

Kitchen Sink Anomaly Detection

ML4Jets 2026

[arXiv:2604.20965](https://arxiv.org/abs/2604.20965)

Lukas Lang

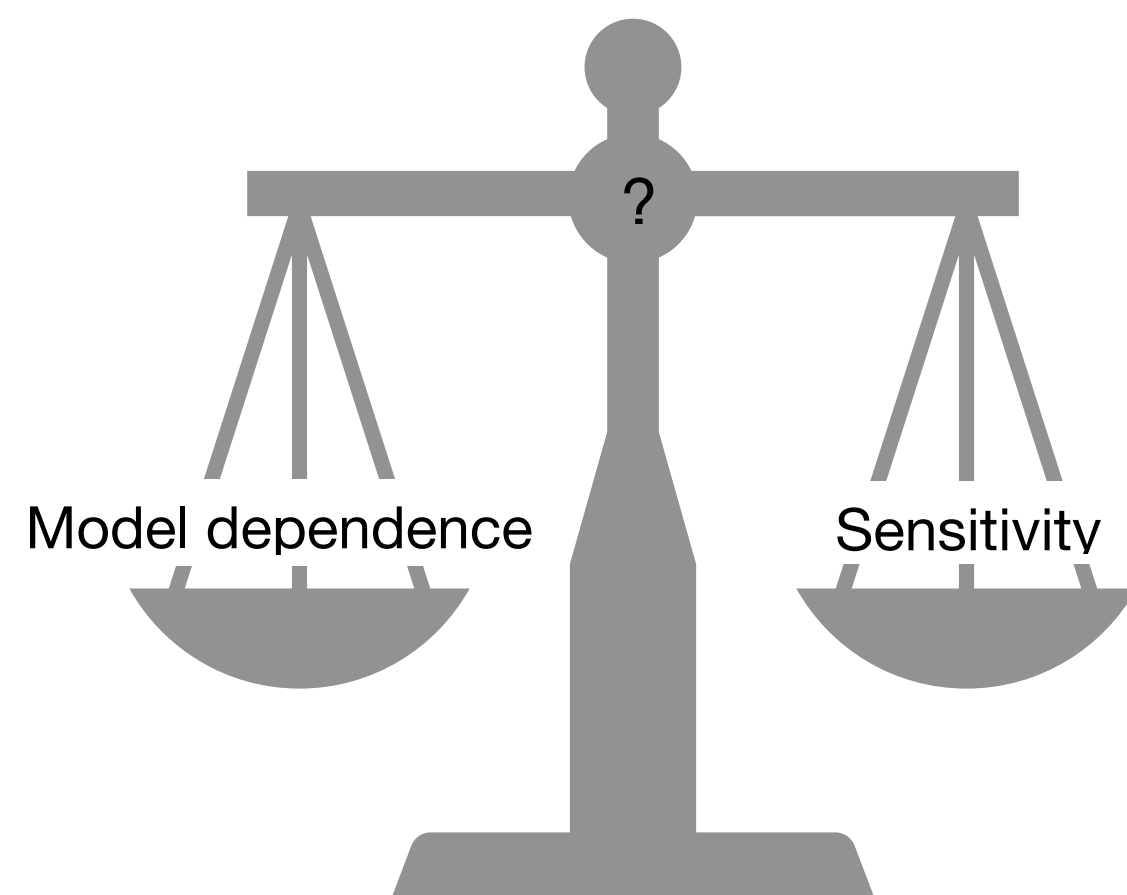
17 Sep 2026

In collaboration with R. Das, M. Hein, G. Kasieczka, M. Krämer, R. Mastandrea, L. Moureaux, A. Mück, D. Shih

Model Agnostic Anomaly Detection

Two Goals:

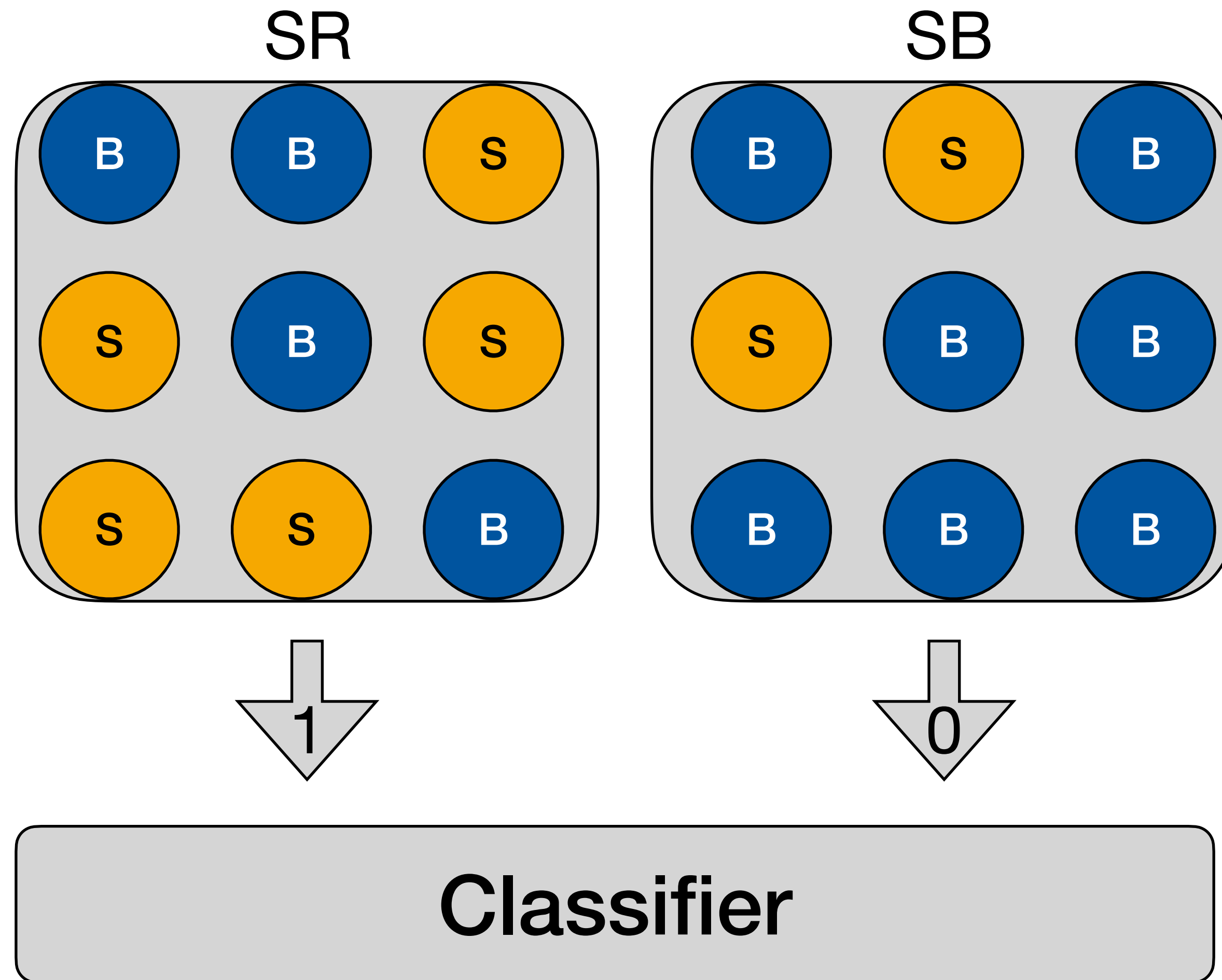
- **Model Agnostic:**
As little assumption on the signal as possible
- **High Sensitivity:**
Find signals with low cross-section



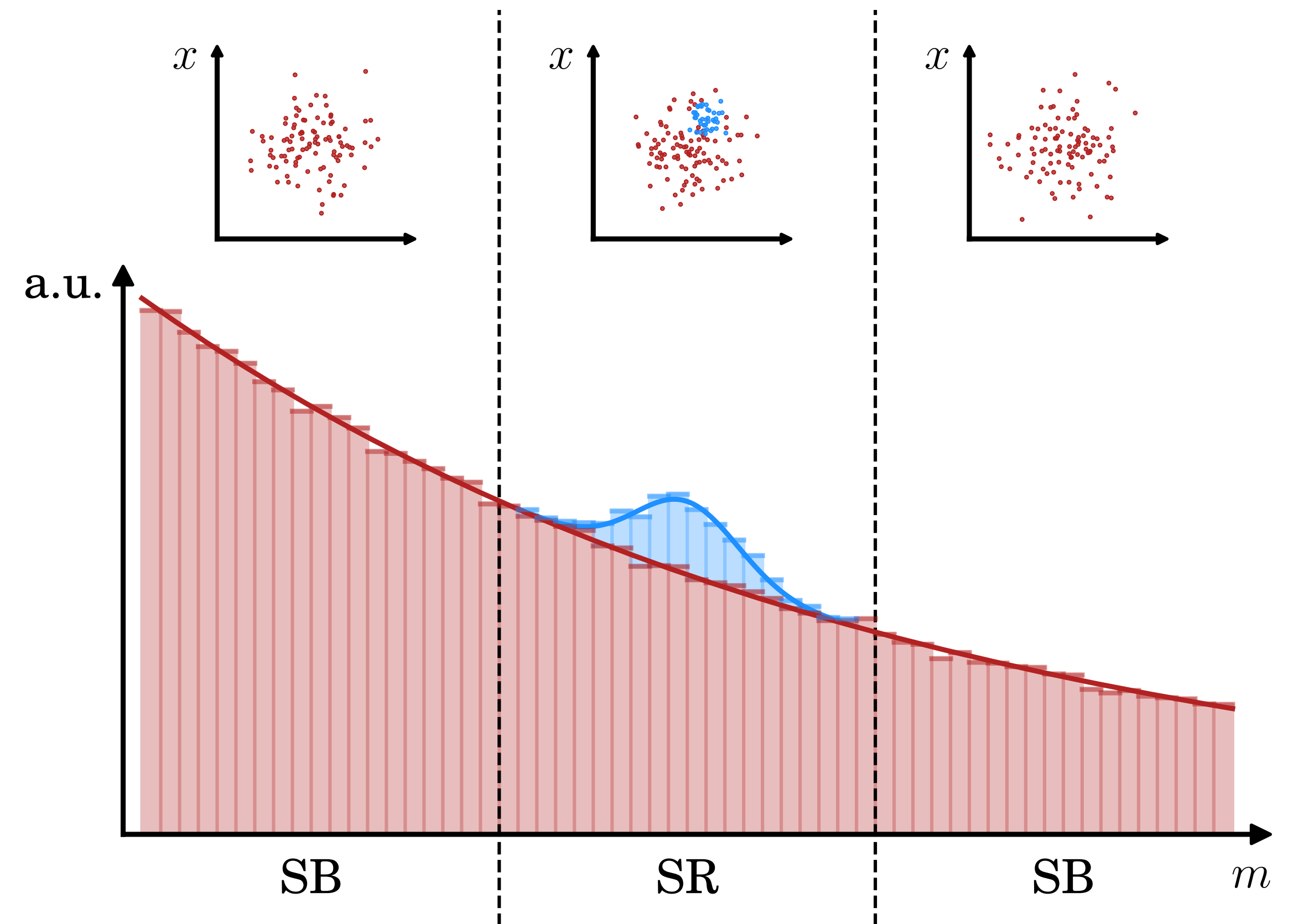
Where do we make assumptions?

- **Method**
 - **Weakly supervised anomaly detection** → contracting a background template (BT)
- **Features**
 - Choosing specific observables

Method: CWoLa Hunting



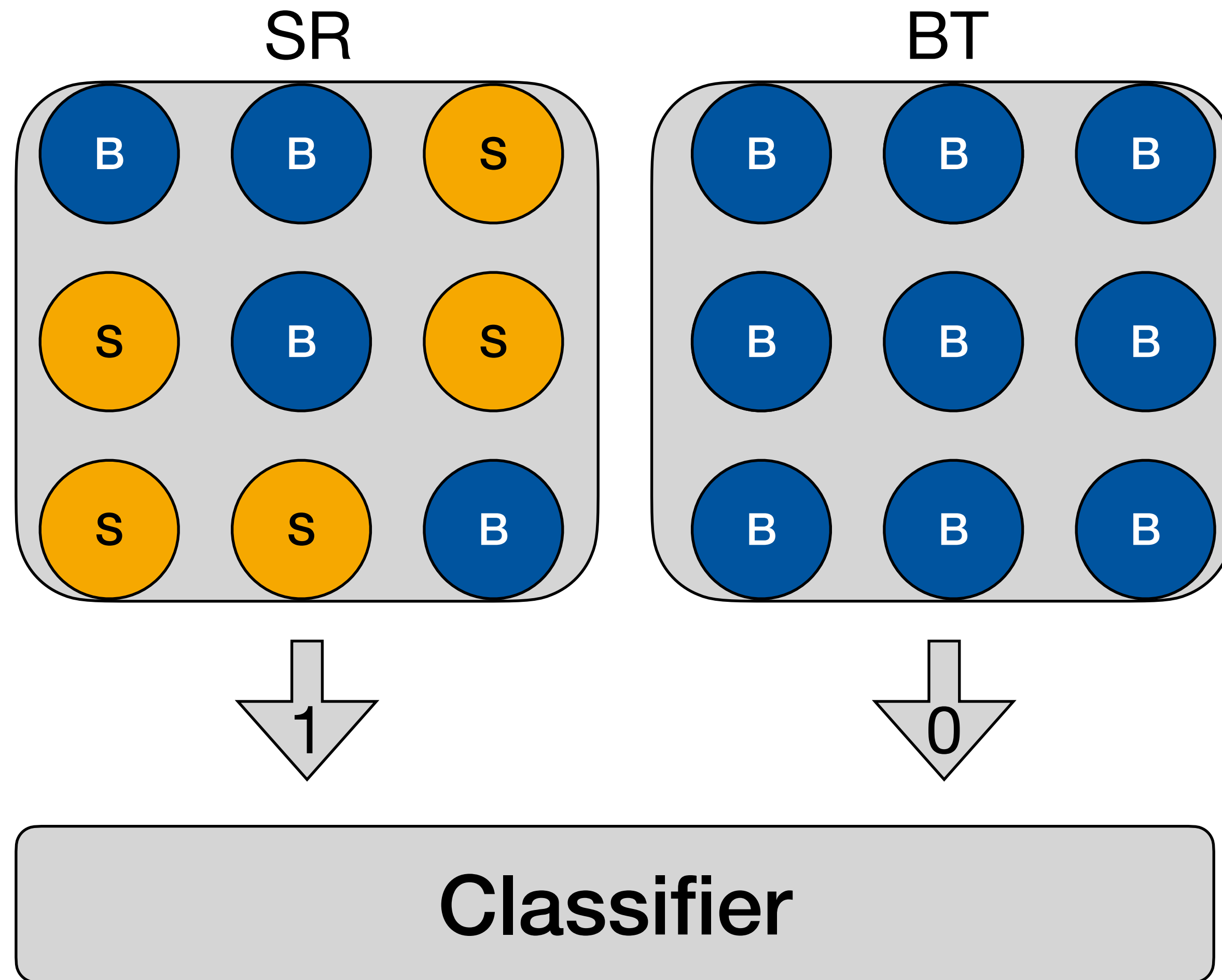
Here ensemble of 50 Boosted Decision Trees



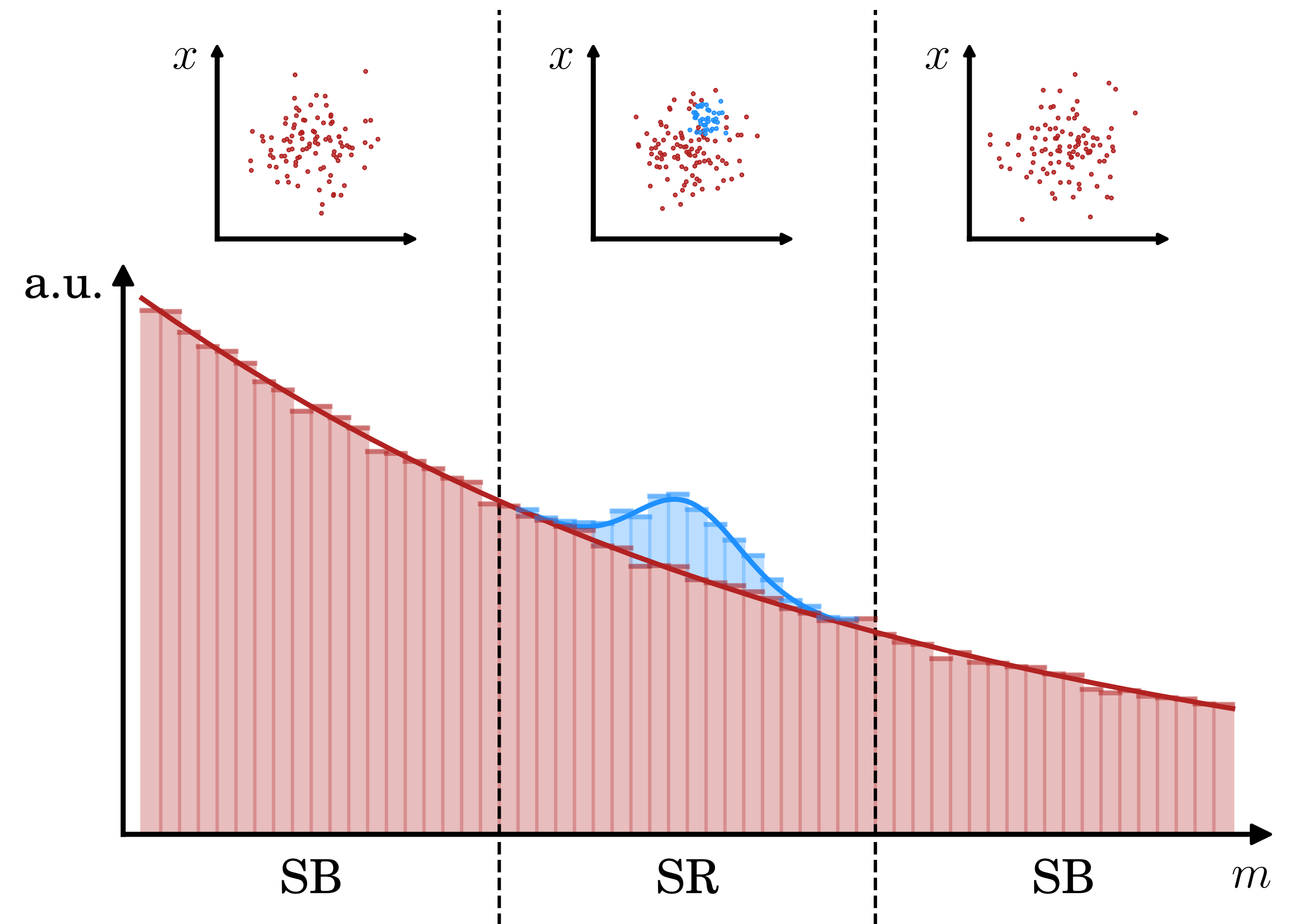
$$p_{\text{data}}(x|m \in SB) = p_{\text{bg}}(x|m \in SB) \quad p_{\text{data}}(x|m \in SR) \quad p_{\text{data}}(x|m \in SB) = p_{\text{bg}}(x|m \in SB)$$

[arXiv:1708.02949, arXiv:2109.00546, arXiv:1805.02664]

Method: Idealized Anomaly Detector (IAD)



Here ensemble of 50 Boosted Decision Trees



$$p_{\text{data}}(x|m \in SB) = p_{\text{bg}}(x|m \in SB)$$

$$p_{\text{data}}(x|m \in SR)$$

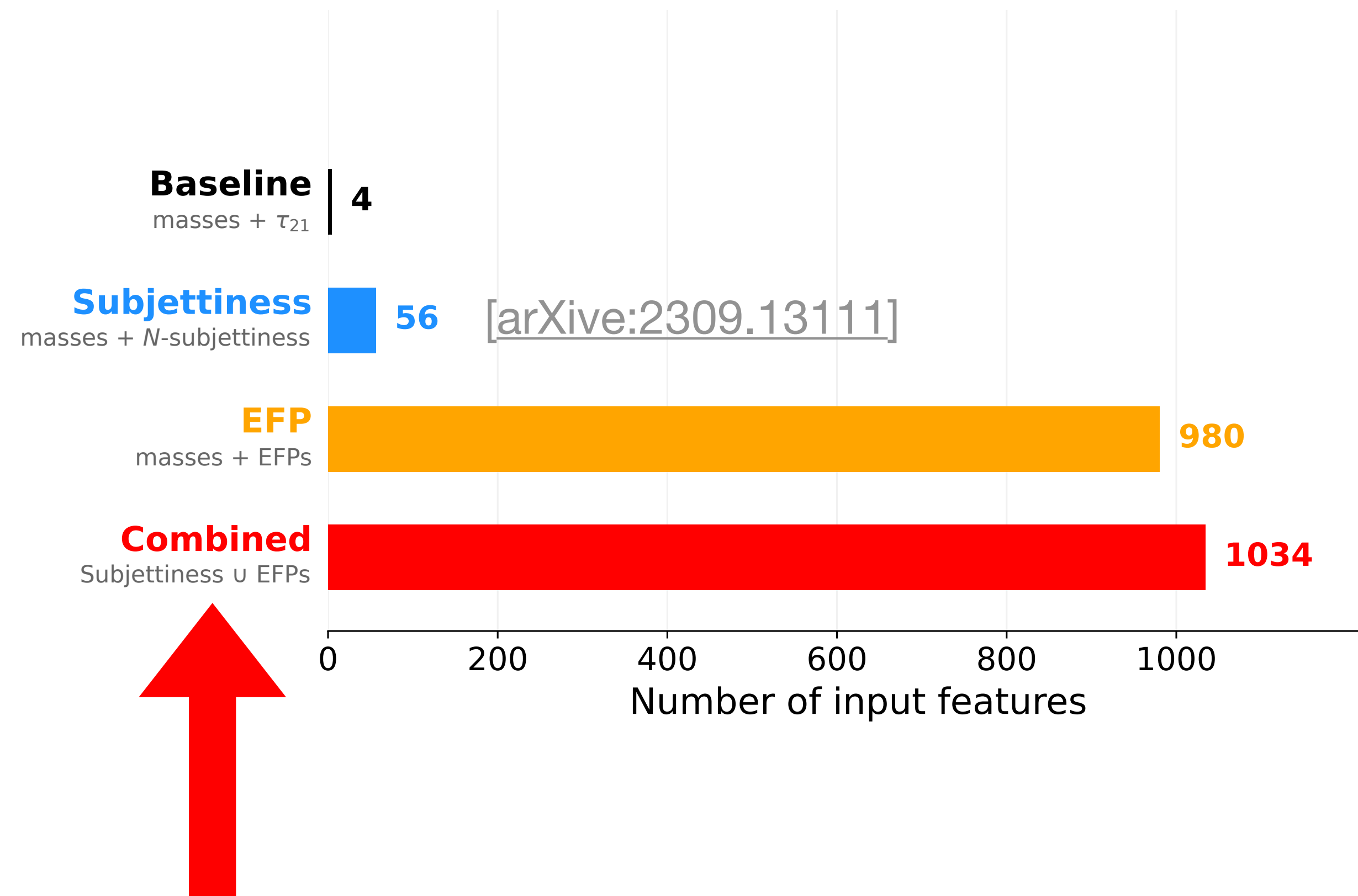
$$p_{\text{data}}(x|m \in SB) = p_{\text{bg}}(x|m \in SB)$$

[[arXiv:1708.02949](https://arxiv.org/abs/1708.02949), [arXiv:2109.00546](https://arxiv.org/abs/2109.00546), [arXiv:1805.02664](https://arxiv.org/abs/1805.02664)]

More Features to be More Model Agnostic

Different families of features

- Jet masses
- Substructure observables
 - N-Subjettiness
 - Energy Flow Polynomials (EFPs)



The "kitchen sink" approach:

➡ Use all the features you have

Data: Six Signals, One Background

Background:

- Taken from the [LHC Olympics R&D Dataset](#) [[arXiv:2101.08320](#)]
 $\Rightarrow 10^6$ QCD dijet events + $\sim 6 \cdot 10^5$ QCD dijet events in SR

Signals:

- [Six different resonance signals](#) inspired by [[arXiv:2412.03747](#)]
 $\Rightarrow$ Hadronic $A \rightarrow BC$ decays with $m_A = 3.5$ TeV

Name	Prongs	Masses
LHCO 2-prong	$2 + 2$	$m_Y = 100$ GeV, $m_X = 500$ GeV
LHCO 3-prong	$3 + 3$	$m_Y = 100$ GeV, $m_X = 500$ GeV
$X \rightarrow YY' \rightarrow 4q$	$2 + 2$	$m_Y = 100$ GeV, $m_{Y'} = 500$ GeV
$W_{KK} \rightarrow WR \rightarrow 3W$	$2 + 4$	$m_R = 500$ GeV
$Z' \rightarrow T'T' \rightarrow tZtZ$	$5 + 5$	$m_{T'} = 400$ GeV
$G_{KK} \rightarrow HH \rightarrow 4t$	$6 + 6$	$m_H = 400$ GeV

Test your Pipeline with the Additional Signals

Published April 23, 2026 | Version v1

Dataset Open

Additional Signal Models for the LHCO2020 R&D

Das, Ranit¹ ; Hein, Marie² ; Kasieczka, Gregor³ ; Krämer, Michael² ; Lang, Lukas² ; Mastandrea, Radha ;
Moureaux, Louis³ ; Mück, Alexander² ; Shih, David¹ 

Show affiliations

This Zenodo contains the data used for the paper "[Kitchen Sink Anomaly Detection](#)". The data used in the paper is based on the LHC Olympics 2020 R&D dataset and the signal models inspired by the CMS study [Rept.Prog.Phys. 88, 067802 \(2024\)](#).

Model generation

For the generated models we provide the Event directories of the MadGraph5_aMC@NLO generation, as well as the clustered events and features calculated from the Delphes output. Furthermore, we provide the MadGraph5_aMC@NLO driver files (.mg5) for producing the signal event samples used in the paper, which can be found in the folder [model_cards.tar.gz](#). The primary generation chain used here was:

- MadGraph5_aMC@NLO 3.6.2
- Pythia 8.313 for parton showering/hadronization
- Delphes 3.5.0 for fast detector simulation

We use at leading order MadGraph5_aMC@NLO 3.6.2 (LO) to simulate the hard process including the decay of all intermediate resonances so that the final state consists of quarks only. The width of the BSM resonances is set to a small value such that they are effectively treated in the narrow width approximation.

The resulting parton-level events are passed to Pythia 8.313 for parton showering and hadronization, followed by Delphes 3.5.0 for detector simulation, using the same detector card as in the LHCO R&D setup. To ensure compatibility with the LHCO configuration, we use the default Pythia tune but switch off multi-parton interactions. For further details we recommend to have a look at the provided .mg5 driver files in the [model_cards.tar.gz](#) folder, which also contain the cards used for the Pythia and Delphes steps.

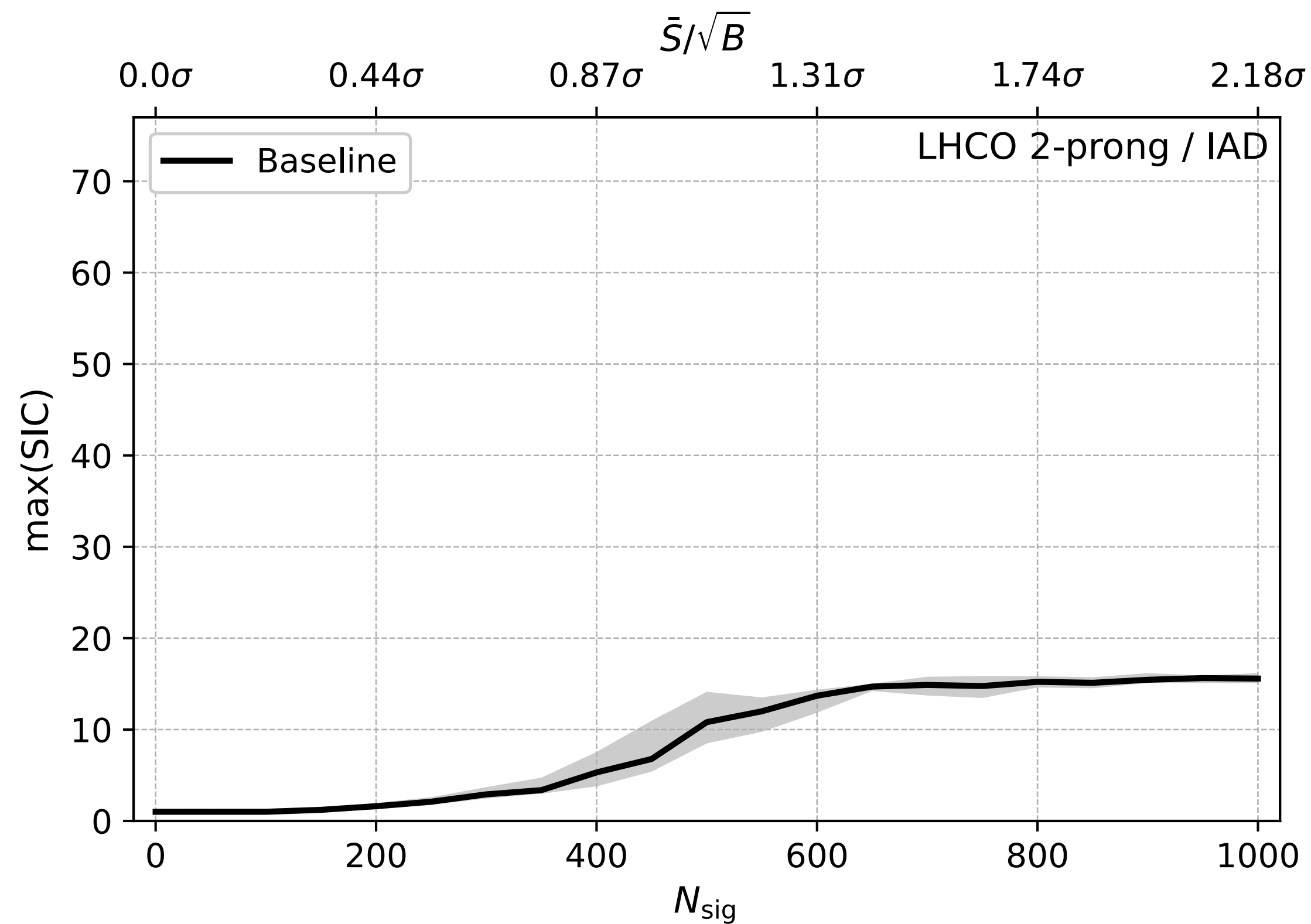
The models are inspired by the CMS study [Rept.Prog.Phys. 88, 067802 \(2024\)](#) and many of the cards come from the [genproductions](#) collection. UFO model tarballs are available from the [CMS project generators site](#).



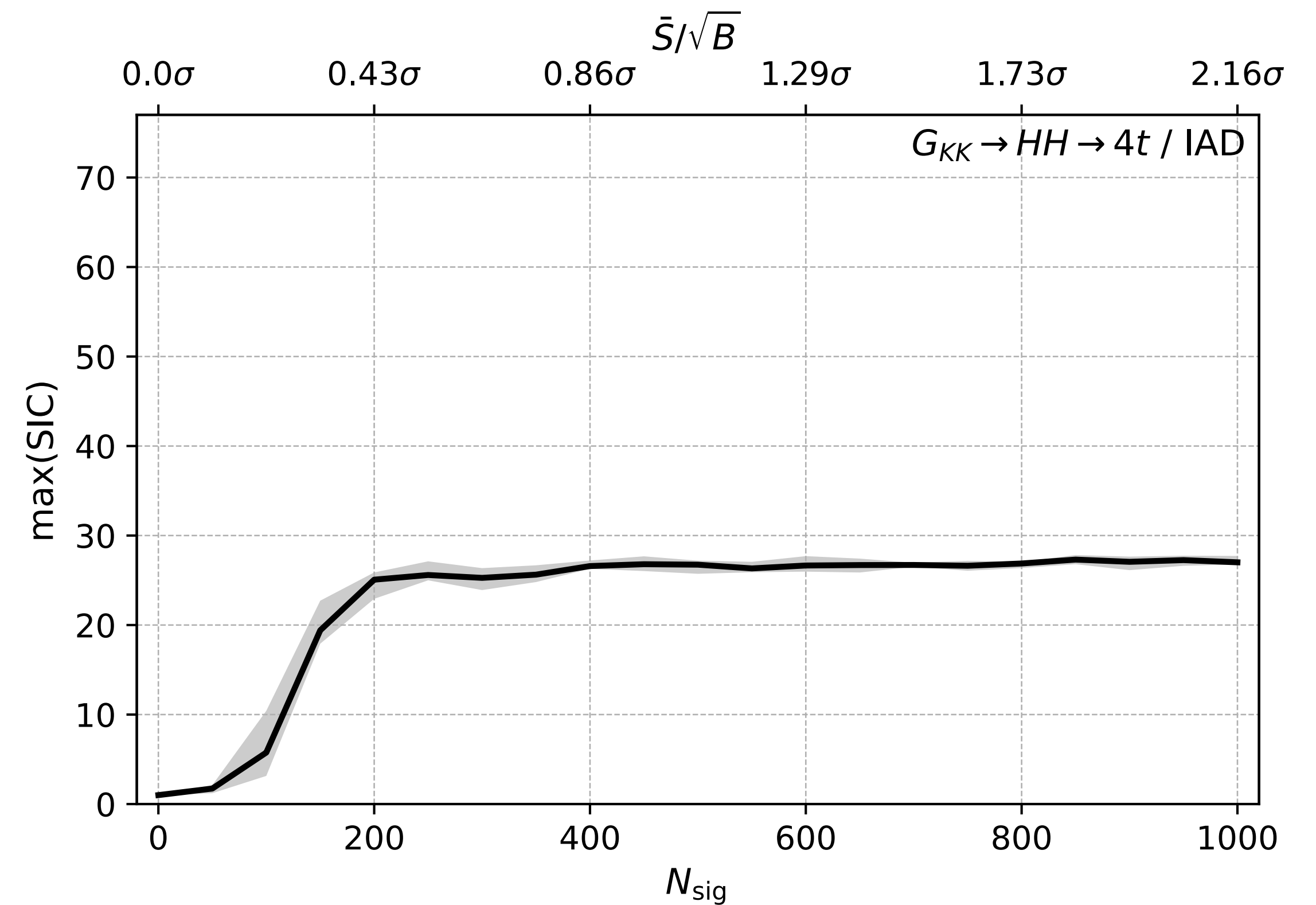
GET THE DATA

Baseline: Some Signals are Easier

LHCO 2-prong

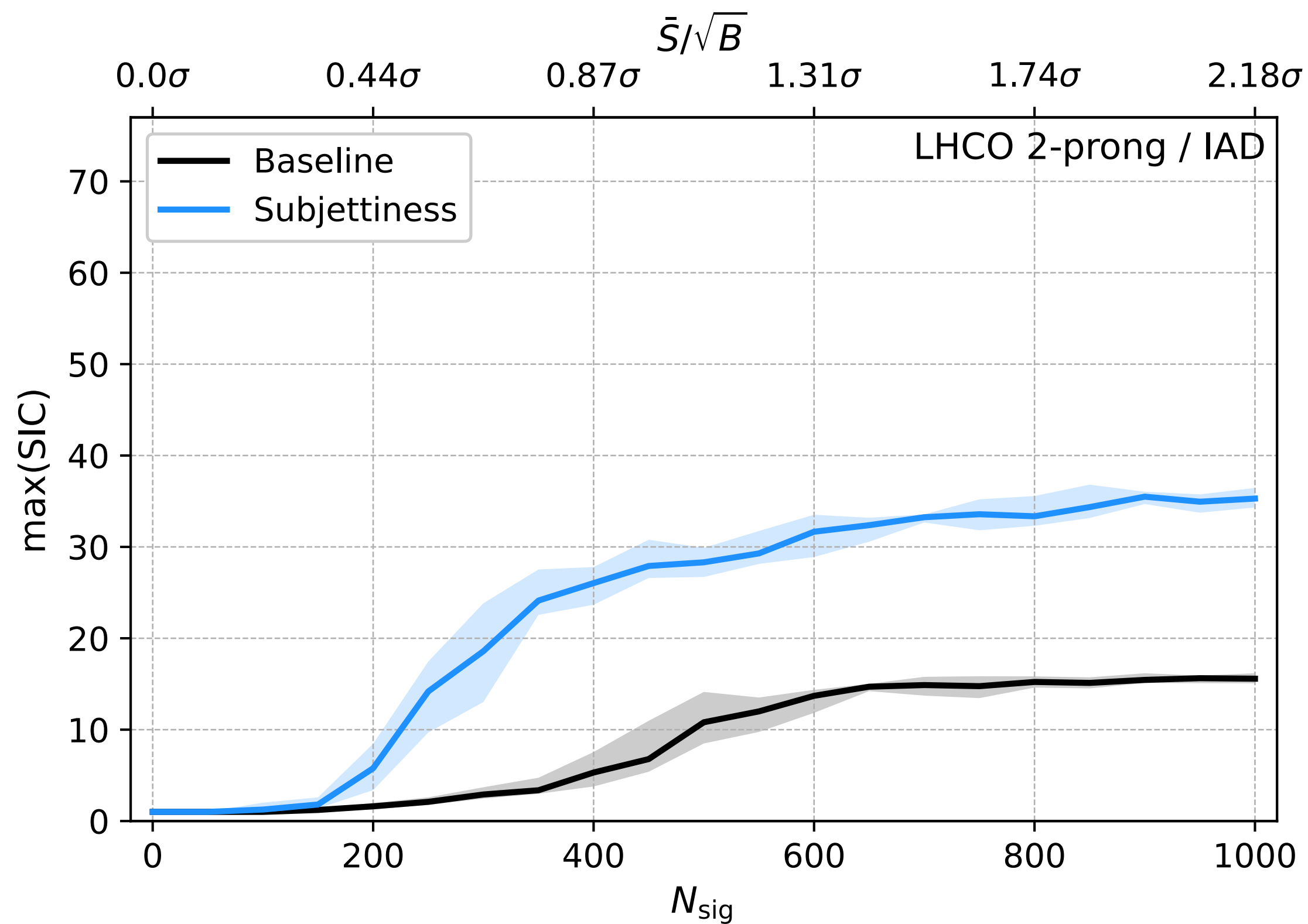


$G_{KK} \rightarrow HH \rightarrow 4t$

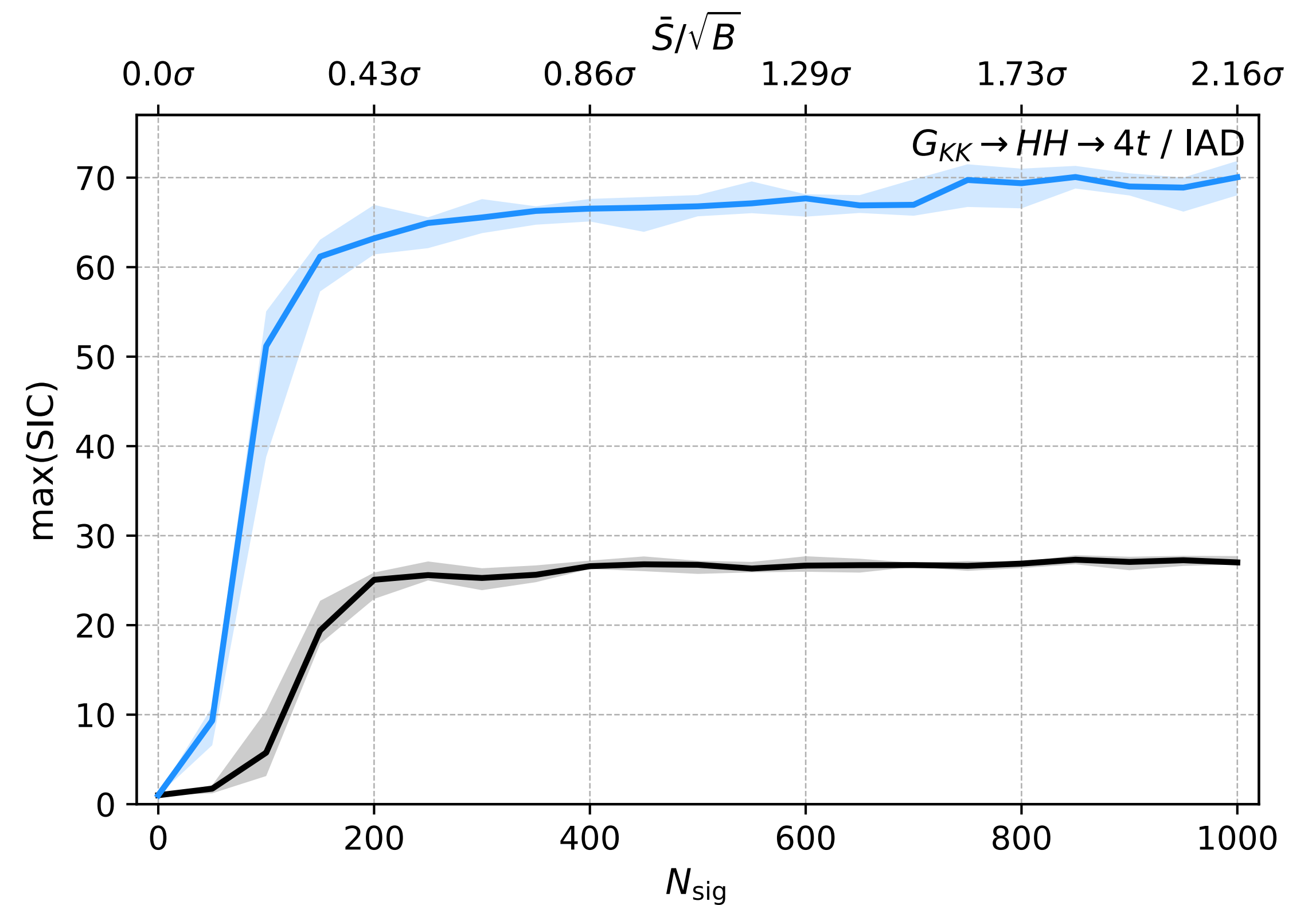


Subjettiness: A Versatile Feature Set

LHCO 2-prong

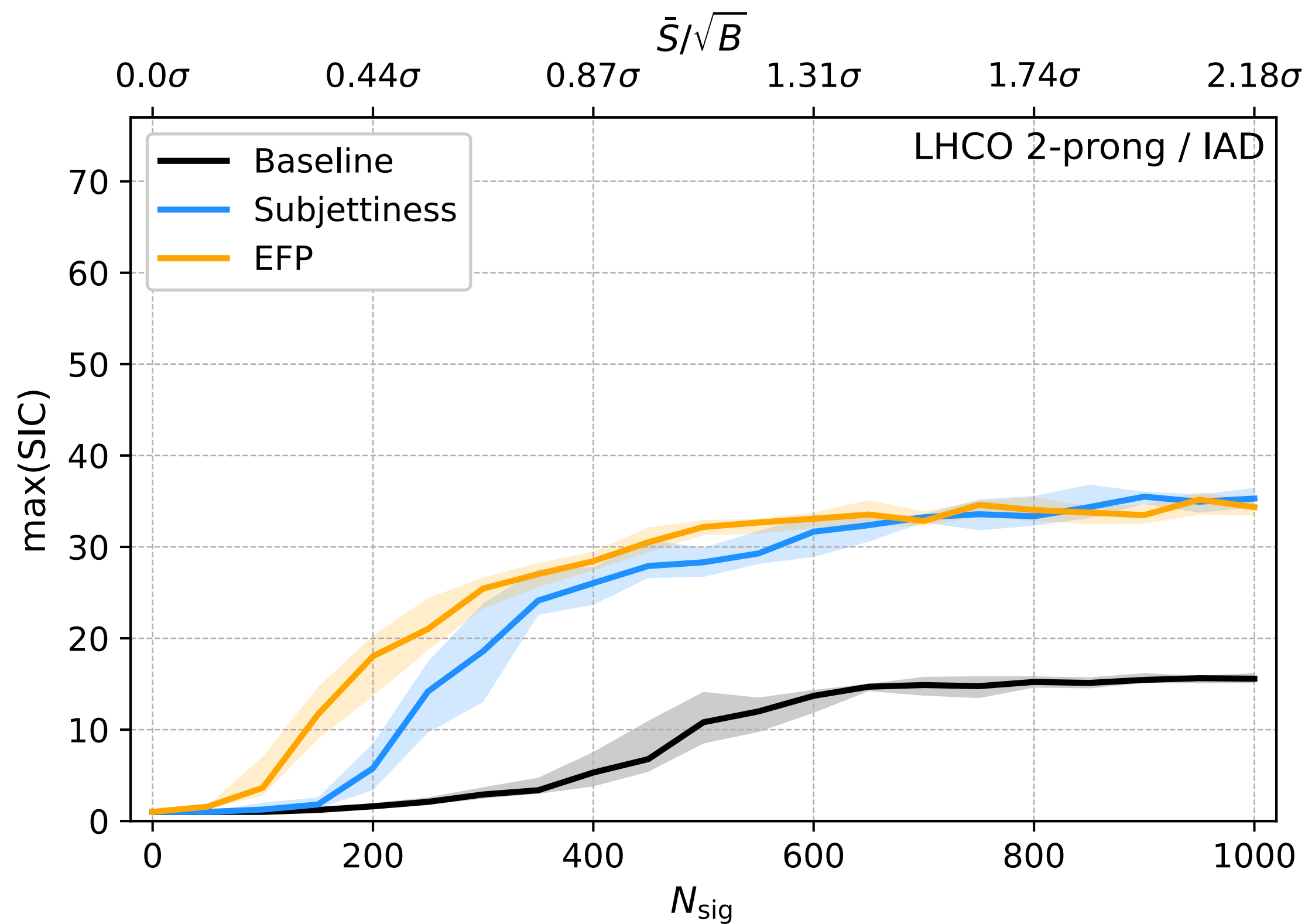


$G_{KK} \rightarrow HH \rightarrow 4t$

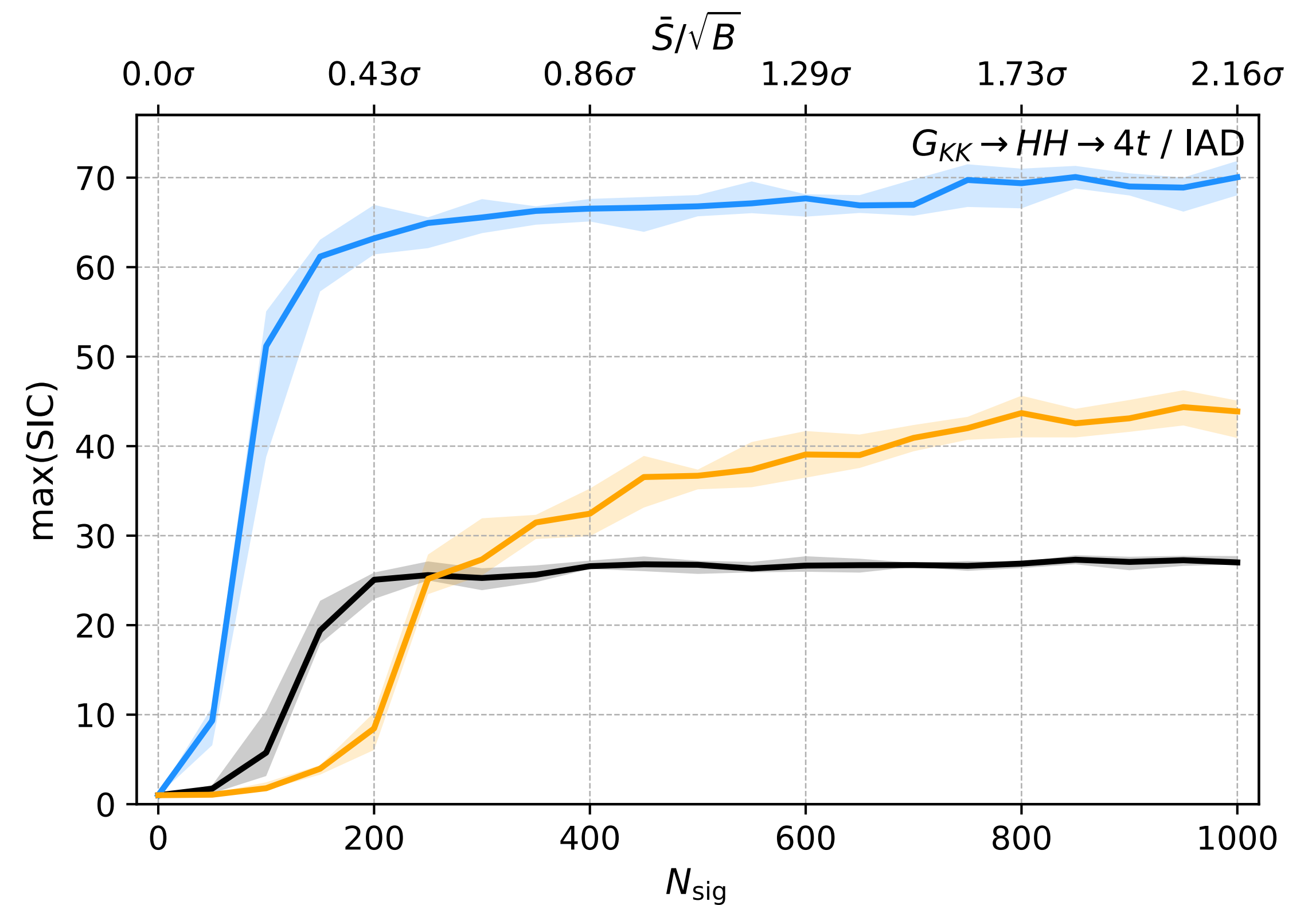


EFP: Sometimes Great, Sometimes Not

LHCO 2-prong

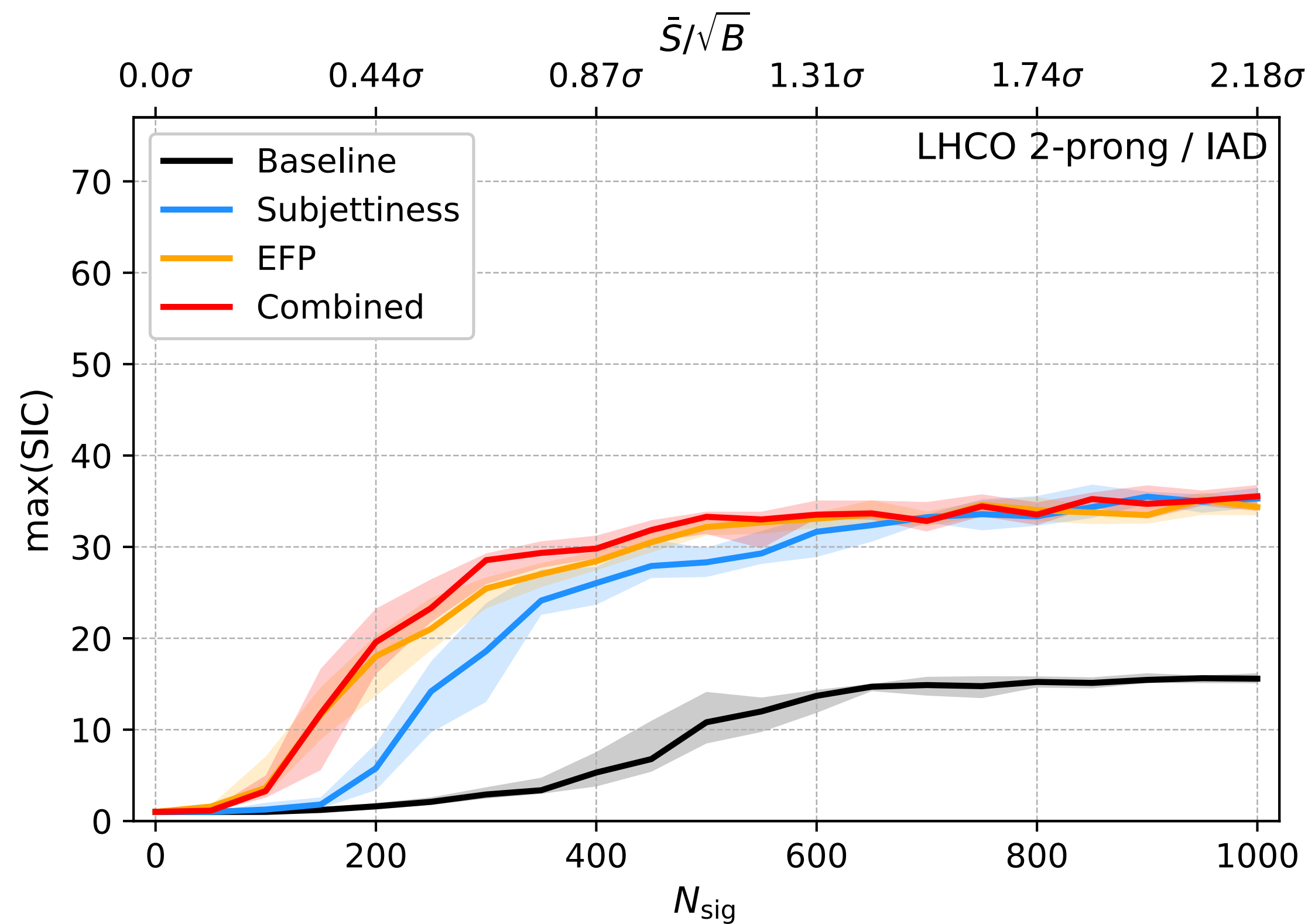


$G_{KK} \rightarrow HH \rightarrow 4t$

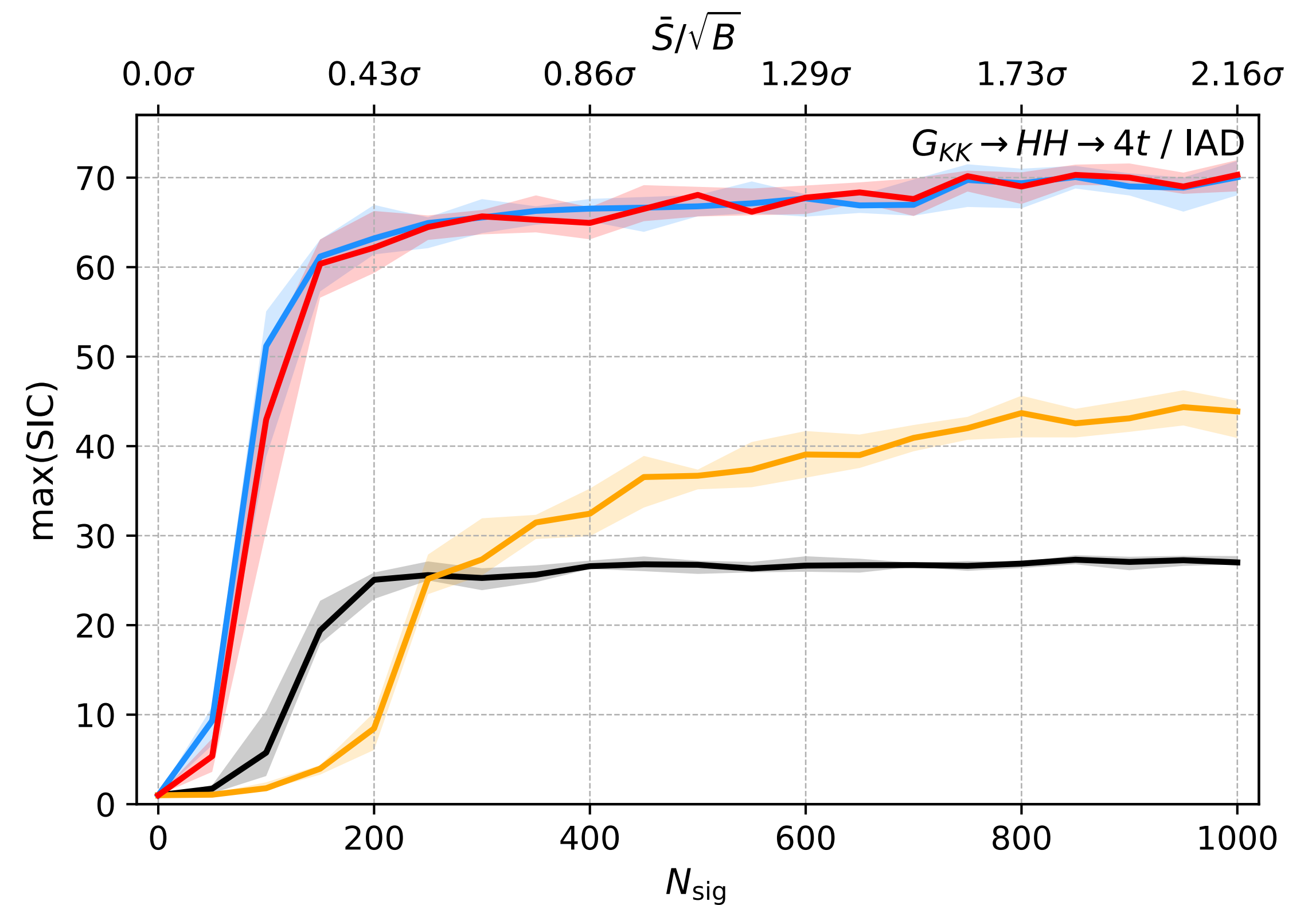


Combined: Most Performant and Model Agnostic

LHCO 2-prong

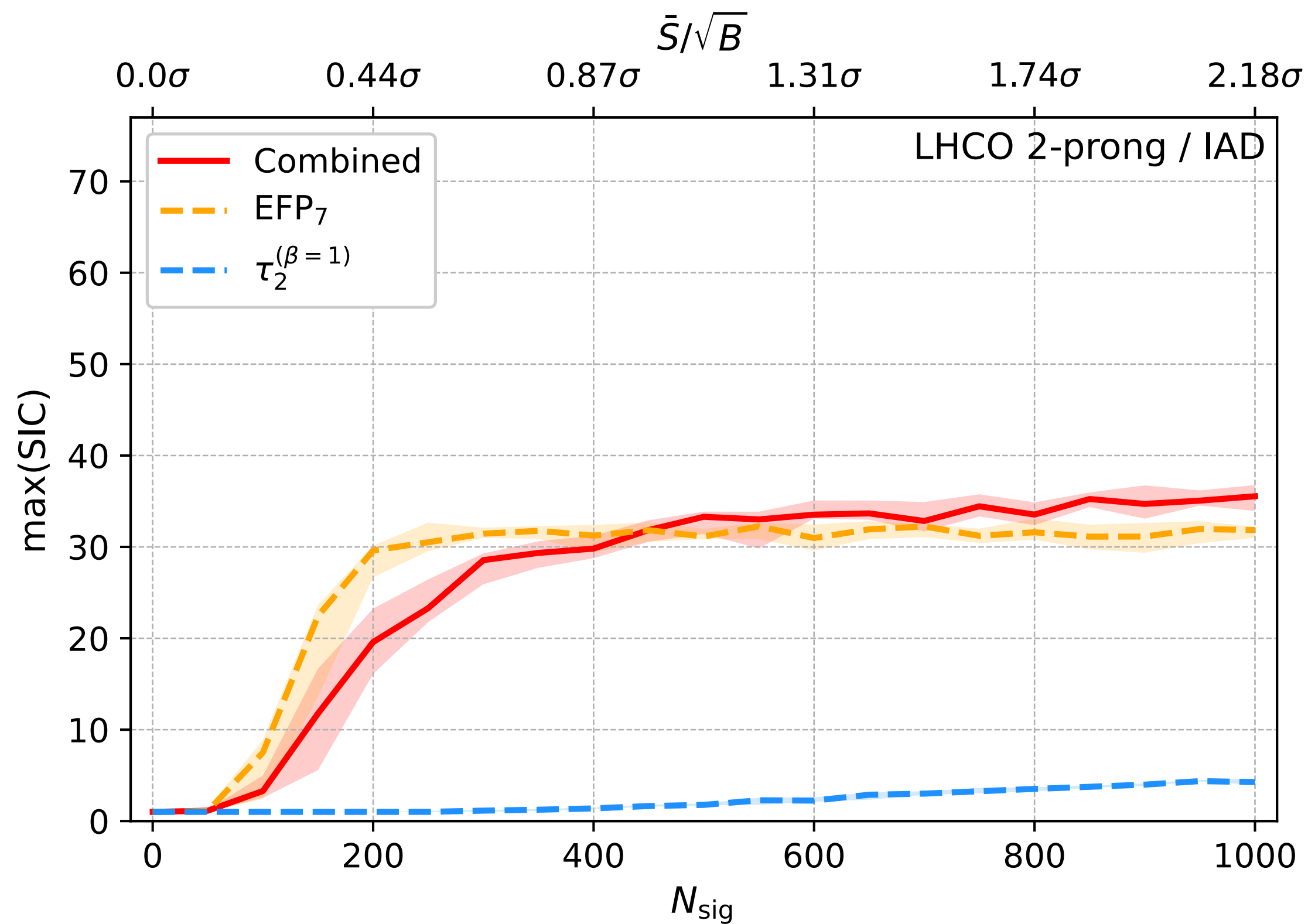


$G_{KK} \rightarrow HH \rightarrow 4t$

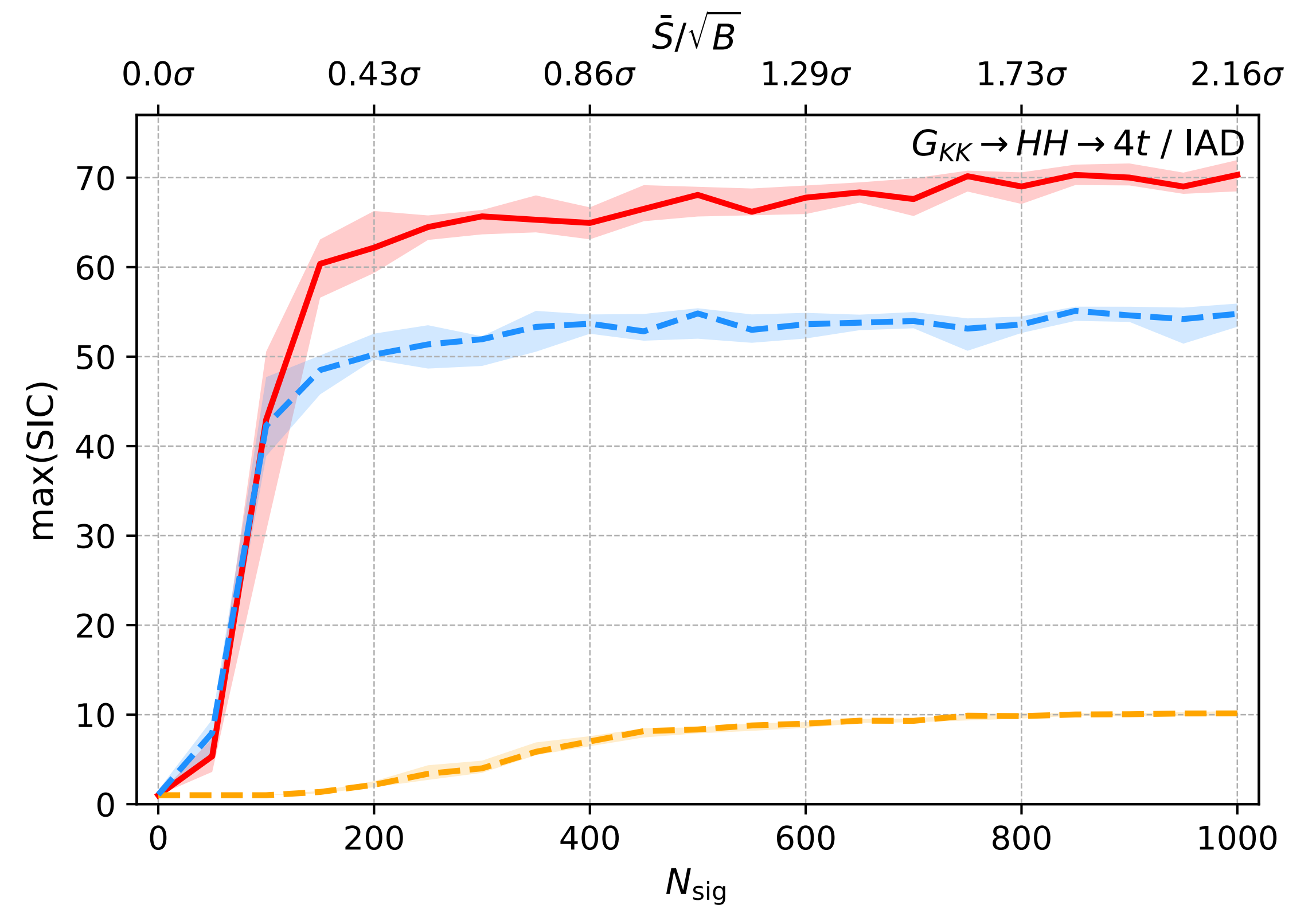


Individual Substructure Feature Drives the Performance

LHCO 2-prong

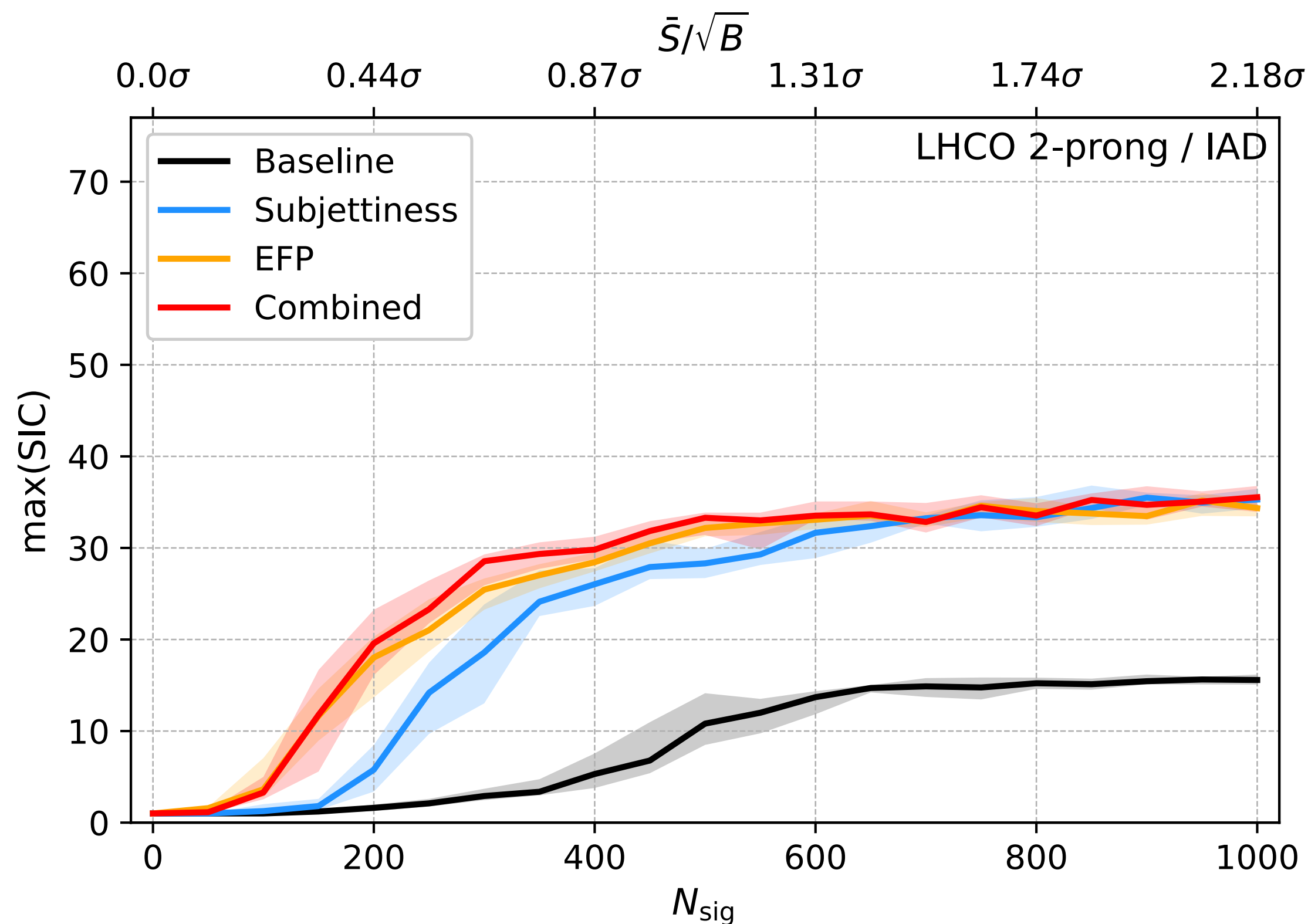


$G_{KK} \rightarrow HH \rightarrow 4t$



Discovery Potential: What do We Need for 5σ ?

LHCO 2-prong



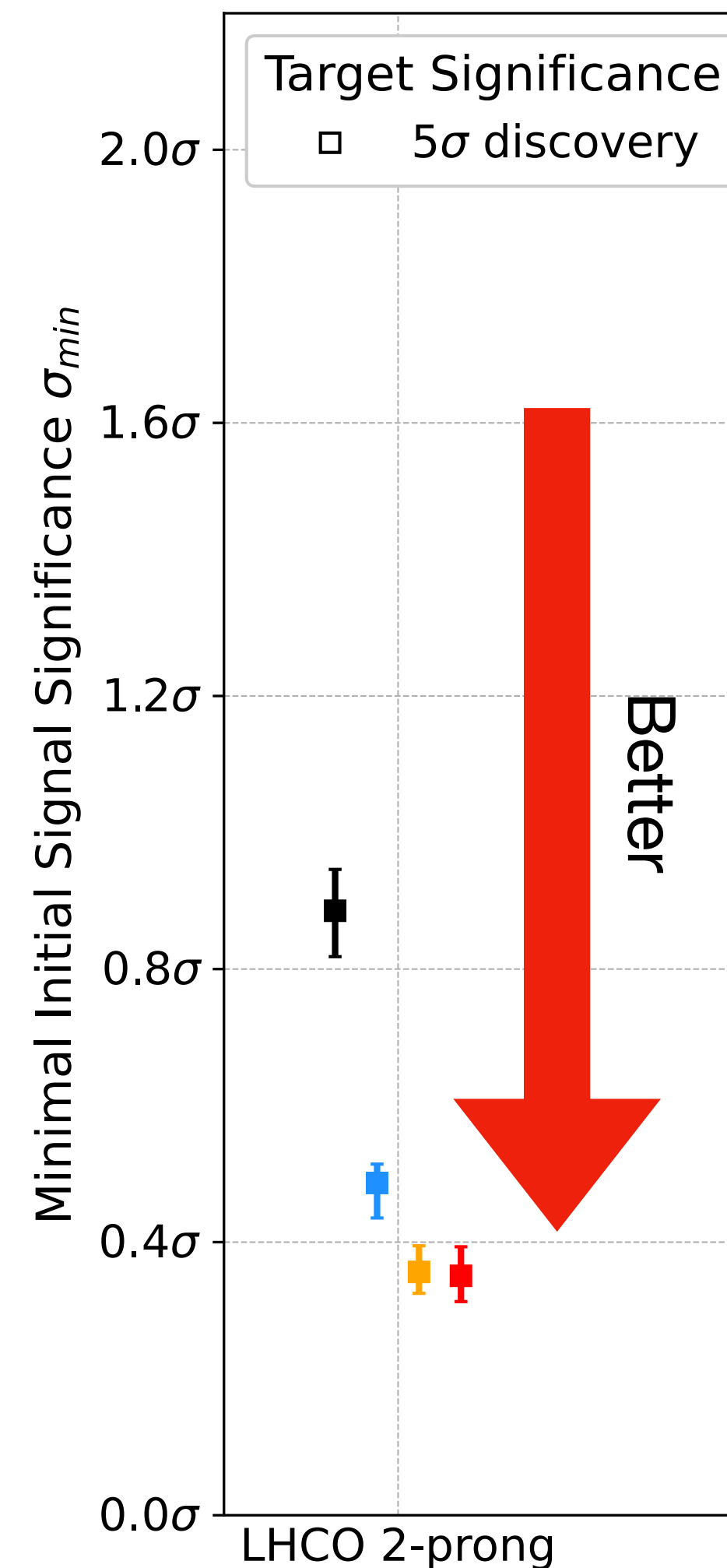
Initial Significance:

$$\sigma = S/\sqrt{B}$$

Significance with Classifier,
Asimov estimate: $\mathcal{S}$

Solve for $\sigma_{\min}$:

$$\max(\mathcal{S})(\sigma_{\min}) = 5\sigma$$



The Combined Set is the Best!

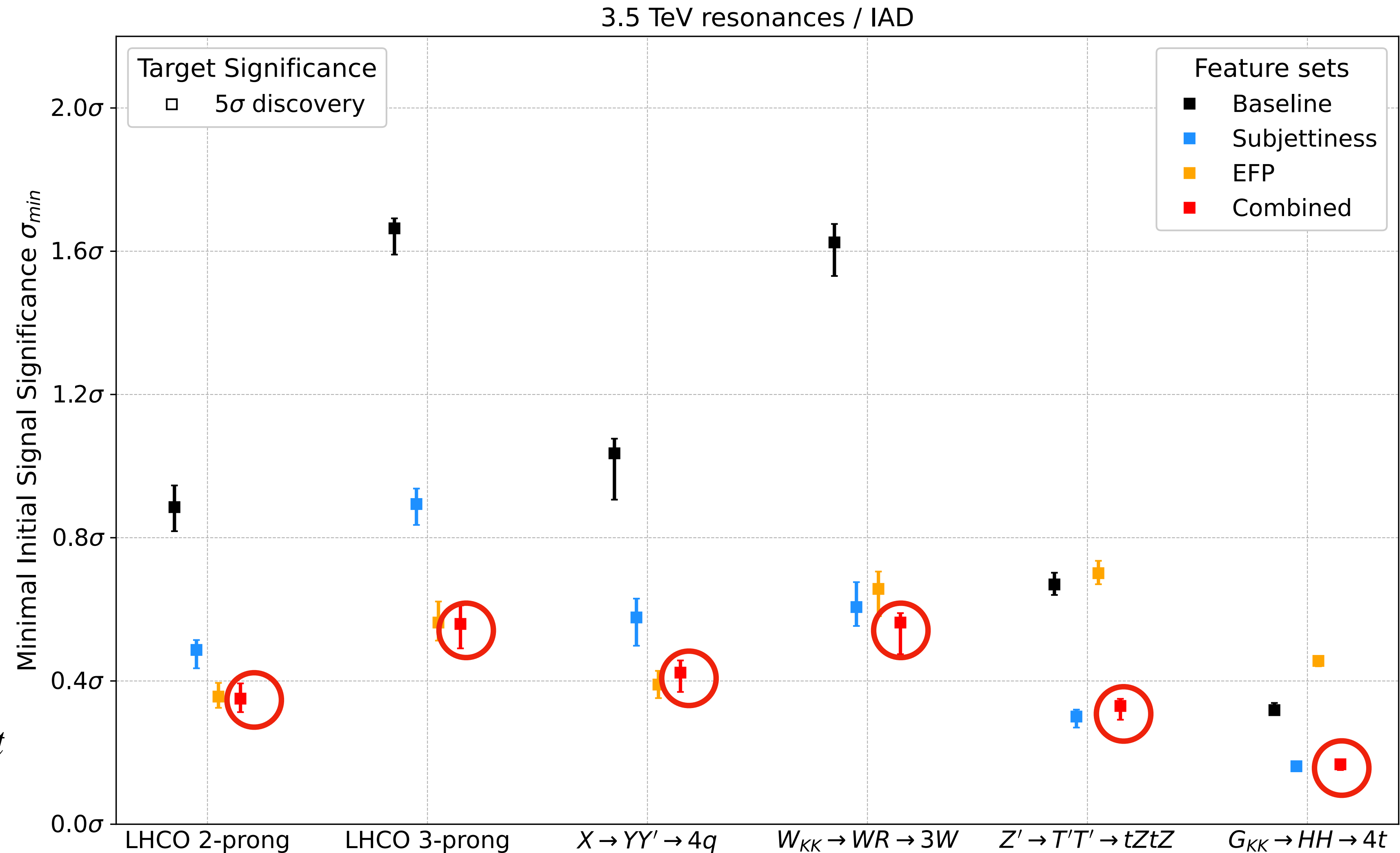
Initial
Significance:

$$\sigma = S/\sqrt{B}$$

Significance with
Classifier $\mathcal{S}$

Solve for σ_{min} :

$$\max(\mathcal{S})(\sigma_{min}) = \sigma_t$$



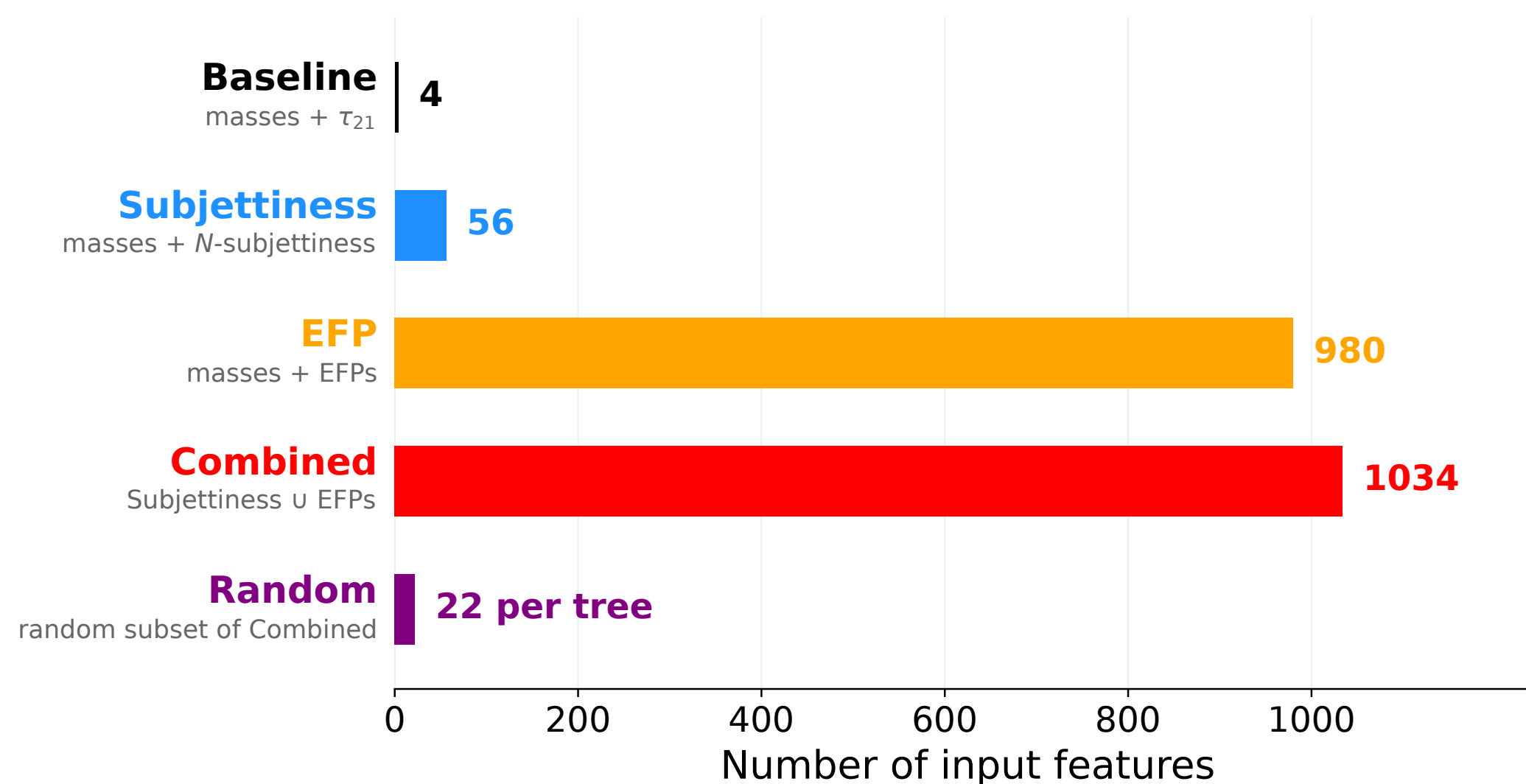
Reducing Training Time with the Random Set

The Problem:

- Training time of a BDT scales linearly with the number of input features

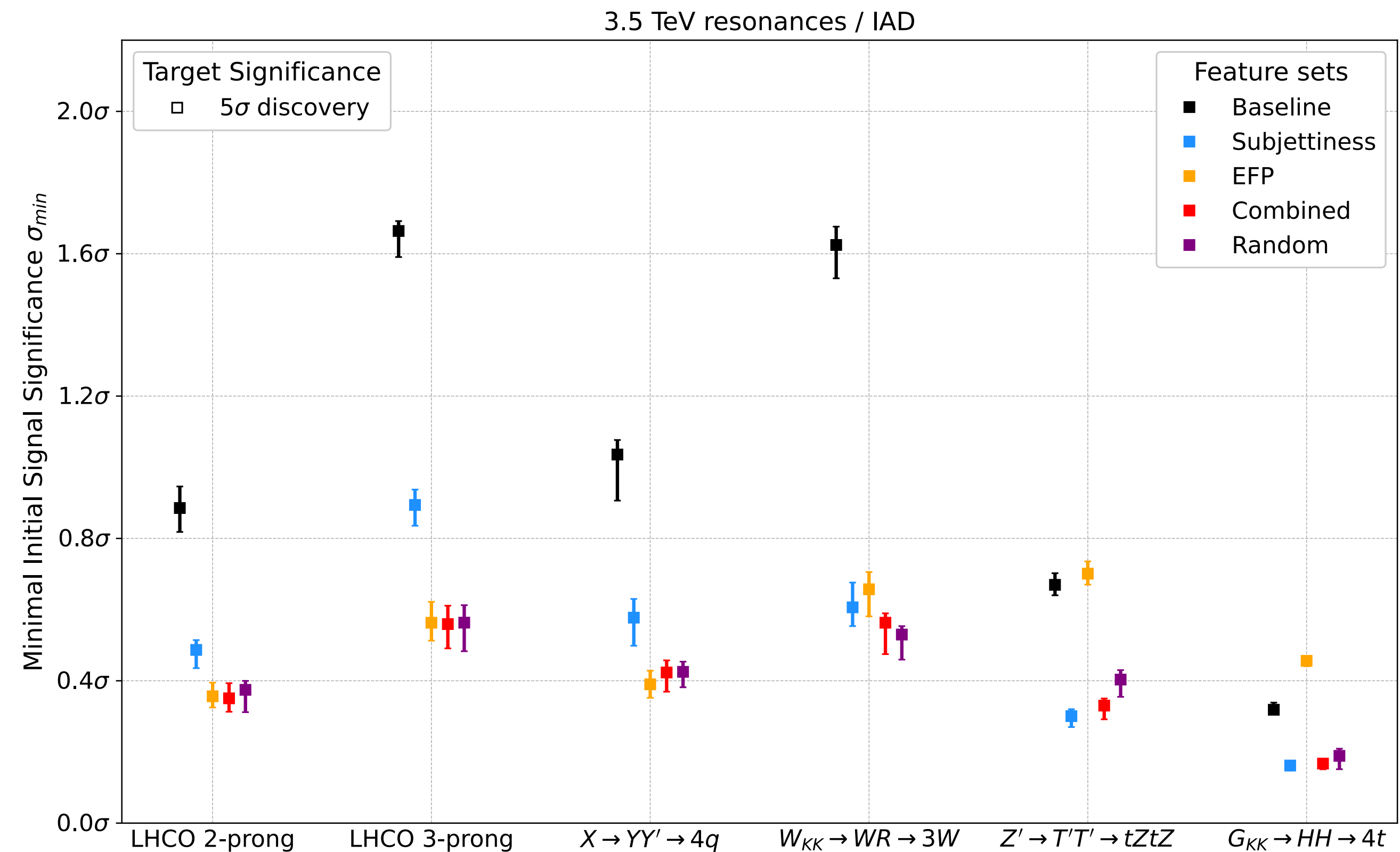
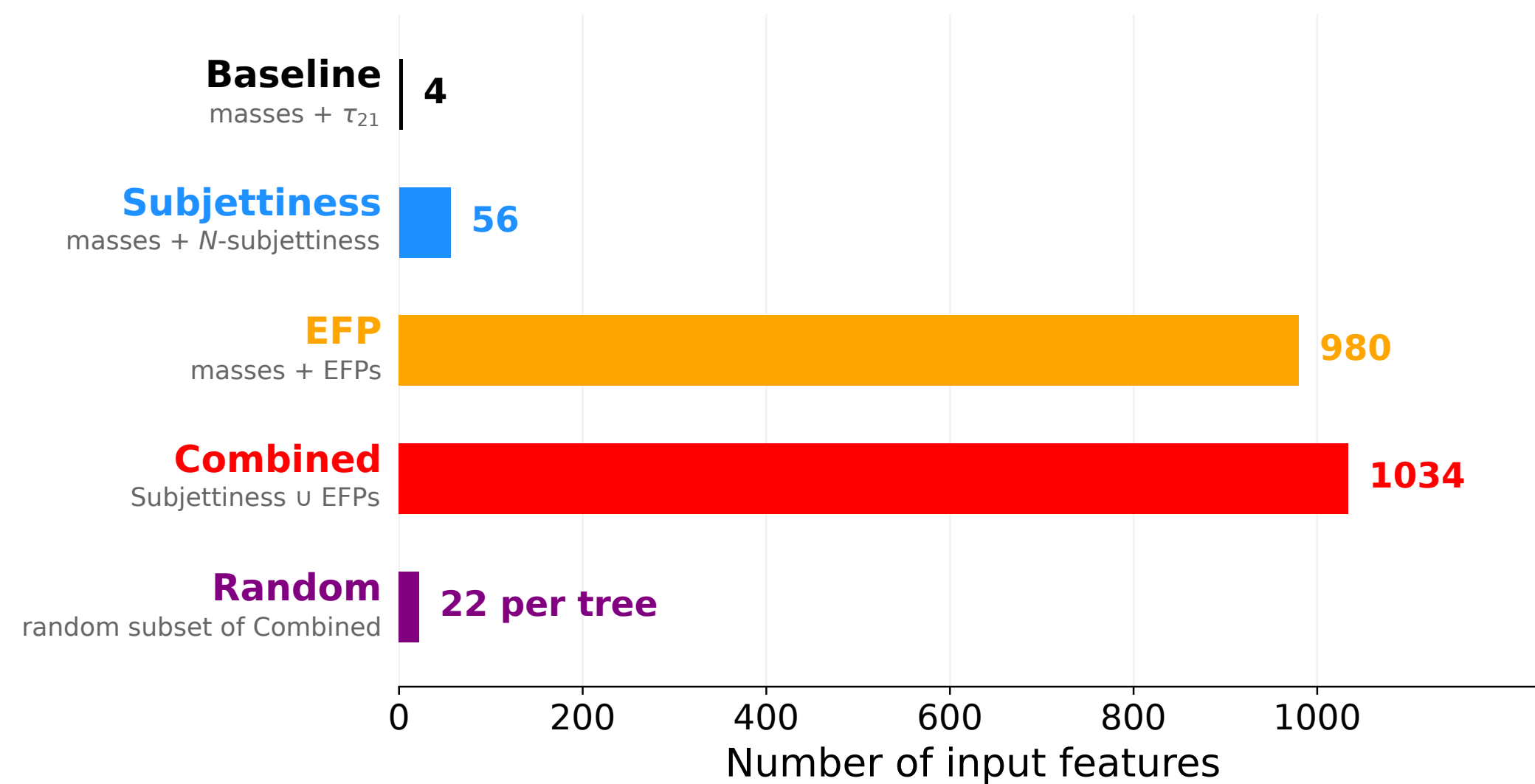
The Idea:

- Leveraging the ensemble in our classifier architecture
- Training each ensemble member on a different random subset of features
- So 50 BDTs each trained on 22 randomly drawn features from the Combined set



Random: Same Performance But Faster

- Training a BDT on 22 features instead of 1034
→ ~ 20 times faster run time
- Performance similar to the **Combine** set





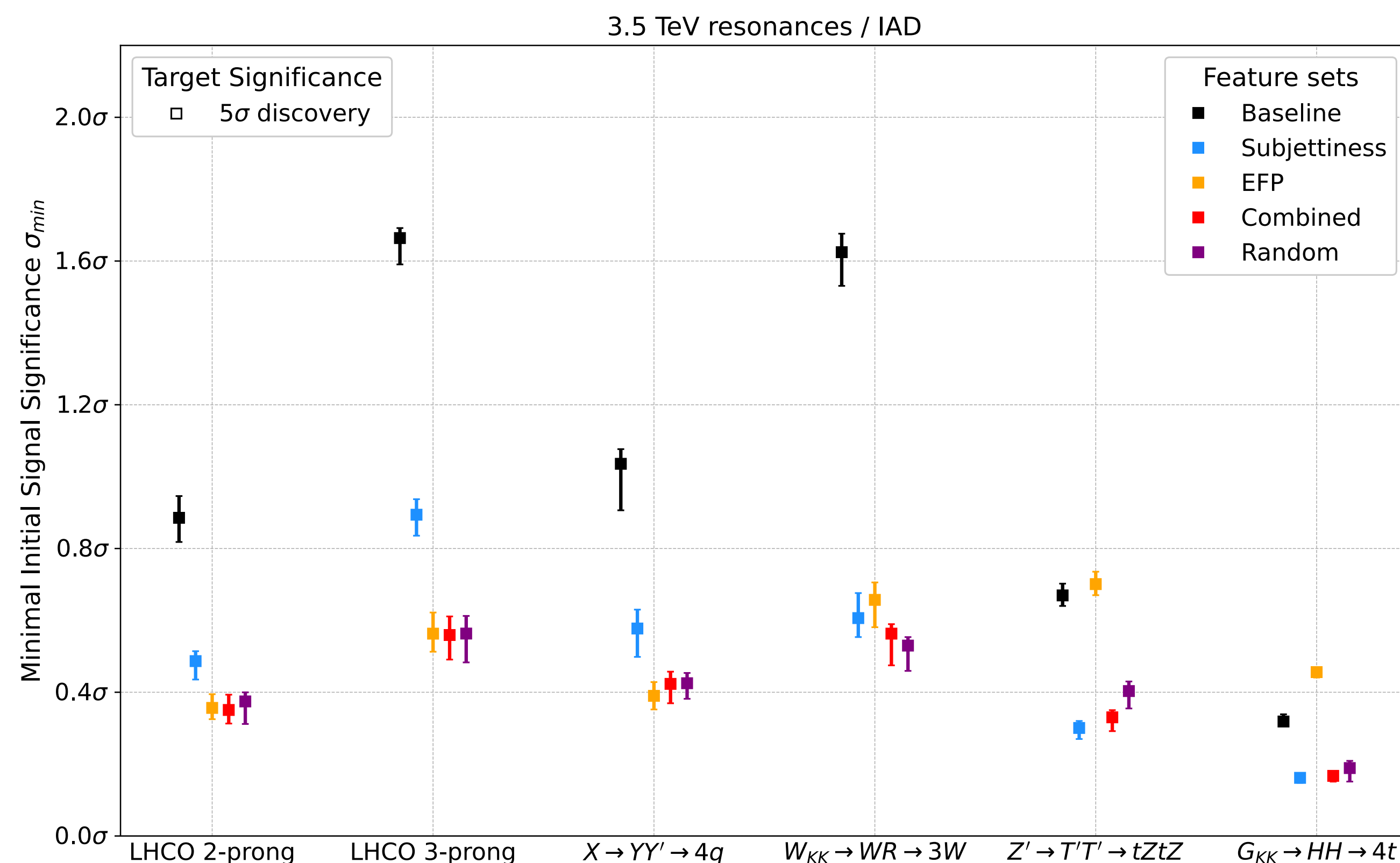
GET THE DATA

Conclusion



READ THE PAPER

- New signal benchmarks compatible with the LHCO background
- The Combined set yielding on average a factor of ~ 2.5 better discovery potential than the Baseline
- Minimizing runtime by a factor of ~ 20 with Random set
- CWoLa hunting produces similar results as the IAD case



Backup

Classifier Significance

- We perform a cut and count experiment using the Asimov estimate for the classifier significance:

$$\mathcal{S} = \sqrt{2[(s + b)\ln\left(1 + \frac{s}{b}\right) - s]}, \text{ where } s = S \cdot \epsilon_S(t) \text{ and } b = B \cdot \epsilon_B(t)$$

are the signal and background events after the classifier cut t , respectively.

- We find the σ_{min} by solving $\max(\mathcal{S})(\sigma_{min}) = 5\sigma$

CWoLa Results

