Boosting the LHC resonance search program with Sophon

Partially based on [arXiv:2405.12972](#) [Github] [Dataset] [Model] [Colab Demo]

Congqiao Li (李聪乔), *Peking University*

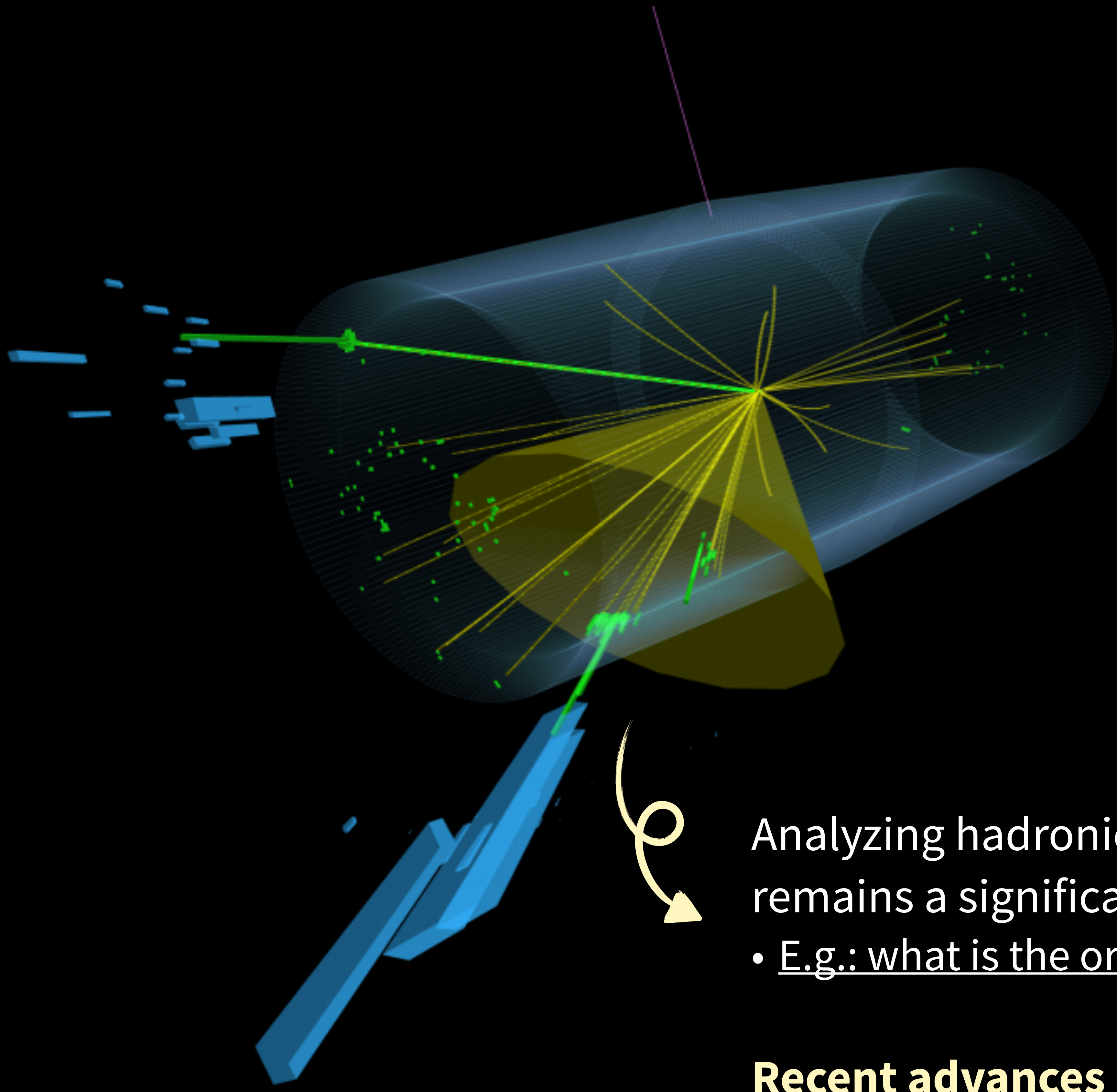
this talk is based mainly on the work with:

Antonios Agapitos[1], Jovin Drews[2], Javier Duarte[3], Dawei Fu[1], Leyun Gao[1], Raghav Kansal[3], Gregor Kasieczka[2], Louis Moureaux[2], Huilin Qu[4], Cristina Mantilla Suarez[5], Qiang Li[1]

1) Peking U. 2) Hamburg U. 3) UC San Diego 4) CERN 5) FNAL

29th Mini-workshop on the frontier of LHC, Fuzhou

15 December, 2024



Analyzing hadronic behavior in LHC events remains a significant challenge

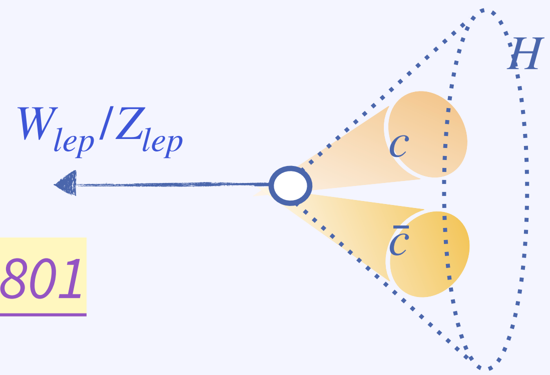
- E.g.: what is the origin of this jet?

Recent advances in ML/AI offer transformative insights!

Inspiring results from CMS *to address hadronic final state*

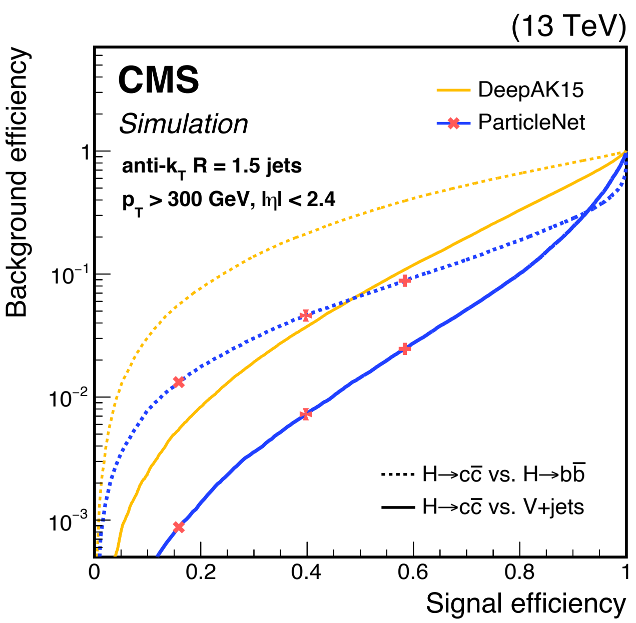
VH→cc search
(merged region)

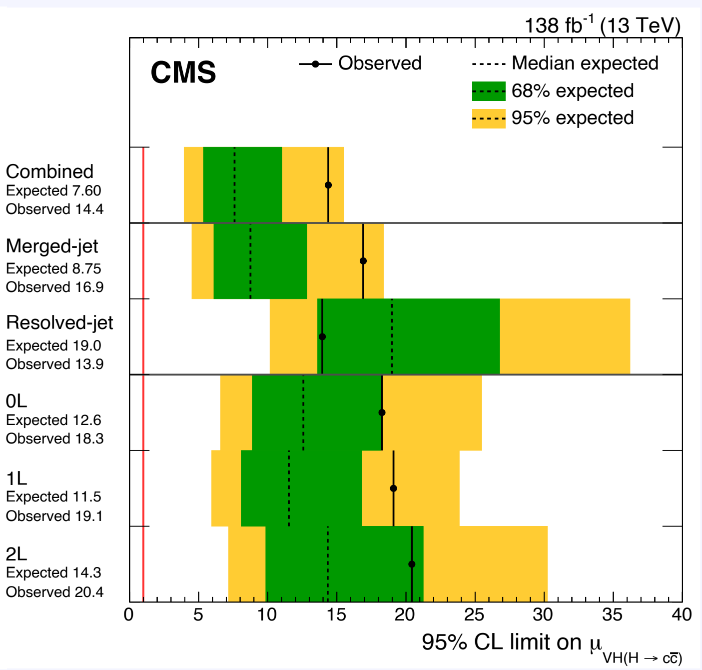
PRL 131 (2023) 061801



- With merged+resolved region combined, achieve most stringent direct limit on κ_c: $1.1 < |\kappa_c| < 5.5$

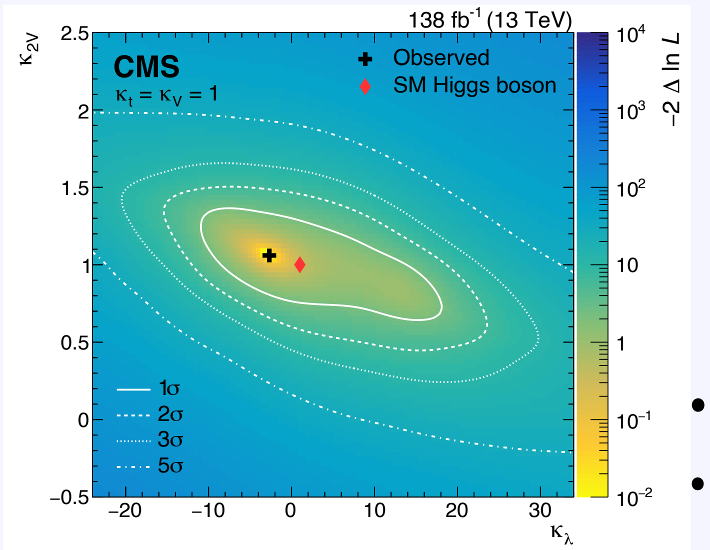
largely improved sensitivity!

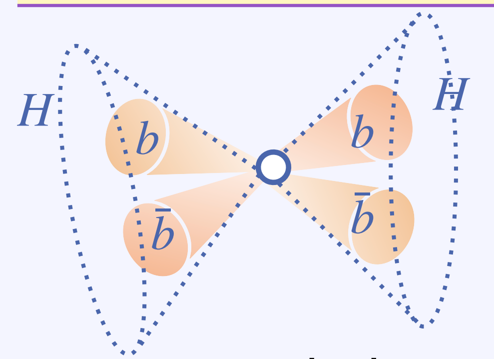




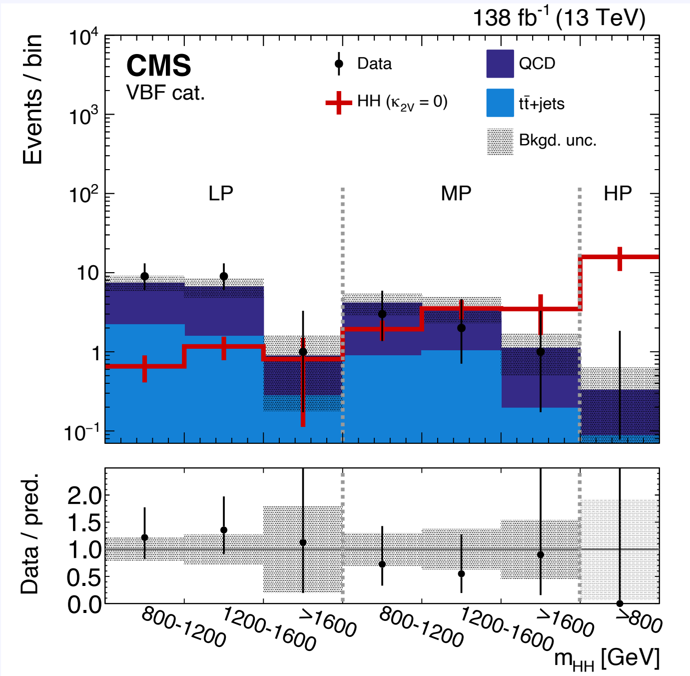
Boosted HH→4b search

PRL 131 (2023) 041803





- First time excluding $\kappa_{2V} = 0$
- Set competitive constraints on κ_λ



Inspiring results from CMS *to address hadronic final state*

VH→cc search (merged region)

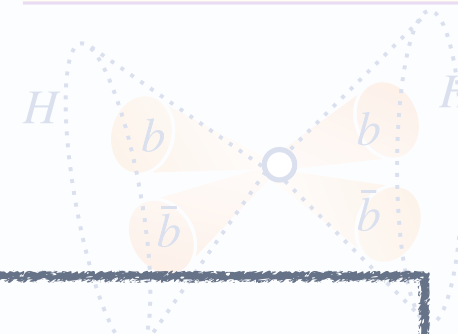
PRL 131 (2023) 061801

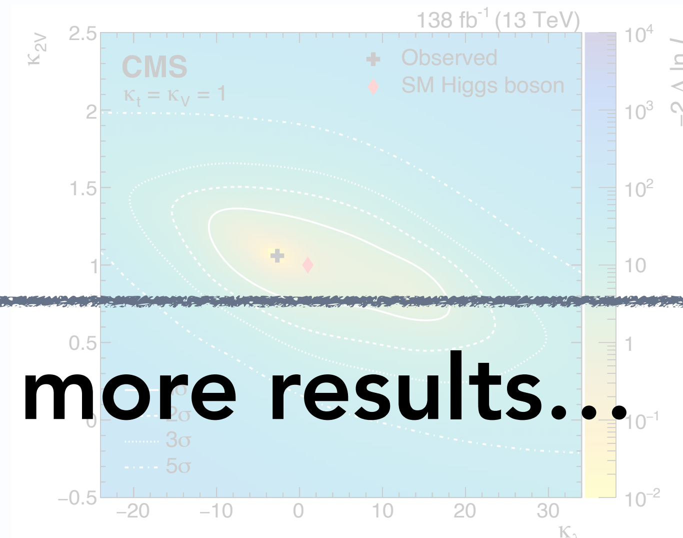
- With merged+resolved region combined, achieve most stringent direct limit on κ_c: $1.1 < |\kappa_c| < 5.5$

largely improved sensitivity!

Boosted HH→4b search

PRL 131 (2023) 041803

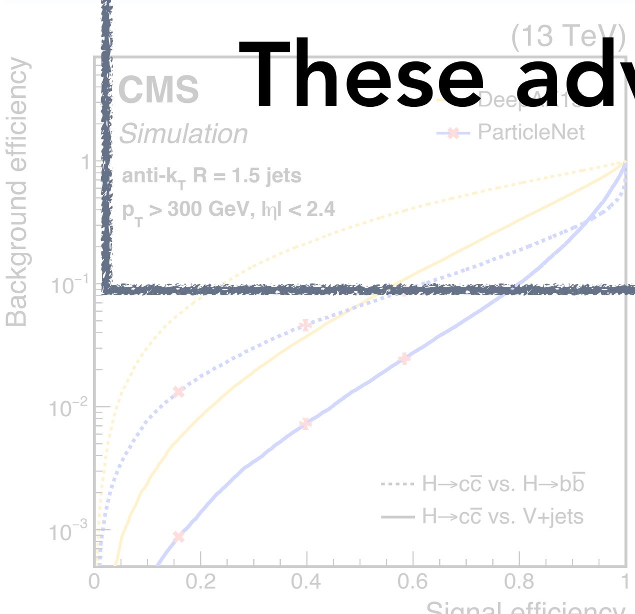


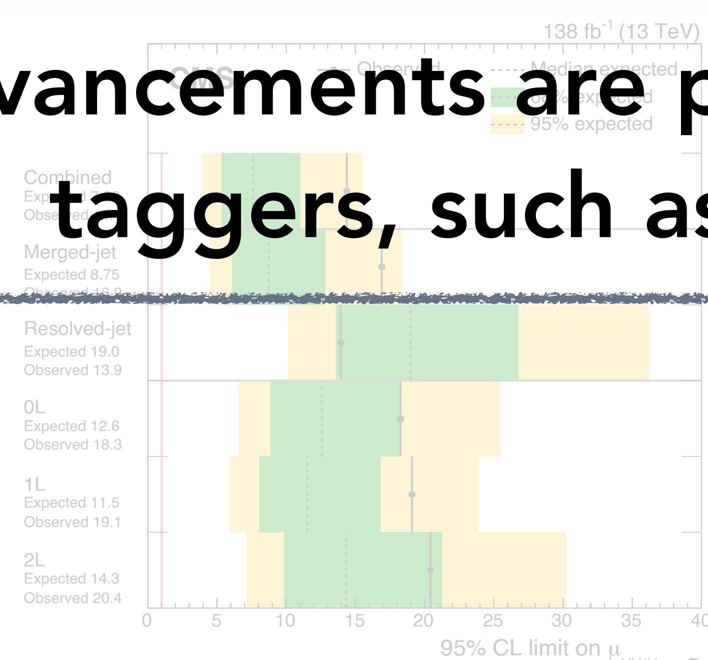


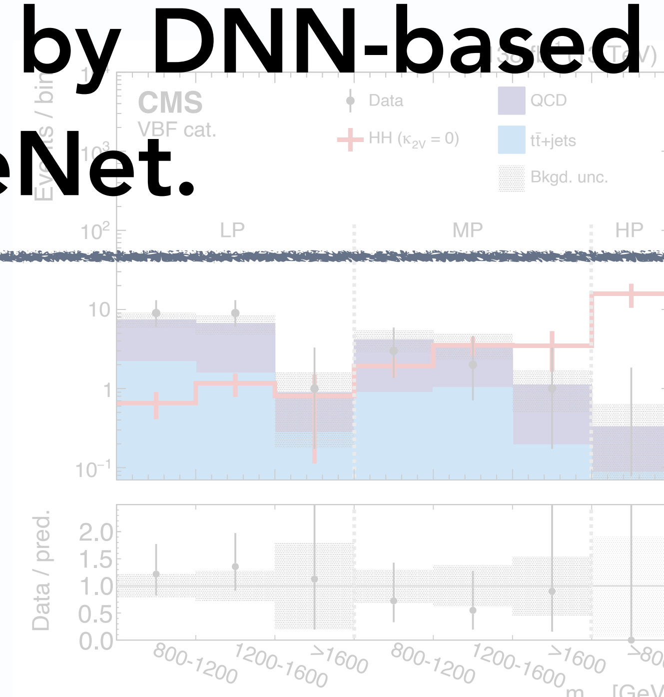
- First time excluding $\kappa_{2V} = 0$
- Set competitive constraints on κ_λ

and many, many more results…

These advancements are powered by DNN-based taggers, such as ParticleNet.

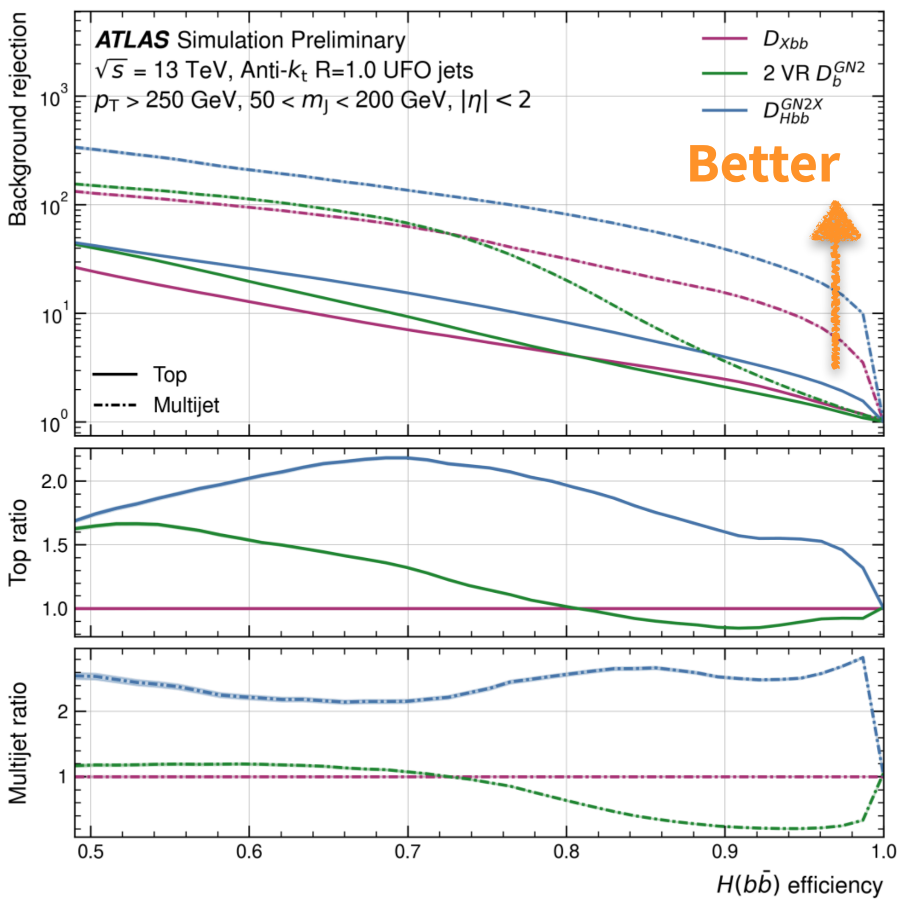




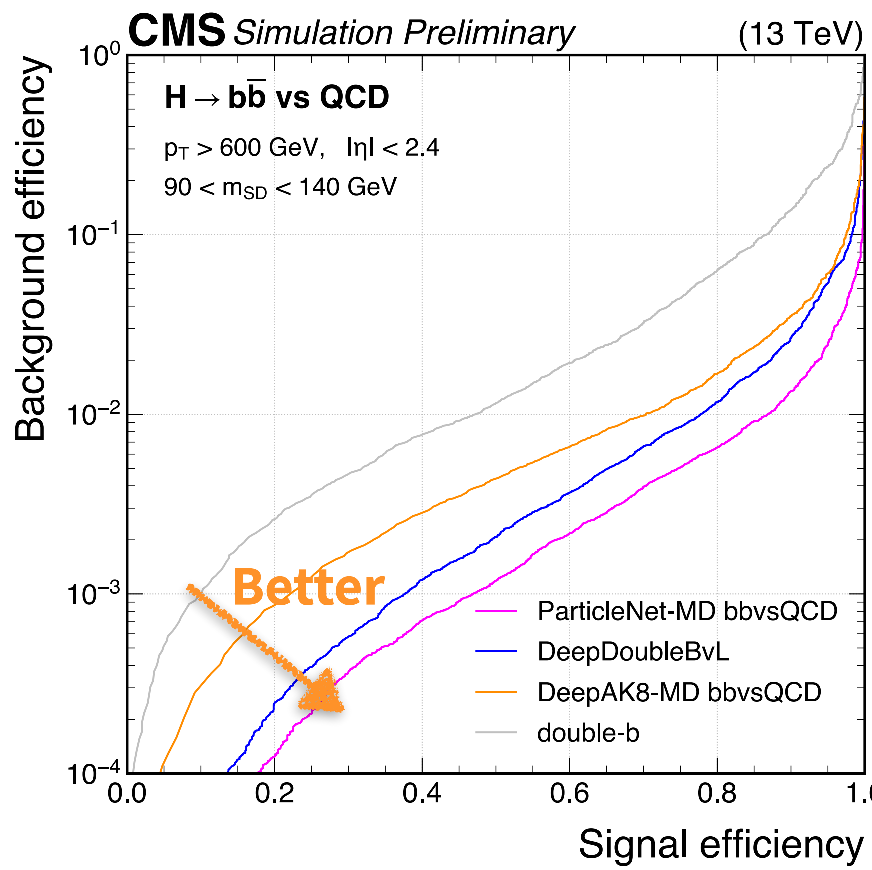


Inspiring progress on identifying H→bb̄/cc̄ jets

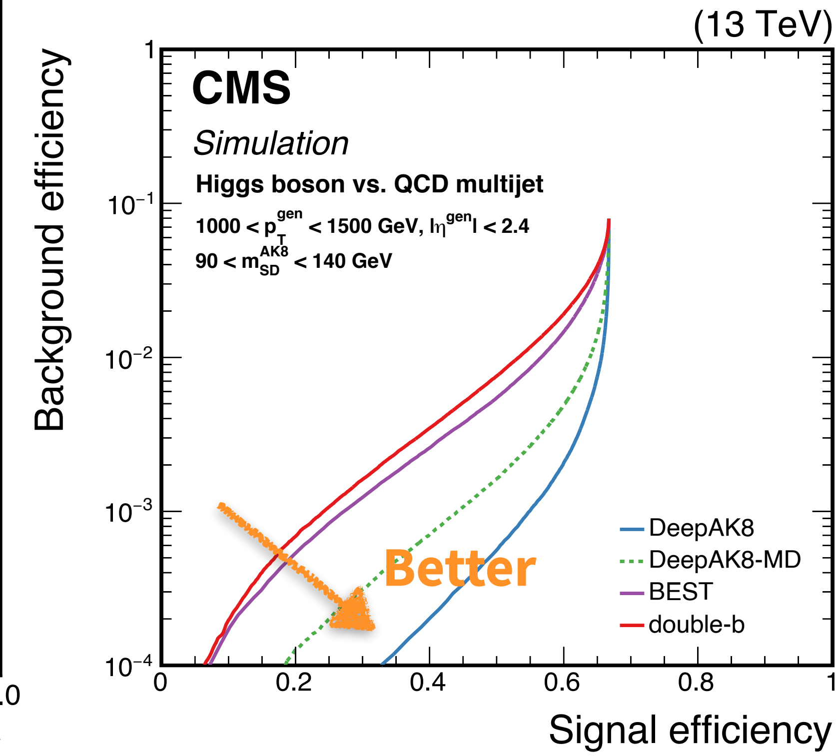
ATL-PHYS-PUB-2023-021



CMS-PAS-BTV-22-001



JINST 15 (2020) P06005



Transformer-based GN2X tagger: **~x3 📈 QCD and x2 top background rejection**

DeepAK8 → ParticleNet: **x5 📈 QCD background rejection**

Comparing with early approaches

Another **~x5** 📈 improvement achieved

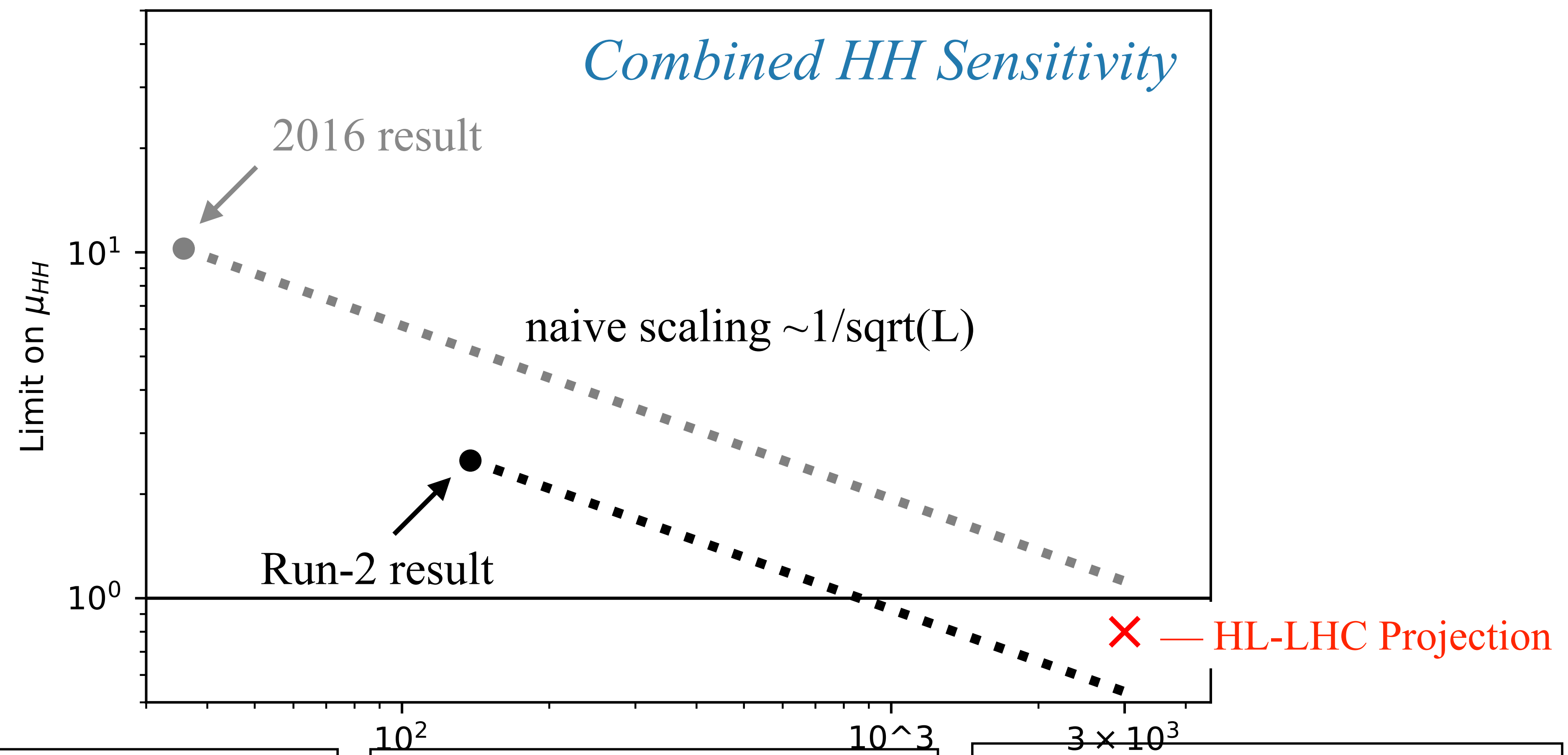
Huge improvements in recent 5 years

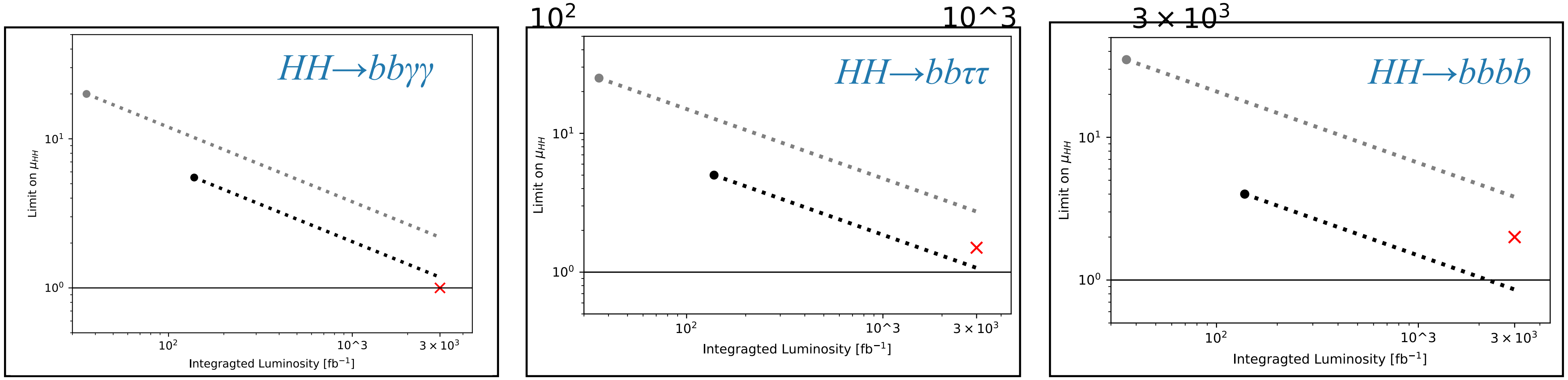
LHC data: Has the mining reached its limits?

Slides by John Alison at LHCP 2024

Future of HH ?

Nature 607, 60–68 (2022).

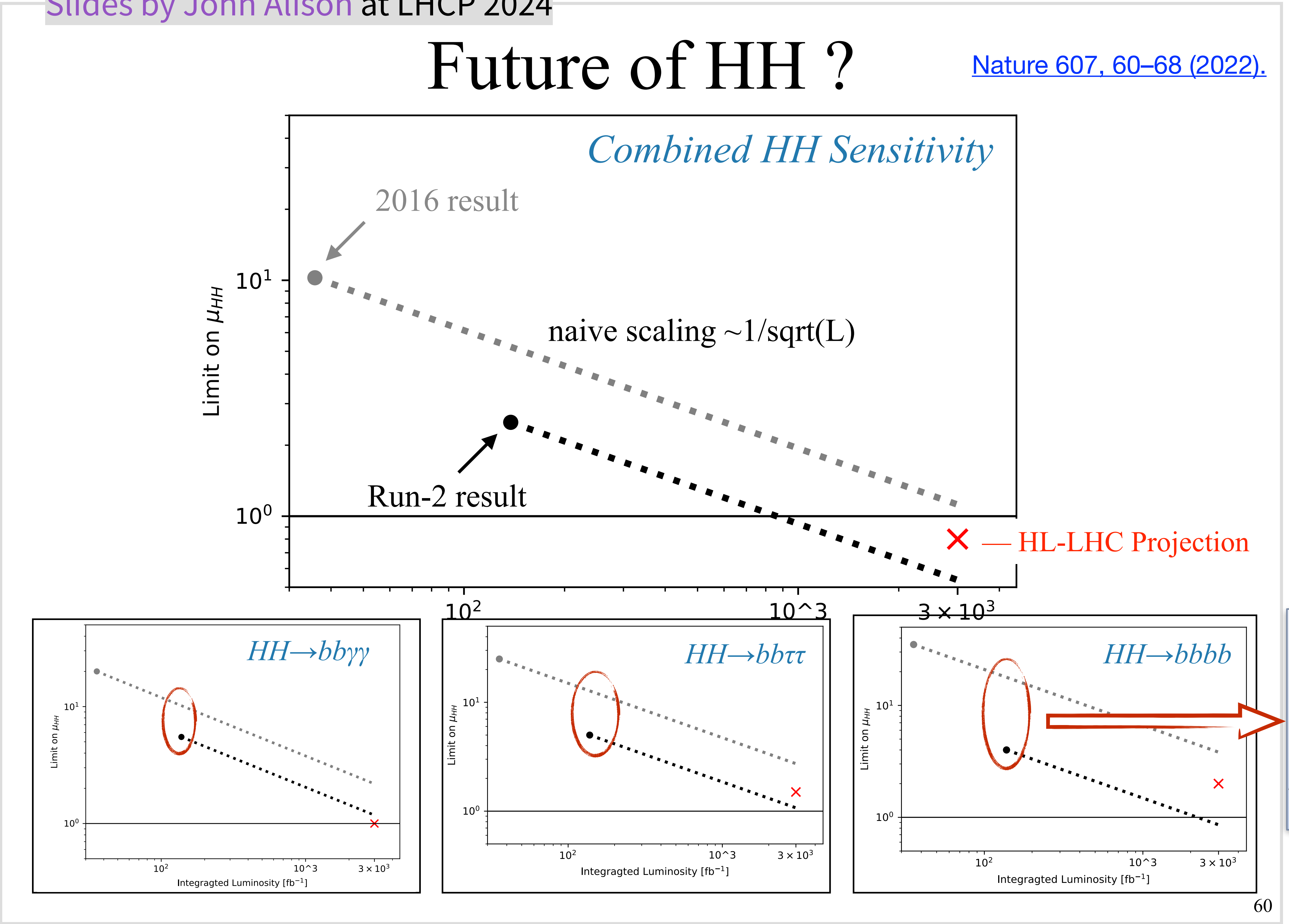




60

LHC data: Has the mining reached its limits?

Slides by John Alison at LHCP 2024



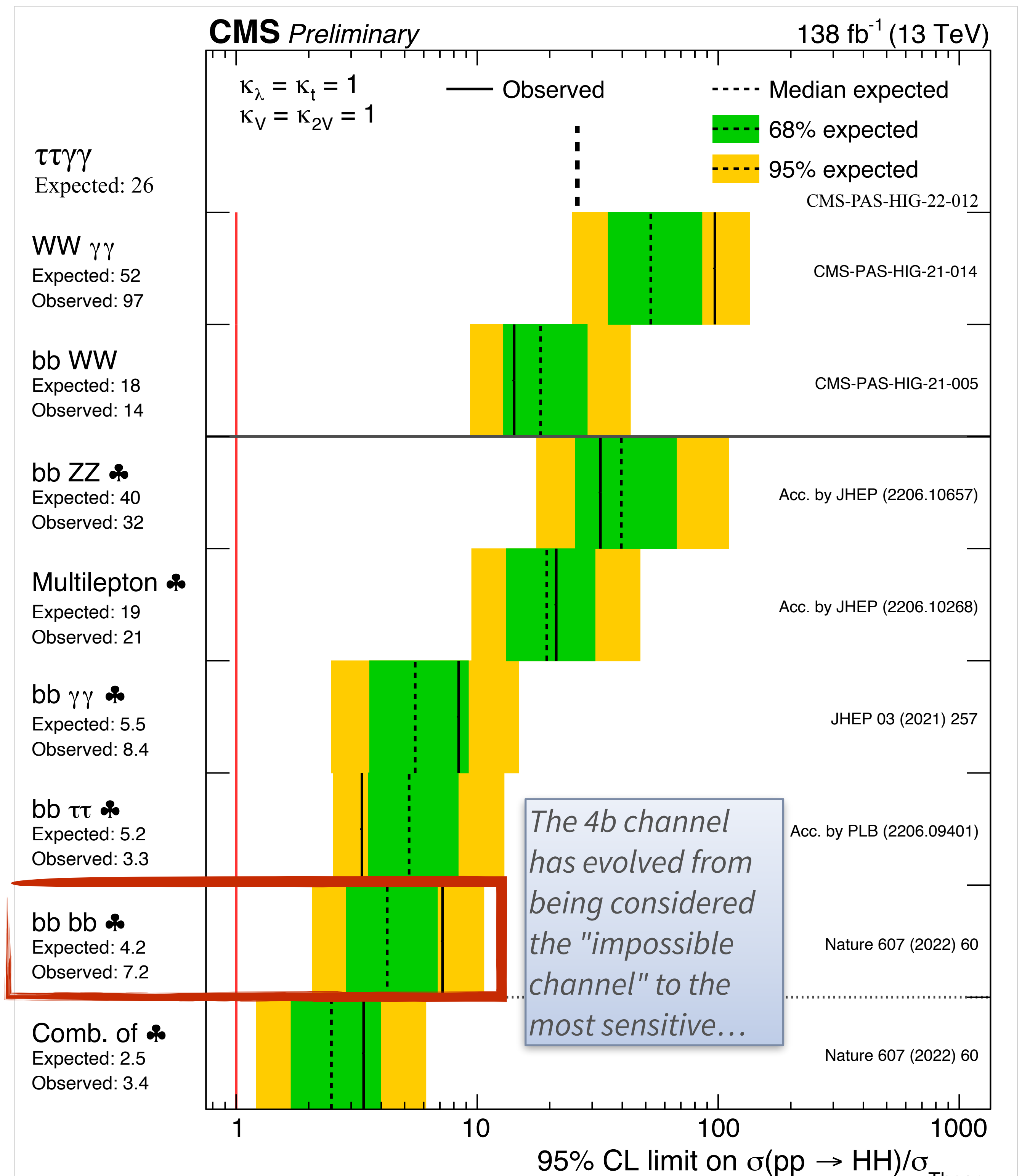
*More improvements achieved through the renovation of analysis techniques, **particularly in more hadronic channels.***

LHC data: Has the mining reached its limits?

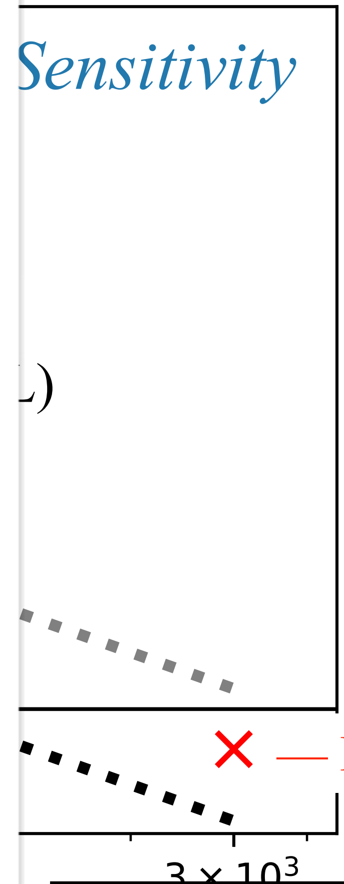
Slides by John Alison at LHCP 2024

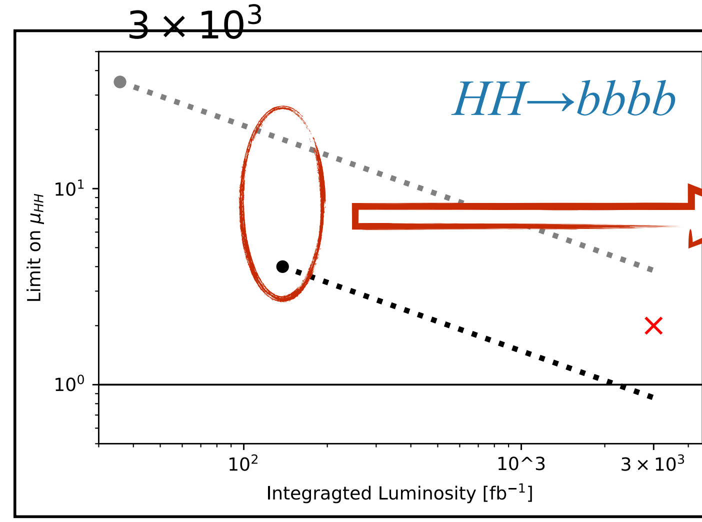
Future of HH ?

Nature 607, 60–68 (2022).



The 4b channel has evolved from being considered the "impossible channel" to the most sensitive…





60

More improvements achieved through the renovation of analysis techniques, ***particularly in more hadronic channels.***

How can we accelerate the pace?

What we possess in ATLAS and CMS

CMS: DeepAK8 and ParticleNet algorithms

JINST 15 (2020) P06005

Output

Category	Label
Higgs	H (bb)
	H (cc)
	H (VV*→qqqq)
Top	top (bcq)
	top (bqq)
	top (bc)
	top (bq)
W	W (cq)
	W (qq)
Z	Z (bb)
	Z (cc)
	Z (qq)
QCD	QCD (bb)
	QCD (cc)
	QCD (b)
	QCD (c)
	QCD (others)

DeepAK8-MD, ParticleNet-MD, and DeepDoubleX algorithms

CMS-PAS-BTV-22-001

- Focus on variable-mass resonance decays
- X→bb, cc, qq and QCD (5 subclasses)

GN2X tagger

ATL-PHYS-PUB-2023-021

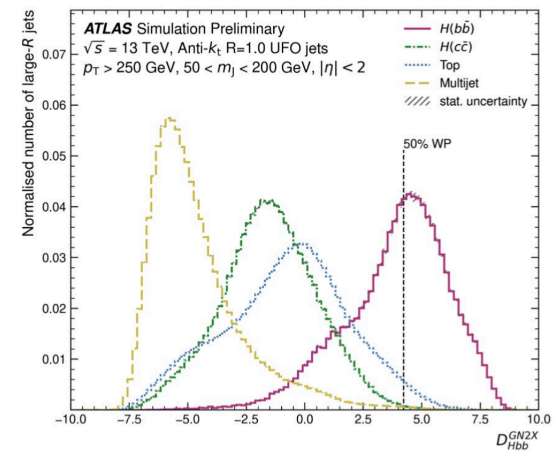
- including flat-mass H→bb, cc and t→bqq samples, with QCD

Jackson's slides

GN2X Outputs

- GN2X adds a H → cc output class in addition to the H → bb, top and QCD classes from the previous tagger
- A discriminant score is built using a weighted log likelihood ratio similar to what's used for small-R tagging
- GN2X also includes the same auxiliary vertexing and track origin classification tasks present in GN1/GN2

$$D_{\mathrm{Hbb}}^{\mathrm{GN2X}} = \ln\left(\frac{p_{\mathrm{Hbb}}}{f_{\mathrm{Hcc}}\cdot p_{\mathrm{Hcc}} + f_{\mathrm{top}}\cdot p_{\mathrm{top}} + (1 - f_{\mathrm{Hcc}} - f_{\mathrm{top}})\cdot p_{\mathrm{QCD}}}\right)$$



How can we accelerate the pace?

What we possess in ATLAS and CMS

CMS: DeepAK8 and ParticleNet algorithms

JINST 15 (2020) P06005

Output

Category	Label
Higgs	H (bb)
	H (cc)
	H (VV*→qqqq)
Top	top (bcq)
	top (bqq)
	top (bc)

DeepAK8-MD, ParticleNet-MD, and DeepDoubleX algorithms

CMS-PAS-BTV-22-001

- Focus on variable-mass resonance decays
- X→bb, cc, qq and QCD (5 subclasses)

GN2X tagger

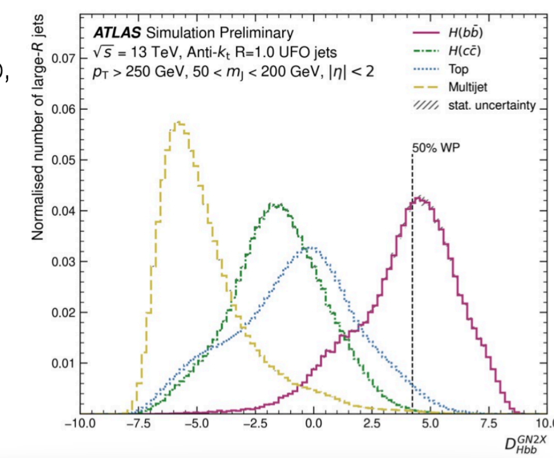
ATL-PHYS-PUB-2023-021

- including flat-mass H→bb, cc and t→bqq samples, with QCD

Jackson's slides

GN2X Outputs

- GN2X adds a H → cc output class in addition to the H → bb, top and QCD classes from the previous tagger
- A discriminant score is built using a weighted log likelihood ratio similar to what's used for small-R tagging
- …2X also includes the same auxiliary vertexing and track …in classification tasks present in GN1/GN2



What we propose: "Large model for large-scale classification!"

We include all possible final states in the search

Major types	Index range	Label names
Resonant jets: $X \to 2$ prong	0–14	bb, cc, ss, qq, bc, cs, bq, cq, sq, gg, ee, $\mu\mu$, $\tau_h\tau_e$, $\tau_h\tau_\mu$, $\tau_h\tau_h$
Resonant jets: $X \to 3$ or 4 prong	15–160	$bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_h\tau_e$, $bb\tau_h\tau_\mu$, $bb\tau_h\tau_h$, bbb, bbc, bbs, bbq, bbg, bbe, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_h\tau_e$, $cc\tau_h\tau_\mu$, $cc\tau_h\tau_h$, ccb, ccc, ccs, ccq, ccg, cce, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_h\tau_e$, $ss\tau_h\tau_\mu$, $ss\tau_h\tau_h$, ssb, ssc, sss, ssq, ssg, sse, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_h\tau_e$, $qq\tau_h\tau_\mu$, $qq\tau_h\tau_h$, qqb, qqc, qqs, qqq, qqg, qqe, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_h\tau_e$, $gg\tau_h\tau_\mu$, $gg\tau_h\tau_h$, ggb, ggc, ggs, ggq, ggg, gge, $gg\mu$, bee, cee, see, qee, gee, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_h\tau_e$, $c\tau_h\tau_e$, $s\tau_h\tau_e$, $q\tau_h\tau_e$, $g\tau_h\tau_e$, $b\tau_h\tau_\mu$, $c\tau_h\tau_\mu$, $s\tau_h\tau_\mu$, $q\tau_h\tau_\mu$, $g\tau_h\tau_\mu$, $b\tau_h\tau_h$, $c\tau_h\tau_h$, $s\tau_h\tau_h$, $q\tau_h\tau_h$, $g\tau_h\tau_h$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, bcs, bcq, bsq, csq, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_h\nu$, $cs\tau_h\nu$, $bq\tau_h\nu$, $cq\tau_h\nu$, $sq\tau_h\nu$, $qq\tau_h\nu$
QCD jets	161–187	$bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, bbc, $bbss$, bbs, bb, $bccss$, $bccs$, bcc, $bcss$, bcs, bc, bss, bs, b, $ccss$, ccs, cc, css, cs, c, ss, s, others

How can we accelerate the pace?

What we possess in ATLAS and CMS

CMS: DeepAK8 and ParticleNet algorithms

JINST 15 (2020) P06005

Output

Category	Label
Higgs	H (bb)
	H (cc)
	H (VV*→qqqq)
Top	top (bcq)
	top (bqq)
	top (bc)

DeepAK8-MD, ParticleNet-MD, and DeepDoubleX algorithms

CMS-PAS-BTV-22-001

- Focus on variable-mass resonance decays
- X→bb, cc, qq and QCD (5 subclasses)

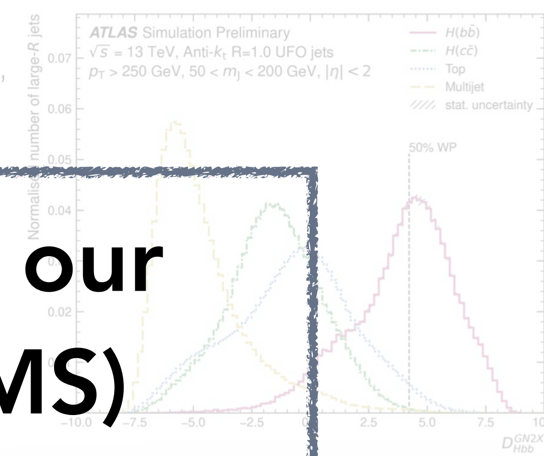
GN2X tagger *ATL-PHYS-PUB-2023-021*

- including flat-mass H→bb, cc and t→bqq samples, with QCD

Jackson's slides

GN2X Outputs

- GN2X adds a H → cc output class in addition to the H → bb, top and QCD classes from the previous tagger
- A discriminant score is built using a weighted log likelihood ratio similar to what's used for small-R tagging
- GN2X also includes the same auxiliary vertexing and track ... in classification tasks present in GN1/GN2



A global jet model suitable to deploy to our general-purpose experiment (ATLAS/CMS)

What we propose: "Large model for large-scale classification!"

We include all possible final states in the search

Major types	Index range	Label names
Resonant jets: $X \to$ 2 prong	0–14	bb, cc, ss, qq, bc, cs, bq, cq, sq, gg, ee, $\mu\mu$, $\tau_h\tau_e$, $\tau_h\tau_\mu$, $\tau_h\tau_h$
Resonant jets: $X \to$ 3 or 4 prong	15–160	$bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_h\tau_e$, $bb\tau_h\tau_\mu$, $bb\tau_h\tau_h$, bbb, bbc, bbs, bbq, bbg, bbe, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_h\tau_e$, $cc\tau_h\tau_\mu$, $cc\tau_h\tau_h$, ccb, ccc, ccs, ccq, ccg, cce, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_h\tau_e$, $ss\tau_h\tau_\mu$, $ss\tau_h\tau_h$, ssb, ssc, sss, ssq, ssg, sse, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_h\tau_e$, $qq\tau_h\tau_\mu$, $qq\tau_h\tau_h$, qqb, qqc, qqs, qqq, qqg, qqe, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_h\tau_e$, $gg\tau_h\tau_\mu$, $gg\tau_h\tau_h$, ggb, ggc, ggs, ggq, ggg, gge, $gg\mu$, bee, cee, see, qee, gee, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_h\tau_e$, $c\tau_h\tau_e$, $s\tau_h\tau_e$, $q\tau_h\tau_e$, $g\tau_h\tau_e$, $b\tau_h\tau_\mu$, $c\tau_h\tau_\mu$, $s\tau_h\tau_\mu$, $q\tau_h\tau_\mu$, $g\tau_h\tau_\mu$, $b\tau_h\tau_h$, $c\tau_h\tau_h$, $s\tau_h\tau_h$, $q\tau_h\tau_h$, $g\tau_h\tau_h$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, bcs, bcq, bsq, csq, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_h\nu$, $cs\tau_h\nu$, $bq\tau_h\nu$, $cq\tau_h\nu$, $sq\tau_h\nu$, $qq\tau_h\nu$
QCD jets	161–187	$bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, bbc, $bbss$, bbs, bb, $bccss$, $bccs$, bcc, $bcss$, bcs, bc, bss, bs, b, $ccss$, ccs, cc, css, cs, c, ss, s, others

How can we accelerate the pace?

Sophon (智子)*: Signature-Oriented Pre-training for Heavy-resonant ObservatioN*

	Index range	Label names
$X \to 2$ prong	0–14	bb, cc, ss, qq, bc, cs, bq, cq, sq, gg, ee, $\mu\mu$, $\tau_\mathrm{h}\tau_e$, $\tau_\mathrm{h}\tau_\mu$, $\tau_\mathrm{h}\tau_\mathrm{h}$
Resonant jets: $X \to 3$ or 4 prong	15–160	$bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_\mathrm{h}\tau_e$, $bb\tau_\mathrm{h}\tau_\mu$, $bb\tau_\mathrm{h}\tau_\mathrm{h}$, bbb, bbc, bbs, bbq, bbg, bbe, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_\mathrm{h}\tau_e$, $cc\tau_\mathrm{h}\tau_\mu$, $cc\tau_\mathrm{h}\tau_\mathrm{h}$, ccb, ccc, ccs, ccq, ccg, cce, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_\mathrm{h}\tau_e$, $ss\tau_\mathrm{h}\tau_\mu$, $ss\tau_\mathrm{h}\tau_\mathrm{h}$, ssb, ssc, sss, ssq, ssg, sse, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_\mathrm{h}\tau_e$, $qq\tau_\mathrm{h}\tau_\mu$, $qq\tau_\mathrm{h}\tau_\mathrm{h}$, qqb, qqc, qqs, qqq, qqg, qqe, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_\mathrm{h}\tau_e$, $gg\tau_\mathrm{h}\tau_\mu$, $gg\tau_\mathrm{h}\tau_\mathrm{h}$, ggb, ggc, ggs, ggq, ggg, gge, $gg\mu$, bee, cee, see, qee, gee, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_\mathrm{h}\tau_e$, $c\tau_\mathrm{h}\tau_e$, $s\tau_\mathrm{h}\tau_e$, $q\tau_\mathrm{h}\tau_e$, $g\tau_\mathrm{h}\tau_e$, $b\tau_\mathrm{h}\tau_\mu$, $c\tau_\mathrm{h}\tau_\mu$, $s\tau_\mathrm{h}\tau_\mu$, $q\tau_\mathrm{h}\tau_\mu$, $g\tau_\mathrm{h}\tau_\mu$, $b\tau_\mathrm{h}\tau_\mathrm{h}$, $c\tau_\mathrm{h}\tau_\mathrm{h}$, $s\tau_\mathrm{h}\tau_\mathrm{h}$, $q\tau_\mathrm{h}\tau_\mathrm{h}$, $g\tau_\mathrm{h}\tau_\mathrm{h}$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, bcs, bcq, bsq, csq, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_\mathrm{h}\nu$, $cs\tau_\mathrm{h}\nu$, $bq\tau_\mathrm{h}\nu$, $cq\tau_\mathrm{h}\nu$, $sq\tau_\mathrm{h}\nu$, $qq\tau_\mathrm{h}\nu$
QCD jets	161–187	$bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, bbc, $bbss$, bbs, bb, $bccss$, $bccs$, bcc, $bcss$, bcs, bc, bss, bs, b, $ccss$, ccs, cc, css, cs, c, ss, s, others

➔ Major concerns:

- Will the model achieve the best performance for each specific task?
- What can we use this model for, beyond its identification task supported by its final states?

How can we accelerate the pace?

Sophon (智子): *Signature-Oriented Pre-training for Heavy-resonant ObservatioN*

	Index range	Label names
$X \to 2$ prong	0–14	bb, cc, ss, qq, bc, cs, bq, cq, sq, gg, ee, $\mu\mu$, $\tau_h\tau_e$, $\tau_h\tau_\mu$, $\tau_h\tau_h$
Resonant jets: $X \to 3$ or 4 prong	15–160	$bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_h\tau_e$, $bb\tau_h\tau_\mu$, $bb\tau_h\tau_h$, bbb, bbc, bbs, bbq, bbg, bbe, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_h\tau_e$, $cc\tau_h\tau_\mu$, $cc\tau_h\tau_h$, ccb, ccc, ccs, ccq, ccg, cce, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_h\tau_e$, $ss\tau_h\tau_\mu$, $ss\tau_h\tau_h$, ssb, ssc, sss, ssq, ssg, sse, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_h\tau_e$, $qq\tau_h\tau_\mu$, $qq\tau_h\tau_h$, qqb, qqc, qqs, qqq, qqg, qqe, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_h\tau_e$, $gg\tau_h\tau_\mu$, $gg\tau_h\tau_h$, ggb, ggc, ggs, ggq, ggg, gge, $gg\mu$, bee, cee, see, qee, gee, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_h\tau_e$, $c\tau_h\tau_e$, $s\tau_h\tau_e$, $q\tau_h\tau_e$, $g\tau_h\tau_e$, $b\tau_h\tau_\mu$, $c\tau_h\tau_\mu$, $s\tau_h\tau_\mu$, $q\tau_h\tau_\mu$, $g\tau_h\tau_\mu$, $b\tau_h\tau_h$, $c\tau_h\tau_h$, $s\tau_h\tau_h$, $q\tau_h\tau_h$, $g\tau_h\tau_h$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, bcs, bcq, bsq, csq, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_h\nu$, $cs\tau_h\nu$, $bq\tau_h\nu$, $cq\tau_h\nu$, $sq\tau_h\nu$, $qq\tau_h\nu$
QCD jets	161–187	$bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, bbc, $bbss$, bbs, bb, $bccss$, $bccs$, bcc, $bcss$, bcs, bc, bss, bs, b, $ccss$, ccs, cc, css, cs, c, ss, s, others

→ Major concerns:

- Will the model achieve the best performance for each specific task?
  ⇒ *Yes, it does. The Particle Transformer can support each task to reach its optimal performance*
- What can we use this model for, beyond its identification task supported by its final states?
  ⇒ *We can regard it as a true **"based model"** and fine-tune it for wider ranges of downstream tasks*

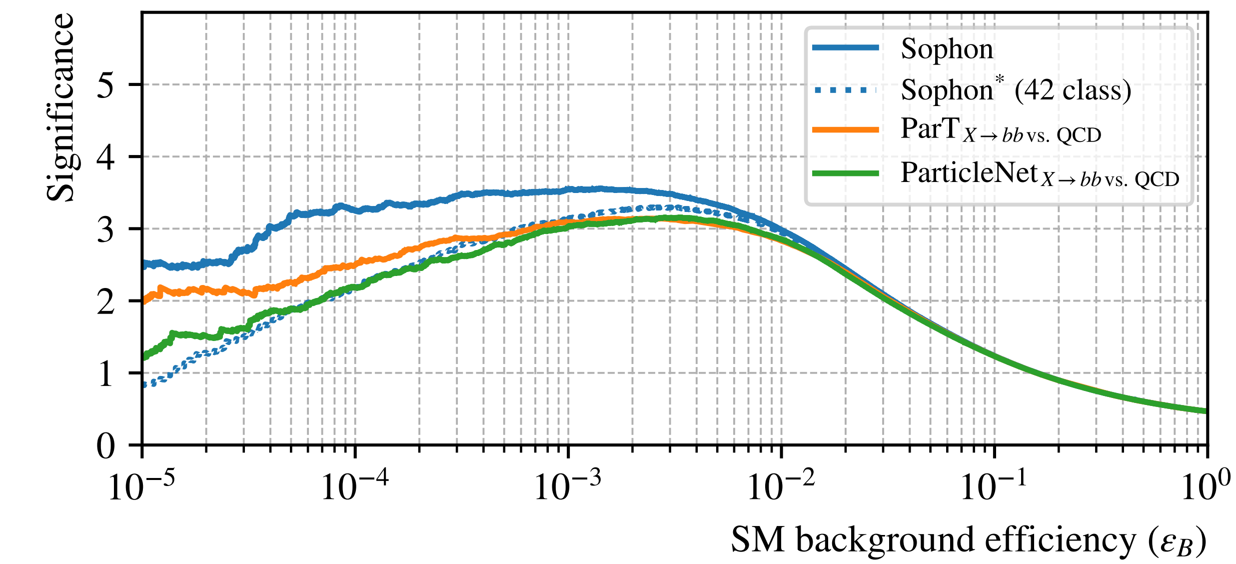
Sophon: performance benchmark

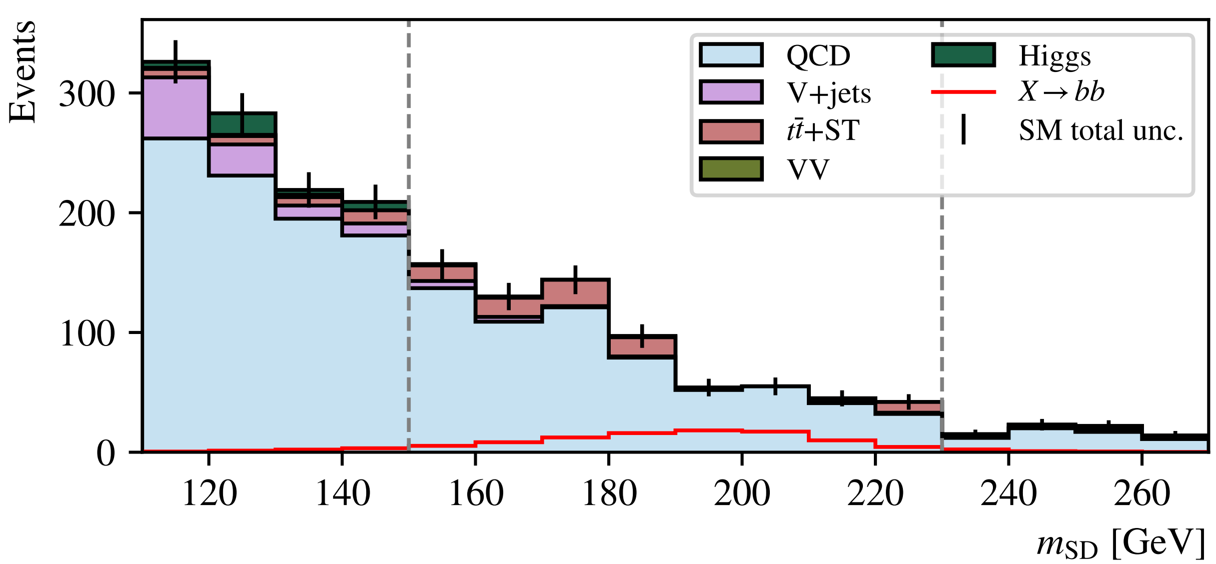
arXiv:2405.12972

Sophon demonstrates the best performance on the direct identification task

Search significance:

$$Z = \sqrt{2((s+b)\log(1+s/b)-s)}$$





- Check discrimination power of X (200 GeV) → **bb** signal vs. all backgrounds

- **Sophon** (training on 188 classes) has best performance

$$\text{discr}(X\to bb\text{ vs. QCD}) = \frac{g_{X\to bb}}{g_{X\to bb} + \sum_{l=1}^{27} g_{\text{QCD}_l}}$$

- Performance gain does come from large-scale classification (compared to **Sophon* (42 classes)**)
- **ParT** and **ParticleNet** for binary classification: they represent the best performance we can reach in experiment now

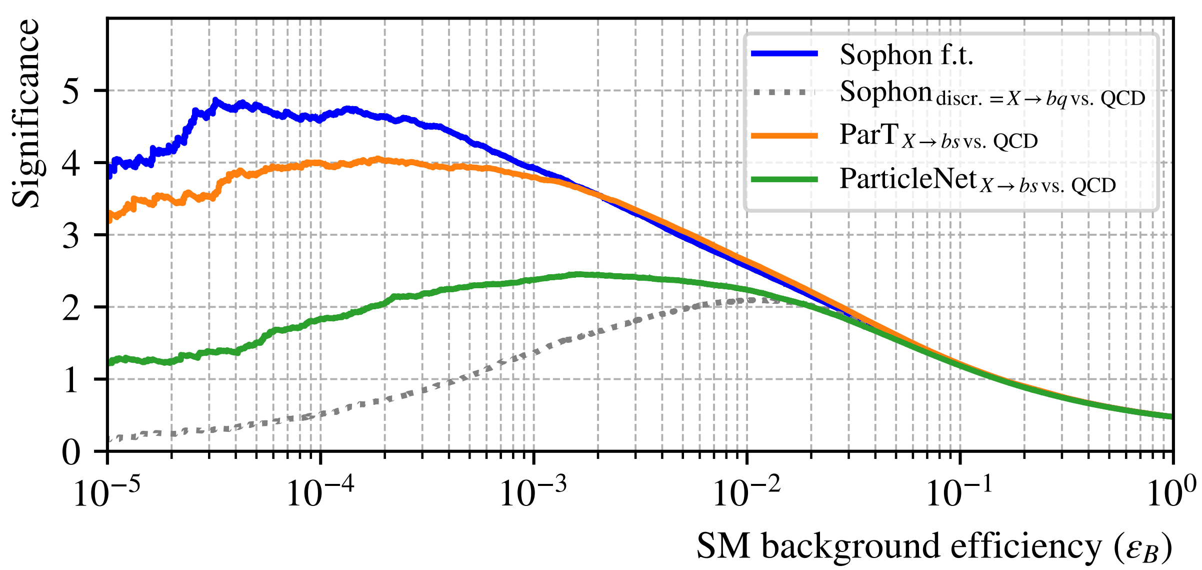
Sophon: performance benchmark

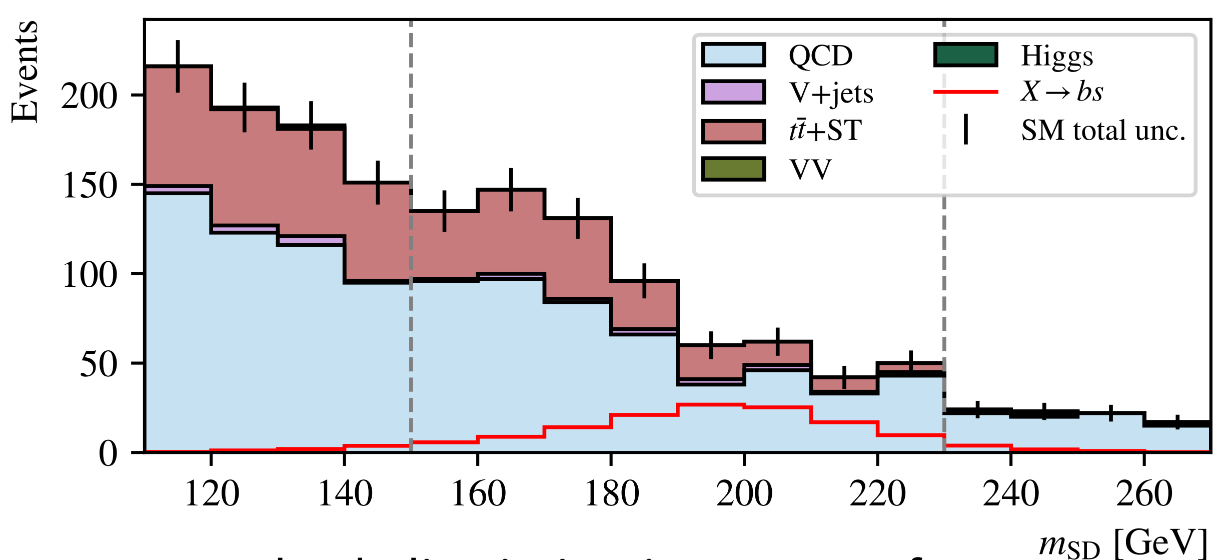
arXiv:2405.12972

Sophon demonstrates the best performance also when fine-tuned for new tasks

Search significance:

$$Z = \sqrt{2((s+b)\log(1+s/b)-s)}$$





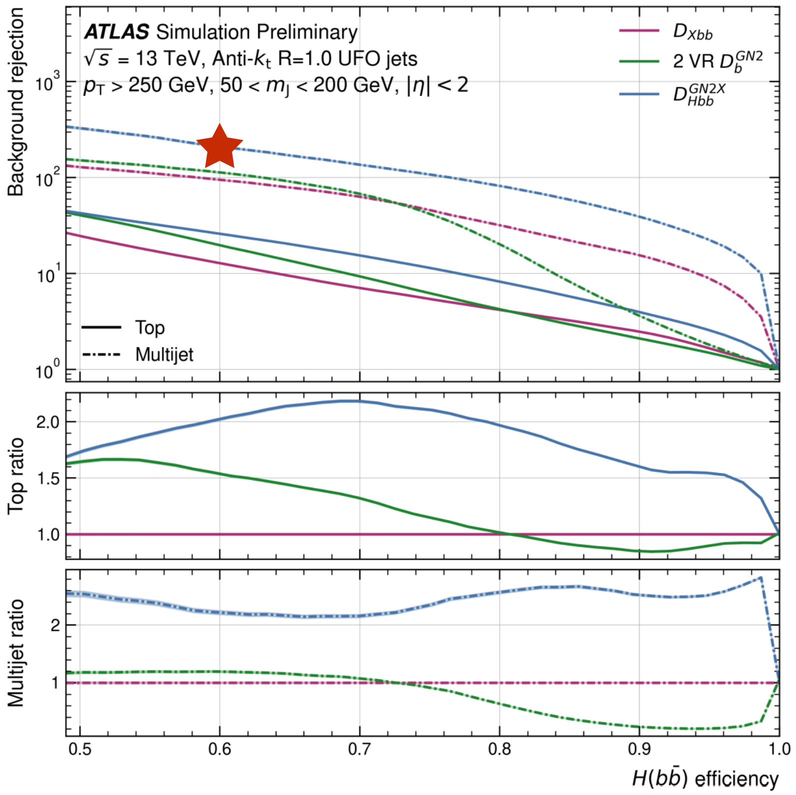
- Check discrimination power of X (200 GeV) → **bs** signal vs. all backgrounds

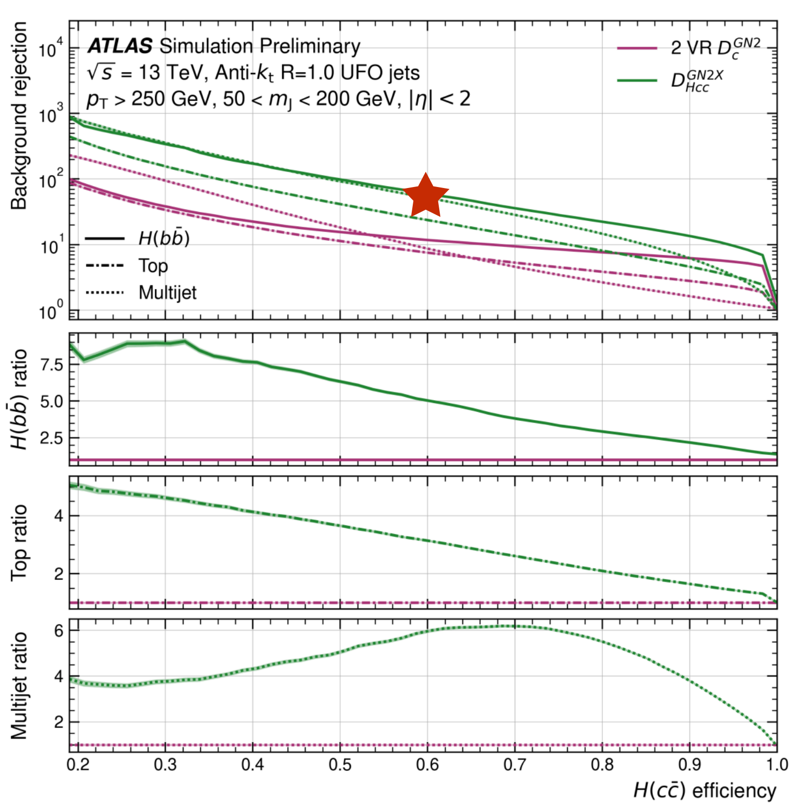
- **Sophon** (training on 188 classes) reaches the best performance **after fine-tuned (via transfer learning)**
- **ParT** and **ParticleNet** for binary X→bs vs QCD classification: they reveal the best performance we can reach in the experiment now

Sophon: close to actual ATLAS/CMS performance?

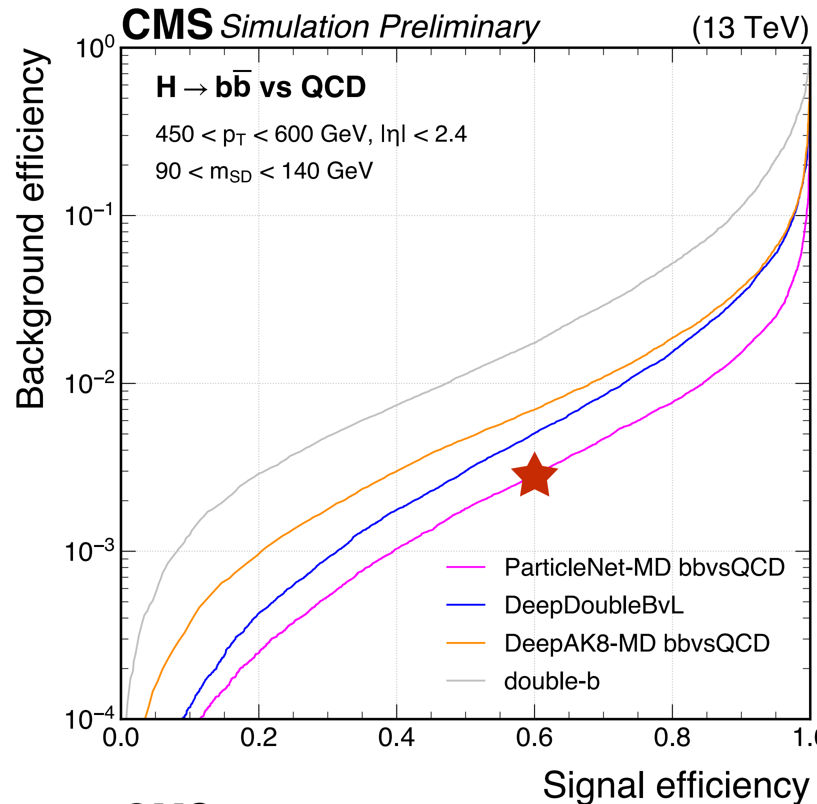
★ marks QCD BKG rej at signal eff. = 60%

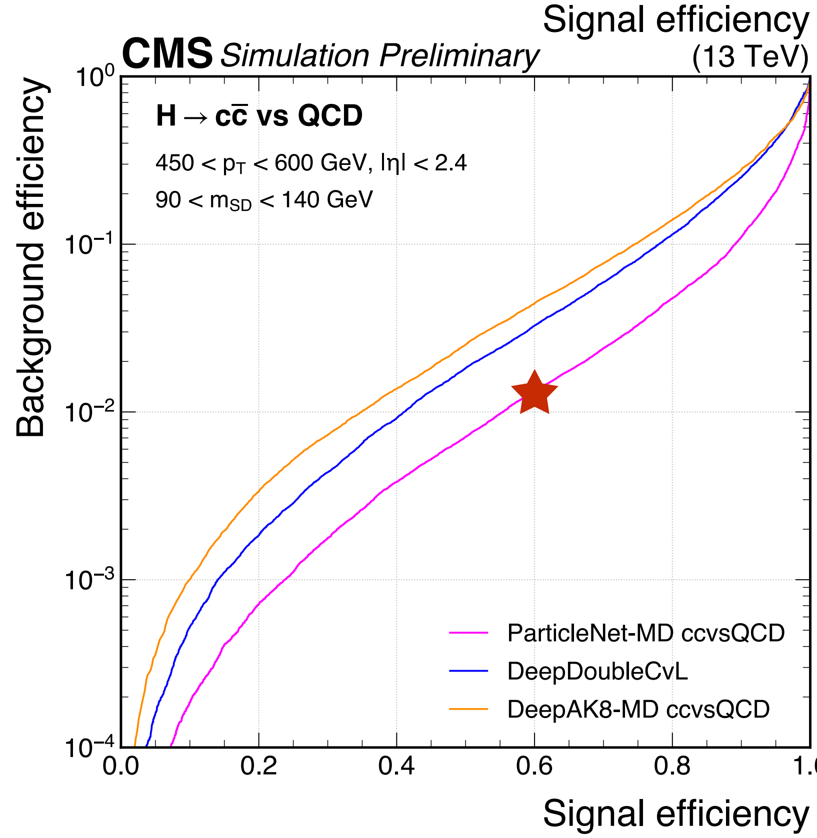
ATLAS results ATL-PHYS-PUB-2023-021



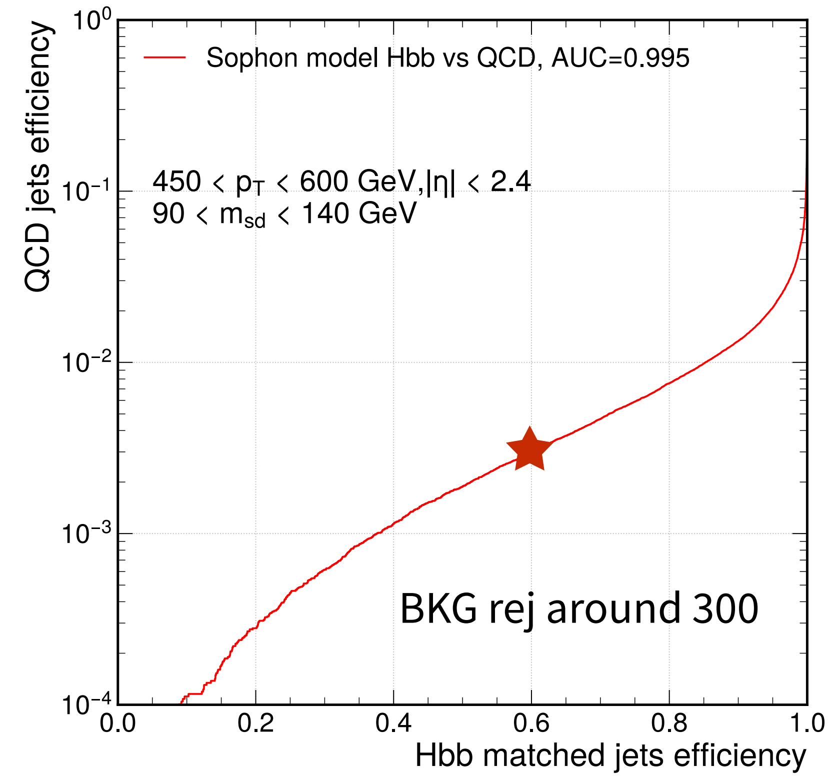


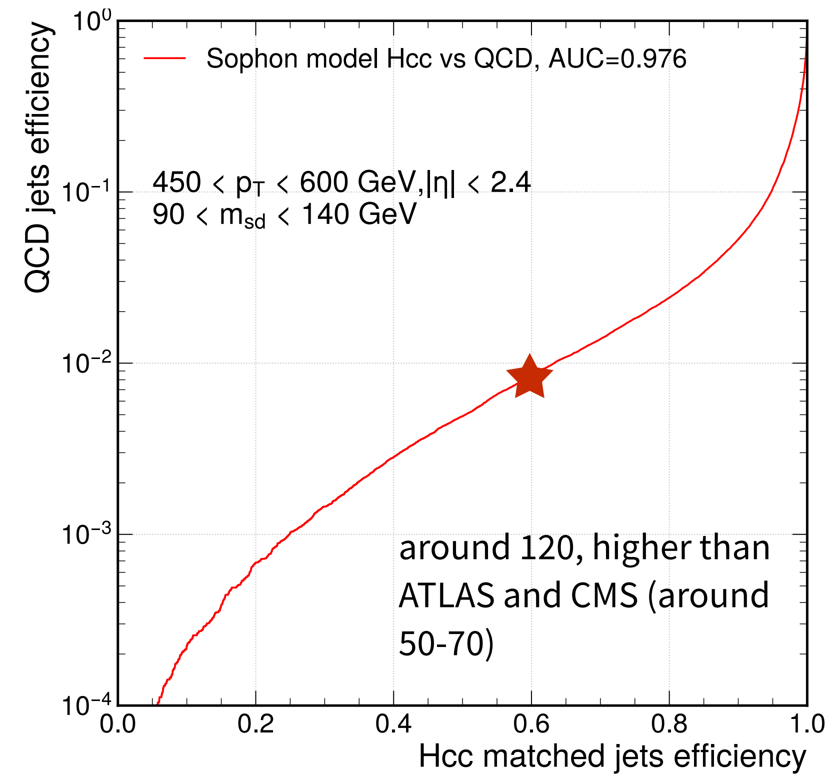
CMS results CMS-PAS-BTV-22-001





Sophon results (performance on Delphes)

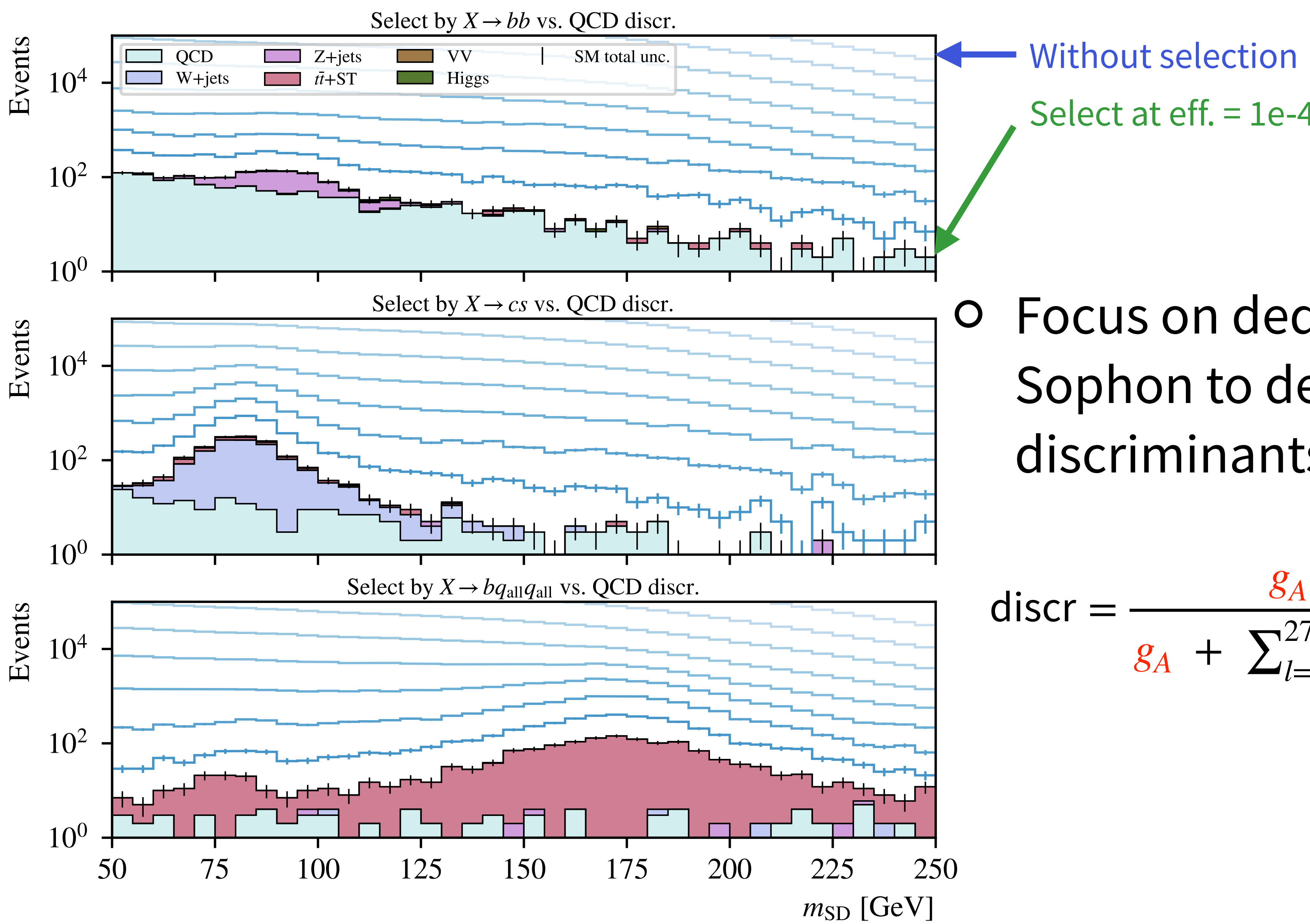




Use cases: rediscover the SM particles?

arXiv:2405.12972

➔ Use Sophon to rediscover **Z/W/t particles**

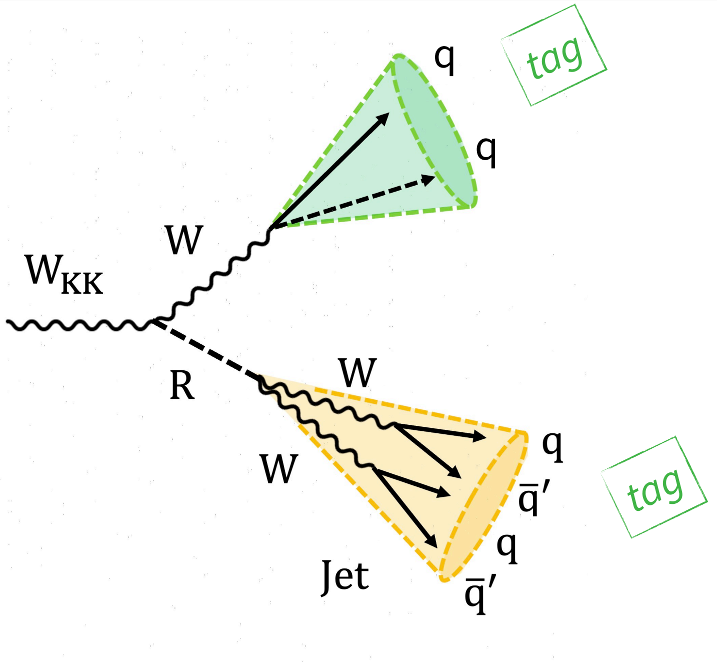


- Focus on dedicated final states, and use Sophon to define corresponding selection discriminants

$$\text{discr} = \frac{g_A}{g_A + \sum_{l=1}^{27} g_{\text{QCD}_l}} \quad \begin{cases} ①: A = \{bb\} \\ ②: A = \{cs\} \\ ③: A = \{ccb, ssb, qqb, bcs, bcq, bsq\} \end{cases}$$

Use cases: discover a BSM resonance

arXiv:2405.12972



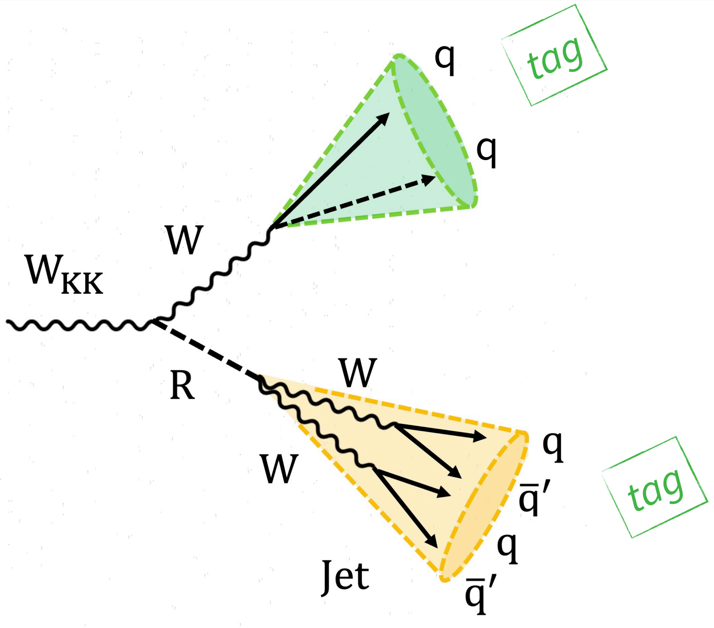
- Consider triboson signal:
  W' ($m_{W'}$ = 3 TeV) → $W\phi$ (m_ϕ = 400 GeV) → WWW → ***6q***
- Optimize an event-level discriminant

$$\text{discr} = \sum_{\text{jet}=1,2} \frac{g_{A,\text{jet}}}{g_{A,\text{jet}} + \sum_{l=1}^{27} g_{\text{QCD}_l,\text{jet}}} \quad \text{(sum for jets 1, 2)}$$

$$A = \begin{cases} 0.3 \times \{cs, qq\} \\ + 0.1 \times \{ccss, qqcs, qqqq\} \\ + 0.6 \times \{ccs, ccq, ssc, ssq, qqc, qqs, qqq\} \end{cases}$$

Use cases: discover a BSM resonance

arXiv:2405.12972



- ➔ Consider triboson signal:
  W' ($m_{W'}$ = 3 TeV) → *Wϕ* (m_ϕ = 400 GeV) → *WWW* → ***6q***
- ➔ Optimize an event-level discriminant

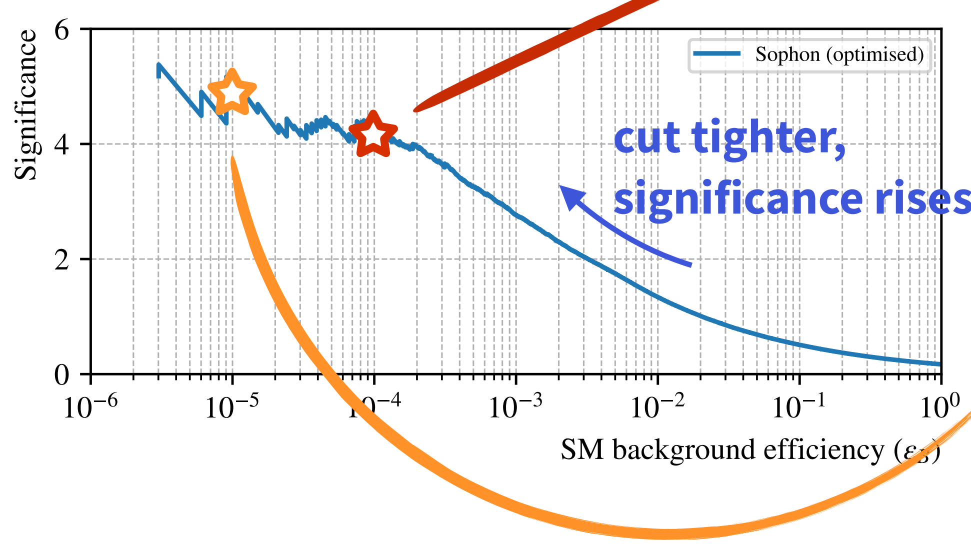
$$\text{discr} = \sum_{\text{jet}=1,2} \frac{g_{A,\text{jet}}}{g_{A,\text{jet}} + \sum_{l=1}^{27} g_{\text{QCD}_l,\text{jet}}} \quad \text{(sum for jets 1, 2)}$$

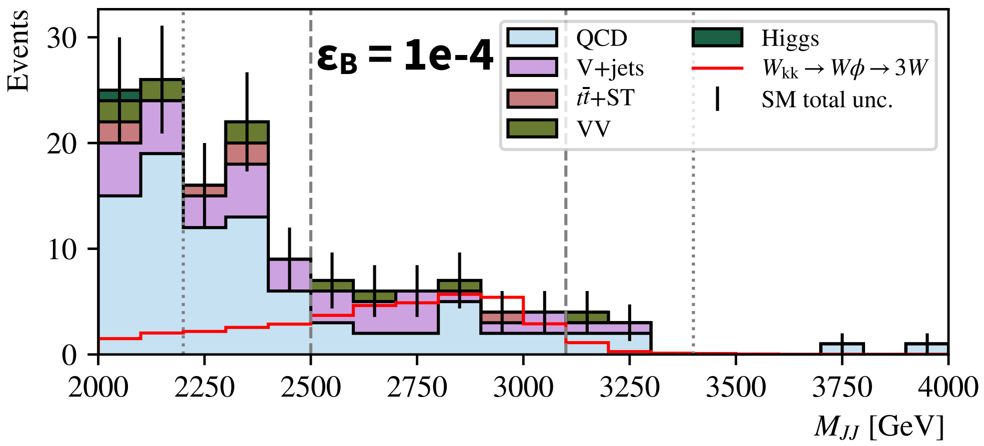
$$A = \begin{cases} 0.3 \times \{cs, qq\} \\ + 0.1 \times \{ccss, qqcs, qqqq\} \\ + 0.6 \times \{ccs, ccq, ssc, ssq, qqc, qqs, qqq\} \end{cases}$$

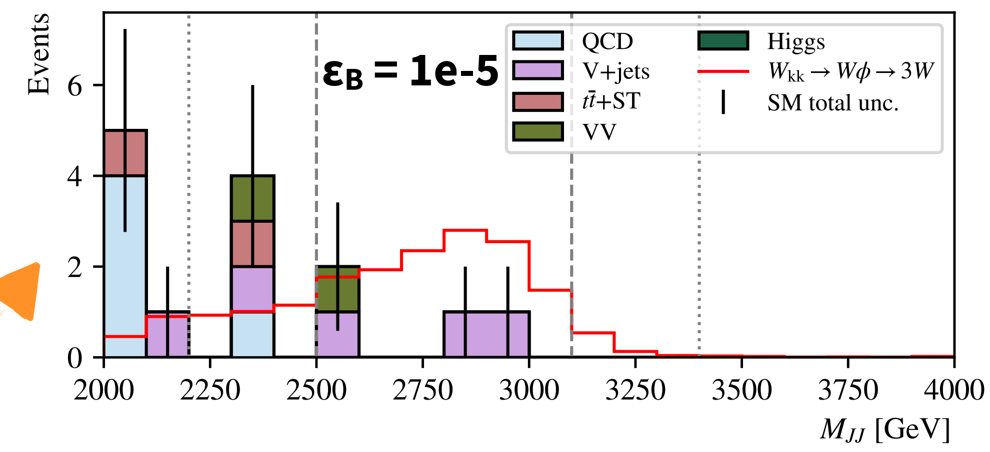
Search significance

$$Z = \sqrt{2((s+b)\log(1+s/b) - s)}$$

in dijet inv. mass window **2500–3100 GeV**





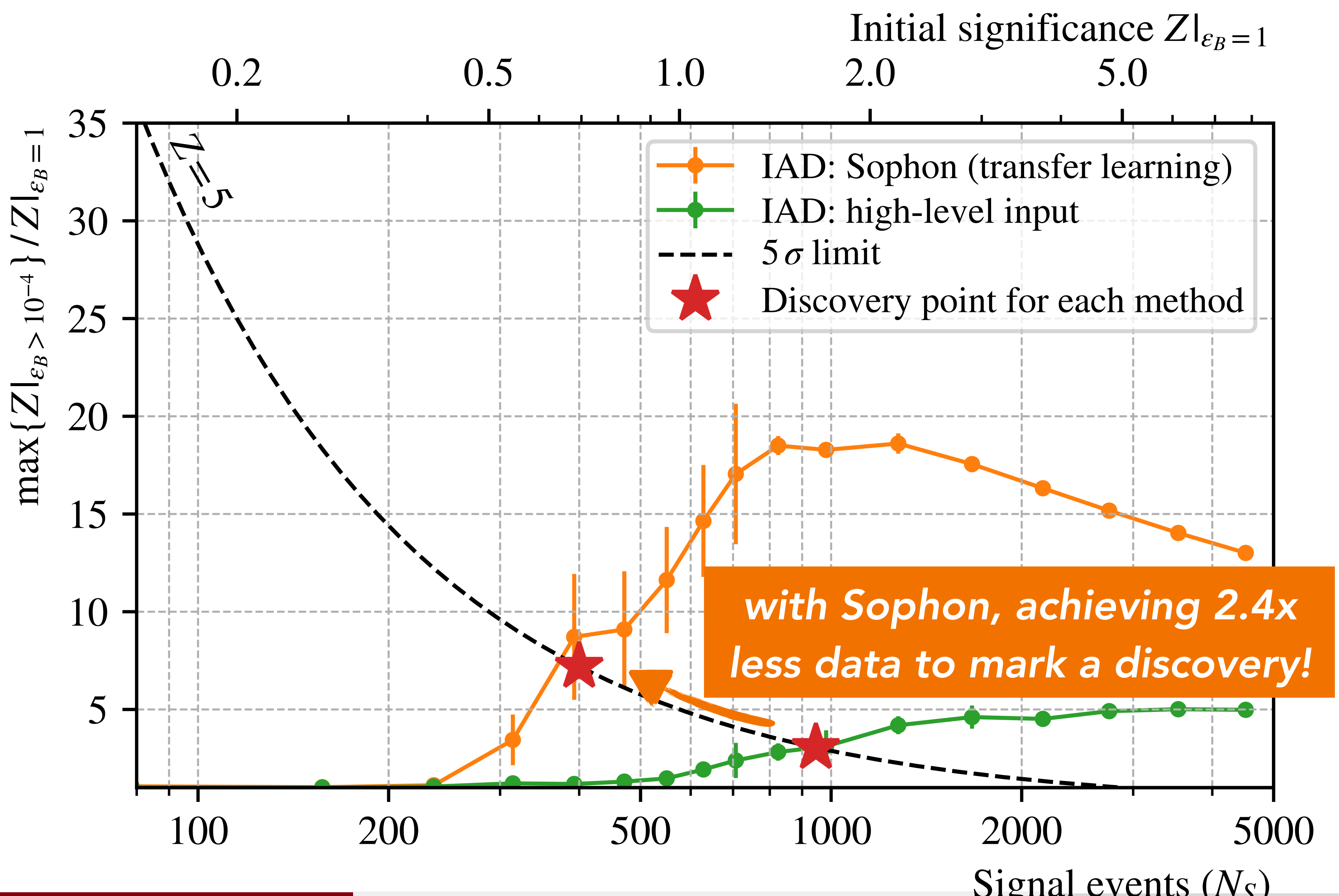


Use cases: anomaly detection

arXiv:2405.12972

→ Background:

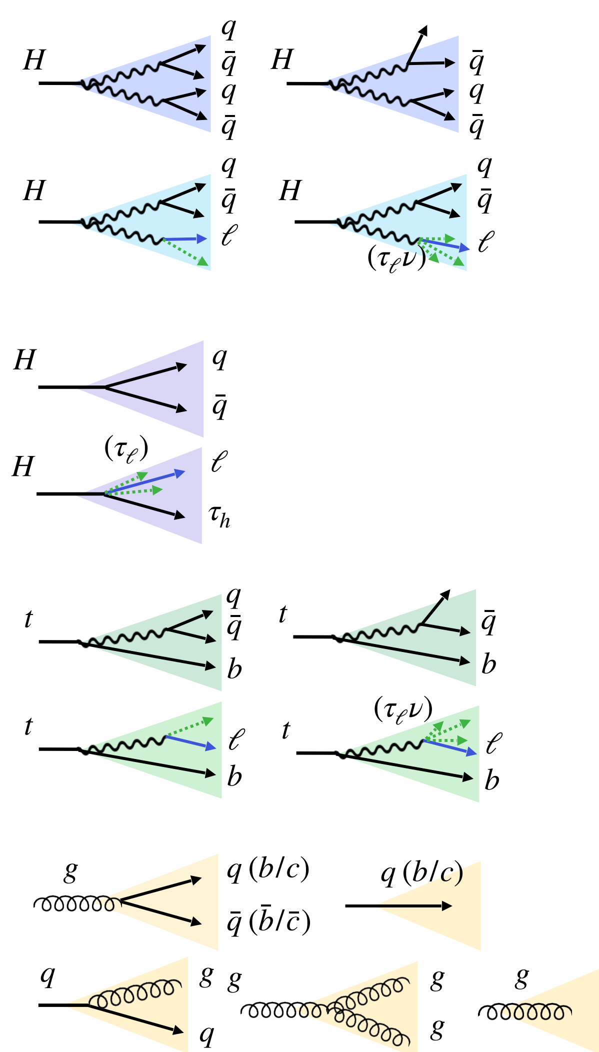
- anomaly detection (via a weakly supervised approach) is a novel experimental technique to explore peculiar resonance structures directly from data
- refer to ATLAS and CMS's established methods & results: PRL 125, 131801 (2020) CMS-PAS-EXO-22-026
- **not working if there are too few signals in data!**



CMS version of “Sophon” is now available!

In CMS, we developed “Global Particle Transformer” (GloParT)
— an effort initiated in mid-2022

Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	$e\nu qq$	0c/1c	2
	$\mu\nu qq$		2
	$\tau_e\nu qq$		2
	$\tau_\mu\nu qq$		2
	$\tau_h\nu qq$		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	$\tau_e\tau_h$		1
	$\tau_\mu\tau_h$		1
	$\tau_h\tau_h$		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	$be\nu$	1b	1
	$b\mu\nu$		1
	$b\tau_e\nu$		1
	$b\tau_\mu\nu$		1
	$b\tau_h\nu$		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1



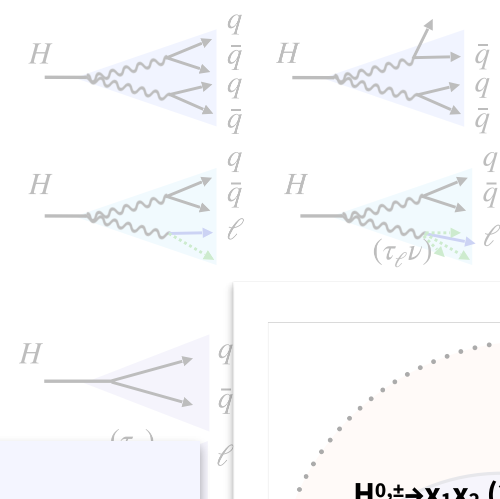
The 1st version has been released:
a fatjet tagger for **37-category classification**

CMS-PAS-HIG-23-012

CMS version of "Sophon" is now available!

In CMS, we developed "Global Particle Transformer" (GloParT)
— an effort initiated in mid-2022

Process	Final state/ prongness	heavy flavour	# of classes
H→VV (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	eνqq	0c/1c	2
	μνqq		2
	τ_eνqq		2
	τ_μνqq		2
	τ_hνqq		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1

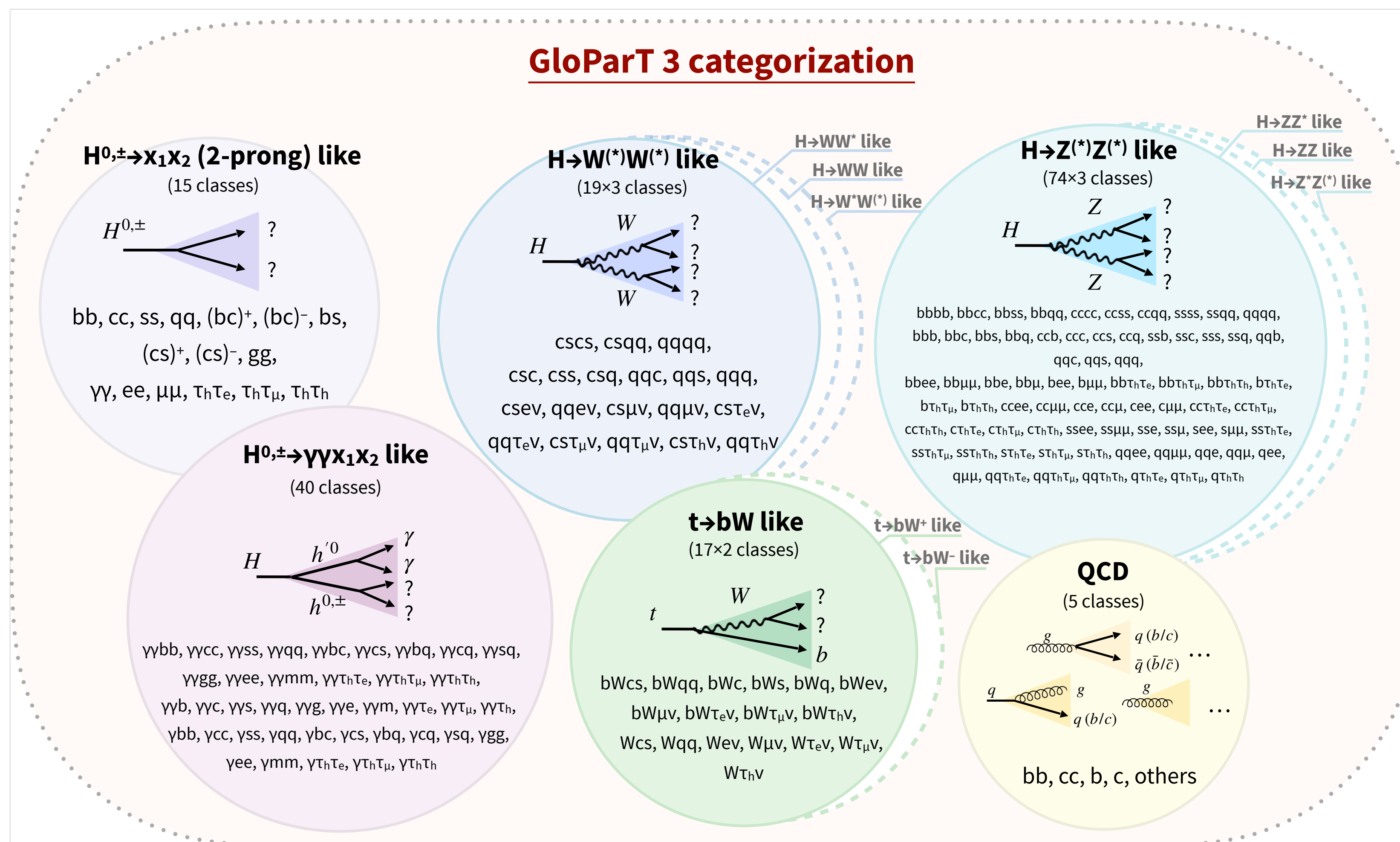


The 1st version has been released:
a fatjet tagger for **37-category classification**

CMS-PAS-HIG-23-012

Now we've finalized the 3rd version of the Global Particle Transformer **(GloParT-3)** and successfully integrated it into the CMS software

- ***750 nodes in total***
- ***expected to benefit all relevant Run 2 & Run 3 analyses through 2030***



Summary: Sophon, GloParT…

- ➔ Sophon: using Sophon on Delphes LHC dataset to (re)study the LHC potential
  - ❖ extensive BSM programs to explore with the cutting-edge tool
  - ❖ we welcome future collaborations with theorists and phenomenologists to explore novel potentials at the LHC

- ➔ GloParT's opportunity to CMS: expanding the resonance search program with GloParT!
  - ❖ $H/X \to bb/cc/\tau\tau/WW$… diverse channels for search
  - ❖ broadening conventional searches with GloParT's fine-tuning capabilities for **less-explored channels**
  - ❖ anomaly detection: the "data-only searches"

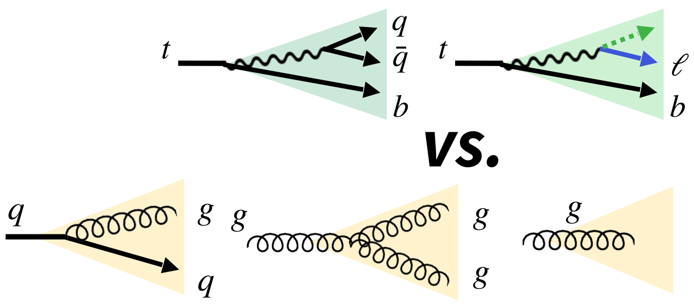
- ➔ A bright journey ahead!

Backup

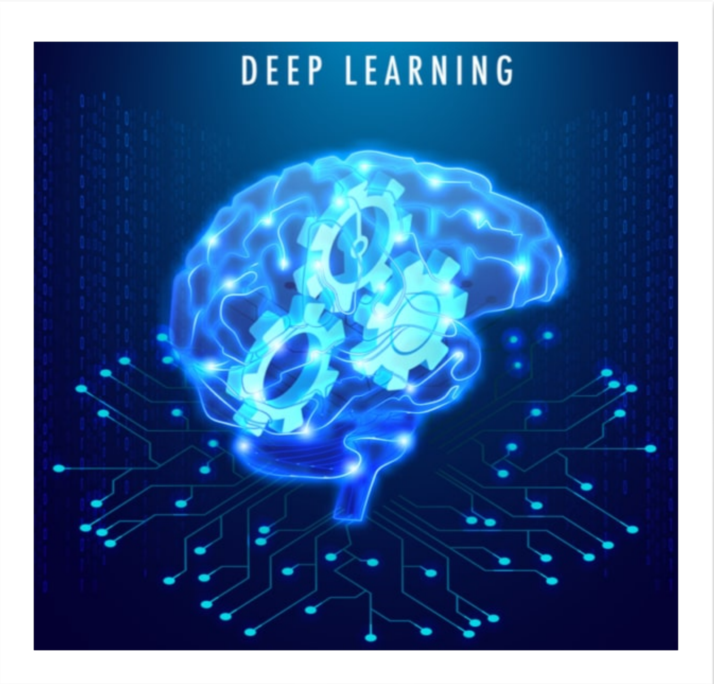
Propose "Large model for large-scale classification"

View from jet tagging

- ➔ Instead of training dedicated jet taggers, we consider **multi-class classification** with N(class) reaches o(100)
  - ❖ statistical insights: an ideal multi-class classifier is a stack of ideal binary classifiers
- ➔ The model should be **large** → carry enough capacity
- ➔ The classes should be comprehensive → **tagging ability can be further generalized by fine-tuning**

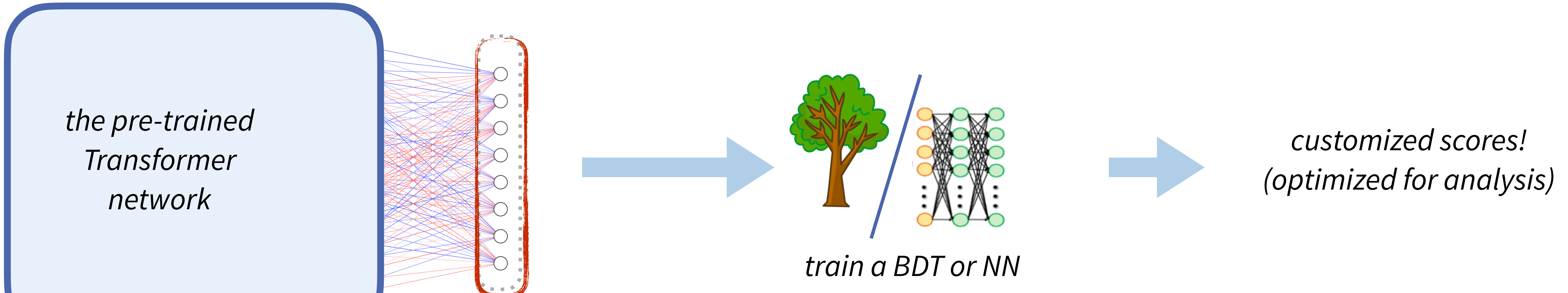


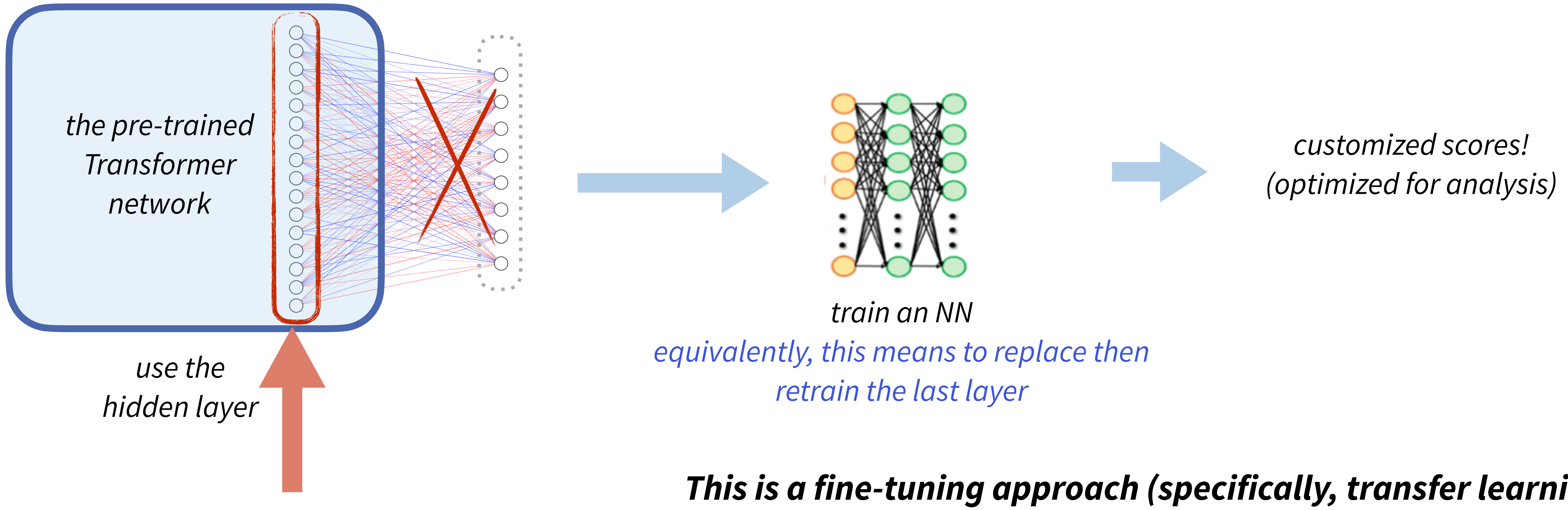
View from a pre-training solution



- ➔ Based on a comprehensive jet dataset, we hope to pre-train **a base model** to facilitate **all** LHC analyses exploring the large-R jet
- ➔ Set the training task: let the model learn to connect **"what a jet is like"** to **"which truth signature the jet reveals"** (= jet label in our case)
  - ❖ "jet labels" are simple signatures to explore → pre-training it as a classifier is just a starting point in this sense!

A glance into fine-tuning spirits



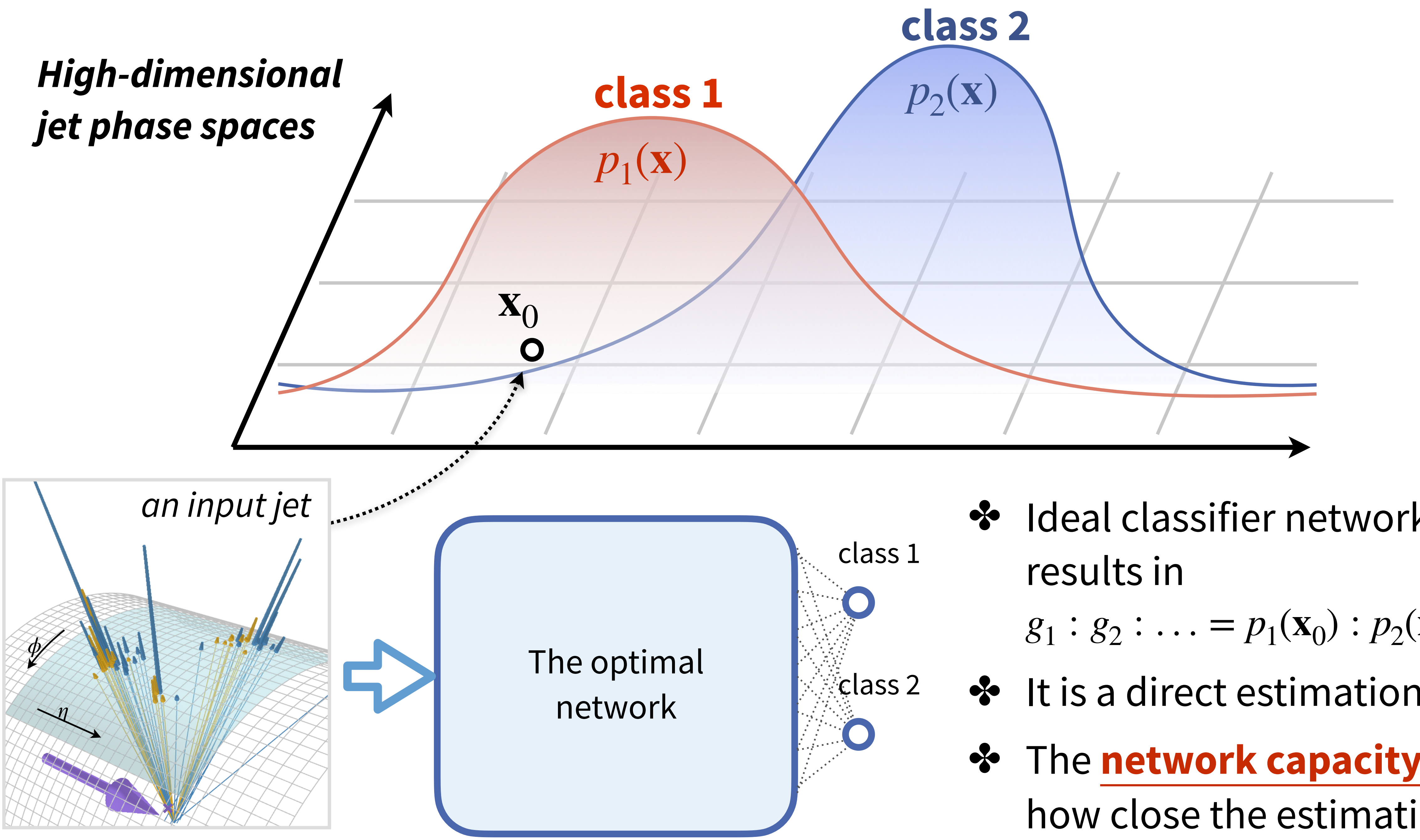


This is a fine-tuning approach (specifically, transfer learning) in its equivalent form

Statistical essence of jet tagging problem

→ **Question: where is the limit of jet tagging?**

→ Answer: the probability density ratio of two classes provides the optimal tagging



- Ideal classifier network results in $g_1 : g_2 : \ldots = p_1(\mathbf{x}_0) : p_2(\mathbf{x}_0) : \ldots$
- It is a direct estimation of p
- The **network capacity** decides how close the estimation is

Statistical property of multi-class classifier

➔ Statistical theory shows that:

A **multi-class** classifier with minimum **cross-entropy loss** **estimates the probability ratios** on the input classes:

$$g_i(\mathbf{x}) = \frac{p(\text{class} = i \,|\, \mathbf{x})}{\sum_{j=1}^{N_{\text{out}}} p(\text{class} = j \,|\, \mathbf{x})}$$

→ hence it contains all the information the ideal $N(N-1)$ binary classifiers can do

Statistical property of multi-class classifier

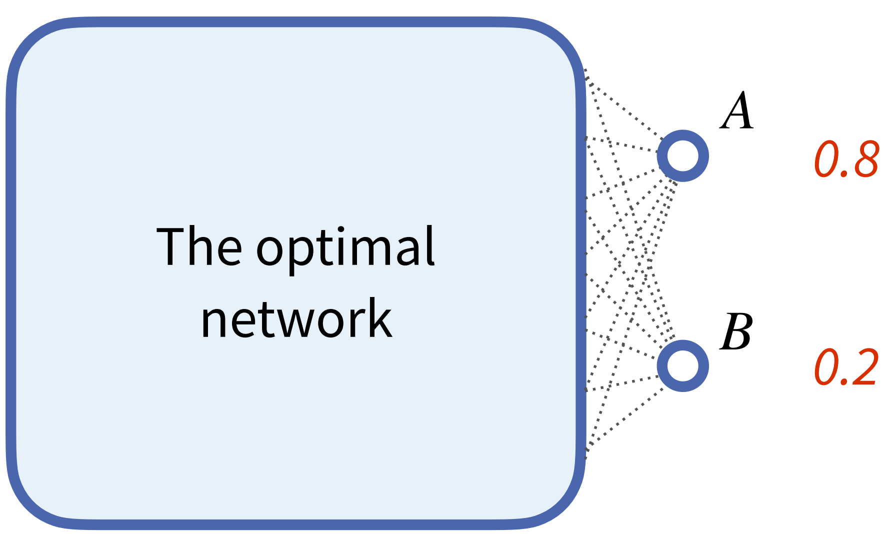
→ Statistical theory shows that:

A **multi-class** classifier with minimum **cross-entropy loss** estimates the probability ratios on the input classes:

$$g_i(\mathbf{x}) = \frac{p(\text{class} = i\,|\,\mathbf{x})}{\sum_{j=1}^{N_{\text{out}}} p(\text{class} = j\,|\,\mathbf{x})}$$

hence it contains all the information the ideal $N(N-1)$ binary classifiers can do

Two properties:



splitting class A

- A_1 0.55
- A_2 0.25
- B 0.2

$p_A = p_{A_1} + p_{A_2}$ remains the same

adding class C

- A 0.6
- B 0.15
- C 0.25

p_A / p_B remains the same

Statistical property of multi-class classifier

→ Statistical theory shows that:

A **multi-class** classifier with minimum **cross-entropy loss** estimates the probability ratios on the input classes:

$$g_i(\mathbf{x}) = \frac{p(\text{class} = i \,|\, \mathbf{x})}{\sum_{j=1}^{N_{\text{out}}} p(\text{class} = j \,|\, \mathbf{x})}$$

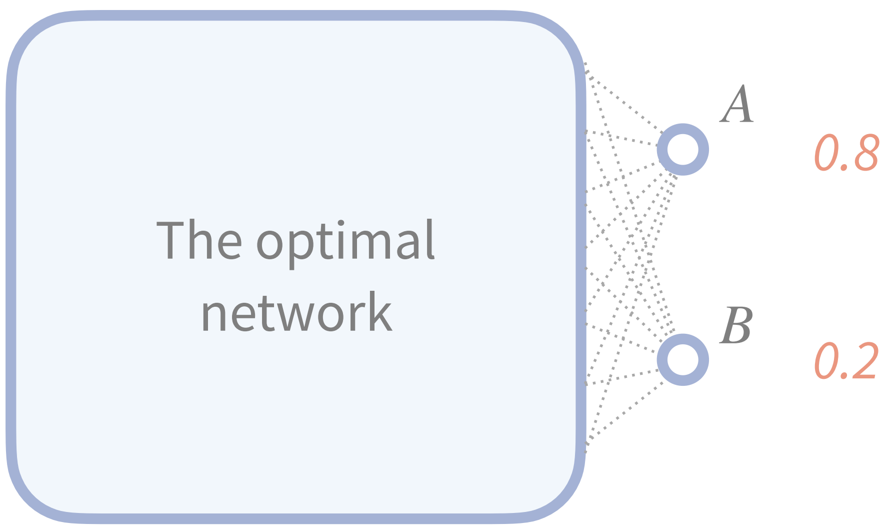
hence it contains all the information the ideal $N(N-1)$ binary classifiers can do

The key question in this context

Does the model's capacity still enable us to reach the best achievable performance in existing tasks?

Our result will show: Yes.

Two properties:



splitting class A

- A_1 0.55
- A_2 0.25
- B 0.2

⇒ $p_A = p_{A_1} + p_{A_2}$ remains the same

adding class C

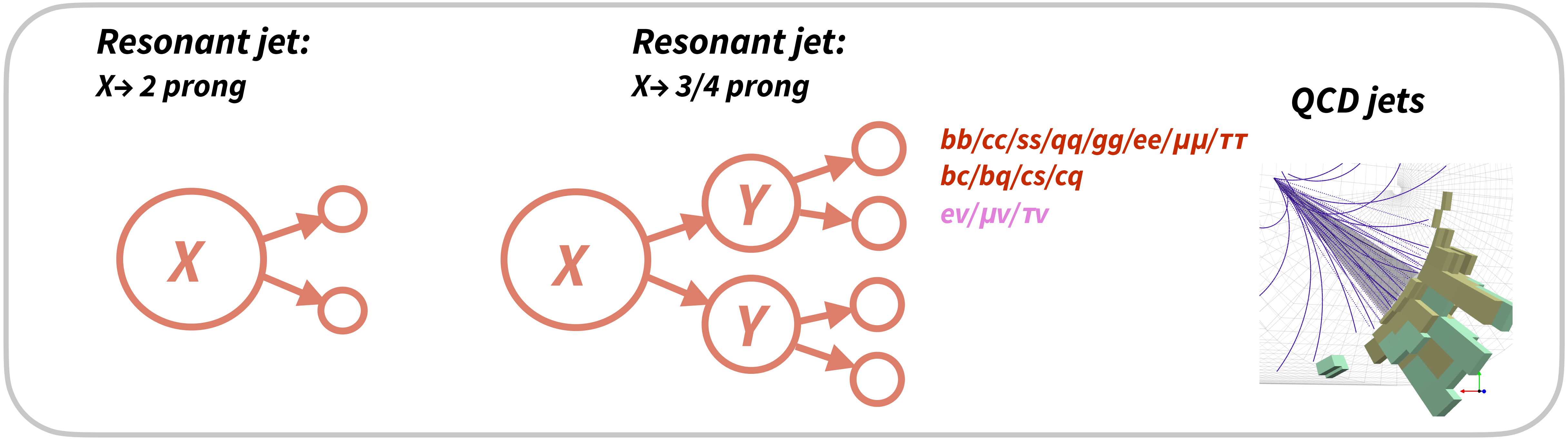
- A 0.6
- B 0.15
- C 0.25

⇒ p_A / p_B remains the same

Introducing Sophon

arXiv:2405.12972
https://github.com/jet-universe/sophon

- ➔ **Signature-Oriented Pre-training for Heavy-resonant ObservatioN**
- ➔ the model is based on Particle Transformer (**ParT)** architecture
- ➔ a pre-trained model on a newly developed comprehensive dataset: **JetClass-II**
  - ***finely categorized labels:***



contributed final states: ***bb/cc/ss/qq/gg/ee/μμ/ττ bc/bq/cs/cq*** *all combination of Y decays, resulting to 4-prong or 3-prong*

Key property: we do not focus on any specific *X* and *Y* masses
Their masses are variables: ranges from 20-500 GeV

Introducing Sophon

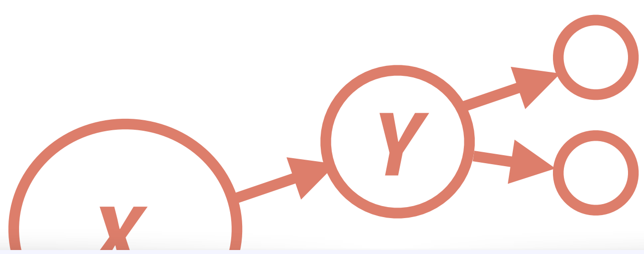
arXiv:2405.12972

https://github.com/jet-universe/sophon

- ➔ **Signature-Oriented Pre-training for Heavy-resonant ObservatioN**
- ➔ the model is based on Particle Transformer (**ParT)** architecture
- ➔ a pre-trained model on a newly developed comprehensive dataset: **JetClass-II**
  - ***finely categorized labels:***

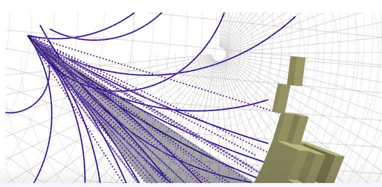
Resonant jet:
X→ 2 prong

Resonant jet:
X→ 3/4 prong



bb/cc/ss/qq/gg/ee/μμ/ττ
bc/bq/cs/cq
eν/μν/τν

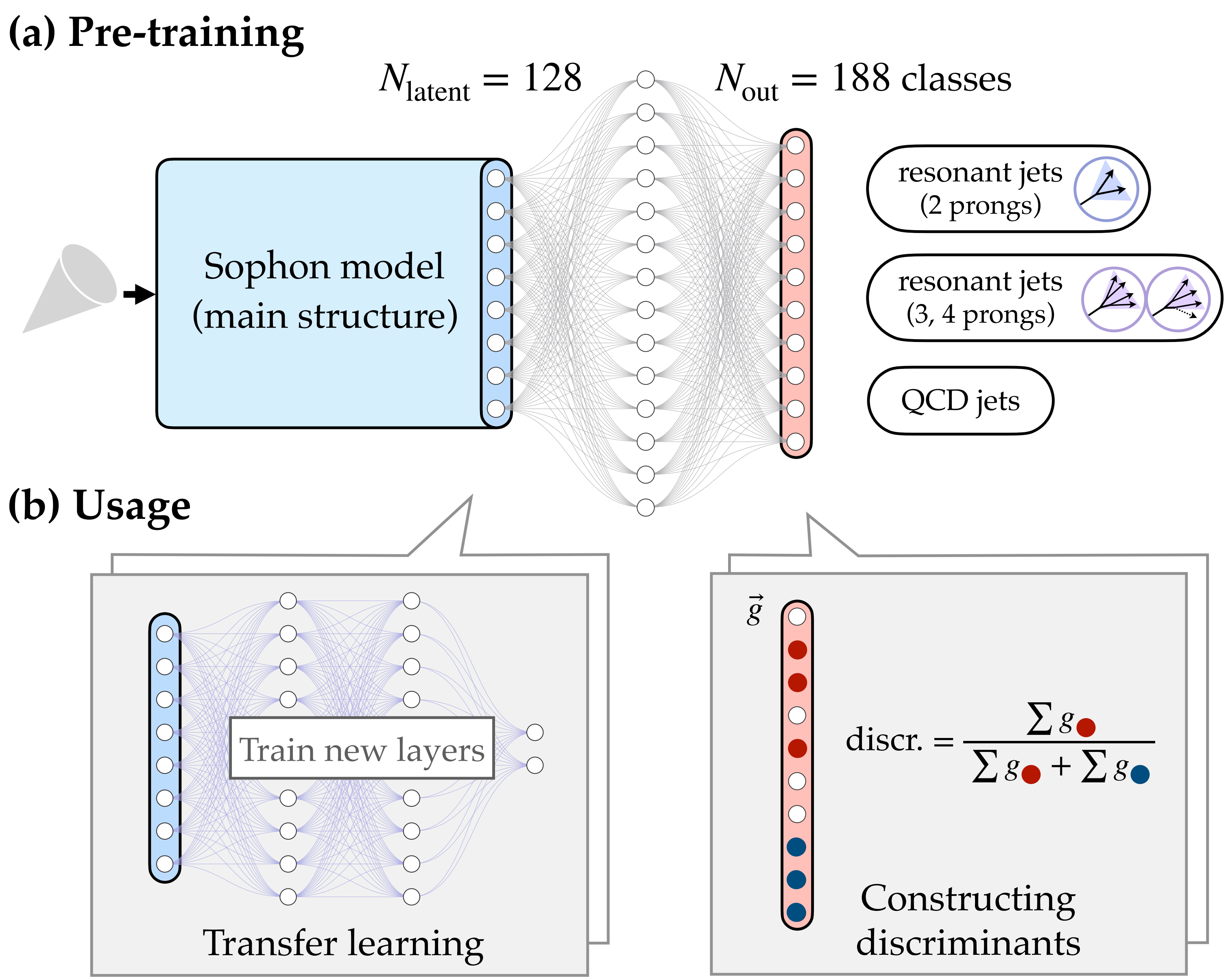
QCD jets



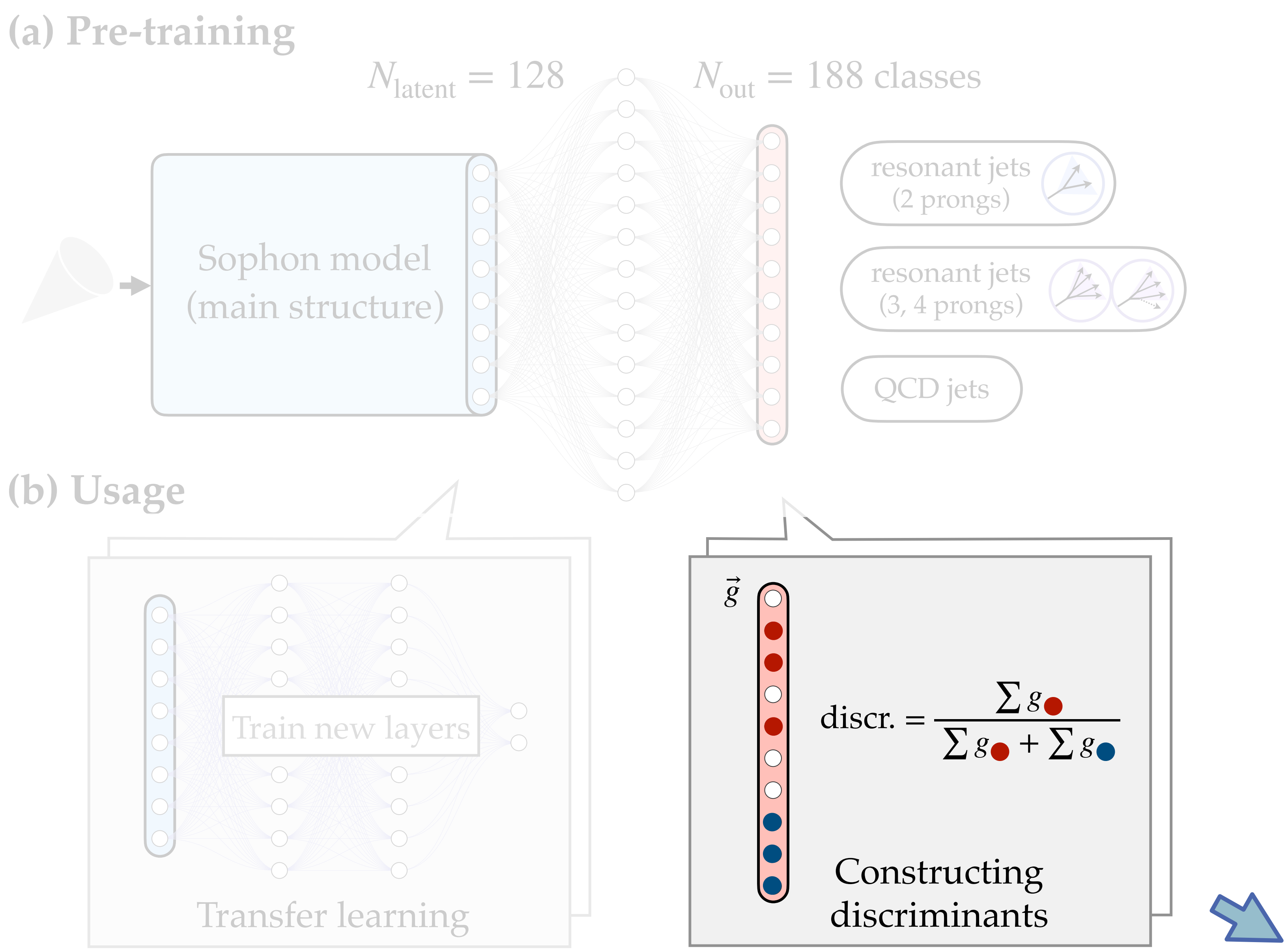
All final states!

Major types	Index range	Label names
Resonant jets: $X \to 2$ prong	0–14	bb, cc, ss, qq, bc, cs, bq, cq, sq, gg, ee, $\mu\mu$, $\tau_h\tau_e$, $\tau_h\tau_\mu$, $\tau_h\tau_h$
Resonant jets: $X \to 3$ or 4 prong	15–160	$bbbb$, $bbcc$, $bbss$, $bbqq$, $bbgg$, $bbee$, $bb\mu\mu$, $bb\tau_h\tau_e$, $bb\tau_h\tau_\mu$, $bb\tau_h\tau_h$, bbb, bbc, bbs, bbq, bbg, bbe, $bb\mu$, $cccc$, $ccss$, $ccqq$, $ccgg$, $ccee$, $cc\mu\mu$, $cc\tau_h\tau_e$, $cc\tau_h\tau_\mu$, $cc\tau_h\tau_h$, ccb, ccc, ccs, ccq, ccg, cce, $cc\mu$, $ssss$, $ssqq$, $ssgg$, $ssee$, $ss\mu\mu$, $ss\tau_h\tau_e$, $ss\tau_h\tau_\mu$, $ss\tau_h\tau_h$, ssb, ssc, sss, ssq, ssg, sse, $ss\mu$, $qqqq$, $qqgg$, $qqee$, $qq\mu\mu$, $qq\tau_h\tau_e$, $qq\tau_h\tau_\mu$, $qq\tau_h\tau_h$, qqb, qqc, qqs, qqq, qqg, qqe, $qq\mu$, $gggg$, $ggee$, $gg\mu\mu$, $gg\tau_h\tau_e$, $gg\tau_h\tau_\mu$, $gg\tau_h\tau_h$, ggb, ggc, ggs, ggq, ggg, gge, $gg\mu$, bee, cee, see, qee, gee, $b\mu\mu$, $c\mu\mu$, $s\mu\mu$, $q\mu\mu$, $g\mu\mu$, $b\tau_h\tau_e$, $c\tau_h\tau_e$, $s\tau_h\tau_e$, $q\tau_h\tau_e$, $g\tau_h\tau_e$, $b\tau_h\tau_\mu$, $c\tau_h\tau_\mu$, $s\tau_h\tau_\mu$, $q\tau_h\tau_\mu$, $g\tau_h\tau_\mu$, $b\tau_h\tau_h$, $c\tau_h\tau_h$, $s\tau_h\tau_h$, $q\tau_h\tau_h$, $g\tau_h\tau_h$, $qqqb$, $qqqc$, $qqqs$, $bbcq$, $ccbs$, $ccbq$, $ccsq$, $sscq$, $qqbc$, $qqbs$, $qqcs$, $bcsq$, bcs, bcq, bsq, csq, $bce\nu$, $cse\nu$, $bqe\nu$, $cqe\nu$, $sqe\nu$, $qqe\nu$, $bc\mu\nu$, $cs\mu\nu$, $bq\mu\nu$, $cq\mu\nu$, $sq\mu\nu$, $qq\mu\nu$, $bc\tau_e\nu$, $cs\tau_e\nu$, $bq\tau_e\nu$, $cq\tau_e\nu$, $sq\tau_e\nu$, $qq\tau_e\nu$, $bc\tau_\mu\nu$, $cs\tau_\mu\nu$, $bq\tau_\mu\nu$, $cq\tau_\mu\nu$, $sq\tau_\mu\nu$, $qq\tau_\mu\nu$, $bc\tau_h\nu$, $cs\tau_h\nu$, $bq\tau_h\nu$, $cq\tau_h\nu$, $sq\tau_h\nu$, $qq\tau_h\nu$
QCD jets	161–187	$bbccss$, $bbccs$, $bbcc$, $bbcss$, $bbcs$, bbc, $bbss$, bbs, bb, $bccss$, $bccs$, bcc, $bcss$, bcs, bc, bss, bs, b, $ccss$, ccs, cc, css, cs, c, ss, s, others

Using Sophon



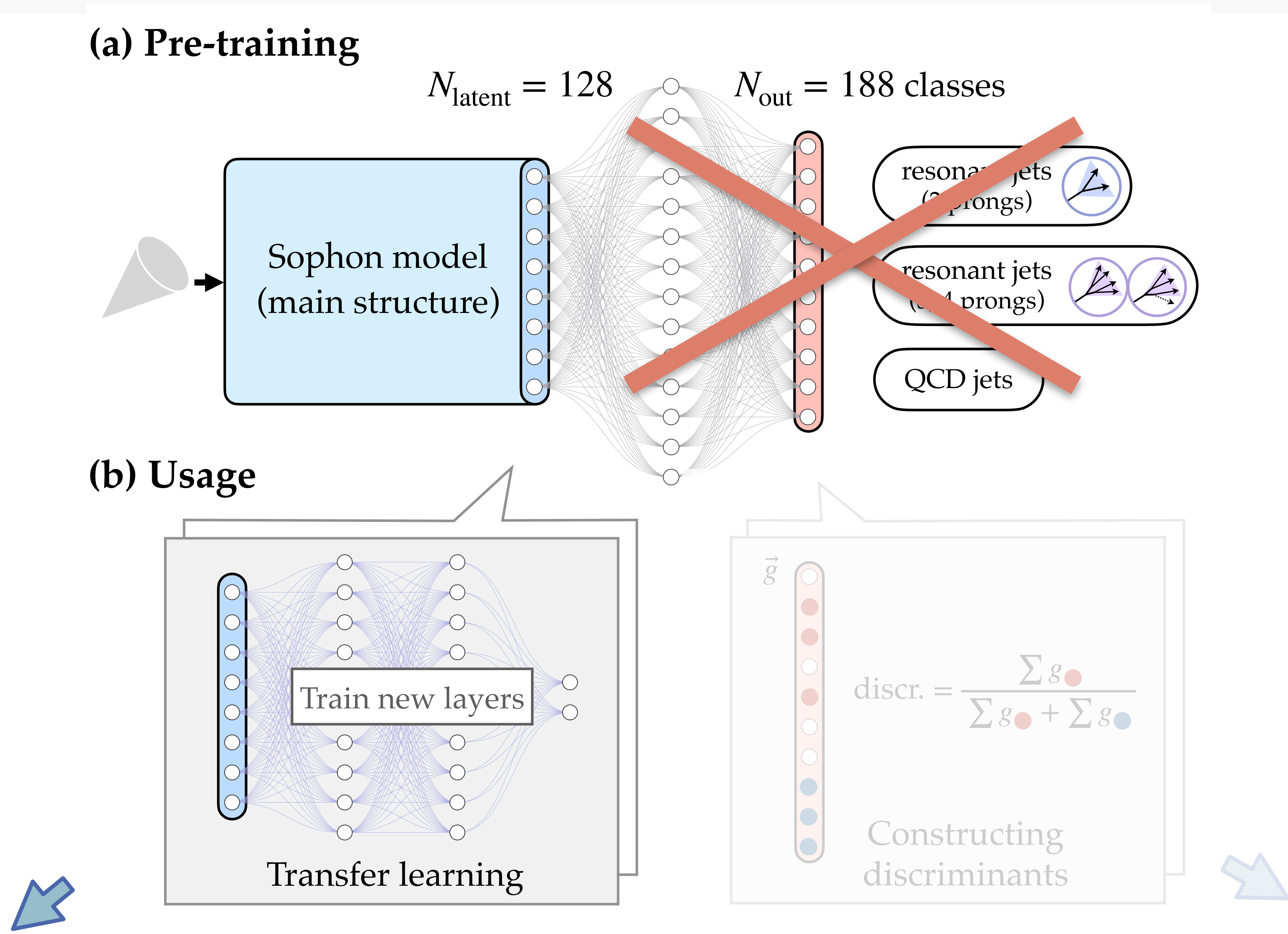
Using Sophon



Use it out of the box!

Construct a dedicated discr.
→ perform a bump hunt

Sophon's transfer learning



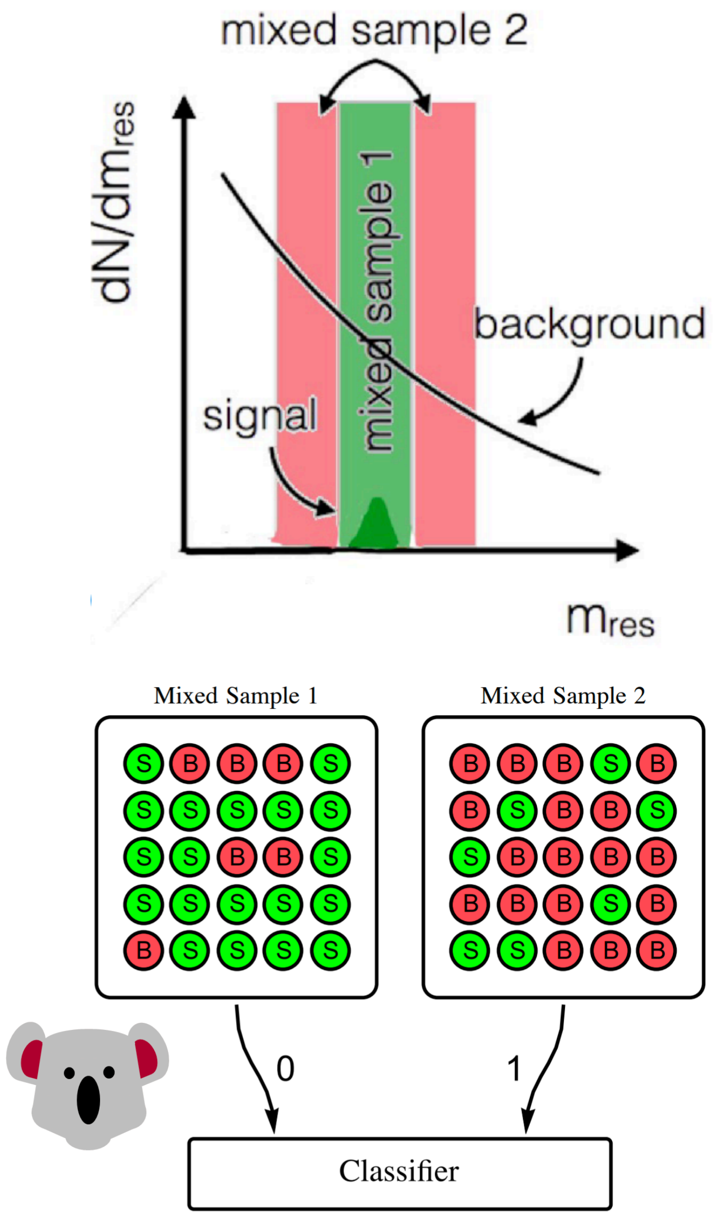
- Transfer to uncovered tagging scenarios…
- facilitate anomaly detection (weakly-supervised, autoencoder)…
- *more potential to unlock!*

Use it out of the box!

Construct a dedicated discr.
→ perform a bump hunt

Background: anomaly detection in weakly-supervised approach

JHEP 10 (2017) 174



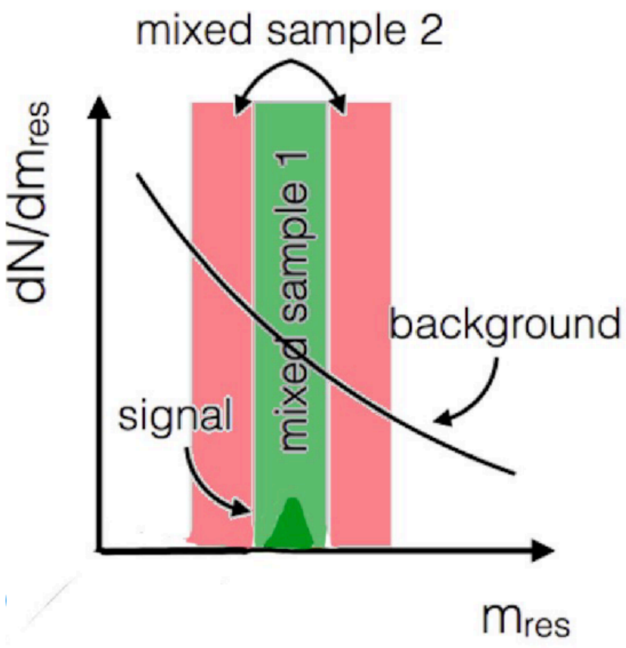
Equivalent effect for training **S** vs **B**

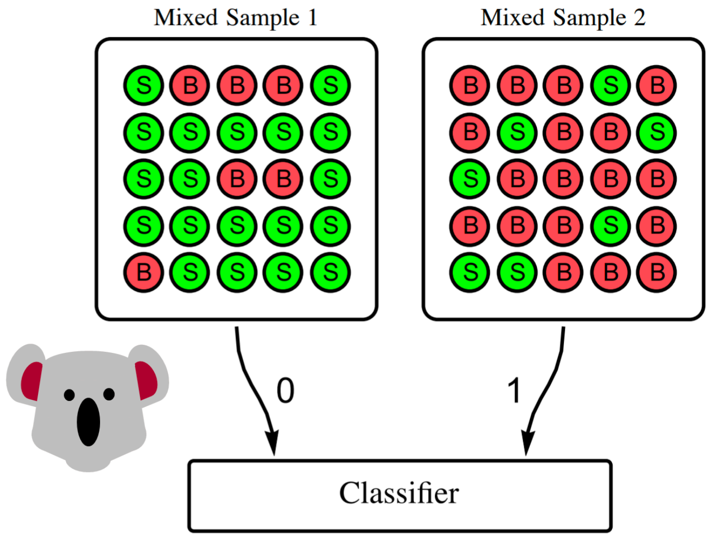
➔ Recall the early work: CWoLa (classification without labels) Hunting

- allow to detect anomalies purely from data
- train a classifier for mass window vs mass sideband (mixed sample 1 vs 2)
- many improved approaches in recent years → very active field

Background: anomaly detection in weakly-supervised approach

JHEP 10 (2017) 174

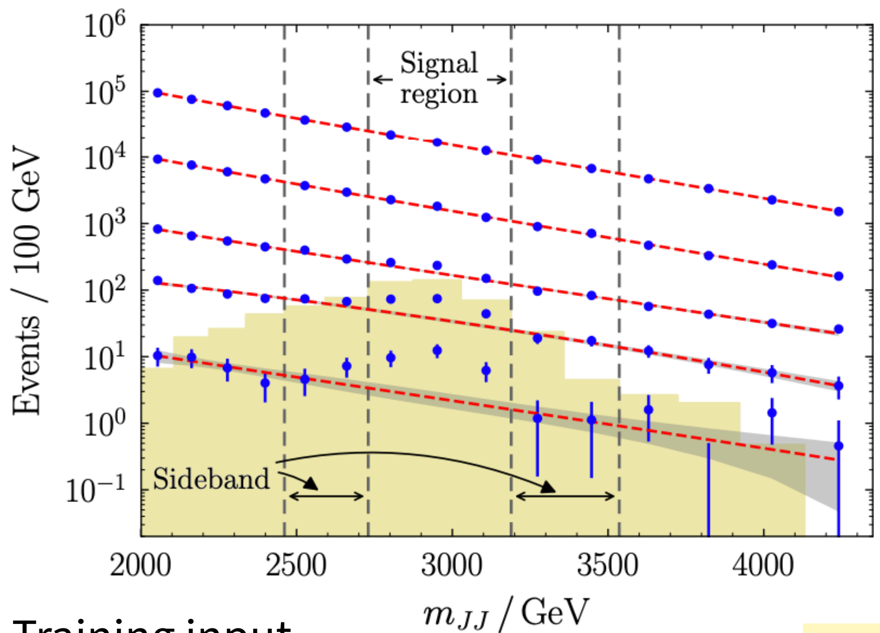




Equivalent effect for training **S** vs **B**

➔ Recall the early work: CWoLa (classification without labels) Hunting

- allow to detect anomalies purely from data
- train a classifier for mass window vs mass sideband (mixed sample 1 vs 2)
- many improved approaches in recent years → very active field

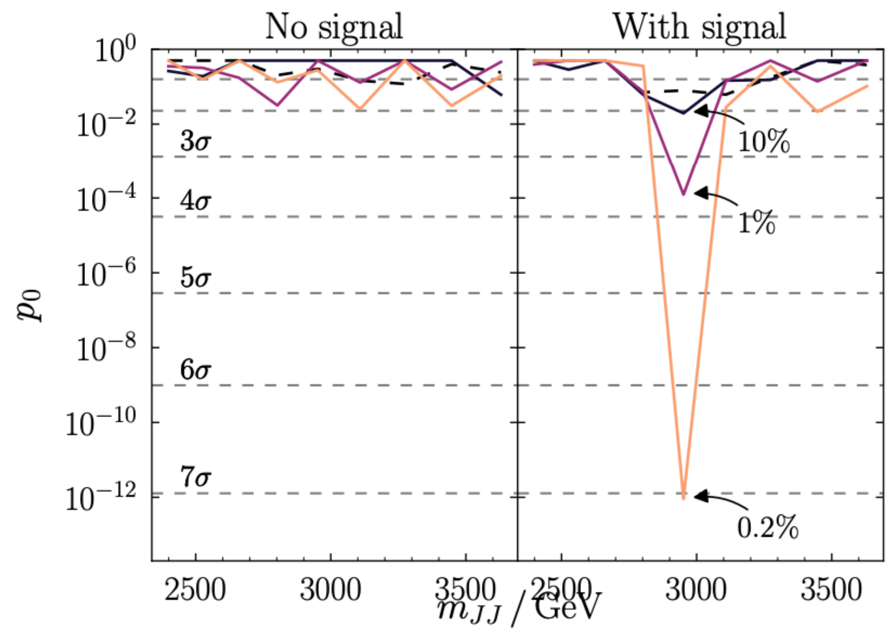


Training input

$m_J, \sqrt{\tau_1^{(2)}/\tau_1^{(1)}}, \tau_{21}, \tau_{32}, \tau_{43}, n_{\text{trk}},$

can discover $W' \to W\phi \to WWW$ signals

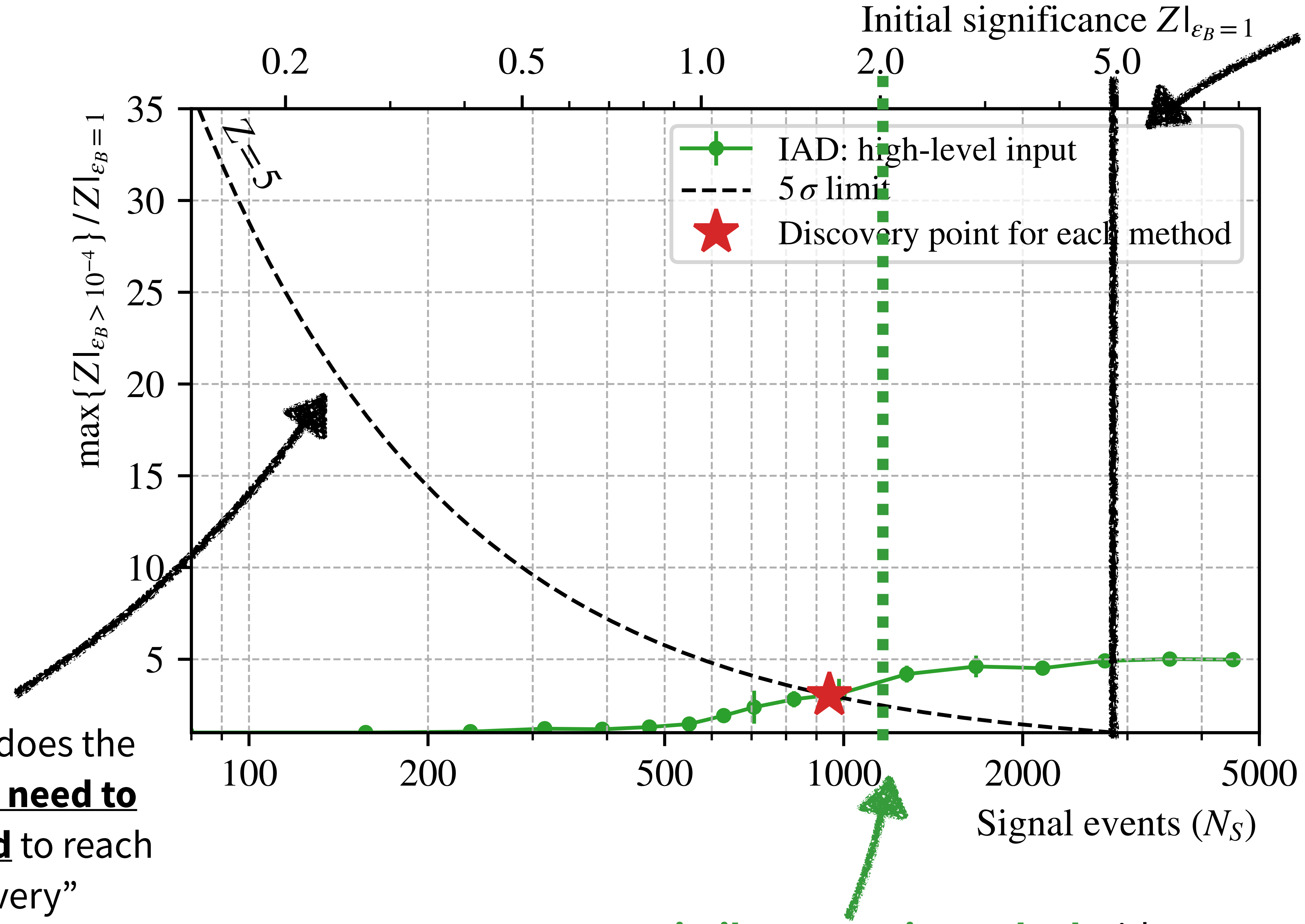
see 2σ → 7σ improvement



PRL, 121 (2018) 24, 241803

PRD, 99 (2019) 1, 014038

Dijet search capabilities

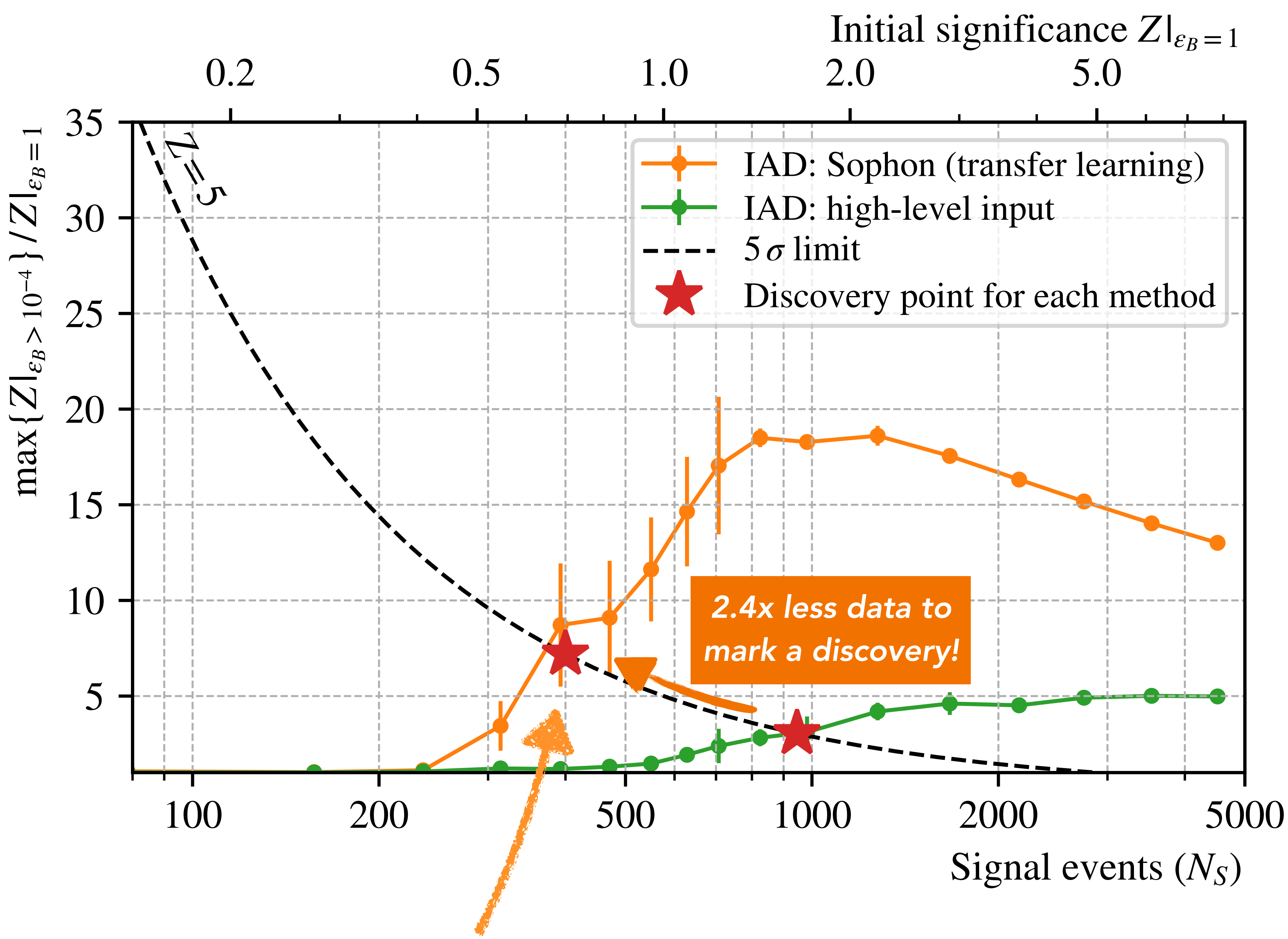


"If signal events reach this point, **with initial *Z*=5**, then we have already discovered the signal without needing to make a cut"

"How much does the **significance need to be increased** to reach the 5σ discovery"

a similar 2σ→7σ is reached with conventional AD approach; roughly reproduce the result in *PRL, 121 (2018) 24, 241803*

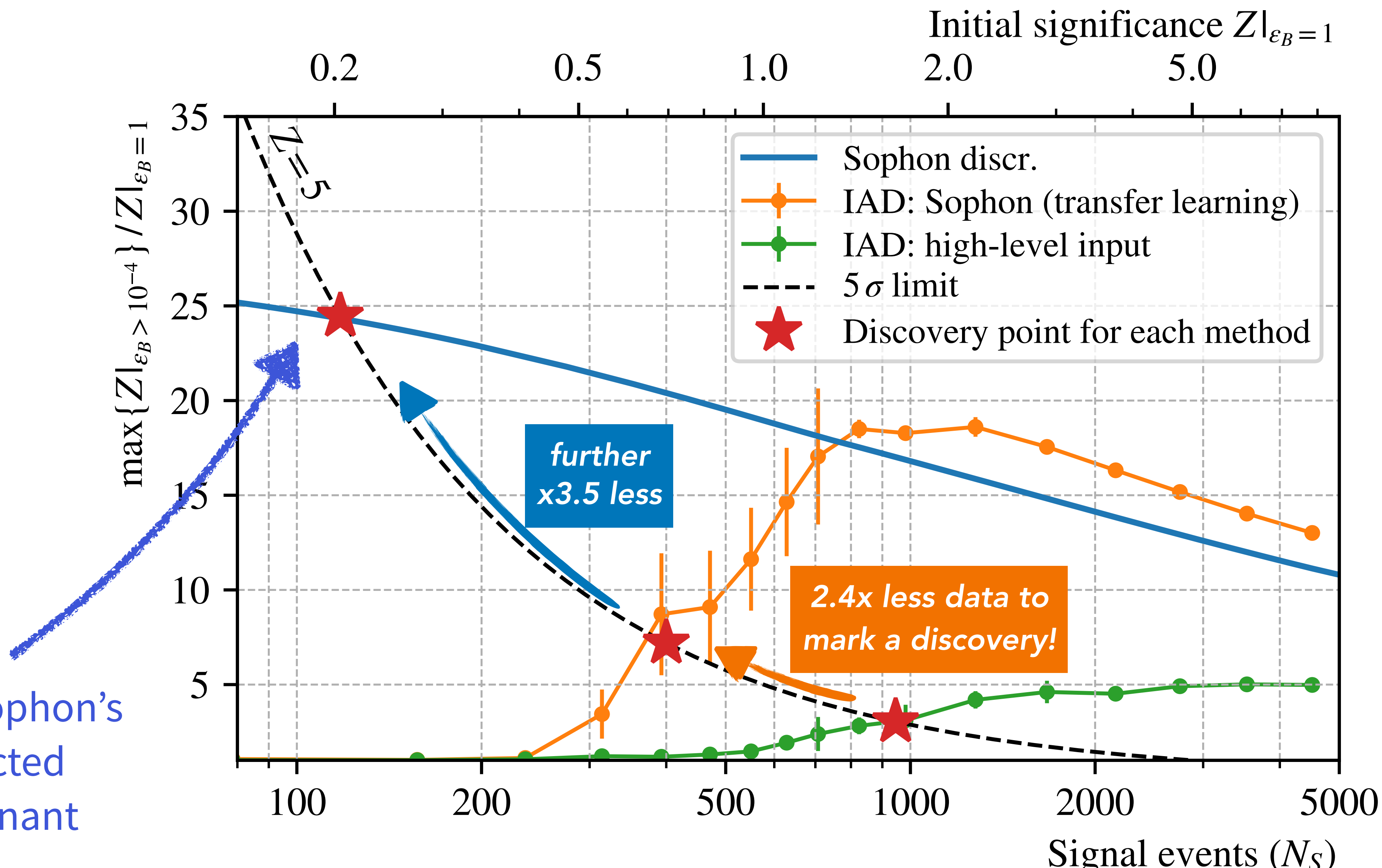
Dijet search capabilities



Combining Sophon's transfer learning (using Sophon's "knowledge") with AD marks a success

- More sensitive at low signal injection (**even starting at ~0.6σ**)
- Much improved S vs B distinguishability than using high-level input

Dijet search capabilities



$$\text{discr} = \sum_{\text{jet}=1,2} \frac{g_{A,\text{jet}}}{g_{A,\text{jet}} + \sum_{l=1}^{27} g_{\text{QCD}_l,\text{jet}}}$$

$$A = \begin{cases} 0.3 \times \{cs, qq\} \\ + 0.1 \times \{ccss, qqcs, qqqq\} \\ + 0.6 \times \{ccs, ccq, ssc, ssq, qqc, qqs, qqq\} \end{cases}$$

Implications to ATLAS/CMS experiments?

- "Sophon methodology" releases a lot of new opportunities for future LHC experiments
  - it creates a **"global large-*R* jet tagger"** → bring benefits of the advanced NN to ~all hadronic final-state searches
  - **also viewed as a pre-trained jet model**: a base model tailored for a broad range of LHC analyses
- How to use the experimental version of the Sophon model?
  - used in conventional analyses: except for some well-calibrated nodes, the major challenge will be the calibration of peculiar signals (not easy to find proxies)
  - data-only analysis: develop discriminants dedicated to different signals → cut tight on the data events → peak finding on some mass observable (single jet / di-jet / jet+lepton…)
    - **could be helpful in broadly searching for BSM resonance!**
  - anomaly detection: weakly-supervised approaches / further improvements?

JetClass-II and Sophon

arXiv:2405.12972

- Developed the **JetClass-II** dataset and the **Sophon** model
- **JetClass-II** **[Hugging Face dataset]** covers more comprehensive phase spaces and can be a good playground to develop future foundation models
  - how to use it?
    - can be used to train models for various jet-related tasks, e.g. jet classification, regression, generation or reconstruction…
    - its extensive phase space coverage and high statistics enable model developers to **focus on specific regions of interest**, or **work with the entire dataset**
  - include detailed low-level information (particle kinematics, PID and IP features), the same as JetClass
  - also include generator-level information (GEN particles within the GEN-jet, and the GEN resonances)
  - generation details can be found in this repository
- The **Sophon model** **[Hugging Face]** can be helpful to deliver future LHC pheno research
  - optimizing sensitivity for dedicated searches/anomaly detection/novel paradigms
  - performing studies on the pheno dataset/model can inspire how we do real experiments at the LHC