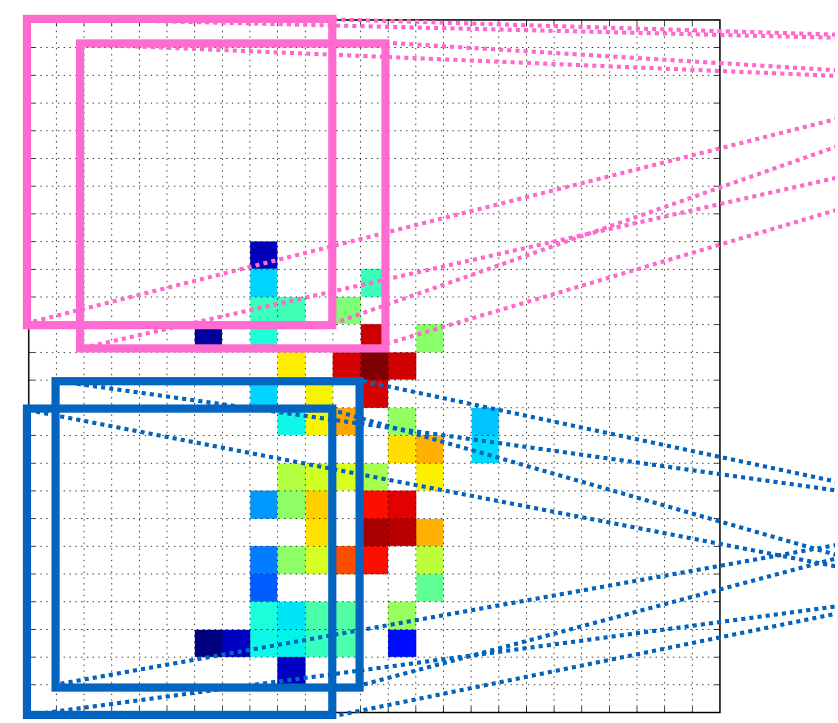
Landscape of model independent searches

Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com
bpnachman@lbl.gov
@bpnachman
bnachman



PHYSTAT-Anomalies
May 2022

Questions in particle physics

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

Questions in particle physics

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Hierarchy problem

Strong CP

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at accessible energies

Questions in particle physics

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Strong CP

Hierarchy problem

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at accessible energies

(2) Patience! (new physics is rare)

Questions in particle physics

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Strong CP

Hierarchy problem

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at accessible energies

(2) Patience! (new physics is rare)

(3) We are not looking in the right place

Questions in particle physics

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter

Strong CP

Hierarchy problem

Dark energy

Flavor puzzles

Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

(1) There is nothing new at accessible energies

(2) Patience! (new physics is rare)

(3) We are not looking in the right place

Questions in particle physics

Theoretical and **experimental** questions motivate a deep exploration **of the fundamental structure of nature**

Dark matter
Strong CP
Hierarchy problem
Dark energy
Flavor puzzles
Baryogengesis

We have performed thousands of hypothesis tests & have no significant evidence for physics beyond the Standard Model

Three possibilities

This is what keeps me up at night!

(3) We are not looking in the right place

The Path Forward

There are two complementary paths forward:

(1) Identify new, specific, well-motivated places to look

This is still an incredibly important direction and has resulted in new directions like long-lived particle searches

(2) Look in many places all at once

Focus of today's talk!

This is what keeps me up at night!

(3) We are not looking in the right place

The Path Forward

There are two **complementary** paths forward:

(1) Identify new, specific, well-motivated places to look

This is still an incredibly important direction and has resulted in new directions like long-lived particle searches

(2) Look in many places all at once

Focus of today's talk!

There is no free lunch: for any particular model, (2) will be less sensitive than (1). We need both search paradigms!

Why not just look everywhere?

Why not just look everywhere?

(a) There are a lot of places to look

$A \to BC$		B = SM									B = BSM
		e	μ	τ	q/g	b	t	γ	Z/W	H	
C = SM	e	Z'	$\not{R}$	$\not{R}$	LQ	LQ	LQ	L^*	L^*	L^*	Many
	μ		Z'	$\not{R}$	LQ	LQ	LQ	L^*	L^*	L^*	
	τ			Z'	LQ	LQ	LQ	L^*	L^*	L^*	
	q/g				Z'	W'	T'	Q^*	Q^*	Q'	
	b					Z'	W'	Q^*	Q^*	B'	
	t						Z'	Q^*	T'	T'	
	γ							H	H	Z_{KK}	
	Z/W								H	$H^\pm/A$	
	H									H	
C = BSM		Consider just the di-object search for resonant A → B C									Many

J. Kim, K. Kong, BN, D. Whiteson, *JHEP 04 (2020) 30*, 1907.06659

Why not just look everywhere?

(a) There are a lot of places to look

		e	μ	τ	q/g	b	t	γ	Z/W	H	BSM → $SM_1 \times SM_1$				BSM → $SM_1 \times SM_2$				BSM → complex			
											q/g	γ/π^0's	b	⋯	tZ/H	bH		⋯	$\tau qq'$	eqq'	$\mu qq'$	⋯
	e	[37, 38]	[39, 40]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅		∅	[43, 44]	∅	
	μ		[37, 38]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅		∅	∅	[43, 44]	
	τ			[45, 46]	∅	[47]	∅	∅	∅	∅	∅	∅	∅		∅	∅	∅		[48, 49]	∅	∅	
	q/g				[29, 30, 50, 51]	[52]	∅	[53, 54]	[55]	∅	∅	∅	∅		∅	∅	∅		∅	∅	∅	
	b					[29, 52, 56]	[57]	[54]	[58]	[59]	∅	∅	∅		[60]	∅	∅		∅	∅	∅	
	t						[61]	∅	[62]	[63]	∅	∅	∅		[64]	[60]	∅		∅	∅	∅	
	γ							[65, 66]	[67–69]	[68, 70]	∅	∅	∅		∅	∅	∅		∅	∅	∅	
	Z/W								[71]	[71]	∅	∅	∅		∅	∅	∅		∅	∅	∅	
	H									[72, 73]	[74]	∅	∅		∅	∅	∅		∅	∅	∅	
BSM → $SM_1 \times SM_1$	q/g										∅	∅	∅		∅	∅	∅		∅	∅	∅	
	γ/π^0's											[75]	∅		∅	∅	∅		∅	∅	∅	
	b												[76, 77]		∅	∅	∅		∅	∅	∅	
	⋮																					
	⋮																					
BSM → $SM_1 \times SM_2$	tZ/H																					
	bH																					
	⋮																					
	⋮																					
BSM → complex	$\tau qq'$																					
	eqq'																					
	$\mu qq'$																					
	⋮																					
	⋮																					

Consider just the di-object search for resonant A → B C

…most cases are uncovered!

J. Kim, K. Kong, BN, D. Whiteson, *JHEP 04 (2020) 30*, 1907.06659

Why not just look everywhere?

(a) There are a lot of places to look

(b) You would find a lot of excesses

Best to cast a wide net in a smart way !

Outline: Casting a Wide **Net(work)**

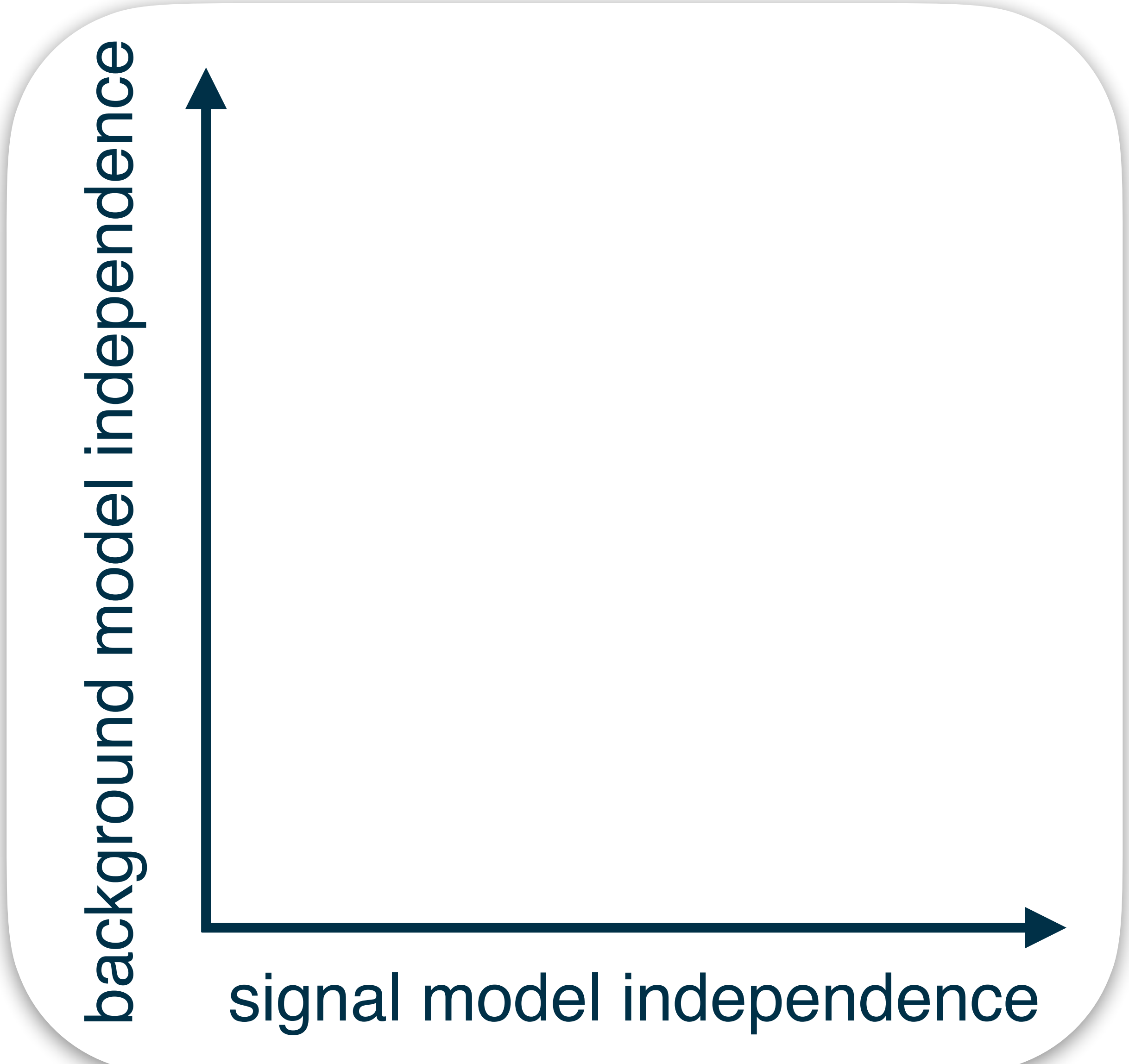
1. The landscape of model dependence
2. Overview of new ideas
3. Resonant anomaly detection
4. The future (and why you should be part of it!)

***Anomaly* in this talk means unanticipated new physics (!)**

Landscape of Model Dependence

Suppose you want to search for a new signal process

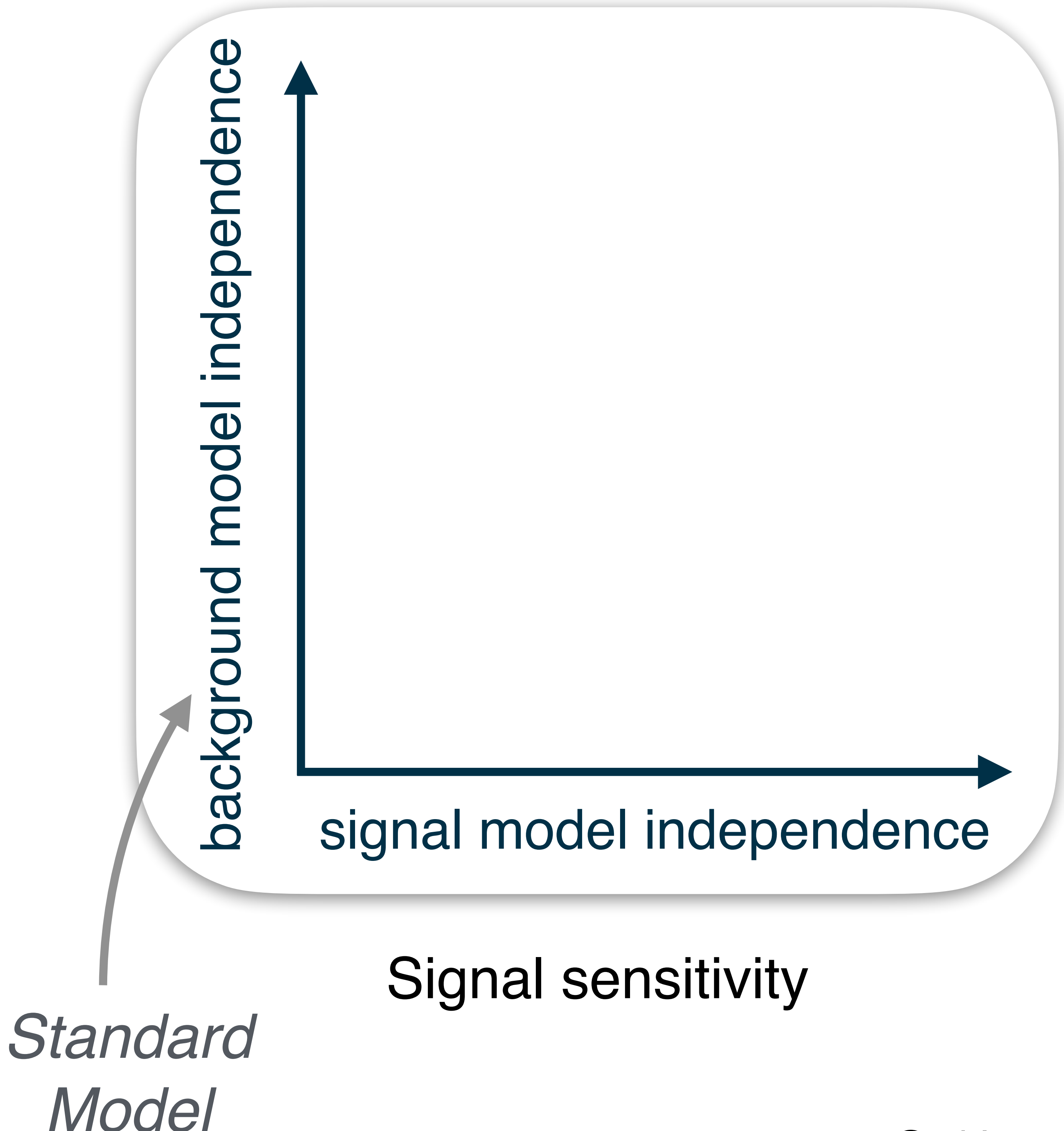
Landscape of Model Dependence



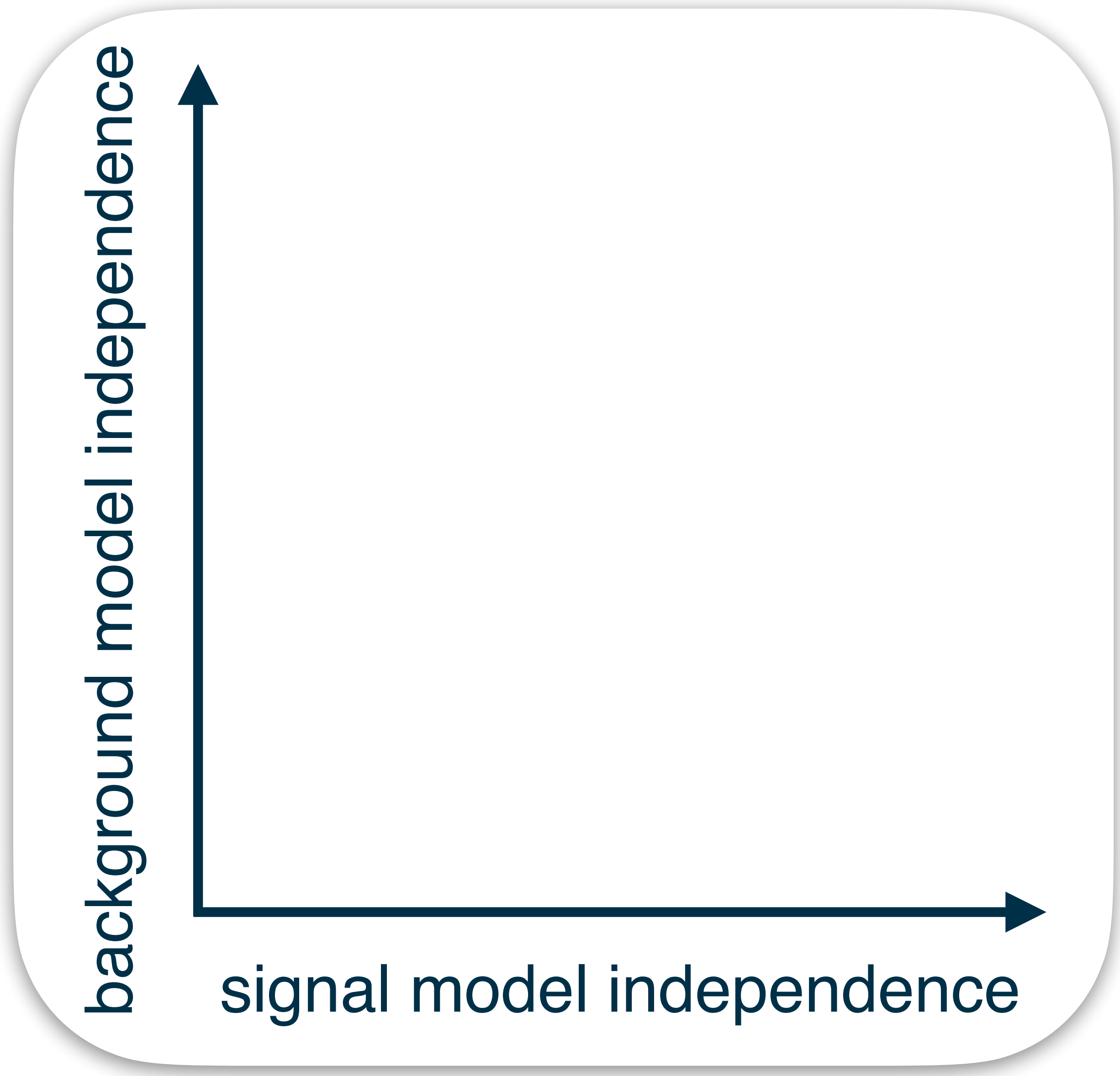
Signal sensitivity

Suppose you want to search for a new signal process

Landscape of Model Dependence



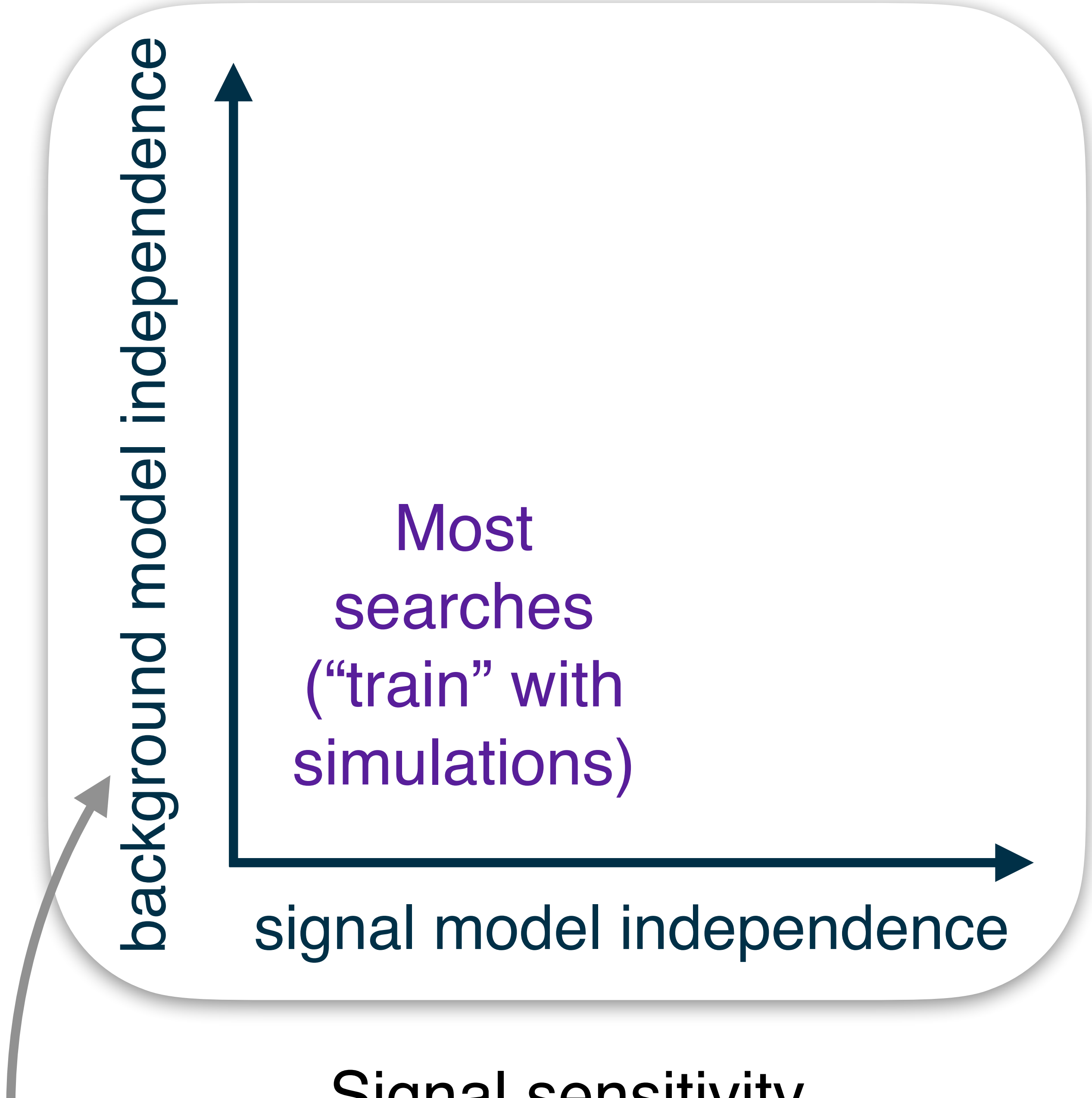
Signal sensitivity



Background specificity

G. Karagiorgi, G. Kasieczka, S. Kravitz, BN, D. Shih, Nature Reviews Physics (2022), 2112.03769

Landscape of Model Dependence



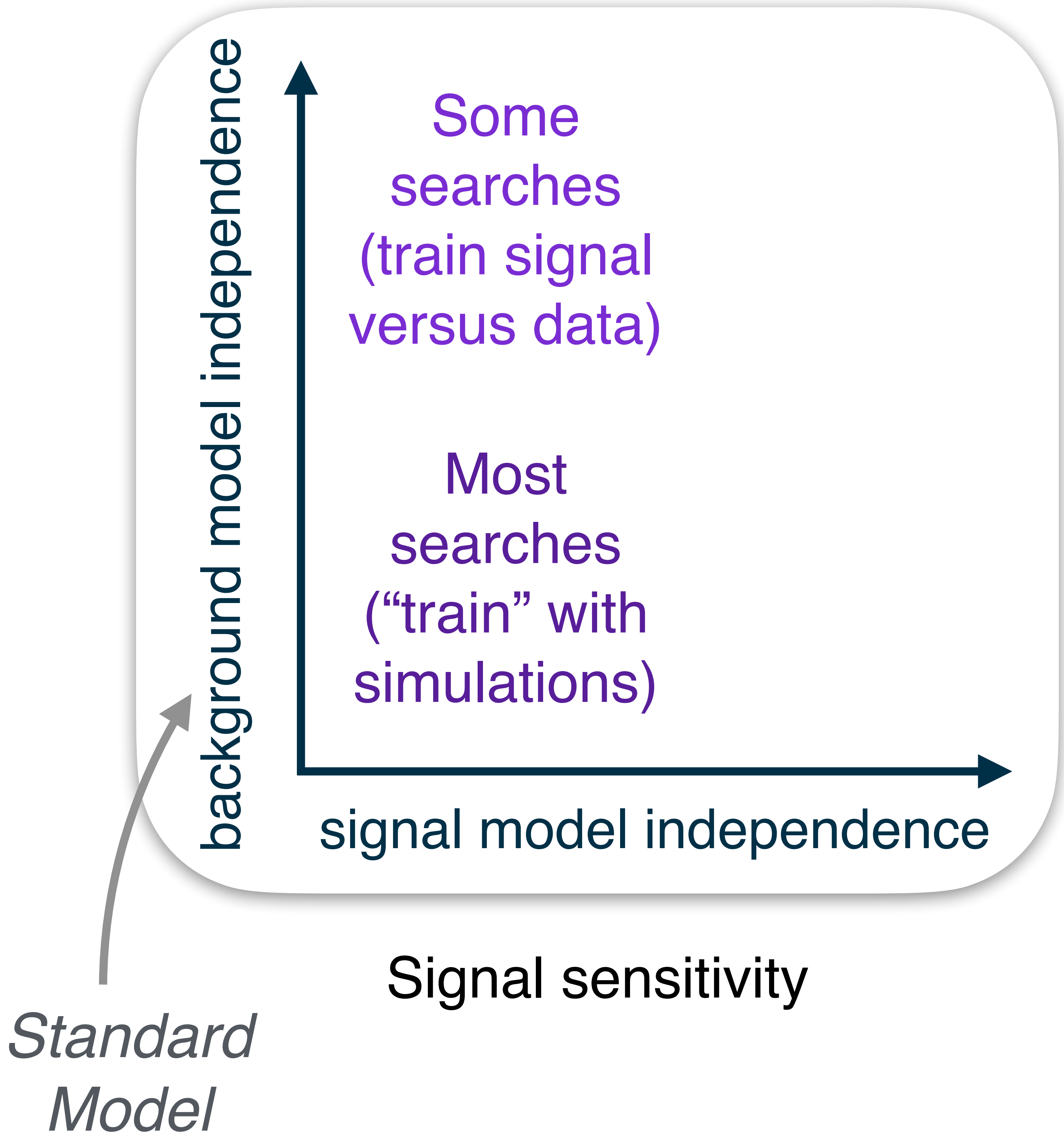
Signal sensitivity

Standard Model

> 99% of searches at the LHC and elsewhere are of this type

"train" is in quotes because such searches may or may not use machine learning

Landscape of Model Dependence

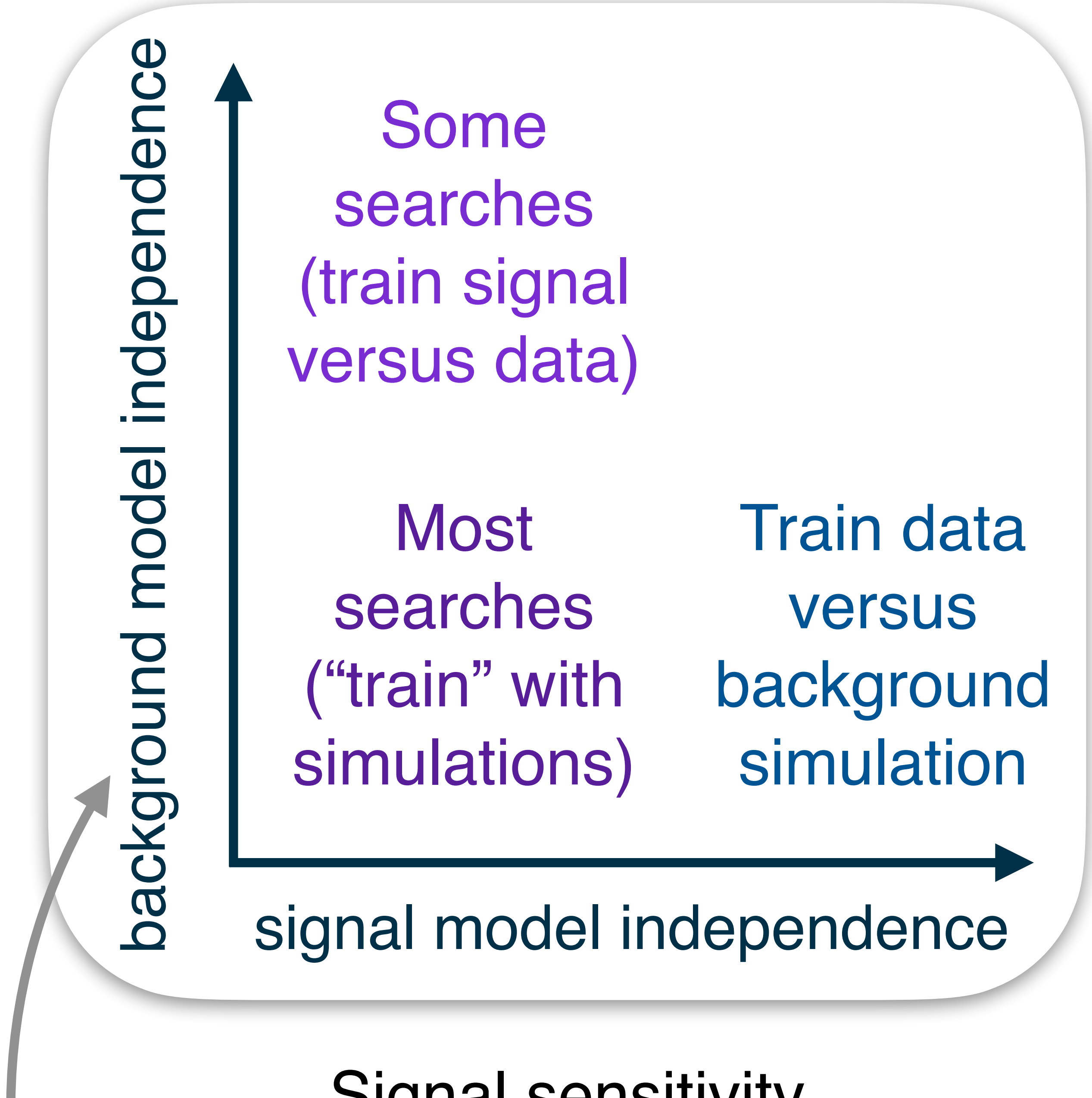


Signal sensitivity

e.g. signal simulation versus calibration data

standard approach when signal is clean and well-understood, but background is not, e.g. $h \rightarrow \gamma\gamma$

Landscape of Model Dependence



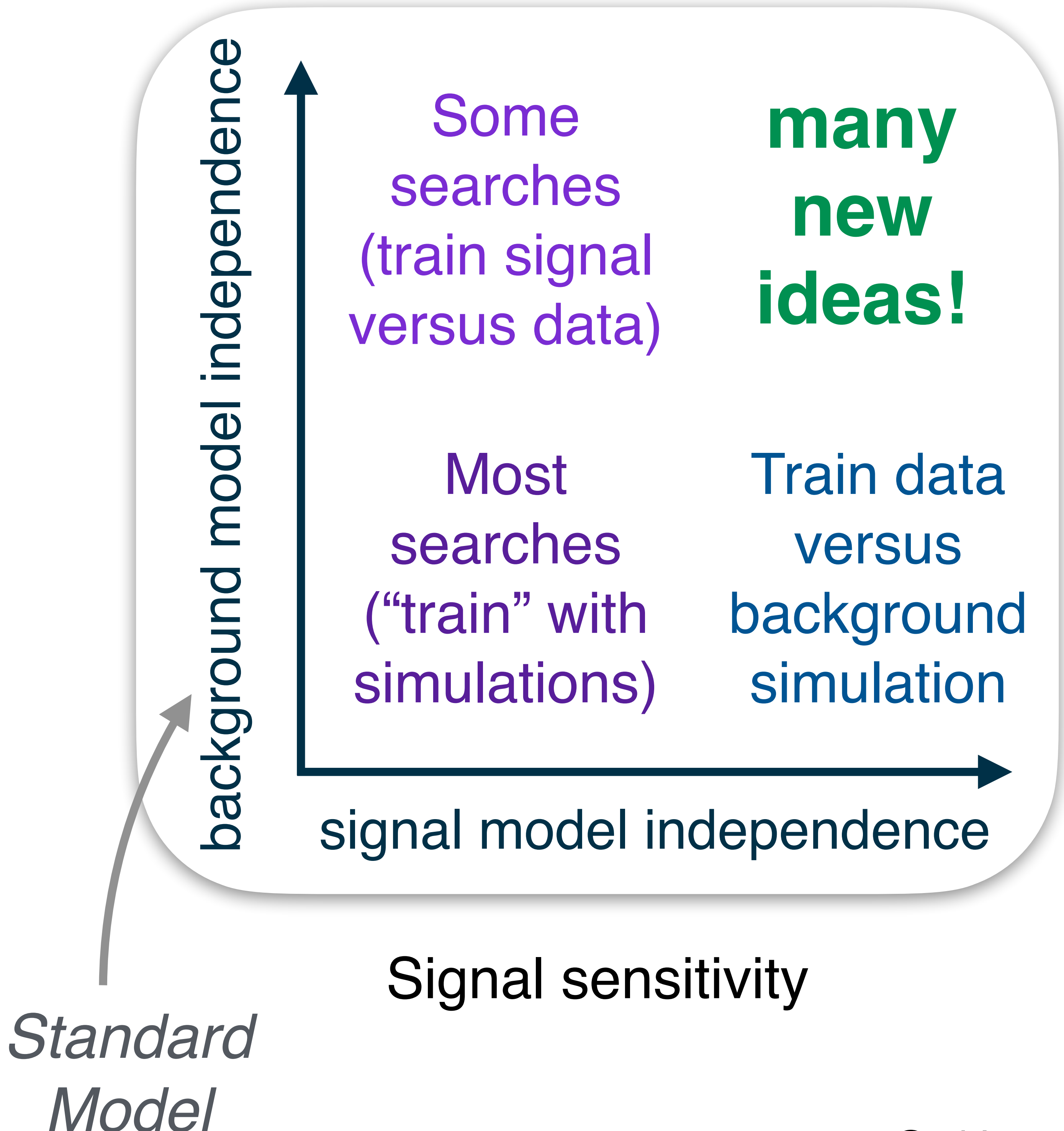
Standard Model

Signal sensitivity

signal model *independent*
background model *dependent*

There is a history of these searches at the LHC, Tevatron, HERA, LEP

Landscape of Model Dependence

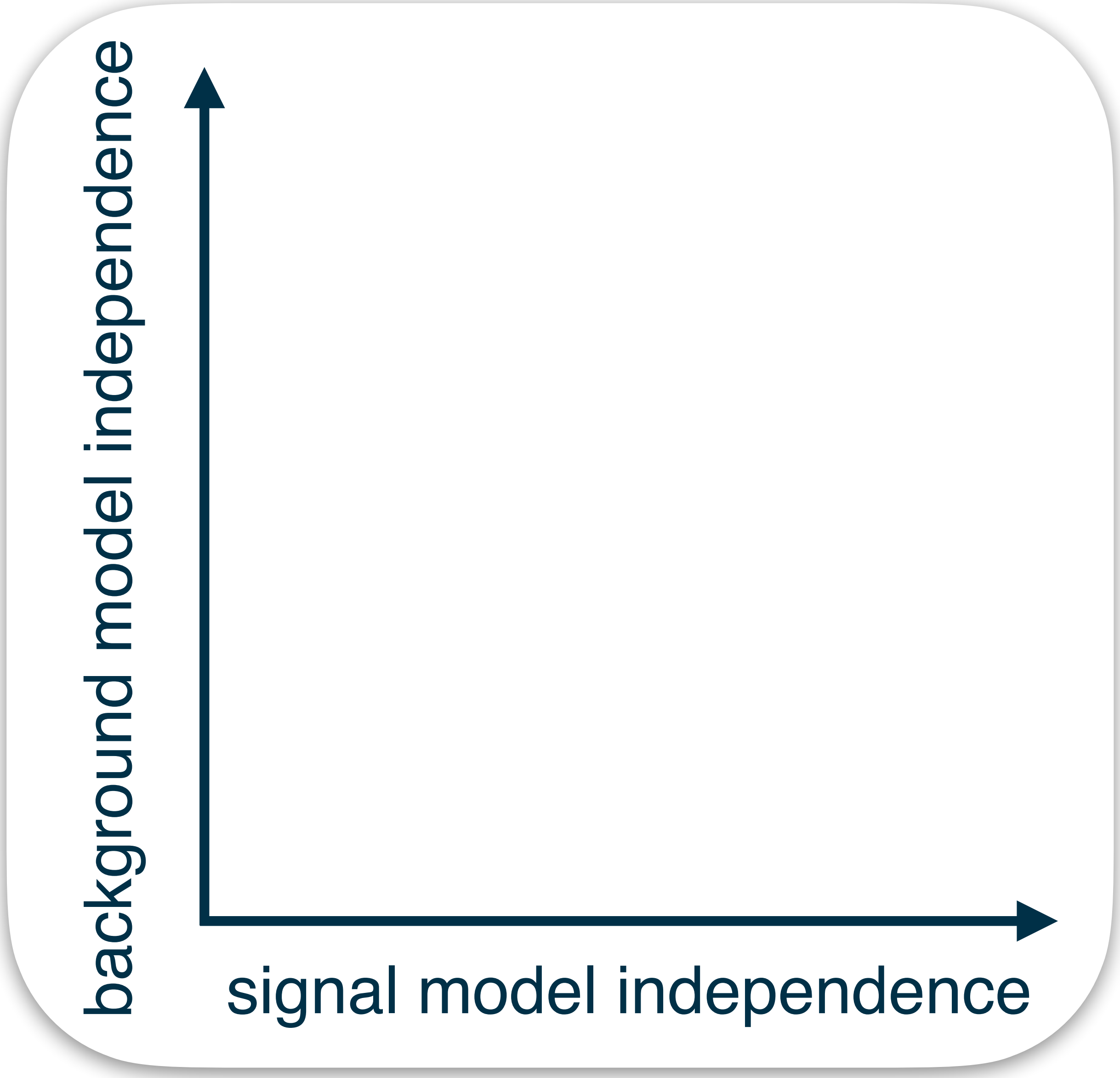


There are many new ideas that make use of modern machine learning

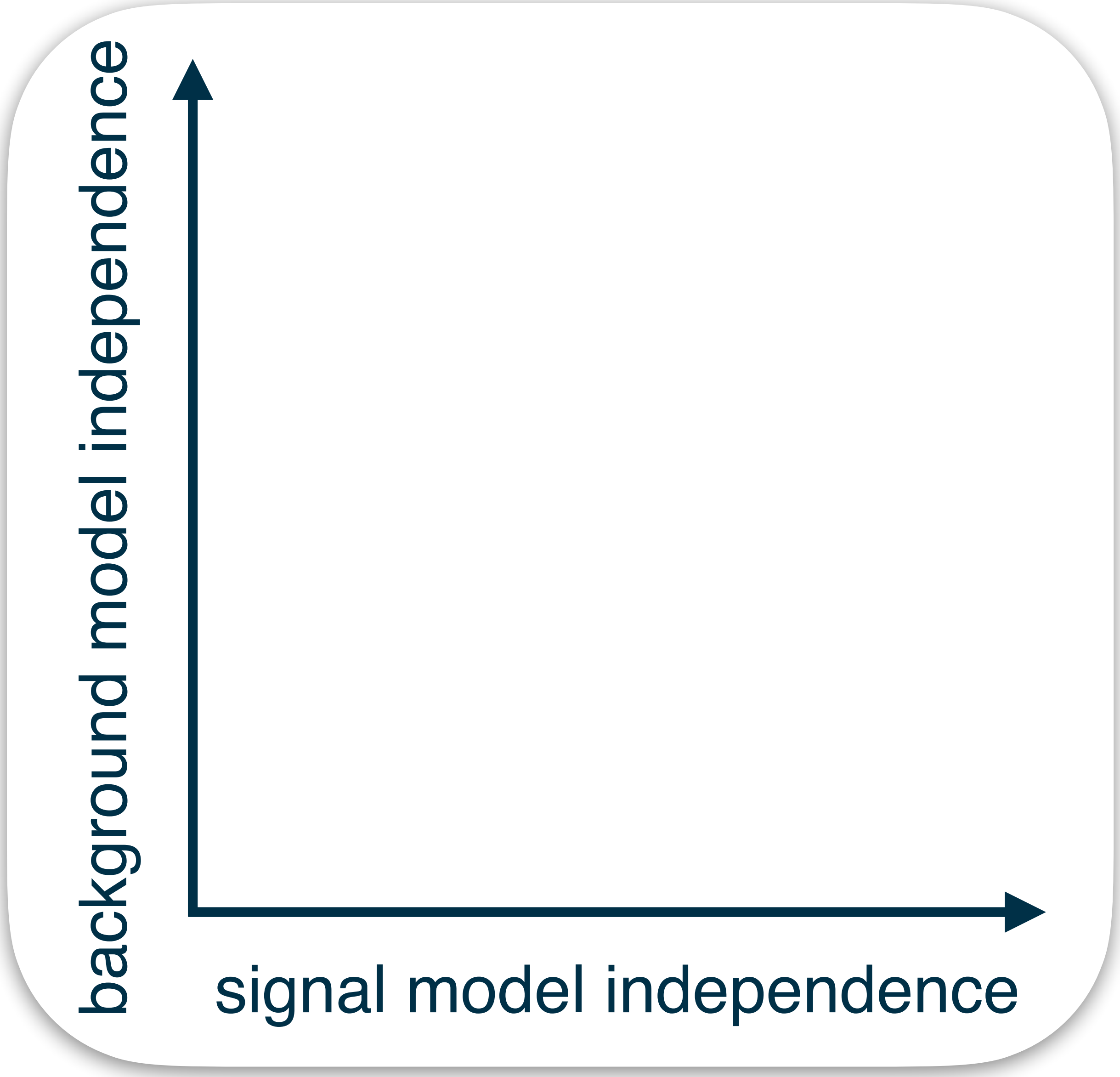
The goal is to learn **directly from data**, injecting as little bias as possible

G. Karagiorgi, G. Kasieczka, S. Kravitz, BN, D. Shih, Nature Reviews Physics (2022), 2112.03769

Landscape of Model Dependence



Signal sensitivity



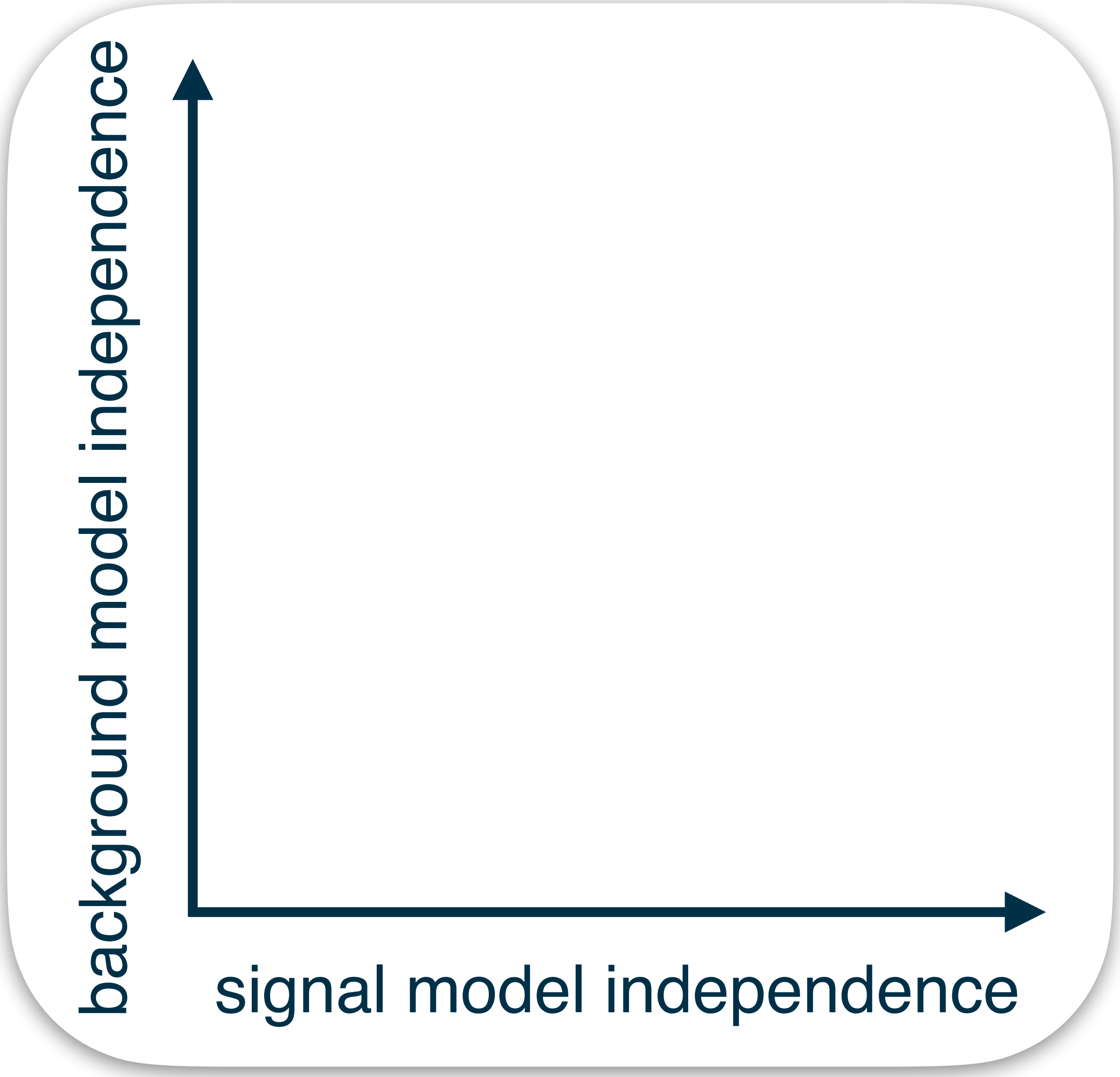
Background specificity

Suppose you want to search for a new signal process

Landscape of Model Dependence

Core idea: create a reference sample and see if our target and reference are the same; if yes, limits; if no, discovery!

See Andreas's talk about how to "compare" using ML



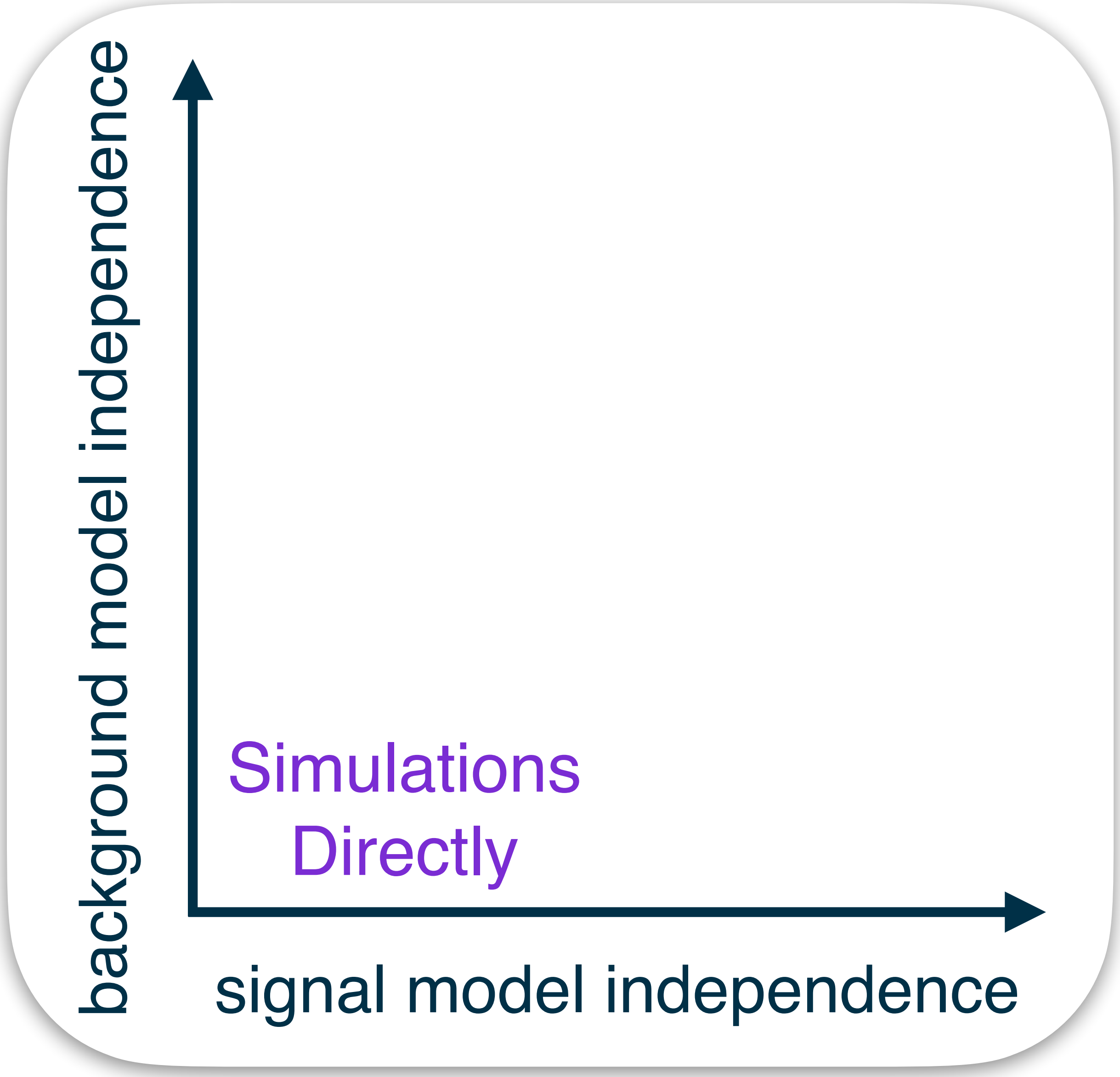
Background specificity

Suppose you want to search for a new signal process

Landscape of Model Dependence

Almost no searches at the LHC are of this type, with a few exceptions for very well-known processes like 4-leptons

(See K. Krzyzanska and B. Nachman, 2203.09601)



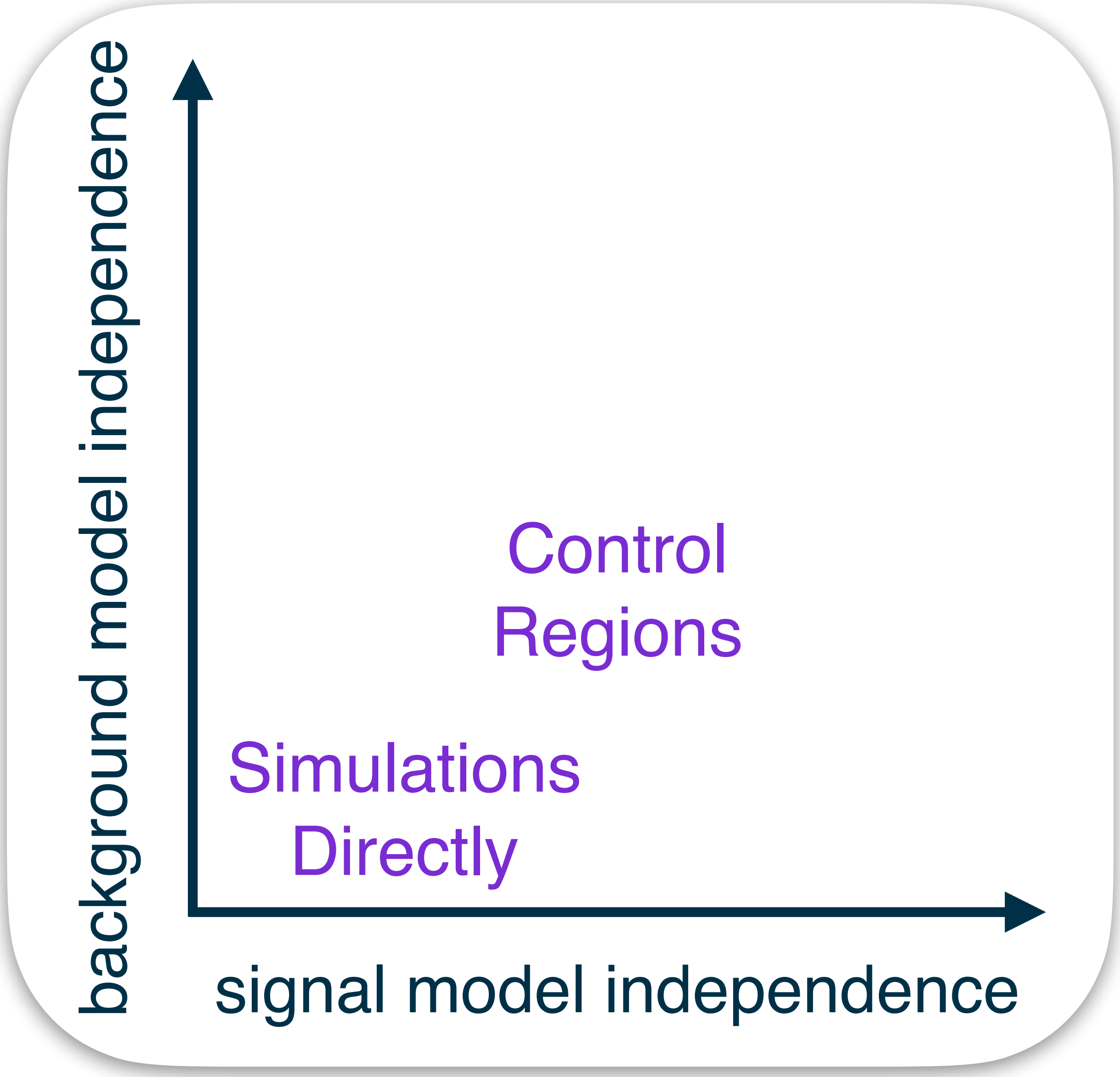
Background specificity

Suppose you want to search for a new signal process

Landscape of Model Dependence

Background-dominated regions are used to constrain the signal-sensitive regions (using simulation to relate the regions)

This is one of the most common approaches



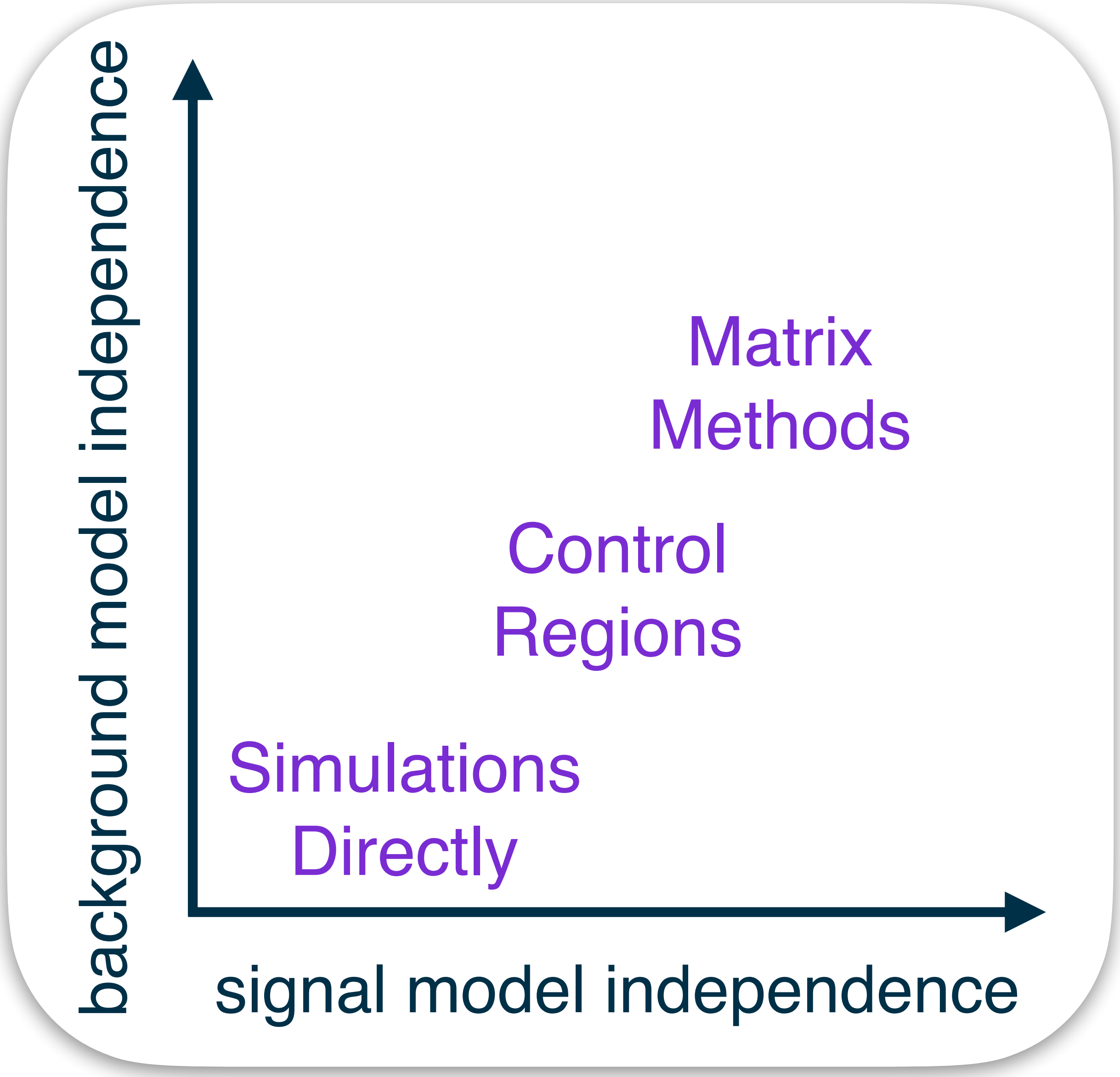
Background specificity

Suppose you want to search for a new signal process

Landscape of Model Dependence

Same as control regions, but the "transfer factors" from the control region to signal region now are derived in data. (use simulations to validate)

Can also be automated with ML! See G. Kasieczka et al., 2007.14400



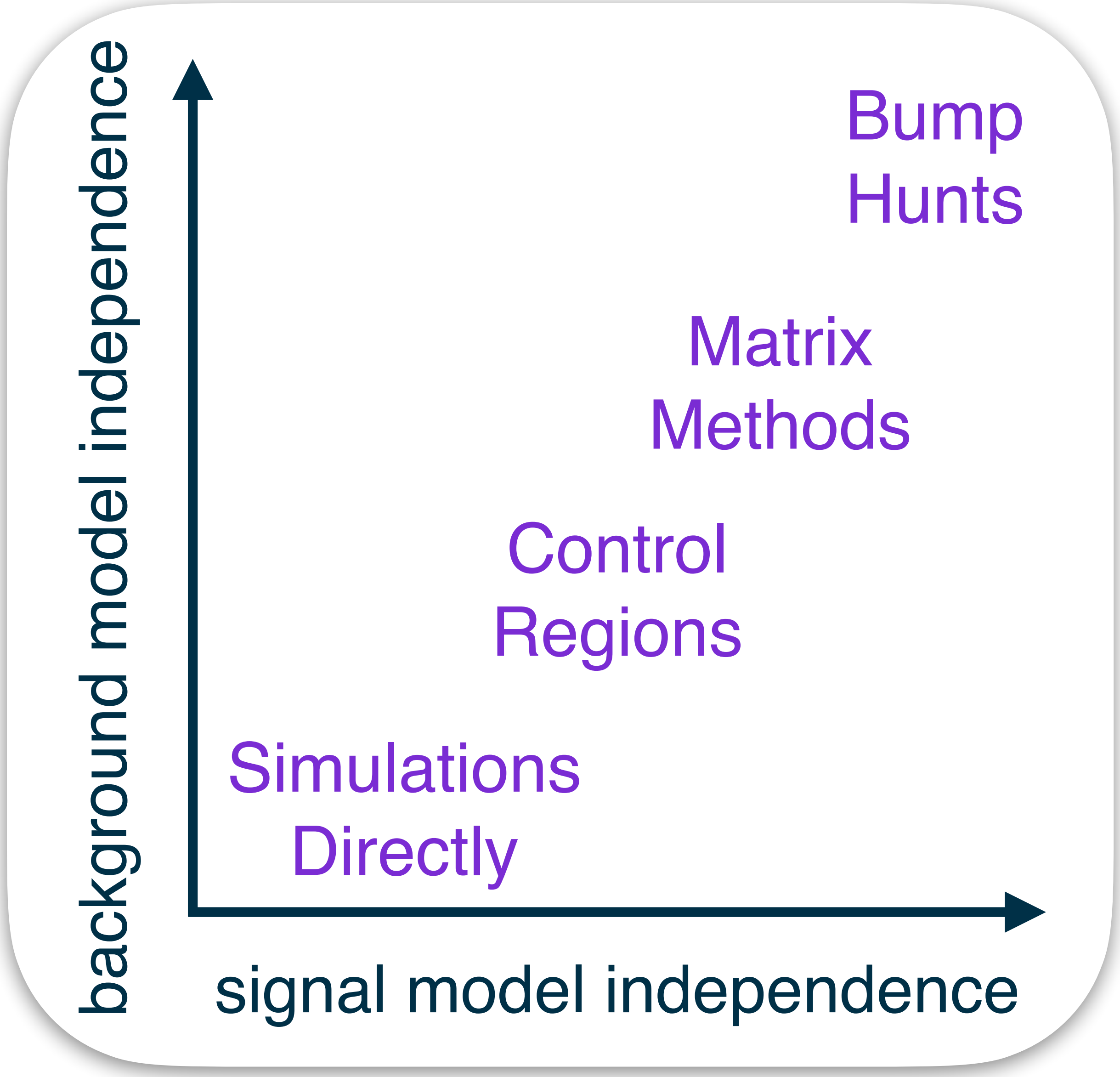
Background specificity

Suppose you want to search for a new signal process

Landscape of Model Dependence

*Many of these searches don't use simulations **at all (!)***

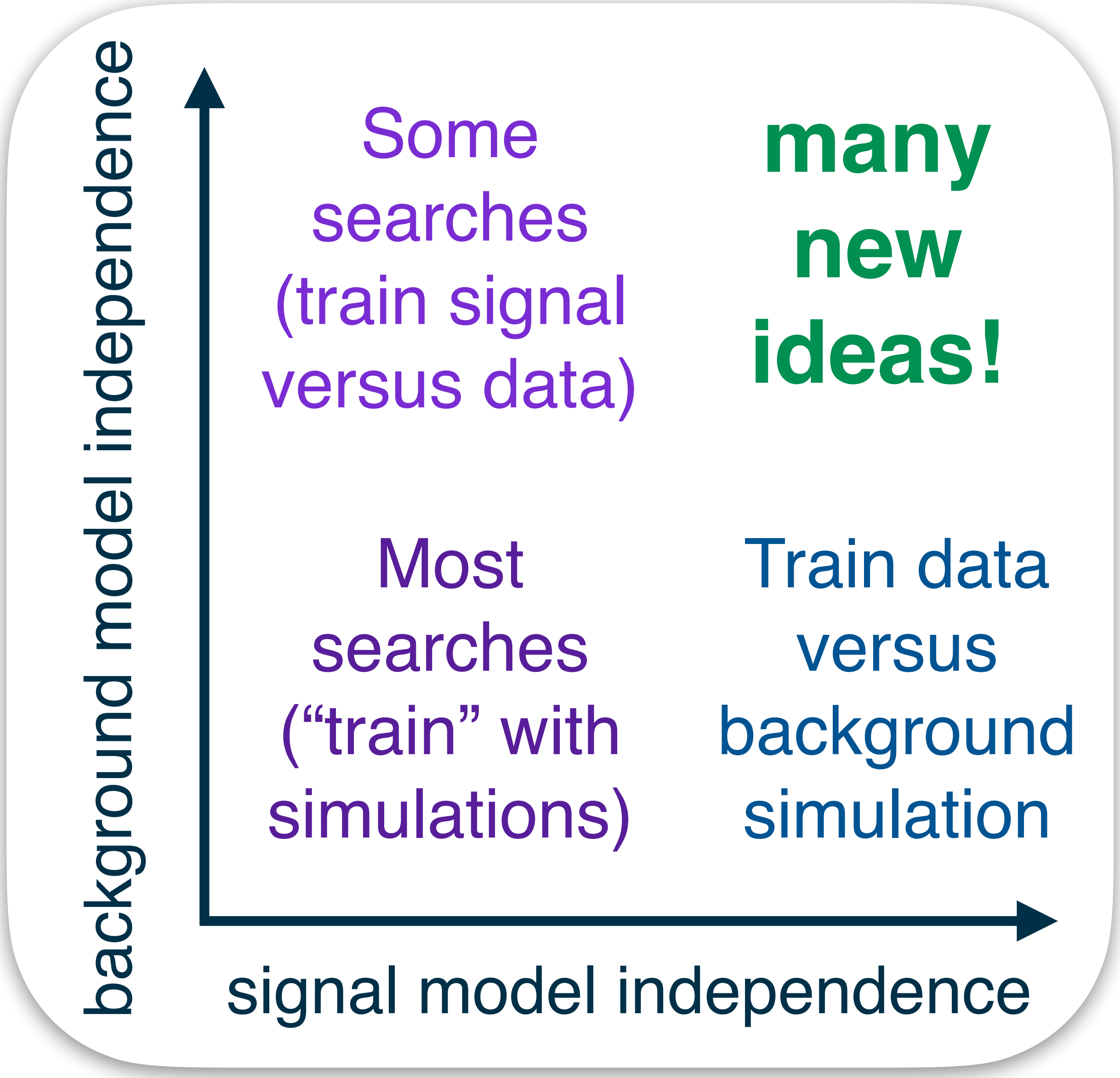
A big challenge is finding the right fit function … ML can also play a role there - see e.g. M. Frate et al., 1709.05681



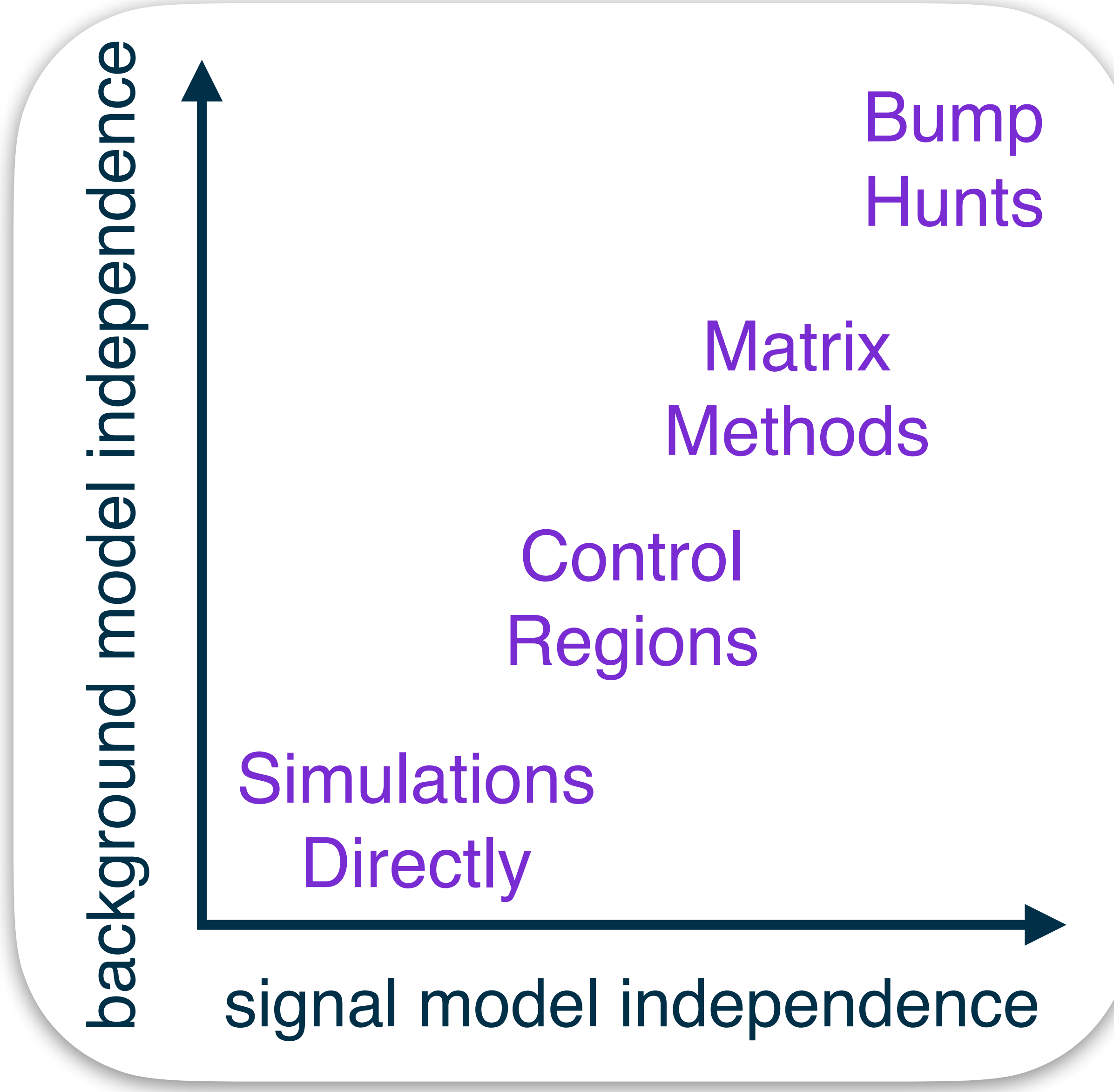
Background specificity

Suppose you want to search for a new signal process

Landscape of Model Dependence



Signal sensitivity



Background specificity

Suppose you want to search for a new signal process

Overview of New Ideas [signal sens.]

I like to categorize new ideas based on the core assumption about the BSM, which is intimately related to the technique ***supervision***

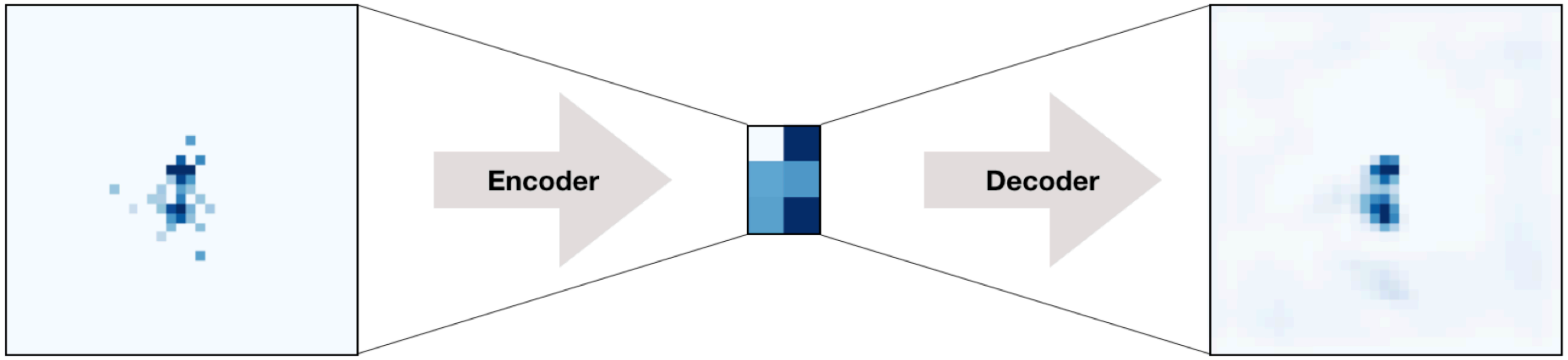
Unsupervised = no labels
Weakly-supervised = noisy labels
Semi-supervised = partial labels
Supervised = full label information

This is most searches. You simulate the signal (label = 1), simulate the background (label = 0) and "train" a classifier to distinguish the 1's from the 0's.

Unsupervised

Unsupervised = no labels

Typically, the goal of these methods is to look for events with **low *p(background)***



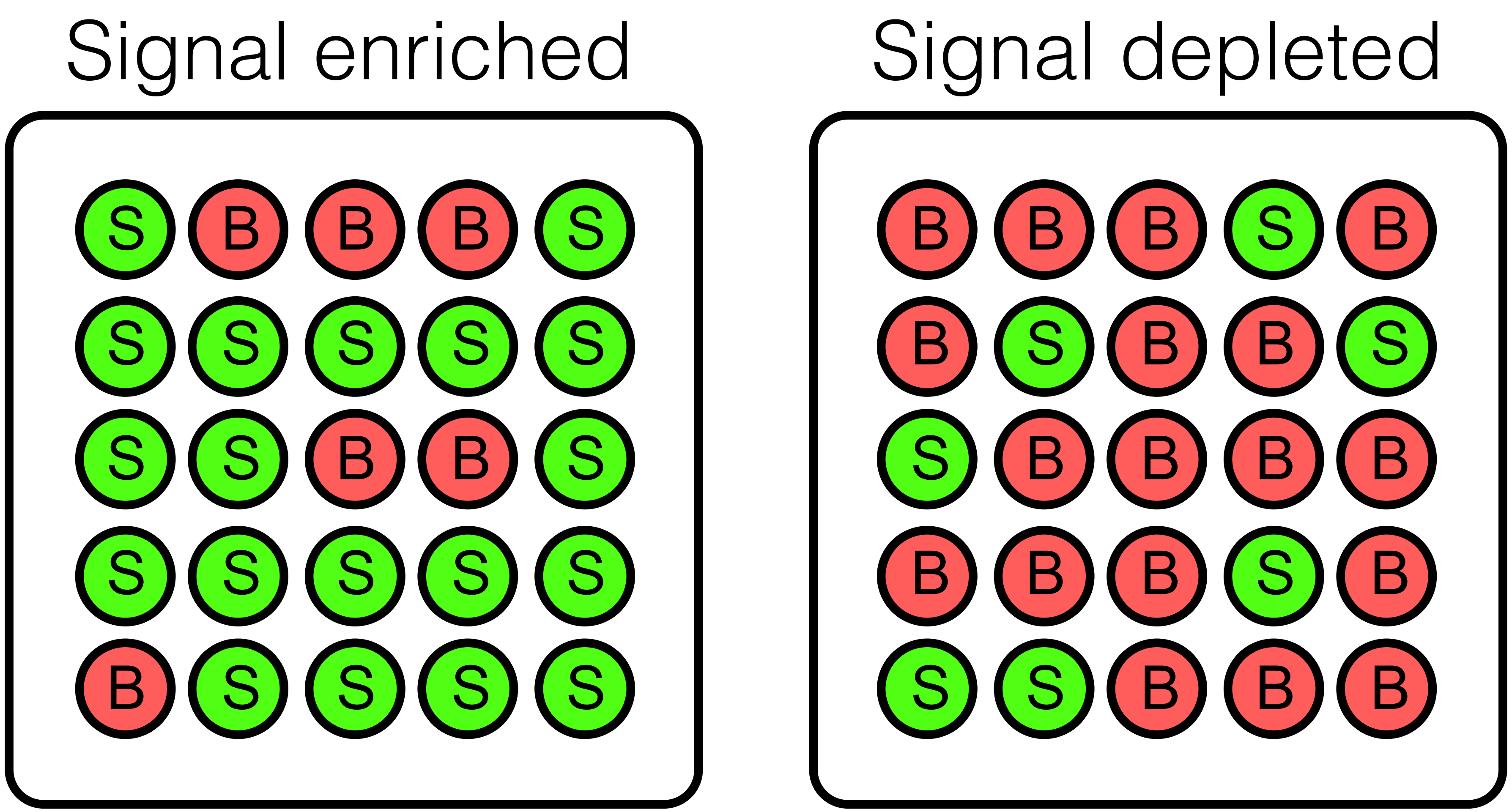
One strategy (autoencoders) is to try to compress events and then uncompress them. When x is far from uncompres(compress(x)), then x probably has low p(x).

M. Farina, Y. Nakai, D. Shih, 1808.08992; T. Heimel, G. Kasieczka, T. Plehn, J. Thompson, 1808.08979; + many more

Weakly-supervised

Weakly-supervised = noisy labels

Typically, the goal of these methods is to look for events with high *p(possibly signal-enriched)/p(possibly signal-depleted)*



E. Metodiev, BN, J. Thaler, 1708.02949; J. Collins, K. Howe, BN, 1805.02664; + many more

Semi-supervised

Semi-supervised = partial labels

Typically, these methods use some signal simulations to build signal sensitivity

Image credit: https://www.particlezoo.net

vs

e.g. SM background versus many signals

Overview of New Ideas

Approach:	**Unsupervised**	**Weakly supervised**
BSM assumption	Signal is rare (low p)	Signal is an over density (high p ratio)
Main drawback	rare is not invariant* under coordinate transformations!	need two samples

*for a detailed discussion about this, see K. Desai, BN, J. Thaler, 2112.05722

Overview of New Ideas

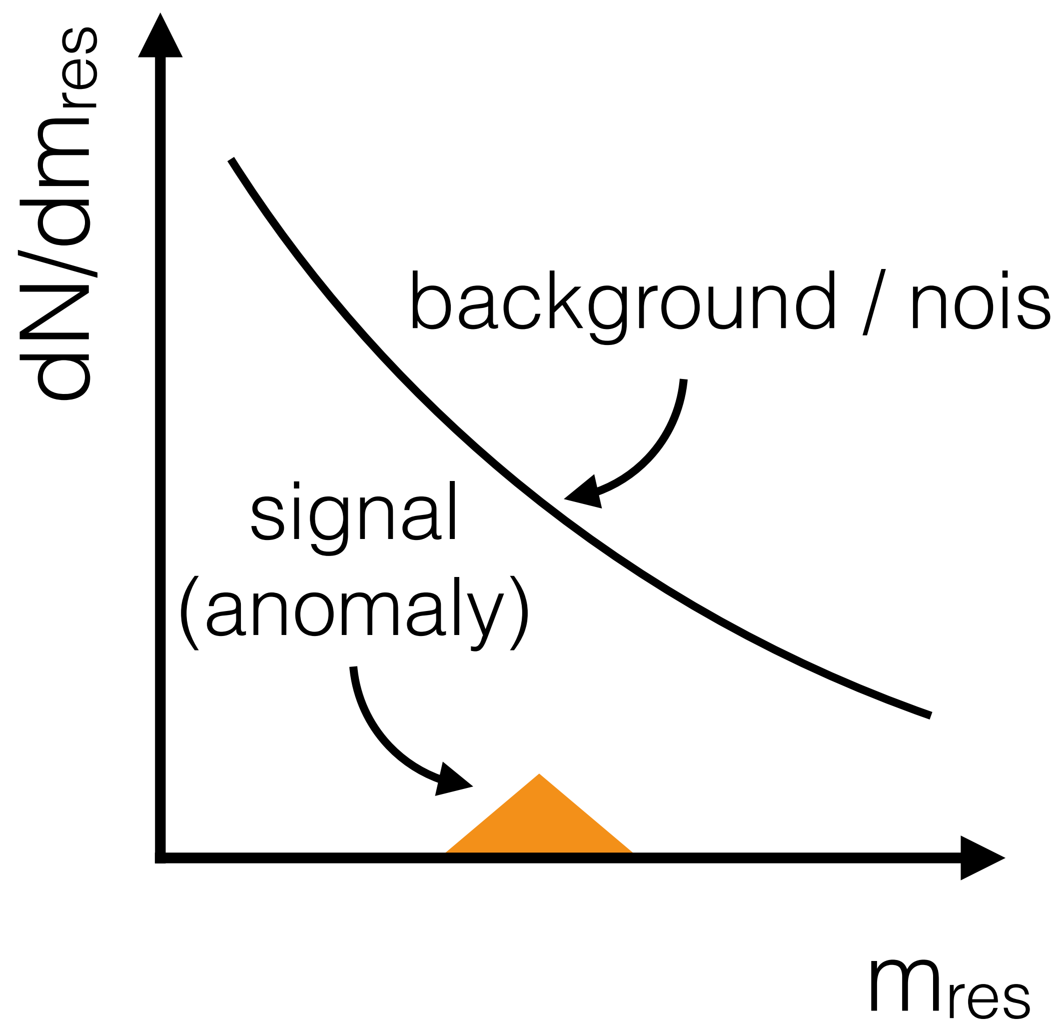
Approach:	**Unsupervised**	**Weakly supervised**
BSM assumption	Signal is rare (low *p*)	Signal is an over density (high *p* ratio)
Main drawback	rare is not invariant* under coordinate transformations!	need tw[o] samples

Cannonical example: resonances!

*for a detailed discussion about this, see K. Desai, BN, J. Thaler, 2112.05722

Resonant Anomalies

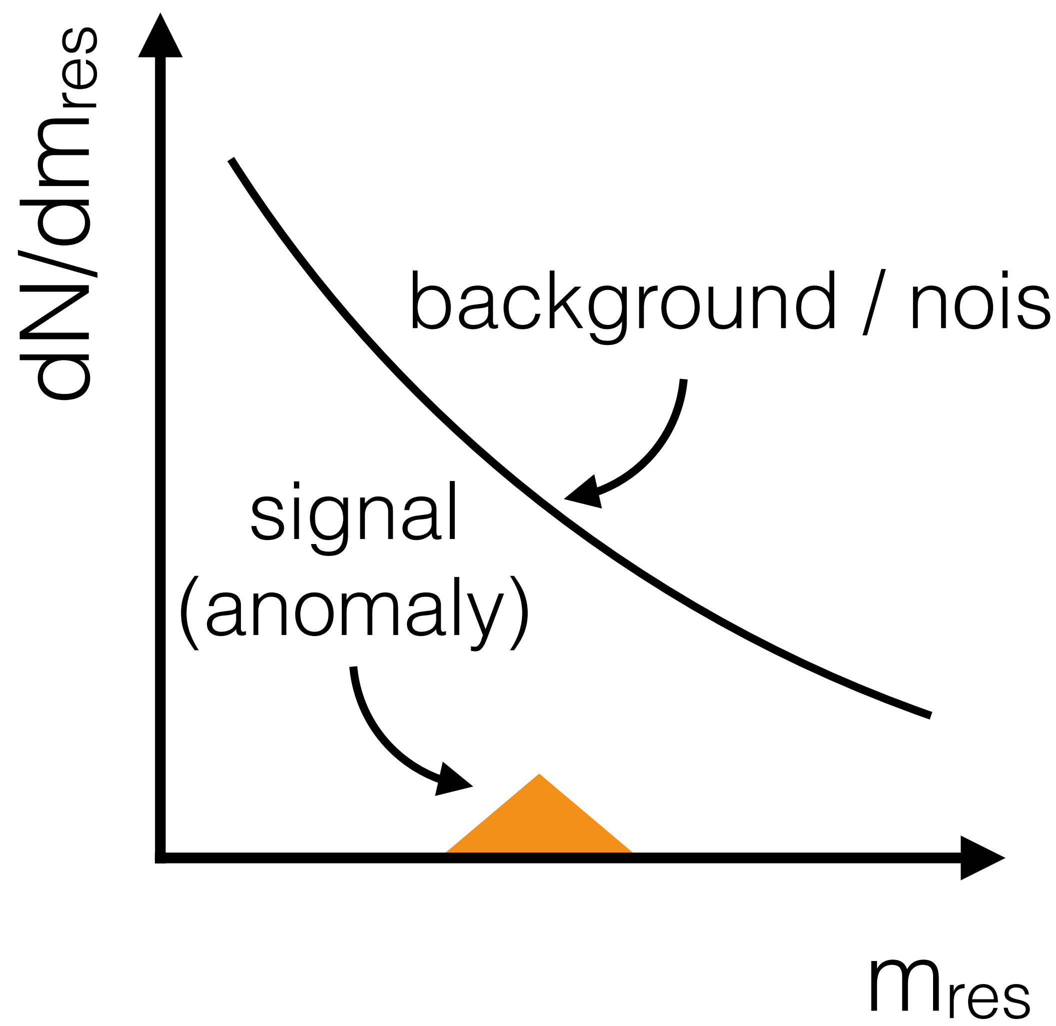
A relatively general, but powerful assumption is that the anomaly is localized somewhere in phase space.



Generically true when there are on-shell new particles.

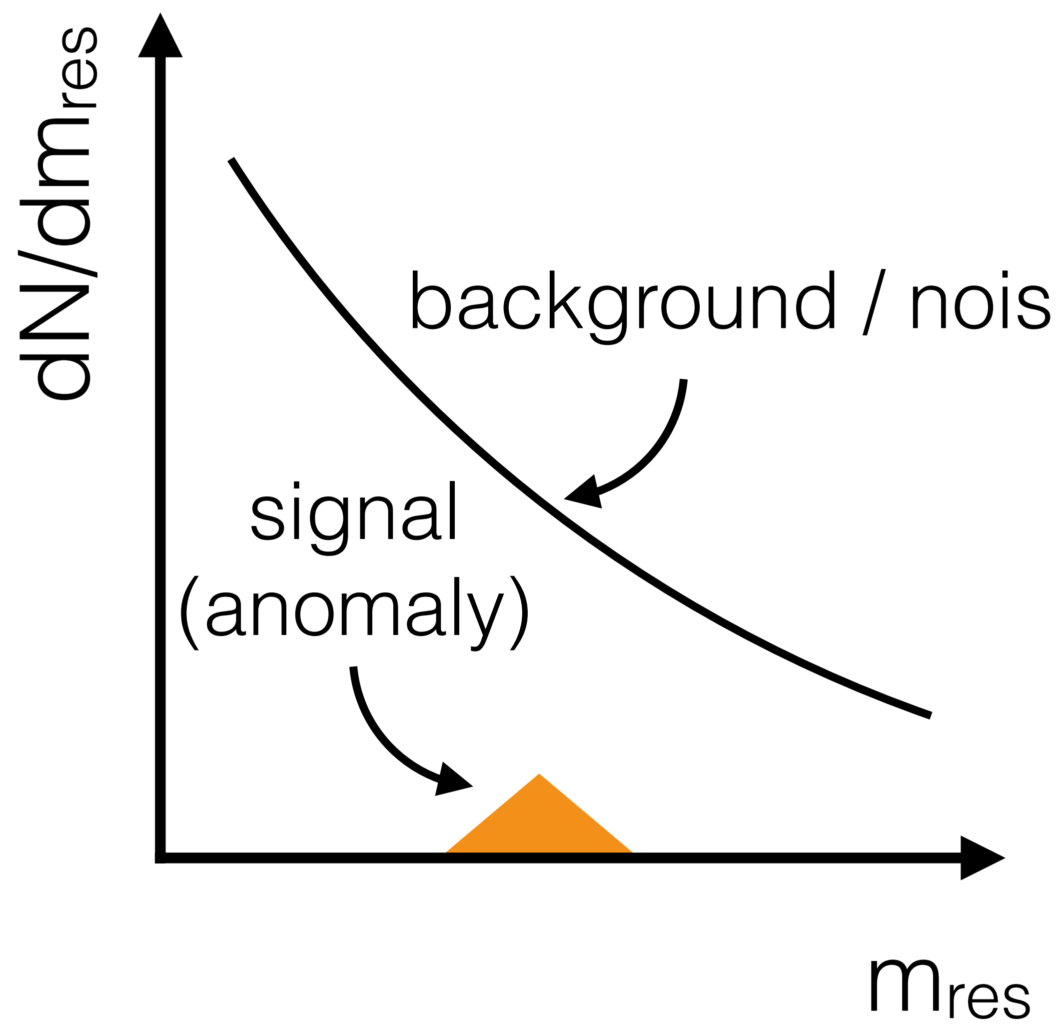
Resonant Anomalies

I’ll walk you through a weakly-supervised approach.



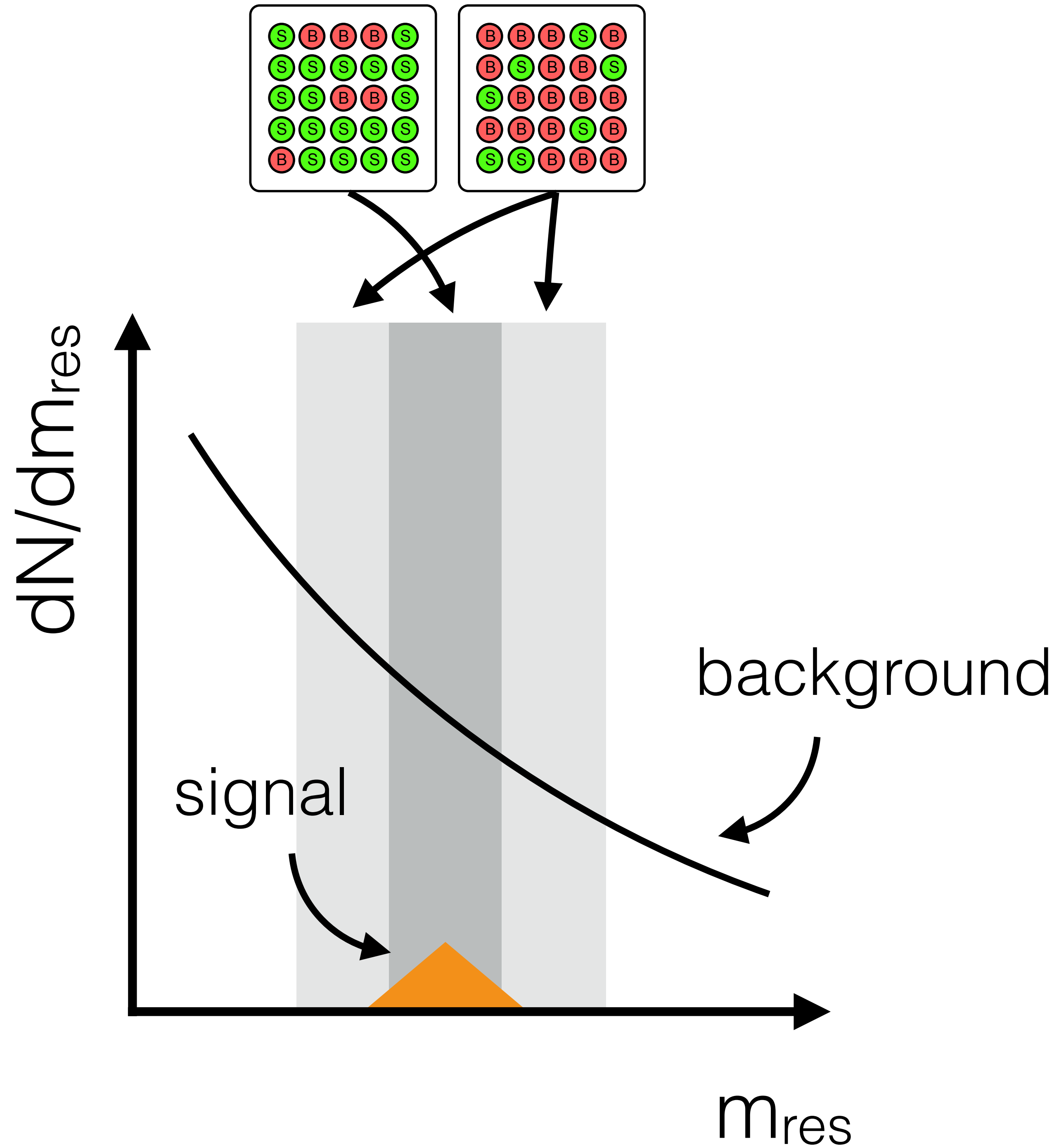
Resonant Anomalies

I'll walk you through a weakly-supervised approach.



First: we will need to generate (noisy) labels.

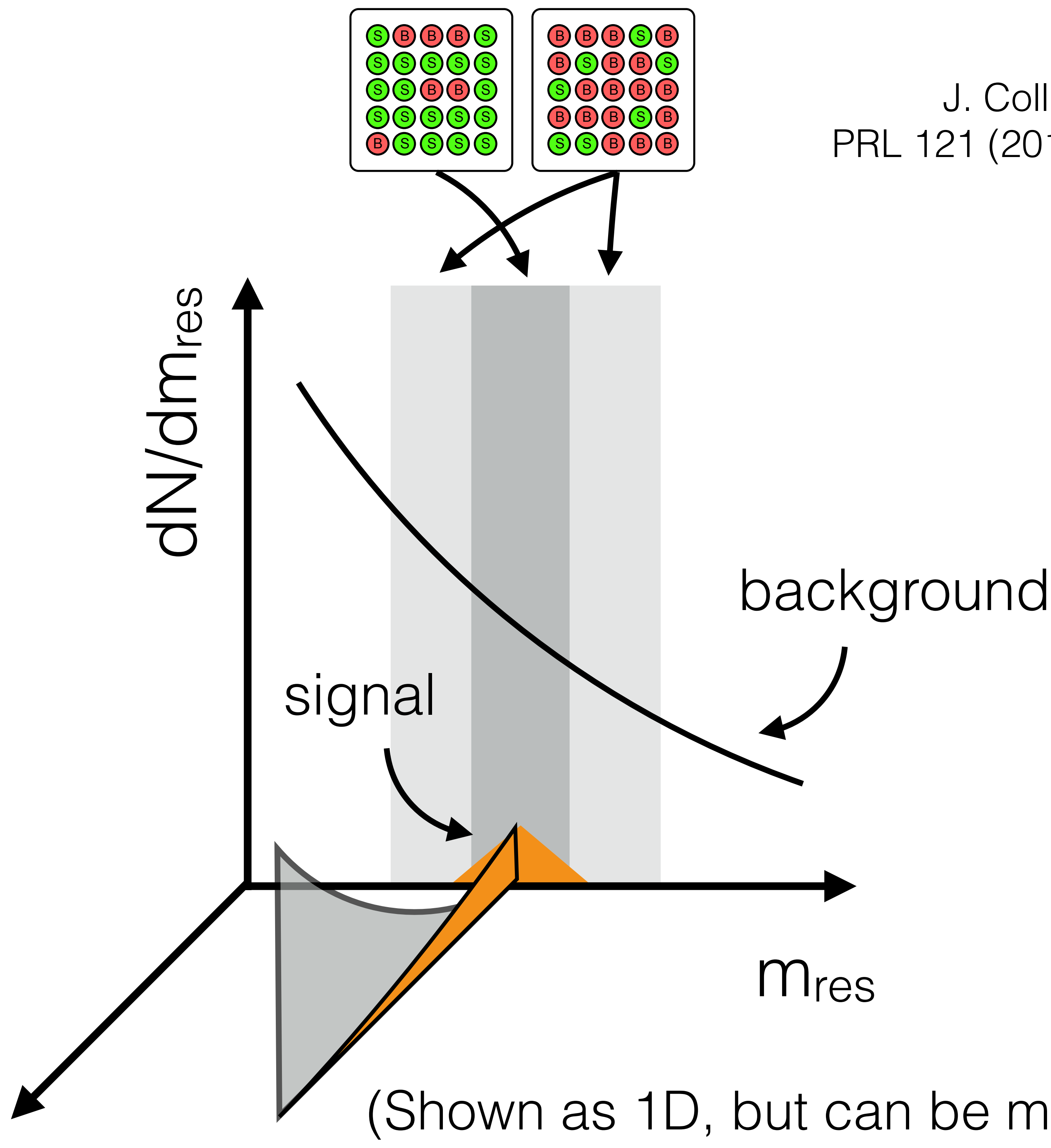
Resonant Anomalies



First: we will need to generate noisy labels.

Resonant Anomalies

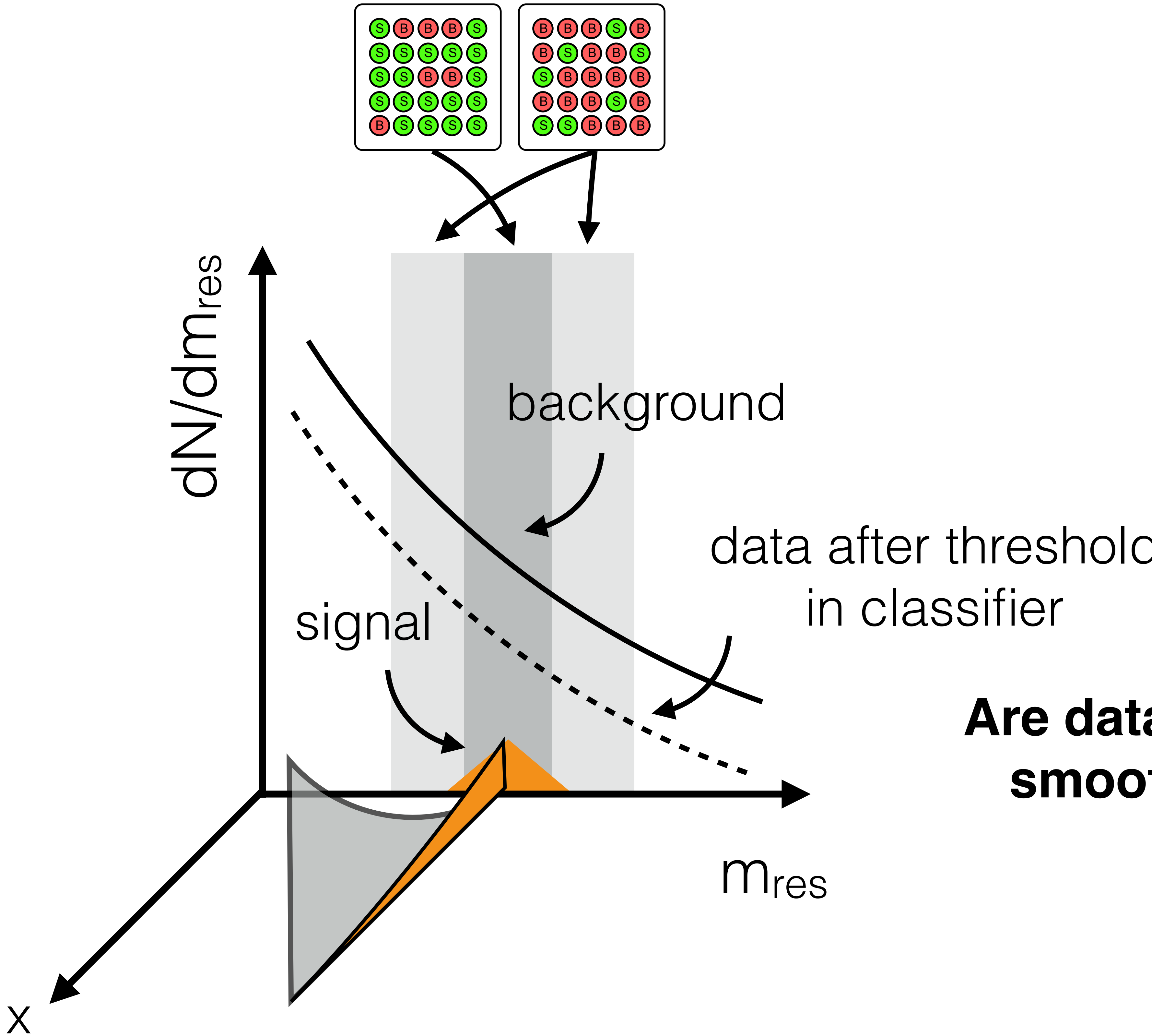
J. Collins, K. Howe, BN,
PRL 121 (2018) 241803, 1805.02664



Extended feature space for machine learning

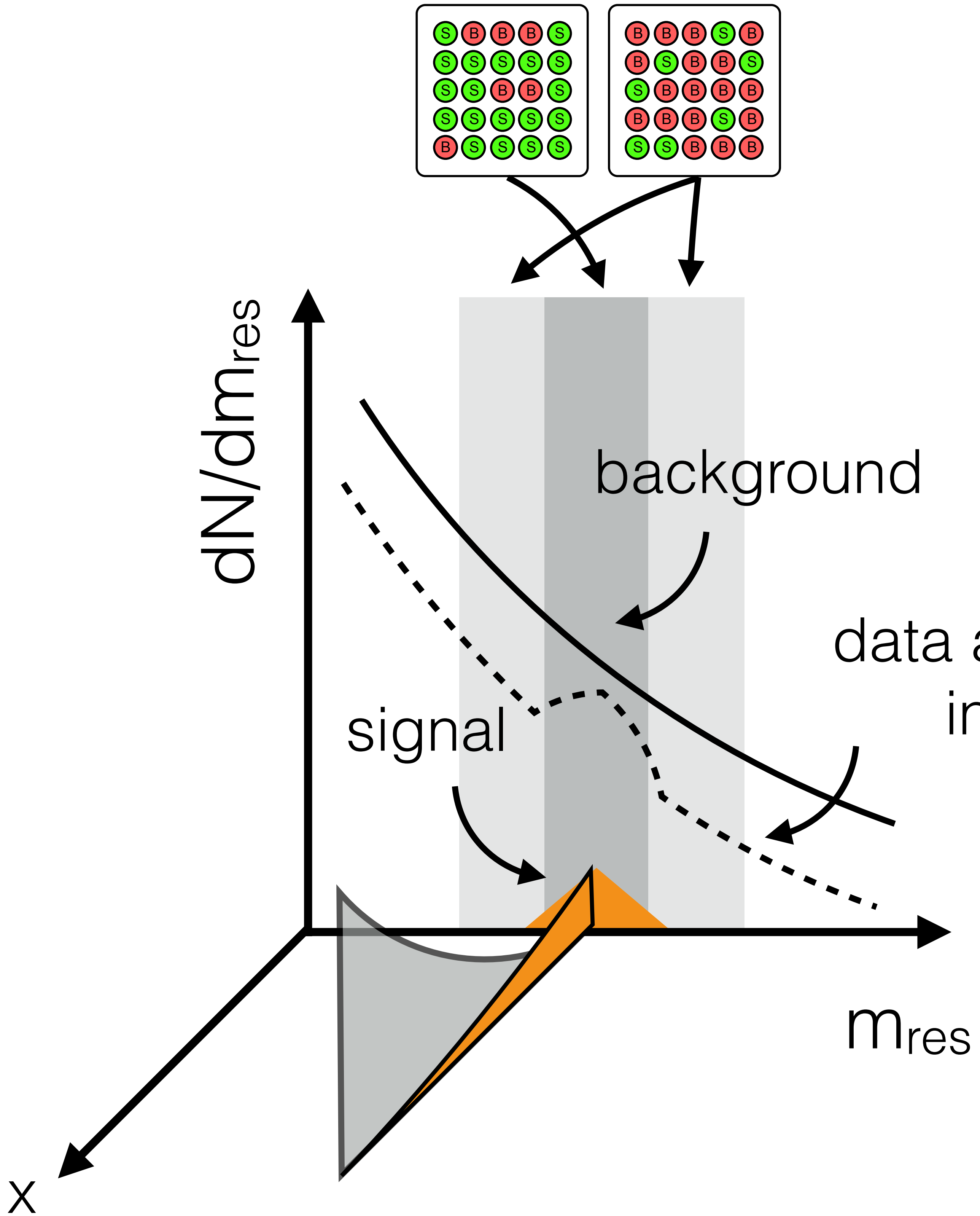
(Shown as 1D, but can be many-dimensional)

Resonant Anomalies



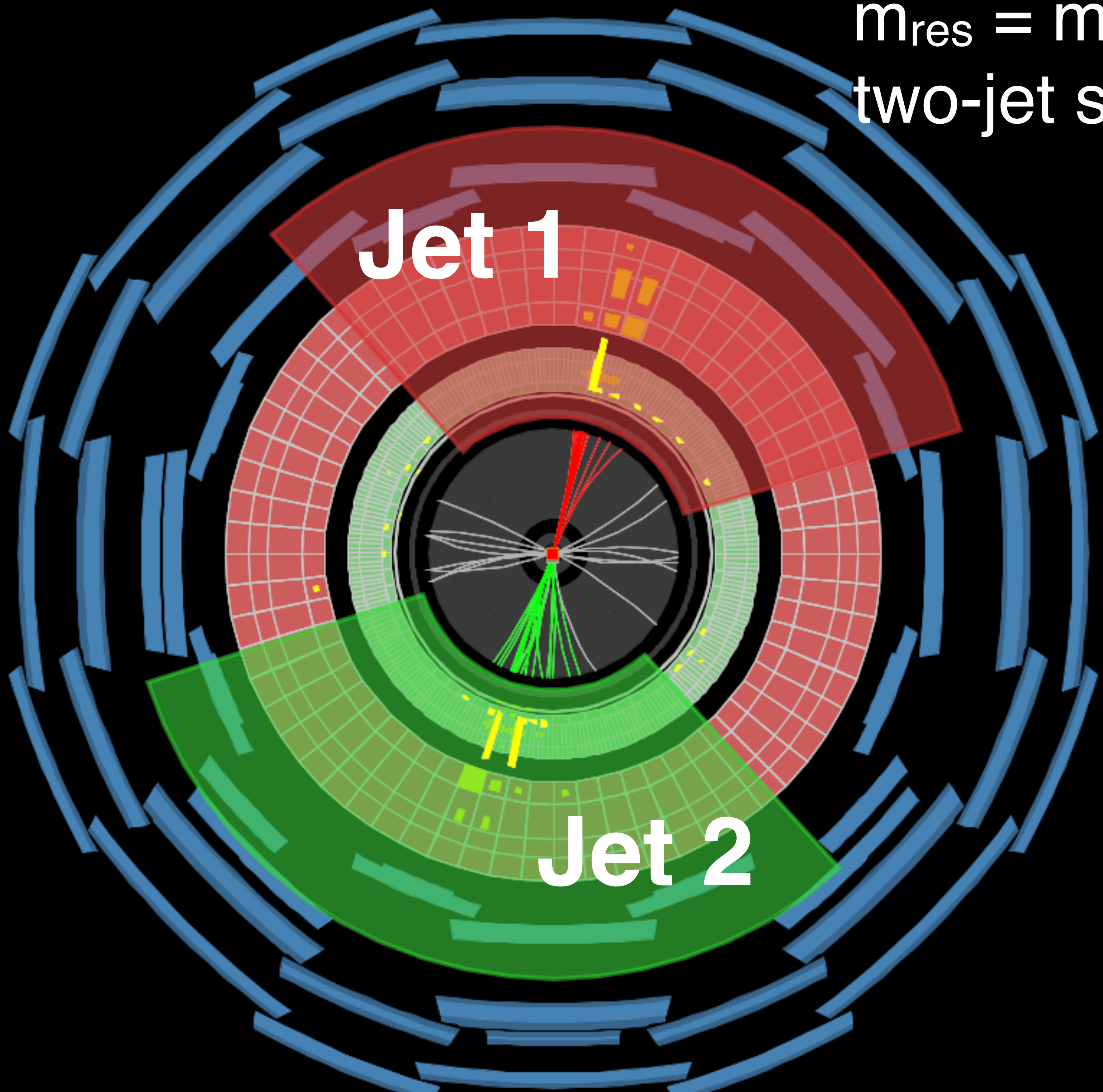
Are data still smooth?

Resonant Anomalies

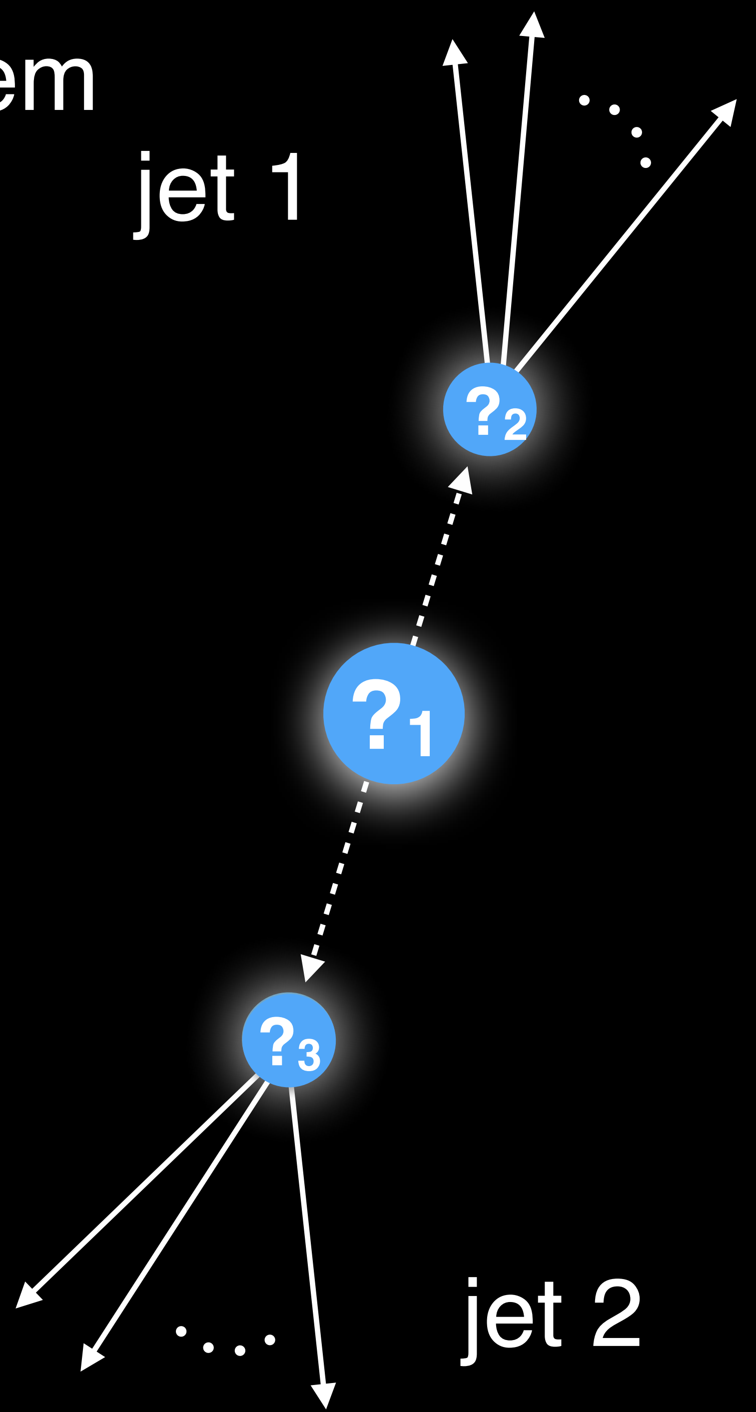


Are data still smooth?

Example: two-jet search



m_{res} = mass of two-jet system

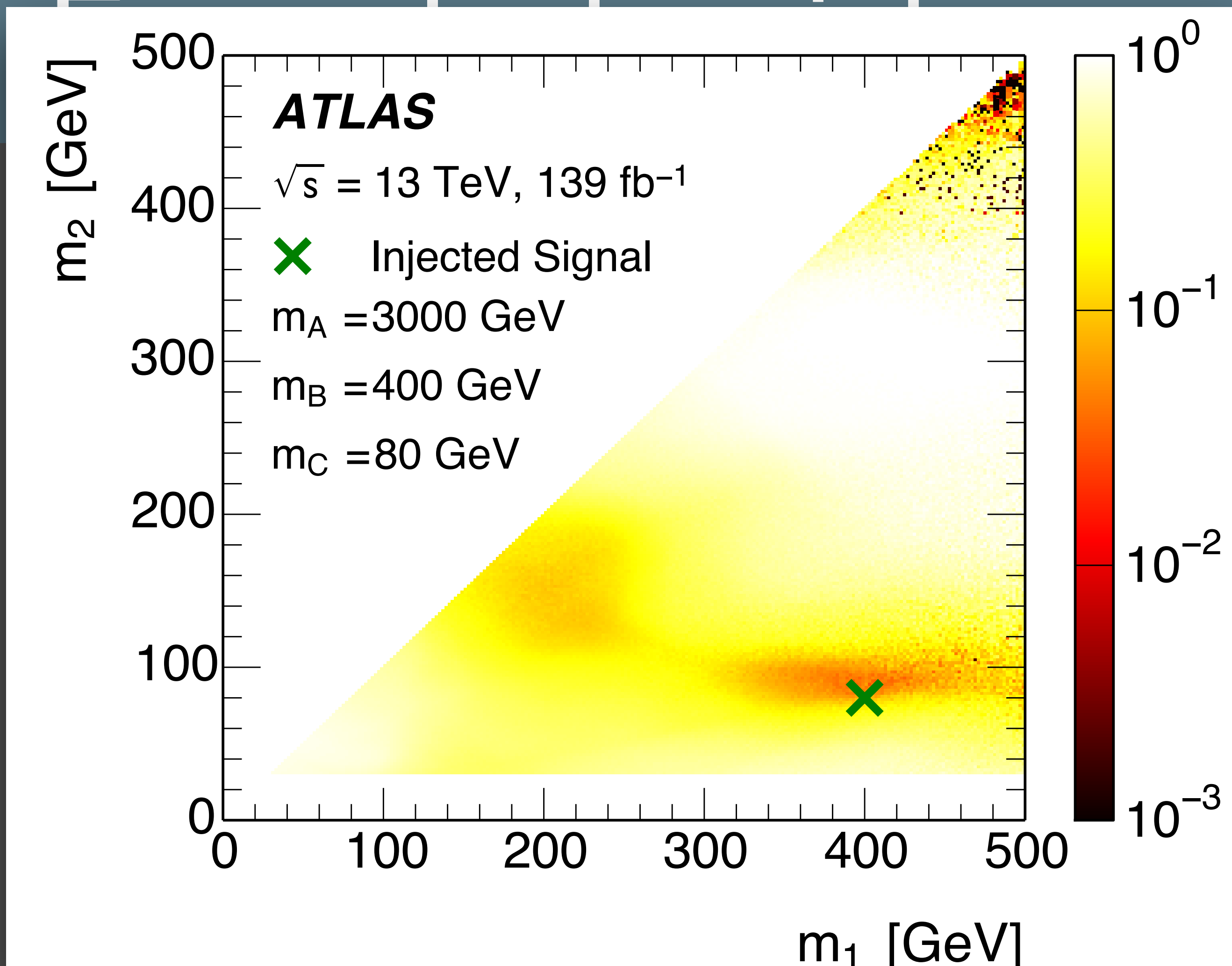


collisions in/out of page

x = many features of the two jets

See Inês' talk

res = mass of



A first version of this search has been performed by ***ATLAS***!

Phys. Rev. Lett. 125 (2020) 131801, 2005.02983

ATLAS EXPERIMENT

Collaboration Site | Physics Results

ABOUT DISCOVER RESOURCES UPDATES SEARCH

All News Briefings Features Portraits Press Blog

updates > briefing > Machine learning qualitatively changes the search for new particles

Physics Briefing

Tags: machine learning, analysis

Machine learning qualitatively changes the search for new particles

13 May 2020 | By ATLAS Collaboration

collisions in/out of page

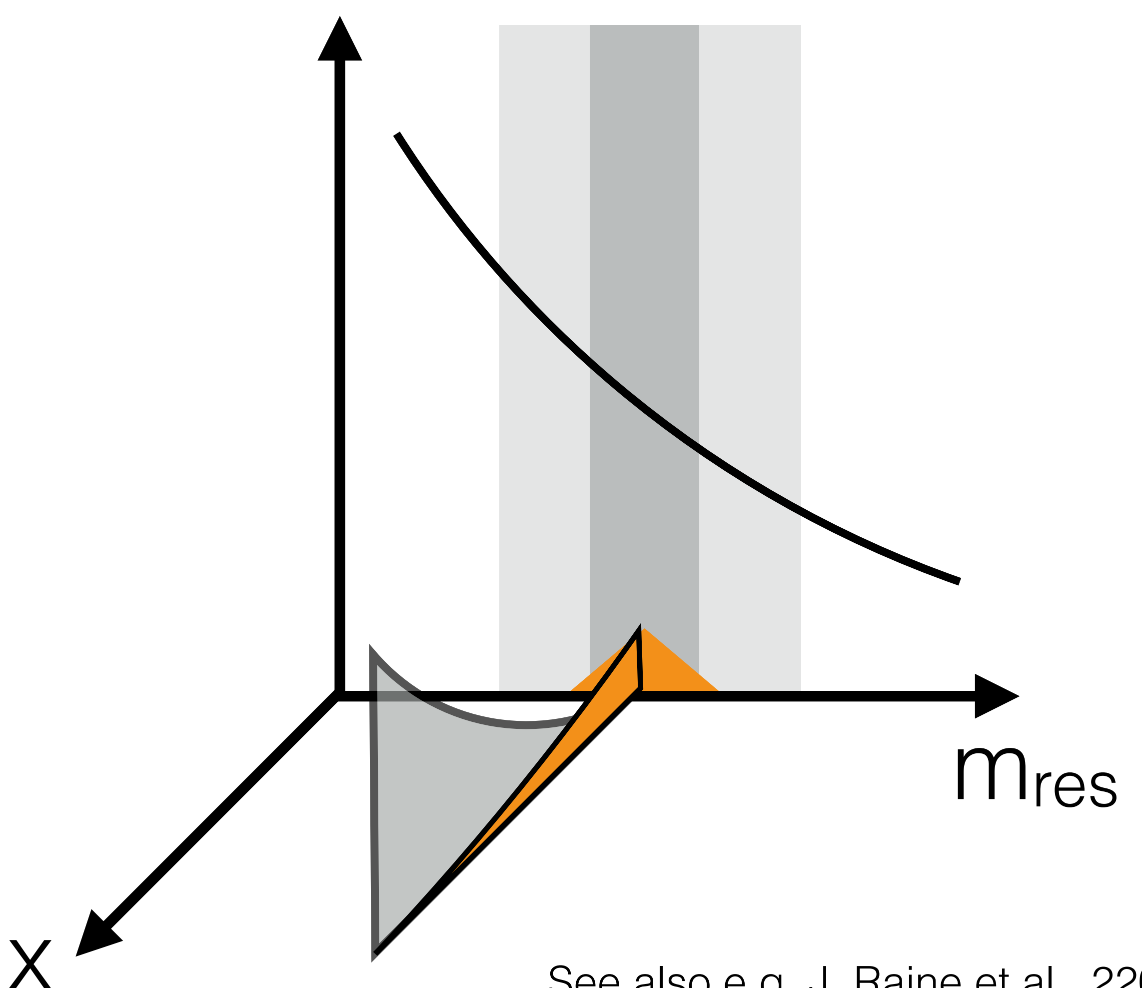
Extensions

While powerful, the approach I've just described has multiple challenges when scaling up the dimension.

Extensions

While powerful, the approach I've just described has multiple challenges when scaling up the dimension.

Example Challenge: Decorrelation



The approach doesn't work if m_{res} and x are strongly related.

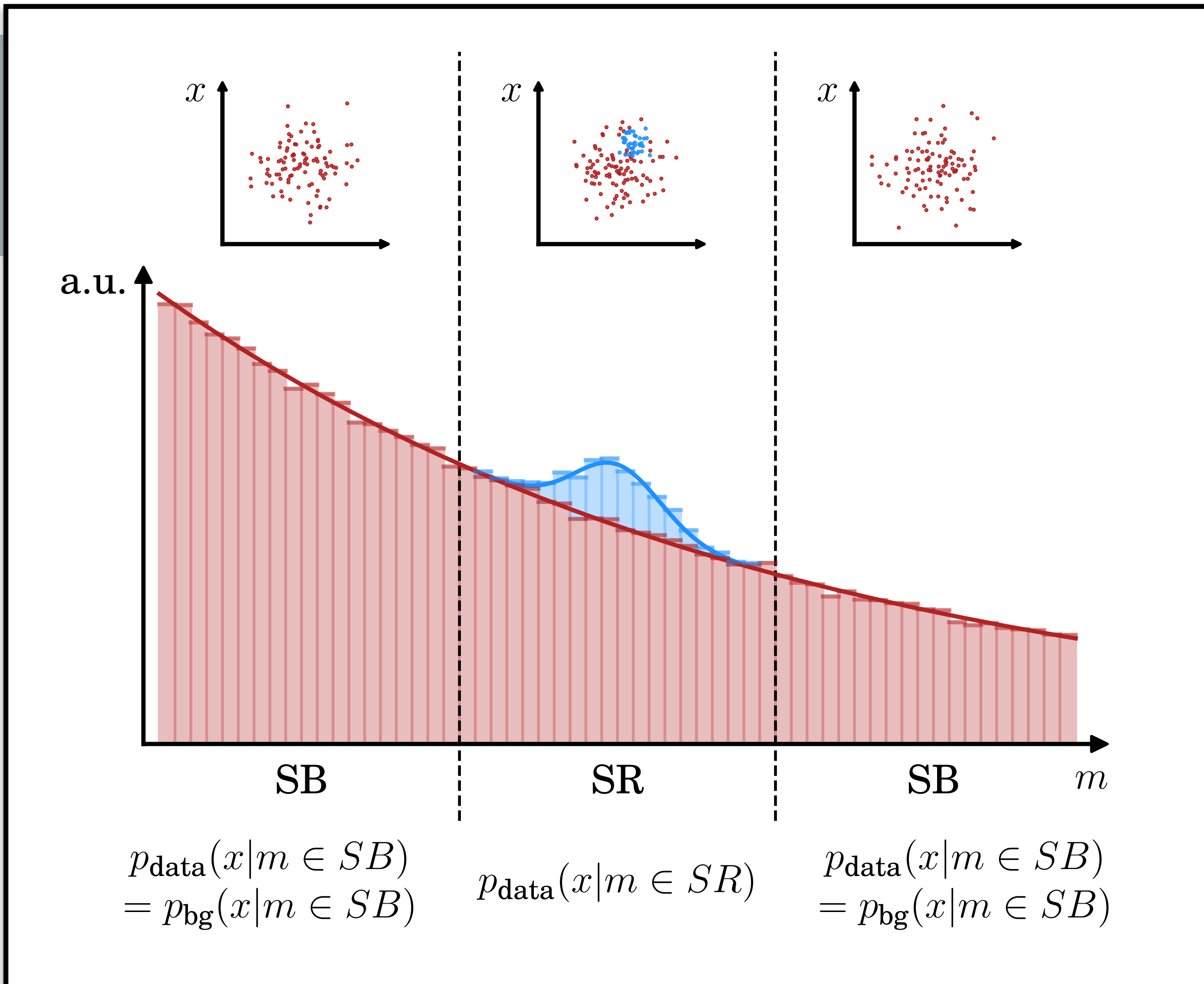
For instance, consider the extreme case where m_{res} is part of x.

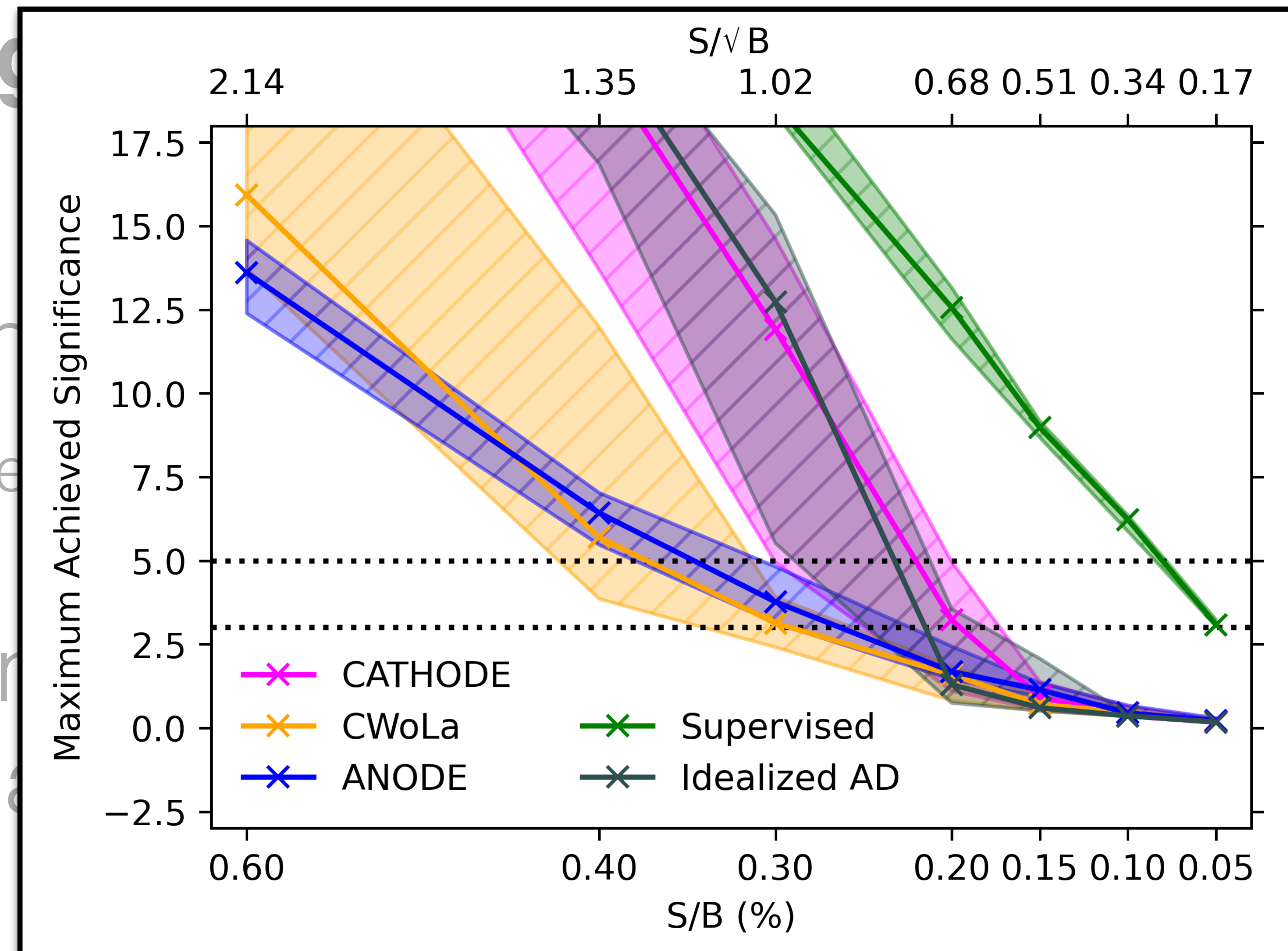
See also e.g. J. Raine et al., 2203.09470

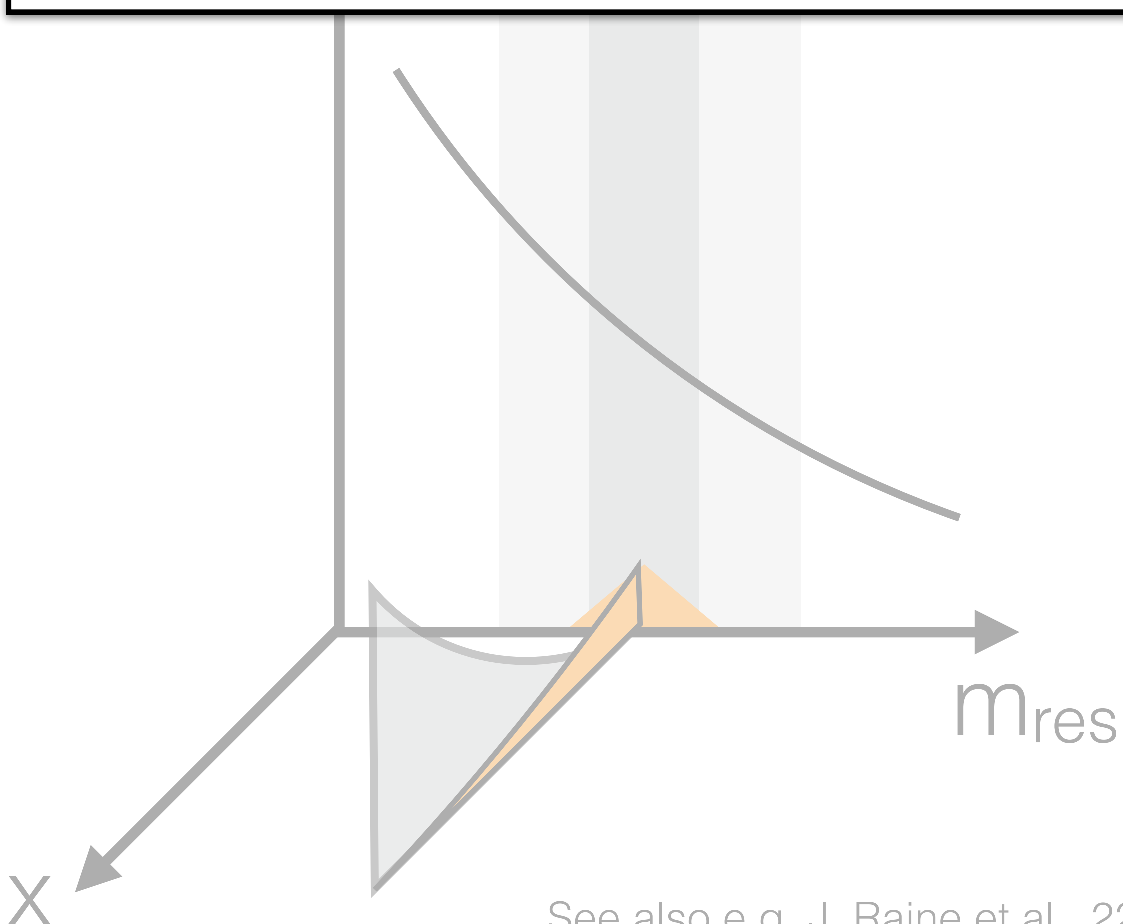
K. Benkendorfer, L. Le Pottier, BN, 2009.02205
A. Hallin et al., 2109.00546
A. Andreassen, BN, D. Shih, PRD 101 (2020) 095004, 2001.05001
BN and D. Shih, PRD 101 (2020) 075042, 2001.04990

Solution: never directly compare SR &SB!

ach I've just described has
scaling up the dimension.







See also e.g. J. Raine et al., 2203.09470

K. Benkendorfer, L. Le Pottier, BN, 2009.02205
A. Hallin et al., 2109.00546
A. Andreassen, BN, D. Shih, PRD 101 (2020) 095004, 2001.05001
BN and D. Shih, PRD 101 (2020) 075042, 2001.04990

Extensions

While powerful, the approach I've just described has multiple challenges when scaling up the dimension.

We also need to benchmark new approaches.

The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

G. Kasieczka, BN, D. Shih et al., 2101.08320

Dark Machines

The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider

T. Arrestad et al., 2105.14027

(see also ADC2021)

Extensions

While powerful, the approach I've just described has multiple challenges when scaling up the dimension.

We also need to benchmark new approaches.

The LHC Olympics 2020

A Community Challenge for Anomaly Detection in High Energy Physics

See Gregor's talk!

G. Kasieczka, BN, D. Shih et al., 2101.08320

Dark Machines

See Sascha's talk!

The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider

T. Arrestad et al., 2105.14027

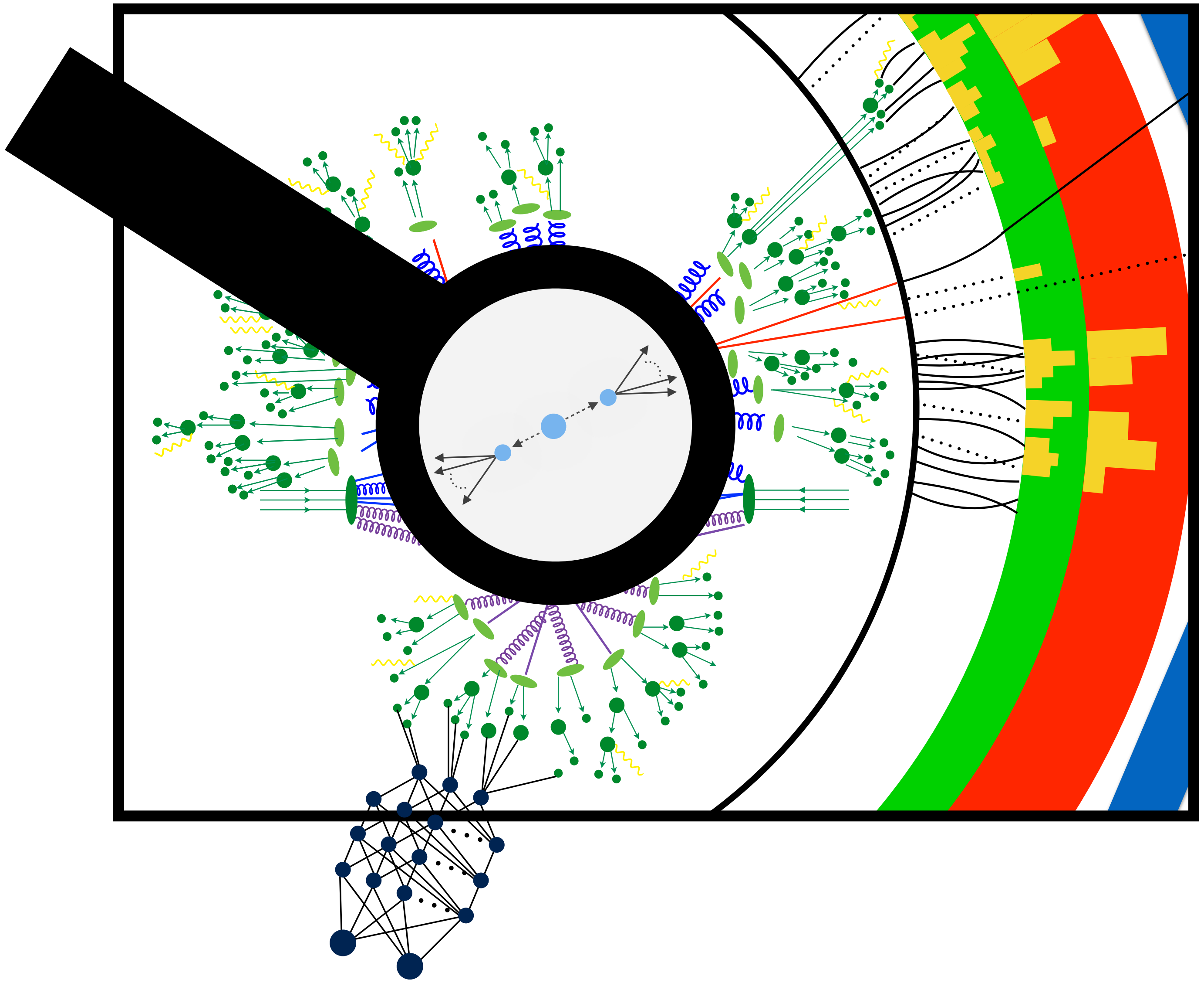
(see also ADC2021)

The Future

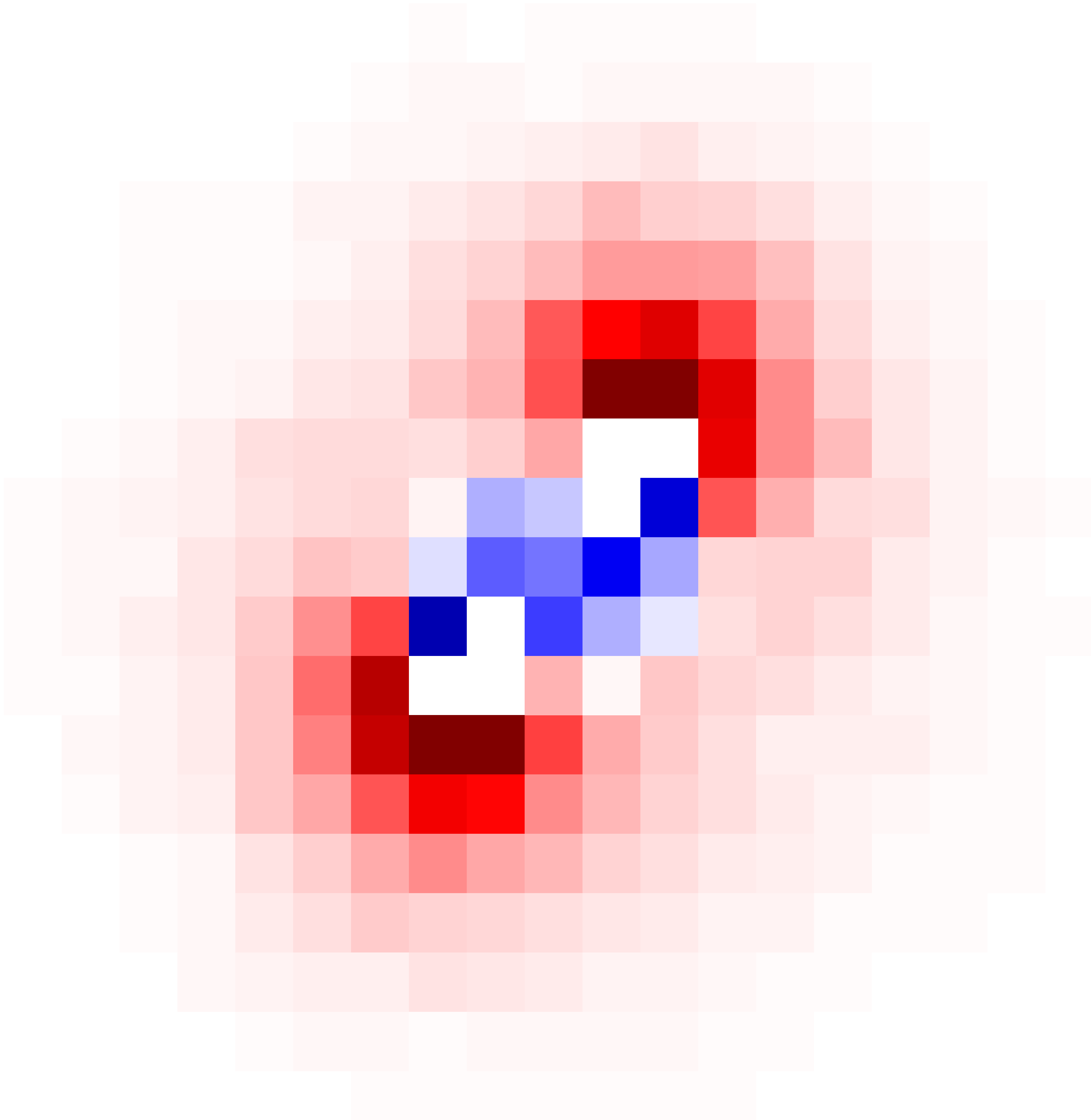
It is an exciting time to work on anomaly detection for the LHC and beyond!

This is a rapidly growing area with lots of room for innovation (and from physicists!)

We will need many approaches to achieve broad coverage



See the Living Review for more refs!



Fin.

Backup

Results with data