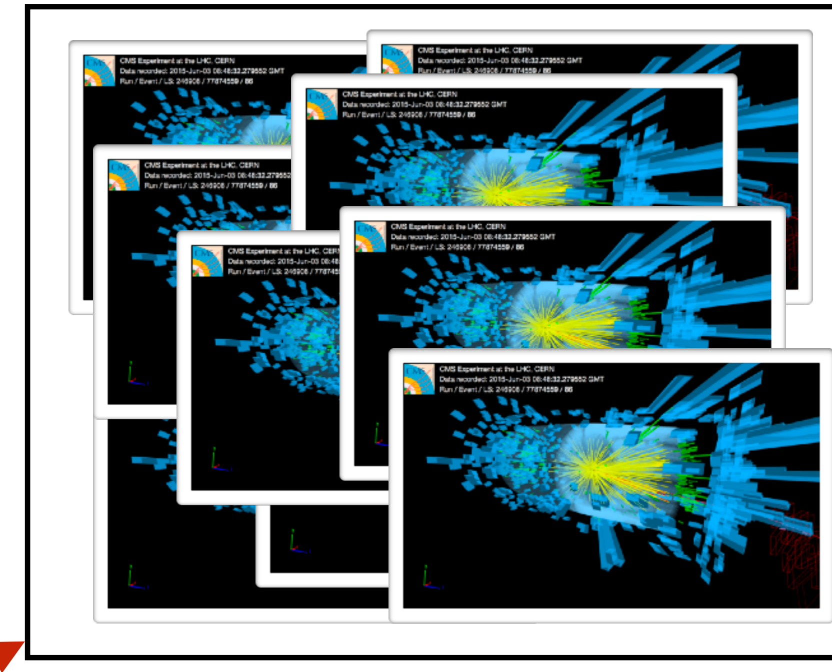
2. Train generative surrogate

3. Oversample



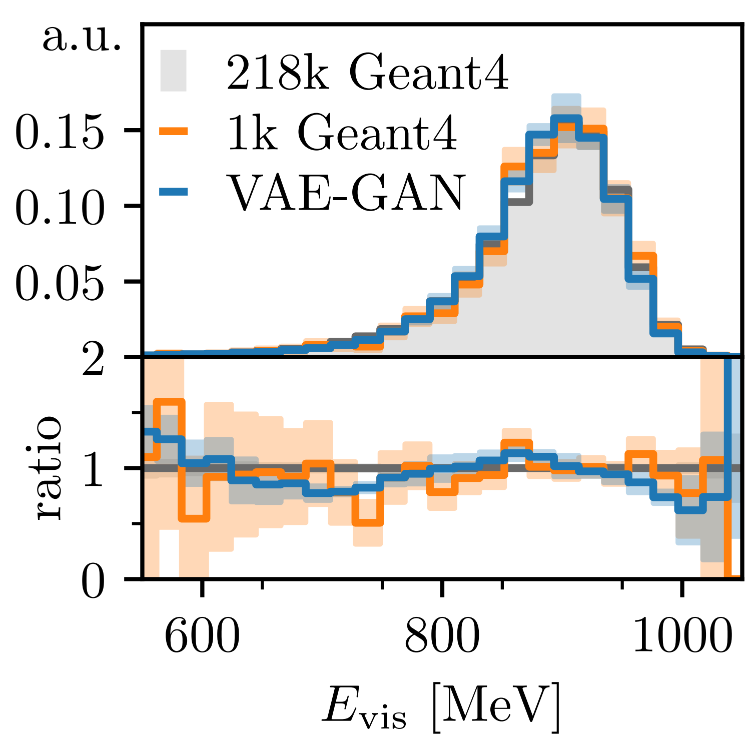
(slow)

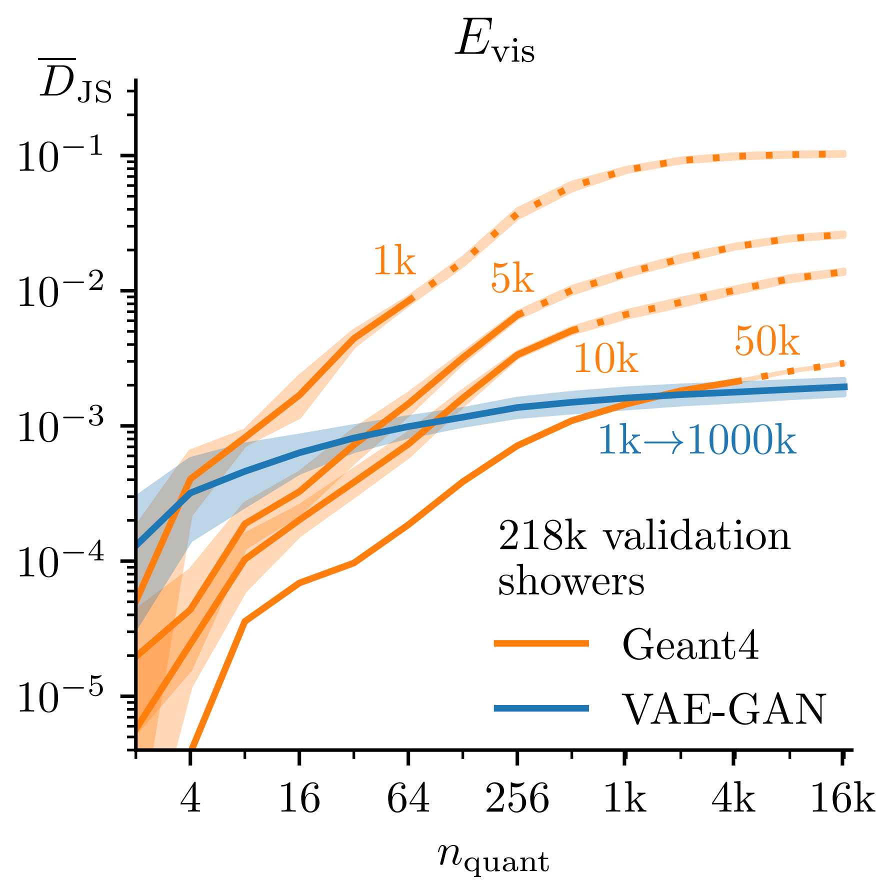




(**fast**)

Does this even make sense?: Yes!





Scaling of difference to ground truth with resolution again better for the generative model.

Follow-up work witg statisticians

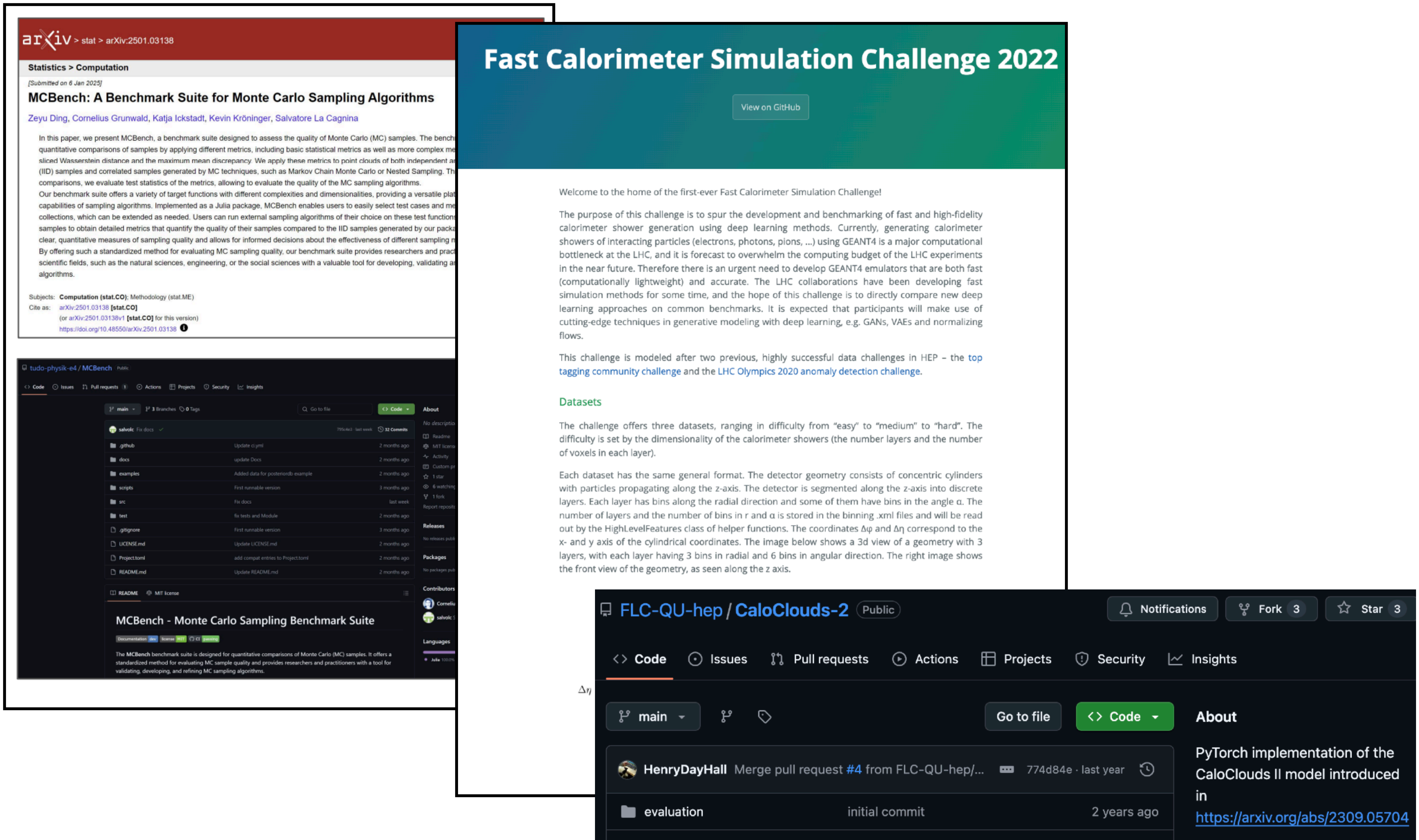
Closing

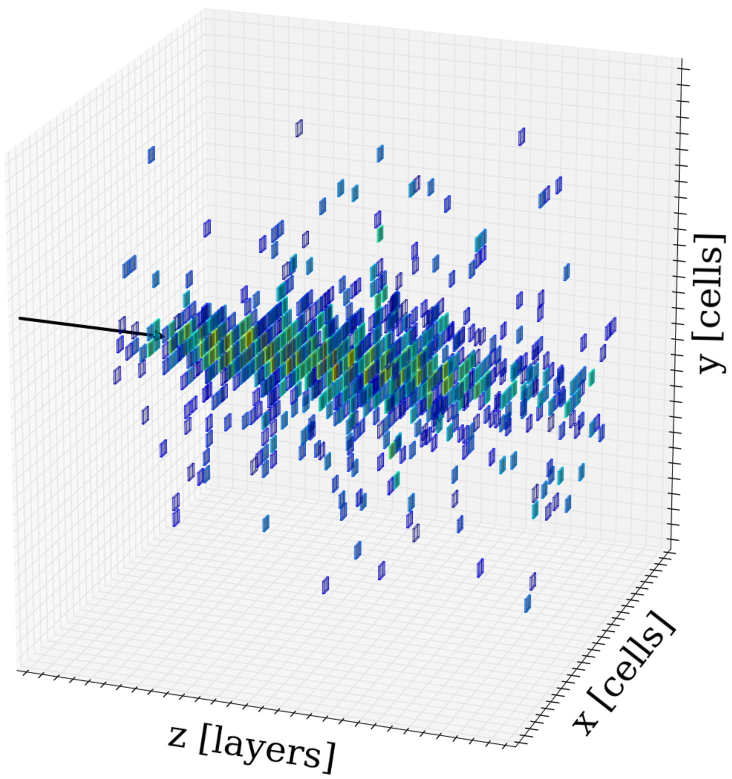
New angles on fast calorimeter shower simulation	https://inspirehep.net/literature/2647716
CaloClouds: fast geometry-independent highly-granular calorimeter	https://inspirehep.net/literature/2657637
CaloClouds II: ultra-fast geometry-independent highly-granular calorimeter	https://inspirehep.net/literature/2696622
EPiC-ly Fast Particle Cloud Generation with Flow-Matching and Diffusion	https://inspirehep.net/literature/2705220
Flow Matching Beyond Kinematics: Generating Jets with Particle-ID and	https://inspirehep.net/literature/2729197
Simulating images of radio galaxies with diffusion models	https://ui.adsabs.harvard.edu/abs/2024arXiv241007794V/abstract
Deep-learning-based radiointerferometric imaging with GAN-aided training	https://ui.adsabs.harvard.edu/abs/2023A%26A...677A.167G/abstract
Convolutional L2LFlows: generating accurate showers in highly granular	https://inspirehep.net/literature/2793021
Calibrating Bayesian Generative Machine Learning for Bayesiamplification	https://inspirehep.net/literature/2814095
How to understand limitations of generative networks	https://inspirehep.net/literature/2663017
Precision-Machine Learning for the Matrix Element Method	https://inspirehep.net/literature/2709868
The MadNIS reloaded	https://inspirehep.net/literature/2718812
Kicking it Off(-shell) with Direct Diffusion	https://inspirehep.net/literature/2727894
Normalizing Flows for High-Dimensional Detector Simulations	https://inspirehep.net/literature/2737698
The Landscape of Unfolding with Machine Learning	https://inspirehep.net/literature/2781602
CaloDREAM -- Detector Response Emulation via Attentive flow Matching	https://inspirehep.net/literature/2787493
Lorentz-Equivariant Geometric Algebra Transformers for High-Energy	https://inspirehep.net/literature/2789600
Differentiable MadNIS-Lite	https://inspirehep.net/literature/2814426
Machine learning study to identify collective flow in small and large colliding	https://inspirehep.net/literature/2660500
Building imaginary-time thermal field theory with artificial neural networks	https://inspirehep.net/literature/2787839
Phase Transition Study Meets Machine Learning	https://inspirehep.net/literature/2721834
Diffusion models as stochastic quantization in lattice field theory	https://inspirehep.net/literature/2704849
Mass and tidal parameter extraction from gravitational waves of binary	https://inspirehep.net/literature/2673529
Exploring QCD matter in extreme conditions with Machine Learning	https://inspirehep.net/literature/2646083
Improved selective background Monte Carlo simulation at Belle II with graph	https://inspirehep.net/literature/2676588
Ultra-high-granularity detector simulation with intra-event aware generative	https://inspirehep.net/literature/2642136
Improved selective background Monte Carlo simulation at Belle II with graph	https://indico.cern.ch/event/1253794/contributions/5588582/
QCD Equation of State of Dense Nuclear Matter from a Bayesian Analysis of	https://inspirehep.net/literature/2512939
Efficient phase-space generation for hadron collider event simulation	https://inspirehep.net/literature/2630465
A Portable Parton-Level Event Generator for the High-Luminosity LHC	https://inspirehep.net/literature/2721108
Unweighting multijet event generation using factorisation-aware neural	https://inspirehep.net/literature/2628385
Development of the time-of-flight particle identification for future Higgs	https://inspirehep.net/literature/2720464
Improving Monte Carlo simulations in high energy physics using machine	https://inspirehep.net/literature/2698487
Event generation with Sherpa 3	https://inspirehep.net/literature/2843469
Learning Optimal and Interpretable Summary Statistics of Galaxy Catalogs	https://inspirehep.net/literature/2848384
MCBench: A Benchmark Suite for Monte Carlo Sampling Algorithms	https://arxiv.org/abs/2501.03138
Full phase space resonant anomaly detection	https://arxiv.org/abs/2310.06897
Generative Diffusion Models for Lattice Field Theory	https://arxiv.org/abs/2311.03578
Accurate Surrogate Amplitudes with Calibrated Uncertainties	https://inspirehep.net/literature/2860406
Phase space sampling with Markov Chain Monte Carlo methods	https://inspirehep.net/literature/2860425
Advancing Tools for Simulation-Based Inference	https://inspirehep.net/literature/2838939

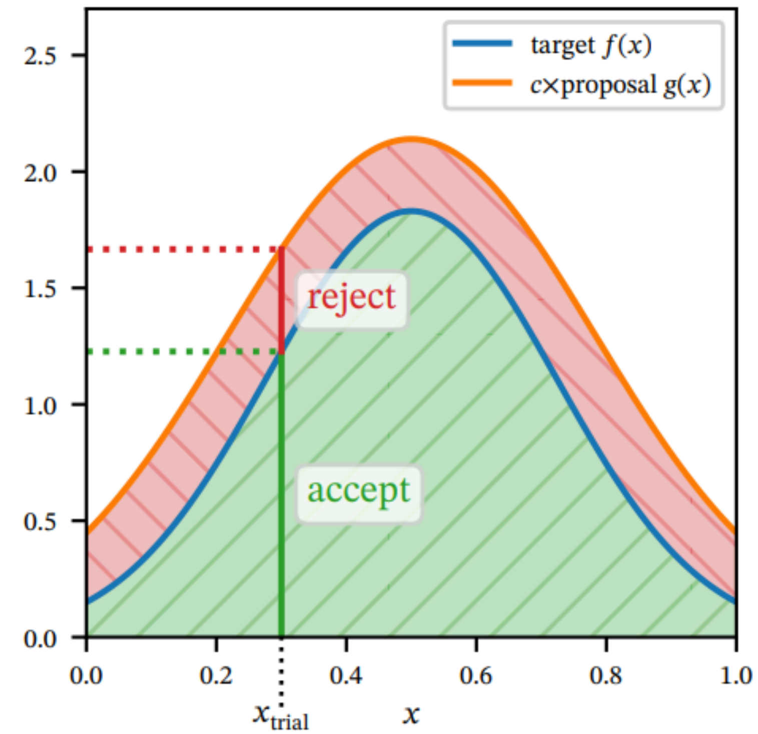
40+ KISS papers since 1.3.2023

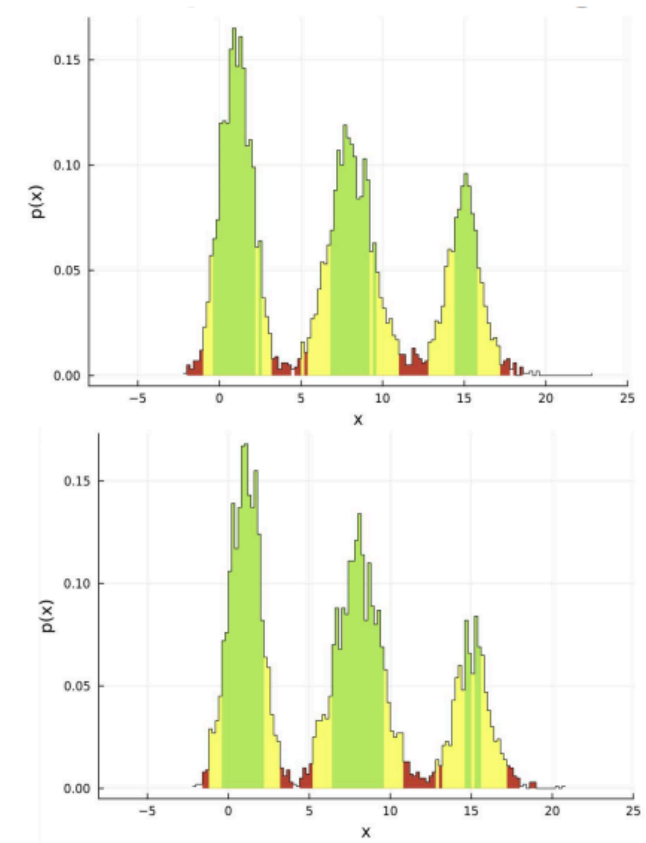
https://kiss.pages.desy.de/website/page/activities/

Code & data for KISS projects









Surrogates

Sampling

Quality

Major improvements (x10-x100) of efficiency for key simulation needs

Substantial publications and public data & code

Generally usable tools and foundational work with statistics

KISS