

THIS WAY

TO THE

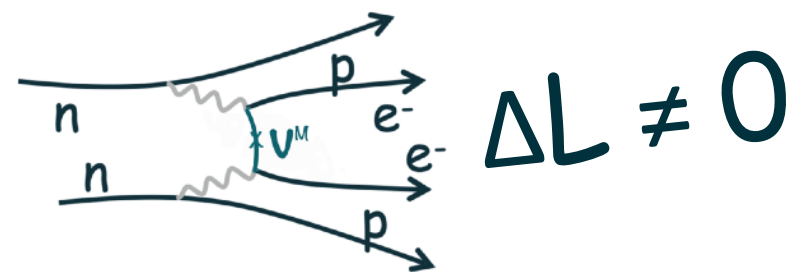
UNIVERSE

A Journey
into Physics

From LEGEND's Neutrons to LIGO's Binary Black Holes:

A Rare-Event Journey Across Physics

Ann-Kathrin Schuetz
Lawrence Berkeley National Laboratory
12/10/2025



Finding a needle in a haystack...



"Rare Events across Physics"

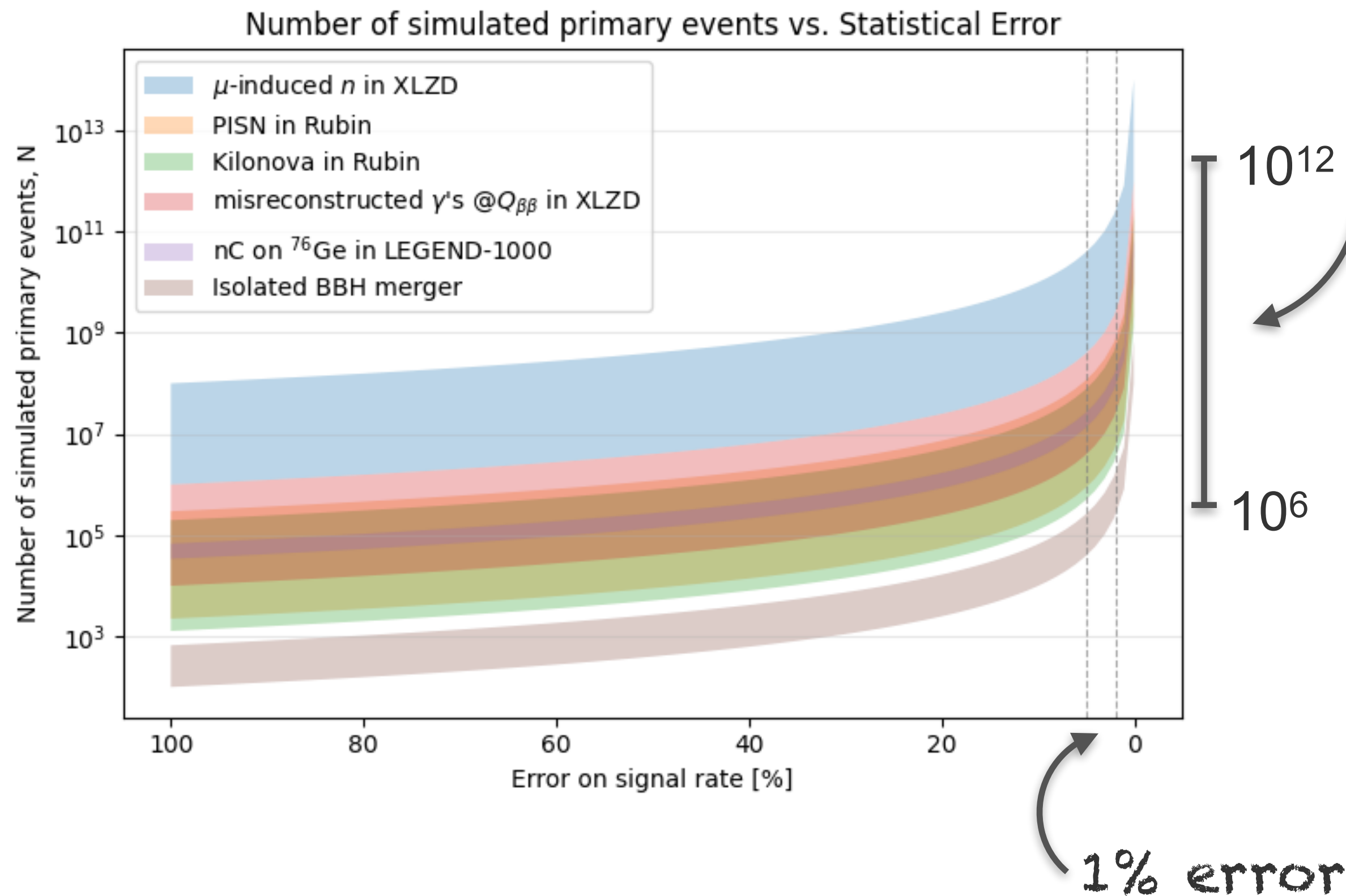
Examples:

- Direct detection of WIMP dark matter
 $< O(1)$ events/(30 kg years)
- Neutrinoless double beta decay ($0\nu\beta\beta$)
 $< O(1)$ events/(500 kg year)
- Neutrino coherent scattering (CEvNS)
 $\sim O(1)$ events/(10 kg years)
- Binary Black Hole Mergers

Every field has a needle – only the haystack changes.

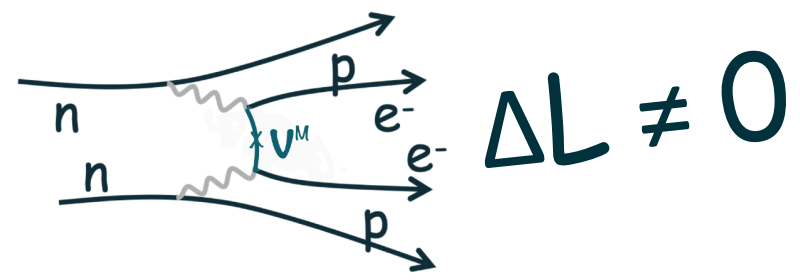
Rare Event Simulations in Fundamental Physics

Finding a needle in a haystack...



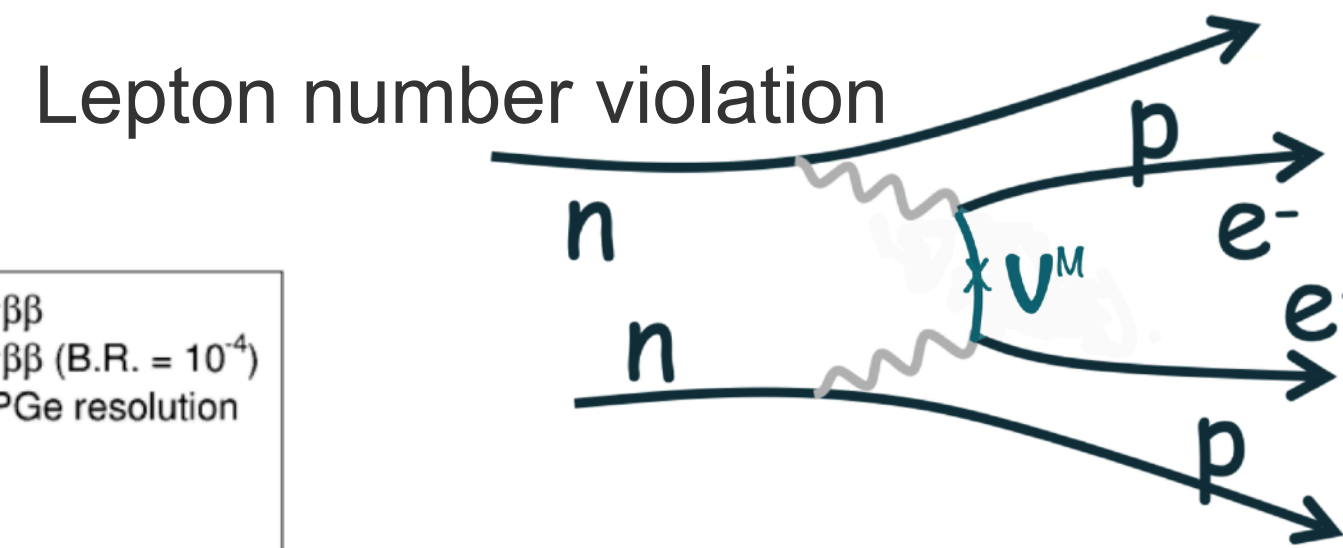
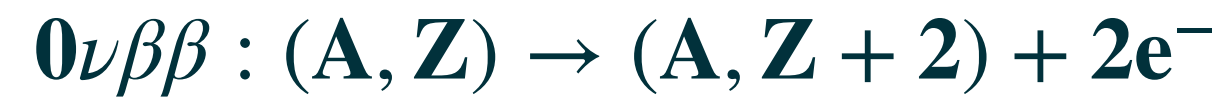
Scaling Rare-Event Simulations: The Compute Bottleneck

- 1% precision $\Rightarrow 10^6$ – 10^{12} events per configuration
- Large search space (optimization, inference,...): $\Rightarrow 10^8$ – 10^{15} total events
- 0.001–1 CPU-sec/event
 \Rightarrow useful event cost 10^3 – 10^{12} CPU-sec
 \Rightarrow using 100 cores: 10^{-3} – 10^{-10} rare events per hour
- Traditional workflows become **compute-bound** and **too slow to iterate**
- **Surrogate models** are required to
 - cut simulation cost by orders of magnitude
 - enable fast optimization and inference at scale



Example: Bayesian Inference and $0\nu\beta\beta$ decay

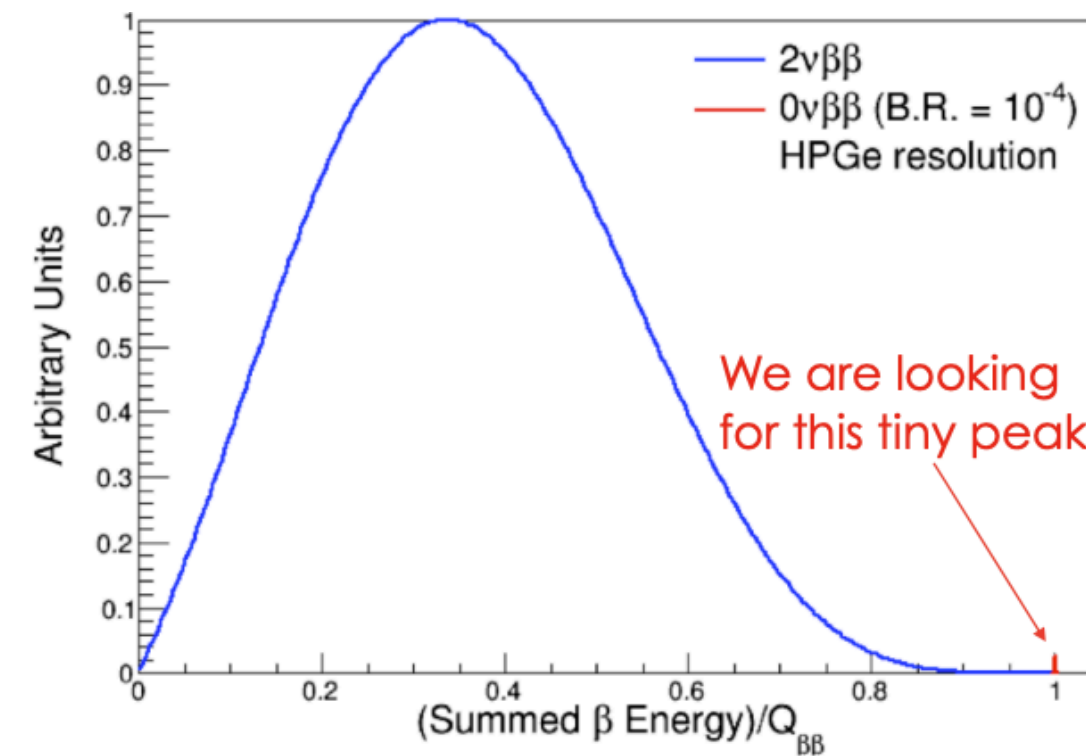
Experimental goal is to measure mono-energetic peak at $Q_{\beta\beta}$



Experimental sensitivity:

background (BI) > 1:

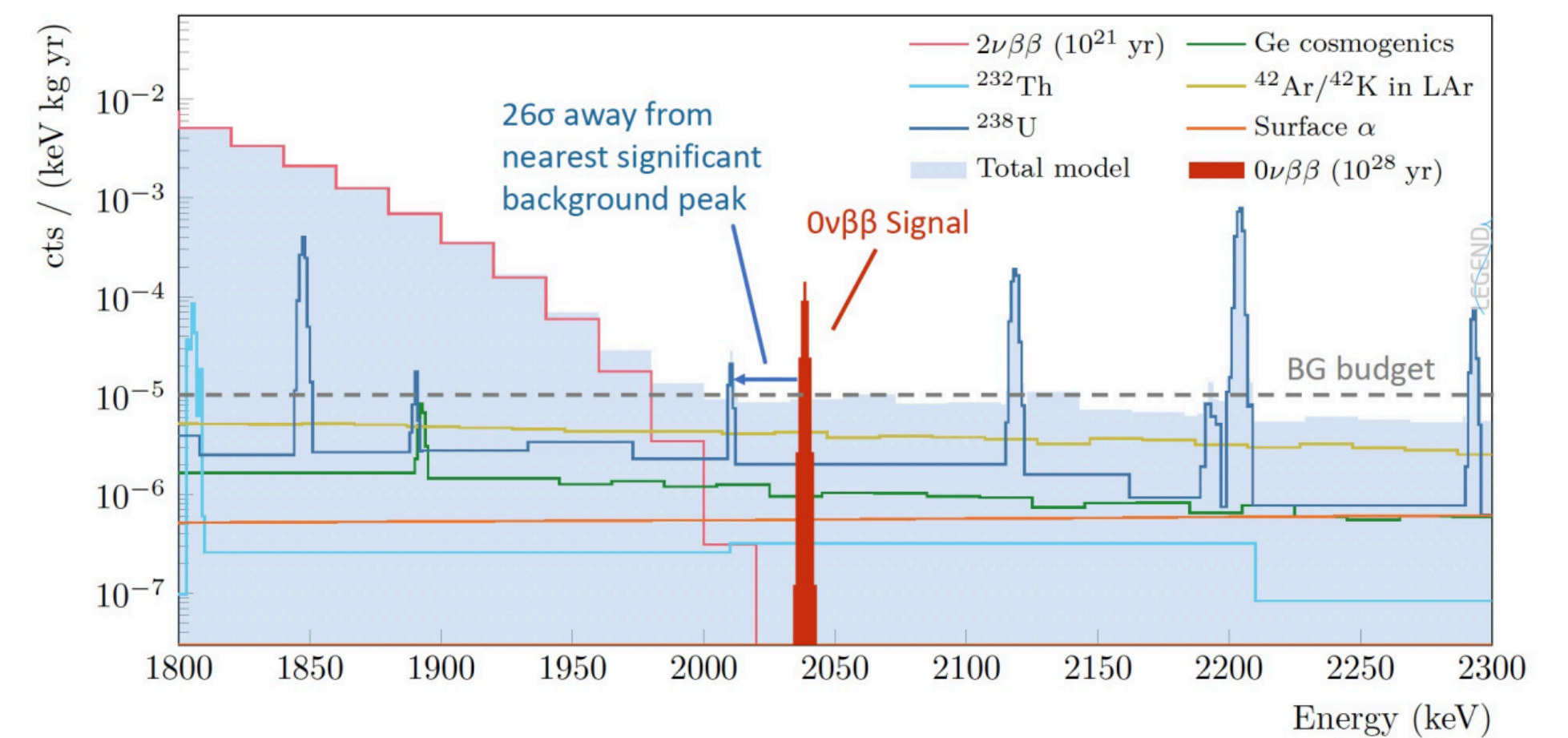
$$T_{1/2}^{0\nu} \propto \varepsilon \cdot a \cdot \sqrt{\frac{M \cdot t}{BI \cdot \Delta E}}$$



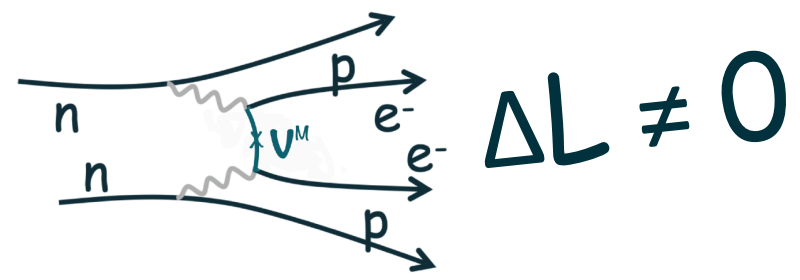
measure sum energy spectrum of electrons

- $2\nu\beta\beta \rightarrow$ continuum
- $0\nu\beta\beta \rightarrow$ mono-energetic peak @ $Q_{\beta\beta}$

But this signal is buried under other backgrounds...



→ increase sensitivity by **background reduction (BI)** at $Q_{\beta\beta}$ and simultaneous increase of mass (M) and improvement of the energy resolution (ΔE)



Background reduction for LEGEND-1000

0νββ decay - Experimental sensitivity

$$T_{1/2}^{0\nu} \propto \varepsilon \cdot a \cdot \sqrt{\frac{M \cdot t}{BI \cdot \Delta E}}$$

Background index

LEGEND-1000 background goal:

$< 10^{-5}$ cts/keV/kg/yr

$^{77(m)}\text{Ge}$ background at LNGS:

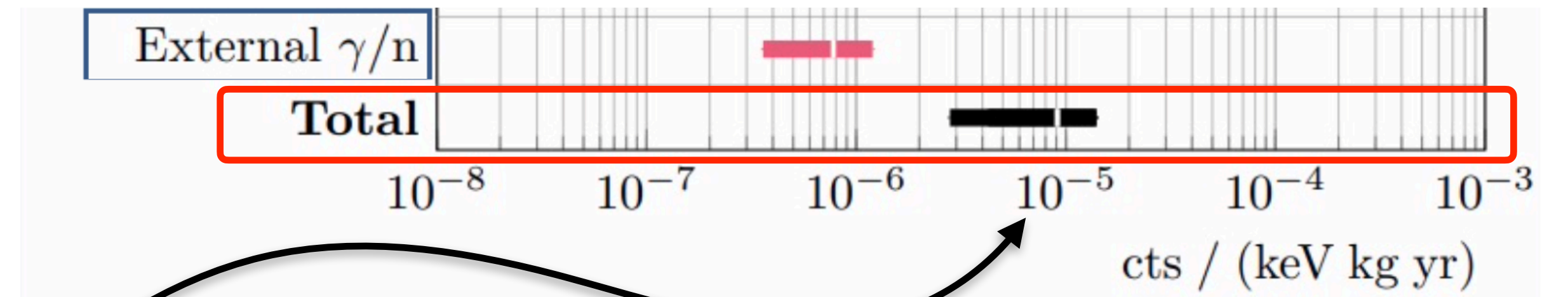
$> 10^{-5}$ cts/keV/kg/yr

[arXiv:2107.11462](https://arxiv.org/abs/2107.11462)

BUT: many opportunities to reduce and actively suppress this background

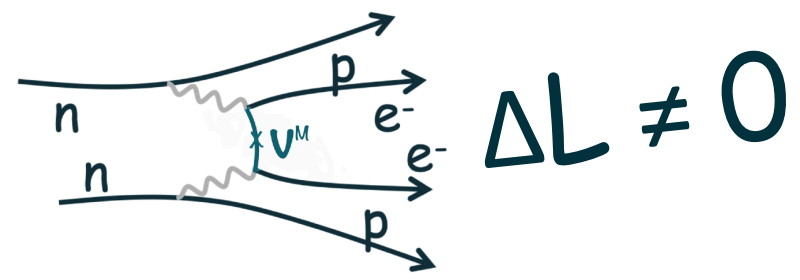
[arXiv:1802.05040](https://arxiv.org/abs/1802.05040)

Why do we have to reduce the cosmogenic background at LNGS?



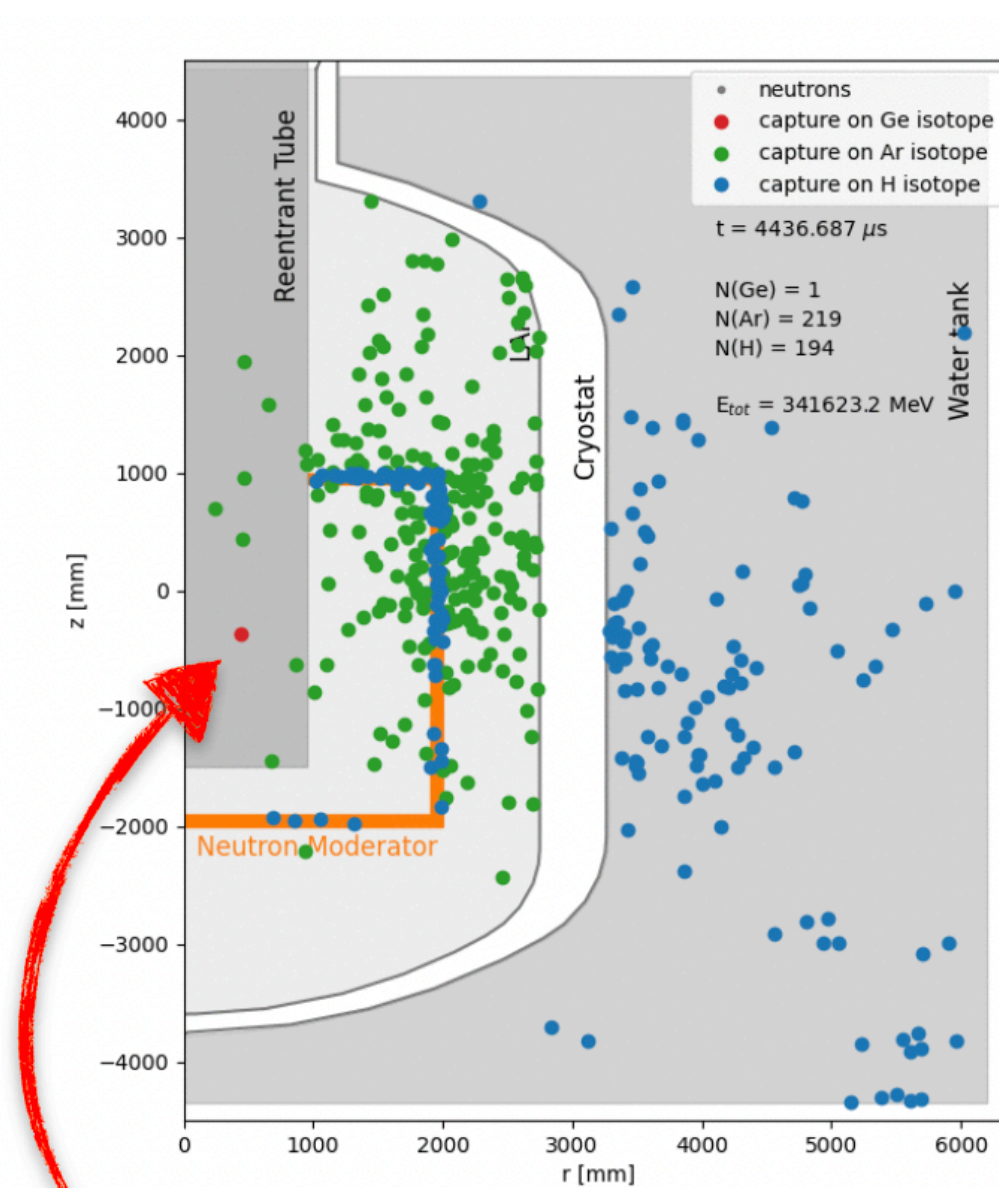
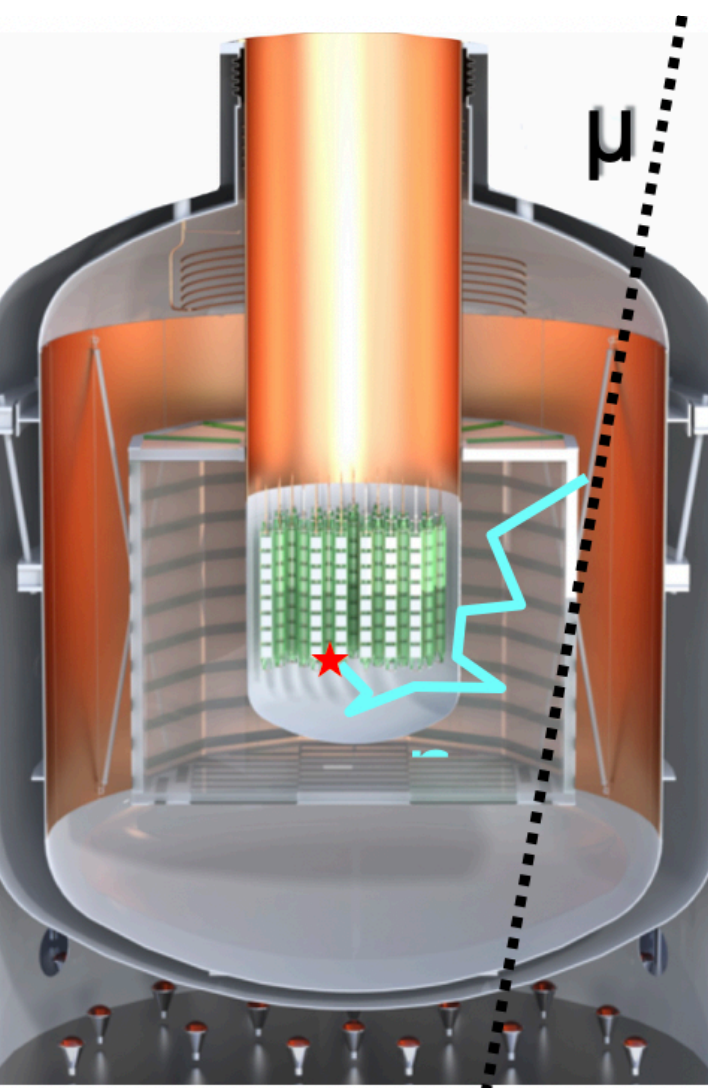
Location	Depth [km.w.e.]	$^{77(m)}\text{Ge}$ background contribution (w/o new cuts*) [cts/keV/kg/yr]
SNOLab (Reference Site)	6	4.2×10^{-7} [0]
LNGS (Alternative Site)	3.5	2.7×10^{-5}

* standard background rejection are applied which strongly supresses ^{77}Ge



Cosmogenic background reduction

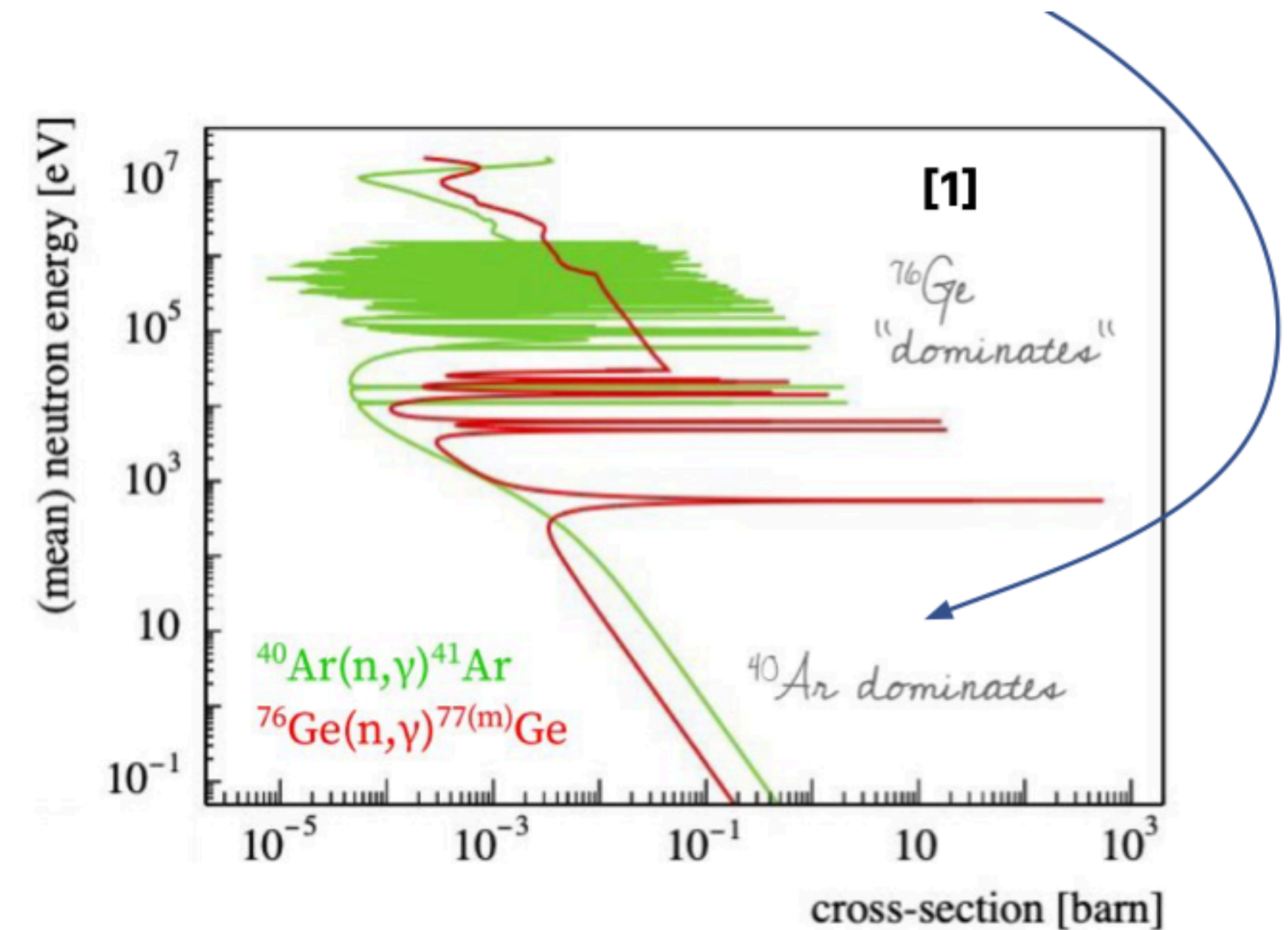
What options are there to reduce the impact of cosmogenic background?



1. Reduce the muon flux → increase overburden.
2. Reduce the neutron flux around the detectors.
3. Tag the $^{77(m)}\text{Ge}$ production and apply a delayed coincidence cut.

Reduce the neutron flux around the detectors - *Idea:*

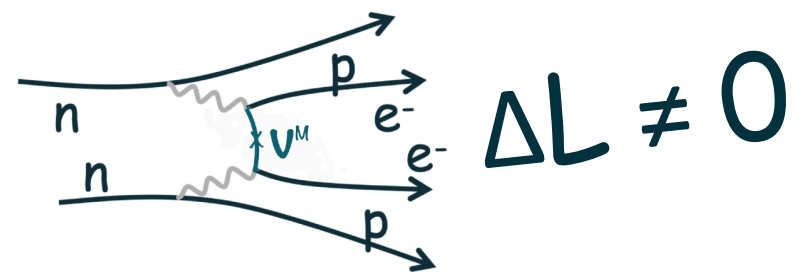
add neutron moderators to slow neutrons down and increase their likelihood to be captured by LAr instead of ^{76}Ge .



LEGEND

dangerous $^{77(m)}\text{Ge}$ event

Credits to M. Neuberger



How to find the optimal design parameter?

- MC studies using a custom simulation module^[3] based on LEGEND-1000 and GERDA setup^[3] implementation
- Solid neutron moderator design: enclosing tube or turbine-like structure
- 5 design parameters: Radius r , n Panels, Thickness d , Length L and Angle θ
- Each simulation run has **2 sets of parameters**:
 - Fixed, **global control parameters** θ (e.g. detector design, hyperparameter...)
 - Event-wise, **local nuisance parameters** ϕ (e.g. initial conditions like particle energy, momentum,...)

- ➔ “High”-dimensional parameter spaces
- ➔ “High” computational cost of Geant4 MC simulations (~200 CPUh)
- ➔ Traditional methods like grid searches are impractical

Run a few simulations at different parameters

turbine-like a geometry

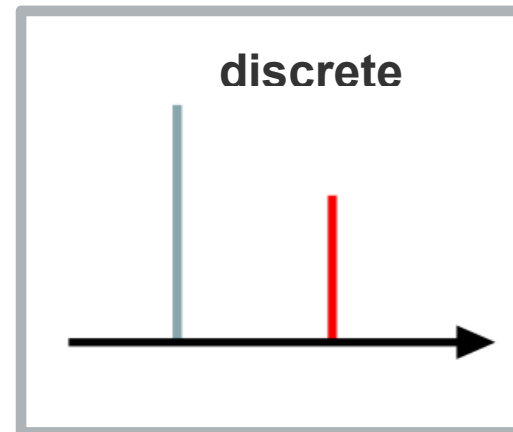
mass further out over larger area

no moderator	enclosure	turbine-like structure

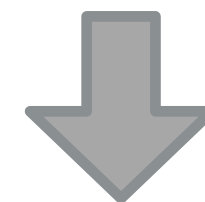
Rare Event Simulations & Rates (Compute Perspective)

Rare Event Probability

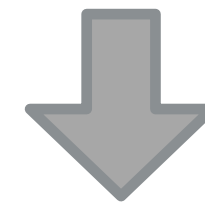
$t(\theta_k, \phi_{i,k})$ small



$$X_{ik} \sim \text{Bernoulli}(p = t(\theta_k, \phi_{ik}))$$



$$y = \frac{1}{N} \sum X_{ik}$$



$$y \ll 1$$

Small $N \rightarrow$ high variance,
noise-dominated estimates

Goal: Estimate $\bar{t}(\theta)$ for many
values of θ

- Each simulation run produces **N binary outcomes** (rare = 1, otherwise = 0).

- The **Observed Rare Event Rate** is

$$y = \frac{1}{N} \sum X_{ik} \ll 1$$

and is **dominated by statistical noise when N is small**.

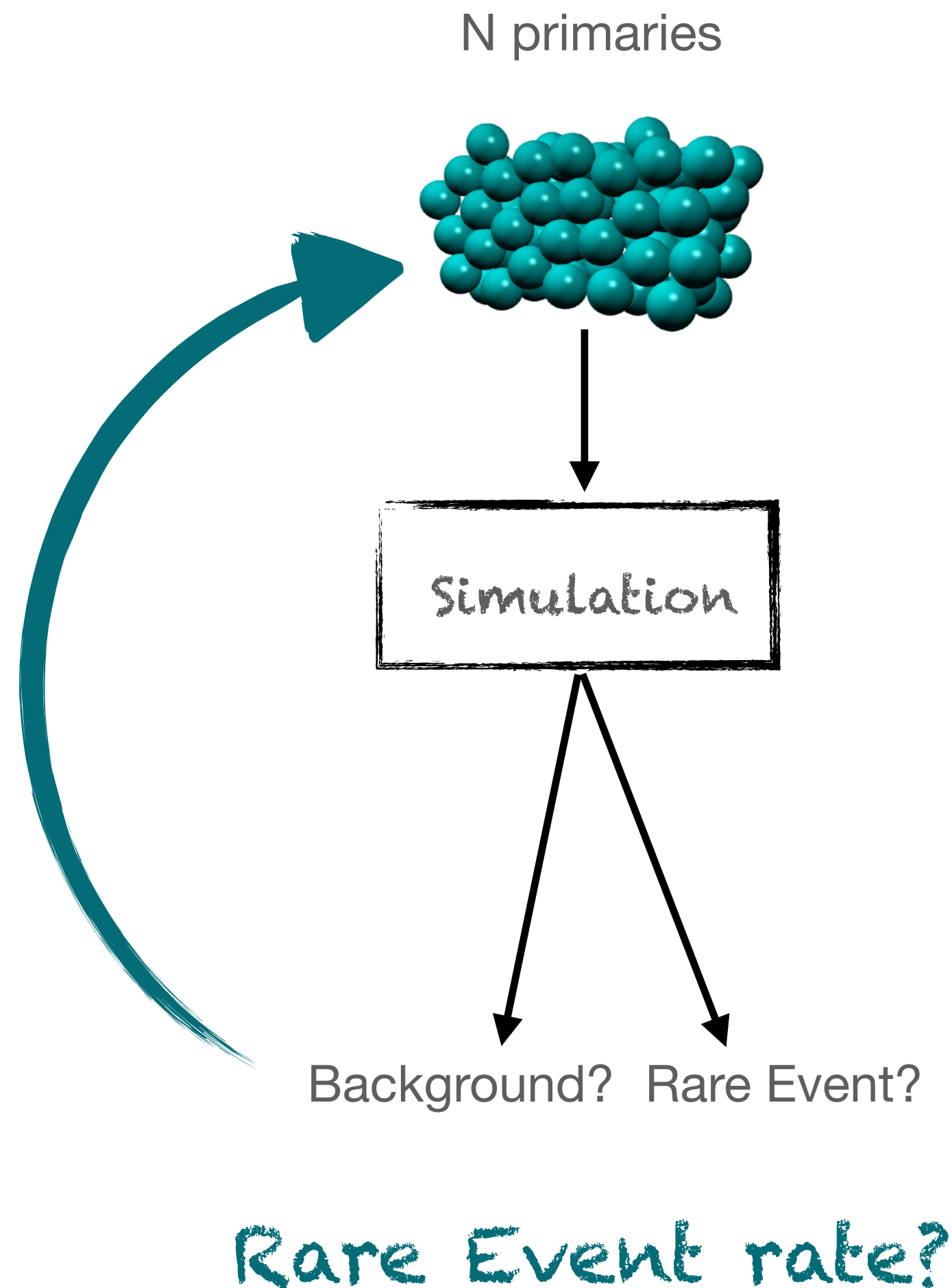
- The true quantity we need is the **Expected Rare-Event Rate**

$$\bar{t}(\theta) = \mathbb{E}_{\phi}[t(\theta, \phi)]$$

evaluated for many configurations θ .

- **Challenge:** Direct Monte Carlo estimation has high variance,
requiring massive N .

Why Brute-force Monte Carlo fails

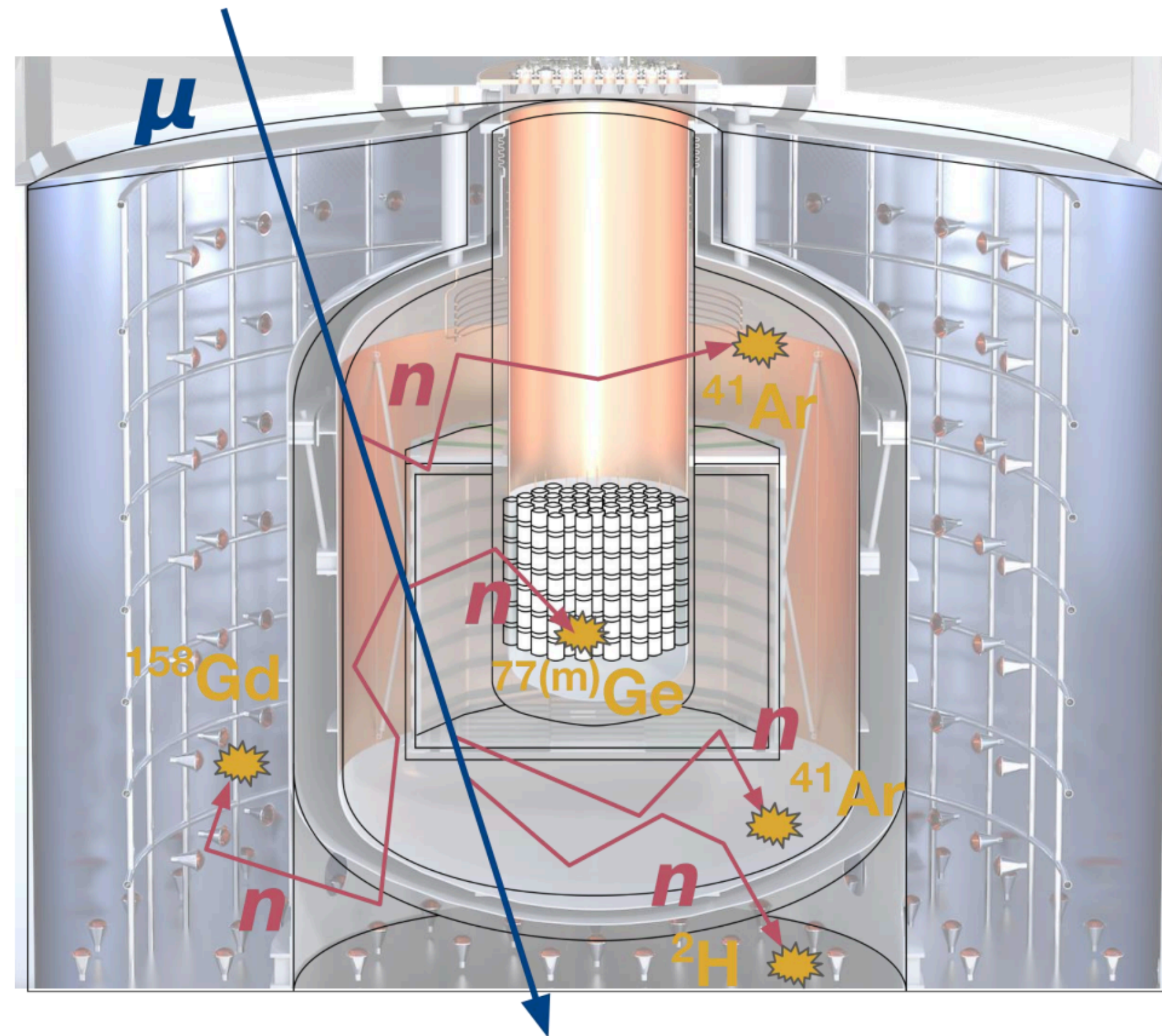


Brute-force Monte Carlo estimate requires rerunning the simulation many times:

- Rerun simulations for **many different physical models**
- Rerun simulations **for many different parameter settings θ** (e.g., *detector geometry, astrophysical hyperparameters, material properties*)
- Rerun simulations **for many different initial conditions ϕ** (*up to tens of event-specific parameters*)
- Rerun simulations **to reduce statistical noise**

→ This leads to an explosion in computational cost. Estimating $\bar{t}(\theta)$ becomes infeasible with brute-force simulation.

Experiment 1: Physics Design Optimization

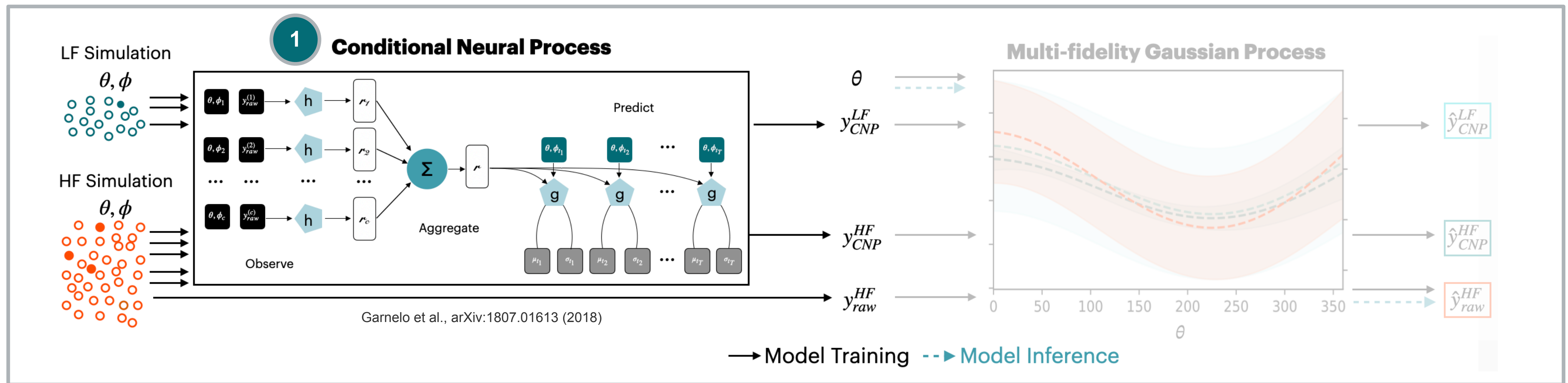


Why brute-force simulation is infeasible

- **Extremely low signal rate:** 10^{-6} - 10^{-5} captures per primary
- **High variance:** Each design point needs $\sim 10^7$ primaries to stabilize estimates
- **Expensive HF run:** ~ 185 CPUh per configuration, statistical noise 5.2%
- **Large search space:** 5D geometry \rightarrow 100-1000 HF evaluations
- **Total cost:** 10^4 - 10^5 CPUh \rightarrow far too slow for iterative design, far beyond practical turnaround time for detector design cycles.

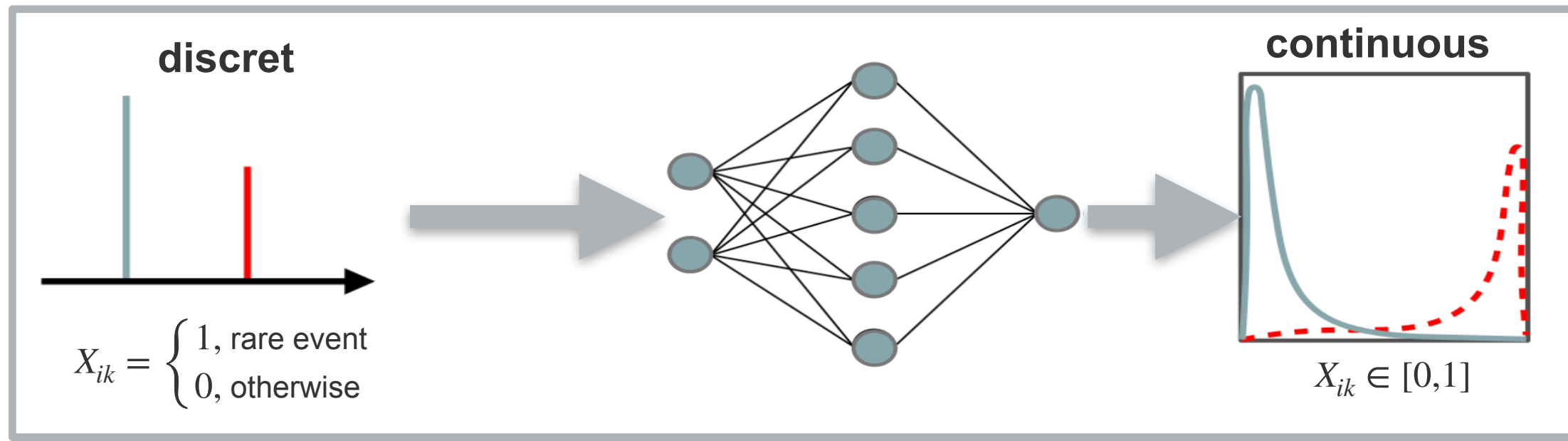
Most compute is spent generating irrelevant background \rightarrow **HF-only optimization is not practical.**

Rare Event Surrogate Model (ReSUM)



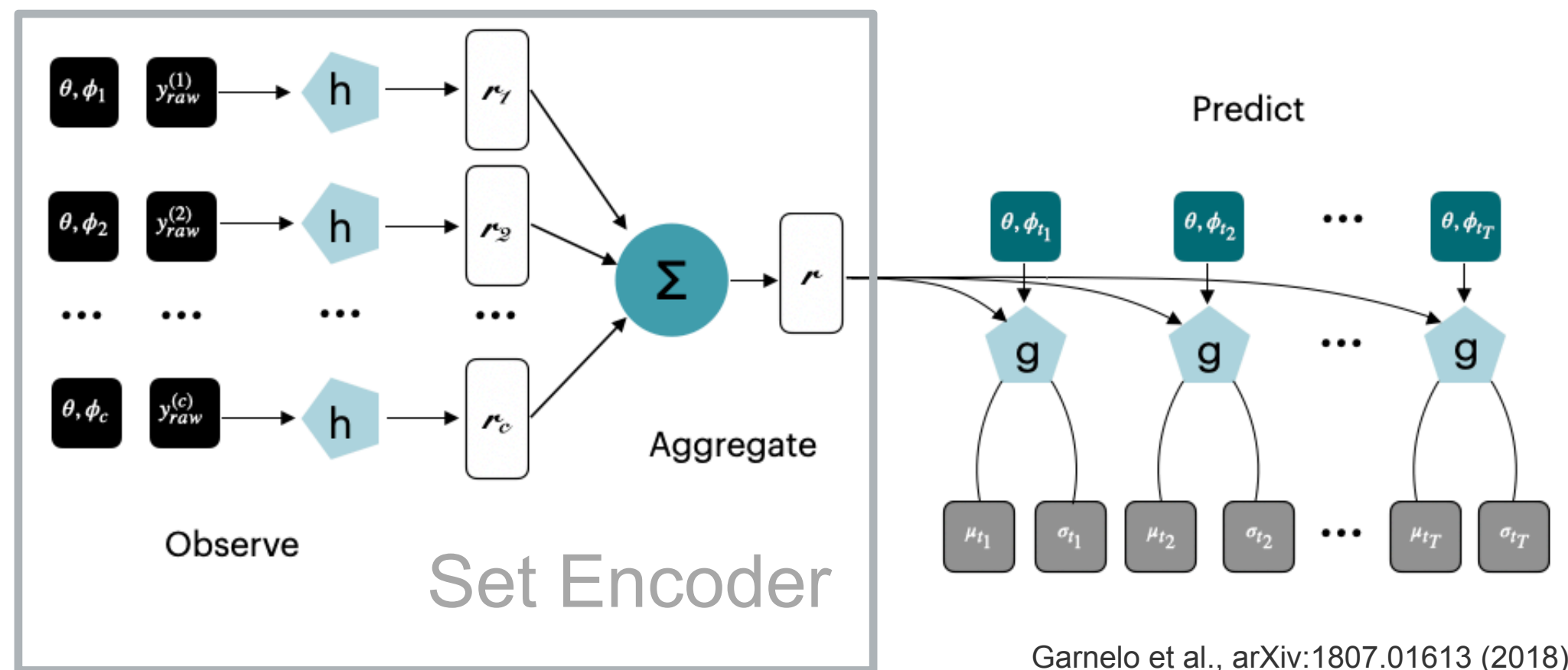
1. Conditional Neural Process

Mitigate statistical noise



- Purpose: Reduce variance in rare-event simulations**
- Converts binary event outcomes into **smooth, continuous estimates**
- Learns correlations across events to **stabilize noisy Monte Carlo outputs**
- Produces **uncertainty-aware predictions** for downstream surrogate modeling

Conditional Neural Process



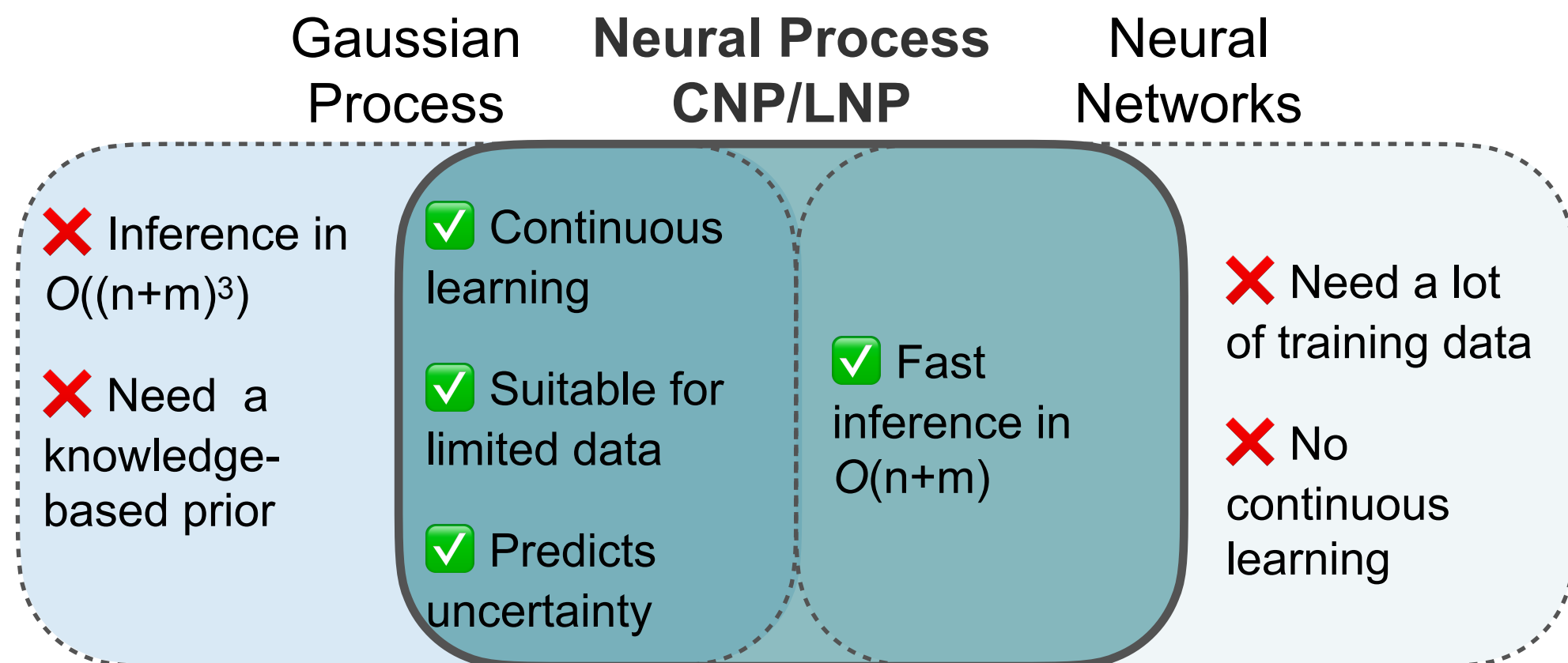
→ a function by conditioning on a set of observed input–output pairs and predicting a distribution for new, unseen inputs.

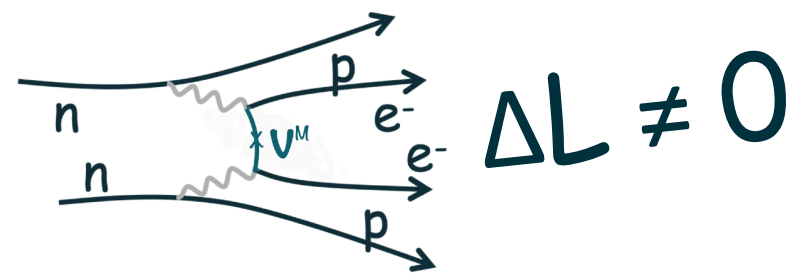
Implementation & scale

- Implemented in PyTorch; **~500k trainable parameters**.
- **GPU memory footprint:** ~1–2 GB during training (fits on any modern GPU)
- **Training samples processed:** $3 \times 10^6 - 1.5 \times 10^7$.
- **Dataset size:** ~0.9–1.2 GB → entire dataset loaded into RAM at startup, eliminating I/O bottlenecks.
- **Training throughput:** 10^5 samples/sec on 1 GPU (~5 GPU-hours total).
- **Inference throughput:** 6×10^6 samples/sec on 1 GPU ($\approx 200\times$ faster than Geant4 event simulation).

Data Augmentation

- Mixup: Linear combination of signal and background event features.

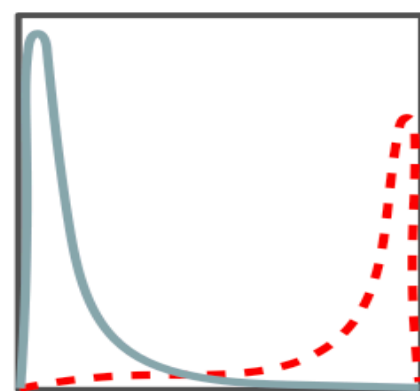
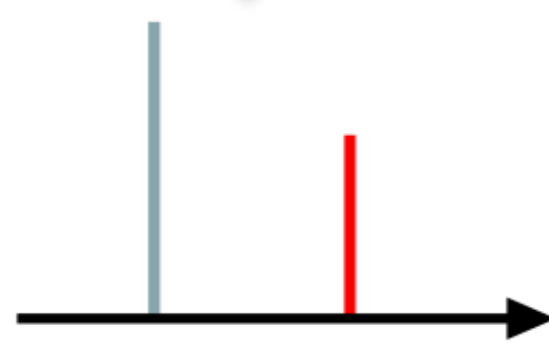




Run Geant4 LF simulations for different moderator configurations



Count number of neutrons being captured given the configuration



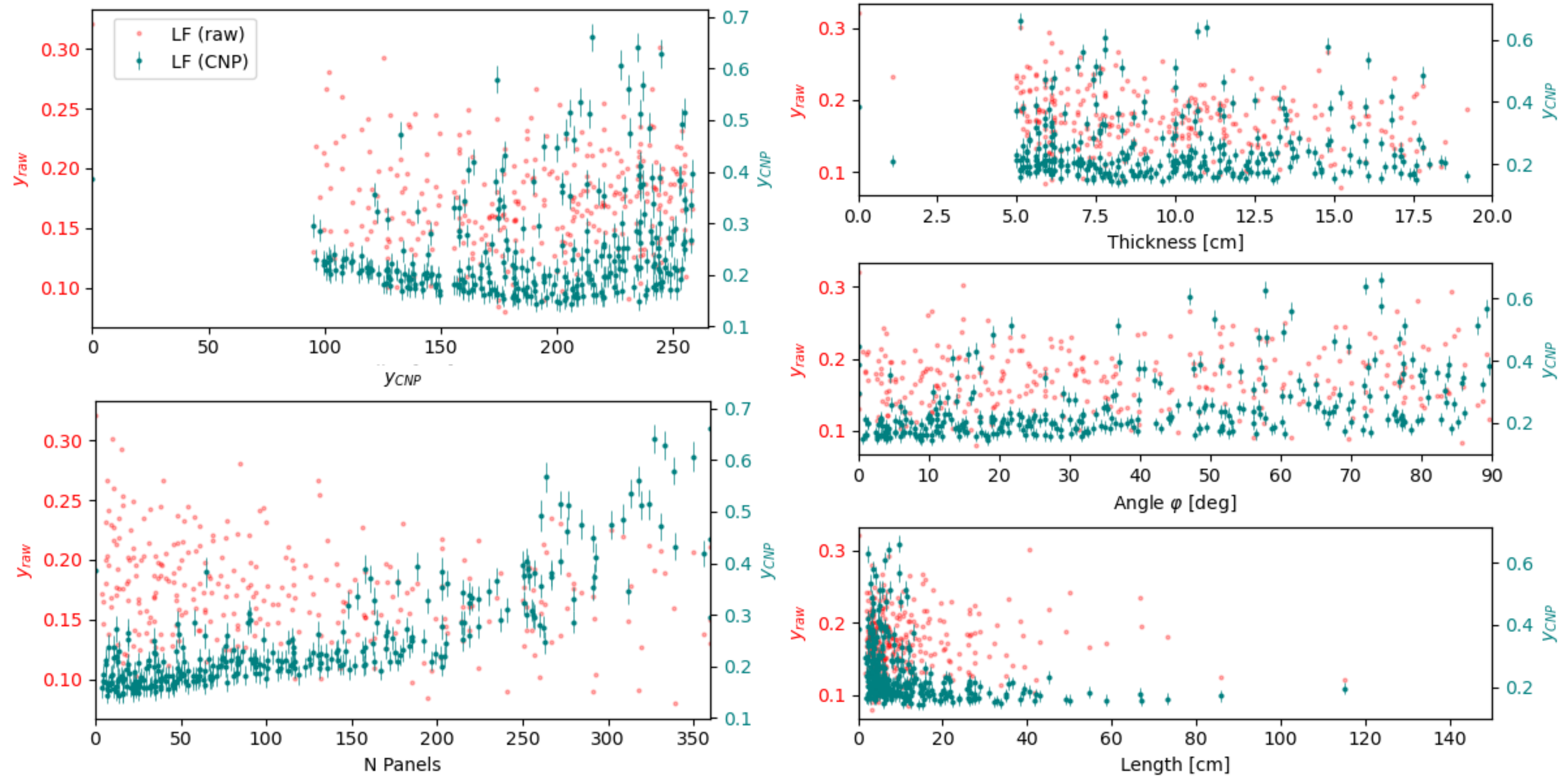
Aggregate

$$y = \frac{1}{N} \sum X_{ik}$$

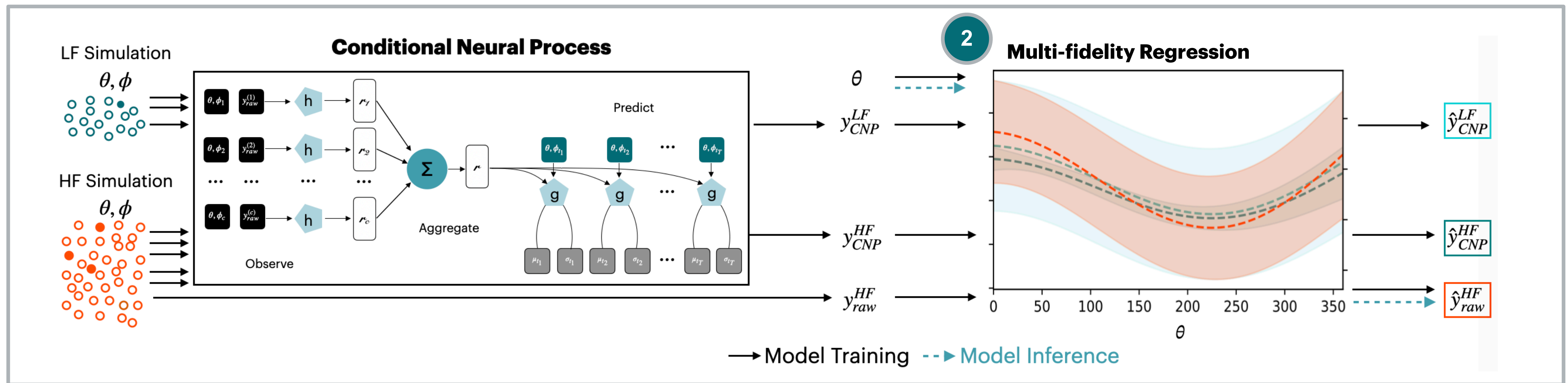
Neutron capture is shown as **1-dim projection**

→ **Strong fluctuations in the LF training data before**

→ **Smoothing with CNP network**



Rare Event Surrogate Model



Reduce computational cost

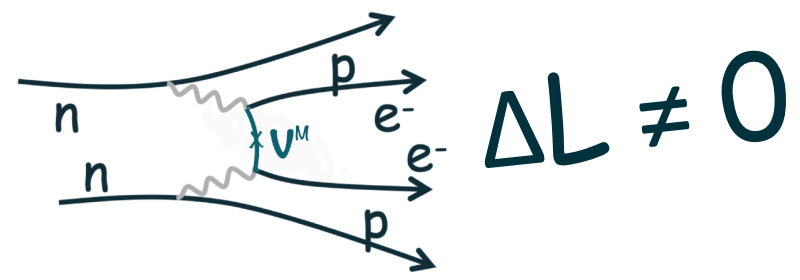
2. Multi-fidelity Regression

- Combine **cheap LF simulations** with **sparse HF simulations** to model the target function.
- LF covers the parameter space; HF provides **accurate anchors**.
- Regression learns correlation + discrepancy between fidelities.
- Active learning selects the **most informative HF points**, minimizing total compute.

based on the Multi-Fidelity approach idea

$$y^{(f_i)} = \rho y^{(f_{i-1})} + \delta$$

↑ *model correlation to lower fidelity*
↑ *model discrepancy term*



HF & LF simulation

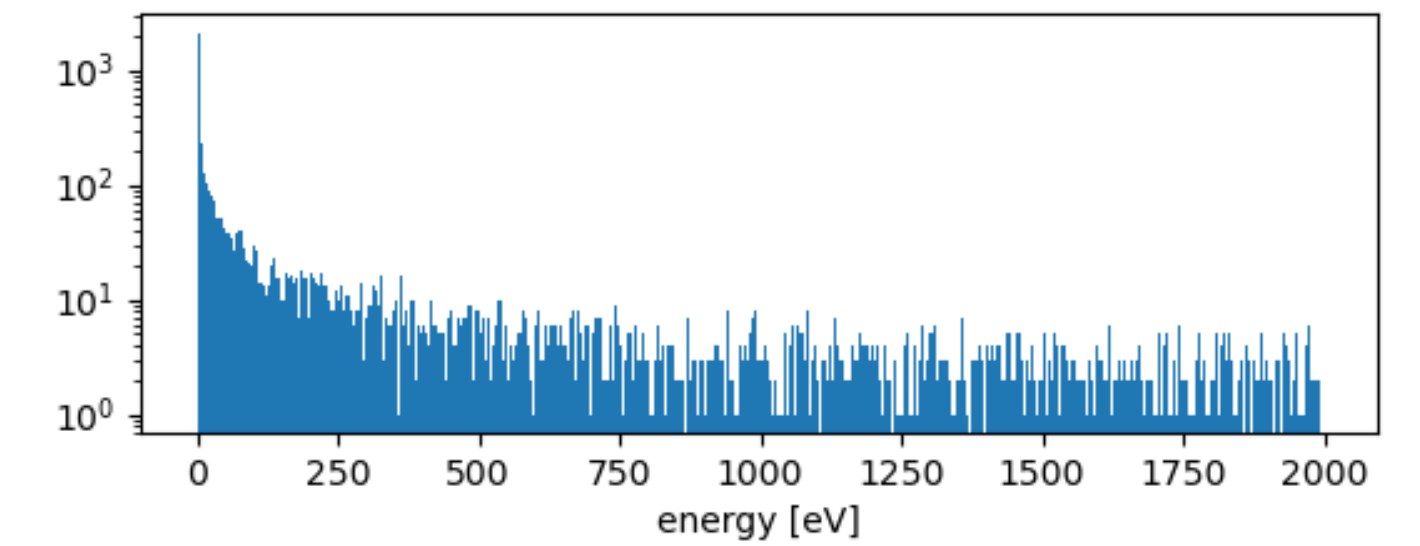
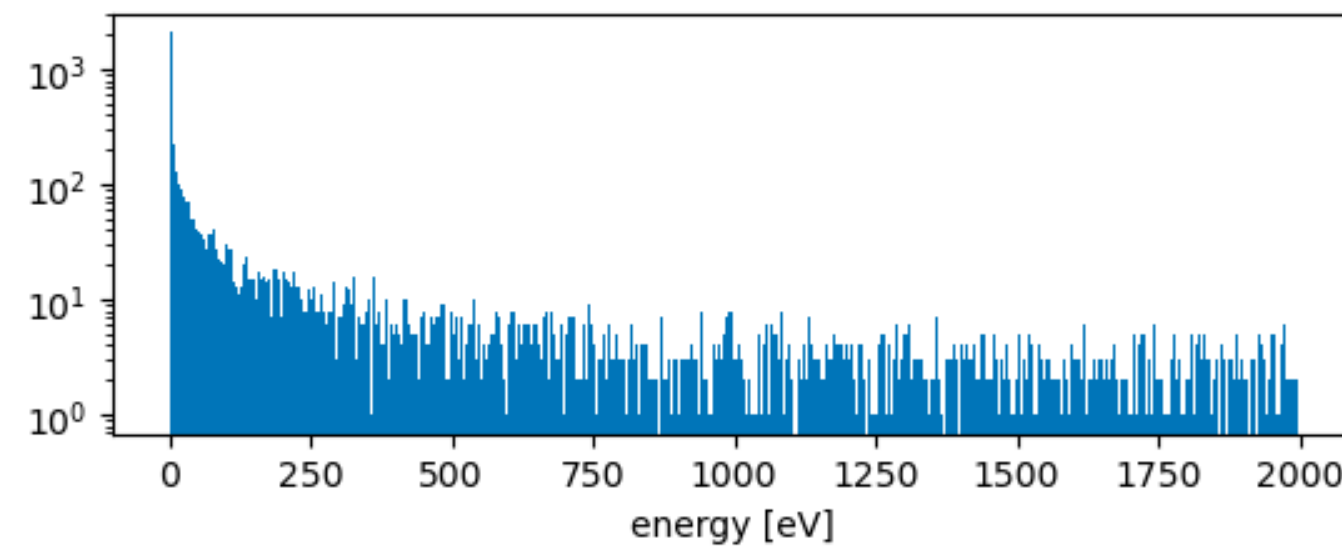
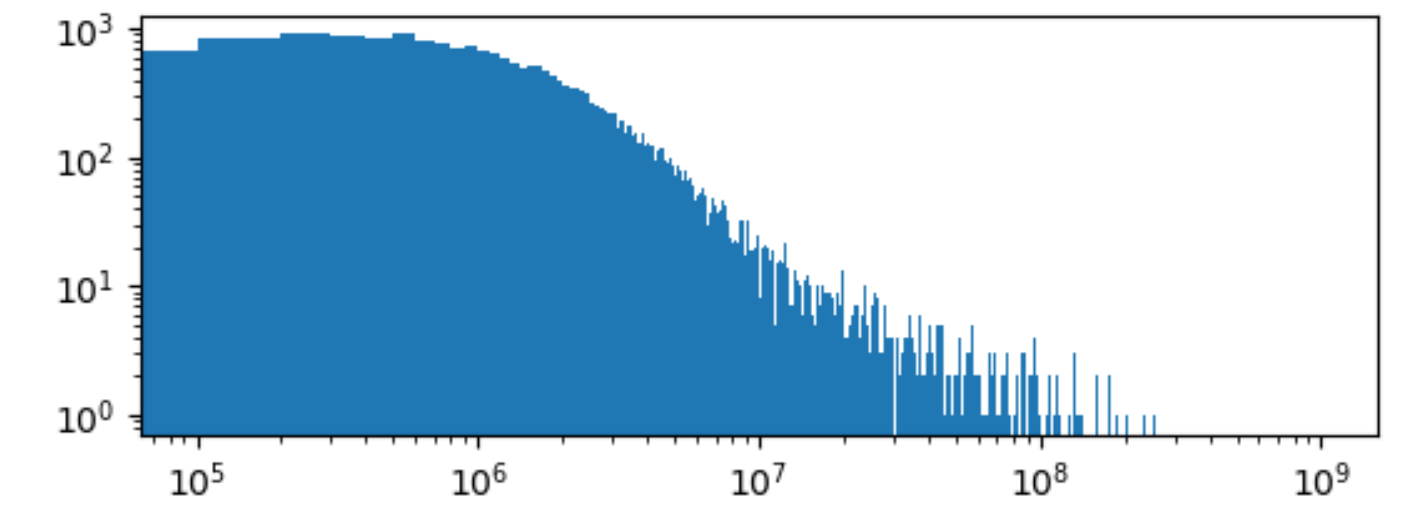
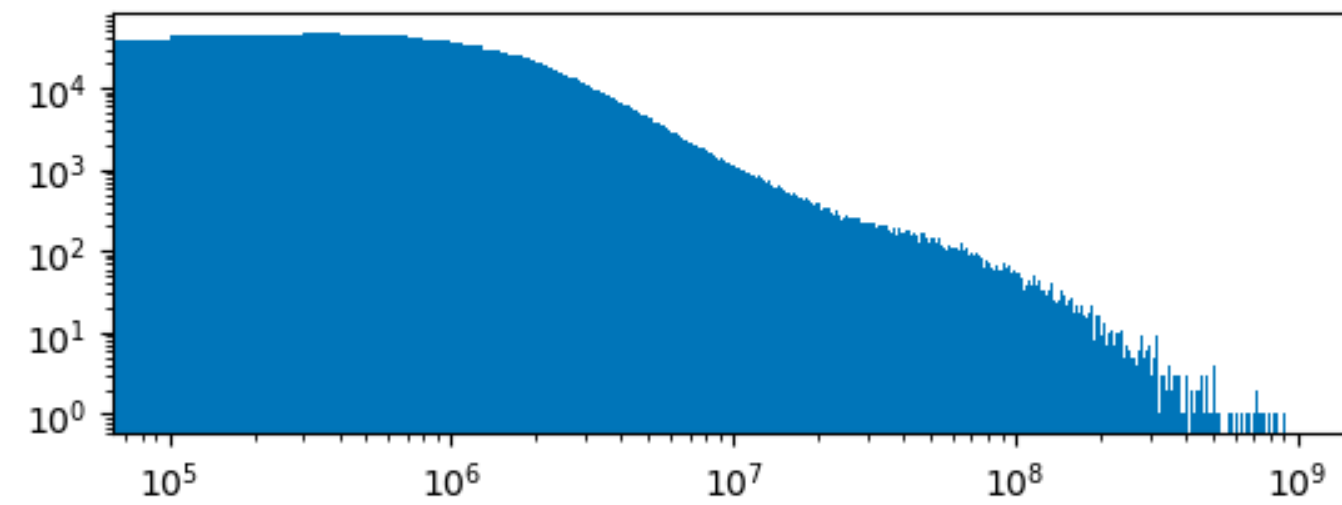
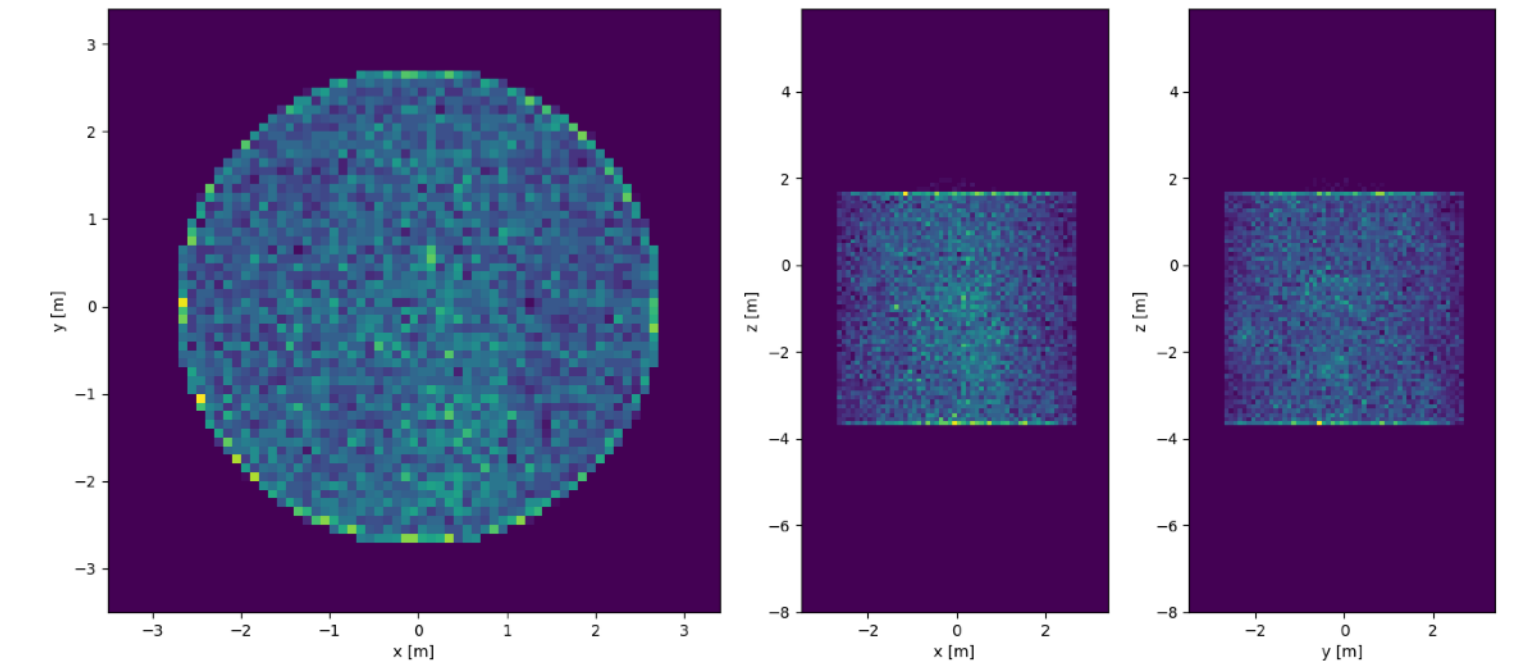
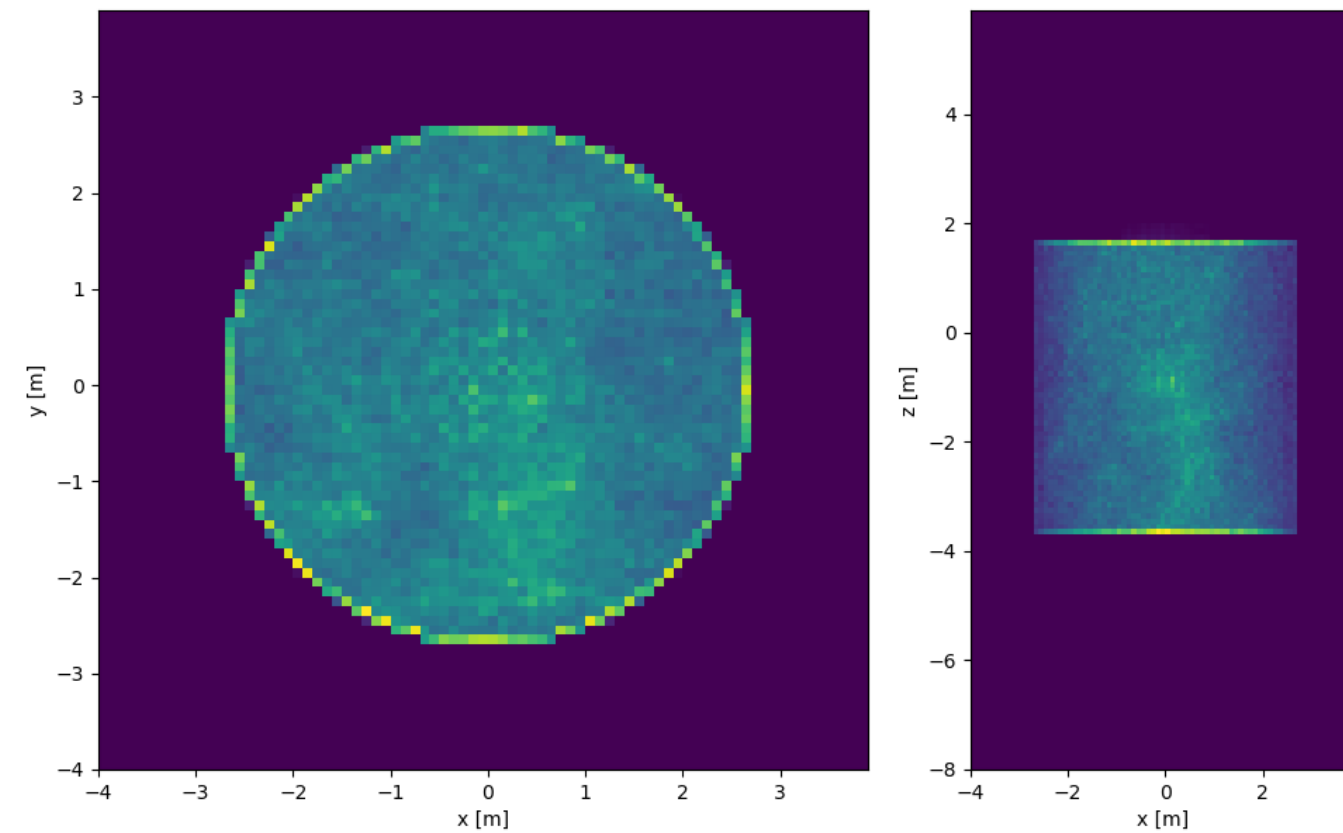
Event specific distribution

$$g(\phi) = g_{\mu}(\phi_{\mu}) d\phi_{\mu} \cdot g_n^{ini}(\phi_n^{ini} | \phi_{\mu}) \cdot g_n^{fin}(\phi_n^{fin} | \phi_n^{ini}) \quad \text{LF}$$

muon in out
neutron

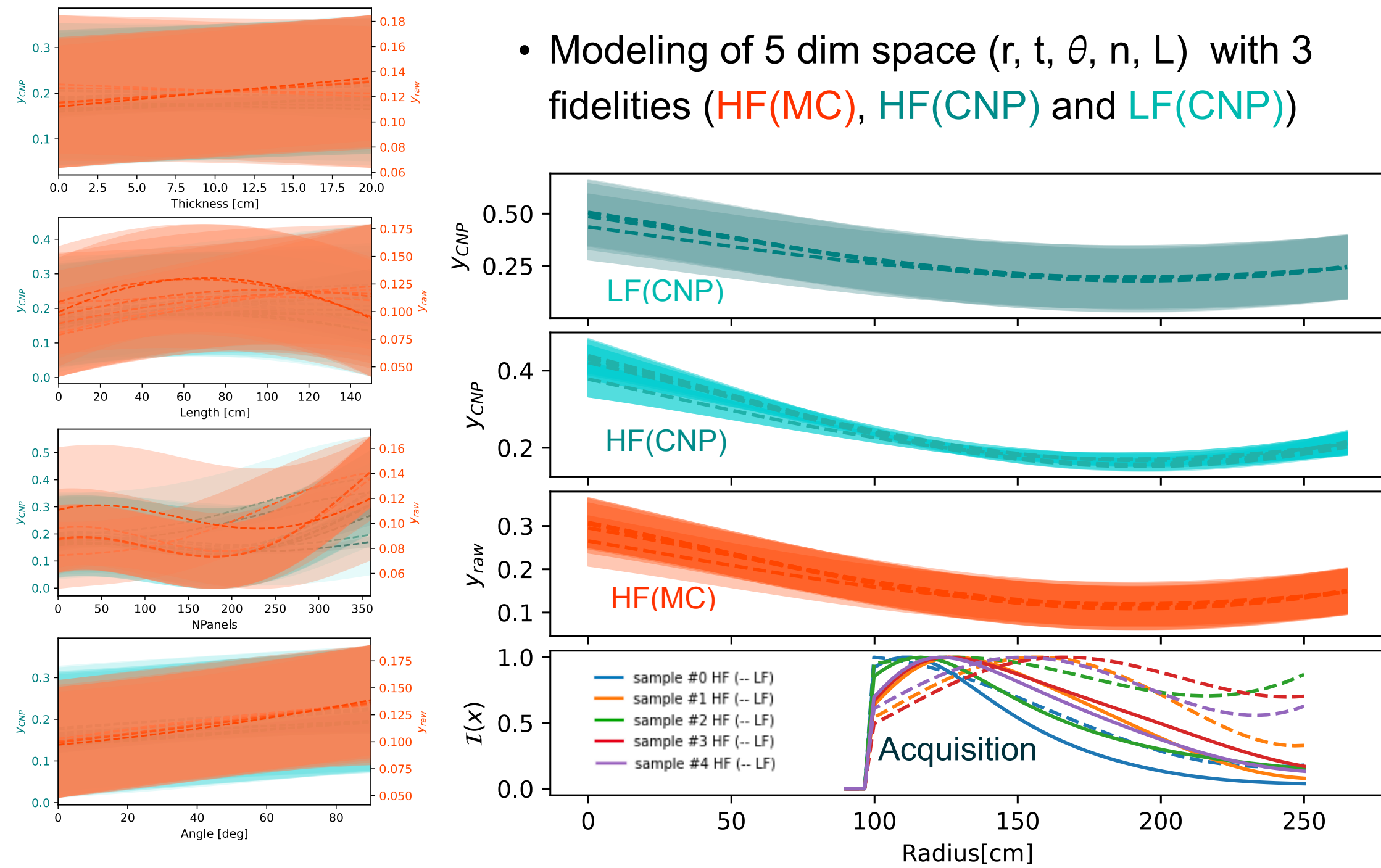
draw distributions with random neutron starting points from high fidelity simulation

LF input



	HF	LF
Primary particle	Muon	Neutron
CPUh per neutron	$1.5 \cdot 10^{-4}$ ($2 \cdot 10^{-5}$ per muon)	$3 \cdot 10^{-6}$
Full detector geometry	✓	✓
Full neutron physics lists	✓	✓
Timing to primary and production info	✓	✗

Experiment I: Results

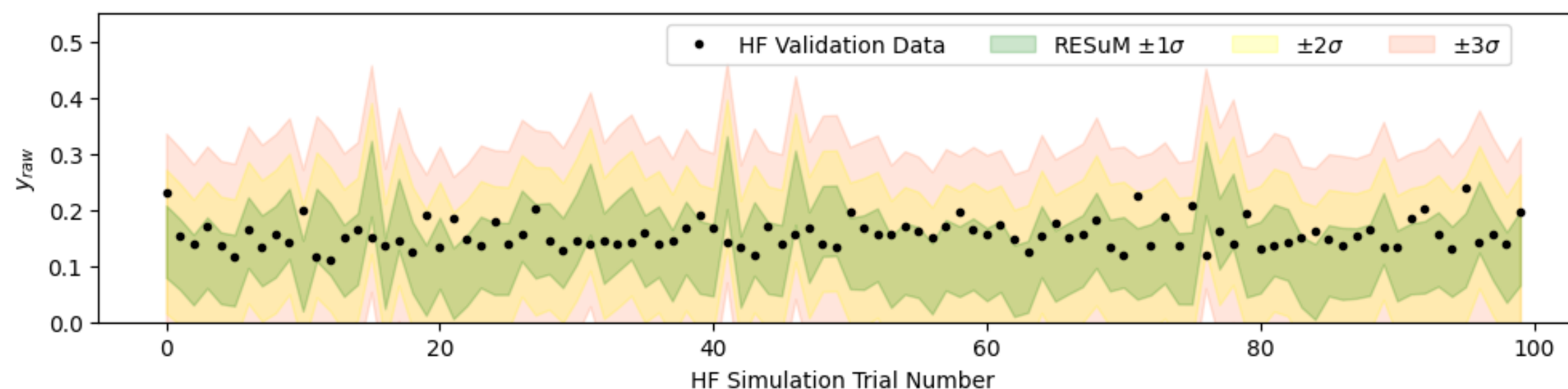


- **Setup:** 310 LF points, 10 HF points
- **Accuracy:** HF outputs fall within
 - $\pm 1\sigma$: 69% · $\pm 2\sigma$: 95% · $\pm 3\sigma$: 100% (MSE = 0.0024)
- **Surrogate performance:**
 - Accurately reproduces HF behavior across a 5D design space
 - Captures LF/HF correlation with 3-fidelity modeling
- **Impact on compute:**
 - 66.5% reduction in neutron-background metric
 - Uses **only 3.3% of HF compute vs. brute-force simulation**



Summary: The Rare Event Surrogate Model for Physics Design (ReSUM) enables reliable design optimization with orders-of-magnitude lower computational cost.

Validation: Comparison of 100 HF Simulation Trials vs Model Prediction



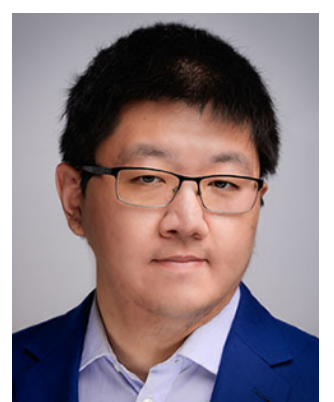
Spotlight 2025



Ann-Kathrin Schuetz¹



Alan Poon¹

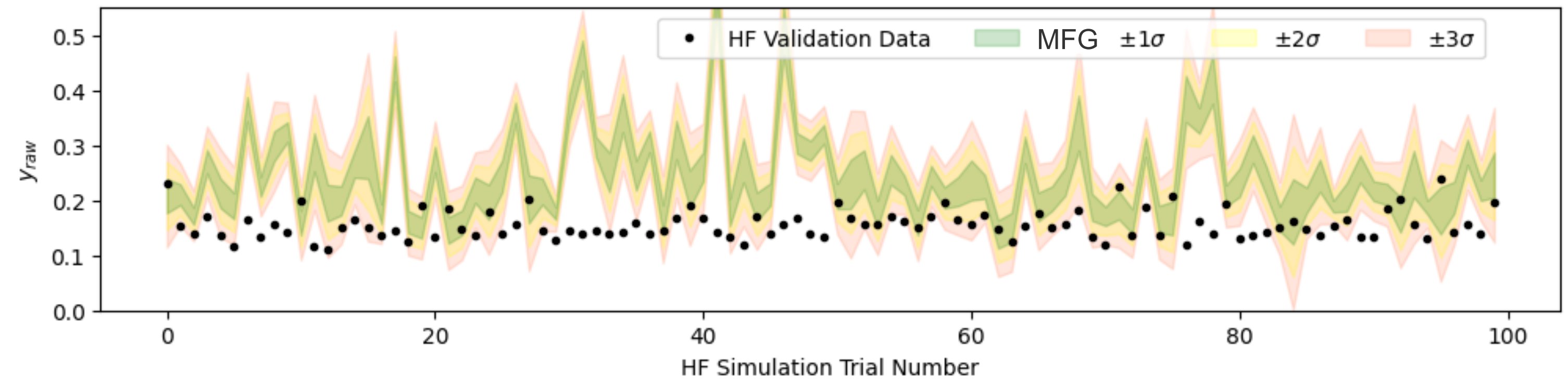


Aobo Li²

Naive Multi-Fidelity GP Fails on Noisy LF Data

$\pm 1\sigma$ [%]	$\pm 2\sigma$ [%]	$\pm 3\sigma$ [%]	MSE
2	4	5	0.0095

- LF \rightarrow small N regime \rightarrow **high variance**
- Discrepancy term learns systematic bias, not variance
- HF is too sparse to counteract LF noise

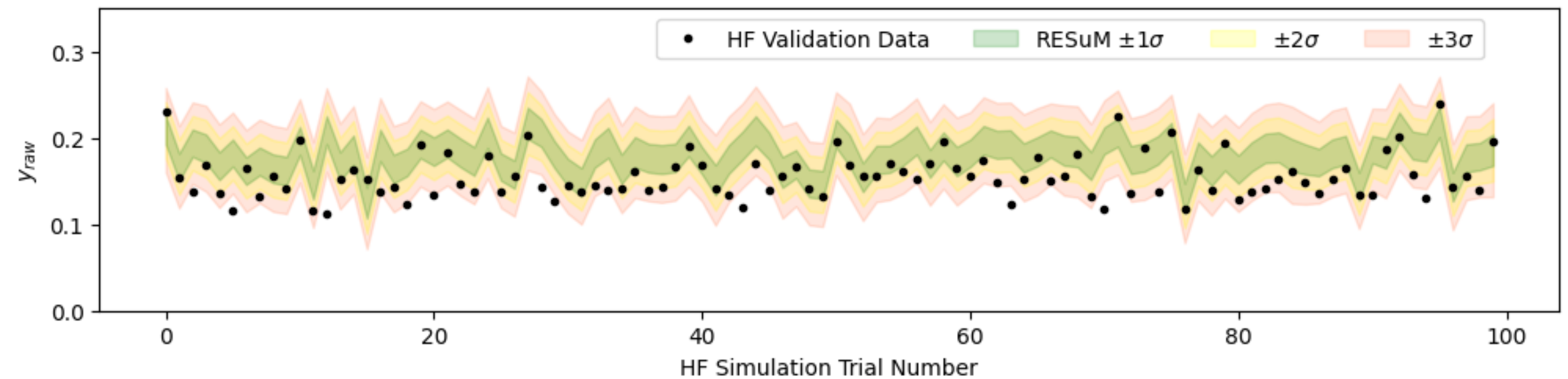


Adaptive Importance Sampling + Multi-Fidelity GP

$\pm 1\sigma$ [%]	$\pm 2\sigma$ [%]	$\pm 3\sigma$ [%]	MSE
33	75	95	0.0012

AIS can fail in cases of

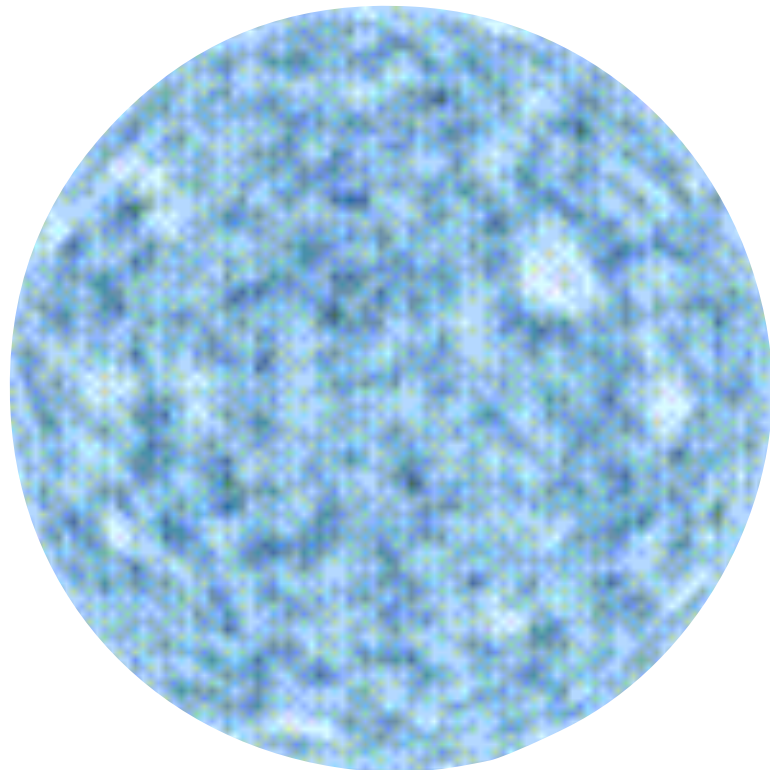
- Not learnable rare event distributions
- Tiny/ disconnected modes
- Too few rare events
- Sample from a local mode only



AIS helps reduce variance but cannot denoise LF completely

"Rare Events are everywhere"

It is challenging to observe Massive Stars ...



Massive Stars are Rare

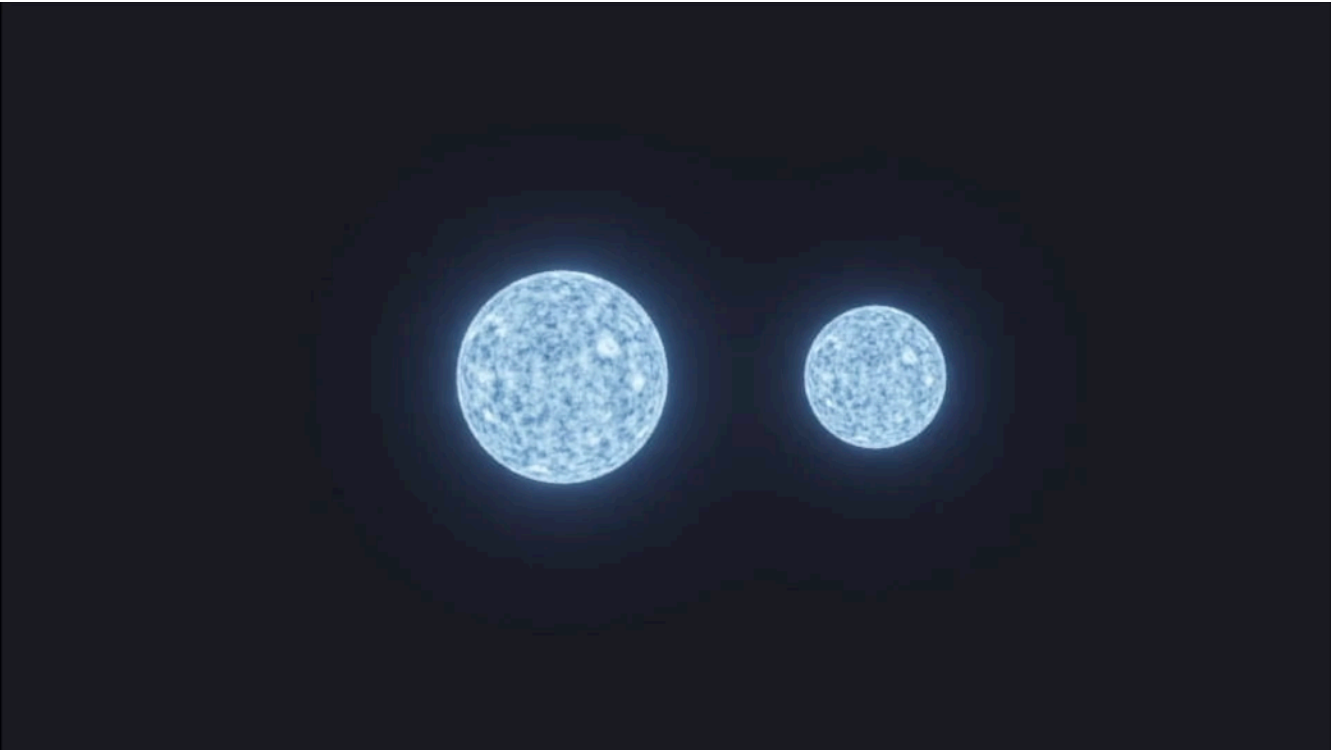
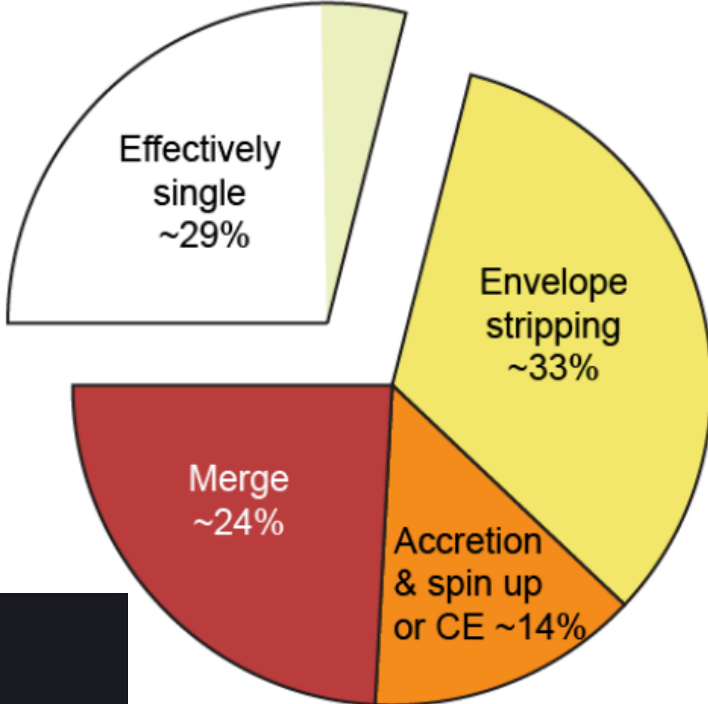
Number of stars $\propto \text{Mass}^{-2.3}$
(Salpeter 1955)

Massive stars live fast and die young

Lifetime $\propto \text{Mass}^{-2.8}$
(Kippenhanh & Weigert 1990)

They live in binaries

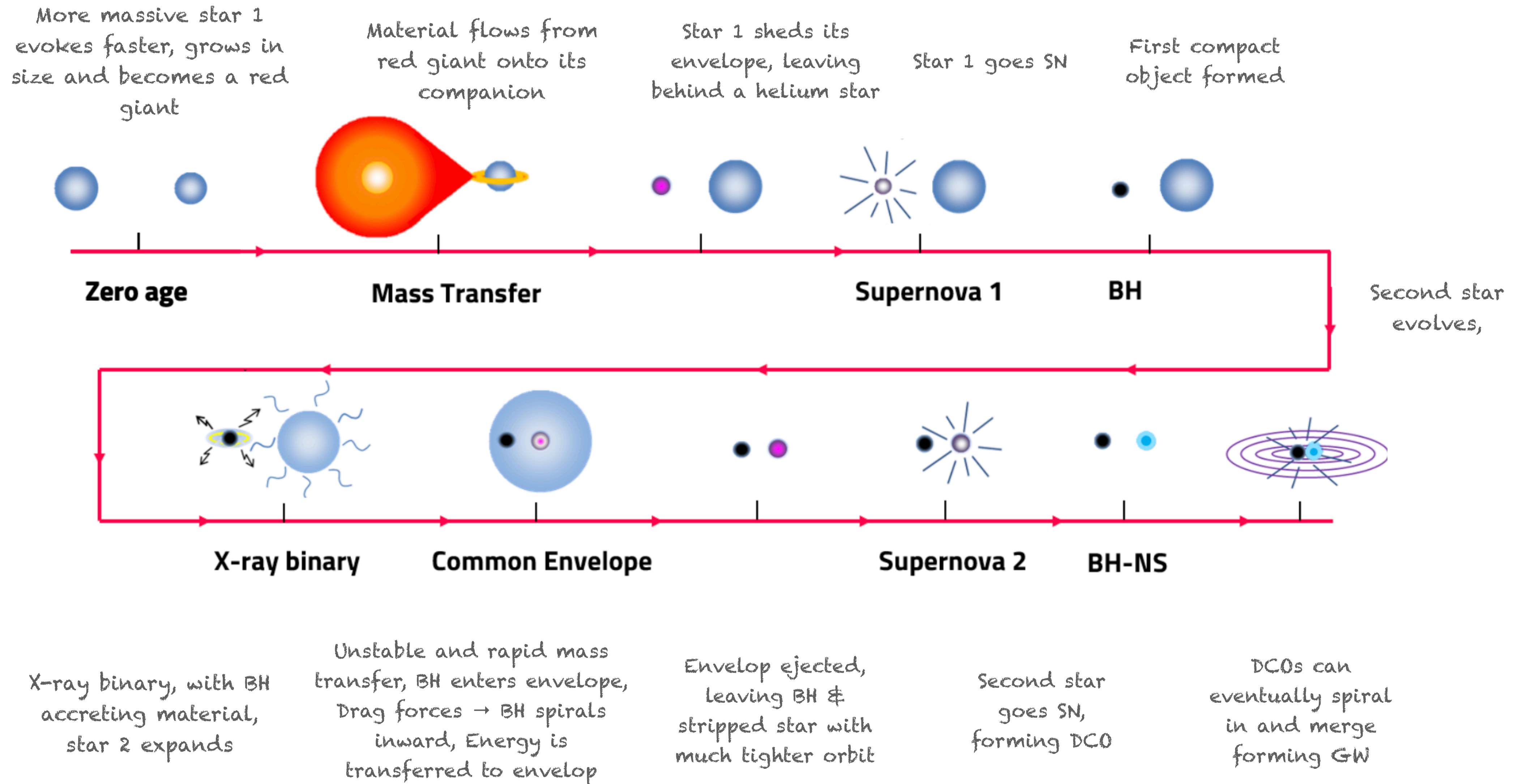
~70% of massive stars will interact with a companion during their lifetime (Sana + 2012)



Sana + (2012), Moe & DiStefano + (2017)

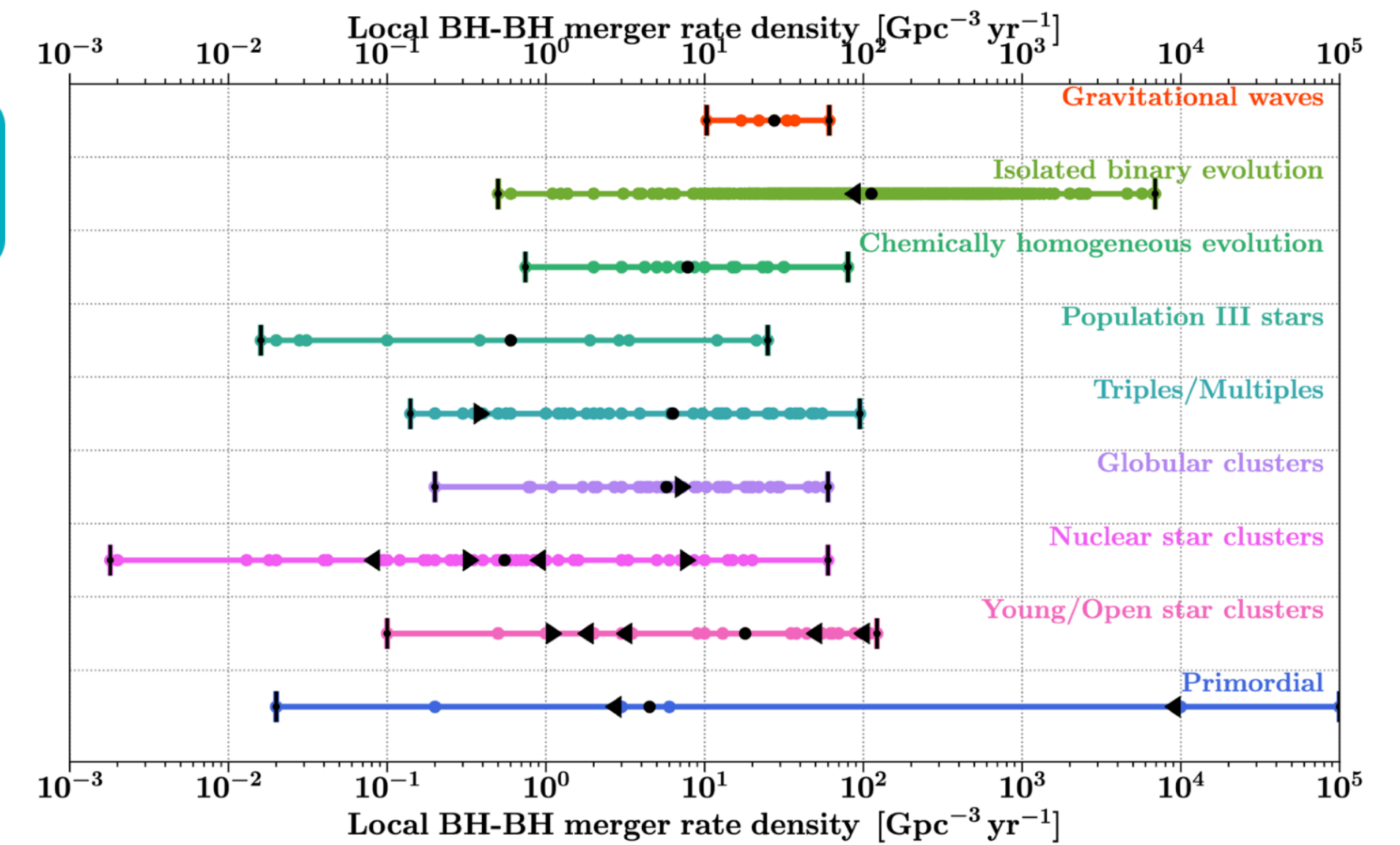
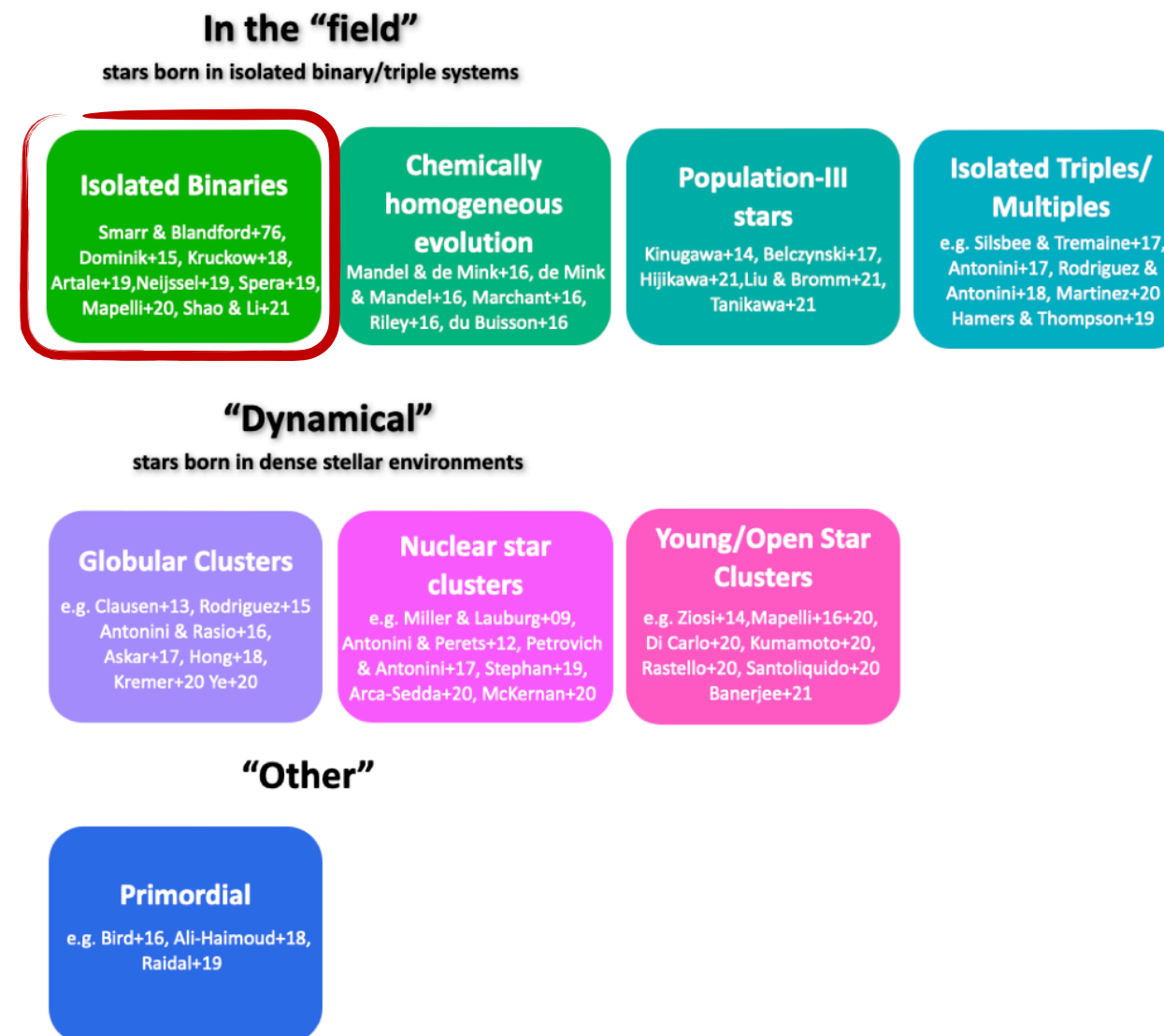
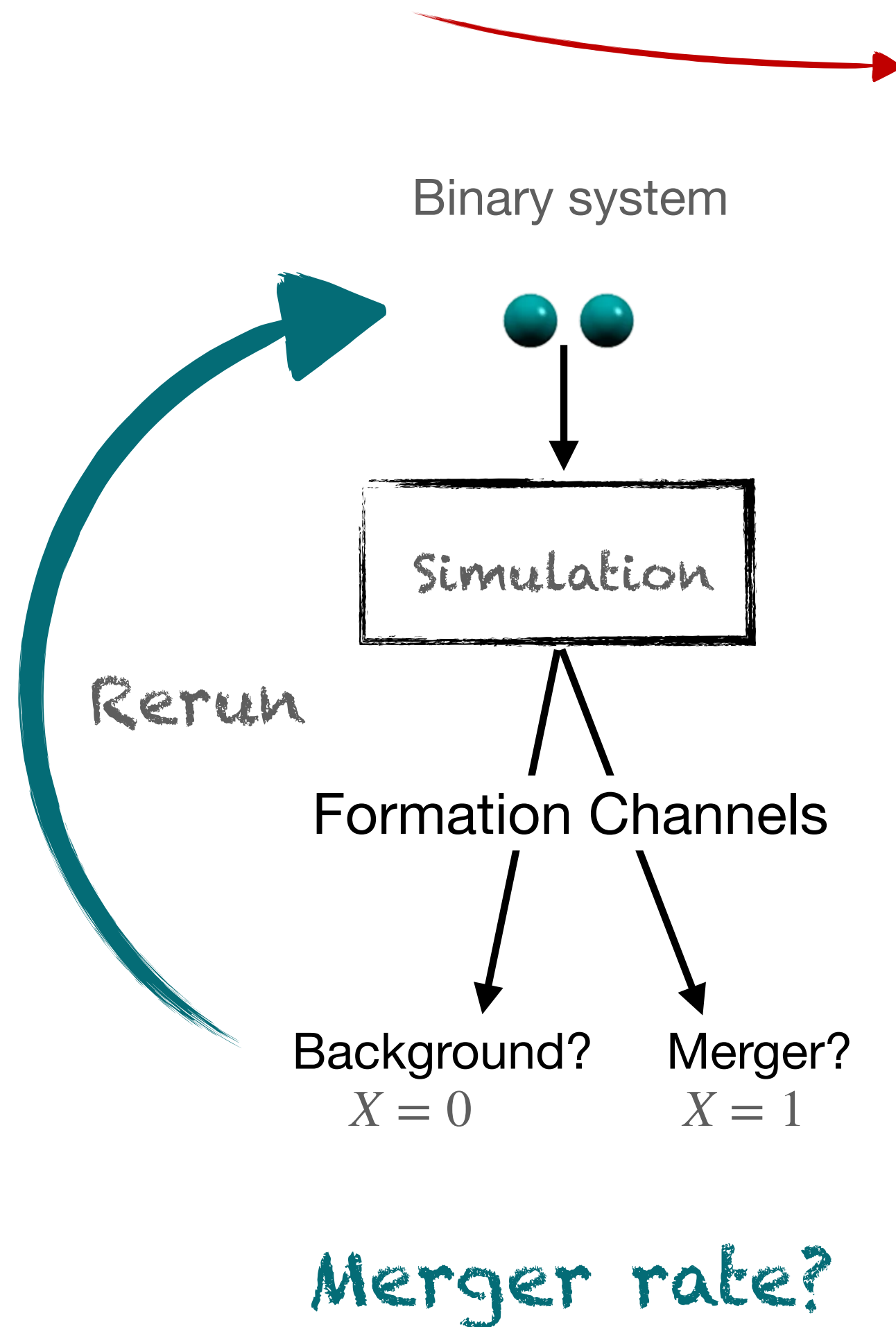
Binary interactions are crucial in forming close black hole binaries which can merge and create gravitational waves.

Isolated binaries



Formation Channels

This talk **only isolated Binaries** - Inclusion of additional formation channels → *future work*

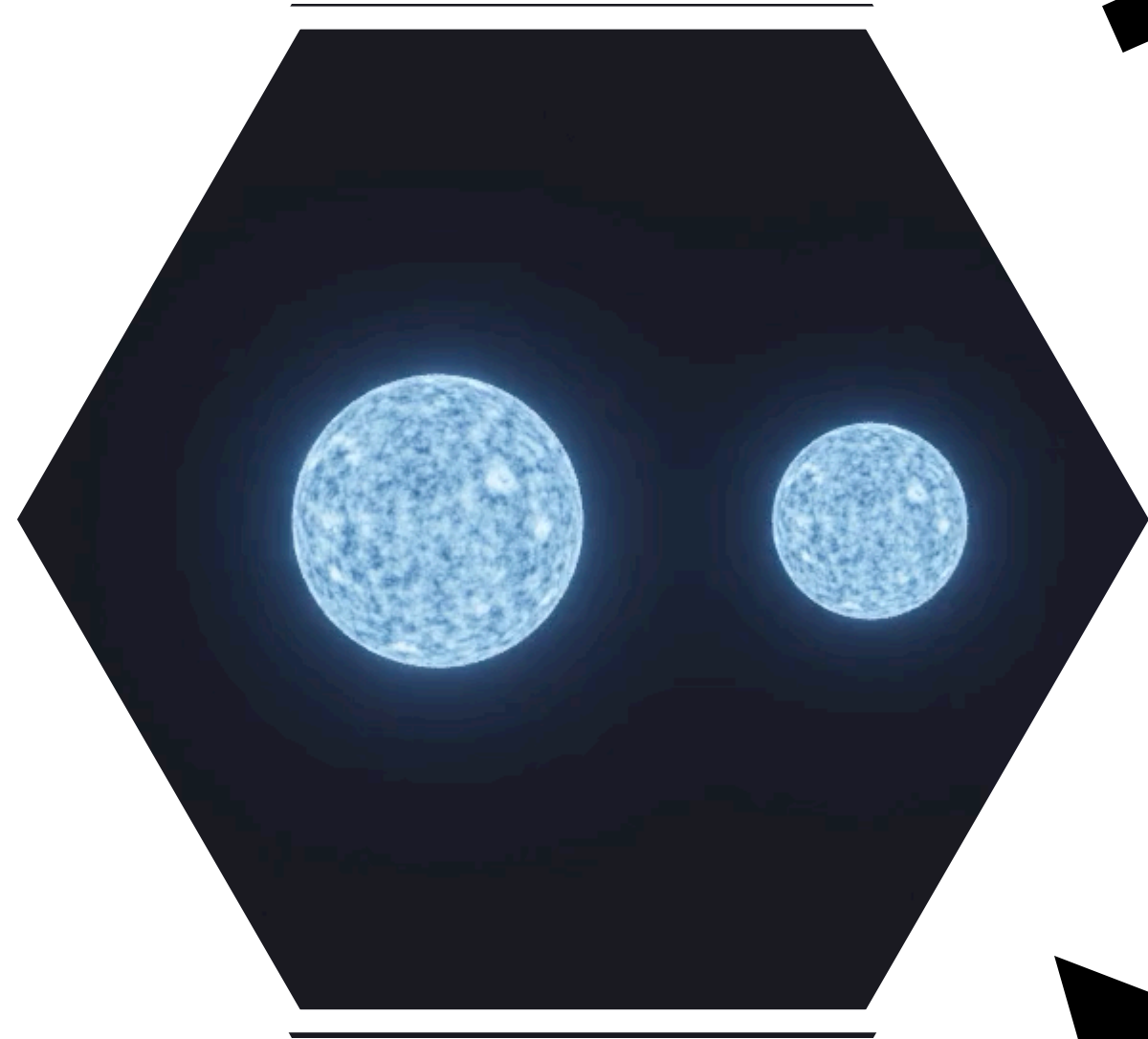


[Broekgaarden et al. (2021)]

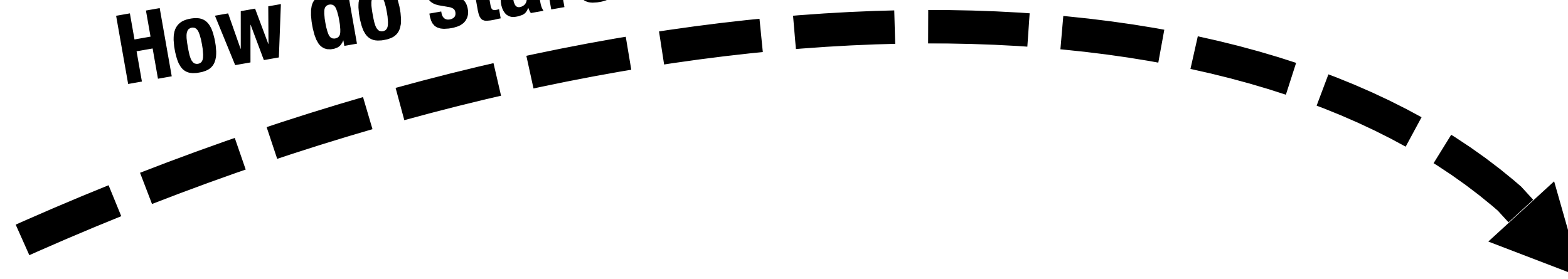
Many different formation channels!

!! Rerun simulations for many different formation channels !!

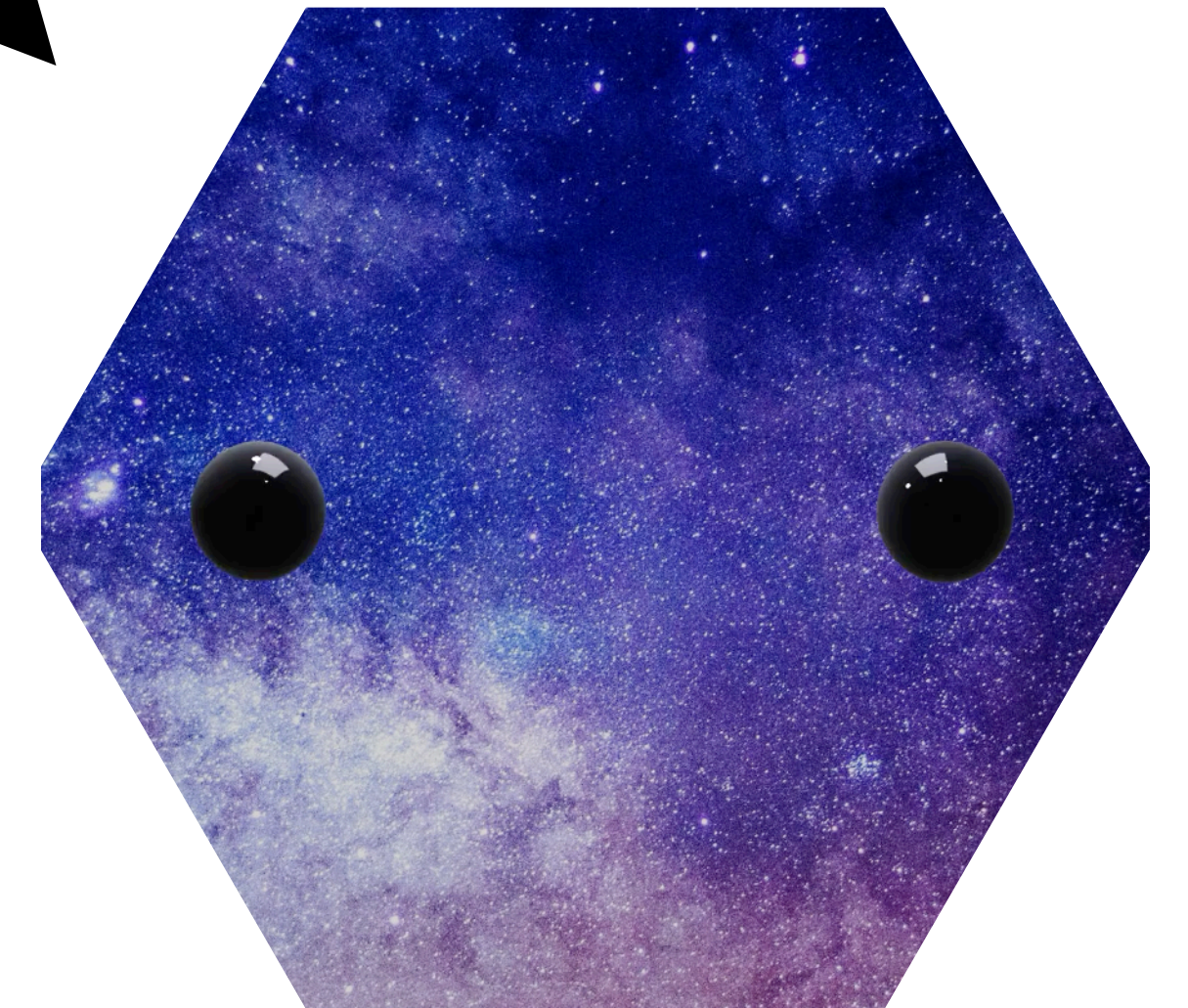
STARS IN THE PAST



How do stars evolve?



WHAT WE SEE



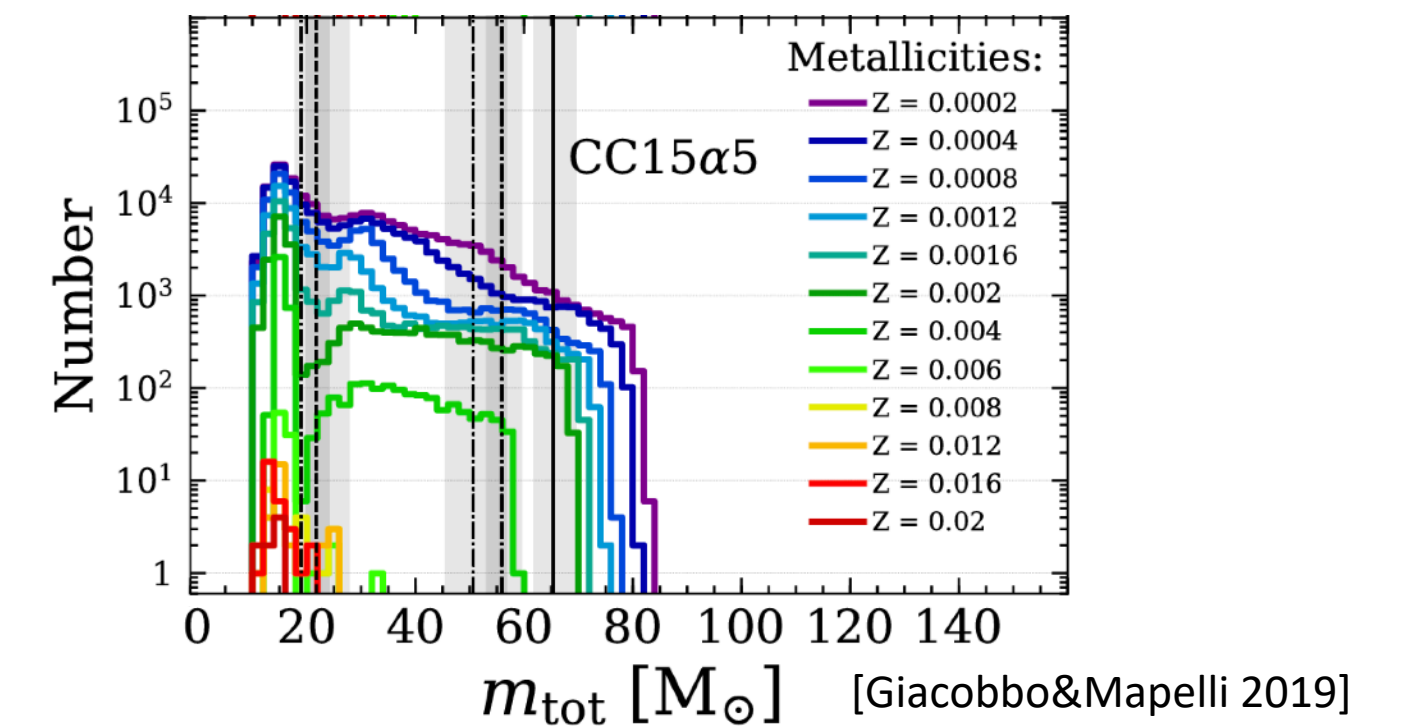
Gravitational Wave Paleontology

How did they form?



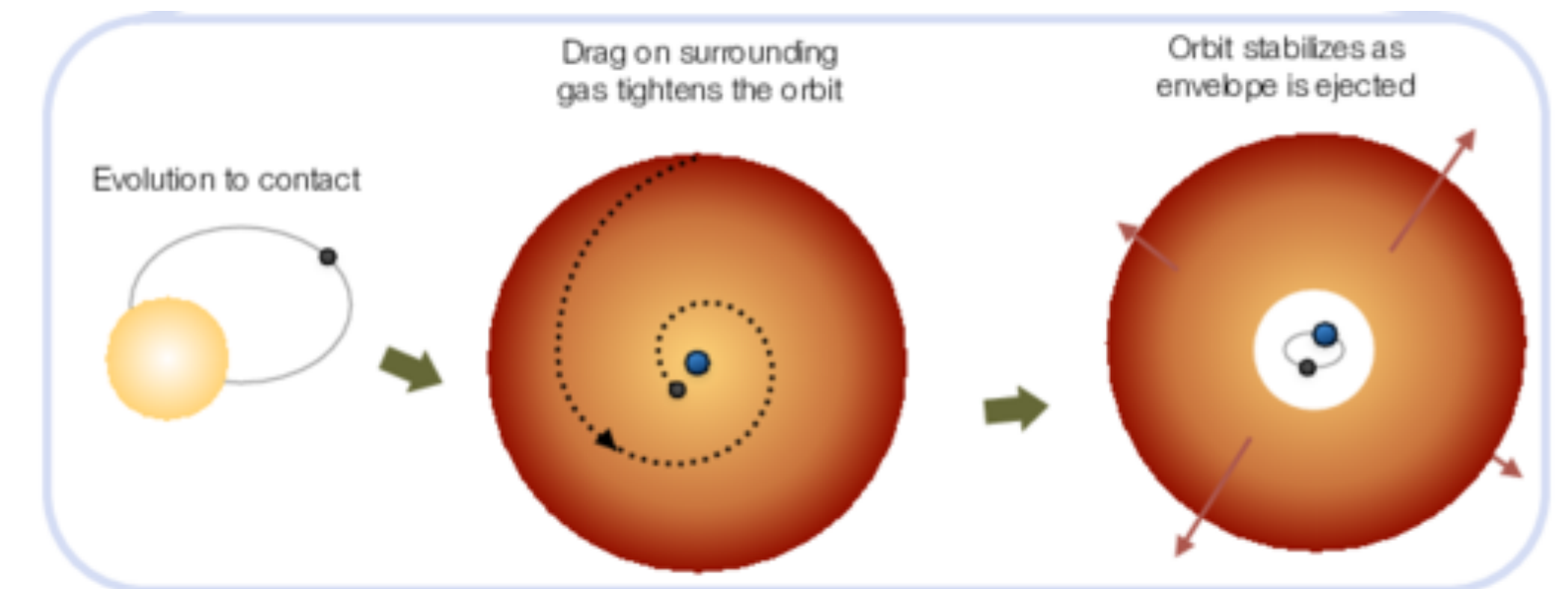
Metallicity

- Metallicity strongly affects both **single and binary star evolution** through its impact on **stellar winds** (e.g. higher metallicity → stronger stellar winds → greater mass loss during stellar evolution)



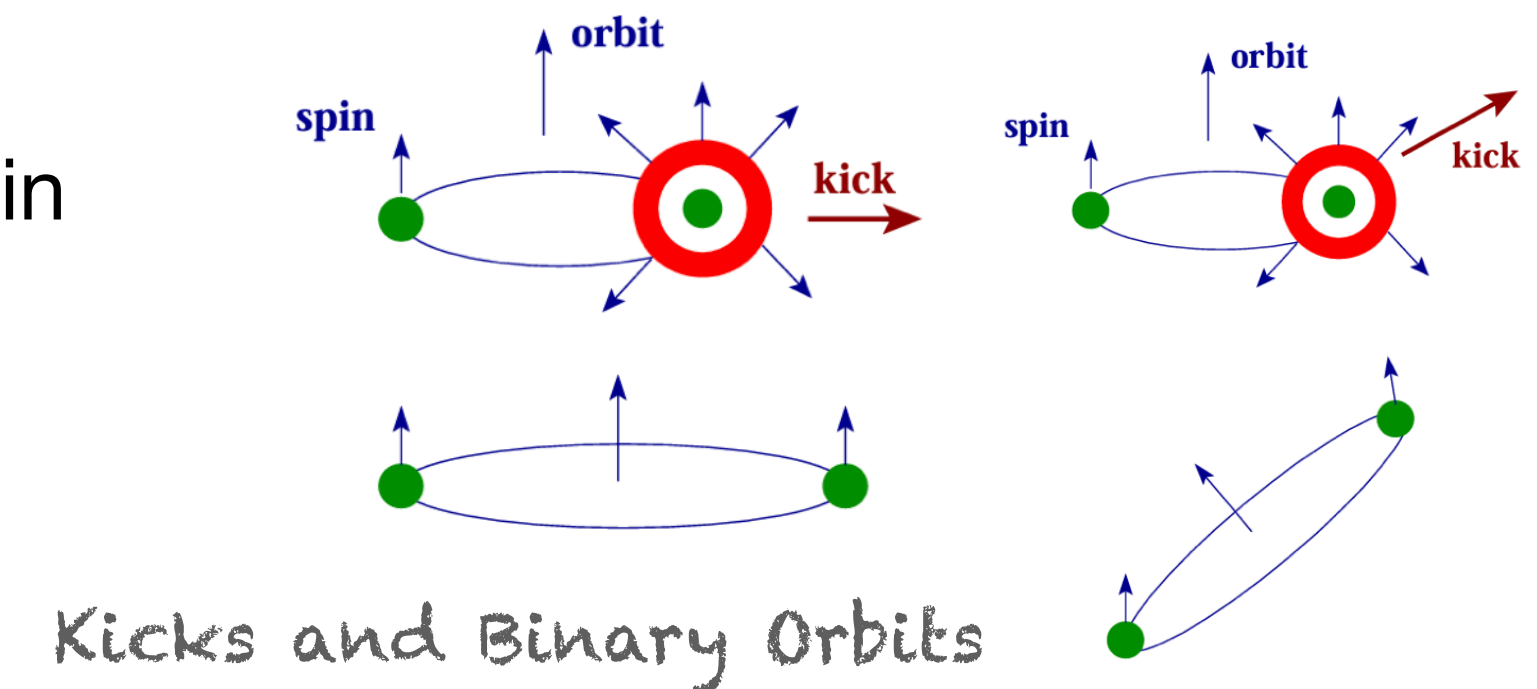
Common Envelope

- Common Envelope determines whether the binary will survive and how tight its orbit will be after envelope ejection



BH natal kick

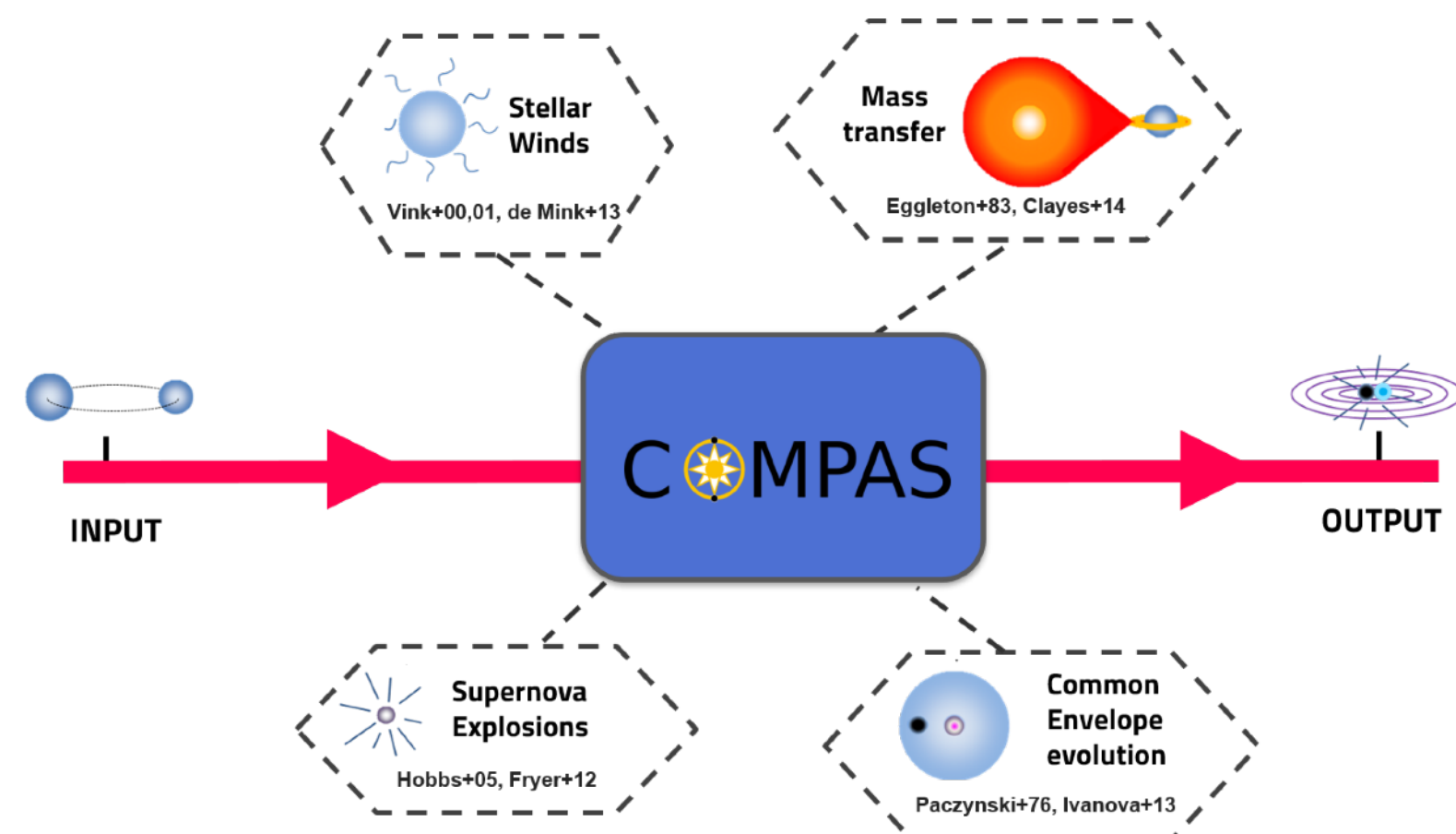
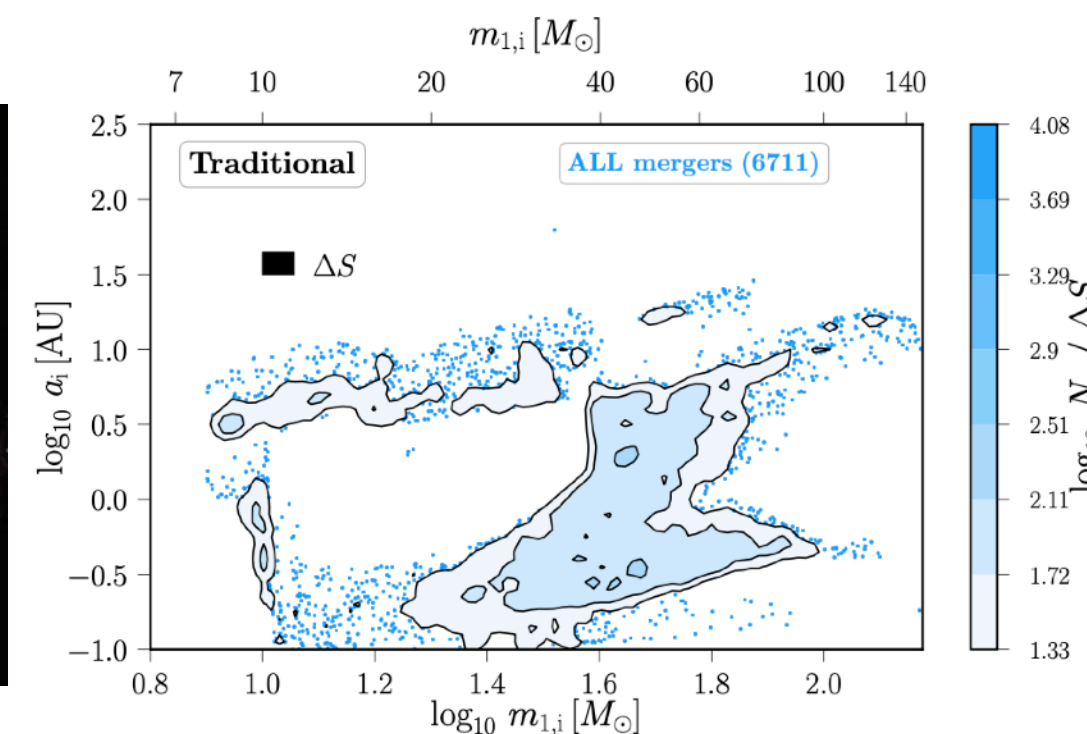
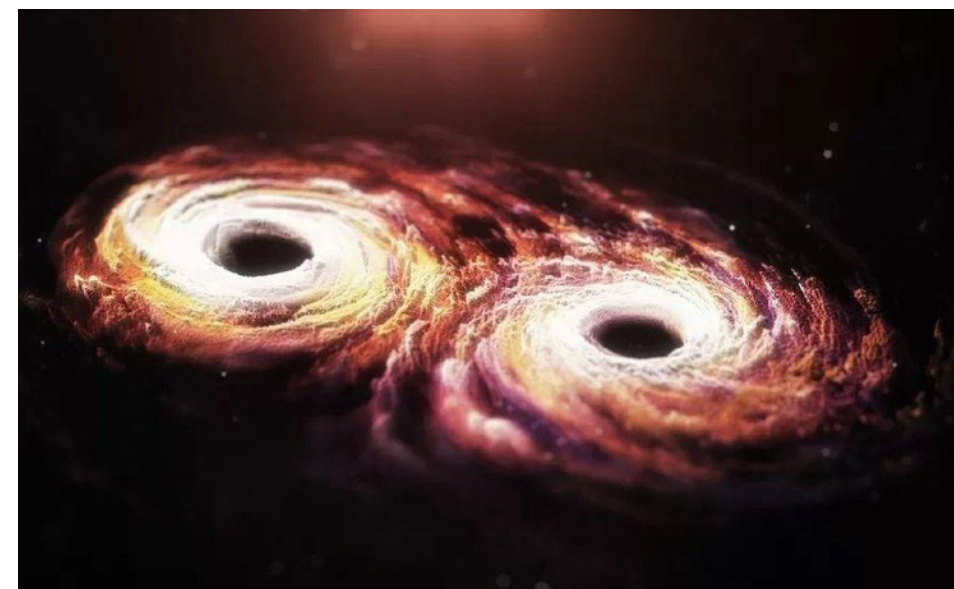
- Compact objects are expected to receive a natal kick from the parent SN explosion (arise from asymmetries in SN ejecta and neutrino flux)
- Kicks can **unbind binaries** or alter orbital properties (eccentricity, spin misalignment)



Binary Black Hole Population Synthesis

Binary Black Hole Merger

Gravitational-wave sources are a Rare outcome among stars



Why brute-force simulation is infeasible

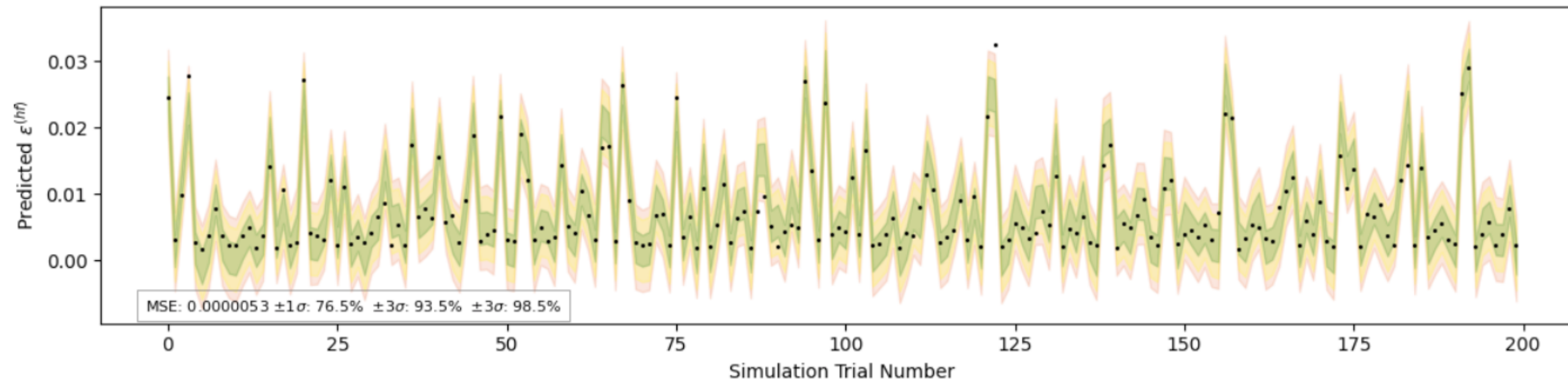
- **Extremely low signal rate:** 10^{-5} - 10^{-4} captures per primary
- **High variance:** Each design point needs $\sim 10^6$ primaries to stabilize estimates
- **Expensive HF run:** ~ 200 CPUh per configuration, statistical noise $\sim 1.6\%$
- **Cheap LF run:** ~ 0.2 CPUh per configuration, reduced number of primaries ($N=5000$), statistical noise $\sim 16.5\%$
- **Large search space:** Hyperparameter inference spans a multi-dimensional astrophysical parameter space, needing $O(10^3-10^4)$ model evaluations.
- **Total cost:** 10^5 CPUh \rightarrow far beyond practical turnaround time for parameter searches or parameter inference.

In practice, groups can only afford a small number of high-fidelity population models, far too few for scientific inference.

Most is spent generating irrelevant background \rightarrow **HF-only optimization is not practical.**

Experiment II: Result

Validation: Comparison of 100 HF Simulation Trials vs Model Prediction



- **Setup:** 1000 LF points · 15 HF points
- **Accuracy:**
 - **$\pm 1\sigma$: 76.5%** · **$\pm 2\sigma$: 93.5%** · **$\pm 3\sigma$: 98.5%** · **MSE = $5.3e-06$**
- **Surrogate performance:**
 - Surrogate **matches HF simulations** across the full parameter space

Efficient, accurate population modeling with orders-of-magnitude lower high-fidelity simulation demand.

- **Population-level predictions are stable** even with minimal HF data
- **Impact on compute:**
 - Uses **only 1.6% of HF compute vs. brute-force simulation**
 - Enables **rare-event likelihood evaluation** at a tiny fraction of the original compute
 - Demonstrates that the **approach scales to larger astrophysical models** without increasing HF cost

RESOLVE: Rare Event Surrogate Likelihood for Gravitational Wave Paleontology Parameter Estimation



Alex Migala
UCSD



Adam Boesky
Harvard



Alan Poon
LBNL

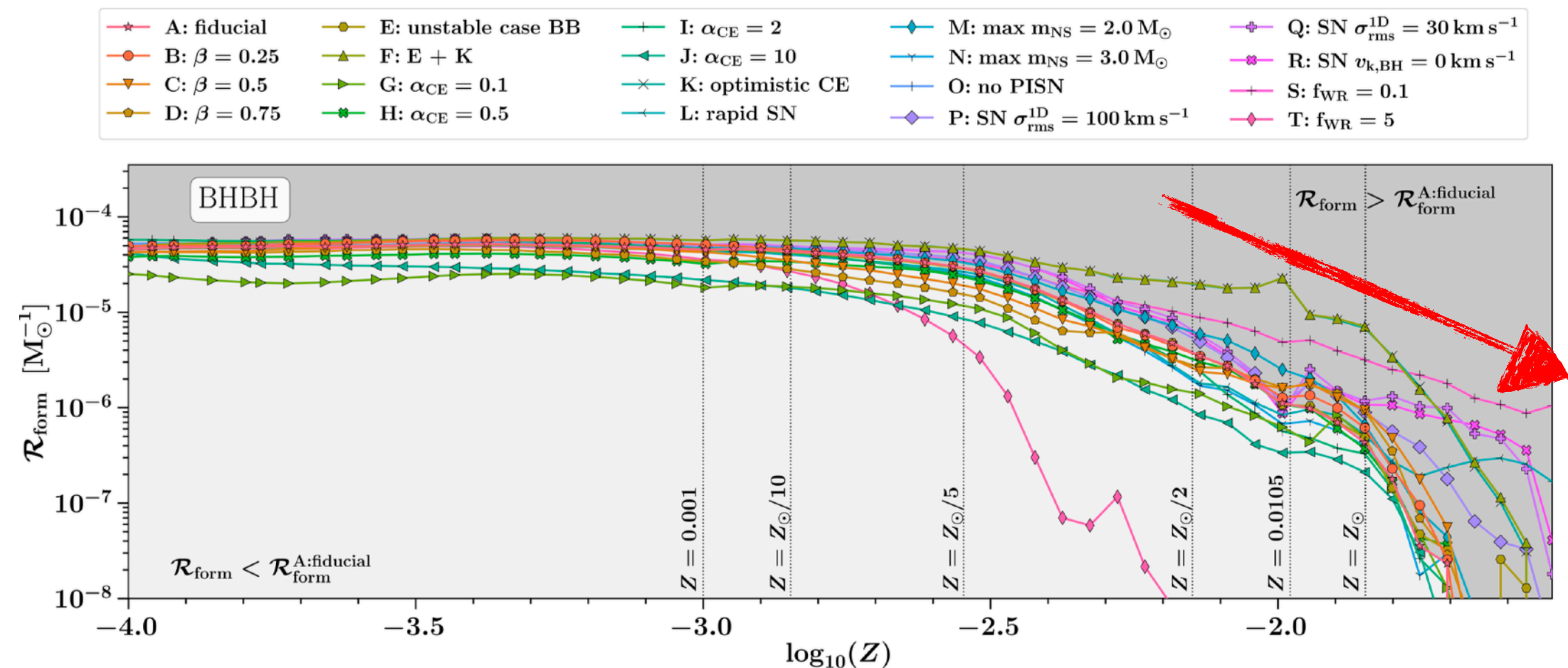


Floor Broekgaarden
UCSD



Aobo Li
UCSD

What do we know about the dependence of the BBH merger rate on metallicity?



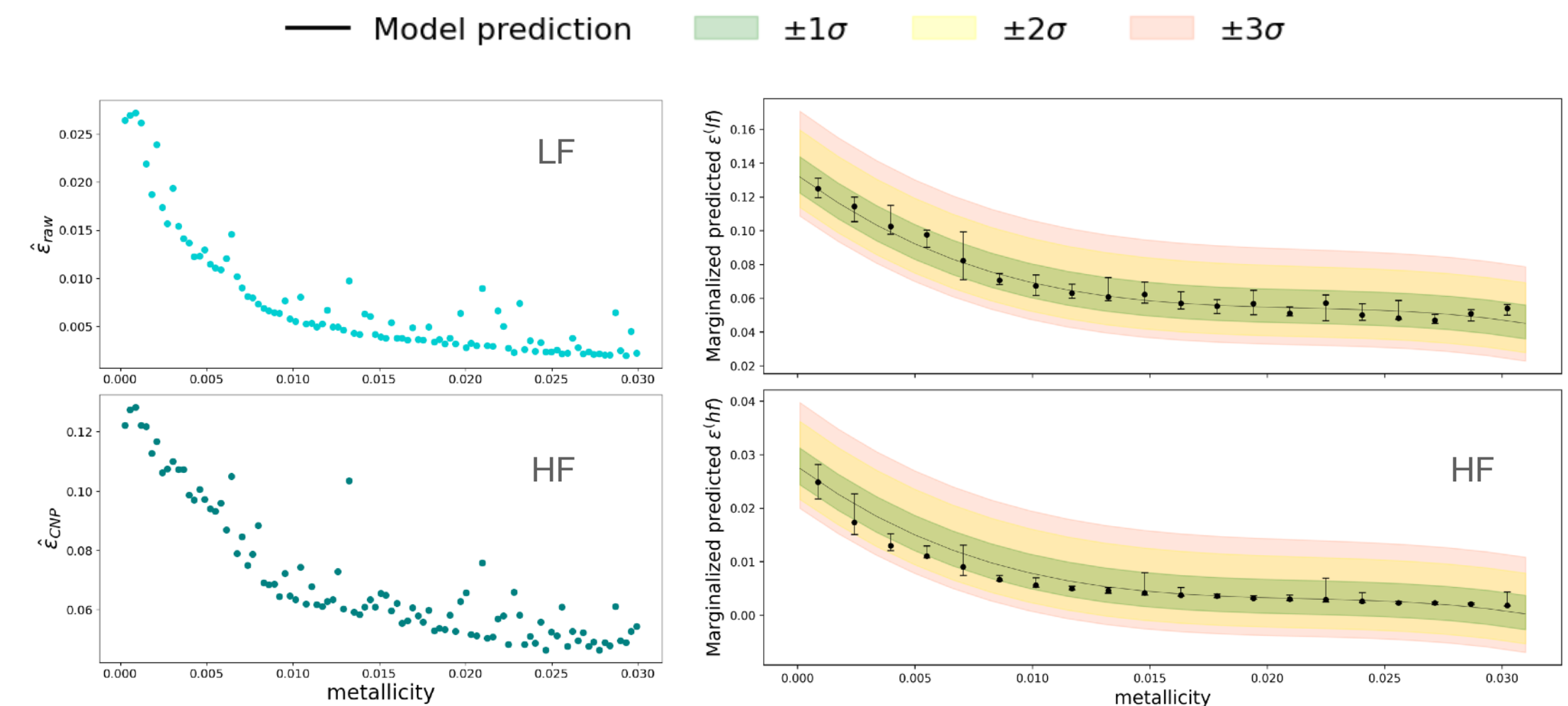
[Broekgaarden et al. (2022)]

- BBH formation rate **decreases strongly with metallicity** (due to weaker stellar winds and tighter binaries)
- The **local universe is more metal-rich** → lower BBH formation rates
- **Early low-metallicity** star formation **dominates present-day BBH** mergers due to delay time

Metallicity strongly affects both **single and binary star evolution** through its impact on **stellar winds**

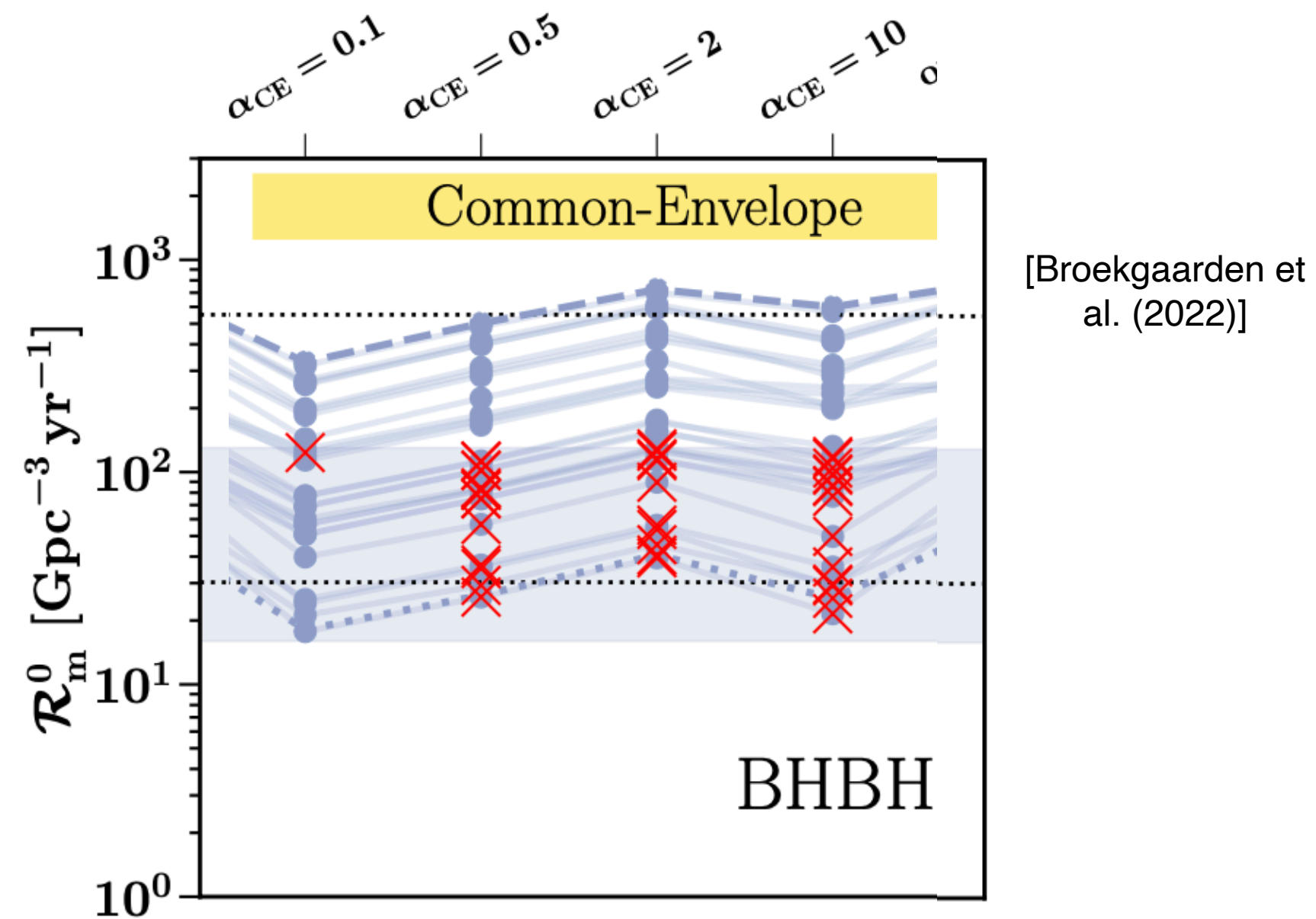
(e.g. higher metallicity → stronger stellar winds → greater mass loss during stellar evolution)

Model Result



- Model confirms strong metallicity dependence
- HF result shows no leveling at high Z, likely due to data sparsity

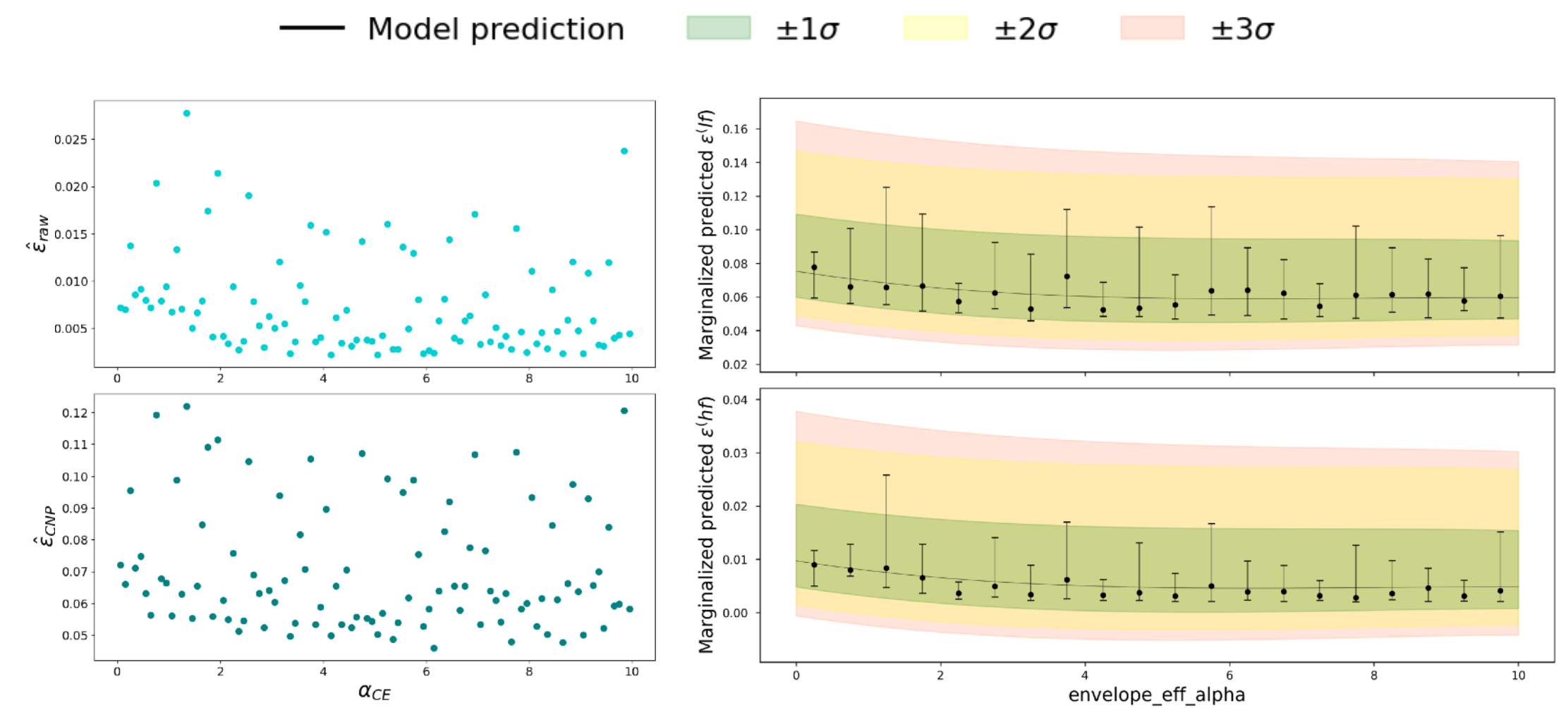
What do we know about the dependence of the BBH merger rate on the common envelope efficiency?



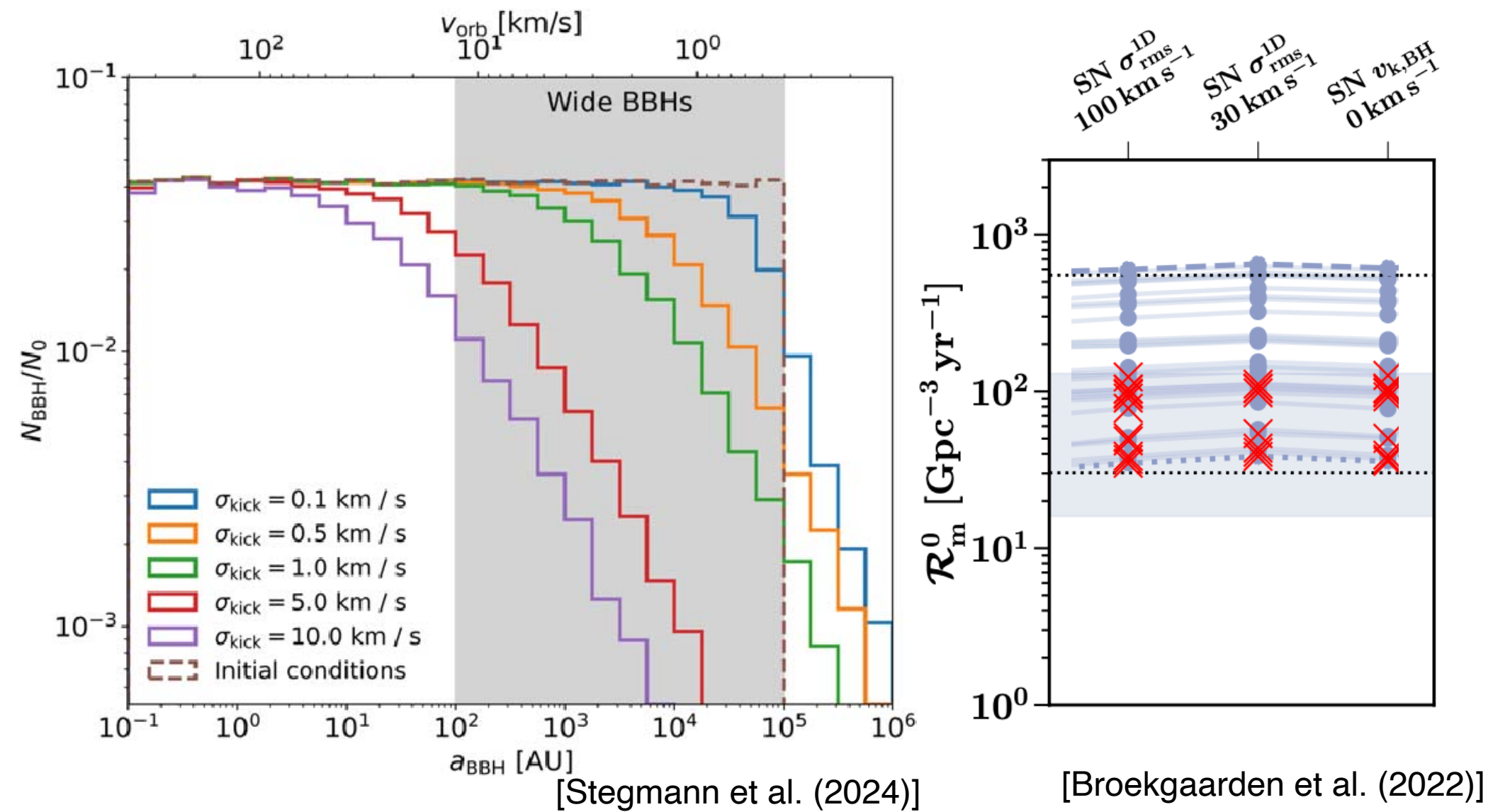
- Too low \rightarrow binaries merge during CE phase \rightarrow few DCOs
- Too high \rightarrow post-CE orbits too wide \rightarrow binaries don't merge in Hubble time
- Sweet spot $\alpha_{CE} \sim 0.5 - 2.0 \rightarrow$ most BBHs form and merge

Common Envelope determines whether the binary will survive and how tight its orbit will be after envelope ejection

Model Result



What do we know about the dependance of the BBH merger rate on the natal kick?

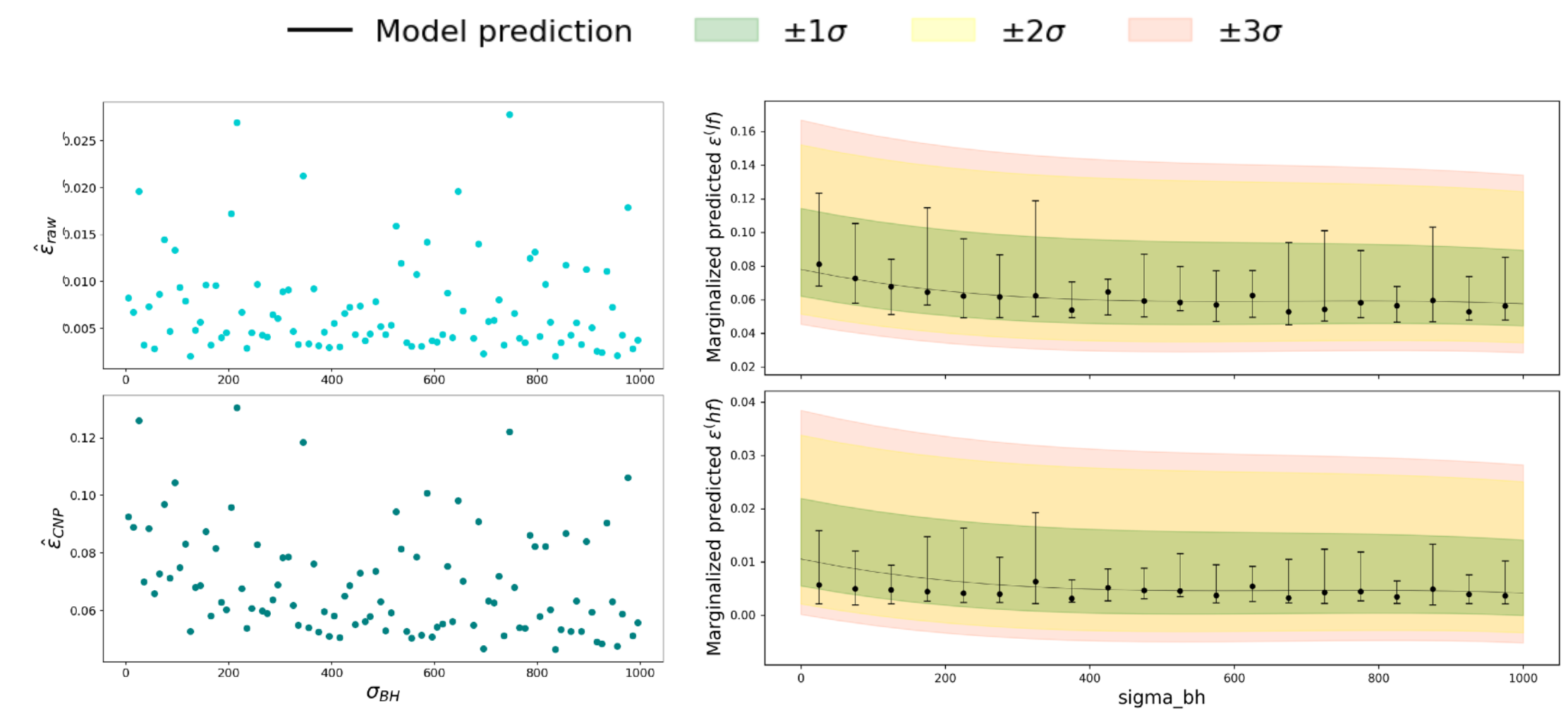


- Higher natal kicks tend to suppress BBH merger rate
- Low kicks help binaries stay bound; high kicks disrupt systems
- Formation efficiency shows non-monotonotic dependence

Compact objects are expected to receive a natal kick from the parent SN explosion (arise from asymmetries in SN ejecta and neutrino flux)

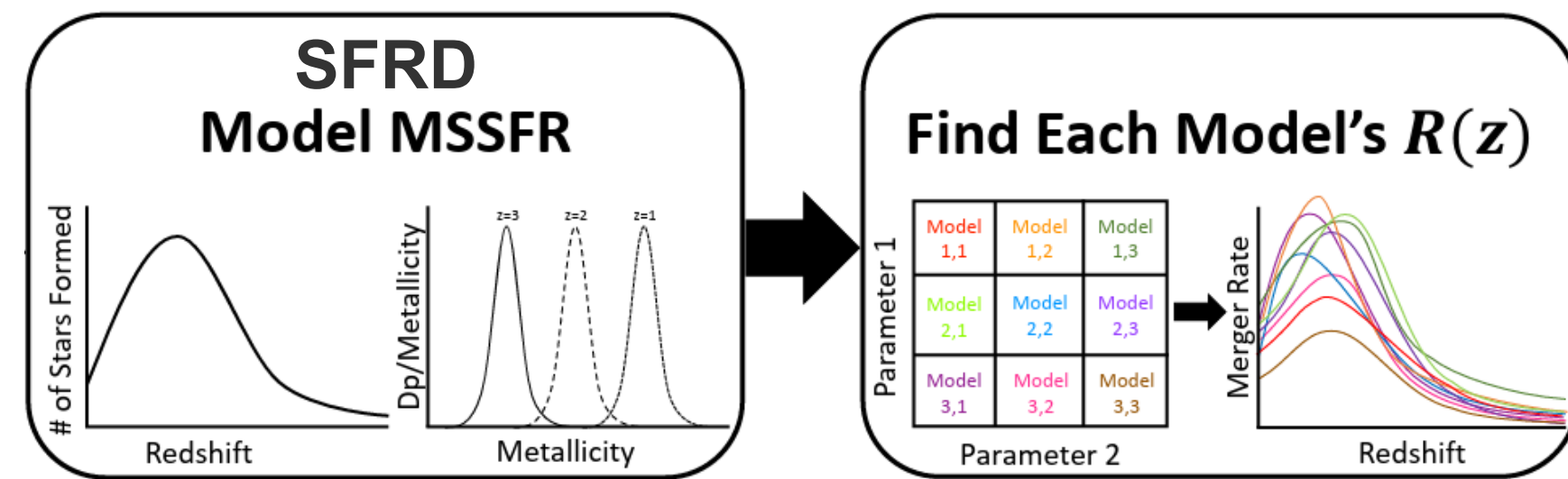
Kicks can **unbind binaries** or alter orbital properties (eccentricity, spin misalignment)

Model Result



Proof-of-Principle: Parameter Inference

- The predicted BBH merger rate is computed by convoluting the formation efficiency with $SFRD(z)$



We simplified using $z_{obs} = 0$ and $t_d = 0$, scaling $SFRD(z = 0) \approx 0.015 M_{sol} yr^{-1} Mpc^{-3}$

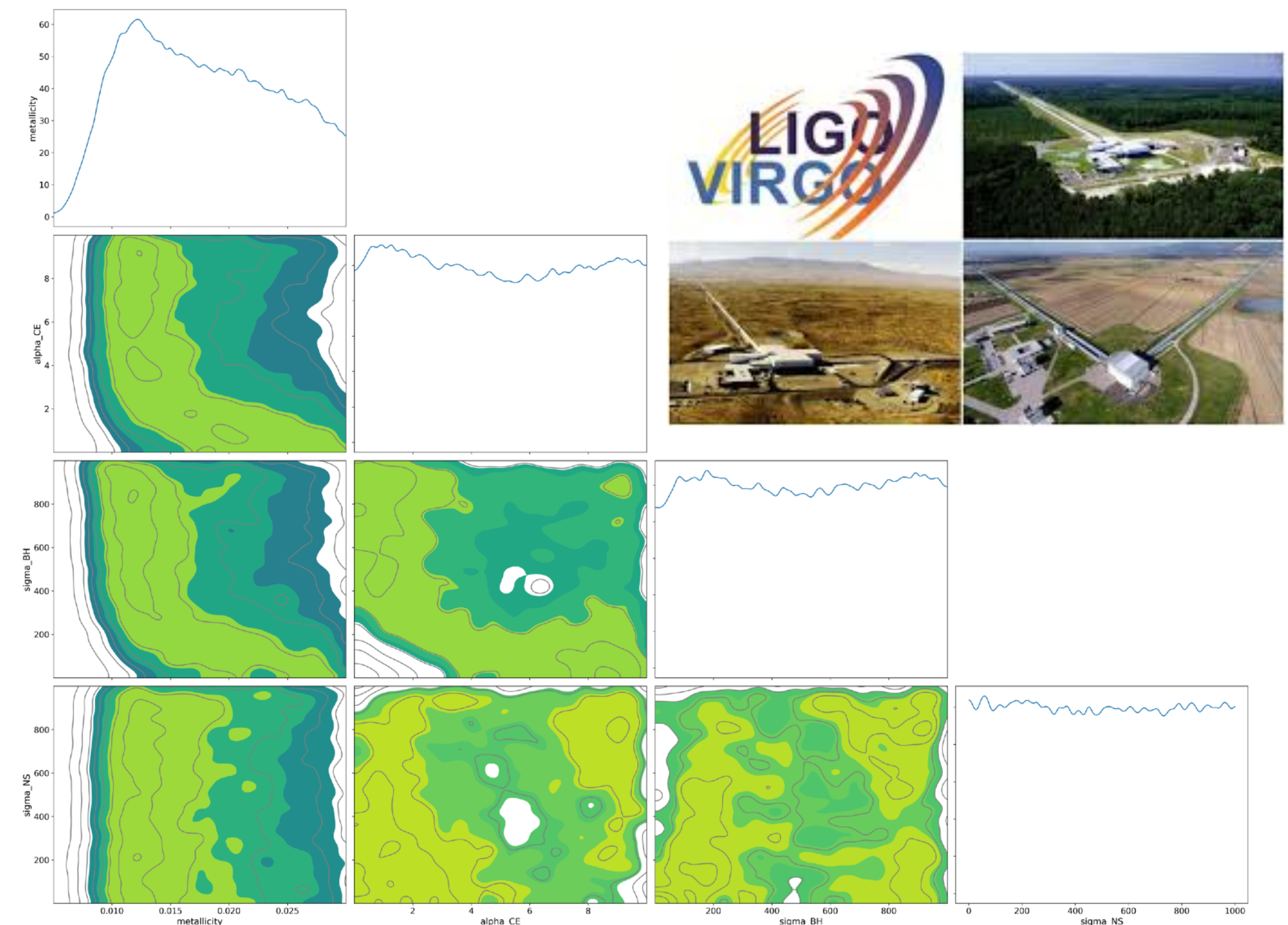
We infer binary evolution parameters by matching predicted BBH merger rate to the observed GW merger rate.

- Metallicity posterior peaks at $0.0172^{+0.0079}_{-0.0062}$ near solar Z_{\odot} , but with large uncertainty.
- α_{CE} , σ_{BH} , σ_{NS} remain **weakly constrained** due to known **degeneracies**.
- Results align with expectations from population synthesis studies** (e.g., Neijssel+2019, Broekgaarden+2021).
- Simplified SFRD** scaling adds further uncertainty, especially in Z

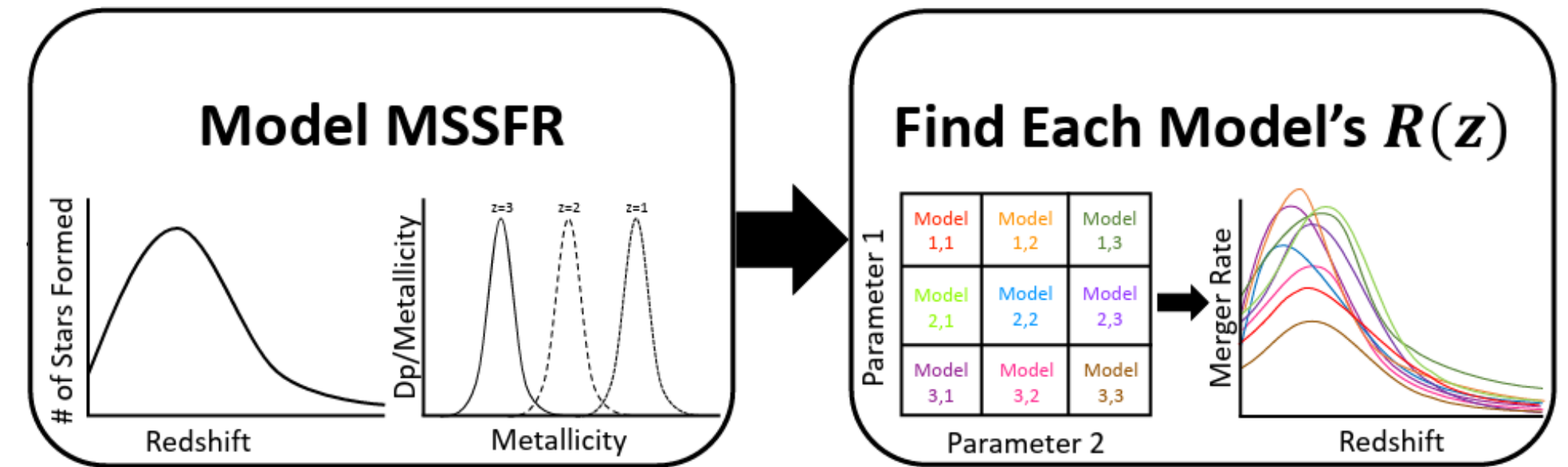
LIGO-Virgo collaboration reported an observed gravitational wave rate of

$$y_{obs} = 17^{+10}_{-6.7} yr^{-1} Gpc^{-3}$$

[GWTC-3, Abbott et al. (2023)]



BBH Merger Rate



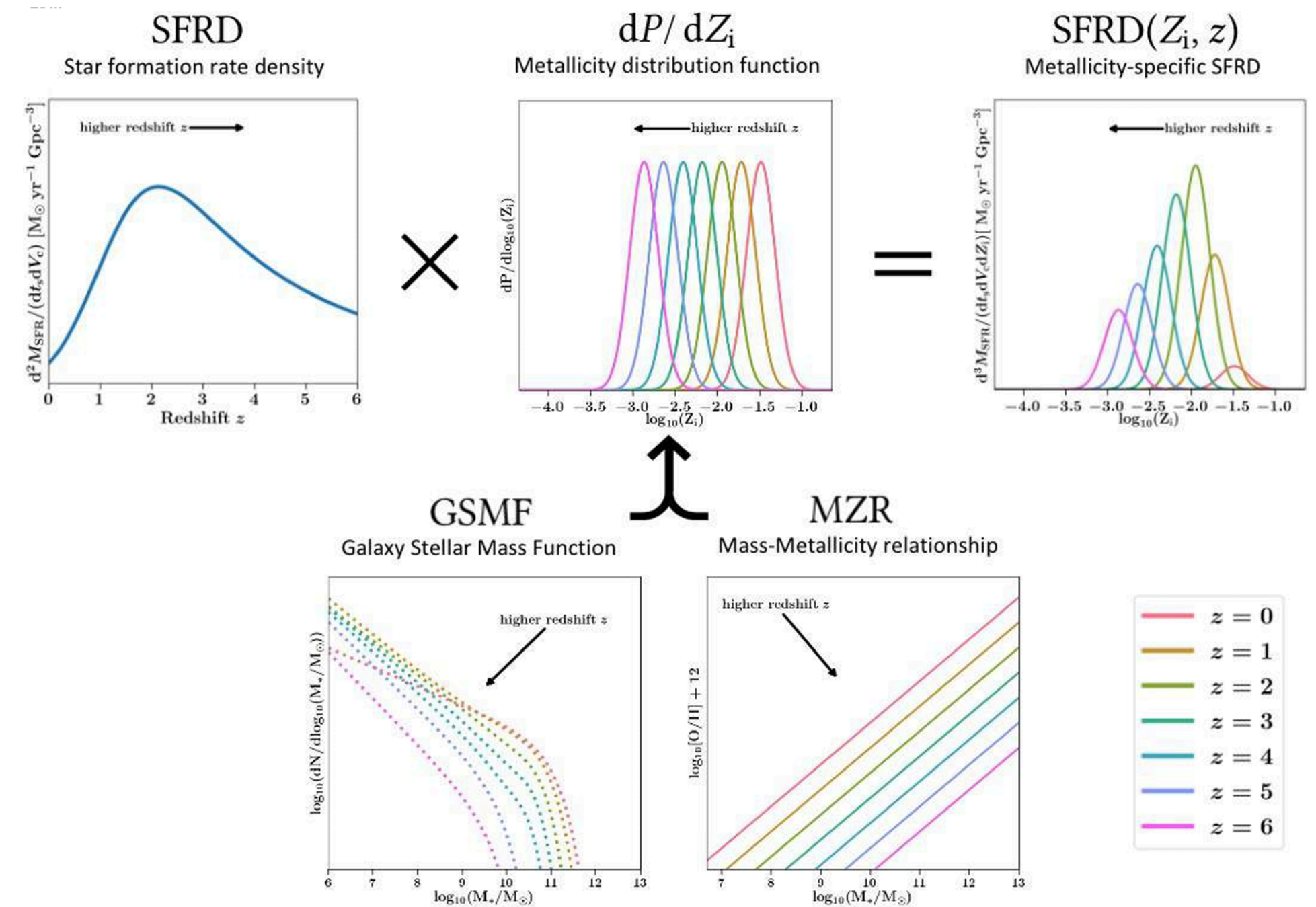
- The BBH merger rate is computed by convoluting the formation efficiency with $SFRD(z)$ with **metallicity-dependent** dP/dZ_i
- Full model integrates over formation redshift with delay time:

$$R_{BBH}(z_{obs}) = \int_{z_{obs}}^{\infty} SFRD(z_{form}) \cdot \epsilon_{BBH}(Z(z_{form})) \cdot P(t(z_{form}) - t(z_{obs})) \frac{dt}{dz_{form}} dz_{form}$$



Full Cosmic Integration requires **event-level modeling** of delay time

Star Formation Rate Distribution

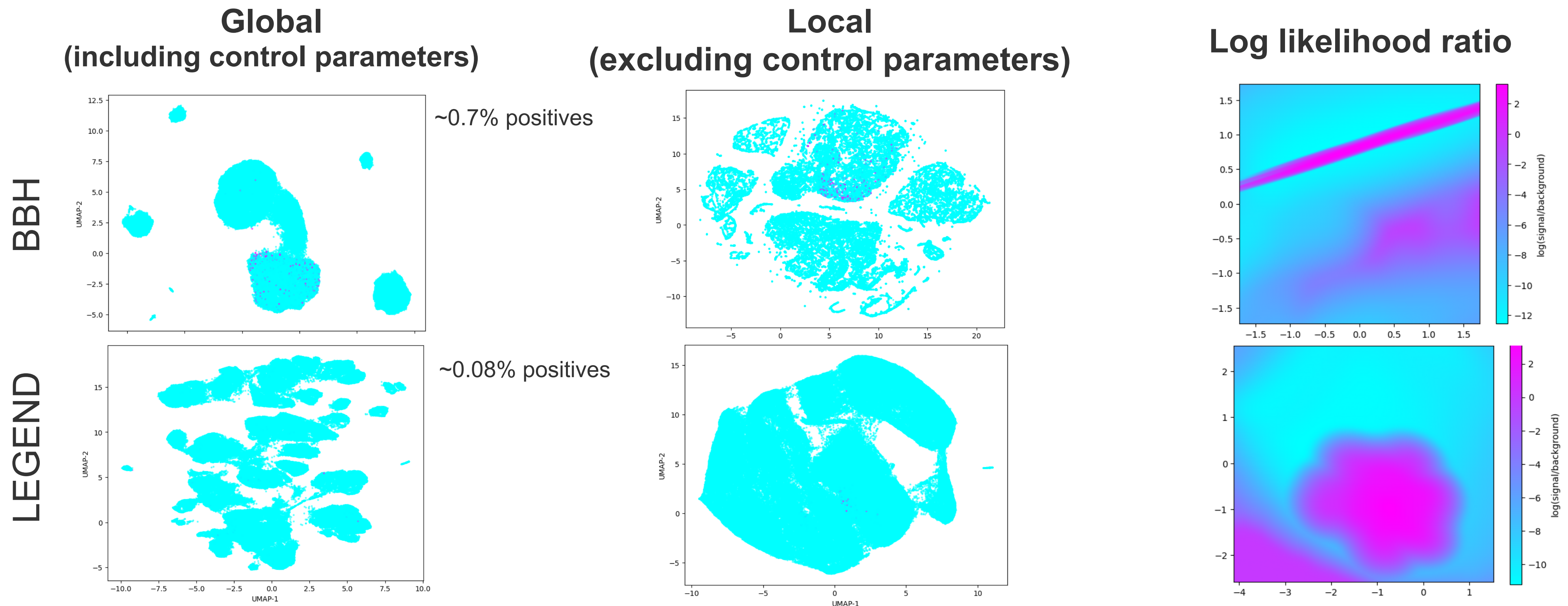


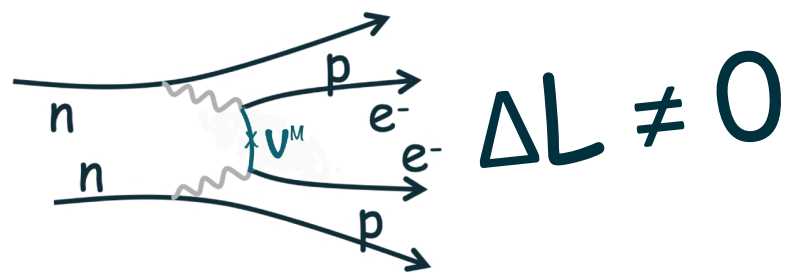
Signal vs Background

- Rare \neq anomalous: signals are statistically scarce, not geometric outliers.
- Signal and background come from the same generative process \rightarrow strong feature overlap.
- No clean separation \rightarrow distance-based or outlier methods fail.

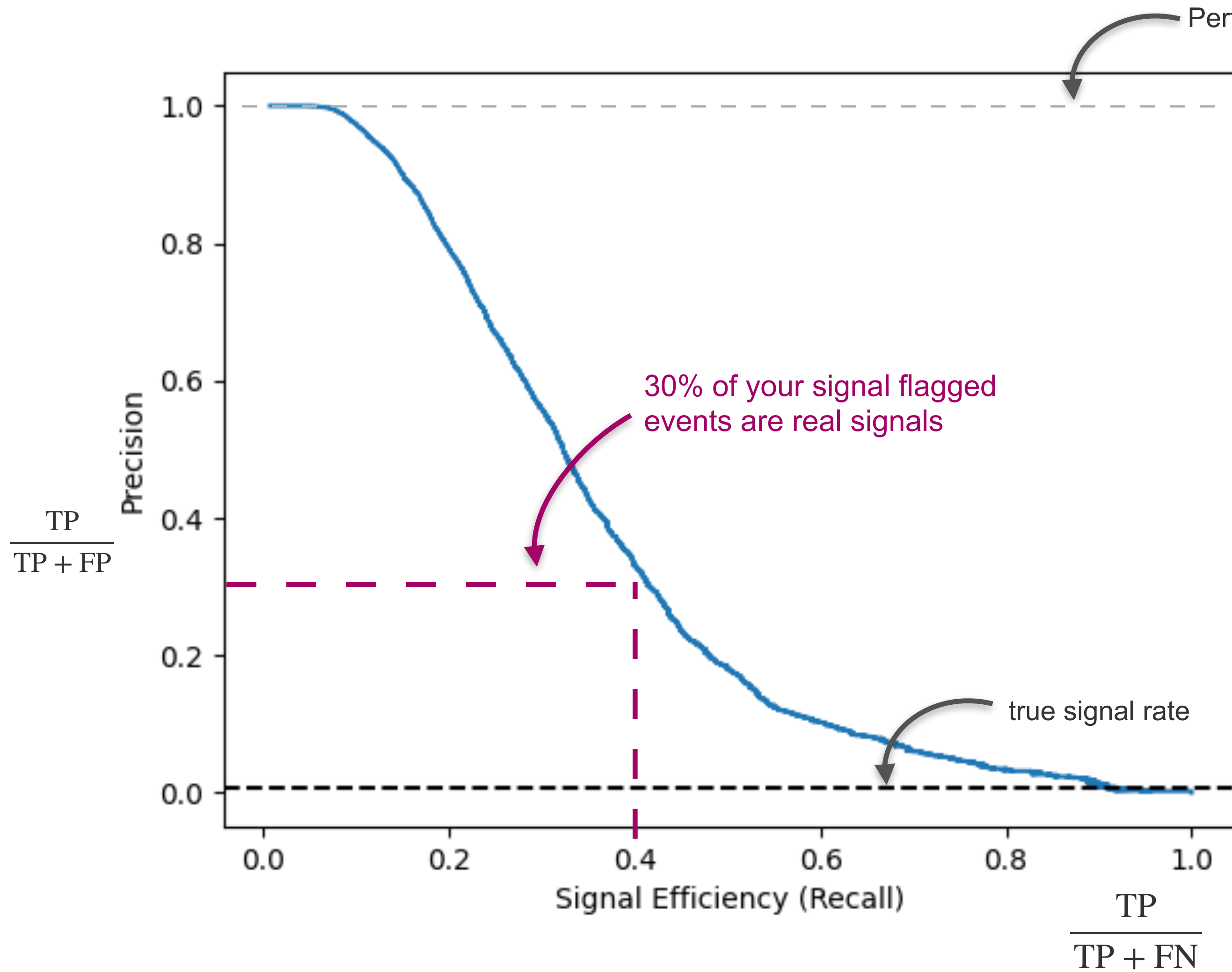
Implication for ML:

We need density-aware models that capture subtle local differences, not standard anomaly detection.





Precision Recall Curve



How to read it

Pick a point on the curve:

- The x value tells you what fraction of signals you're catching.
- The y value tells you how pure your selected events are (how many are true vs. background).

Example:

At recall = 0.4 (you catch 40% of signals), precision = 0.3 means 30% of your flagged events are real signals.

Traditional Machine Learning Fails

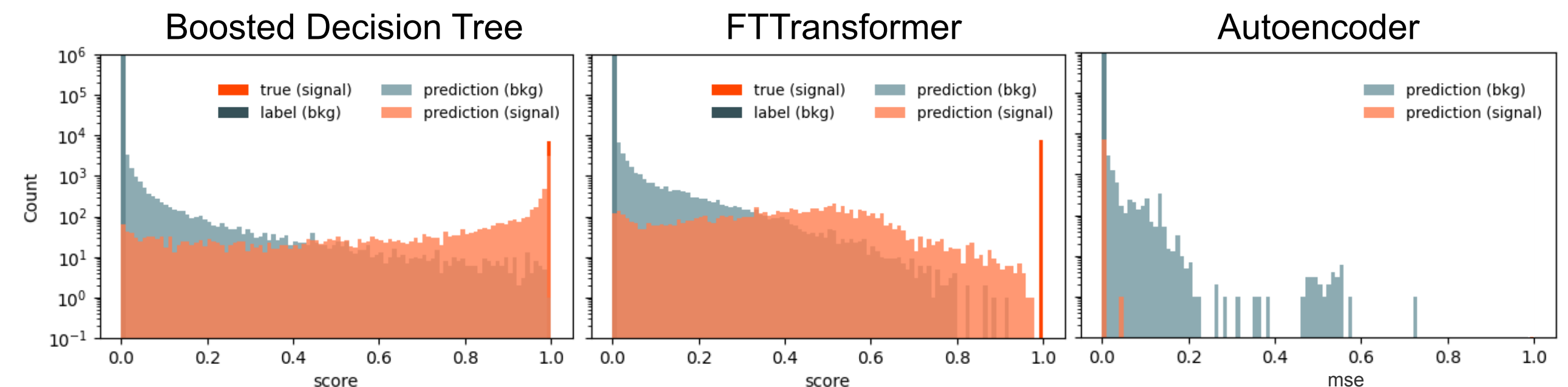
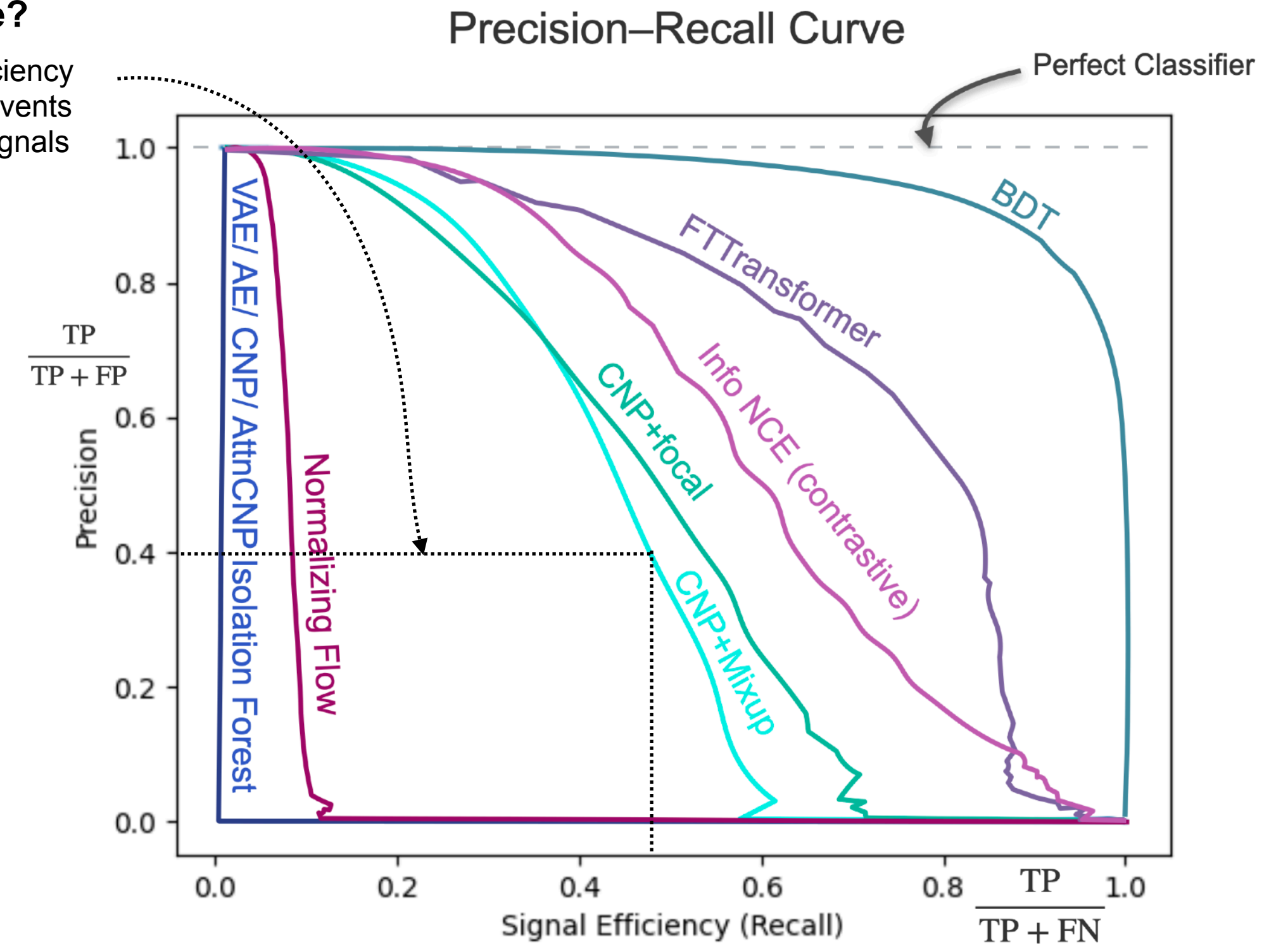
Why Standard ML Methods Fail

- **Representation learning collapses:**
 - Autoencoders, VAEs → learn dominant background; **ignore rare structure.**
 - Normalizing flows → **assign similar likelihoods** to signal & background.
 - Contrastive / graph methods → **cannot separate overlapping manifolds.**
- **Reweighting is insufficient:**
 - Oversampling, focal loss, and importance weights **shift attention** but **do not create** discriminative structure.

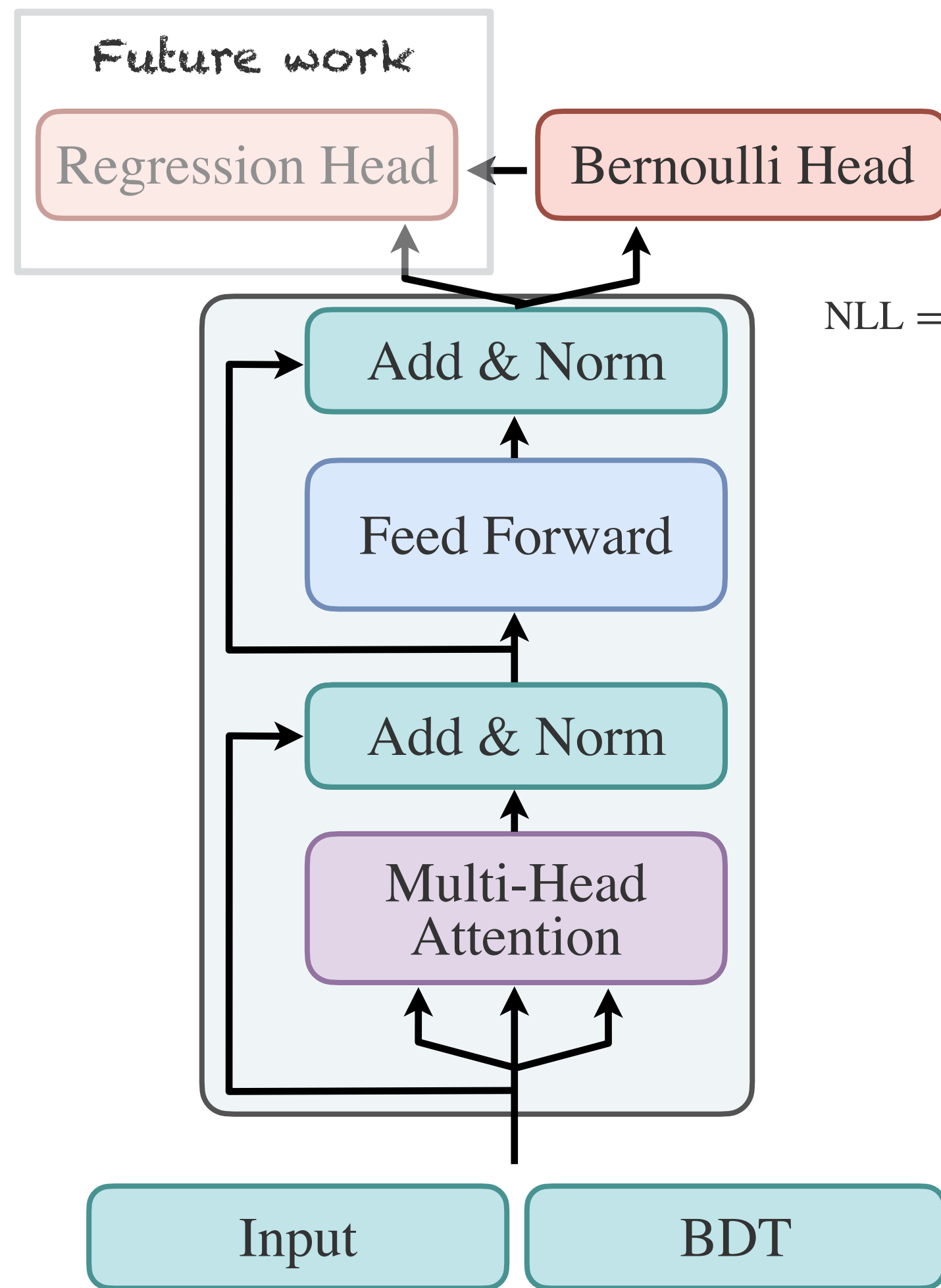
Traditional ML cannot resolve rare events embedded inside high-density backgrounds, the model must learn **local probability structure**, not outlier detection.

How to read curve?

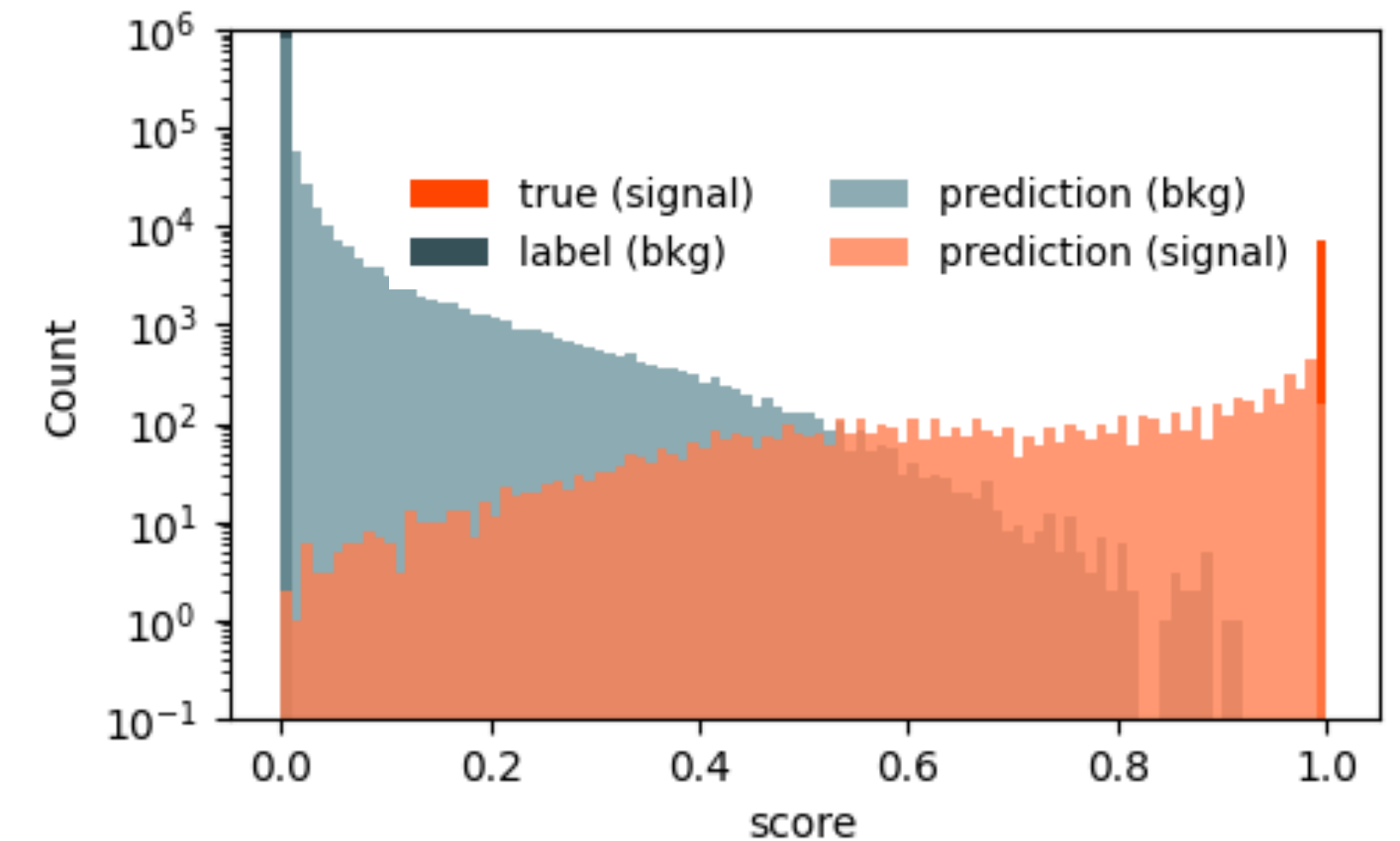
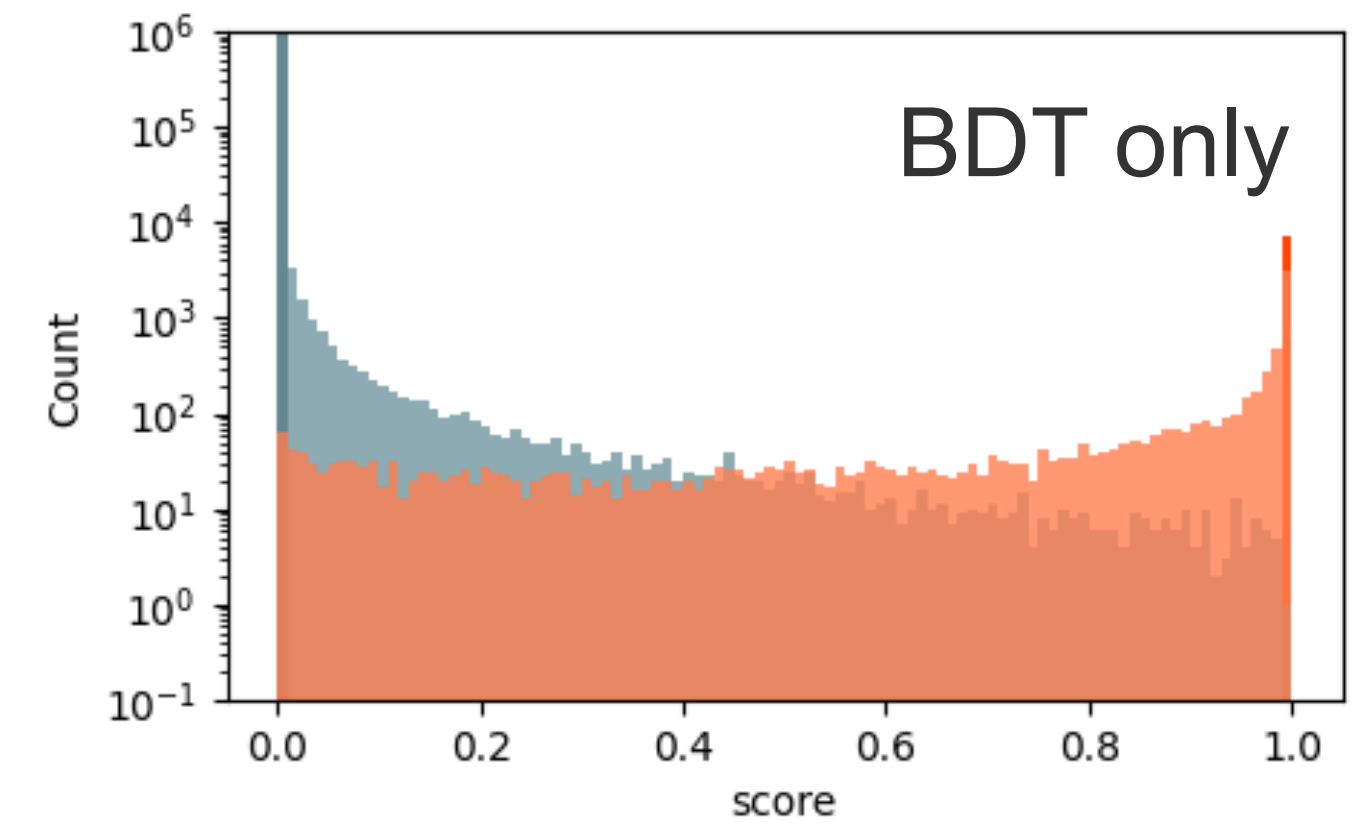
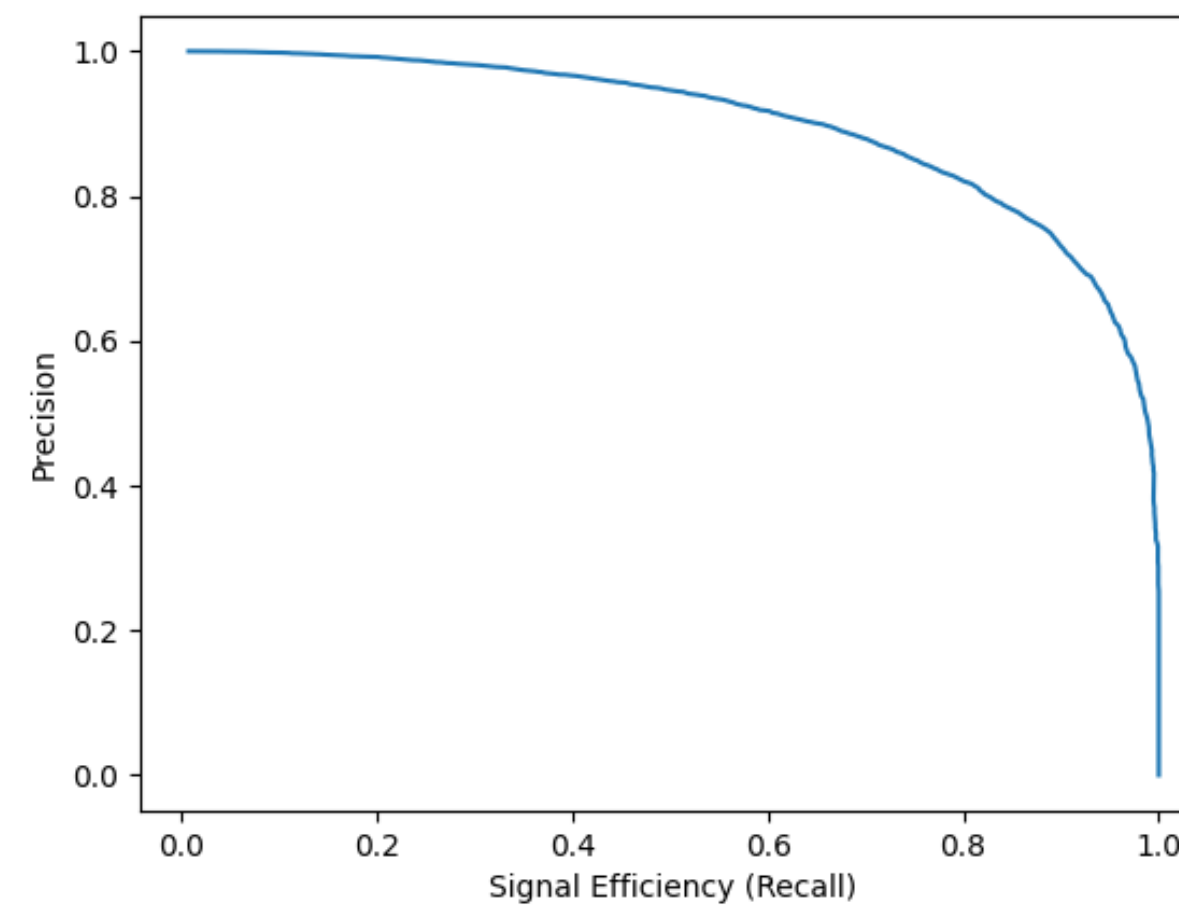
e.g. @ 48% Signal Efficiency
40% of signal flagged events
(score > 0.5) are real signals



FT Transformer with BGT Encodings (Preliminary)



$$\text{NLL} = -\log \int_{-\infty}^{\infty} \frac{\text{Bernoulli}(y | \sigma(\ell)) \mathcal{N}(\ell | \mu, \sigma^2) d\ell}{\sigma(\ell)^y (1 - \sigma(\ell))^{1-y}}$$



→ convert a BDT into a per-feature representation by encoding the decision path that feature follows

- BDT encodings provide structured priors that improve rare-event discrimination beyond standalone transformers or BDTs
- Hybrid model offers a promising path toward more robust scientific ML for rare events.

Resolve: Unified Test Suite for Rare Event Modeling

Modular architecture

- Plug-and-play backends: AE, VAE, CNP, NF, FT-Transformer, custom PyTorch models
- Benchmark different models on HPC datasets without touching pipeline code

Unified training pipeline

- Config-driven runs, checkpointing, logging, reproducibility
- Automatic sharding + batching for large I/O workloads

Contrastive + generative tools

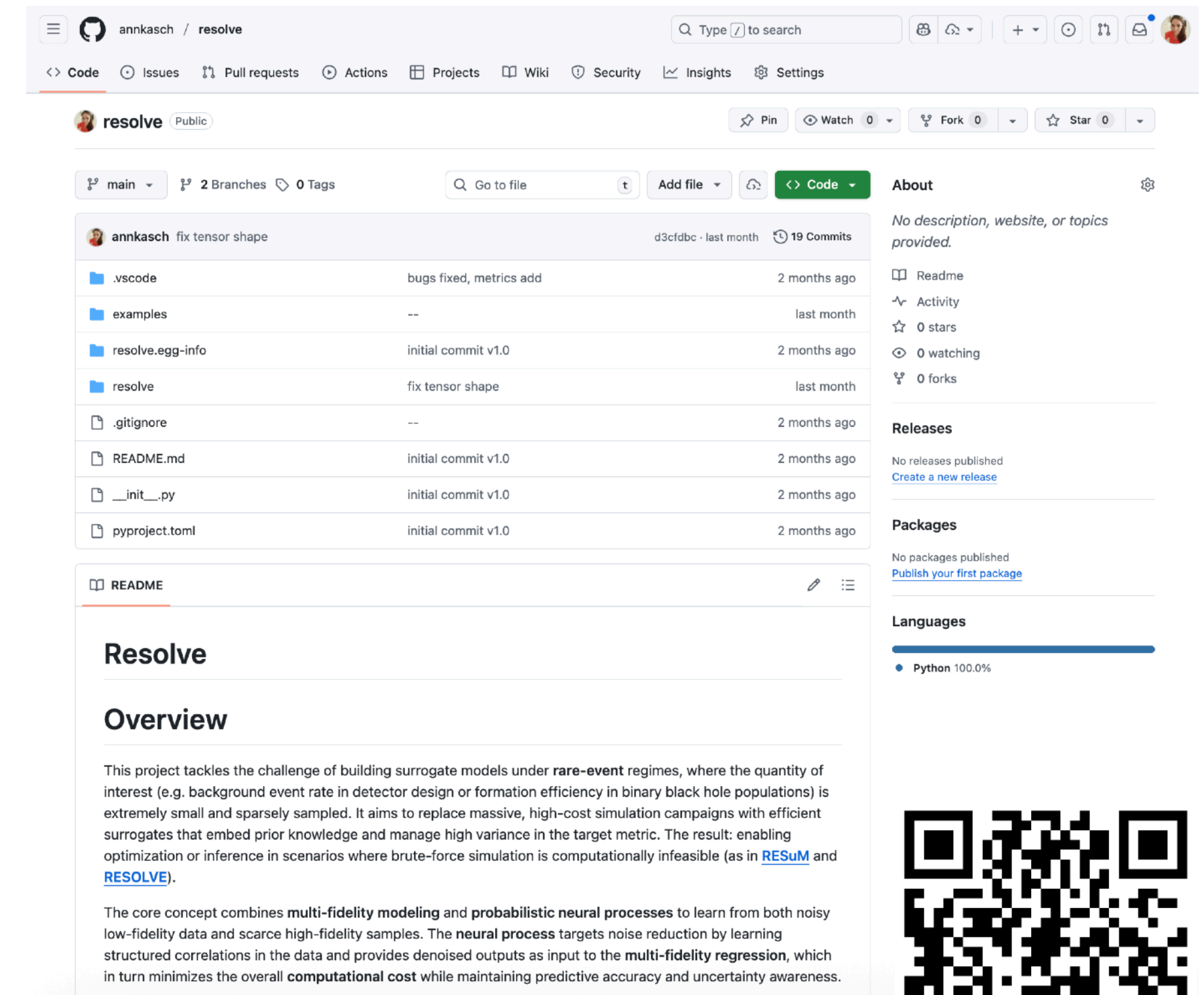
- Built-in SupCon and InfoNCE for representation learning

Flexible loss framework

- MSE, BCE, Gaussian NLL + custom rare-event losses; easily extensible

Class imbalance handling

- Oversampling schedulers + curriculum logic designed for 10^{-7} imbalance
- Reusable across HEP, astrophysics, population modeling
- Designed for integration with HPC simulation workflows



The screenshot shows the GitHub repository page for 'annkasch / resolve'. The repository is public and has 19 commits. The file list includes .vscode, examples, resolve.egg-info, resolve, .gitignore, README.md, __init__.py, and pyproject.toml. The README is open, showing the project title 'Resolve' and an overview section. The overview text describes the project's goal: building surrogate models under rare-event regimes. It mentions the use of multi-fidelity modeling and probabilistic neural processes. A QR code is located in the bottom right corner of the screenshot.

Application Example: XLZD Optimization

Dark Matter (DM) Mode

Signal:

- Nuclear recoils (NR) at keV energies

Detection:

- Prompt scintillation S1
- Delayed charge-induced light S2

Key performance drivers:

- Low energy threshold
- Strong ER/NR separation via S2/S1
- High single-electron efficiency

Fields control:

- Recombination
- Charge yield
- NR band position

Neutrinoless Double Beta Decay

Signal:

- Two-electron decay at fixed energy $Q_{\beta\beta}$

Detection:

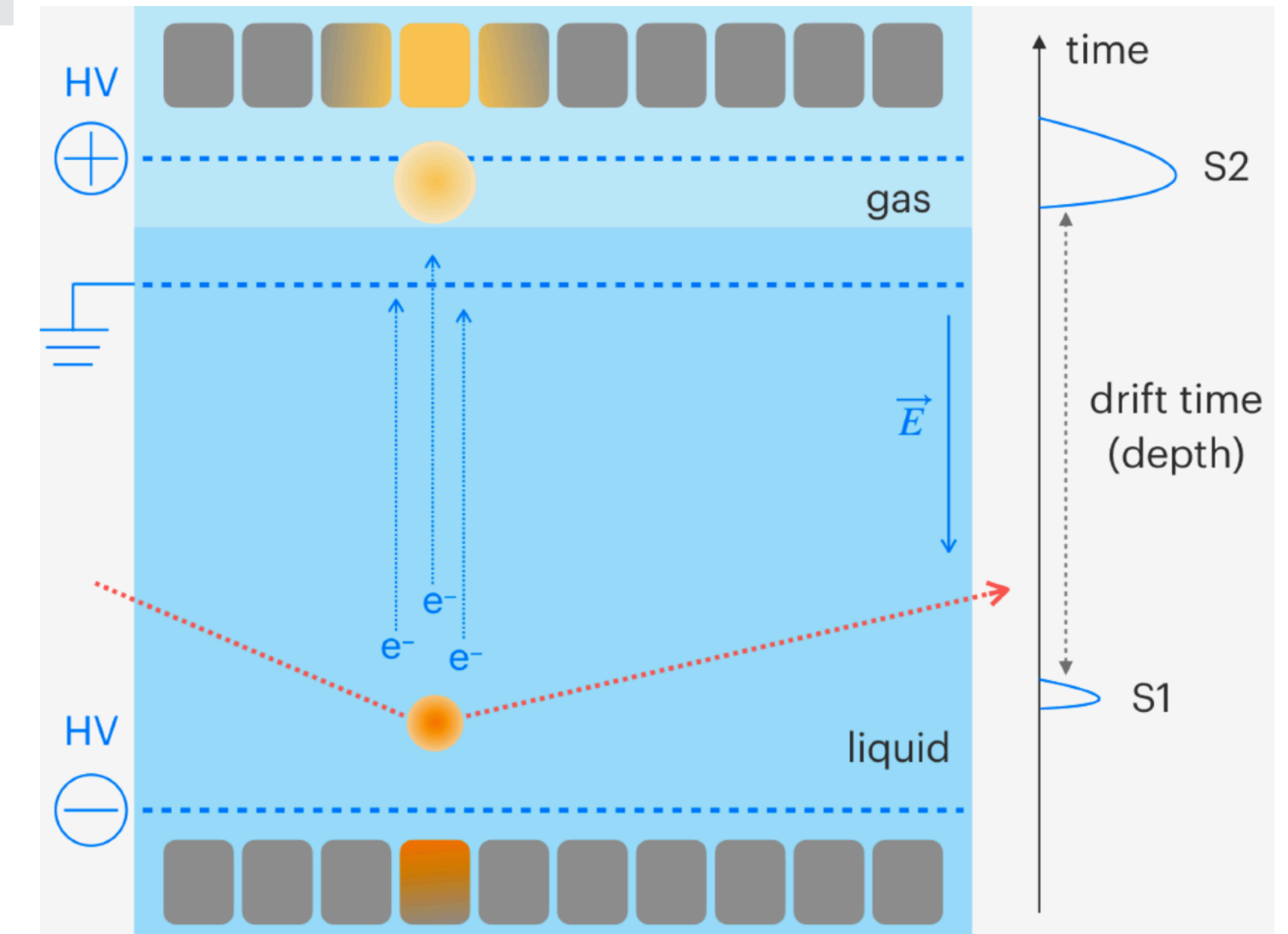
- Total reconstructed energy only

Key performance drivers:

- MeV-scale energy resolution
- Spectral purity in $\pm 1\%$ $Q_{\beta\beta}$ window

Fields control:

- Light/charge anti-correlation
- Fluctuations in combined energy scale



XLZD field tuning is not a single-parameter optimization. It is an intrinsically conflicting, multi-objective inference problem across keV and MeV physics simultaneously.

The Core XLZD Optimization Conflict

DM wants:

- Strong fields \rightarrow better charge extraction & ER/NR separation

$0\nu\beta\beta$ wants:

- Stable fields \rightarrow minimal recombination fluctuations & best energy resolution

From Astrophysical Populations to LSST Alerts: The Simulation

The Forward Chain Rubin Depends On

Population Physics → Radiative Transfer → Light Curves → Survey Cadence + Noise → LSST Alerts

- Defines alert rates, classifier performance, and rare-transient yields
- This entire chain must be simulated to:
- Train ML classifiers
- Validate alert pipelines
- Quantify selection effects

Why This Is Broken Today

- LSST scale: millions of alerts per night
- Truly rare events: $\sim 10^{-5}$ – 10^{-6} of the stream

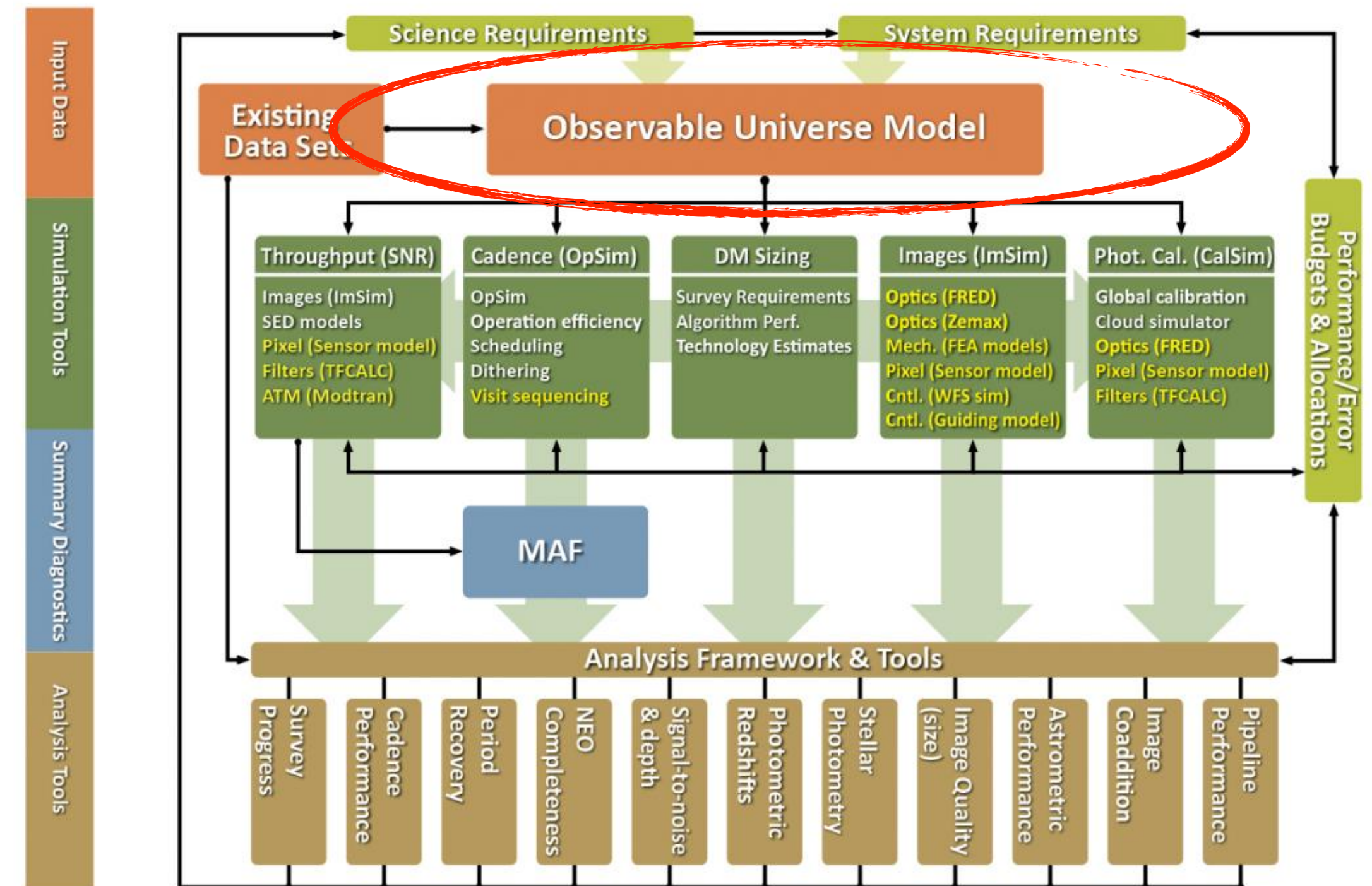
Population synthesis is:

- Too slow to regenerate
- Impossible to sweep over astrophysical uncertainty

Consequence:

- Population assumptions are frozen
- Classifiers are trained on fragile, non-updatable priors
- Rare-transient discovery is statistically inefficient

Sky population input to downstream simulators or ML models



Example of the Legacy CatSim-era architecture (has been replaced by SkyCatalogs in the population → sky layer)

RESuM: Rare-Event Surrogate Model

- Population-level multi-fidelity surrogate for extreme class imbalance
- Replaces brute-force population Monte Carlo with:
 - Uncertainty-aware inference
 - Active learning
 - Orders-of-magnitude speedup
- Validated on:
 - Detector design optimization (LEGEND)
 - Binary black hole populations

Unified Test Suit for Rare Event Modeling

- Unified, modular surrogate–inference stack
- Designed explicitly for rare-event simulation bottlenecks
- Flexible backend architecture:
 - Interchangeable ML models (CNPs, NF, Transformer, etc.)
 - Pluggable training logic (schedulers, sampling, loss functions)
- Core Capabilities
 - Hierarchical Bayesian inference
 - Multi-fidelity autoregressive modeling
 - End-to-end uncertainty propagation

See Details!

The Rare Event Surrogate Model for Physics Design (ReSUM)



RESOLVE: Rare Event Surrogate Likelihood for Gravitational Wave Paleontology Parameter Estimation



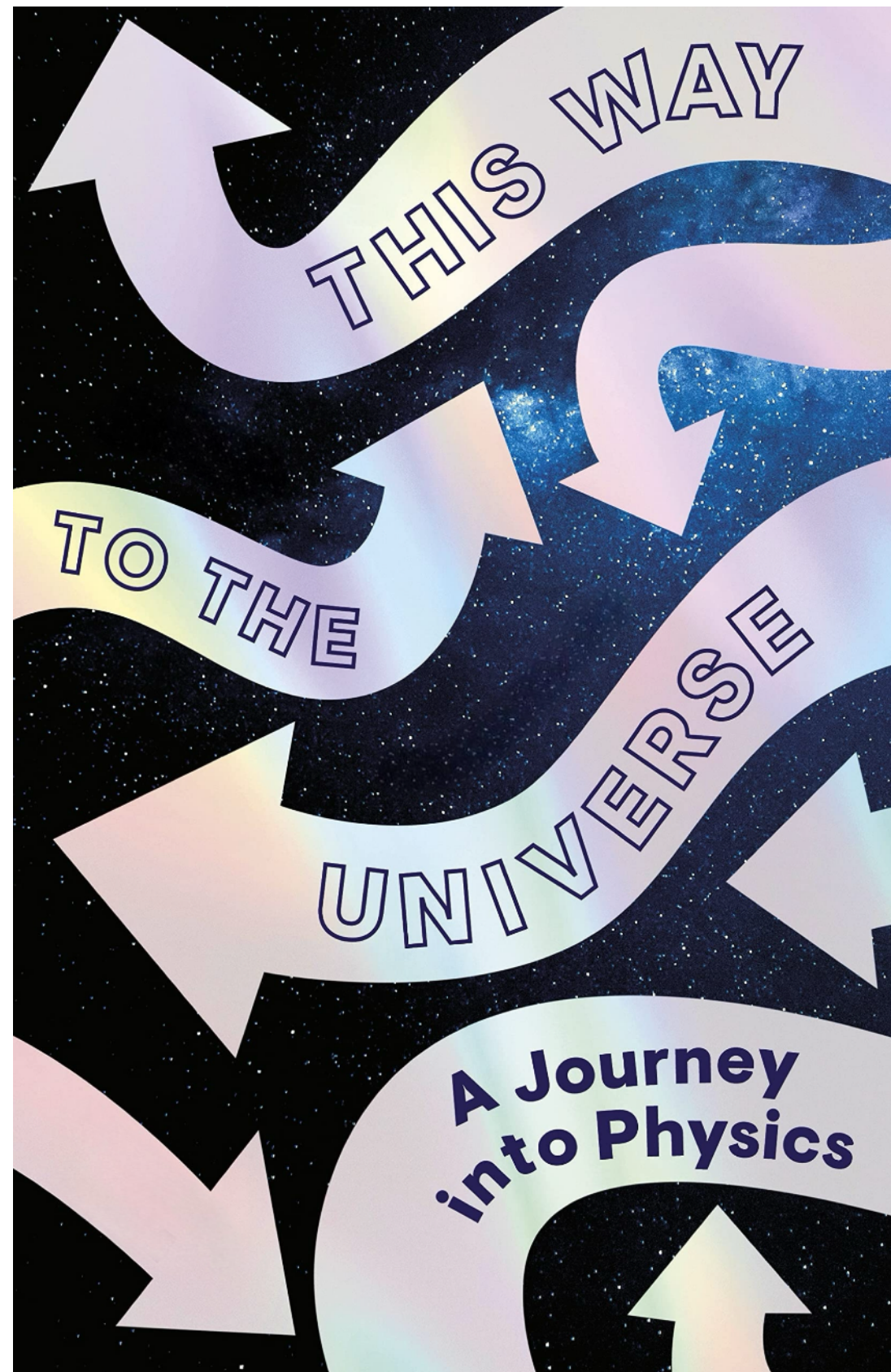
Unified Test Suit for Rare Event Modeling



Future Perspectives (What This Scales To)

- Detector design & field optimization
- Astrophysical population synthesis
- High-dimensional experimental parameter sweeps
- ...and more.

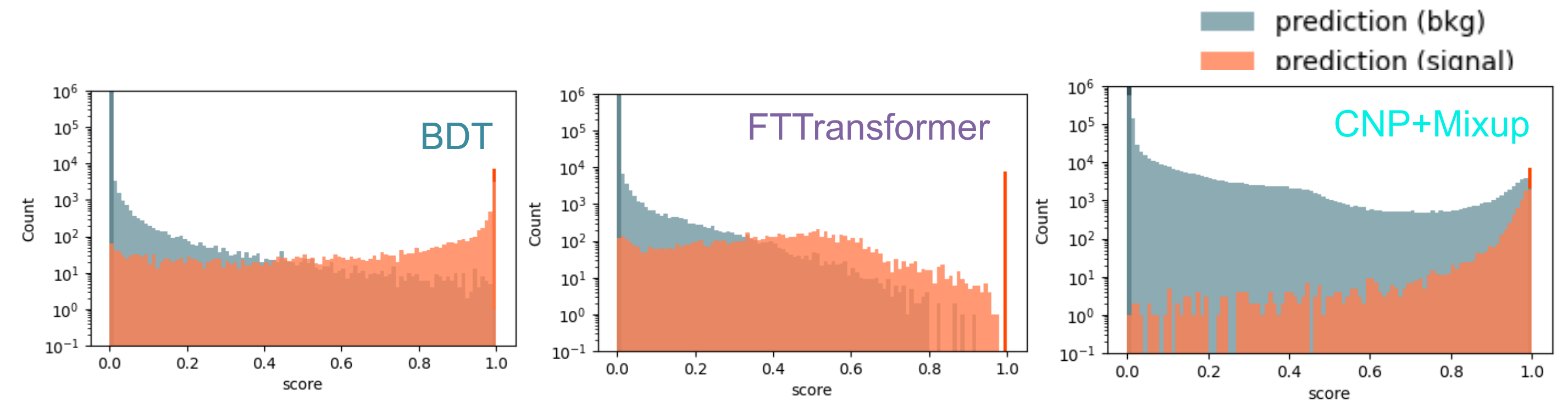
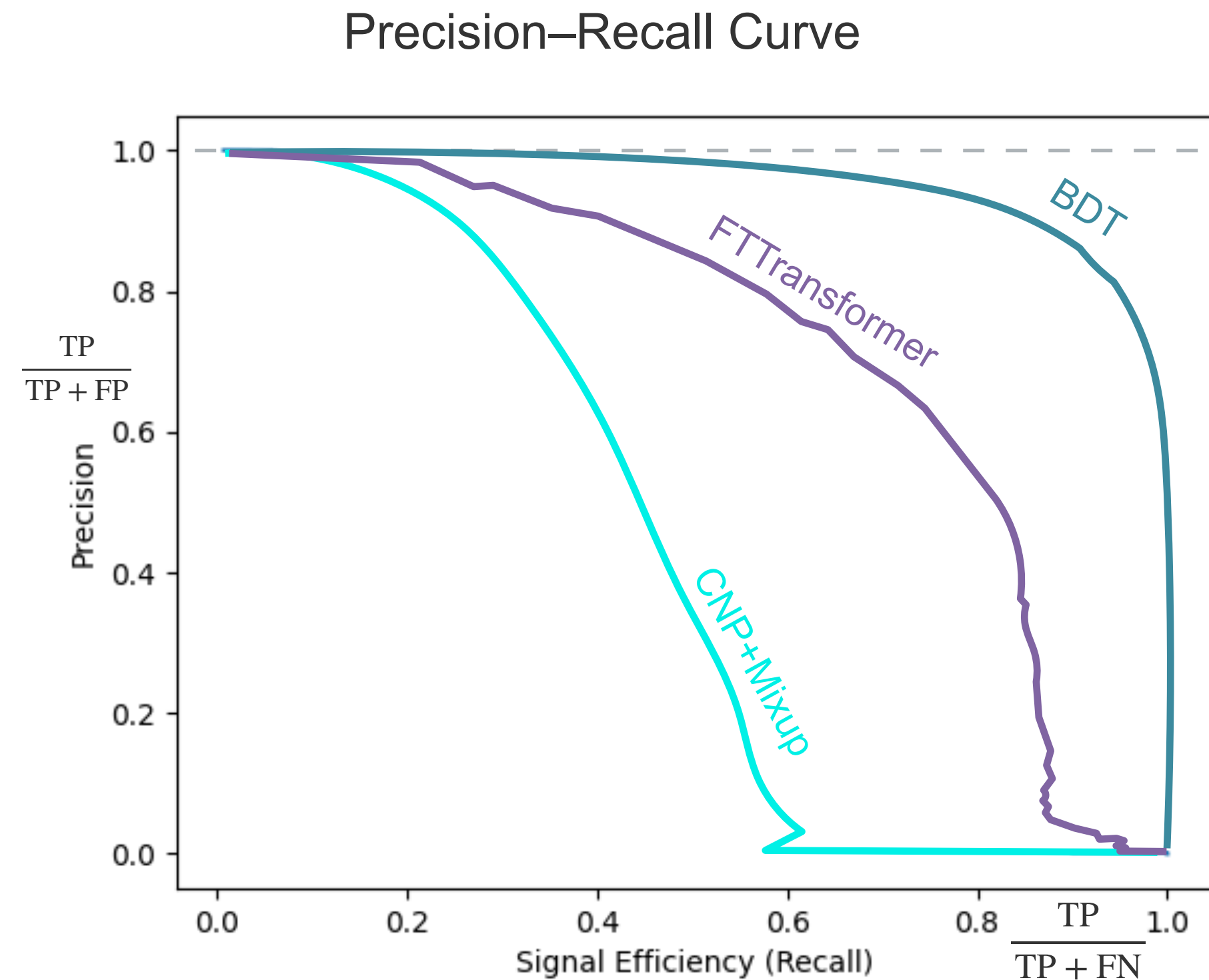
*Rare event modeling connects challenges across
very different areas of physics*



Thank you!

RESuM is a powerful tool for addressing them.

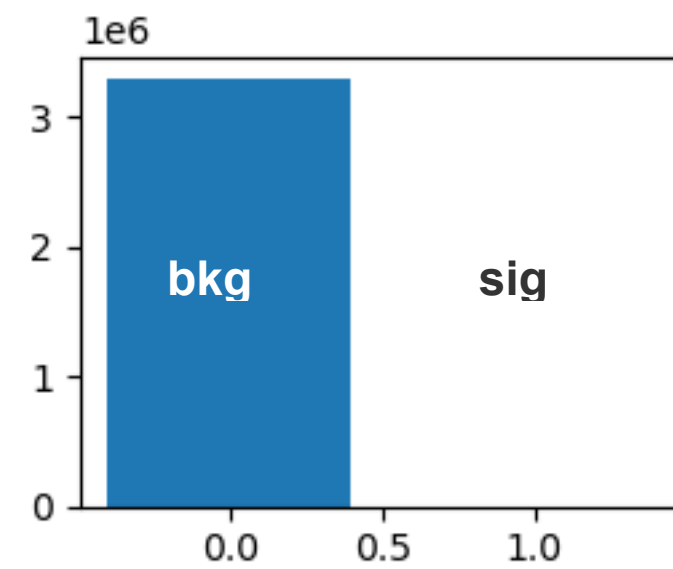
Traditional Machine Learning Fails



- **Inductive bias matched to rare events:**
 - BDTs partition feature space with hard, rule-based splits that naturally isolate sparse signal regions, while neural networks smooth them away.
- **Data-efficient under extreme imbalance:**
 - Tree-based learning concentrates capacity on informative regions and remains stable with limited signal statistics, unlike neural networks that bias toward background.
- **Sharper separation and more reliable ranking:**
 - BDTs produce crisp decision boundaries and well-calibrated scores optimized for discrimination, not averaged regression behavior.

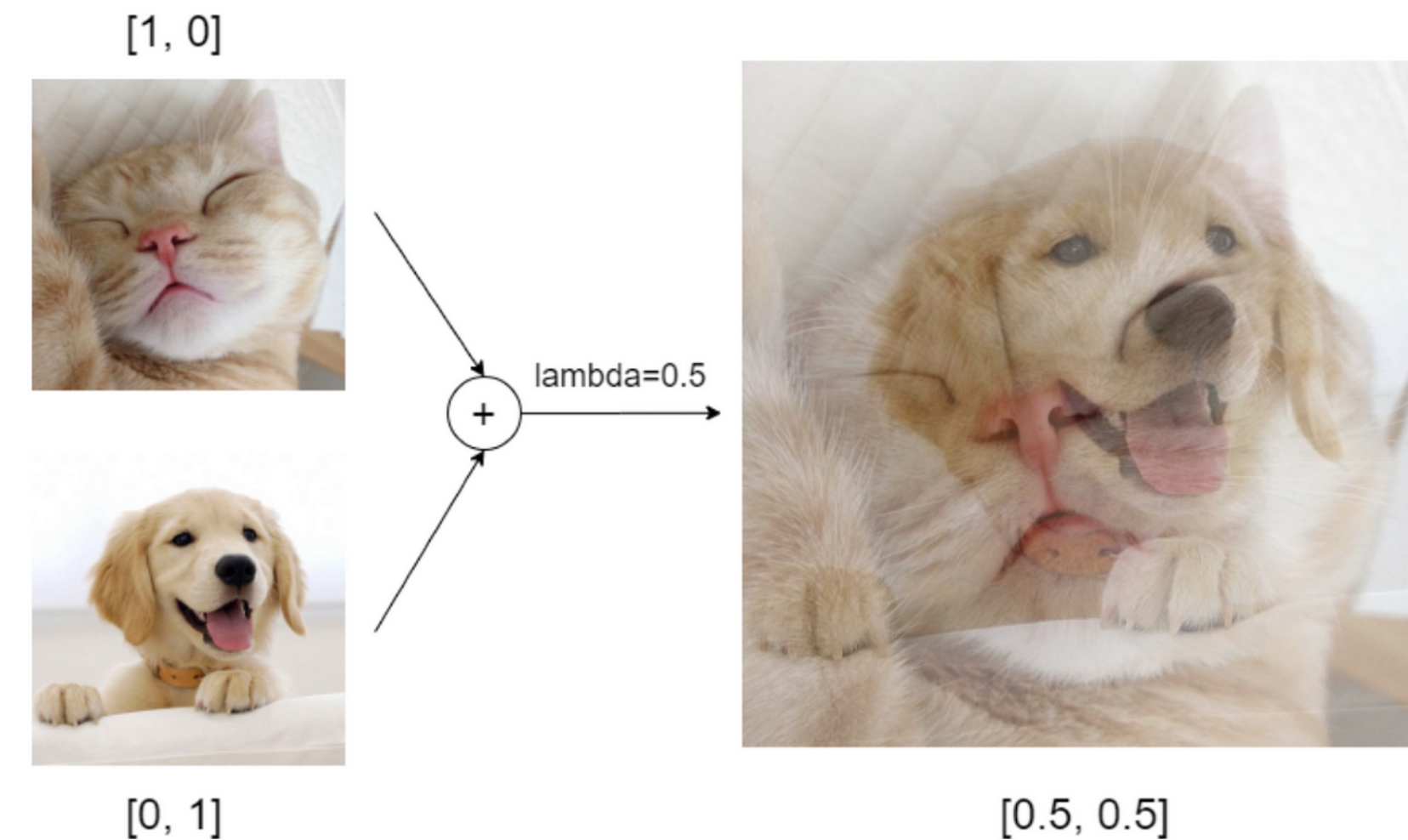
Data Augmentation - Mixup

< 1% signals



Problem: severely imbalanced training dataset

Solution: Linear Combinations with Mixup



$$x = \lambda \cdot x_i + (1 - \lambda) \cdot x_j$$

$$y = \lambda \cdot y_i + (1 - \lambda) \cdot y_j$$

- Create virtual training samples by linearly combining inputs and labels from different, randomly selected classes.
- Linear interpolation in input space should correspond to linear interpolation in output (label) space.
- Smoother decision boundaries and robustness
- Enhances model generalization on test samples that only slightly differ from training data.

Option 1: λ uniformly distr.

Option 2: λ distr. u-shaped

