Modern machine learning methods for new physics searches

Sapienta ex machina?

UCLouvain

14th NExT PhD workshop — Cosener's House 2023

Ramon Winterhalder — UCLouvain

Plan of attack

Today

1. Introduction to Machine Learning

- Classification and Regression
- Example: **MadMiner, Top Tagging**

Lecture I (60min)

COFFEE BREAK

2. Generative Models for the LHC

- Diffusion Models, Normalizing flows
- Example: **MadNIS, MEM-ML**

Lecture II (60min)

Tomorrow

3. Anomaly detection

- Autoencoders and CWoLA
- Examples: **CWoLA-Hunting, ANODE, CATHODE,…**

Lecture III (60min)

Plan of attack

Today

1. Introduction to Machine Learning

- Classification and Regression
- Example: **MadMiner, Top Tagging**

Lecture I (60min)

COFFEE BREAK

2. Generative Models for the LHC

- Diffusion Models, Normalizing flows
- Example: **MadNIS, MEM-ML**

Lecture II (60min)

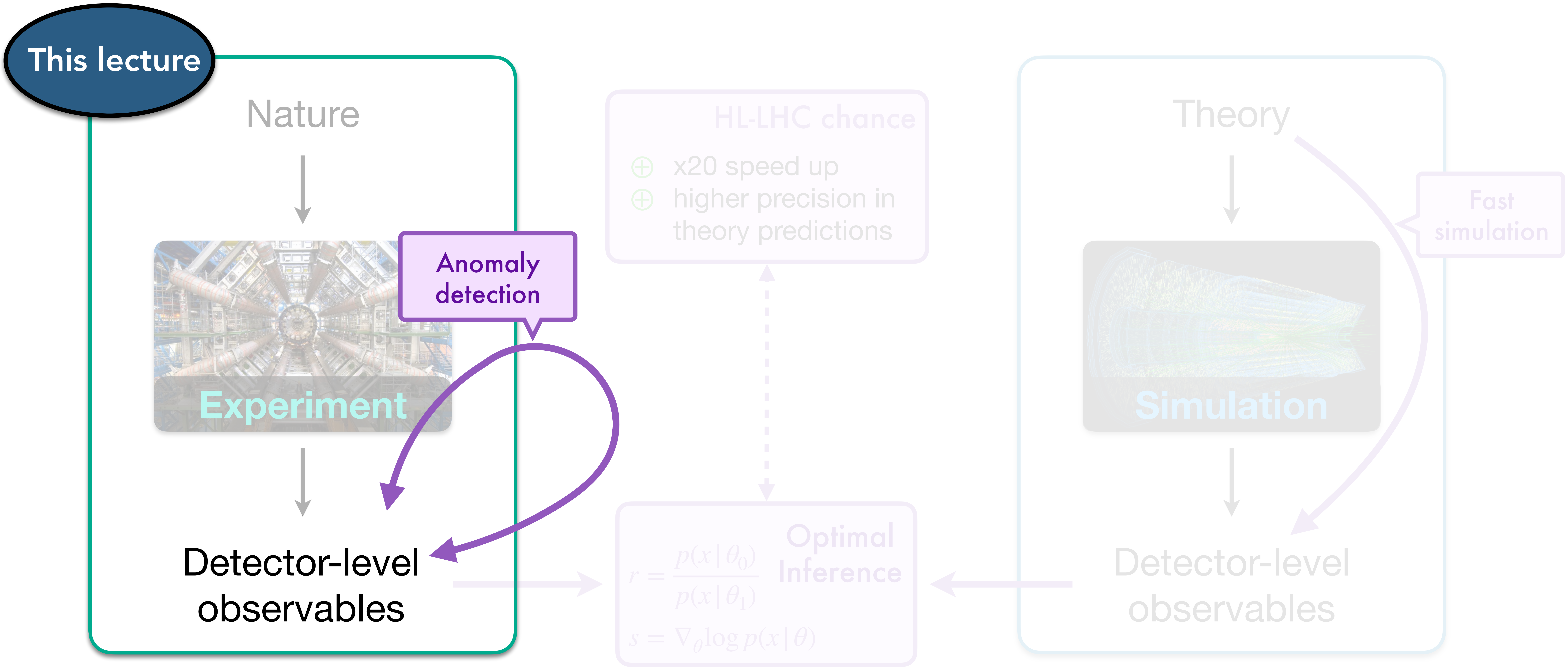
Tomorrow

3. Anomaly detection

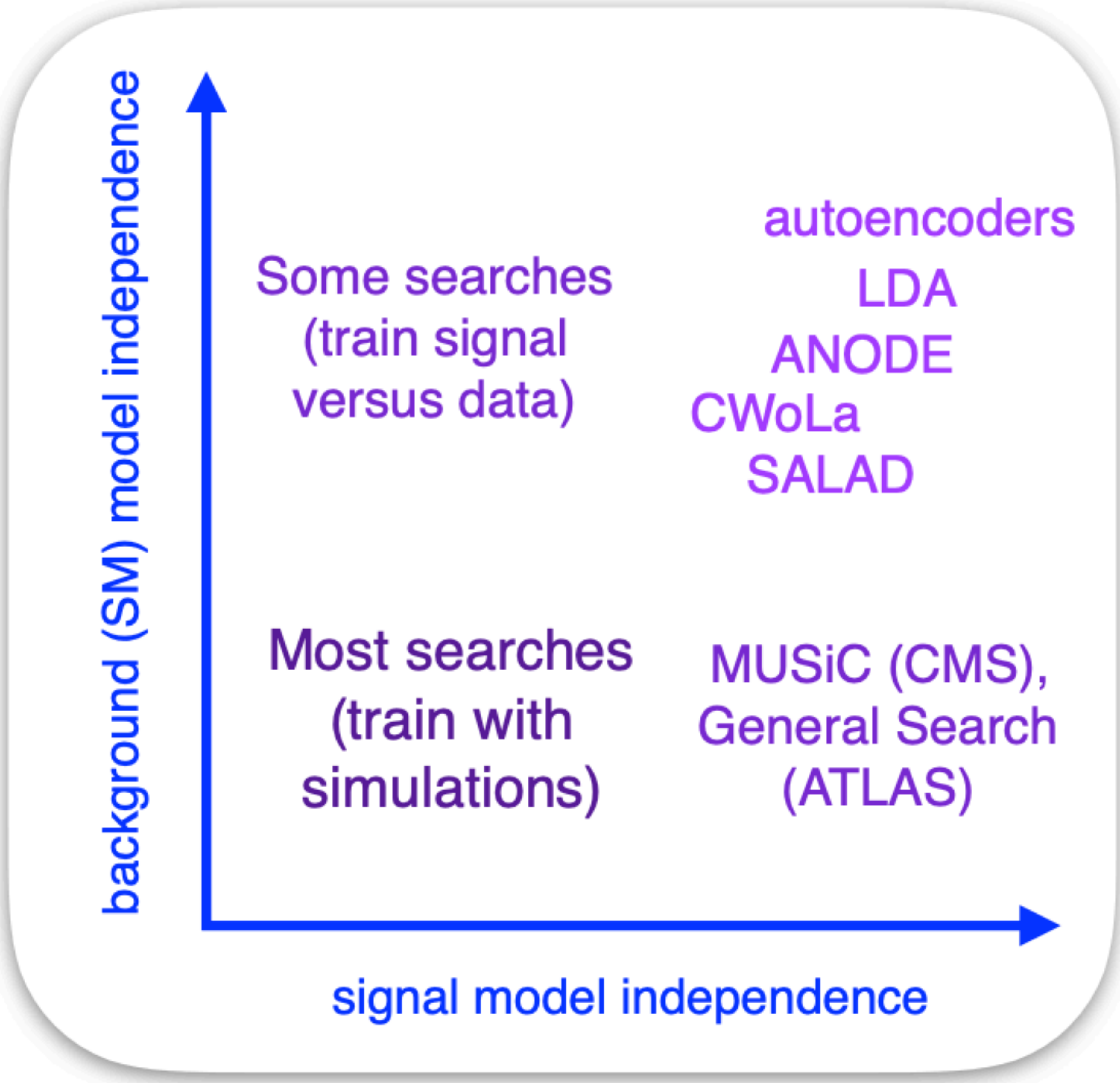
- Autoencoders and CWoLA
- Examples: **CWoLA-Hunting, ANODE, CATHODE,…**

Lecture III (60min)

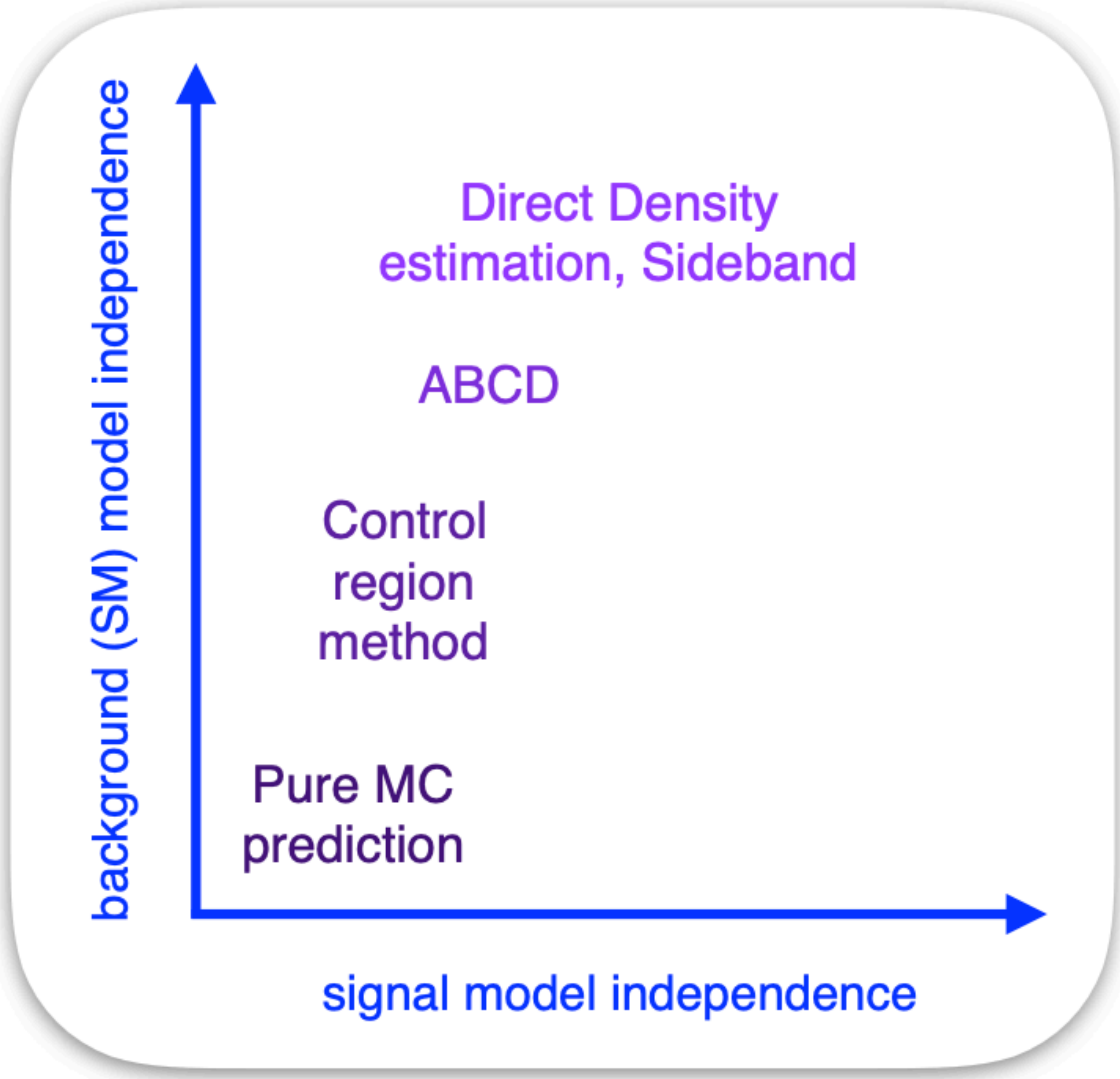
Reminder — LHC analysis



Simulation or data-driven searches



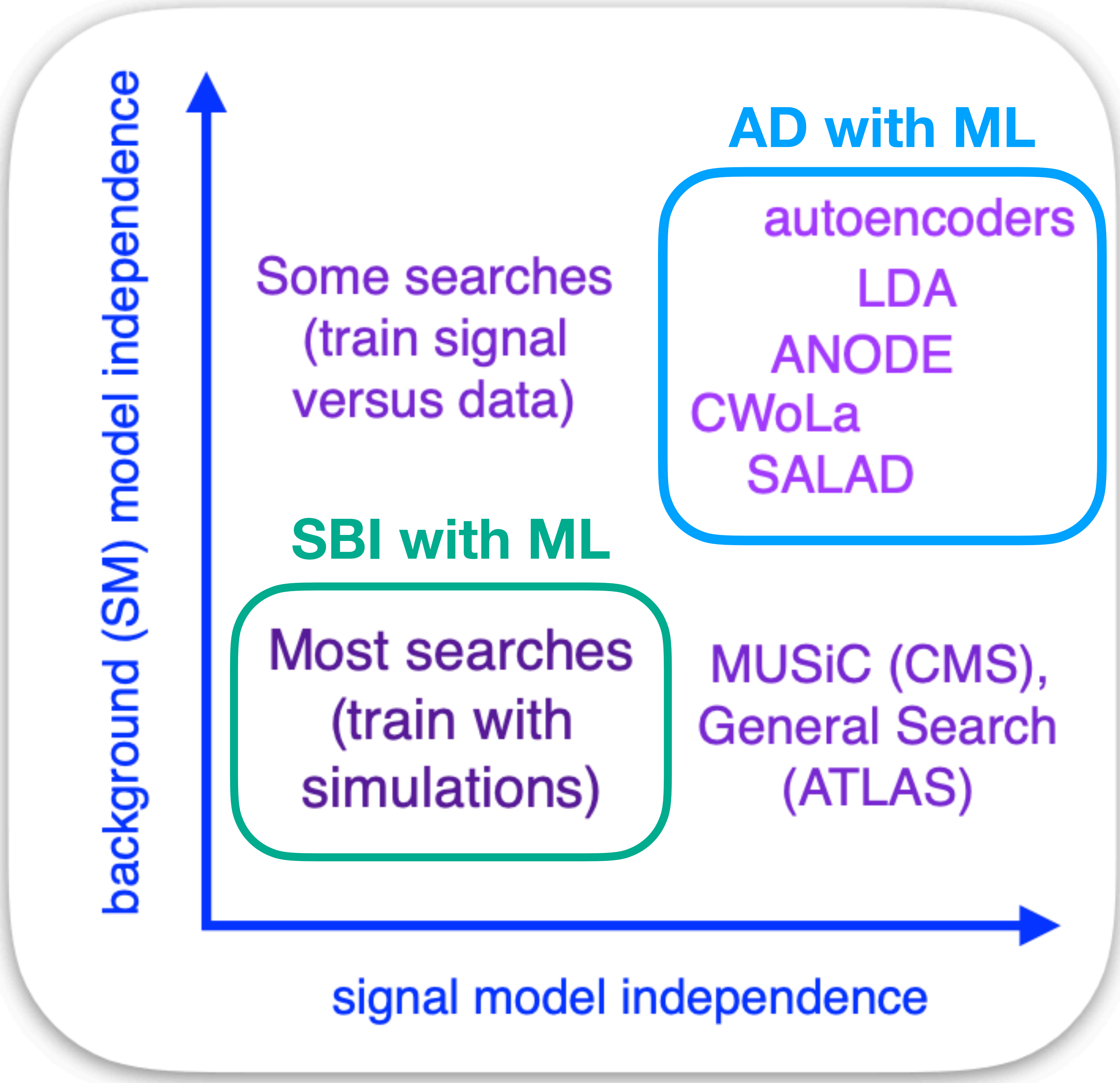
(a) Signal sensitivity



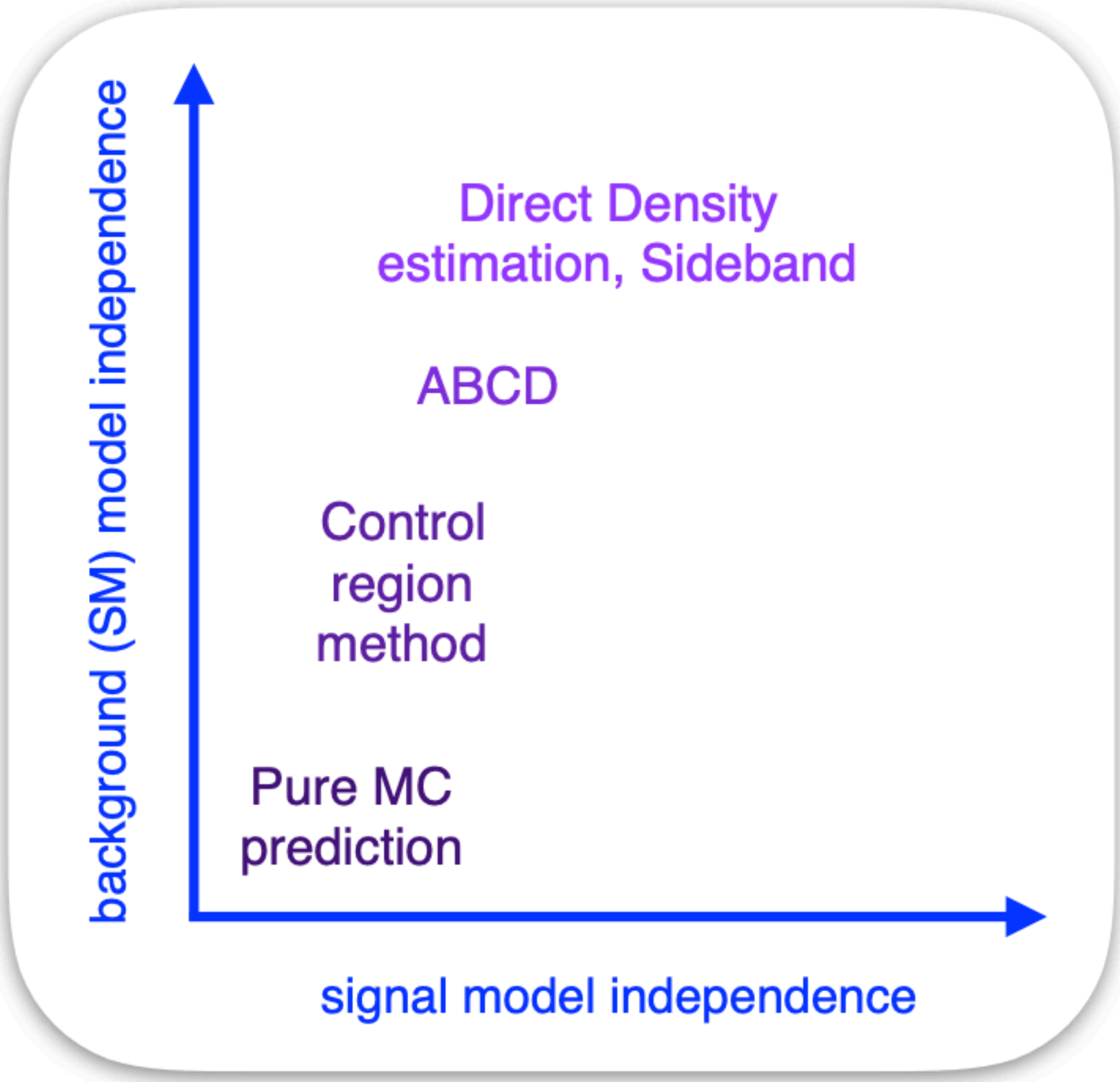
(b) Background specificity

***Taken from [Nachman *et al*: 2001.04990]**

Simulation or data-driven searches



(a) Signal sensitivity



(b) Background specificity

*Taken from [Nachman *et al*: 2001.04990]

Machine Learning
Supervised Learning
Classification
Point Clouds [2102.05073]
PELICAN [2211.00454]
Energy Flow Networks [1810.05165]
Landscape of Top tagger [1902.09914]
Bayesian Tagger [1904.10004]
Regression
MadMiner [1907.10621,…]
NNPDF [2109.02653]
Matrix Elements [2206.14831]
Symbolic regression [2109.10414]
Reinforcement Learning
Flavor structure [2304.14176]
Simplifying Polylogs [2206.04115]
String vacua & landscape [1903.11616, 2111.11466]
Model Building [2103.04759]
Analytic continuation [2112.13011]
Unsupervised Learning
Clustering
Jet Clustering [2008.06064]
3D Pixel Clustering [2007.03083]
Anomaly Detection
Normalized AE [2206.14225]
CATHODE [2109.00546]
CWoLA Hunting [1902.02634]
(R-)ANODE [2001.04990,…]
Generative Models
NF
Precision Generation [2110.13632]
MEM [2210.00019]
MADNIS [2212.06172,..]
CaloFlow I-IV [2106.05285,…]
ELSA [2305.07696]
VAE
OTUS [2101.08944]
Jet Simulation [2203.00520]
GAN
EPiC-GAN [2301.08128]
How to GAN [1907.03764]
CaloGAN [1712.10321]
Transformer
JetGPT [2305.10475]
Diffusion Models
PC-JeDi [2303.05376]
FPCD [2304.01266]
DDPM & CFM [2305.10475]

Machine Learning
Supervised Learning
Classification
Point Clouds [2102.05073]
PELICAN [2211.00454]
Energy Flow Networks [1810.05165]
Landscape of Top tagger [1902.09914]
Bayesian Tagger [1904.10004]
Regression
MadMiner [1907.10621,…]
NNPDF [2109.02653]
Matrix Elements [2206.14831]
Symbolic regression [2109.10414]
Reinforcement Learning
Flavor structure [2304.14176]
Simplifying Polylogs [2206.04115]
String vacua & landscape [1903.11616, 2111.11466]
Model Building [2103.04759]
Analytic continuation [2112.13011]
Unsupervised Learning
Generative Models
NF
Precision Generation [2110.13632]
MEM [2210.00019]
MadNIS [2212.06172,..]
CaloFlow I-IV [2106.05285,…]
ELSA [2305.07696]
VAE
OTUS [2101.08944]
Jet Simulation [2203.00520]
Diffusion Models
PC-JeDi [2303.05376]
FPCD [2304.01266]
DDPM & CFM [2305.10475]
GAN
EPiC-GAN [2301.08128]
How to GAN [1907.03764]
CaloGAN [1712.10321]
Transformer
JetGPT [2305.10475]
Clustering
Jet Clustering [2008.06064]
3D Pixel Clustering [2007.03083]
Anomaly Detection
Normalized AE [2206.14225]
CATHODE [2109.00546]
CWoLA Hunting [1902.02634]
(R-)ANODE [2001.04990,…]

Part III

Anomaly Detection

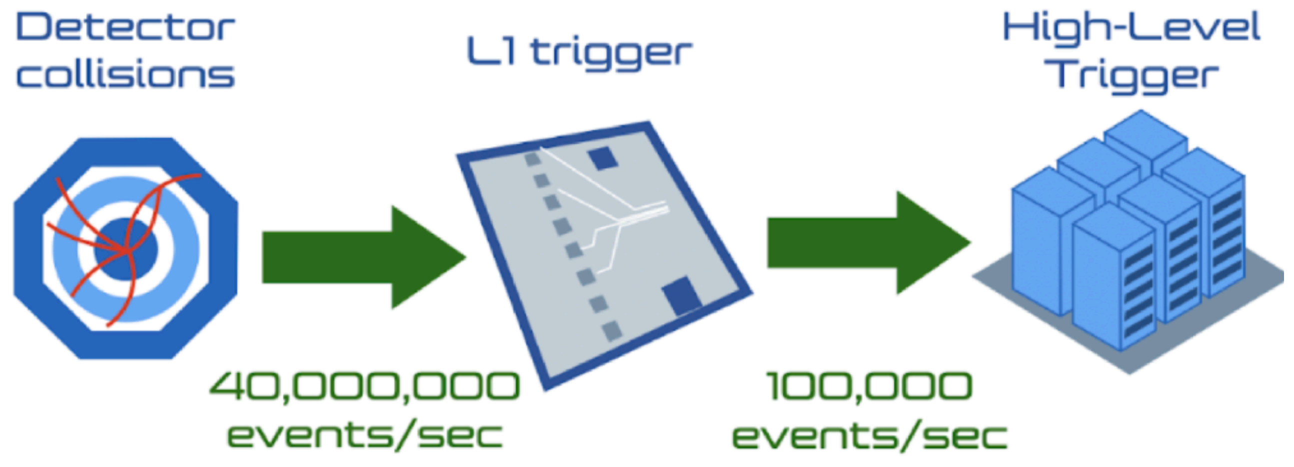
Community interest in AD

LHC Olympics

[Kasieczka et al: 2107.02821, 2101.08320]

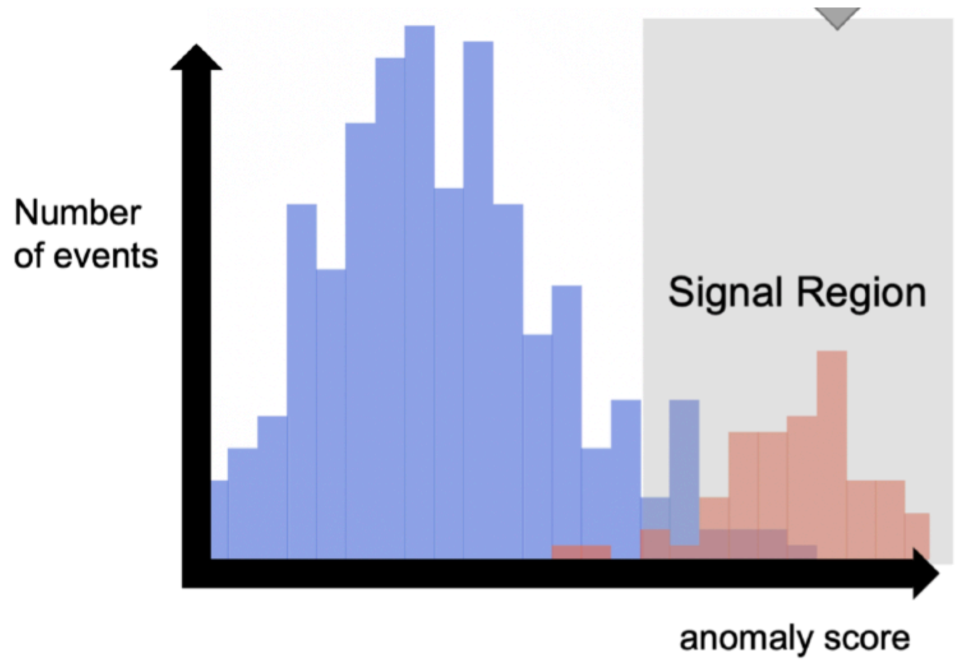
ADC2021

[Govorkova et al: 2107.02157]



Dark Machines

[Ostdiek et al: 2105.14027]



…and **many** papers:

Anomaly detection.

- Learning New Physics from a Machine [DOI]
- Anomaly Detection for Resonant New Physics with Machine Learning [DOI]
- Extending the search for new resonances with machine learning [DOI]
- Learning Multivariate New Physics [DOI]
- Searching for New Physics with Deep Autoencoders [DOI]
- QCD or What? [DOI]
- A robust anomaly finder based on autoencoder
- Variational Autoencoders for New Physics Mining at the Large Hadron Collider [DOI]
- Adversarially-trained autoencoders for robust unsupervised new physics searches [DOI]
- Novelty Detection Meets Collider Physics [DOI]
- Guiding New Physics Searches with Unsupervised Learning [DOI]
- Does SUSY have friends? A new approach for LHC event analysis [DOI]
- Nonparametric semisupervised classification for signal detection in high energy physics
- Uncovering latent jet substructure [DOI]
- Simulation Assisted Likelihood-free Anomaly Detection [DOI]
- Anomaly Detection with Density Estimation [DOI]
- A generic anti-QCD jet tagger [DOI]
- Transferability of Deep Learning Models in Searches for New Physics at Colliders [DOI]
- Use of a Generalized Energy Mover's Distance in the Search for Rare Phenomena at Colliders [DOI]
- Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark [DOI]
- Dijet resonance search with weak supervision using 13 TeV pp collisions in the ATLAS detector [DOI]

https://iml-wg.github.io/HEPML-LivingReview

Community interest in AD

Machine Learning for Anomaly Detection in Particle Physics

Vasilis Belis[1,*], **Patrick Odagiu**[1,*], **and Thea Klæboe Aarrestad**[1,*]

[1]Institute for Particle Physics and Astrophysics, ETH Zürich, 8093 Zürich, Switzerland
[*]e-mail: vbelis@ethz.ch, podagiu@ethz.ch, thea.aarrestad@cern.ch

ABSTRACT

The detection of out-of-distribution data points is a common task in particle physics. It is used for monitoring complex particle detectors or for identifying rare and unexpected events that may be indicative of new phenomena or physics beyond the Standard Model. Recent advances in Machine Learning for anomaly detection have encouraged the utilization of such techniques on particle physics problems. This review article provides an overview of the state-of-the-art techniques for anomaly detection in particle physics using machine learning. We discuss the challenges associated with anomaly detection in large and complex data sets, such as those produced by high-energy particle colliders, and highlight some of the successful applications of anomaly detection in particle physics experiments.

[2312.14190]

Dark Machines

[Ostdiek et al: 2105.14027]

Available on the CERN CDS information server | CMS PAS EXO-22-026

CMS Physics Analysis Summary

Contact: cms-pag-conveners-exotica@cern.ch 2024/03/20

Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS Collaboration

Abstract

This note introduces a model-agnostic search for new physics in the dijet final state. Other than the requirement of a narrow dijet resonance with a mass in the range of 1800-6000 GeV, minimal additional assumptions are placed on the signal hypothesis. Search regions are obtained by utilizing multivariate machine learning methods to select jets with anomalous substructure. A collection of complementary anomaly detection methods – based on unsupervised, weakly-supervised and semi-supervised algorithms – are used in order to maximize the sensitivity to unknown new physics signatures. These algorithms are applied to data corresponding to an integrated luminosity of 138 fb^{-1}, recorded in the years 2016 to 2018 by the CMS experiment at the LHC, at a centre-of-mass energy of 13 TeV. No significant excesses above background expectation are seen, and exclusion limits are derived on the production cross section of benchmark signal models varying in resonance mass, jet mass and jet substructure. Many of these signatures have not previously been searched for at the LHC, making the limits reported on the corresponding benchmark models the first ever and the most stringent to date.

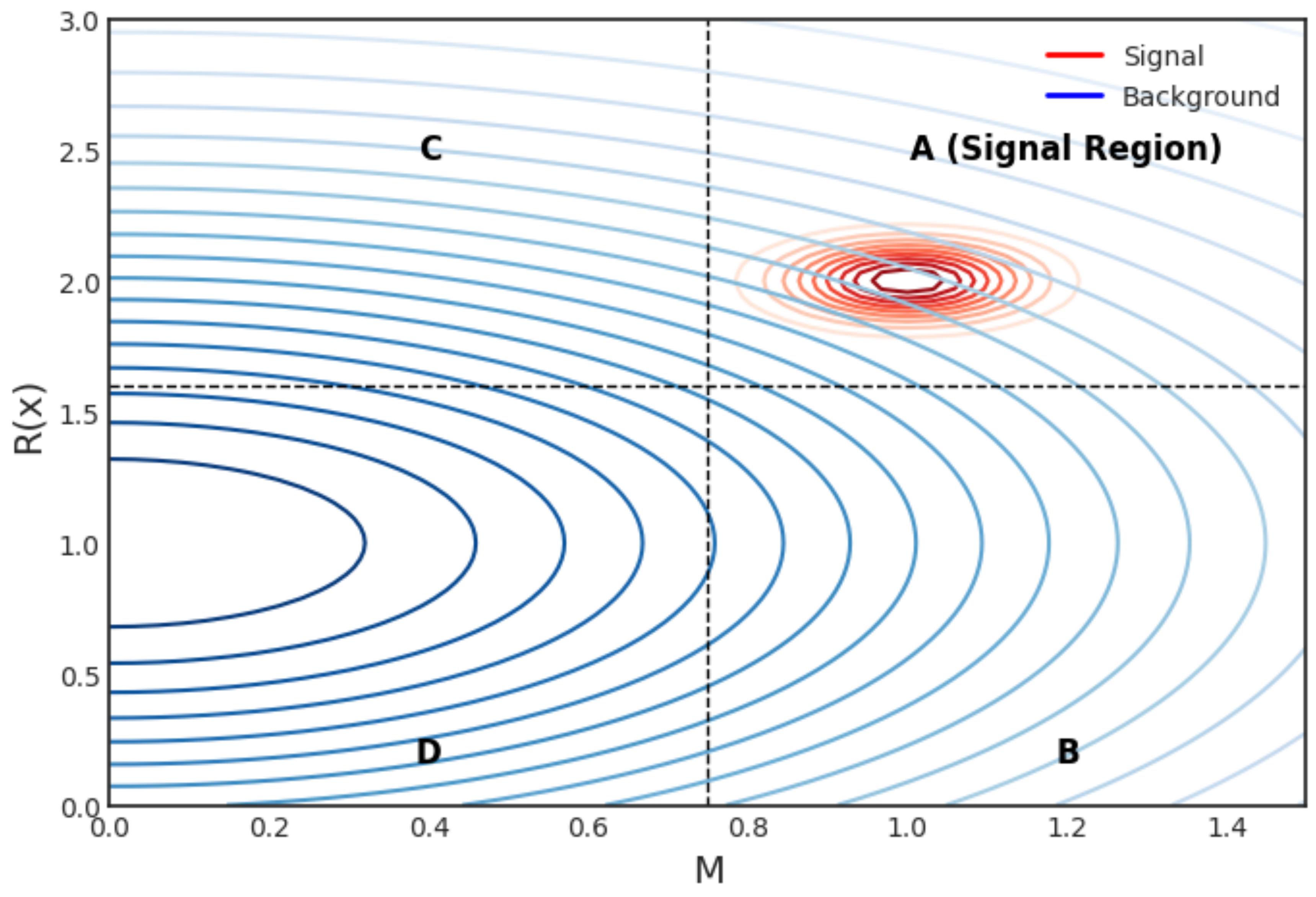
[CMS-PAS-EXO-22-026]

Two Types of Anomaly Detection

Outlier Detection
(non-resonant)

- Searching for unique and unexpected events
- In HEP, this (might) appear in the tails of dist.

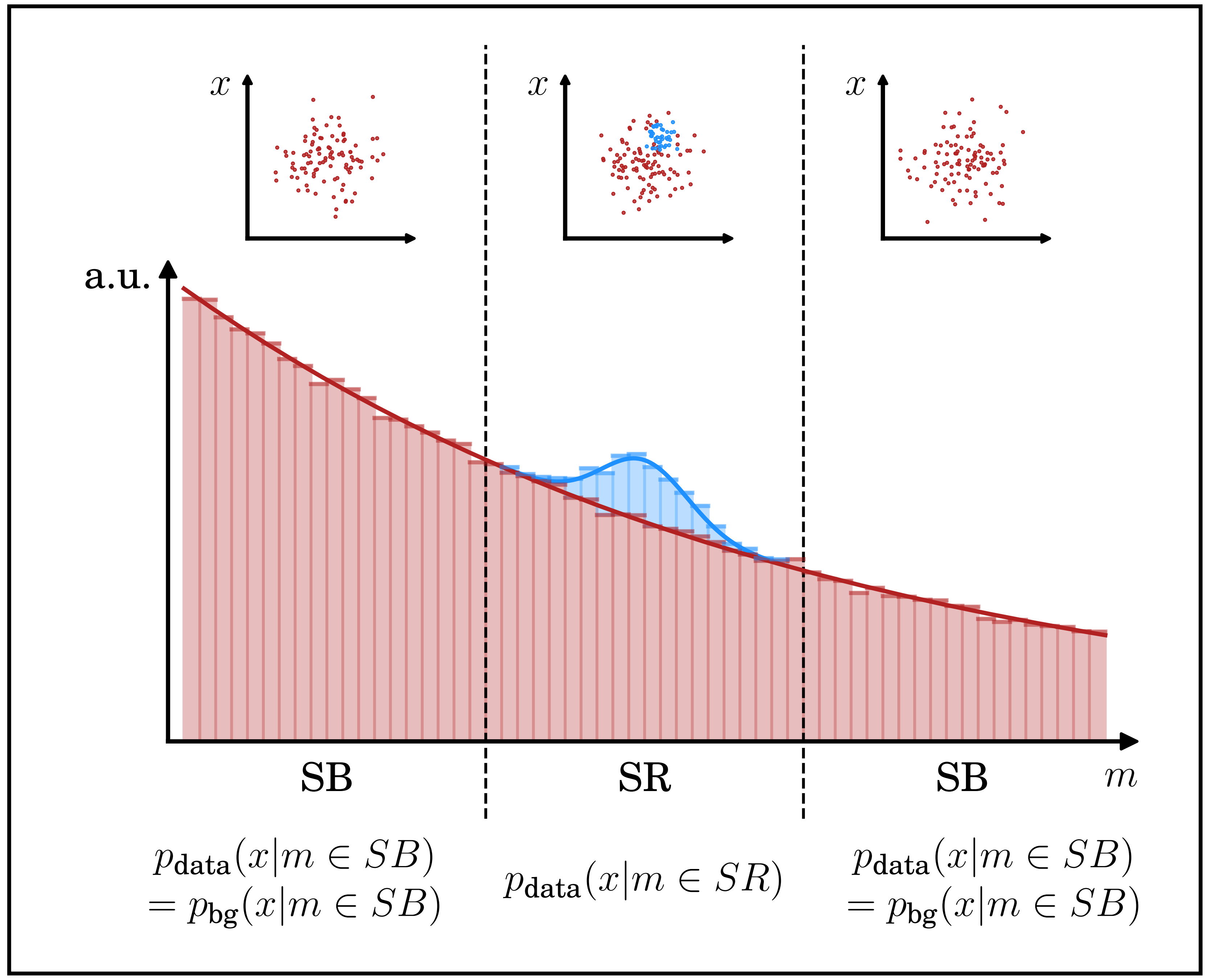
[2404.07258]



Overdensities
(resonant)

- Analagous to traditional bump hunt

[2109.00546]



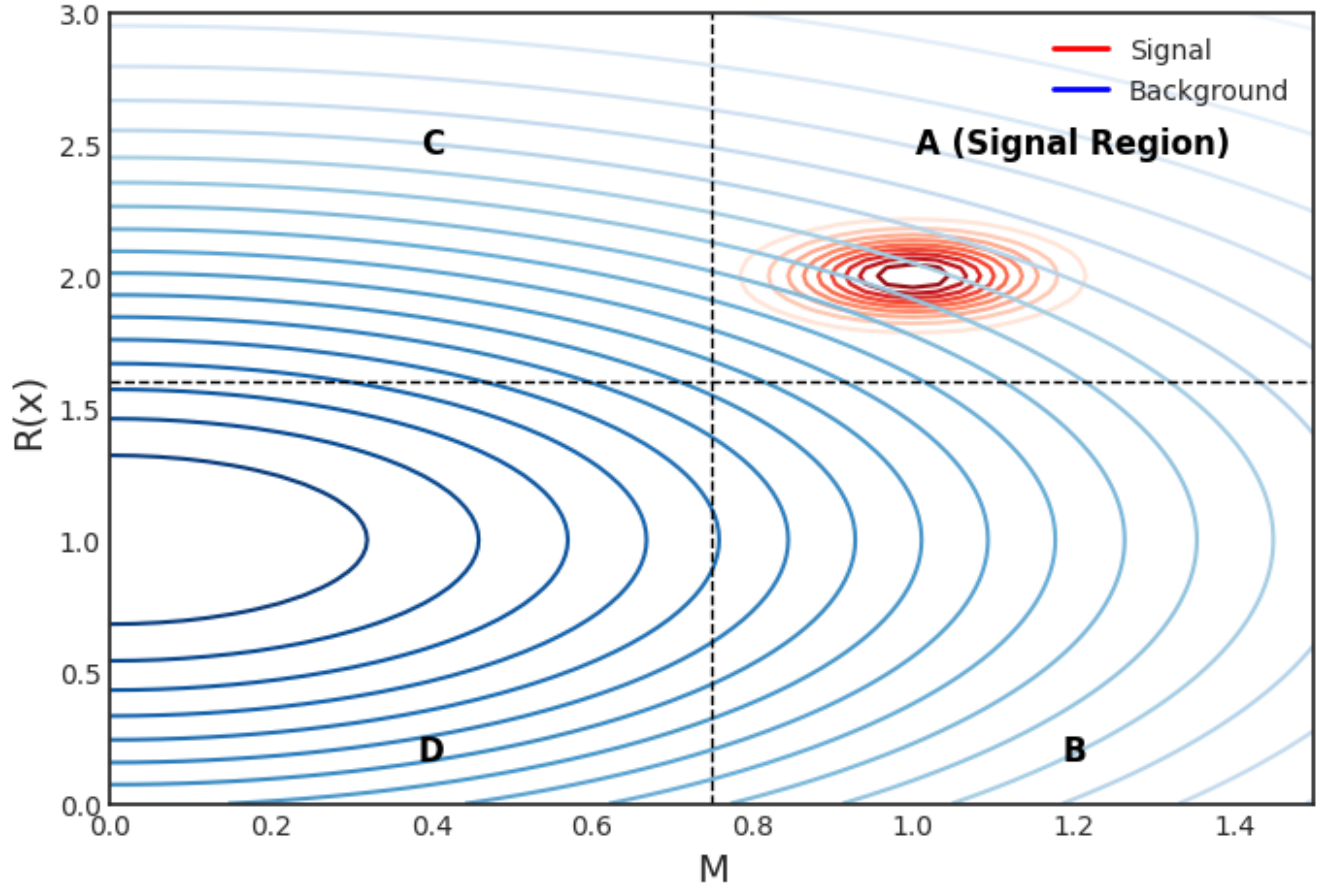
Two Types of Anomaly Detection

Outlier Detection

(non-resonant)

- Searching for unique and unexpected events
- In HEP, this (might) appear in the tails of dist.

[2404.07258]



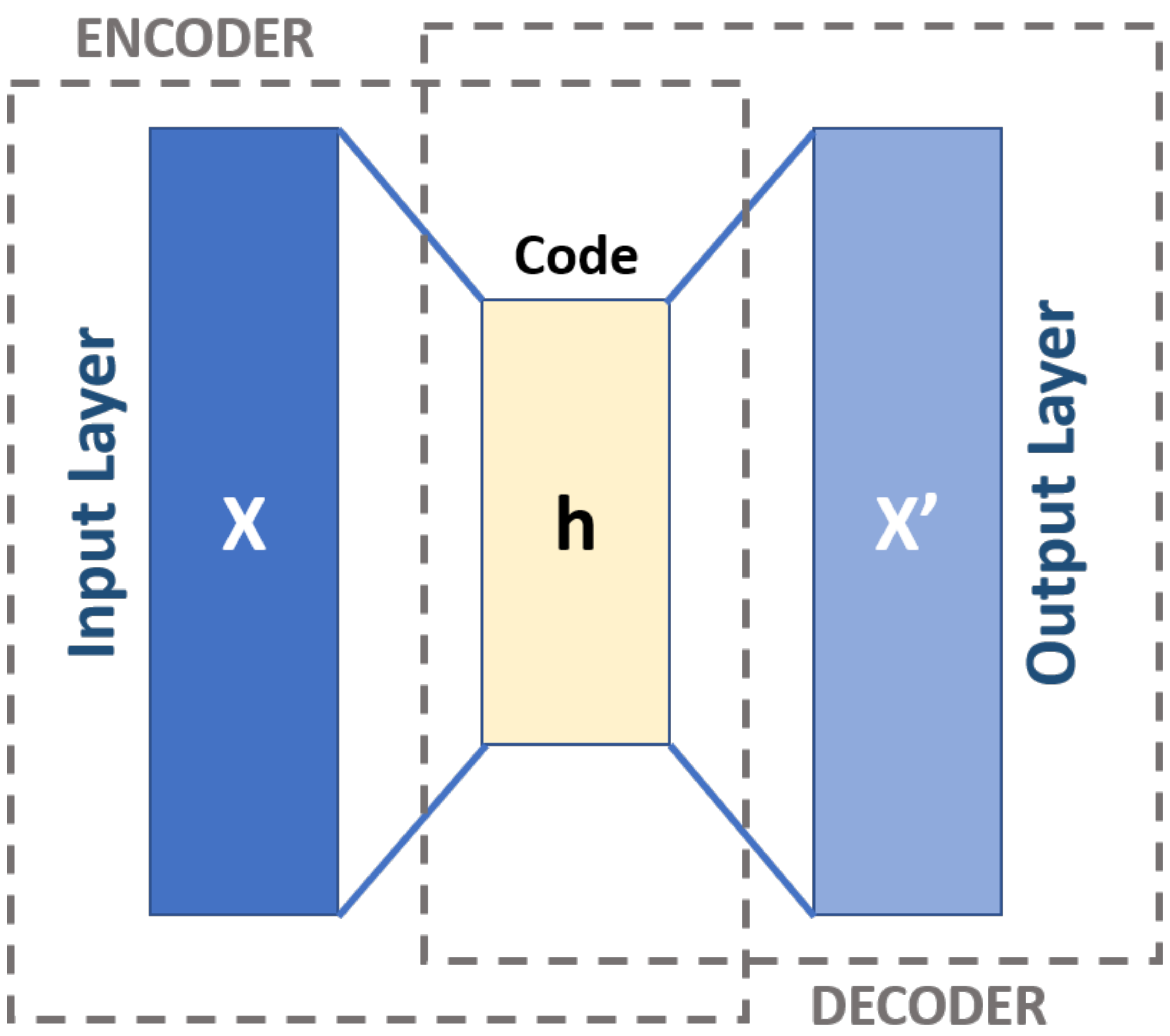
Overdensities

(resonant)

- Analagous to traditional bump hunt

[2109.00546]

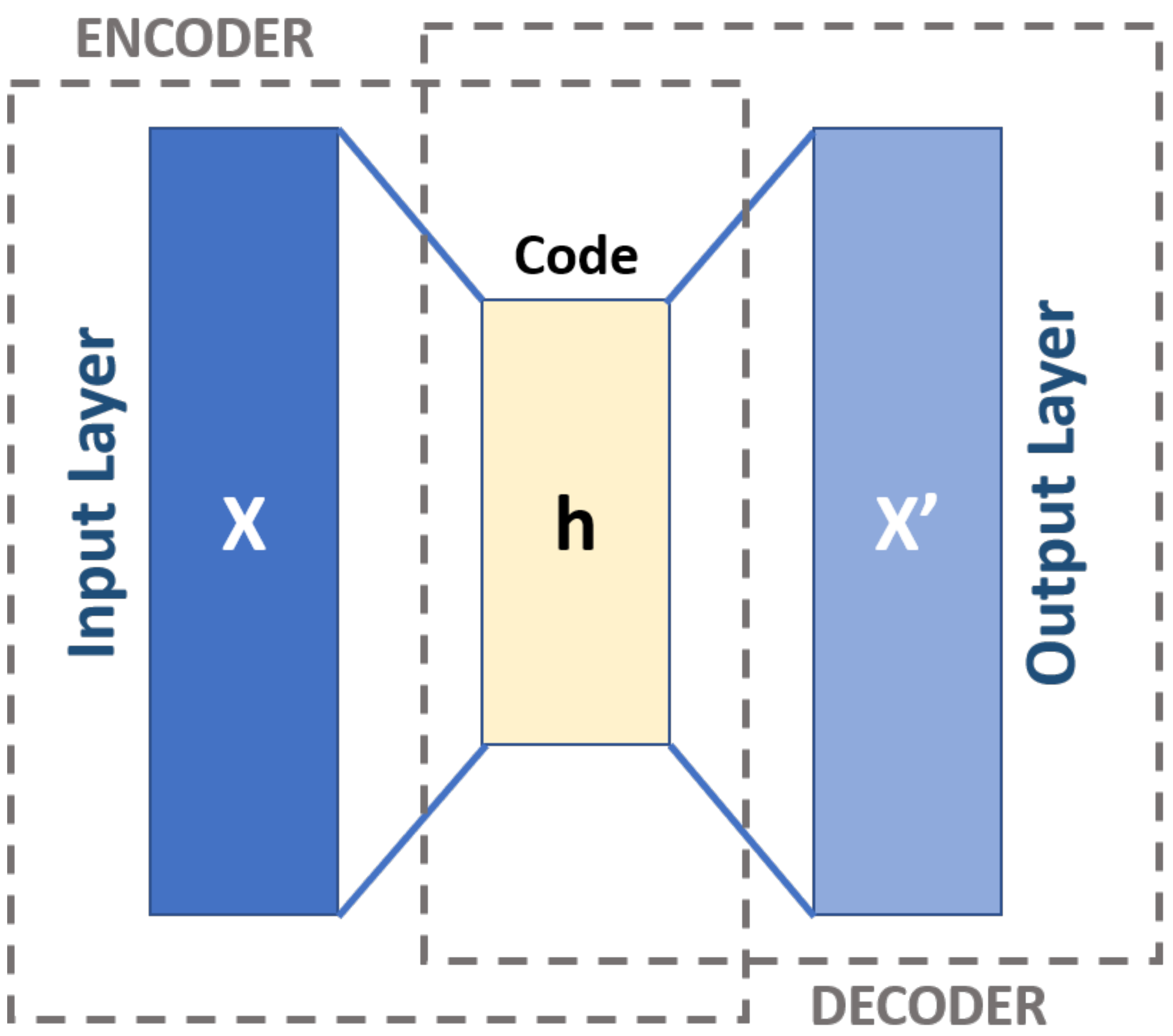
Autoencoder for non-resonant AD



- AE trained on bg.

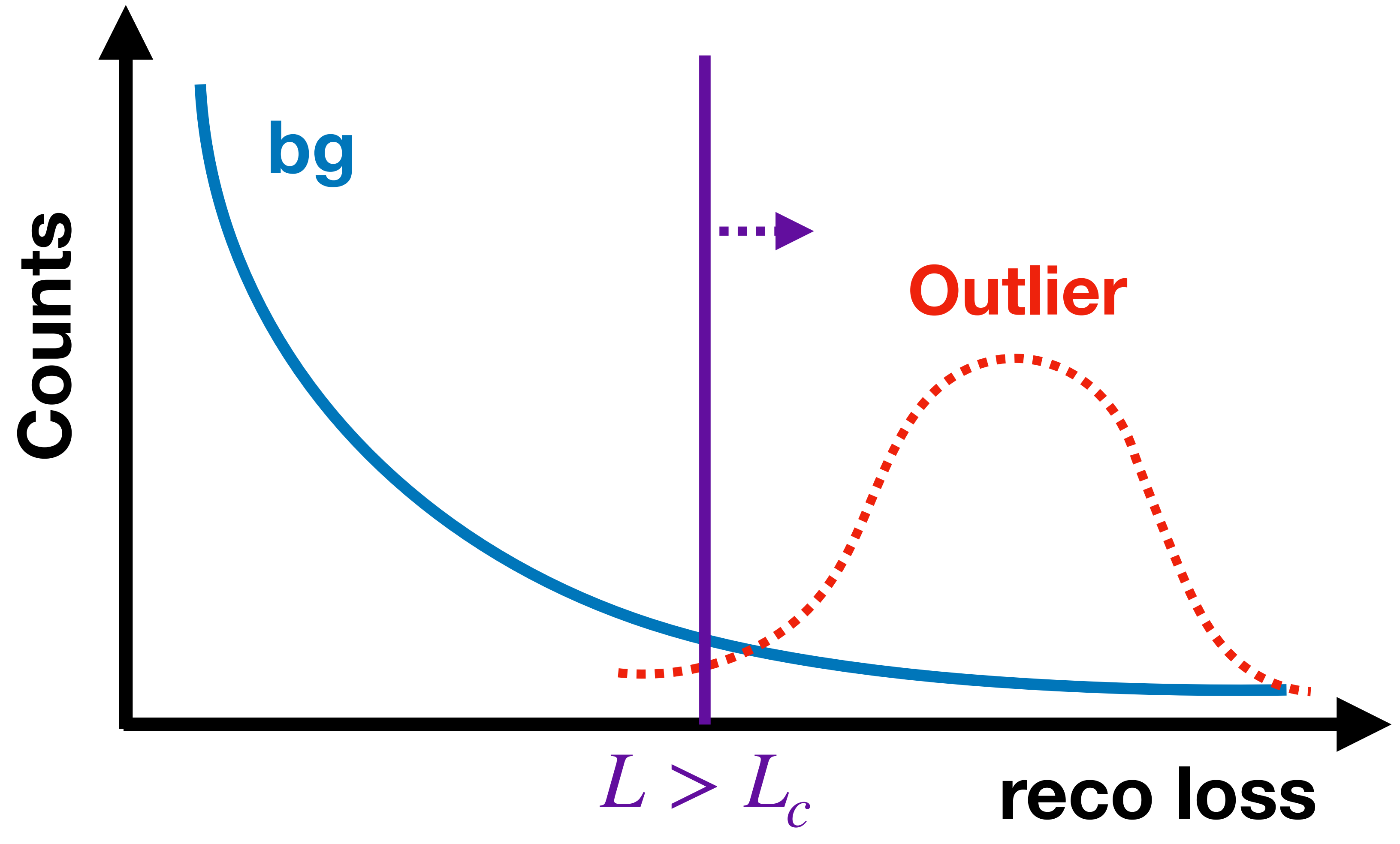
$$L = \frac{1}{N} \sum_i (\mathrm{AE}(x_i) - x_i)^2$$

Autoencoder for non-resonant AD



- AE trained on bg.

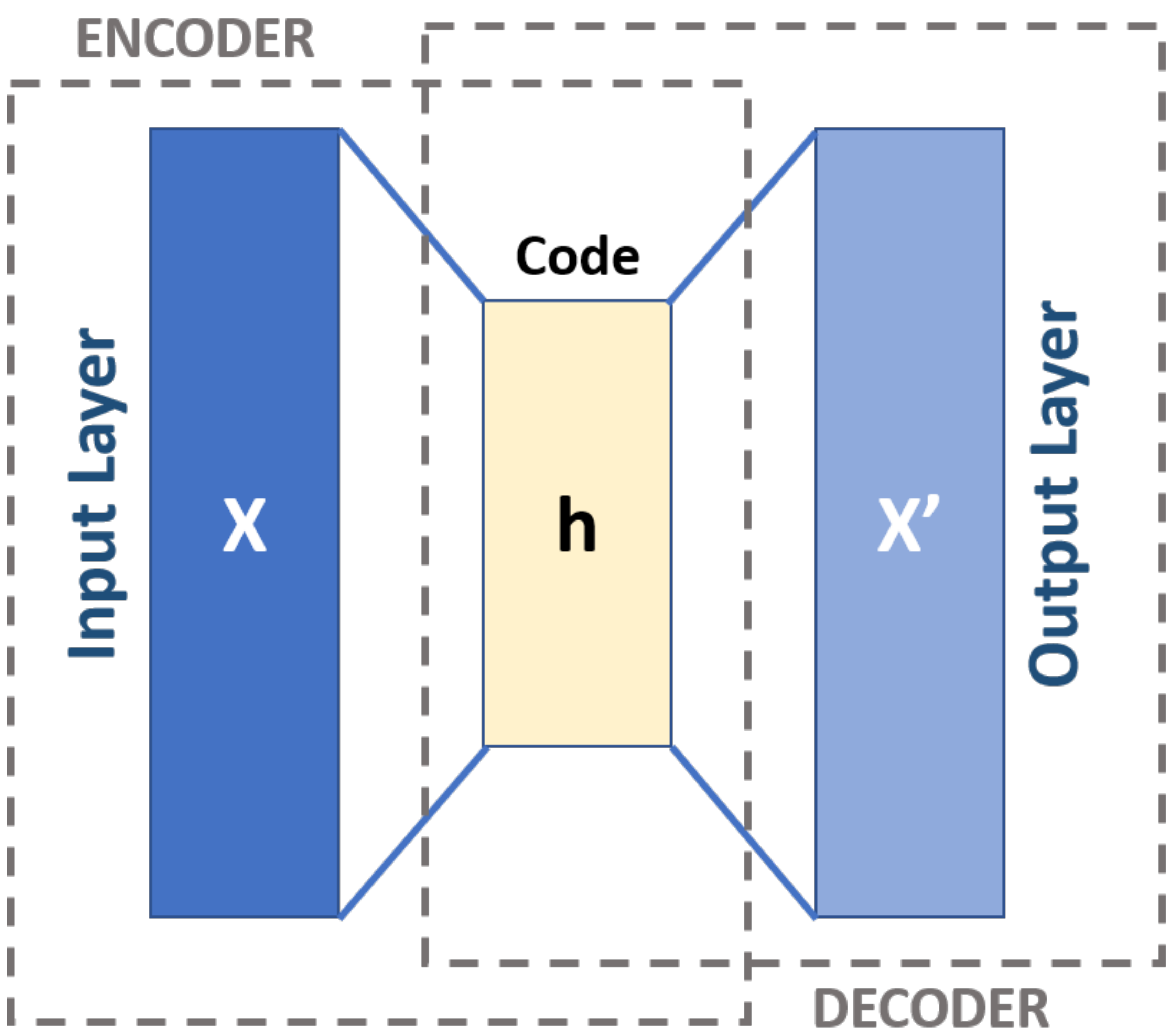
$$L = \frac{1}{N} \sum_i (\mathrm{AE}(x_i) - x_i)^2$$



- Use $L > L_C$ to cut interesting events

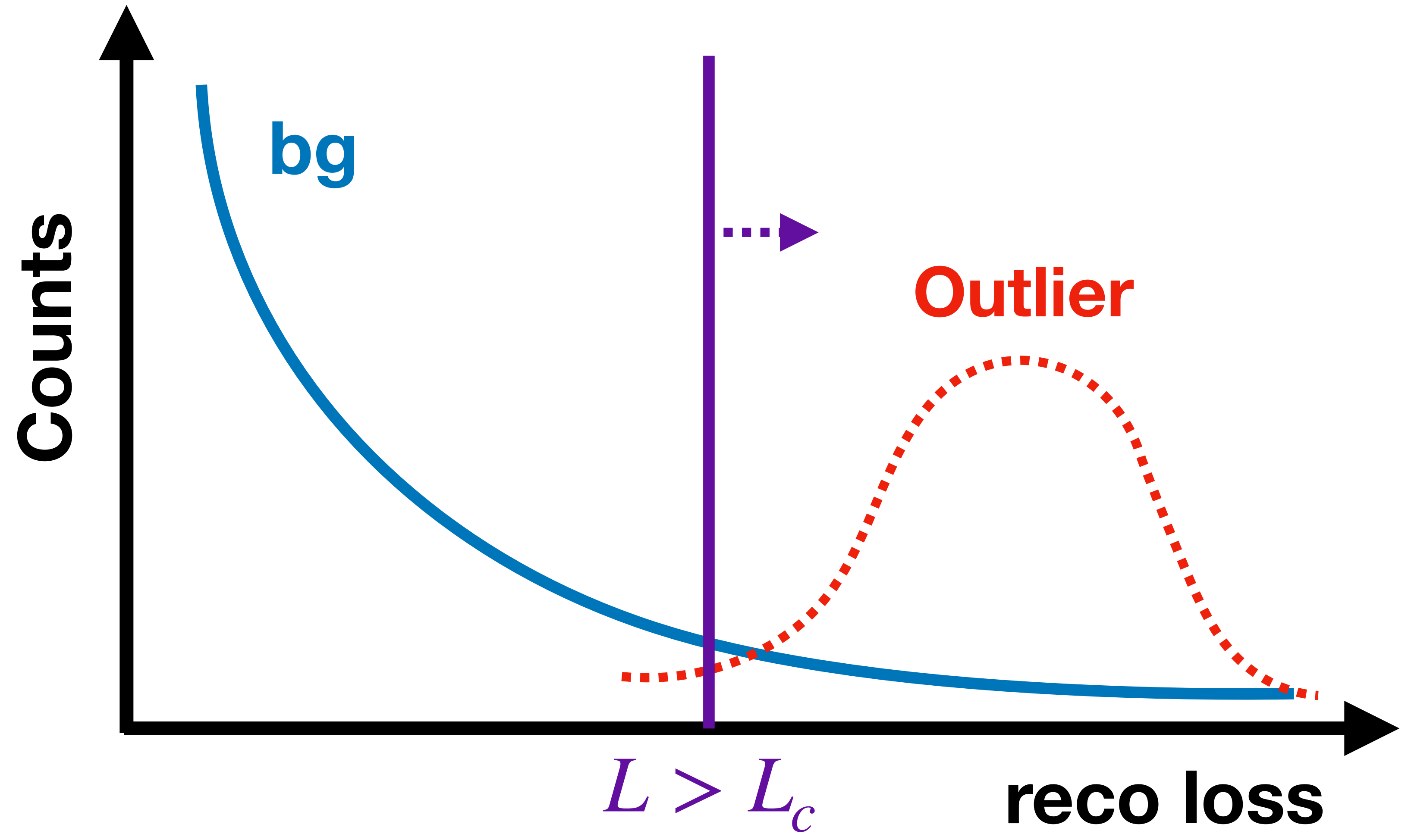
[Heimel et al: 1808.08979] [Farina et al: 1808.08992]

Autoencoder for non-resonant AD



- AE trained on bg.

$$L = \frac{1}{N}\sum_i (\mathrm{AE}(x_i) - x_i)^2$$



- Use $L > L_C$ to cut interesting events

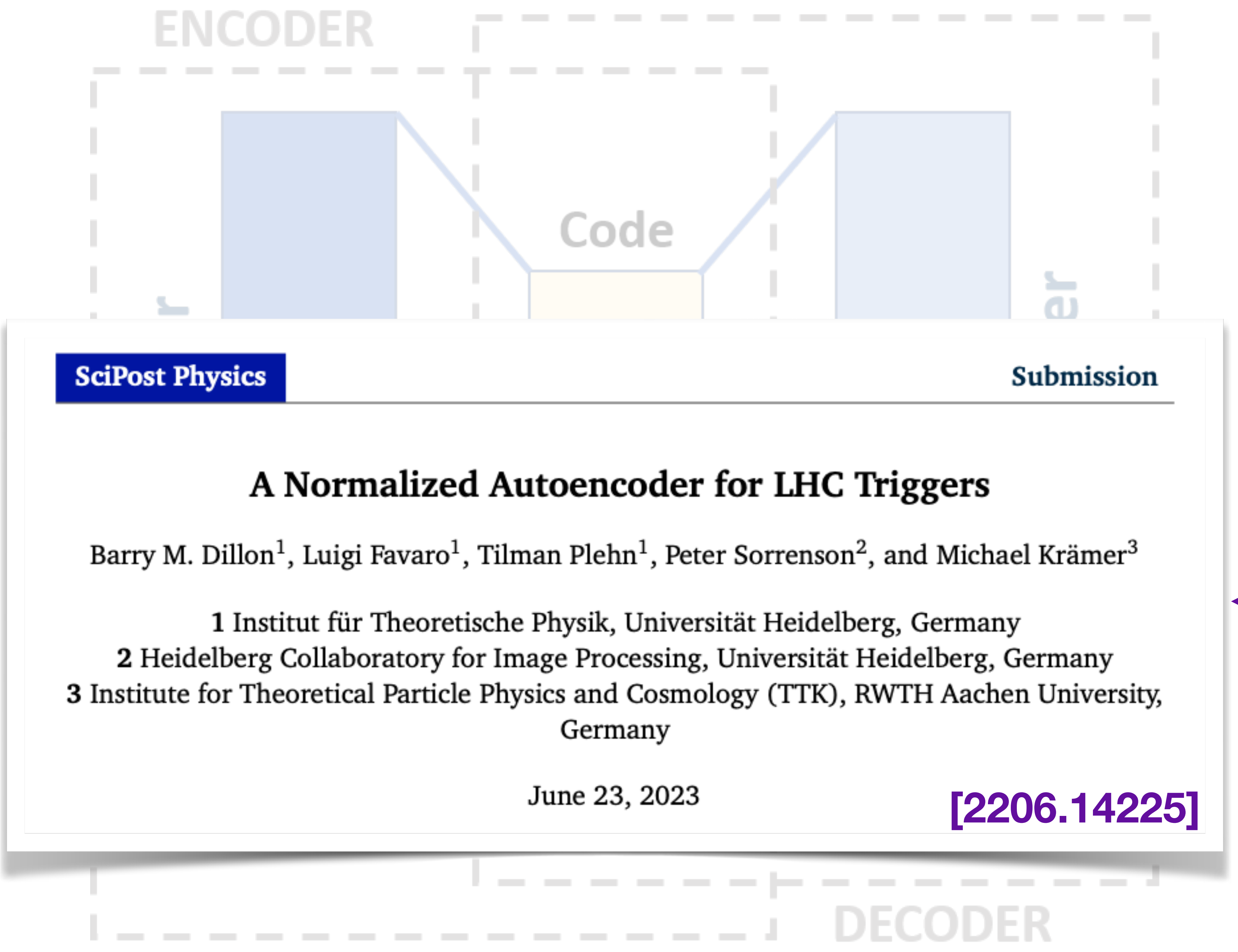
[Heimel et al: 1808.08979] [Farina et al: 1808.08992]

☺ Fully unsupervised

☹ Complexity bias [Finke et al: 2104.09051]

☹ not invariant under coordinate transformations [Kasieczka et al: 2209.06225]

Autoencoder for non-resonant AD



SciPost Physics — Submission

A Normalized Autoencoder for LHC Triggers

Barry M. Dillon[1], Luigi Favaro[1], Tilman Plehn[1], Peter Sorrenson[2], and Michael Krämer[3]

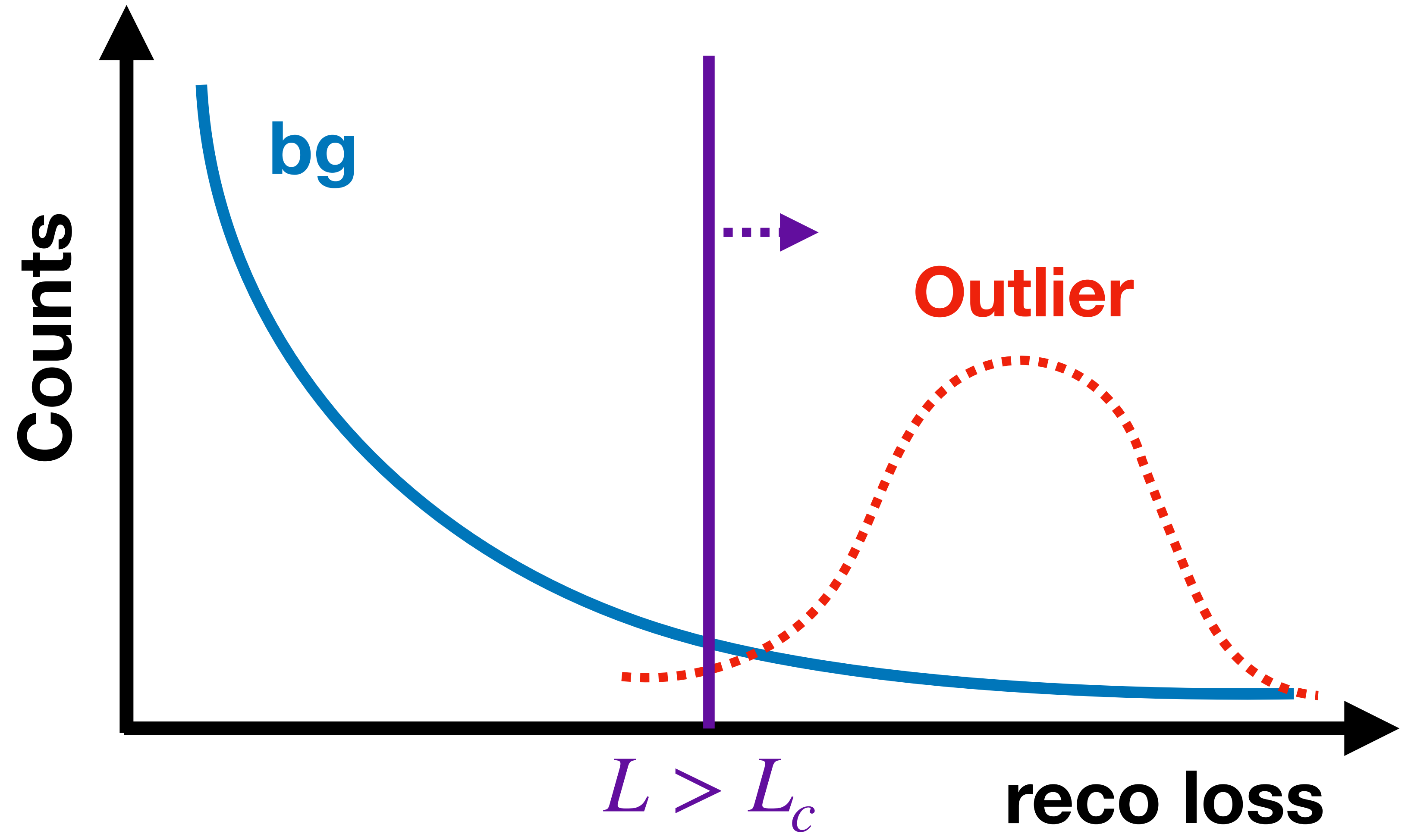
1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany
3 Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, Germany

June 23, 2023

[2206.14225]

- AE trained on bg.

$$L = \frac{1}{N}\sum_i (\mathrm{AE}(x_i) - x_i)^2$$



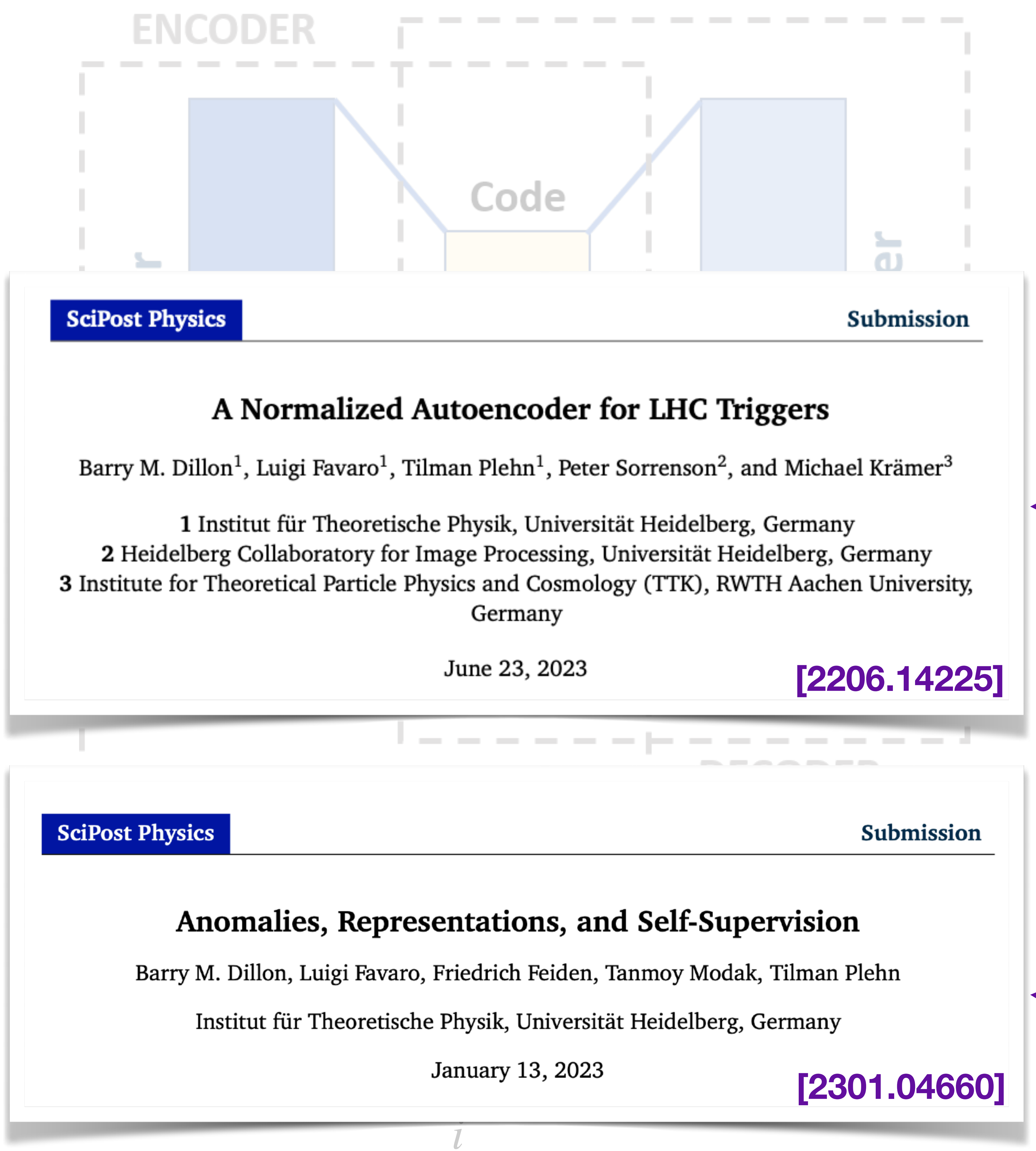
- Use $L > L_C$ to cut interesting events
  [Heimel et al: 1808.08979] [Farina et al: 1808.08992]

☺ Fully unsupervised

☹ Complexity bias [Finke et al: 2104.09051]

☹ not invariant under coordinate transformations [Kasieczka et al: 2209.06225]

Autoencoder for non-resonant AD



SciPost Physics — Submission

A Normalized Autoencoder for LHC Triggers

Barry M. Dillon[1], Luigi Favaro[1], Tilman Plehn[1], Peter Sorrenson[2], and Michael Krämer[3]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany
3 Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, Germany

June 23, 2023

[2206.14225]

SciPost Physics — Submission

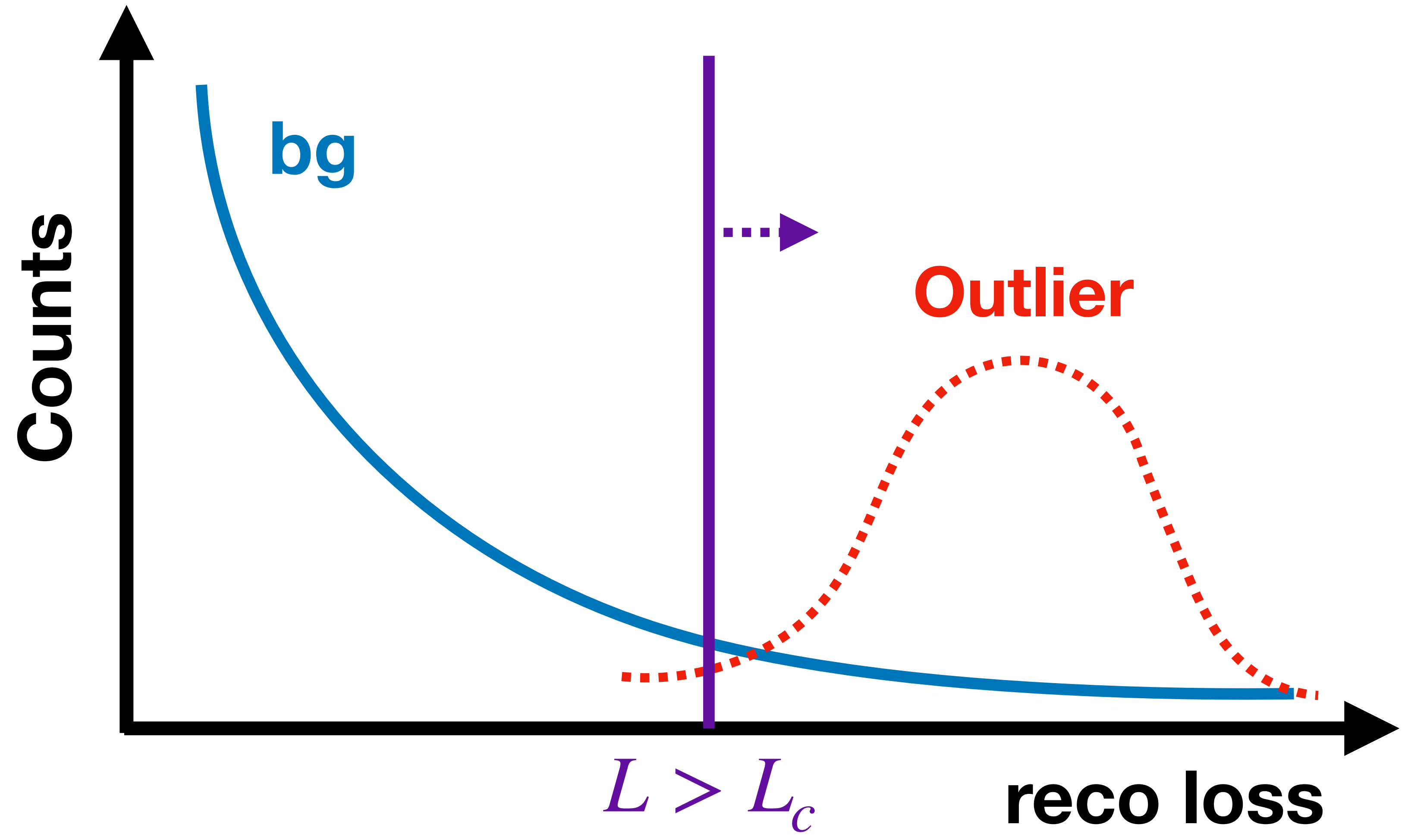
Anomalies, Representations, and Self-Supervision

Barry M. Dillon, Luigi Favaro, Friedrich Feiden, Tanmoy Modak, Tilman Plehn

Institut für Theoretische Physik, Universität Heidelberg, Germany

January 13, 2023

[2301.04660]



- Use $L > L_C$ to cut interesting events
  [Heimel et al: 1808.08979] [Farina et al: 1808.08992]

☺ Fully unsupervised

☹ Complexity bias [Finke et al: 2104.09051]

☹ not invariant under coordinate transformations [Kasieczka et al: 2209.06225]

Two Types of Anomaly Detection

Outlier Detection
(non-resonant)

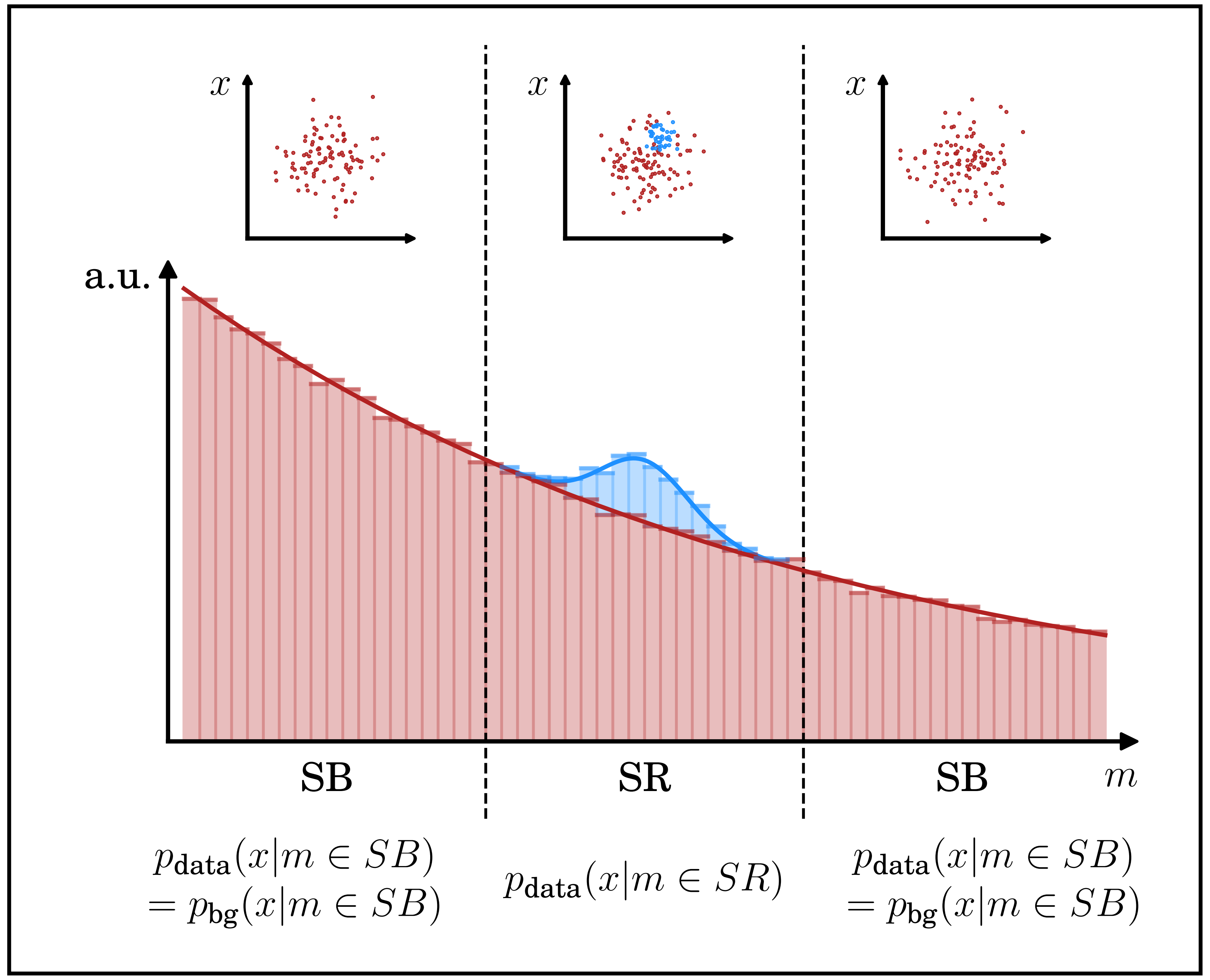
- Searching for unique and unexpected events
- In HEP, this (might) appear in the tails of dist.

[2404.07258]

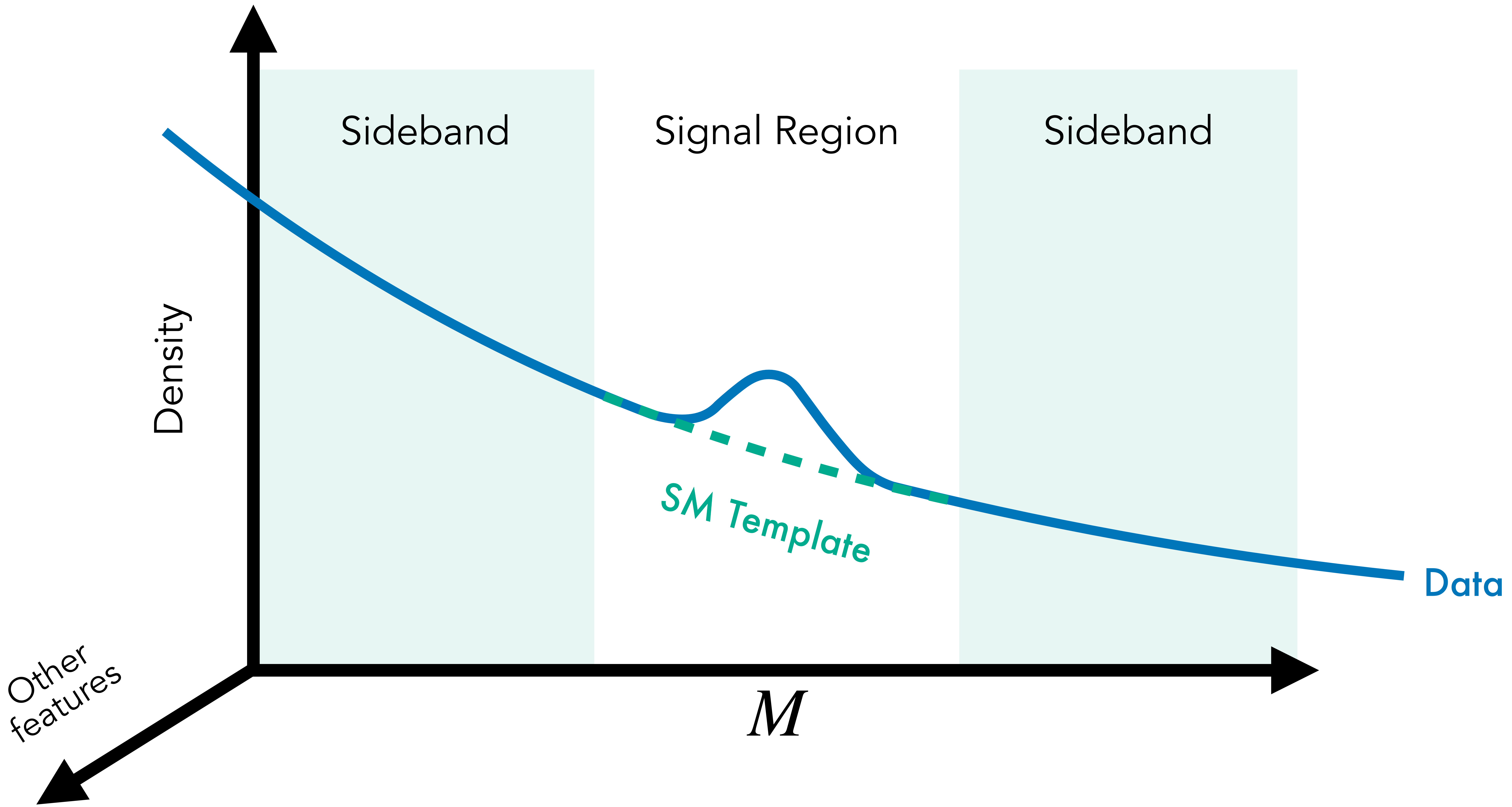
Overdensities
(resonant)

- Analagous to traditional bump hunt

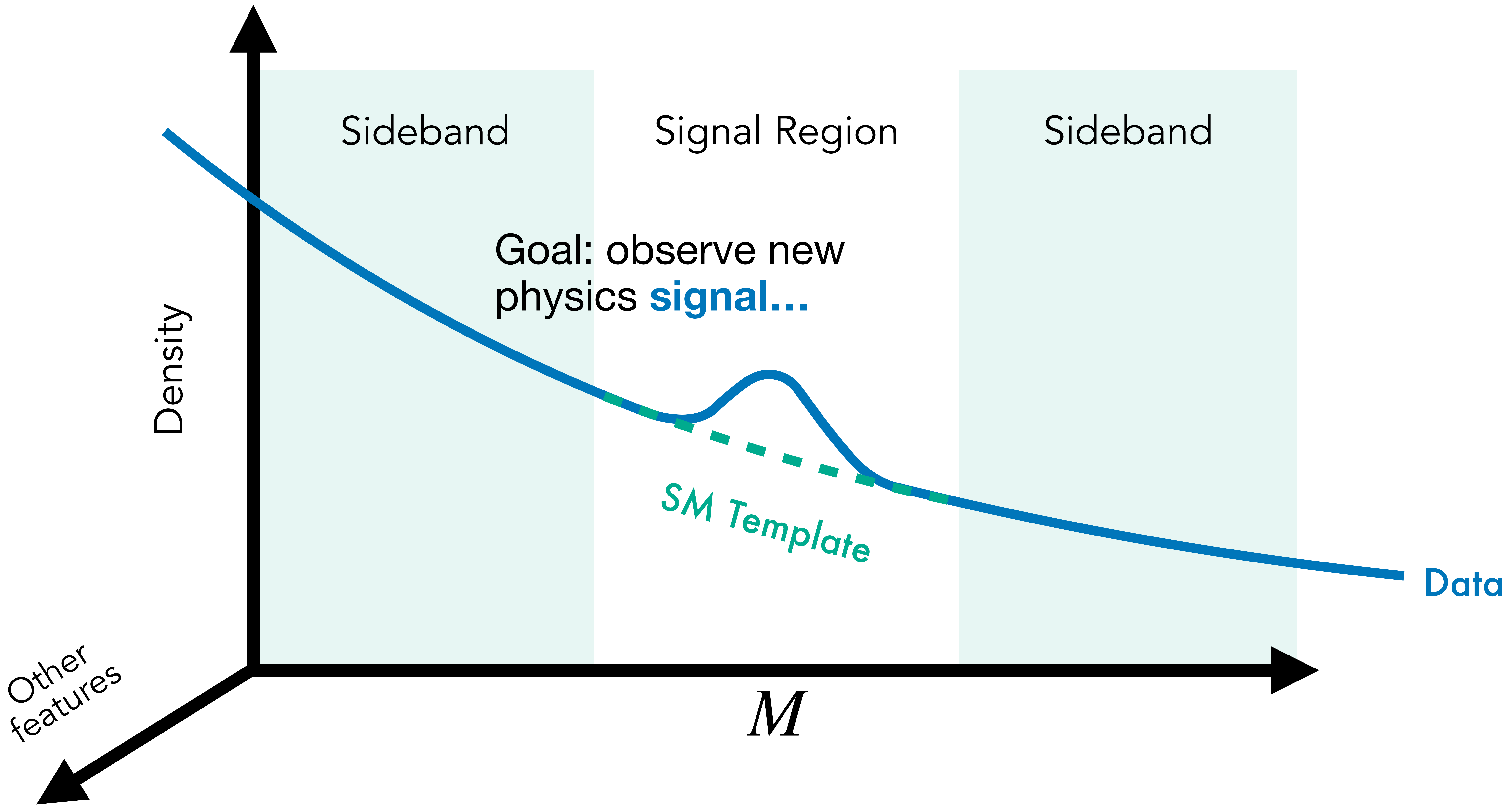
[2109.00546]



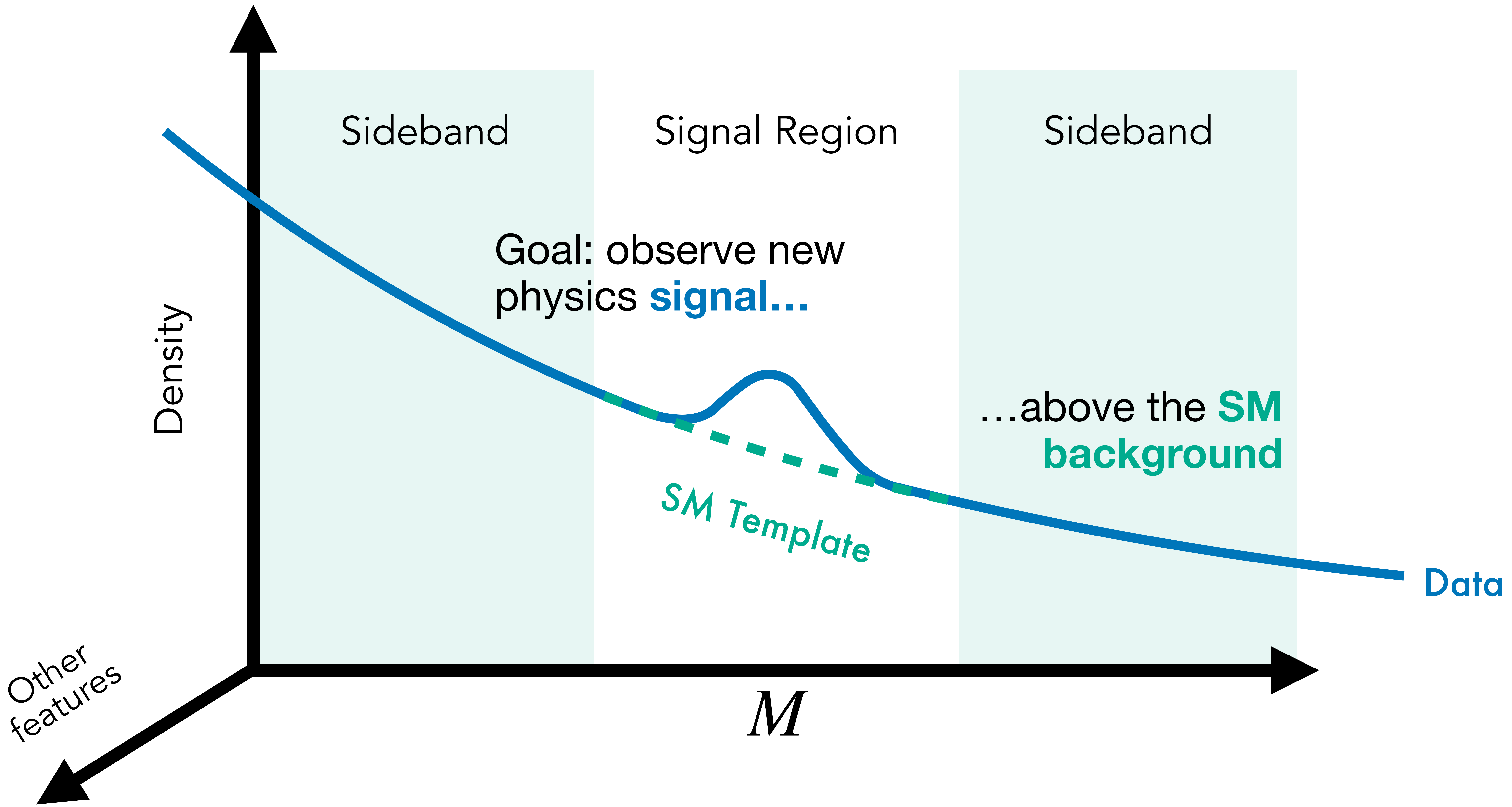
Resonant AD as a search strategy



Resonant AD as a search strategy



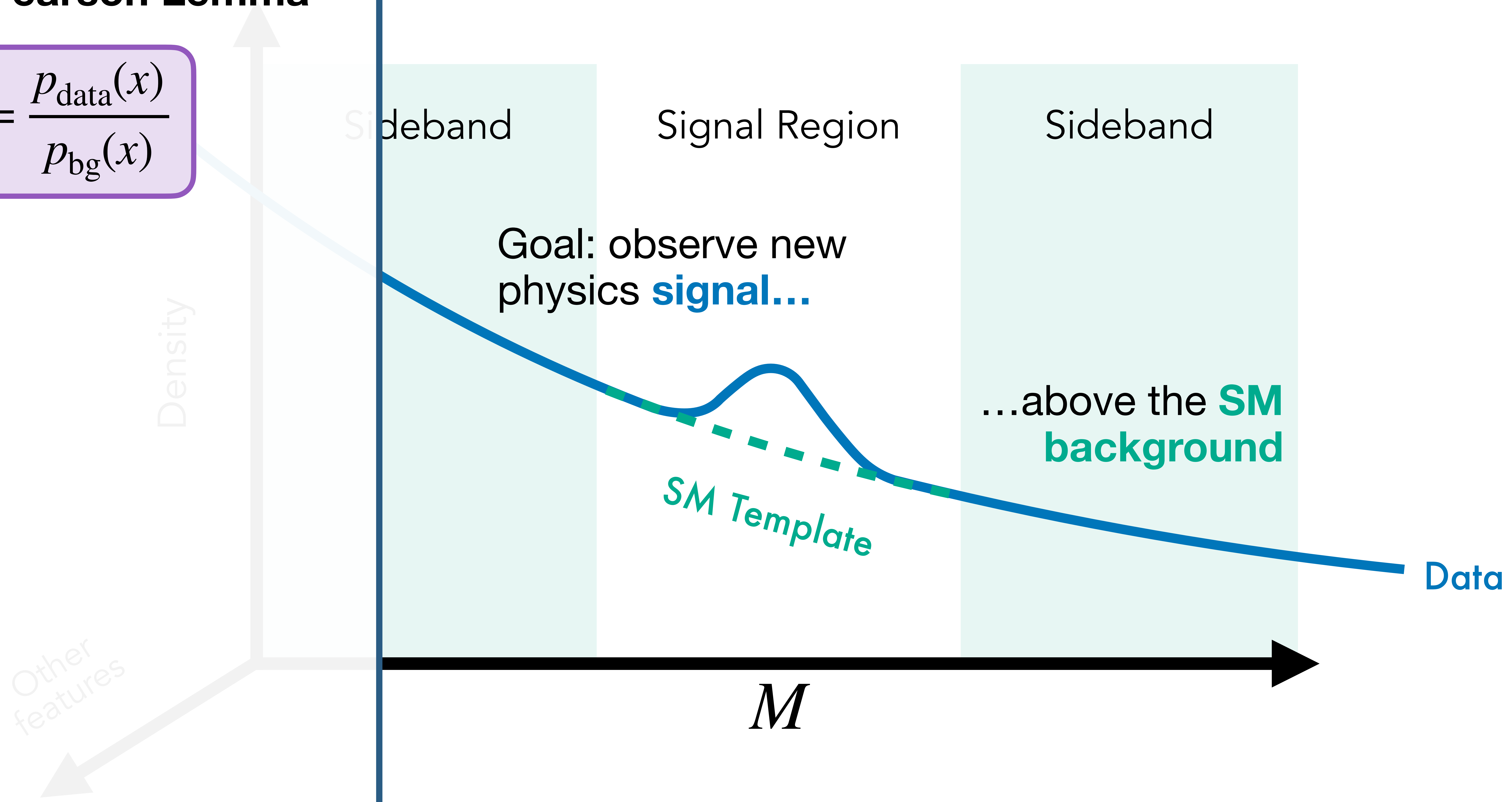
Resonant AD as a search strategy



Resonant AD as a search strategy

Neyman-Pearson Lemma

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

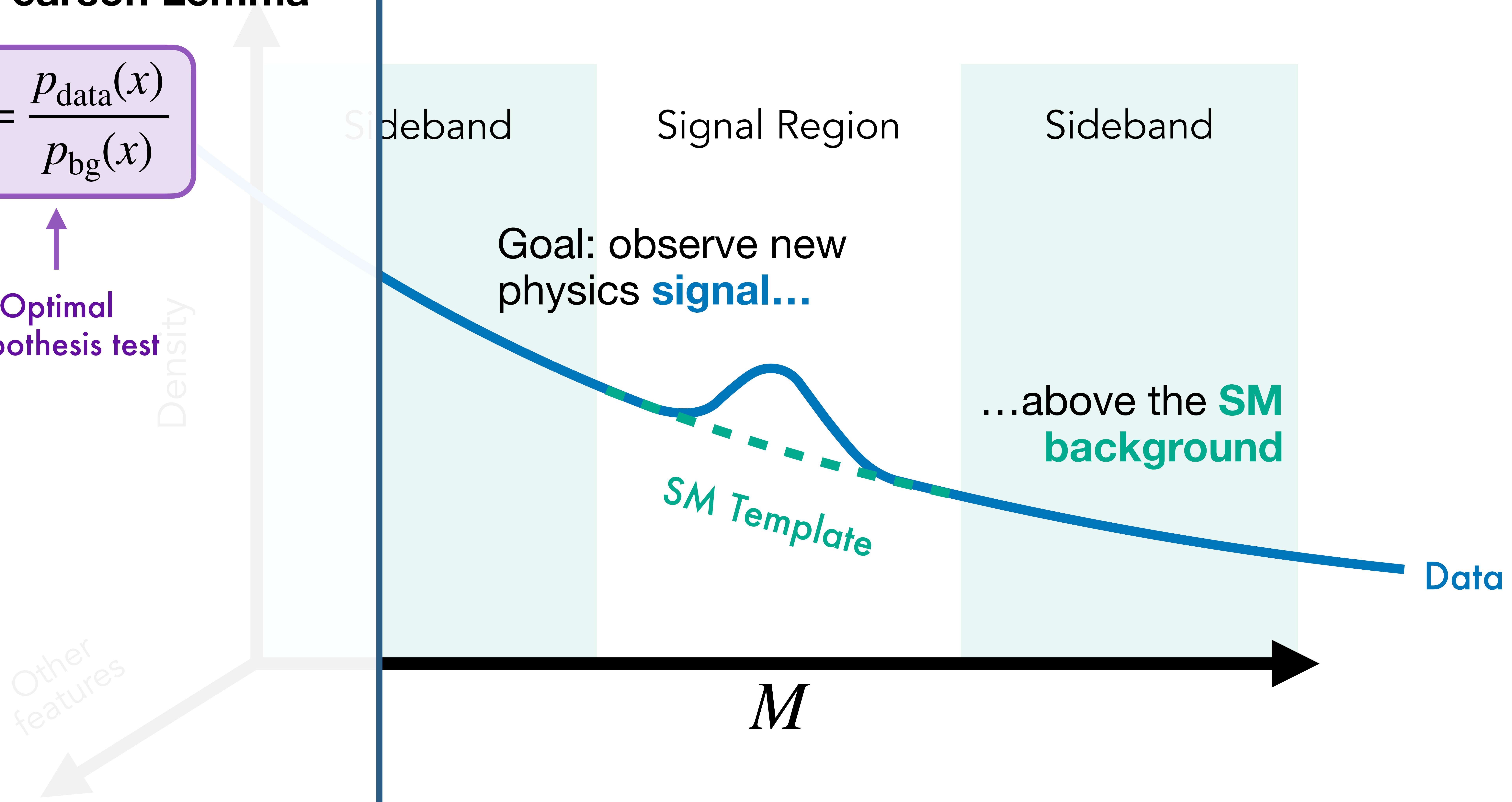


Resonant AD as a search strategy

Neyman-Pearson Lemma

$$R = \frac{p_{\mathrm{data}}(x)}{p_{\mathrm{bg}}(x)}$$

Optimal hypothesis test



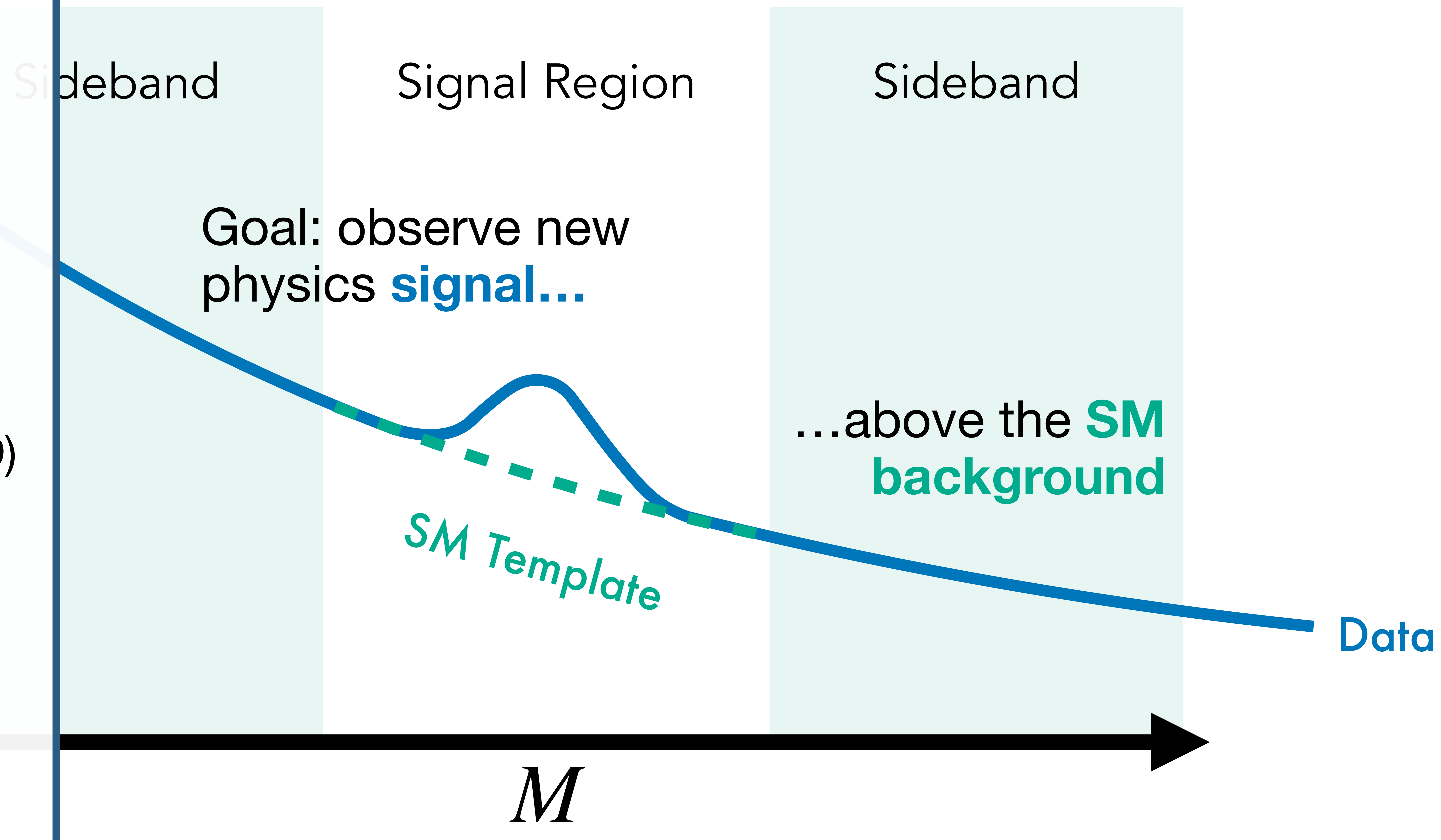
Resonant AD as a search strategy

Neyman-Pearson Lemma

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Optimal hypothesis test

❖ Idealized anomaly detector (IAD)



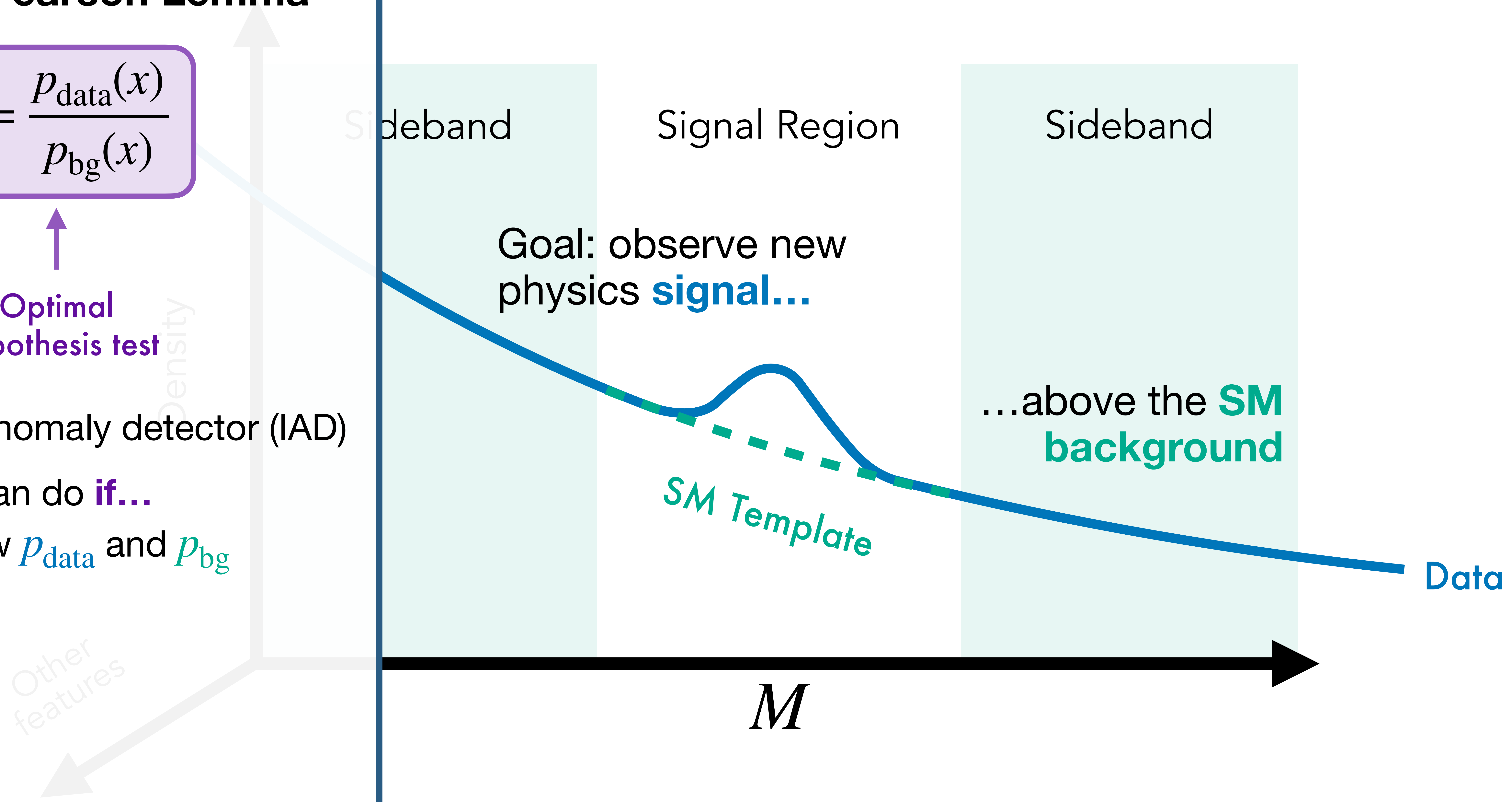
Resonant AD as a search strategy

Neyman-Pearson Lemma

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Optimal hypothesis test

- Idealized anomaly detector (IAD)
- Best you can do **if…**
  …you know p_{data} and p_{bg}



Resonant AD as a search strategy

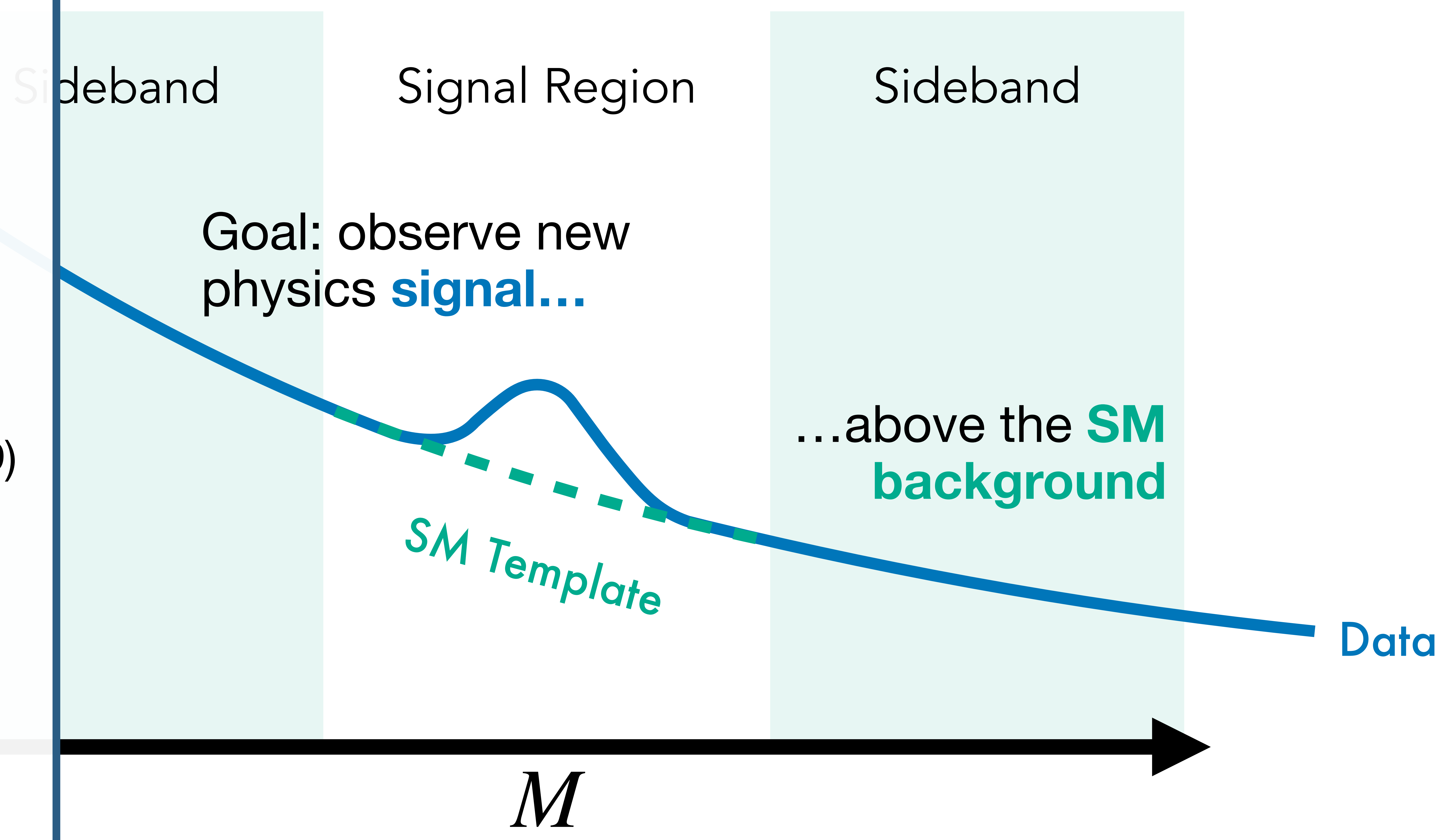
Neyman-Pearson Lemma

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Optimal hypothesis test

- Idealized anomaly detector (IAD)
- Best you can do **if…**
  …you know p_{data} and p_{bg}

ML



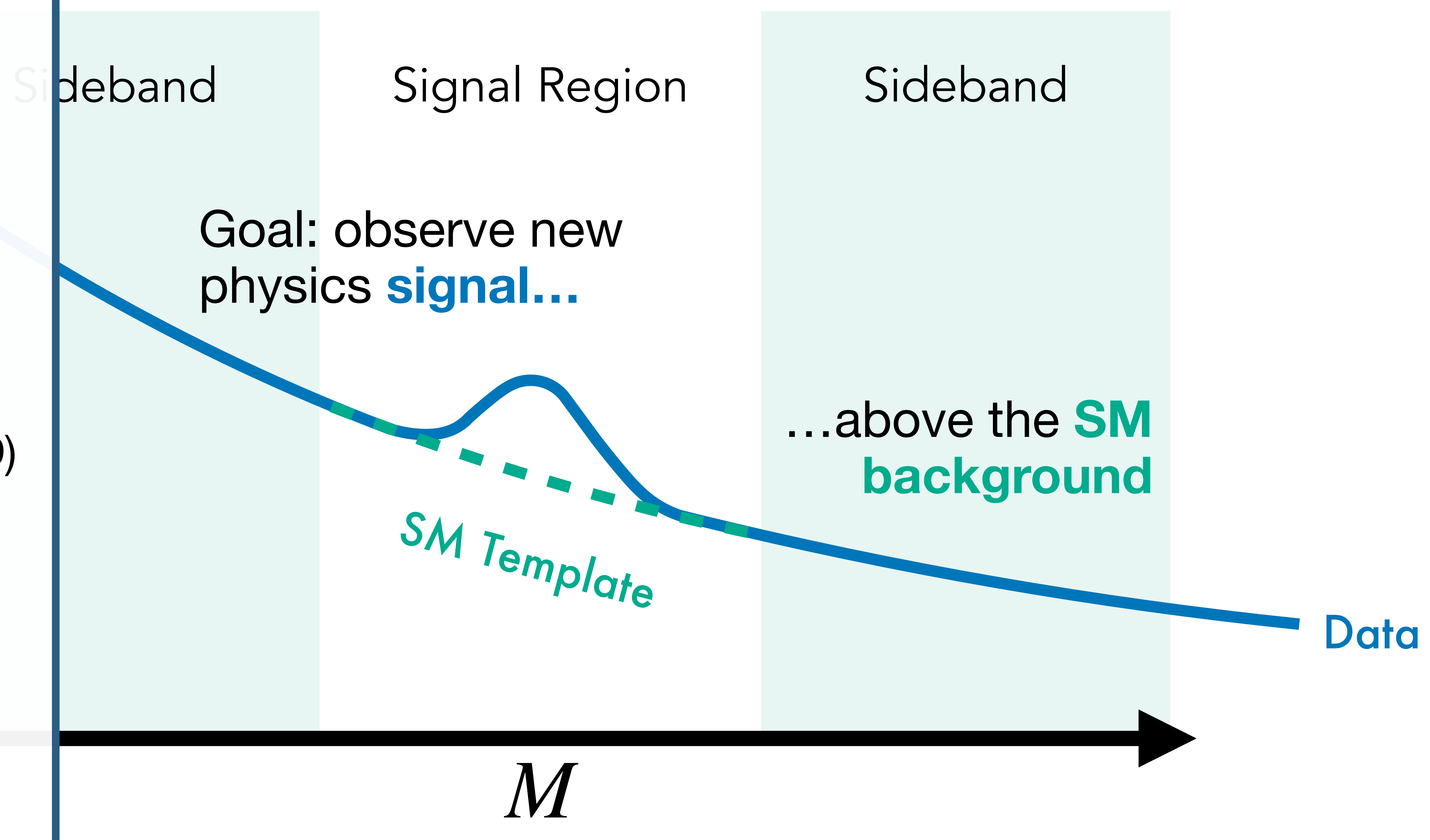
Resonant AD as a search strategy

Neyman-Pearson Lemma

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Optimal hypothesis test

- Idealized anomaly detector (IAD)
- Best you can do **if…**
  …you know p_{data} and p_{bg} (ML)
- Use R as cut discriminant
  → $R > R_c$



How to get the optimal test statistic?

$$R = \frac{p_{\mathrm{data}}(x)}{p_{\mathrm{bg}}(x)}$$

How to get the optimal test statistic?

Classifier

If we have samples from
data and **SM background…**

←

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

How to get the optimal test statistic?

Classifier

If we have samples from **data** and **SM background…**

…an **optimal classifier** yields

$$f(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{bg}}(x)}$$

←

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

How to get the optimal test statistic?

Classifier

If we have samples from **data** and **SM background…**

…an **optimal classifier** yields

$$f(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{bg}}(x)}$$

- Get $x \sim p_{\text{data}}$ and $x \sim p_{\text{bg}}$ from **MC simulations**

←

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

How to get the optimal test statistic?

Classifier

If we have samples from
data and **SM background…**

…an **optimal classifier** yields

$$f(x) = \frac{p_{\mathrm{data}}(x)}{p_{\mathrm{data}}(x) + p_{\mathrm{bg}}(x)}$$

- Get $x \sim p_{\mathrm{data}}$ and $x \sim p_{\mathrm{bg}}$ from **MC simulations**
- Estimate samples from **data**:

$$x \sim p_{\mathrm{data}}(x \,|\, \mathrm{SR})$$

$$x \sim p_{\mathrm{data}}(x \,|\, \mathrm{SB}) \approx p_{\mathrm{bg}}(x)$$

←

$$R = \frac{p_{\mathrm{data}}(x)}{p_{\mathrm{bg}}(x)}$$

How to get the optimal test statistic?

Classifier

If we have samples from **data** and **SM background…**

…an **optimal classifier** yields

$$f(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{bg}}(x)}$$

- Get $x \sim p_{\text{data}}$ and $x \sim p_{\text{bg}}$ from **MC simulations**
- Estimate samples from **data**:

$$x \sim p_{\text{data}}(x \,|\, \text{SR})$$

$$x \sim p_{\text{data}}(x \,|\, \text{SB}) \approx p_{\text{bg}}(x)$$

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Density estimator

Instead of learning the **likelihood ratio** directly…

How to get the optimal test statistic?

Classifier

If we have samples from **data** and **SM background…**

…an **optimal classifier** yields

$$f(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{bg}}(x)}$$

- Get $x \sim p_{\text{data}}$ and $x \sim p_{\text{bg}}$ from **MC simulations**
- Estimate samples from **data**:

$$x \sim p_{\text{data}}(x\,|\,\text{SR})$$

$$x \sim p_{\text{data}}(x\,|\,\text{SB}) \approx p_{\text{bg}}(x)$$

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Density estimator

Instead of learning the **likelihood ratio** directly…

…use a **density estimator** to learn

$$p_\omega(x\,|\,\text{SR}) \simeq p_{\text{data}}(x\,|\,\text{SR})$$

$$p_\omega(x\,|\,\text{SB}) \simeq p_{\text{bg}}(x)$$

How to get the optimal test statistic?

Classifier

If we have samples from **data** and **SM background…**

…an **optimal classifier** yields

$$f(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{bg}}(x)}$$

- Get $x \sim p_{\text{data}}$ and $x \sim p_{\text{bg}}$ from **MC simulations**
- Estimate samples from **data**:

$$x \sim p_{\text{data}}(x \,|\, \text{SR})$$

$$x \sim p_{\text{data}}(x \,|\, \text{SB}) \approx p_{\text{bg}}(x)$$

$$R = \frac{p_{\text{data}}(x)}{p_{\text{bg}}(x)}$$

Density estimator

Instead of learning the **likelihood ratio** directly…

…use a **density estimator** to learn

$$p_\omega(x \,|\, \text{SR}) \simeq p_{\text{data}}(x \,|\, \text{SR})$$

$$p_\omega(x \,|\, \text{SB}) \simeq p_{\text{bg}}(x)$$

- Then **calculate** R directly from the individual likelihoods

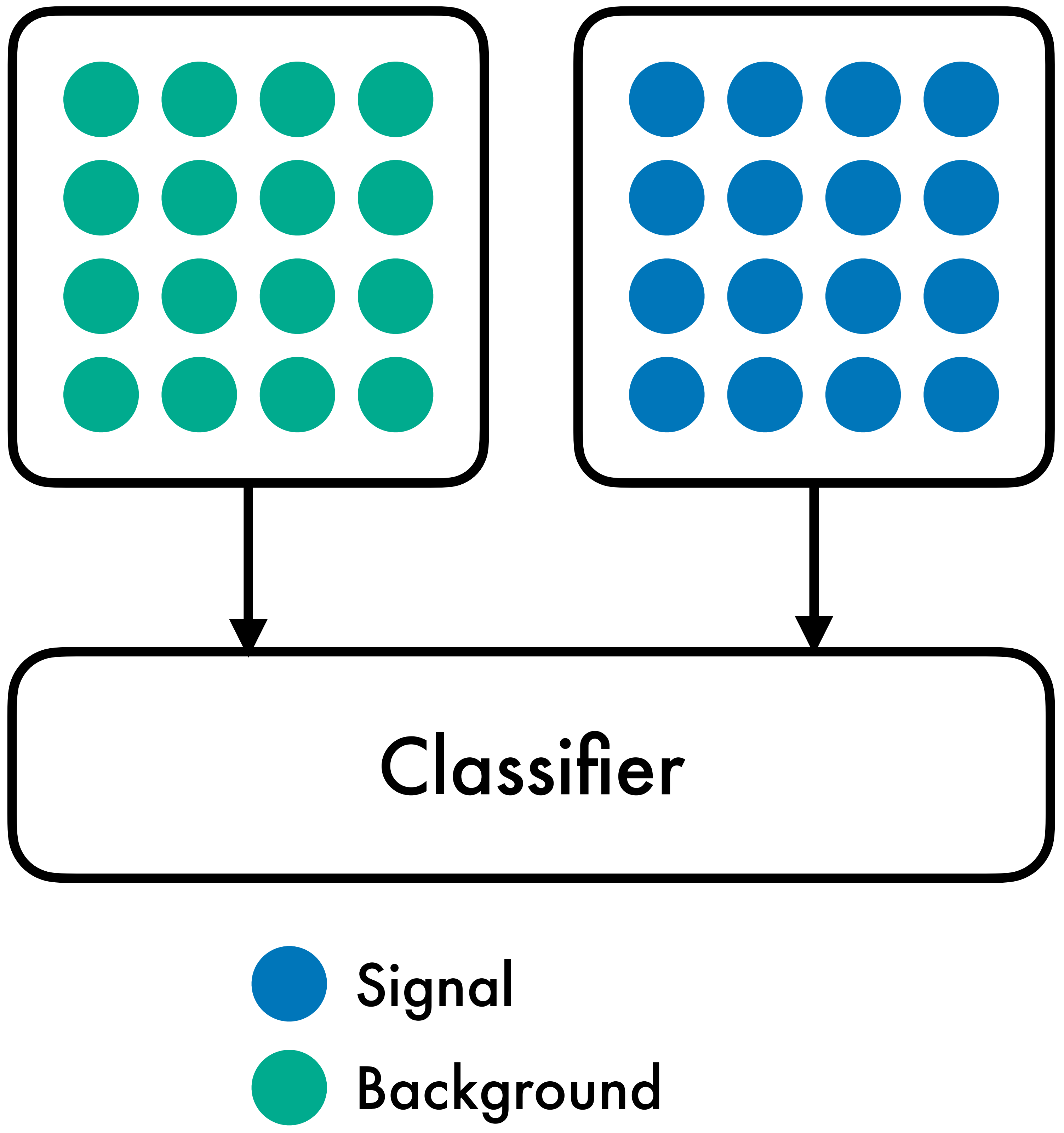
Example I

CWoLa Hunting

Metodiev, Nachman, Thaler [1708.02949]
Collins, Howe, Nachman [1805.02664]

Reminder — Classification Problem

Goal: learn the signal to background ratio

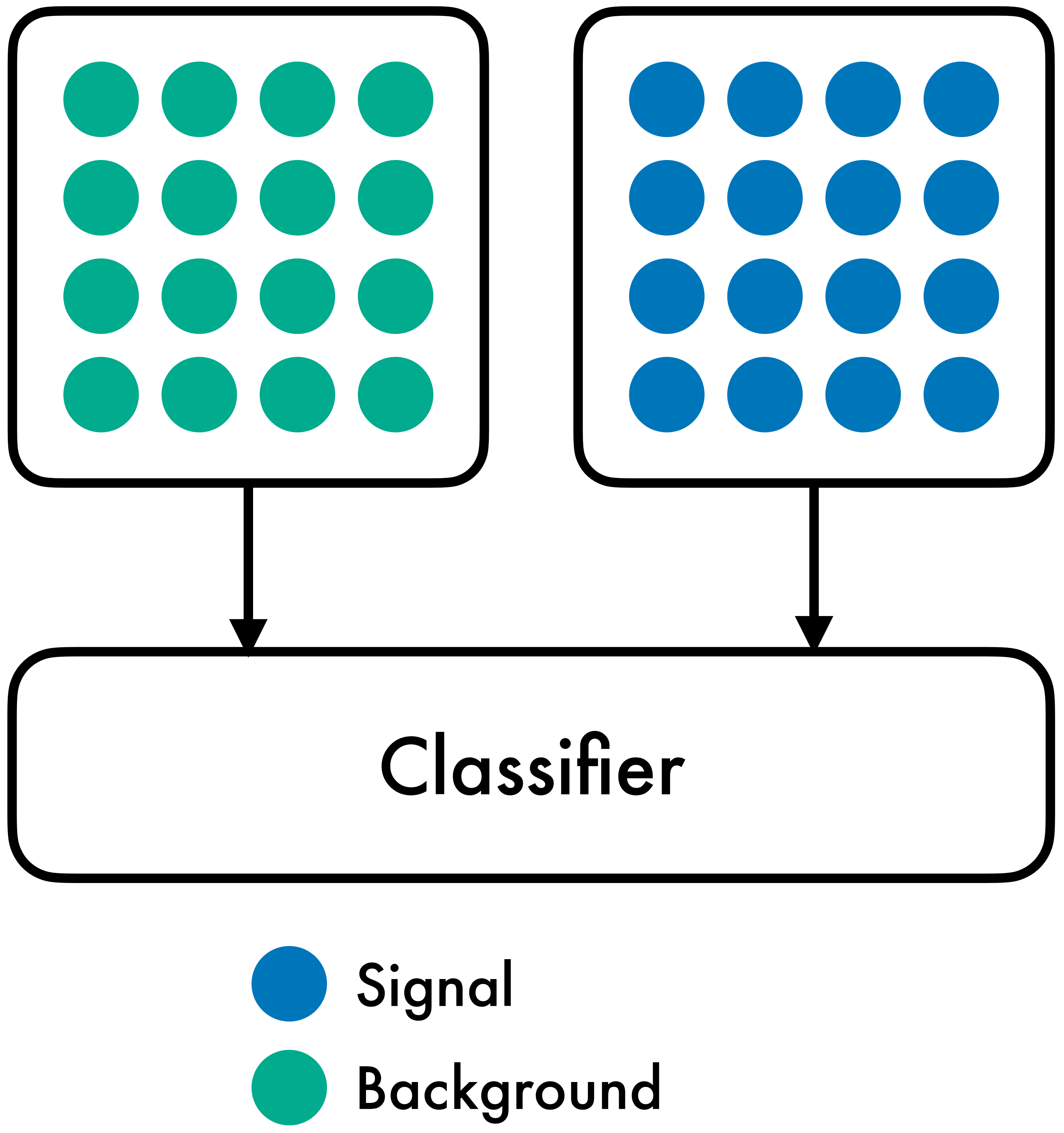


Reminder — Classification Problem

Goal: learn the signal to background ratio

An optimal classifier yields the likelihood ratio

$$R_{\text{optimal}} = \frac{f(x)}{1 - f(x)} = \frac{p_{\text{sig}}(x)}{p_{\text{bg}}(x)}$$



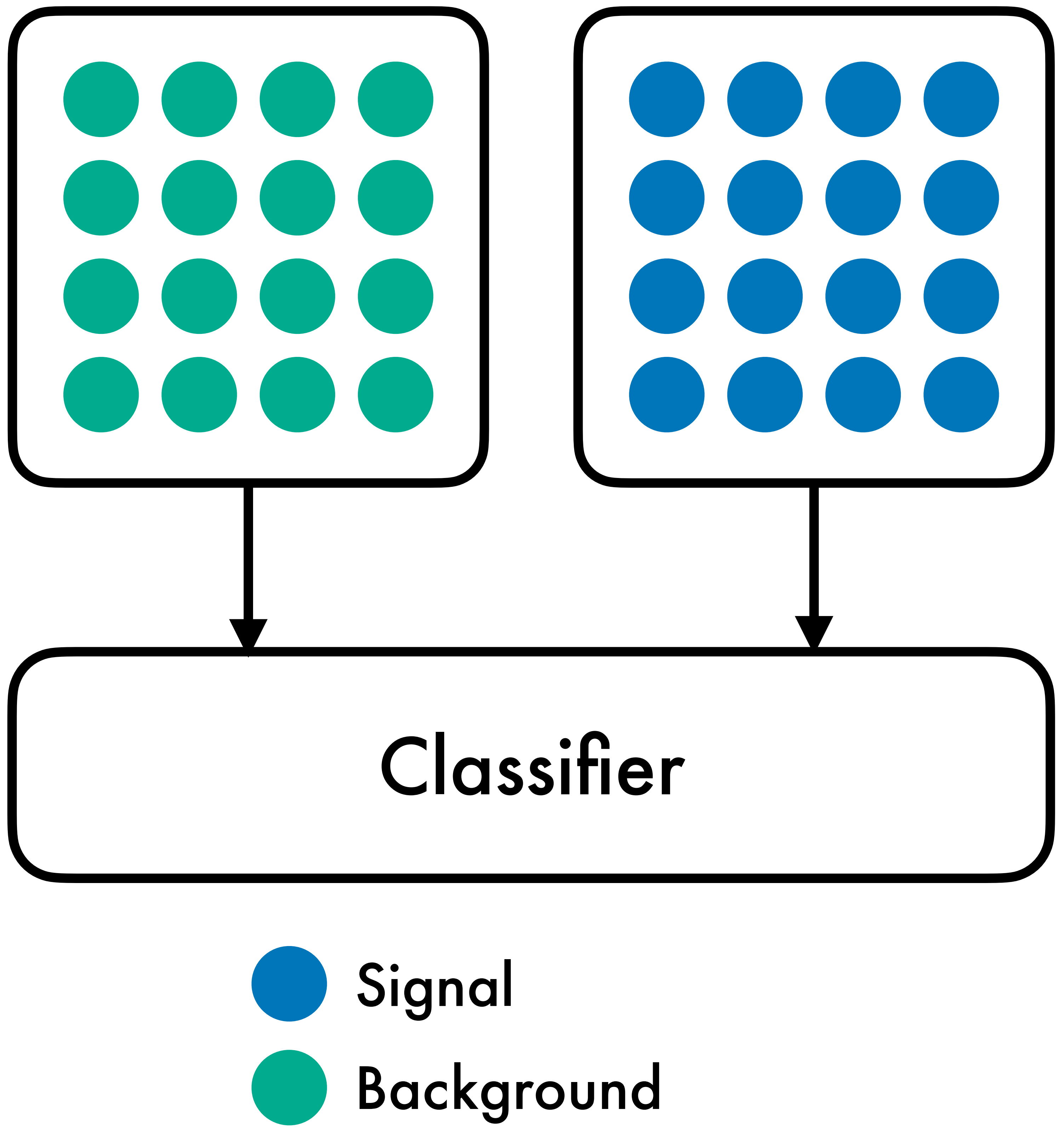
Reminder — Classification Problem

Goal: learn the signal to background ratio

An optimal classifier yields the likelihood ratio

$$R_{\text{optimal}} = \frac{f(x)}{1 - f(x)} = \frac{p_{\text{sig}}(x)}{p_{\text{bg}}(x)}$$

⊕ Can be approximated with a **supervised classifier (ML)**



Reminder — Classification Problem

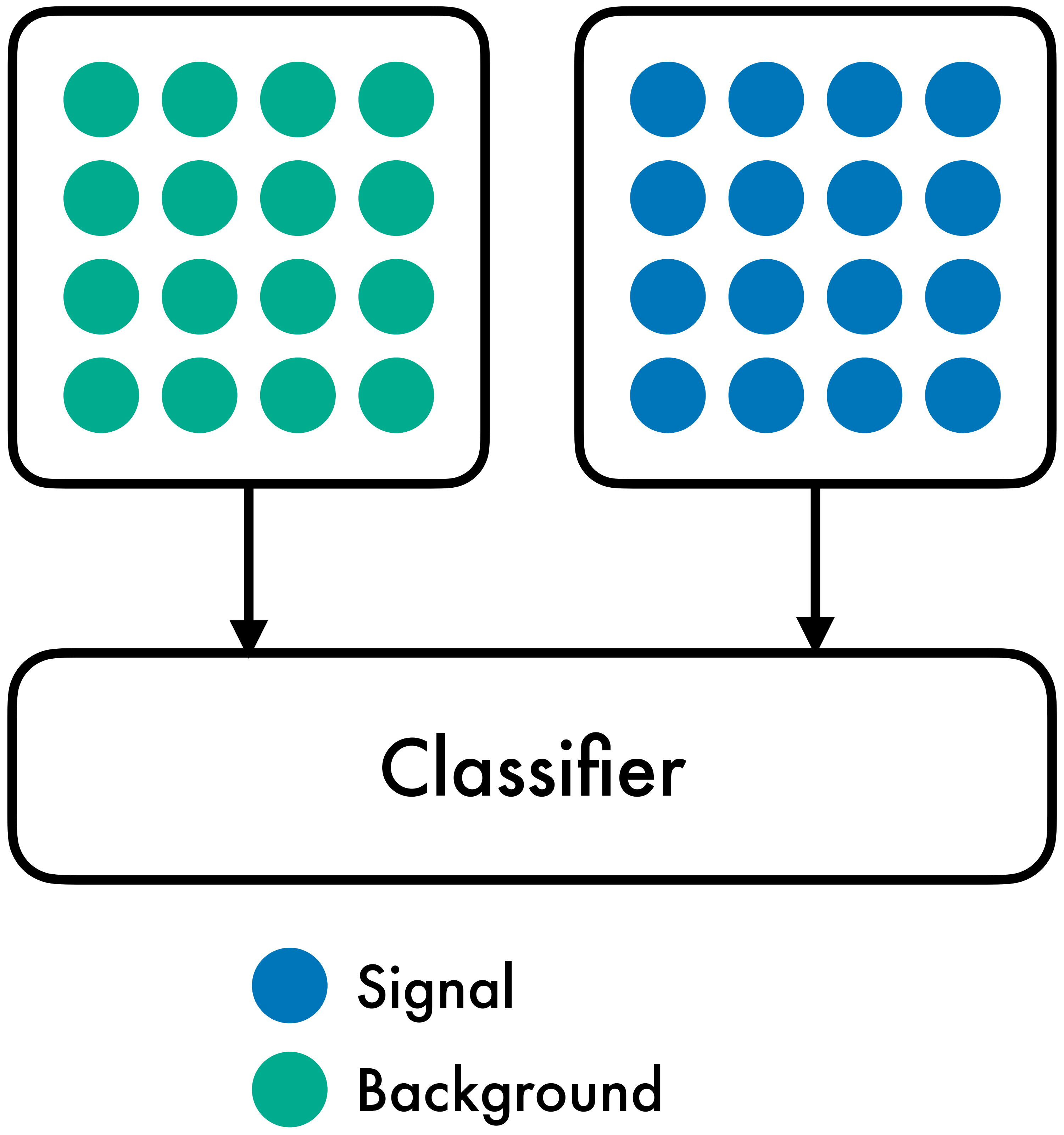
Goal: learn the signal to background ratio

An optimal classifier yields the likelihood ratio

$$R_{\text{optimal}} = \frac{f(x)}{1 - f(x)} = \frac{p_{\text{sig}}(x)}{p_{\text{bg}}(x)}$$

⊕ Can be approximated with a **supervised classifier (ML)**

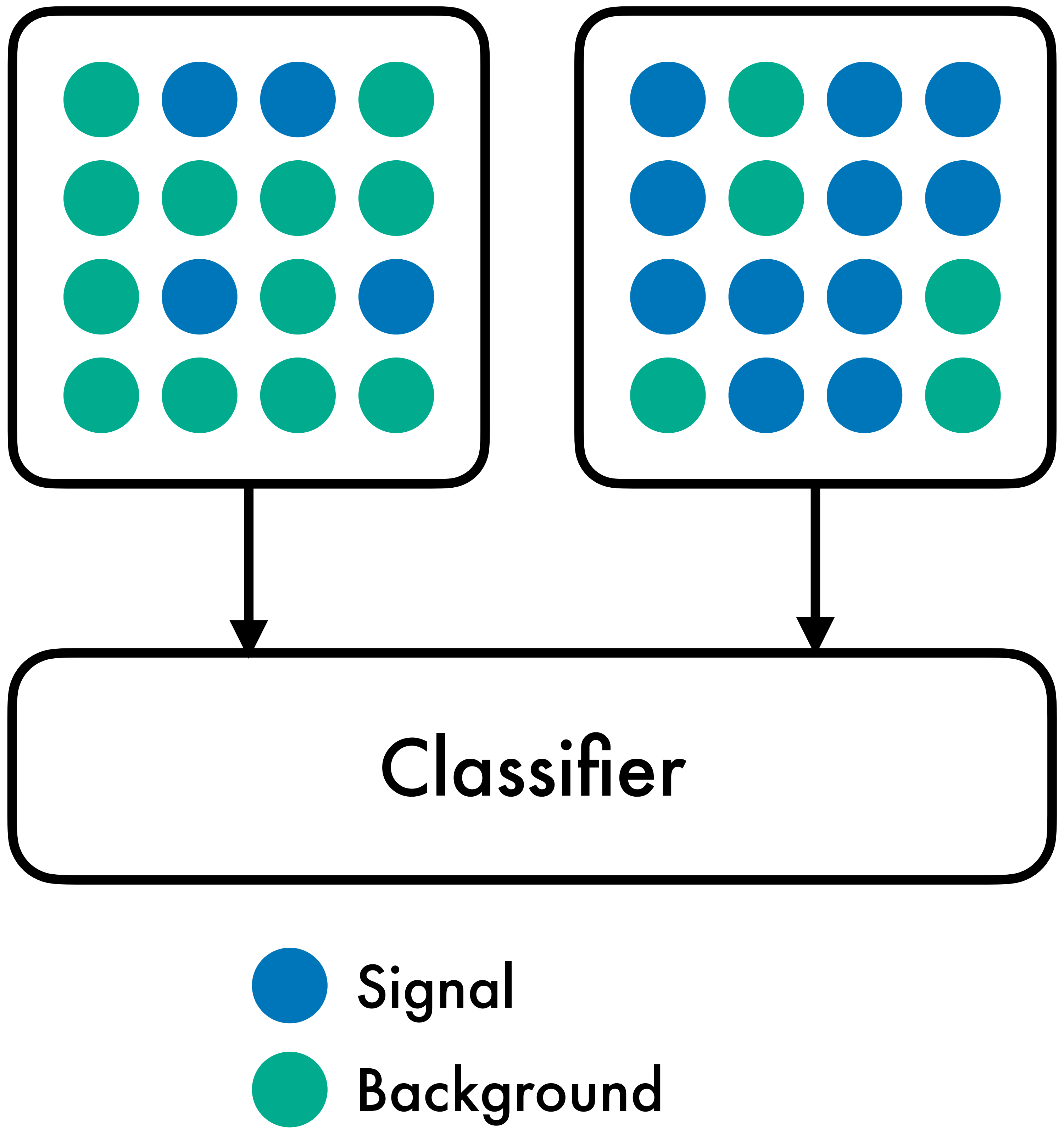
⊖ Labels **are not available** in experimental data



Classification without labels (CWoLa)

Two **mixed datasets** with signal fractions w_i

$$p_i(x) = w_i\, p_{\text{sig}}(x) + (1 - w_i)\, p_{\text{bg}}(x)$$



Metodiev, Nachman, Thaler [1708.02949]

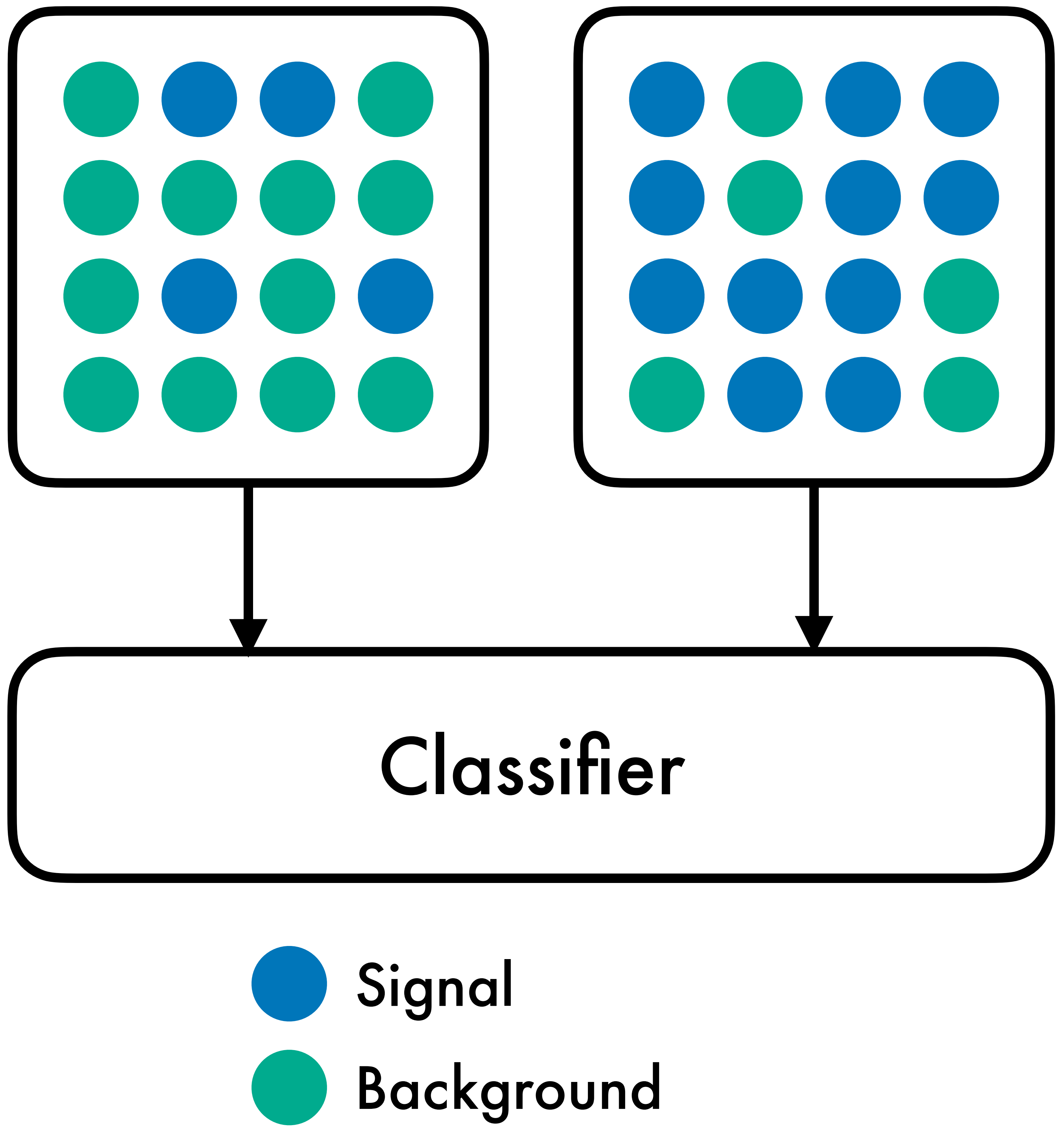
Classification without labels (CWoLa)

Two **mixed** **datasets** with signal fractions w_i

$$p_i(x) = w_i\, p_{\text{sig}}(x) + (1 - w_i)\, p_{\text{bg}}(x)$$

Classifier gives likelihood ratio

$$R_{\text{mixed}} = \frac{w_1\, R_{\text{optimal}}(x) + (1 - w_1)}{w_2\, R_{\text{optimal}}(x) + (1 - w_2)}$$



Metodiev, Nachman, Thaler [1708.02949]

Classification without labels (CWoLa)

Two **mixed datasets** with signal fractions w_i

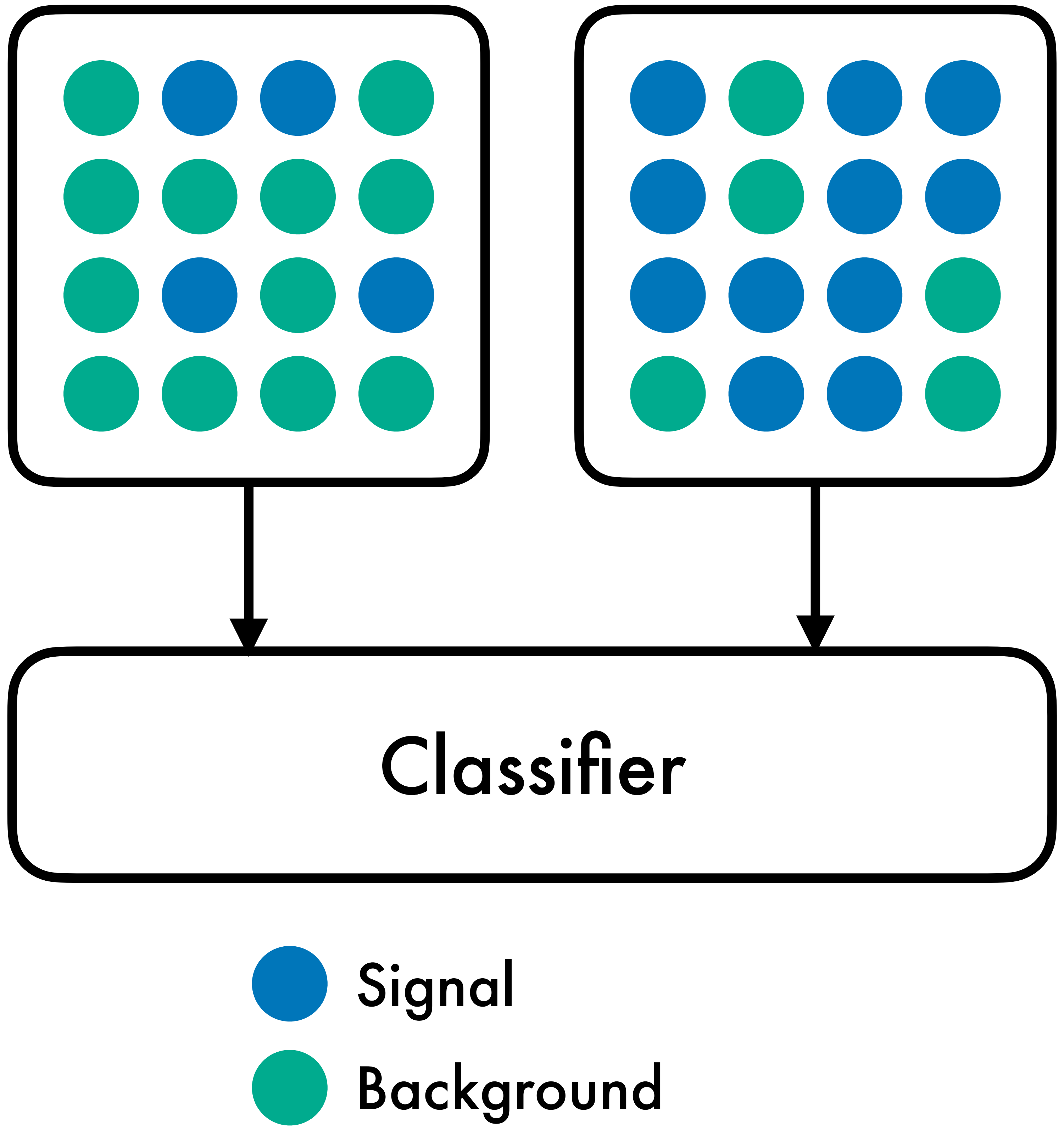
$$p_i(x) = w_i\, p_{\text{sig}}(x) + (1 - w_i)\, p_{\text{bg}}(x)$$

Classifier gives likelihood ratio

$$R_{\text{mixed}} = \frac{w_1\, R_{\text{optimal}}(x) + (1 - w_1)}{w_2\, R_{\text{optimal}}(x) + (1 - w_2)}$$

⊕ Monotonic function

→ optimal on mixed = optimal on pure sample



Metodiev, Nachman, Thaler [1708.02949]

Classification without labels (CWoLa)

Two **mixed datasets** with signal fractions w_i

$$p_i(x) = w_i\, p_{\text{sig}}(x) + (1 - w_i)\, p_{\text{bg}}(x)$$

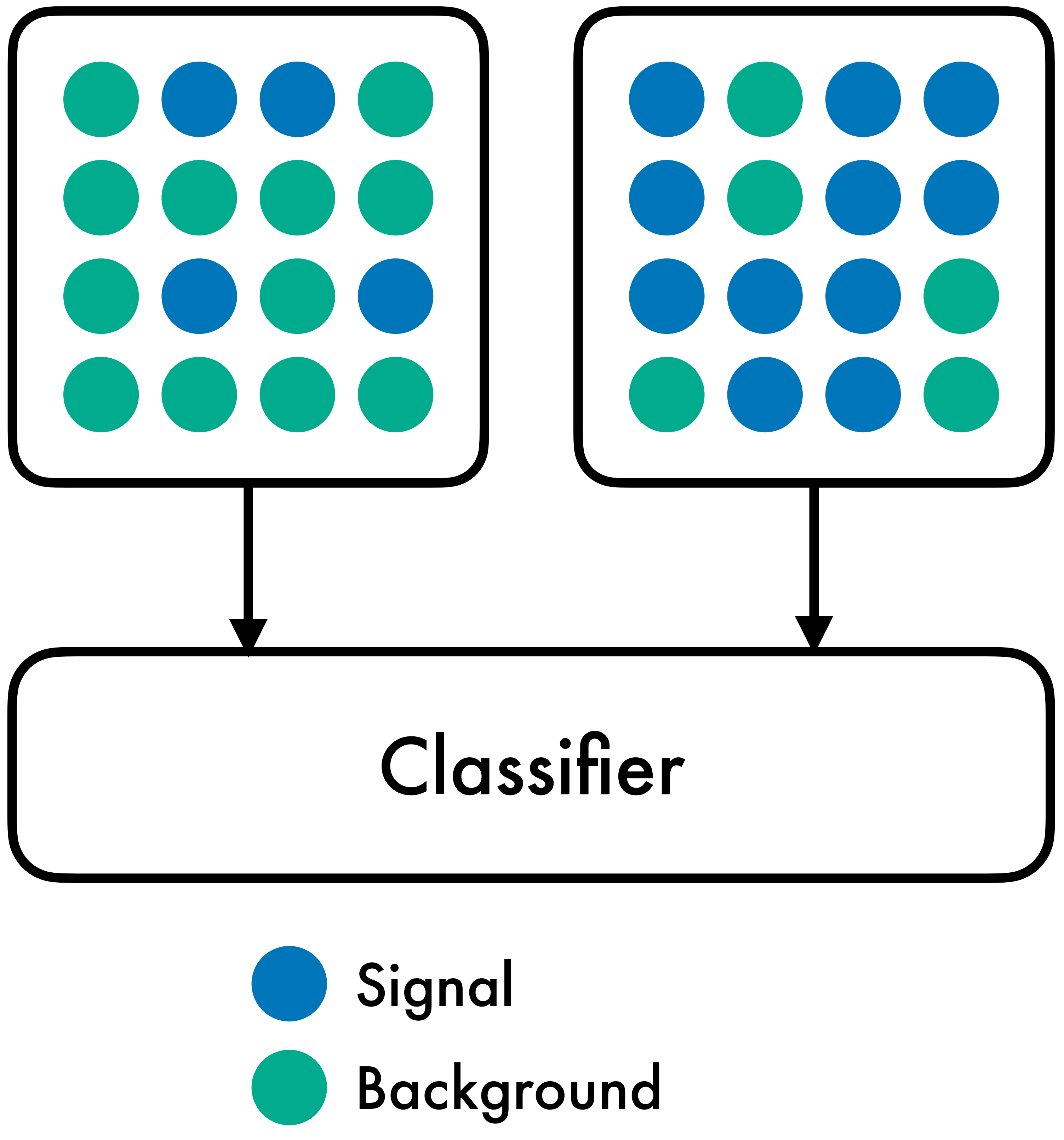
Classifier gives likelihood ratio

$$R_{\text{mixed}} = \frac{w_1\, R_{\text{optimal}}(x) + (1 - w_1)}{w_2\, R_{\text{optimal}}(x) + (1 - w_2)}$$

⊕ Monotonic function

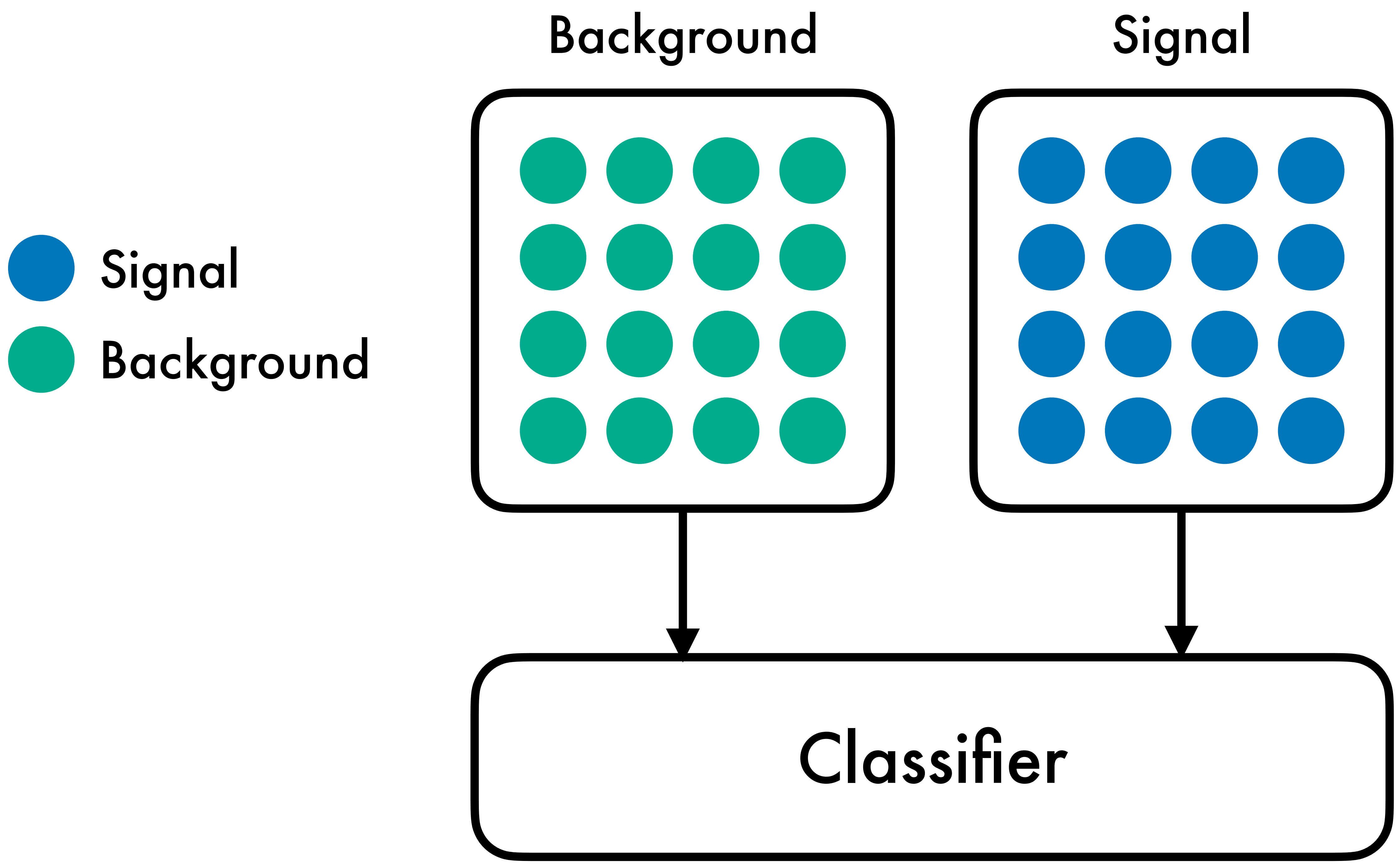
→ optimal on mixed = optimal on pure sample

⟶ Basis of **weak supervised classification**

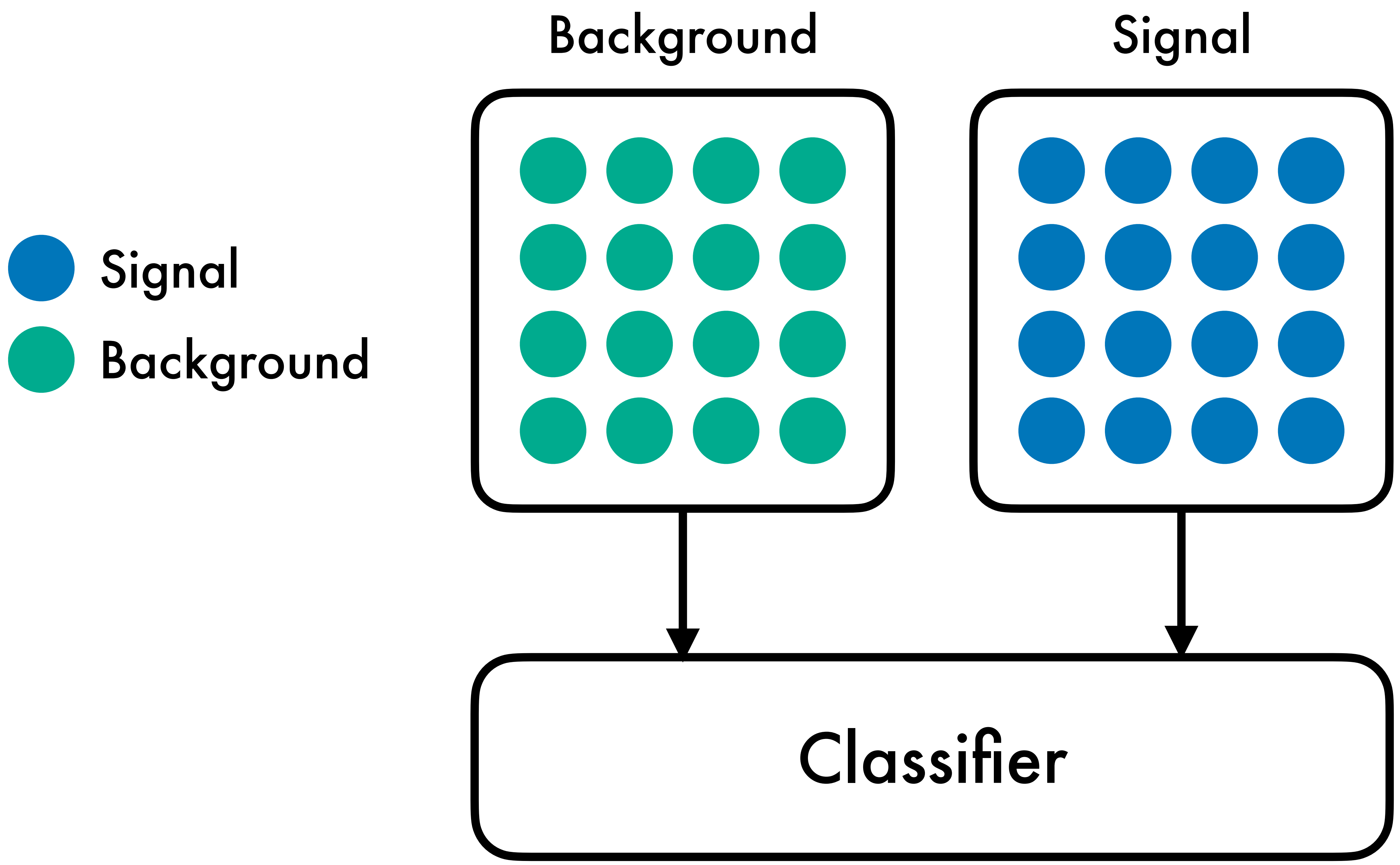


Metodiev, Nachman, Thaler [1708.02949]

Supervised versus IAD

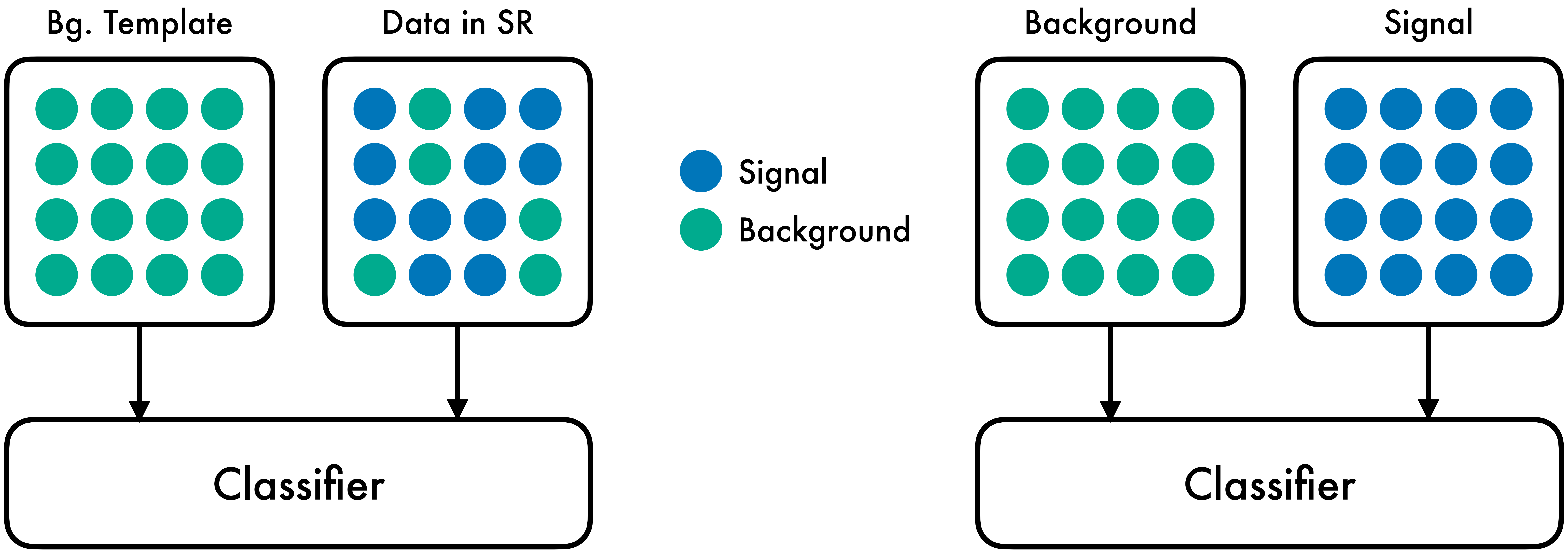


Supervised versus IAD



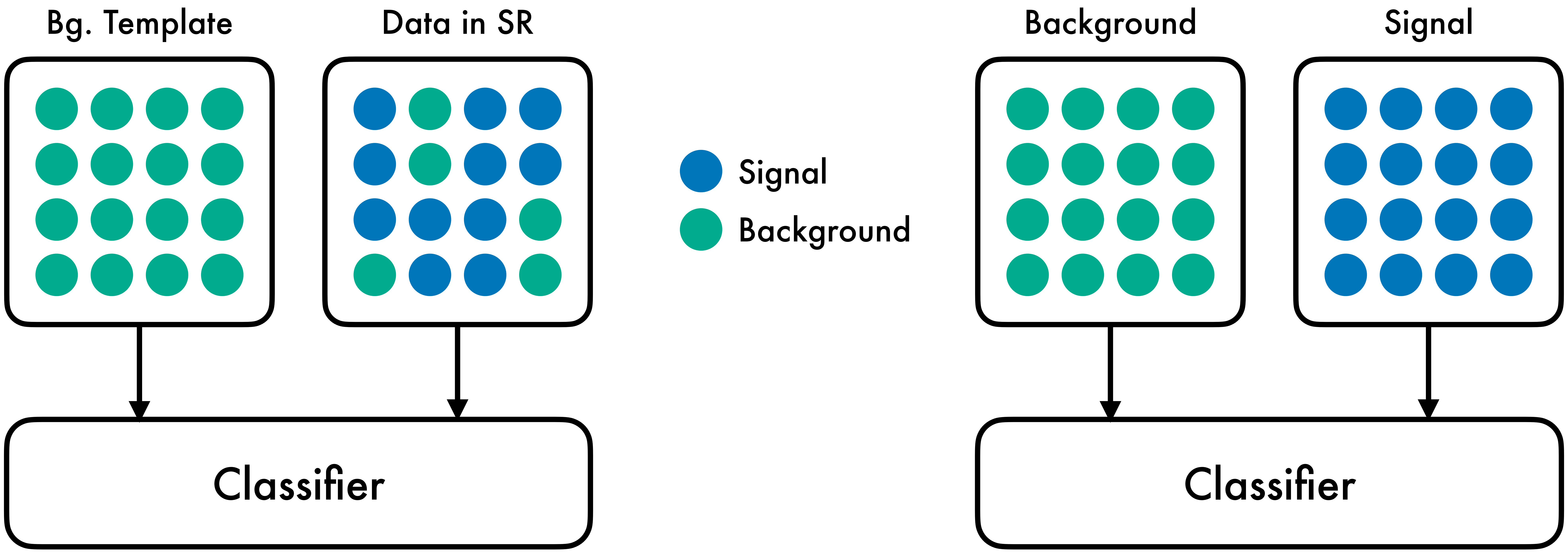
$$R_{\text{supervised}} = \frac{p_{\text{sig}}(x)}{p_{\text{bg}}(x)}$$

Supervised versus IAD



$$R_{\text{supervised}} = \frac{p_{\text{sig}}(x)}{p_{\text{bg}}(x)}$$

Supervised versus IAD

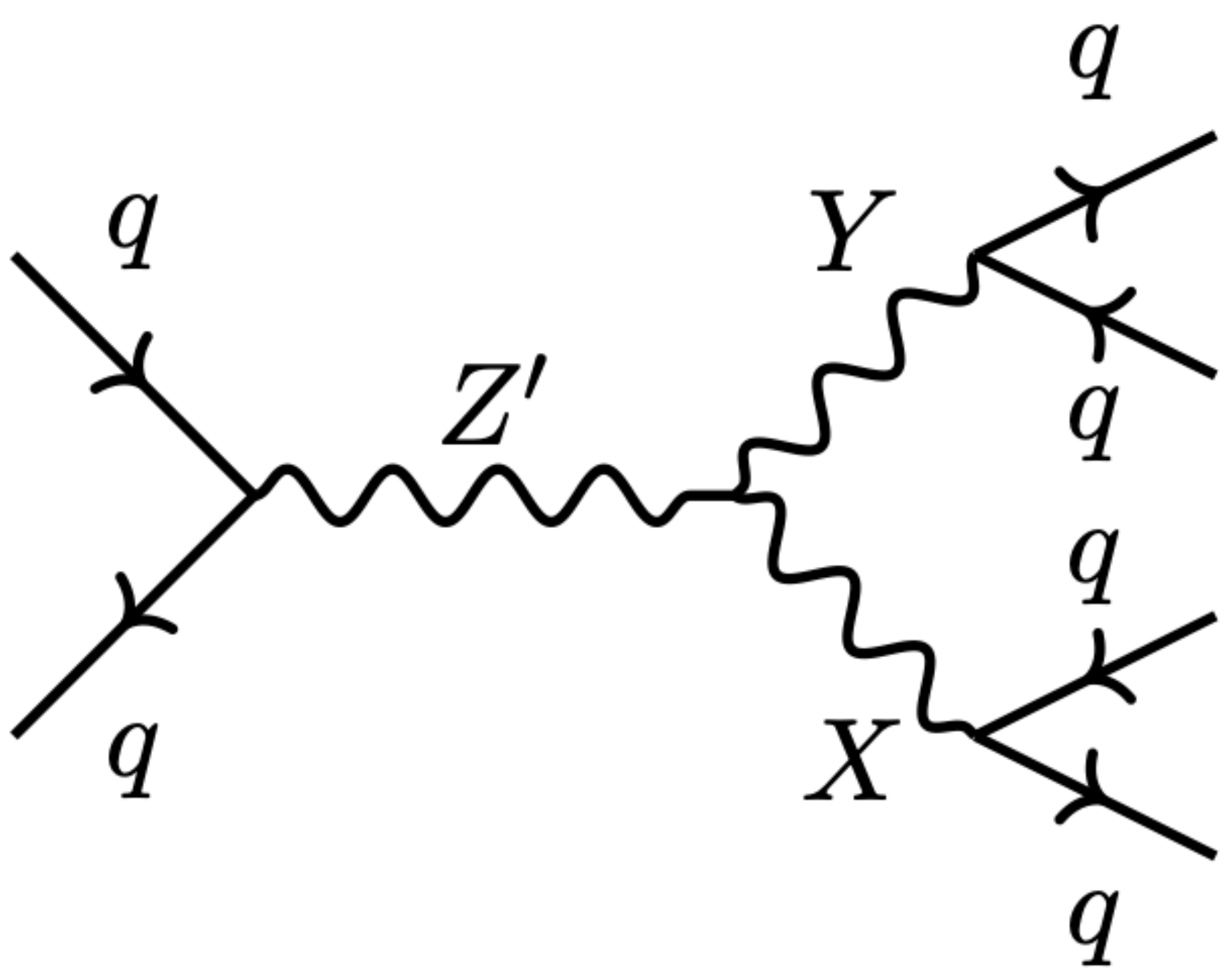


$$R_{\mathrm{IAD}} = \frac{p_{\mathrm{data}}(x)}{p_{\mathrm{bg}}(x)}$$

$$= \epsilon\, R_{\mathrm{supervised}} + (1 - \epsilon)$$

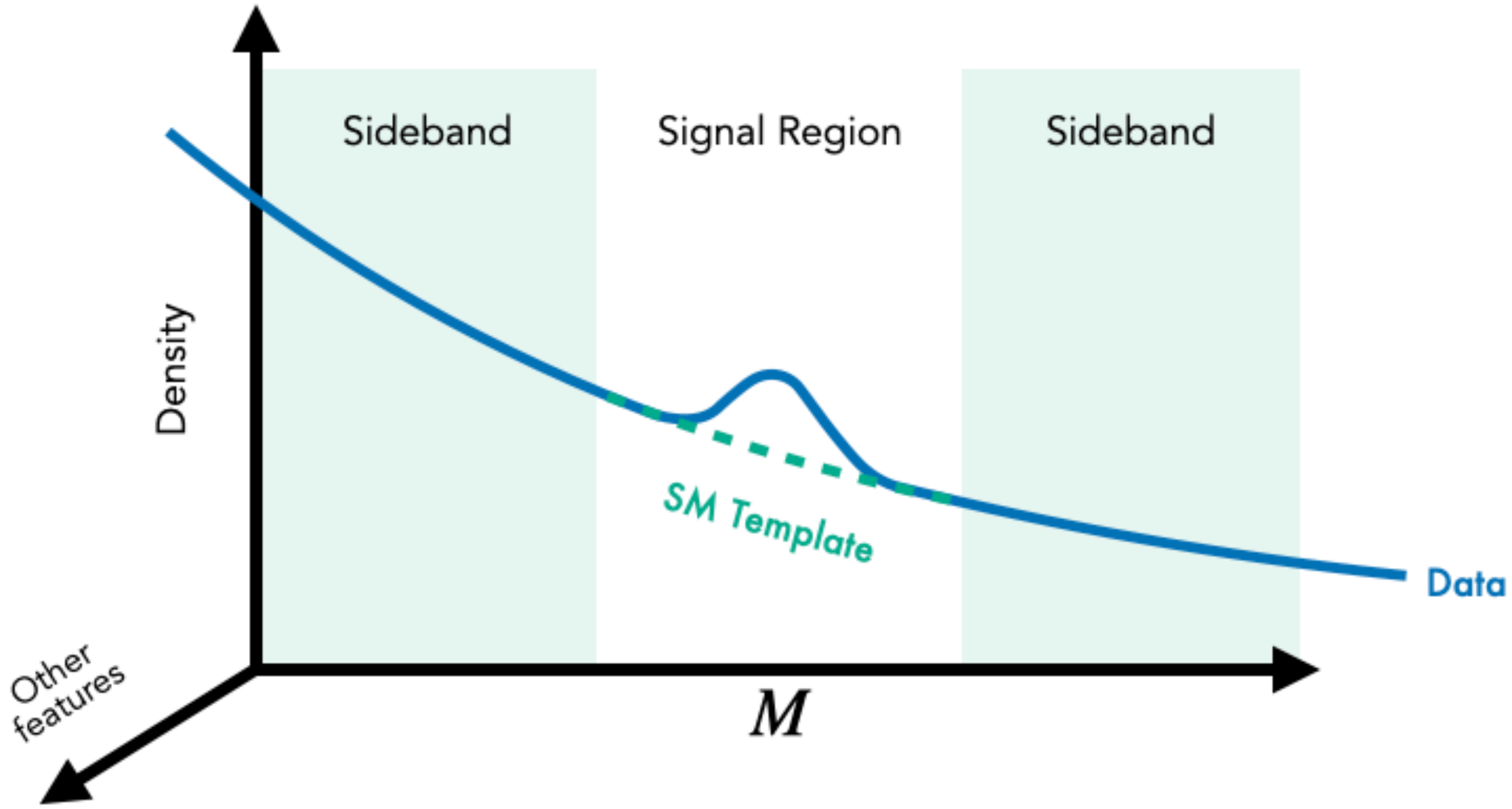
$$R_{\mathrm{supervised}} = \frac{p_{\mathrm{sig}}(x)}{p_{\mathrm{bg}}(x)}$$

CWoLa Hunting



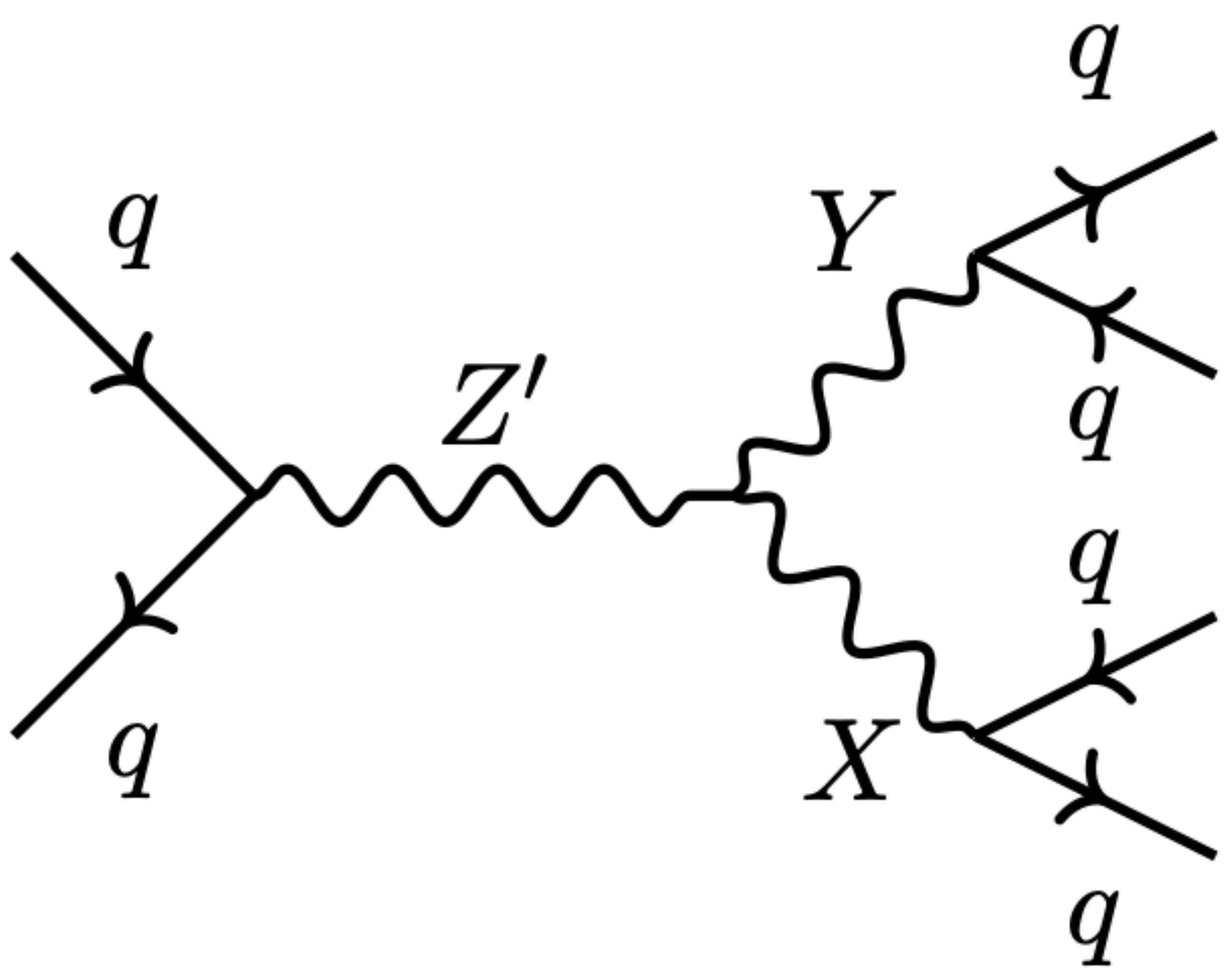
LHC Olympics

[Kasieczka et al: 2107.02821, 2101.08320]



[1902.02634]

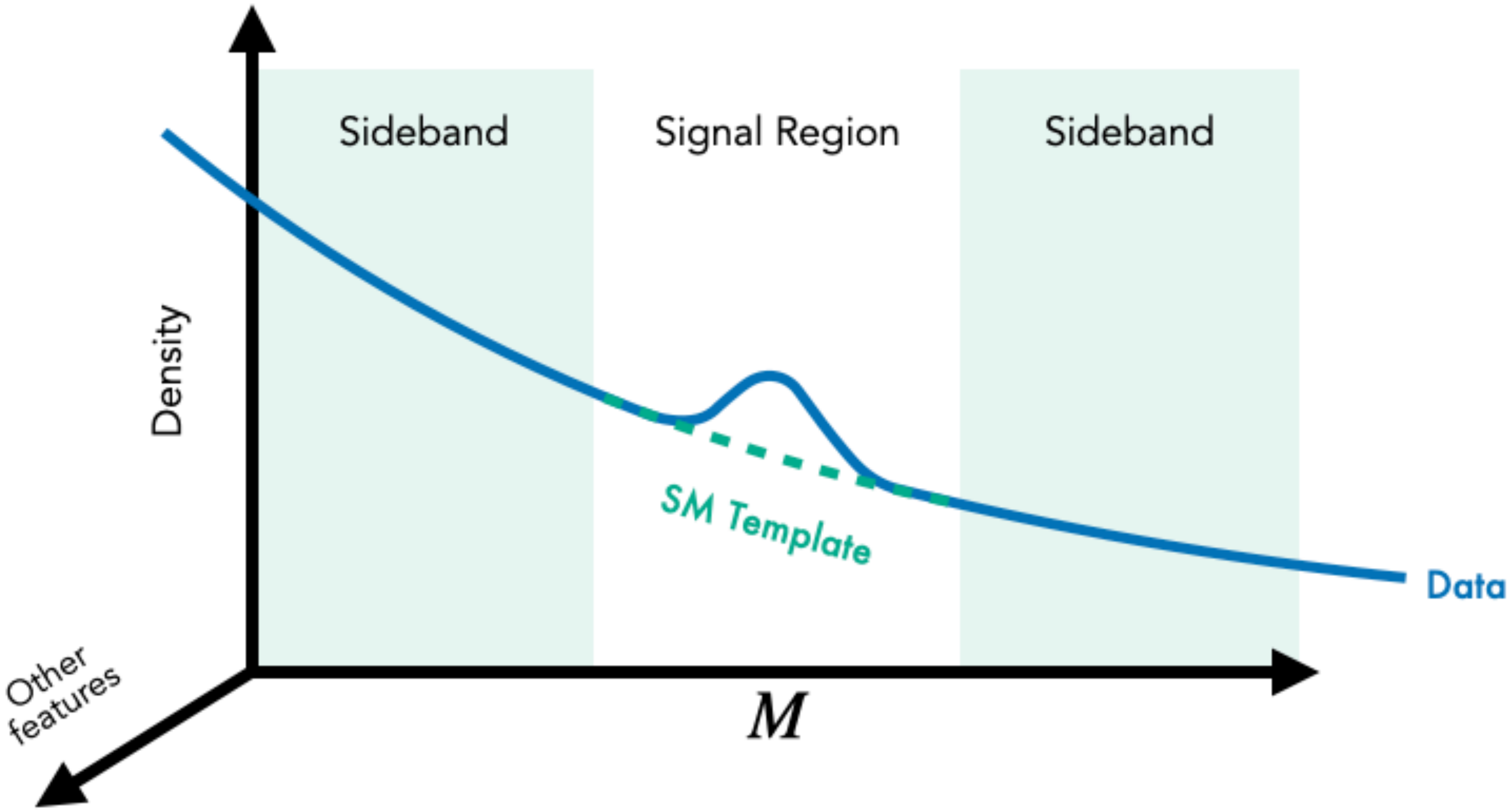
CWoLa Hunting



Resonant observable $m_{jj} = m_{Z'} > m_X, m_Y$

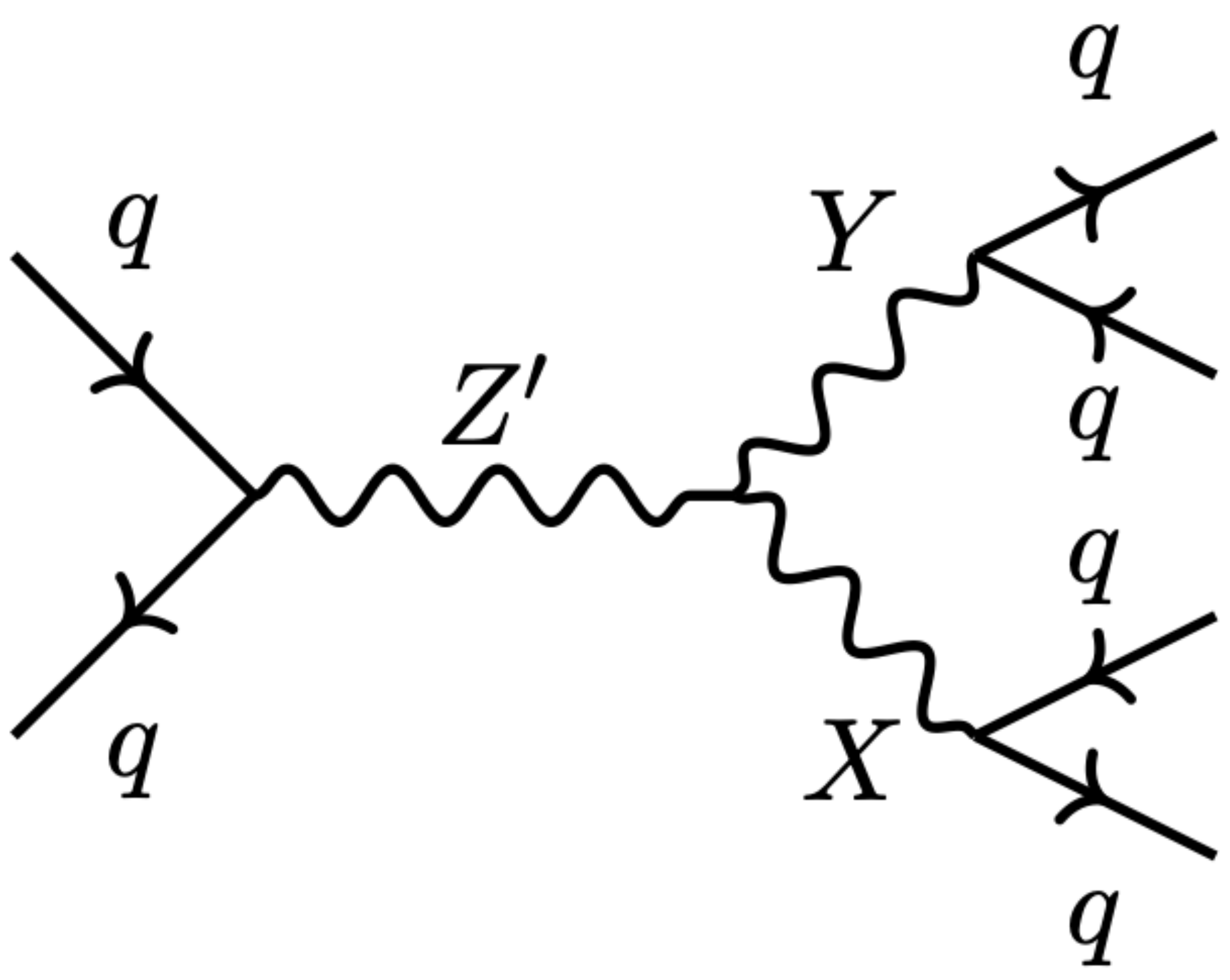
LHC Olympics

[Kasieczka et al: 2107.02821, 2101.08320]



[1902.02634]

CWoLa Hunting

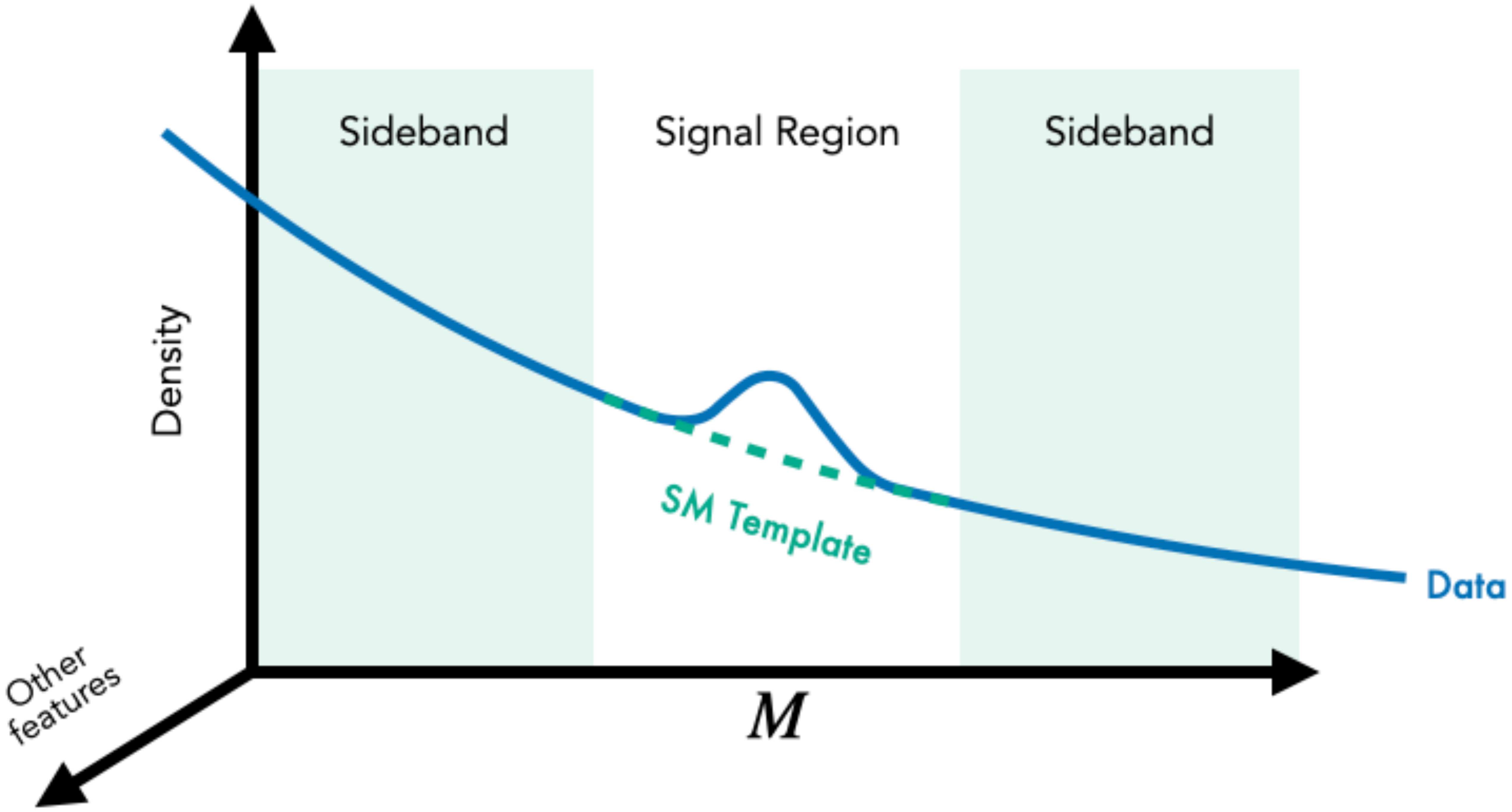


Resonant observable $m_{jj} = m_{Z'} > m_X, m_Y$

Other features $x = \{m_X, m_Y, \Delta m_j, \tau_{21}^{(1)}, \tau_{21}^{(2)}\}$

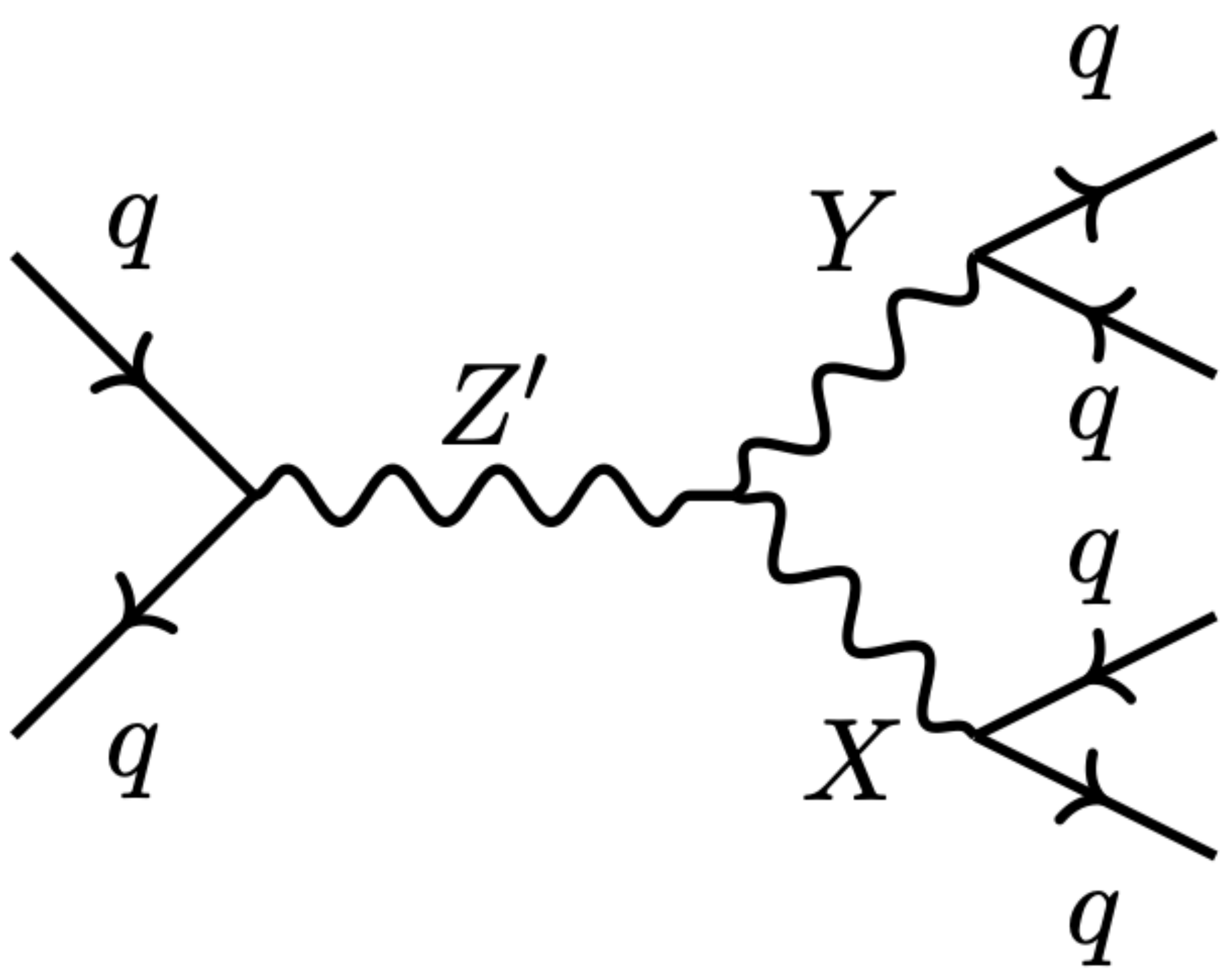
LHC Olympics

[Kasieczka et al: 2107.02821, 2101.08320]



[1902.02634]

CWoLa Hunting



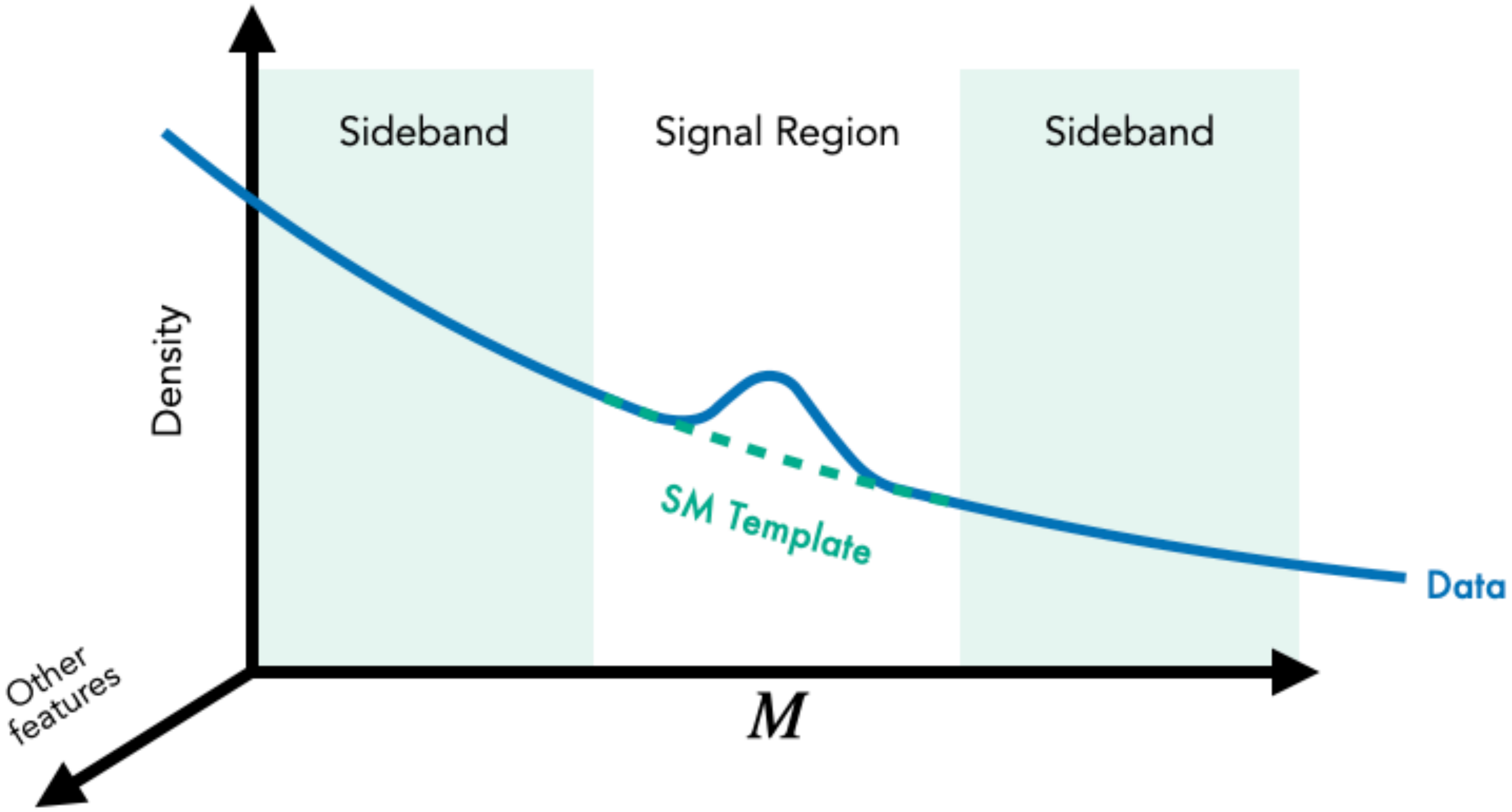
Resonant observable $m_{jj} = m_{Z'} > m_X, m_Y$

Other features $x = \{m_X, m_Y, \Delta m_j, \tau_{21}^{(1)}, \tau_{21}^{(2)}\}$

$$p_{\rm bg}(x \,|\, m_{jj} \in {\rm SR}) \approx p_{\rm bg}(x \,|\, m_{jj} \in {\rm SB}) \approx p_{\rm bg}(x)$$

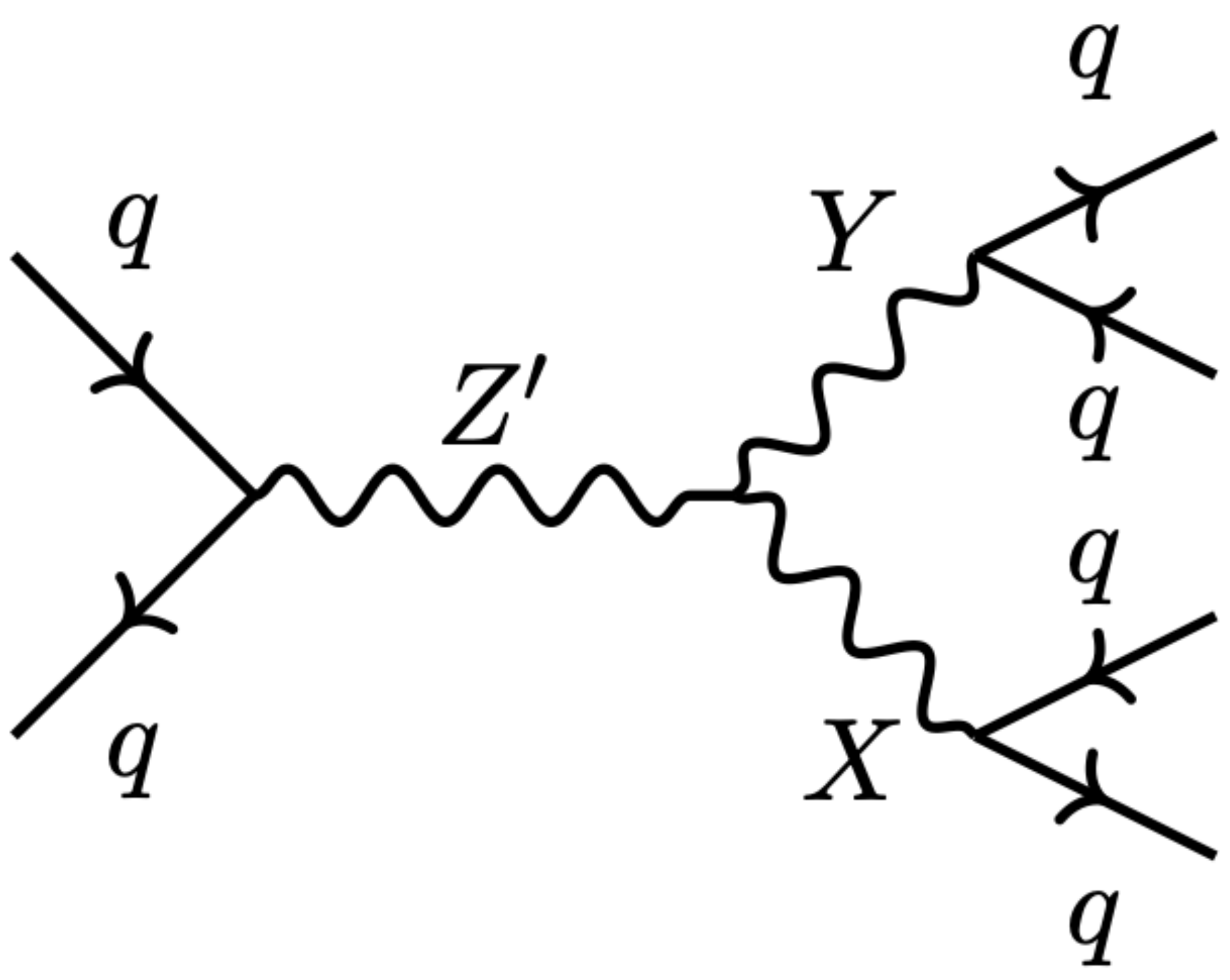
LHC Olympics

[Kasieczka et al: 2107.02821, 2101.08320]



[1902.02634]

CWoLa Hunting



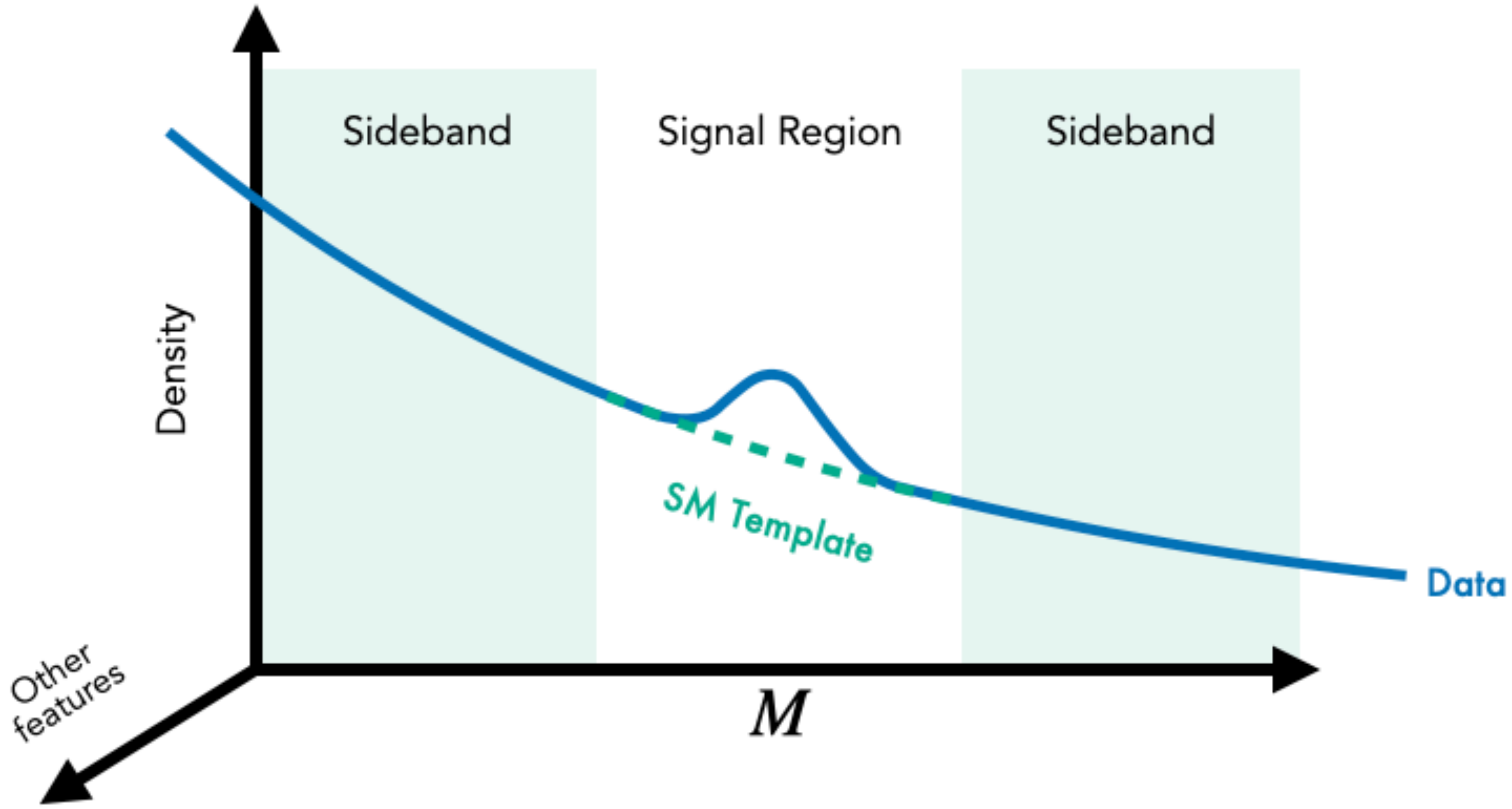
Resonant observable $m_{jj} = m_{Z'} > m_X, m_Y$

Other features $x = \{m_X, m_Y, \Delta m_j, \tau_{21}^{(1)}, \tau_{21}^{(2)}\}$

$$p_{\mathrm{bg}}(x \,|\, m_{jj} \in \mathrm{SR}) \approx p_{\mathrm{bg}}(x \,|\, m_{jj} \in \mathrm{SB}) \approx p_{\mathrm{bg}}(x)$$

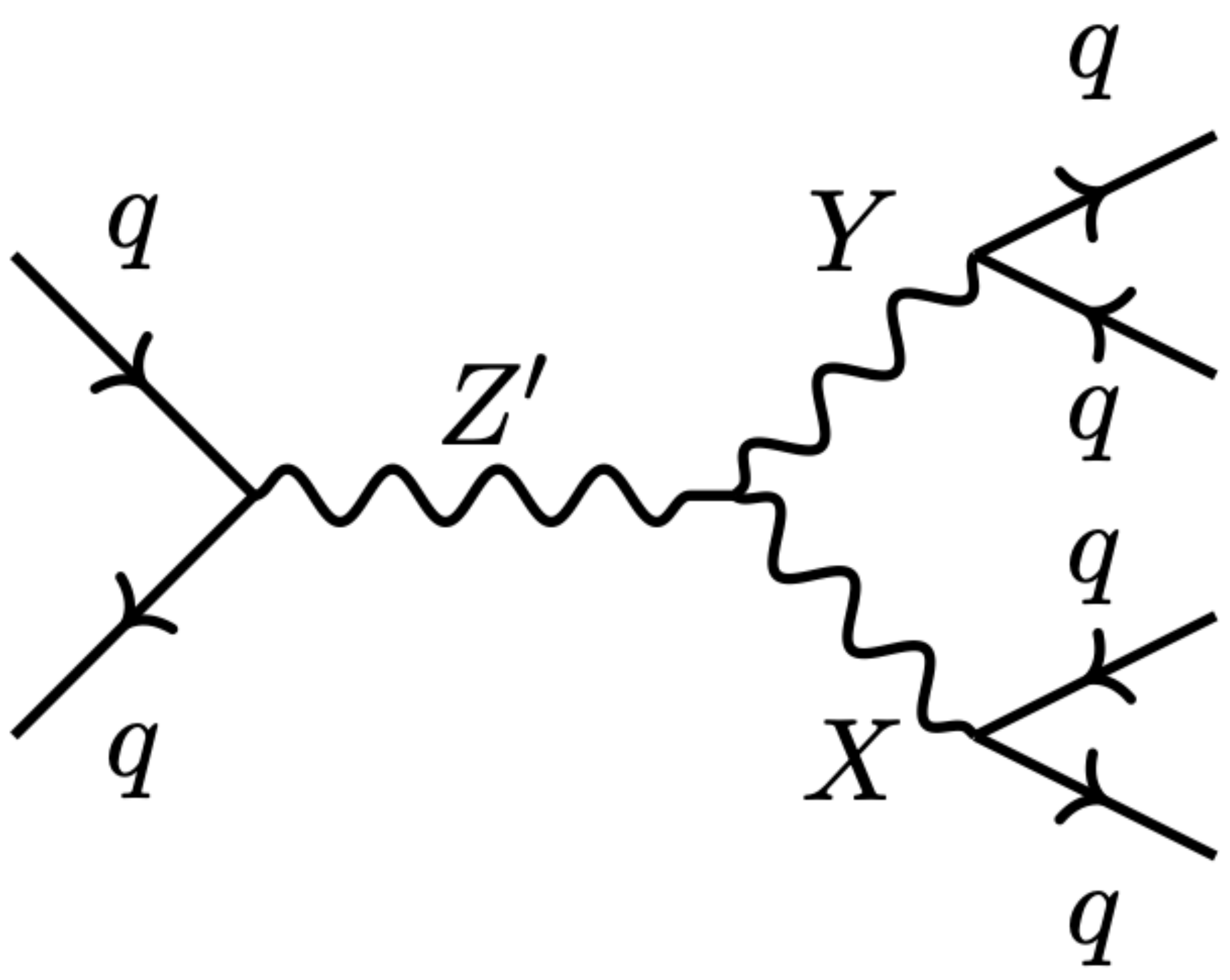
CWoLa Likelihood estimate

$$R_{\mathrm{CWoLa}} = \frac{p_{\mathrm{data}}(x \,|\, \mathrm{SR})}{p_{\mathrm{bg}}(x \,|\, \mathrm{SB})}$$



[1902.02634]

CWoLa Hunting



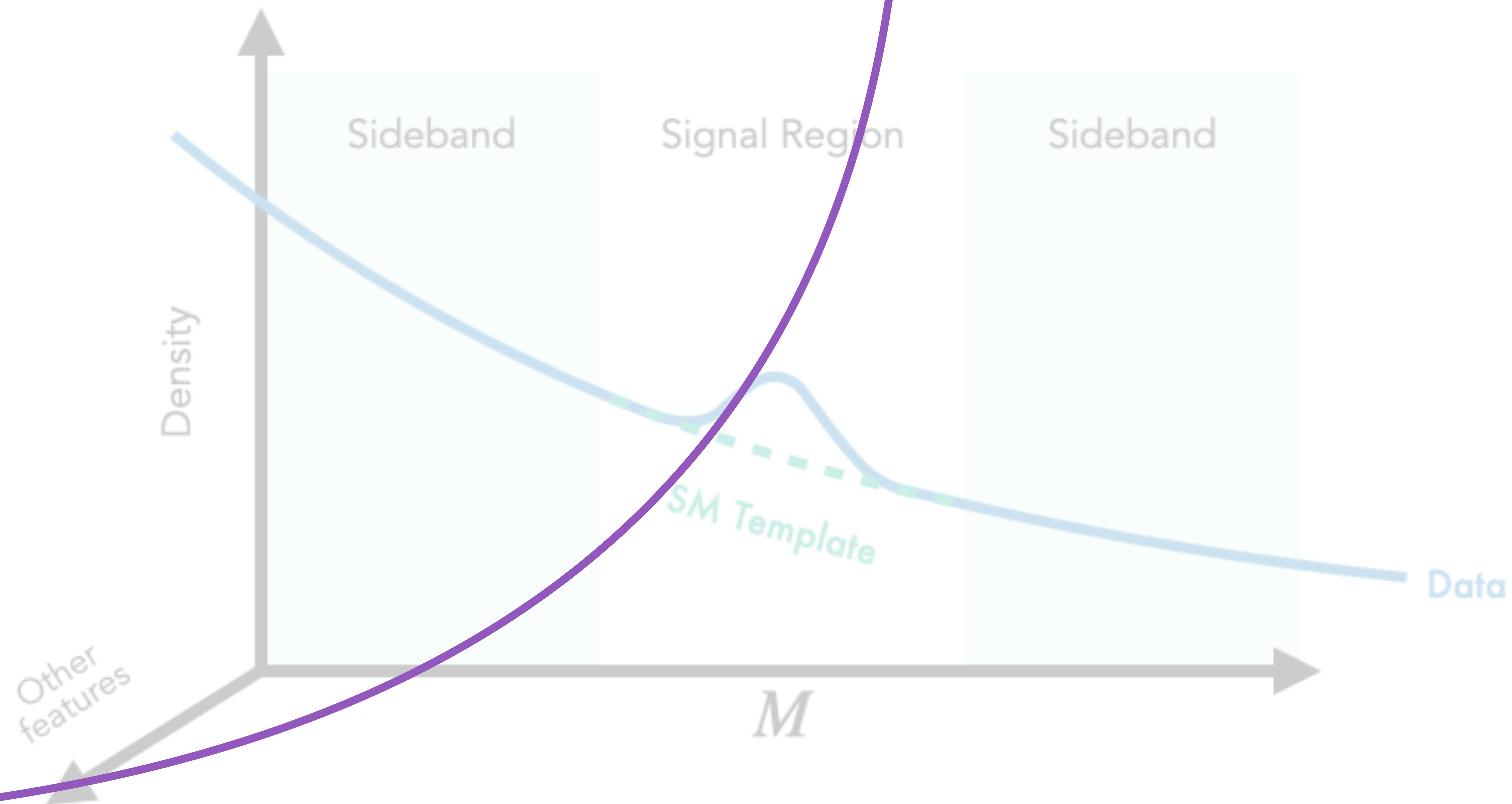
Resonant observable $m_{jj} = m_{Z'} > m_X, m_Y$

Other features $x = \{m_X, m_Y, \Delta m_j, \tau_{21}^{(1)}, \tau_{21}^{(2)}\}$

$$p_{\rm bg}(x\,|\,m_{jj} \in \mathrm{SR}) \approx p_{\rm bg}(x\,|\,m_{jj} \in \mathrm{SB}) \approx p_{\rm bg}(x)$$

CWoLa Likelihood estimate

$$R_{\rm CWoLa} = \frac{p_{\rm data}(x\,|\,\mathrm{SR})}{p_{\rm bg}(x\,|\,\mathrm{SB})} \approx \frac{p_{\rm data}(x\,|\,\mathrm{SR})}{p_{\rm bg}(x\,|\,\mathrm{SR})}$$



[1902.02634]

Can we do better?

Example II

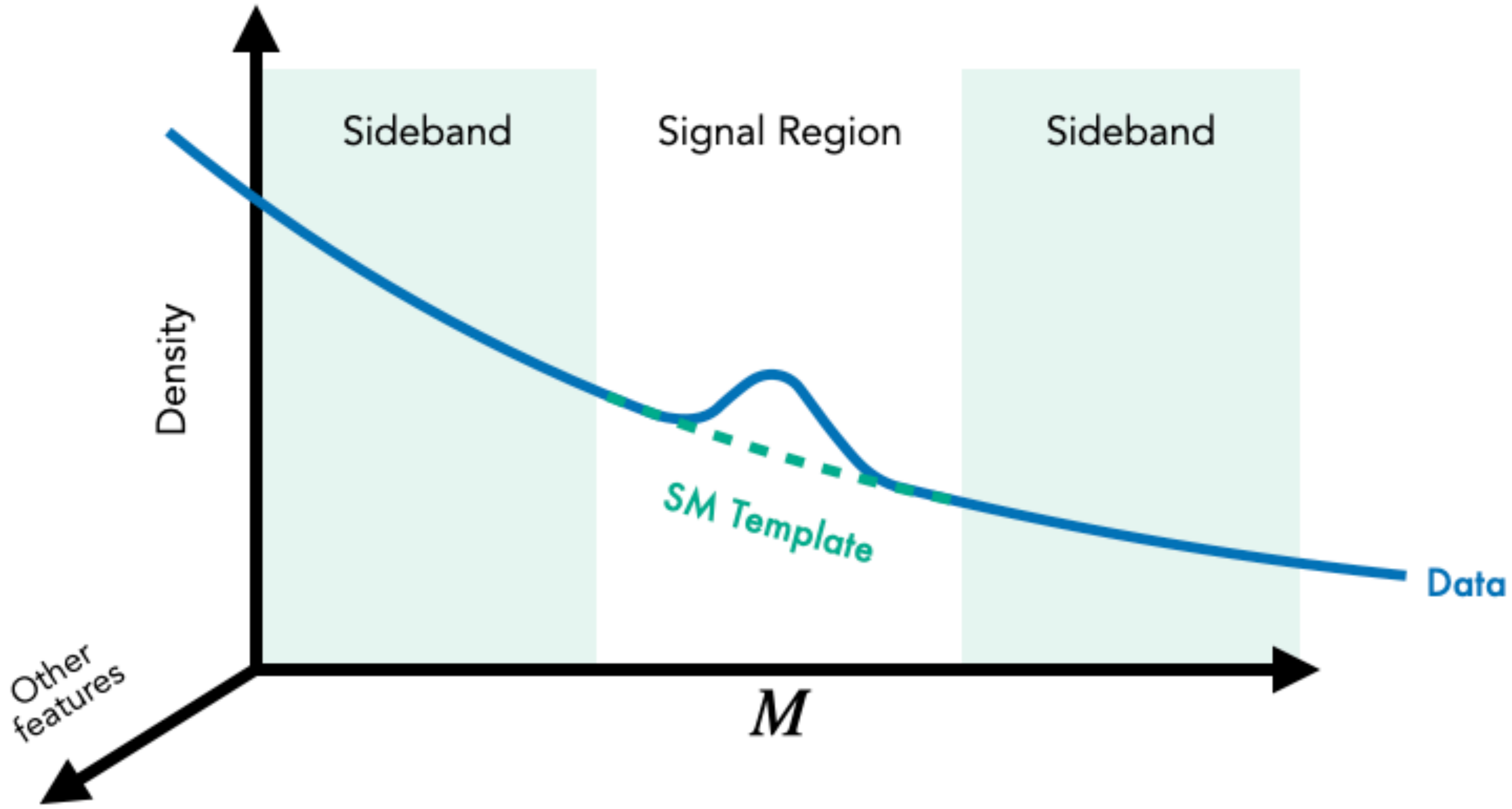
ANOmaly detection with Density Estimation (ANODE)

Nachman, Shih [2001.04990]

ANODE

CWoLa Likelihood estimate

$$R_{\mathrm{CWoLa}} = \frac{p_{\mathrm{data}}(x \mid \mathrm{SR})}{p_{\mathrm{bg}}(x \mid \mathrm{SB})}$$



[2001.04990]

ANODE

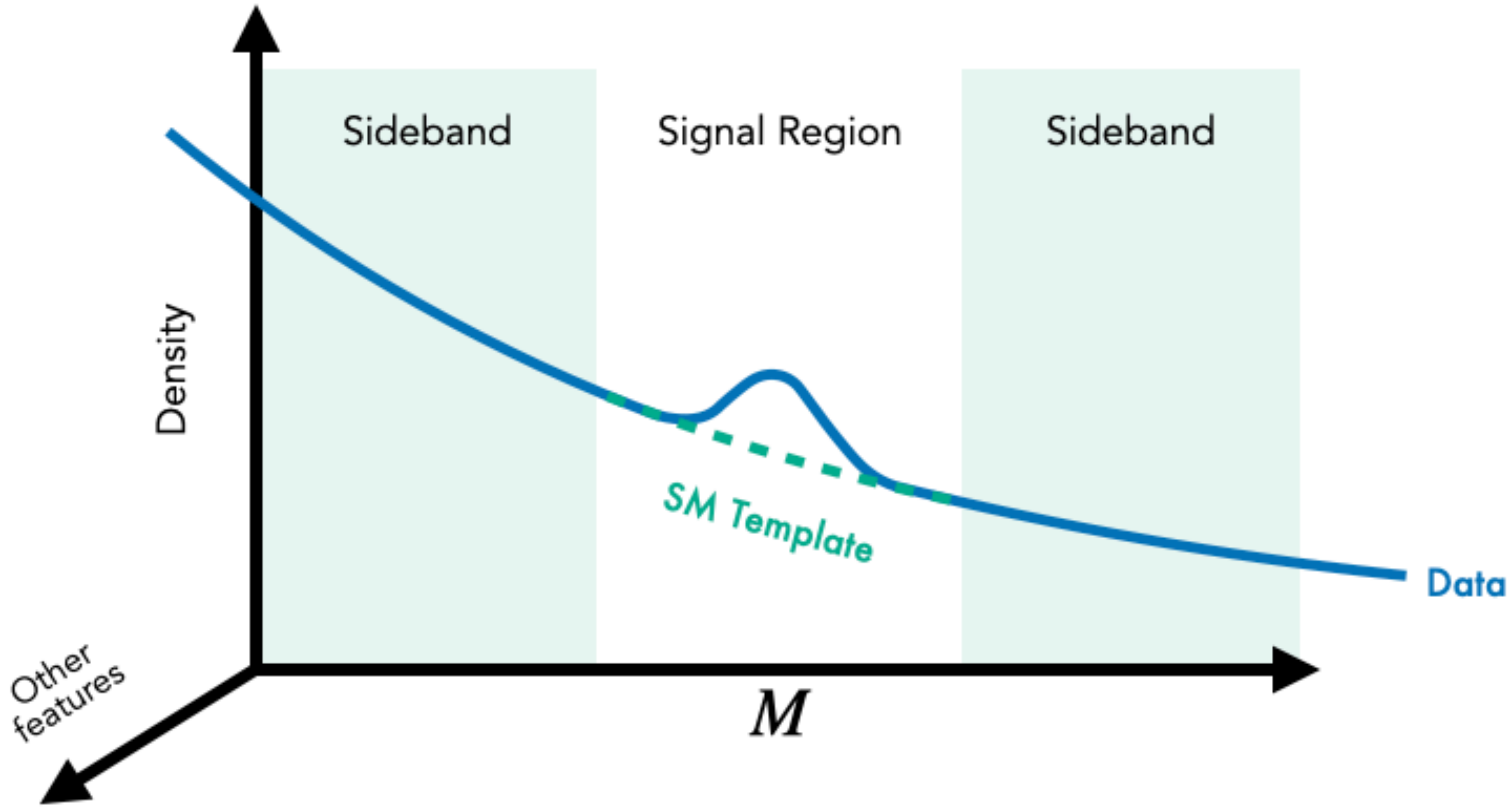
CWoLa Likelihood estimate

$$R_{\mathrm{CWoLa}} = \frac{p_{\mathrm{data}}(x\,|\,\mathrm{SR})}{p_{\mathrm{bg}}(x\,|\,\mathrm{SB})}$$

The ANODE method

$$p_{\omega_0}(x\,|\,m) \simeq p_{\mathrm{bg}}(x\,|\,m)$$

$$p_{\omega_1}(x\,|\,m) \simeq p_{\mathrm{data}}(x\,|\,m)$$



[2001.04990]

ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SB})}$$

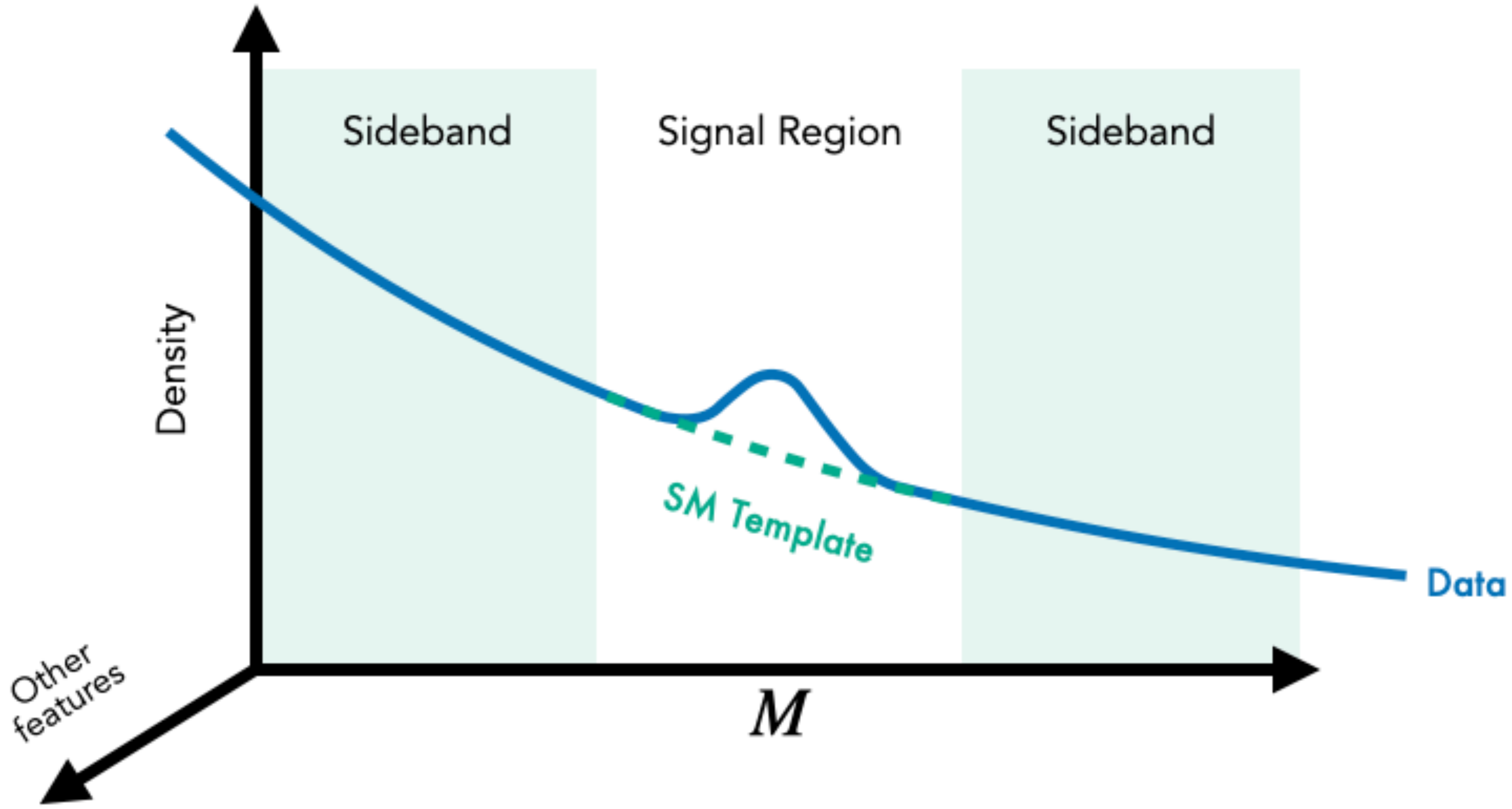
The ANODE method

NF

$$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$$

$$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$$

NF



[2001.04990]

ANDODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x \mid \text{SR})}{p_{\text{bg}}(x \mid \text{SB})}$$

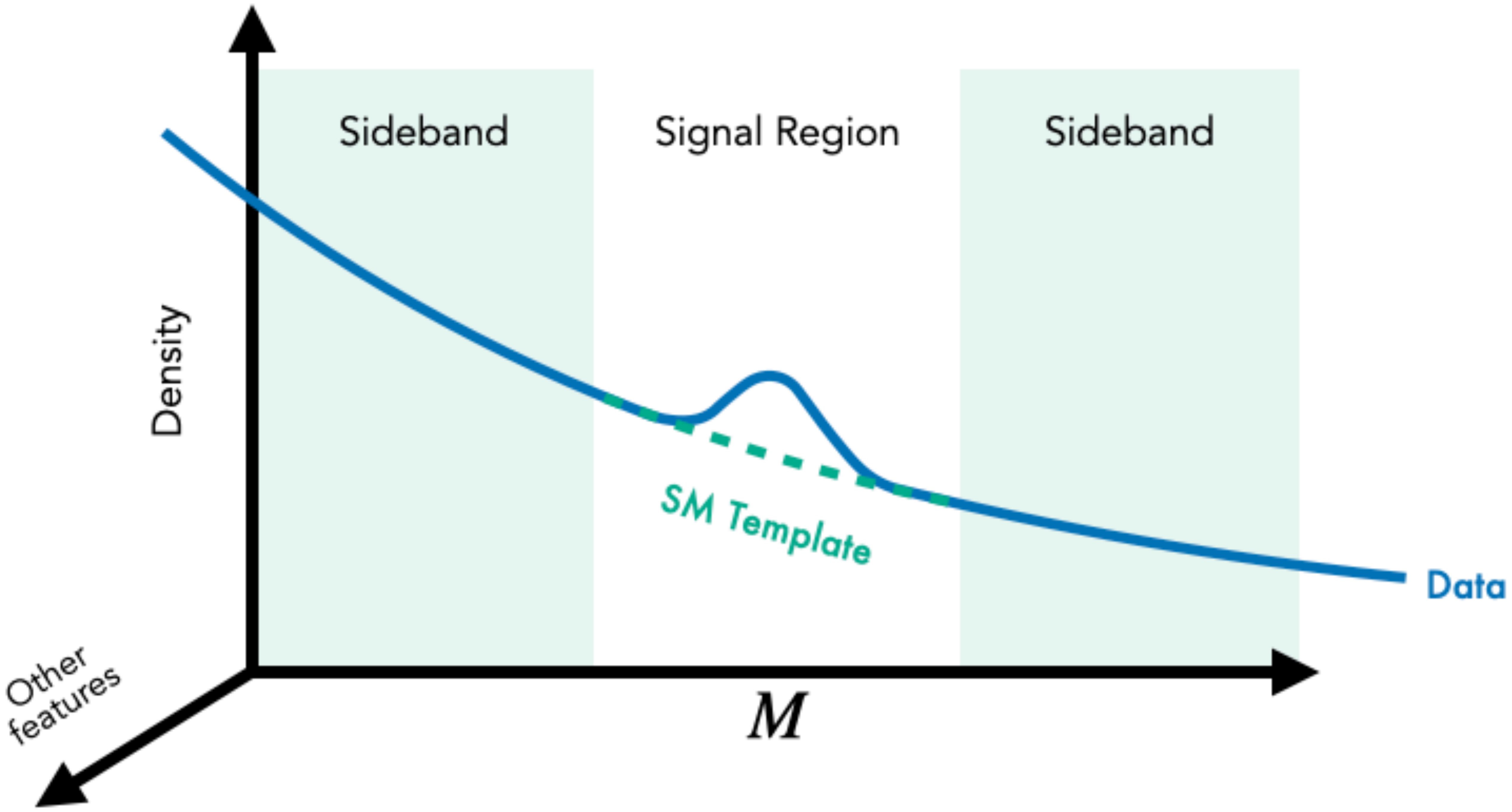
The ANODE method

NF

$$p_{\omega_0}(x \mid m) \simeq p_{\text{bg}}(x \mid m)$$ Trained in $m \in \text{SB}$

$$p_{\omega_1}(x \mid m) \simeq p_{\text{data}}(x \mid m)$$ Trained in $m \in \text{SR}$

NF



[2001.04990]

ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SB})}$$

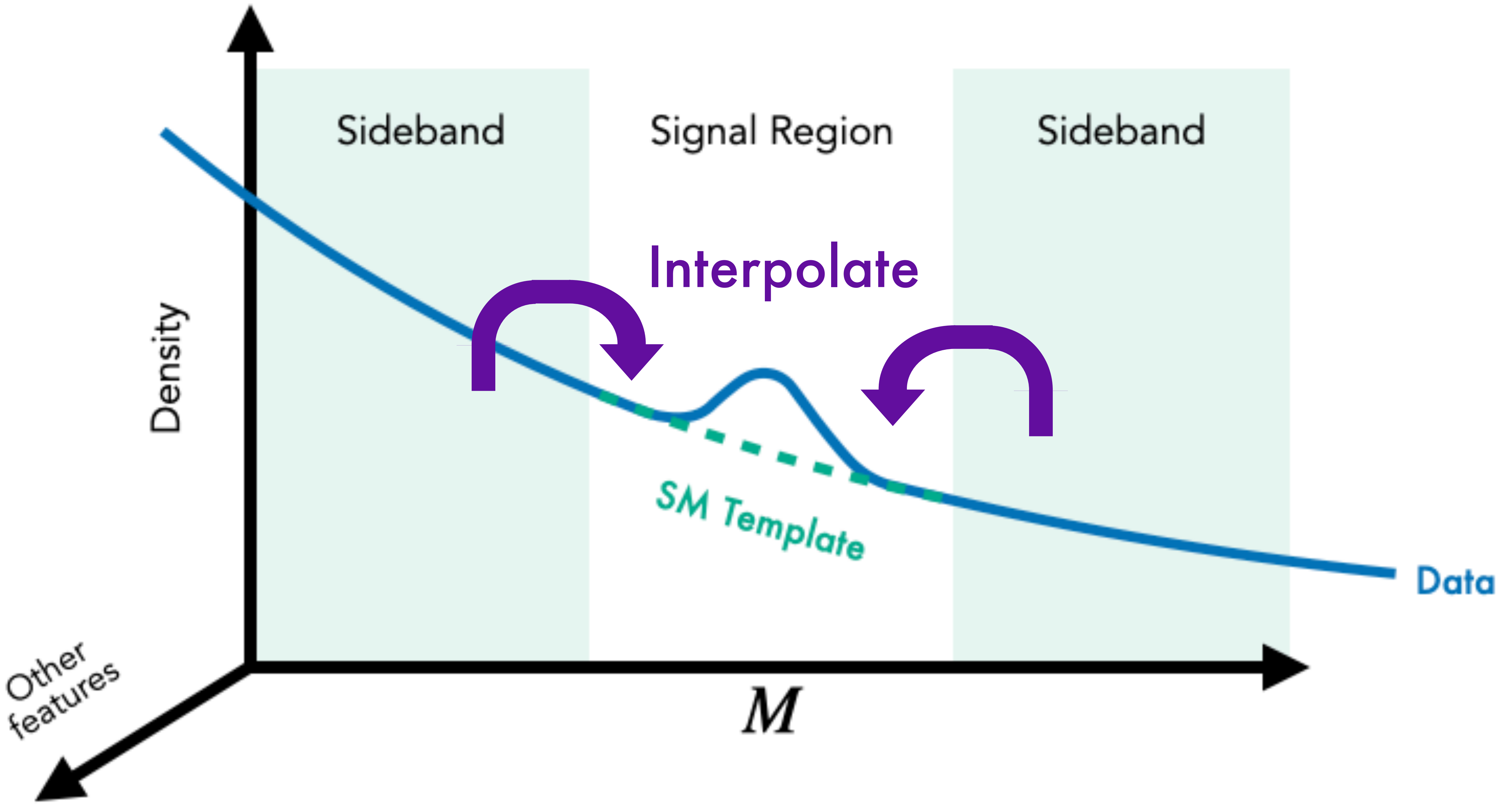
The ANODE method

NF

$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$ Trained in $m \in \text{SB}$

$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$ Trained in $m \in \text{SR}$

NF



[2001.04990]

ANODE

CWoLa Likelihood estimate

$$R_{\mathrm{CWoLa}} = \frac{p_{\mathrm{data}}(x\,|\,\mathrm{SR})}{p_{\mathrm{bg}}(x\,|\,\mathrm{SB})}$$

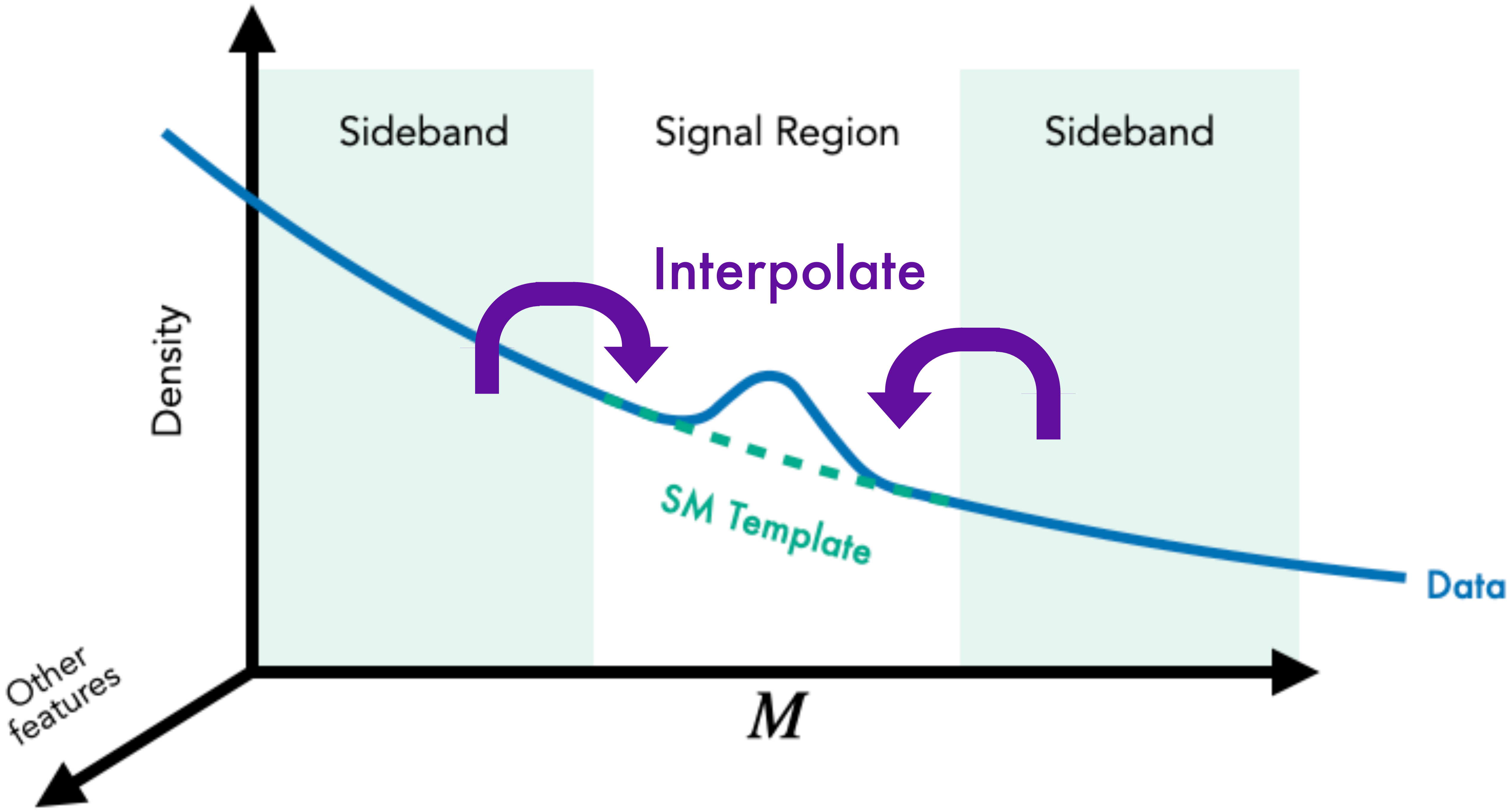
The ANODE method

NF

$$p_{\omega_0}(x\,|\,m) \simeq p_{\mathrm{bg}}(x\,|\,m)$$ Trained in $m \in \mathrm{SB}$

$$p_{\omega_1}(x\,|\,m) \simeq p_{\mathrm{data}}(x\,|\,m)$$ Trained in $m \in \mathrm{SR}$

NF



[2001.04990]

ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SR})}$$

ANODE Likelihood estimate

$$R_{\text{ANODE}} = \frac{p_{\omega_1}(x\,|\,\text{SR})}{p_{\omega_0}(x\,|\,\text{SR})} \simeq \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SR})}$$

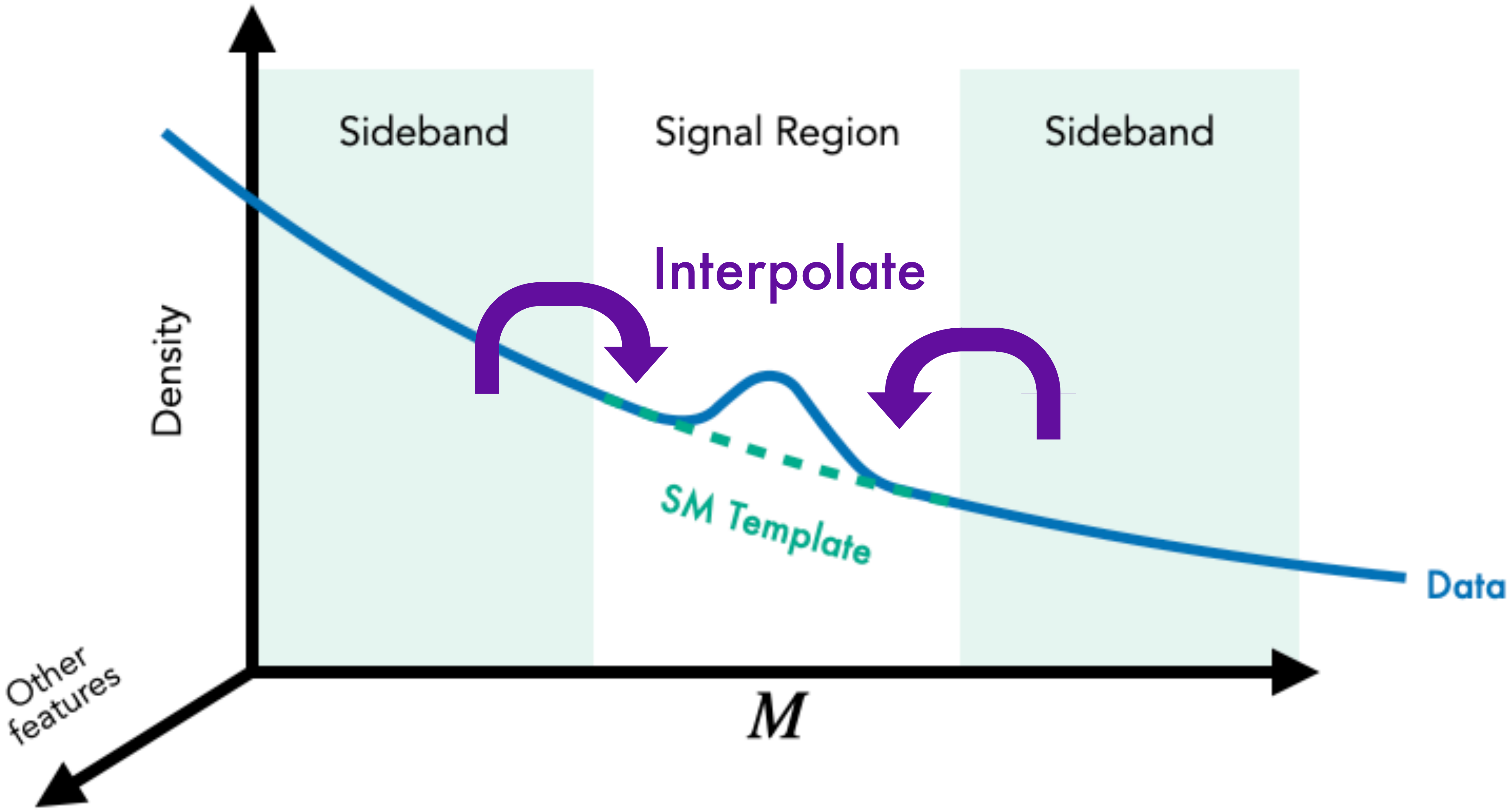
The ANODE method

NF

$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$ Trained in $m \in \text{SB}$

$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$ Trained in $m \in \text{SR}$

NF



[2001.04990]

Are we already happy?

CWoLA versus ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SB})}$$

Pros and cons:

[1902.02634]

ANODE Likelihood estimate

$$R_{\text{ANODE}} = \frac{p_{\omega_1}(x\,|\,\text{SR})}{p_{\omega_0}(x\,|\,\text{SR})}$$

Pros and cons:

[2001.04990]

CWoLA versus ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SB})}$$

Pros and cons:

⊕ Classification is easy and precise

[1902.02634]

ANODE Likelihood estimate

$$R_{\text{ANODE}} = \frac{p_{\omega_1}(x\,|\,\text{SR})}{p_{\omega_0}(x\,|\,\text{SR})}$$

Pros and cons:

[2001.04990]

CWoLA versus ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SB})}$$

Pros and cons:

⊕ Classification is easy and precise

⊖ Sensitive to correlations between m_{jj} and other features x

[1902.02634]

↔

ANODE Likelihood estimate

$$R_{\text{ANODE}} = \frac{p_{\omega_1}(x\,|\,\text{SR})}{p_{\omega_0}(x\,|\,\text{SR})}$$

Pros and cons:

[2001.04990]

CWoLA versus ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x \,|\, \text{SR})}{p_{\text{bg}}(x \,|\, \text{SB})}$$

Pros and cons:

⊕ Classification is easy and precise

⊖ Sensitive to correlations between m_{jj} and other features x

[1902.02634]

ANODE Likelihood estimate

$$R_{\text{ANODE}} = \frac{p_{\omega_1}(x \,|\, \text{SR})}{p_{\omega_0}(x \,|\, \text{SR})}$$

Pros and cons:

⊕ Robust against correlations

[2001.04990]

CWoLA versus ANODE

CWoLa Likelihood estimate

$$R_{\text{CWoLa}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SB})}$$

Pros and cons:

⊕ Classification is easy and precise

⊖ Sensitive to correlations between m_{jj} and other features x

[1902.02634]

ANODE Likelihood estimate

$$R_{\text{ANODE}} = \frac{p_{\omega_1}(x\,|\,\text{SR})}{p_{\omega_0}(x\,|\,\text{SR})}$$

Pros and cons:

⊕ Robust against correlations

⊖ Less powerful and sensitive than classification

[2001.04990]

Can we get the best of both worlds?

Example III

Classifying Anomalies THrough Outer Density Estimation (CATHODE)

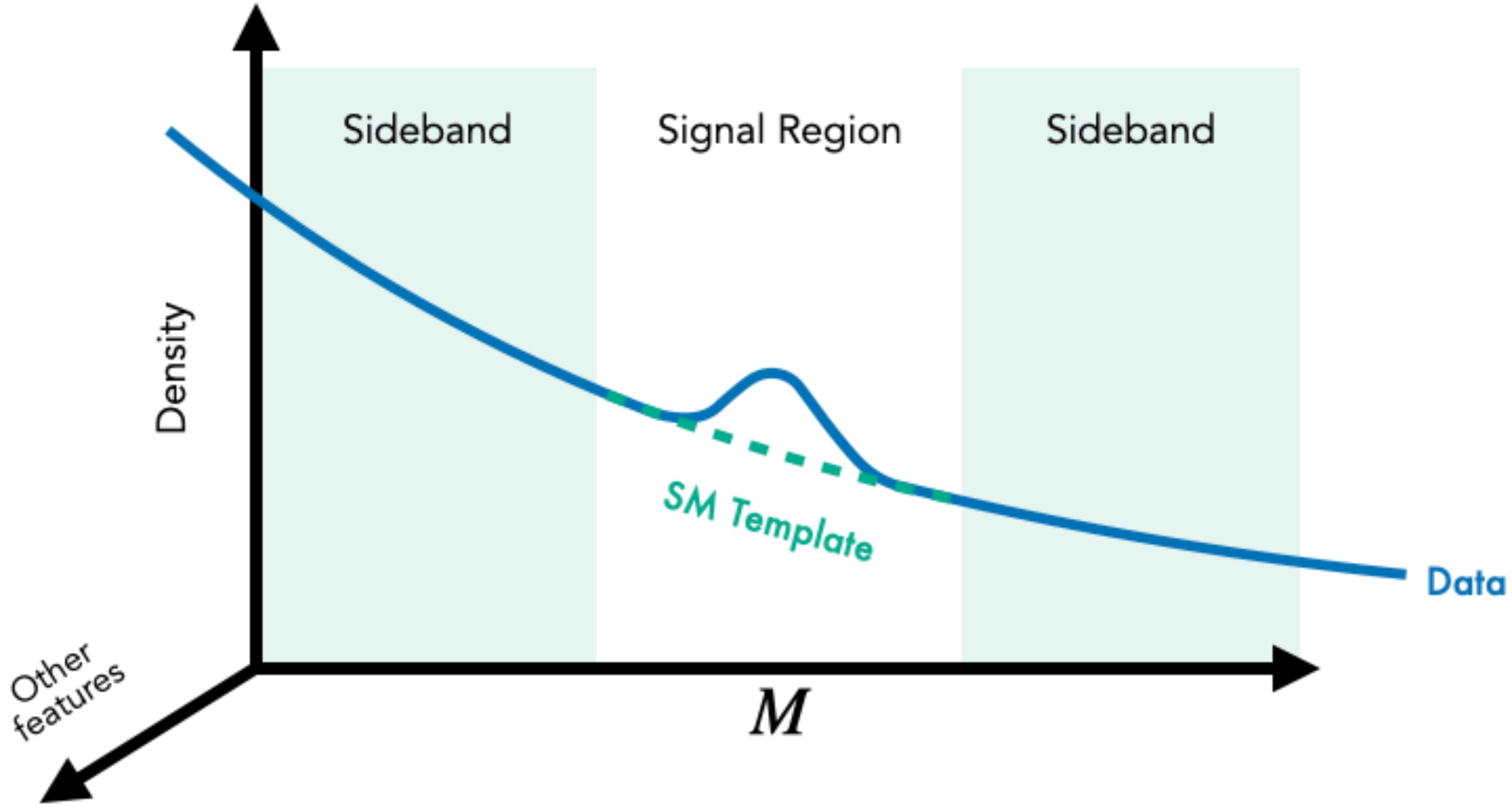
Hallin, Isaacson, Kasieczka, Krause, Nachman, Quadfasel, Schlaffer, Shih, Sommerhalder [2109.00546]

Best of both worlds — CATHODE

The CATHODE method

$p_{\omega_0}(x\,|\,m) \simeq p_{\mathrm{bg}}(x\,|\,m)$ Trained in $m \in \mathrm{SB}$

~~$p_{\omega_1}(x\,|\,m) \simeq p_{\mathrm{data}}(x\,|\,m)$~~



[2109.00546]

Best of both worlds — CATHODE

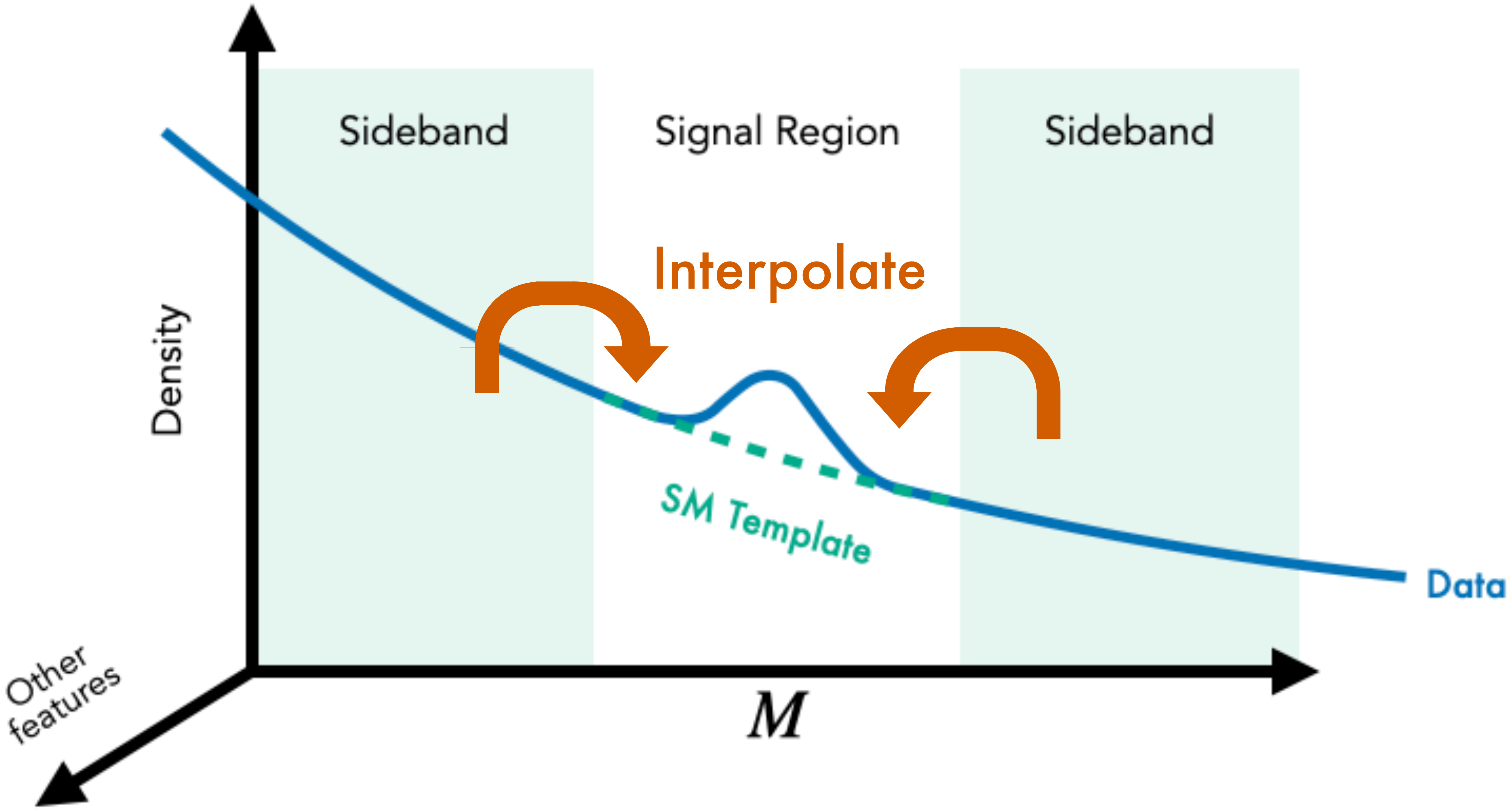
The CATHODE method

$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$ Trained in $m \in \text{SB}$

~~$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$~~

1. Interpolate **SM background template** to SR and sample:

$$\hat{x}_{\text{bg}} \sim p_{\omega_0}(x\,|\,m \in \text{SR}) \simeq p_{\text{bg}}(x\,|\,\text{SR})$$



[2109.00546]

Best of both worlds — CATHODE

The CATHODE method

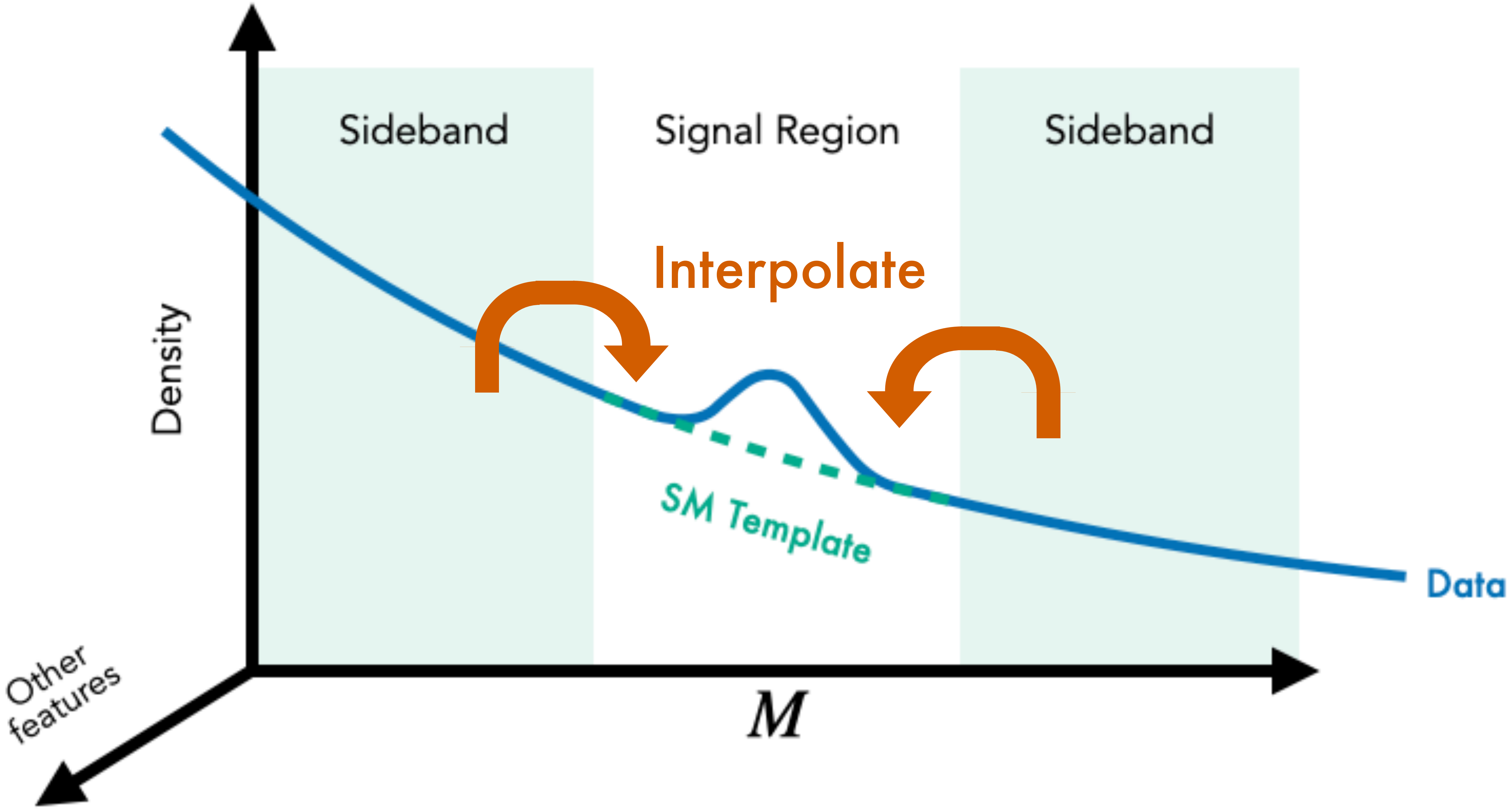
$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$ Trained in $m \in \text{SB}$

~~$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$~~

1. Interpolate **SM background template** to SR and sample:

   $$\hat{x}_{\text{bg}} \sim p_{\omega_0}(x\,|\,m \in \text{SR}) \simeq p_{\text{bg}}(x\,|\,\text{SR})$$

2. Then **train classifier** between $\hat{x}_{\text{bg}}$ and $x \sim p_{\text{data}}(x\,|\,\text{SR})$ as in **CWoLA**



[2109.00546]

Best of both worlds — CATHODE

The CATHODE method

$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$ Trained in $m \in \text{SB}$

~~$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$~~

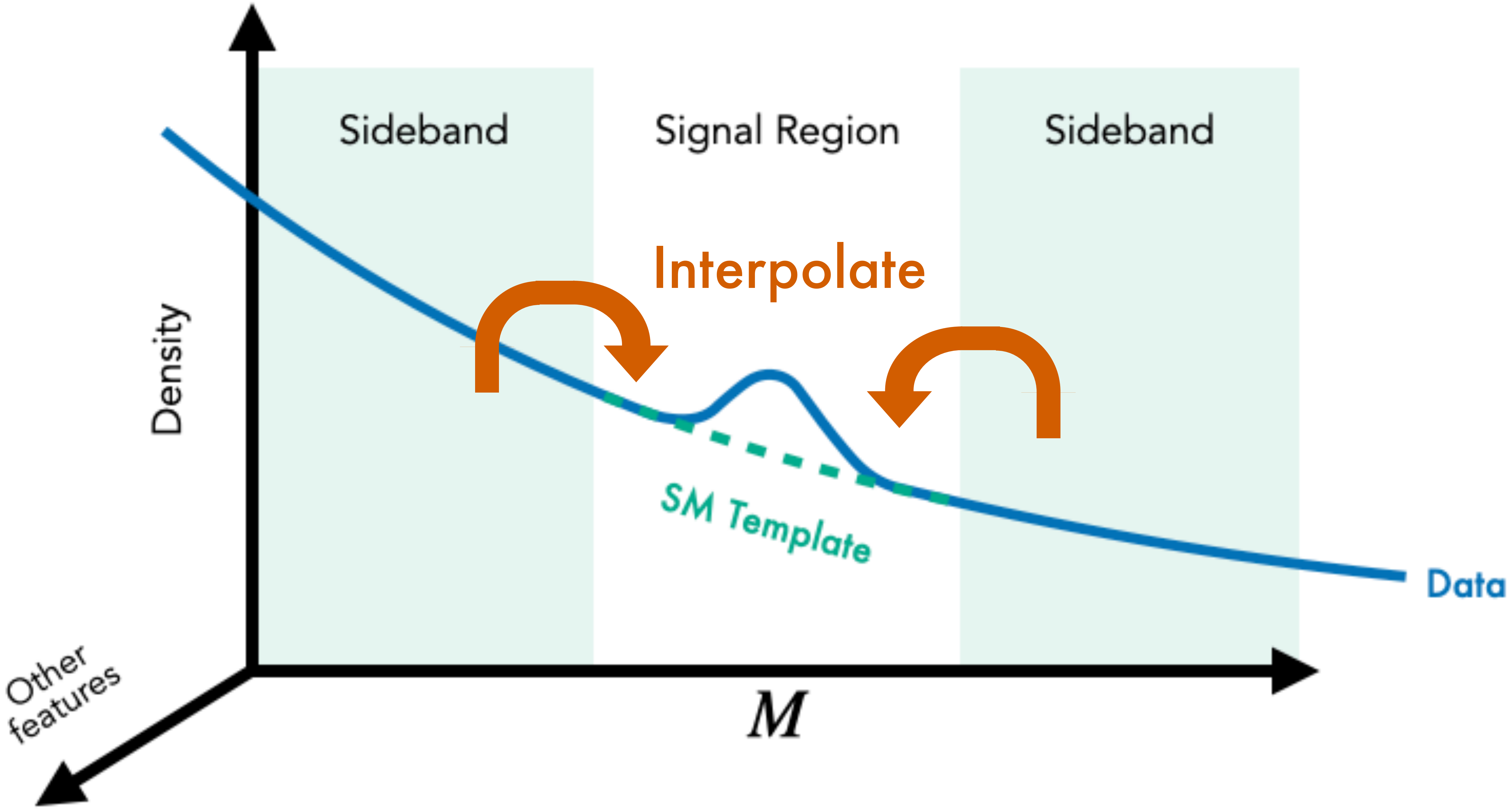
1. Interpolate **SM background template** to SR and sample:

   $$\hat{x}_{\text{bg}} \sim p_{\omega_0}(x\,|\,m \in \text{SR}) \simeq p_{\text{bg}}(x\,|\,\text{SR})$$

2. Then **train classifier** between $\hat{x}_{\text{bg}}$ and $x \sim p_{\text{data}}(x\,|\,\text{SR})$ as in **CWoLA**

CATHODE Likelihood estimate

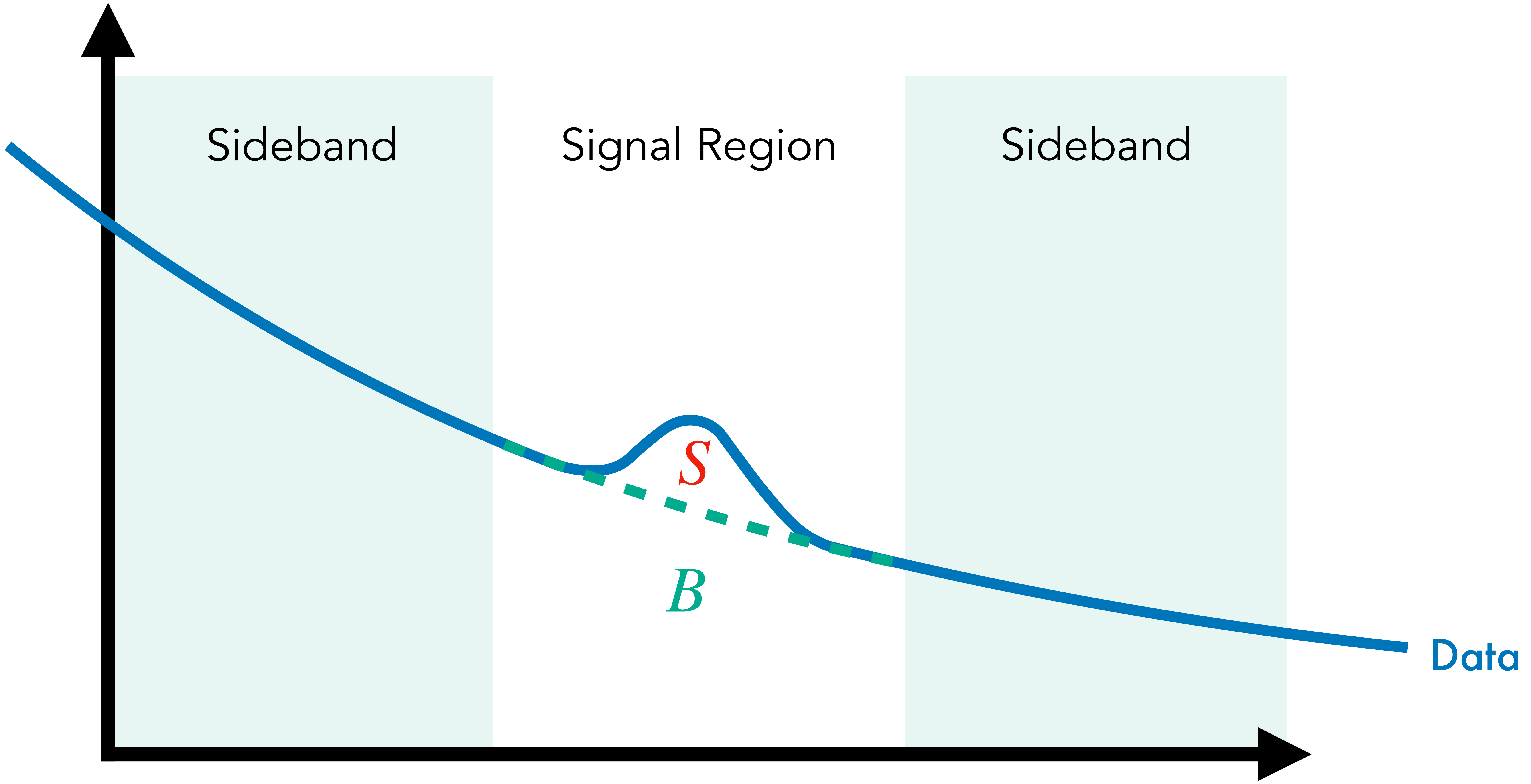
$$R_{\text{CATHODE}} = \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\omega_0}(x\,|\,\text{SR})} \simeq \frac{p_{\text{data}}(x\,|\,\text{SR})}{p_{\text{bg}}(x\,|\,\text{SR})}$$



[2109.00546]

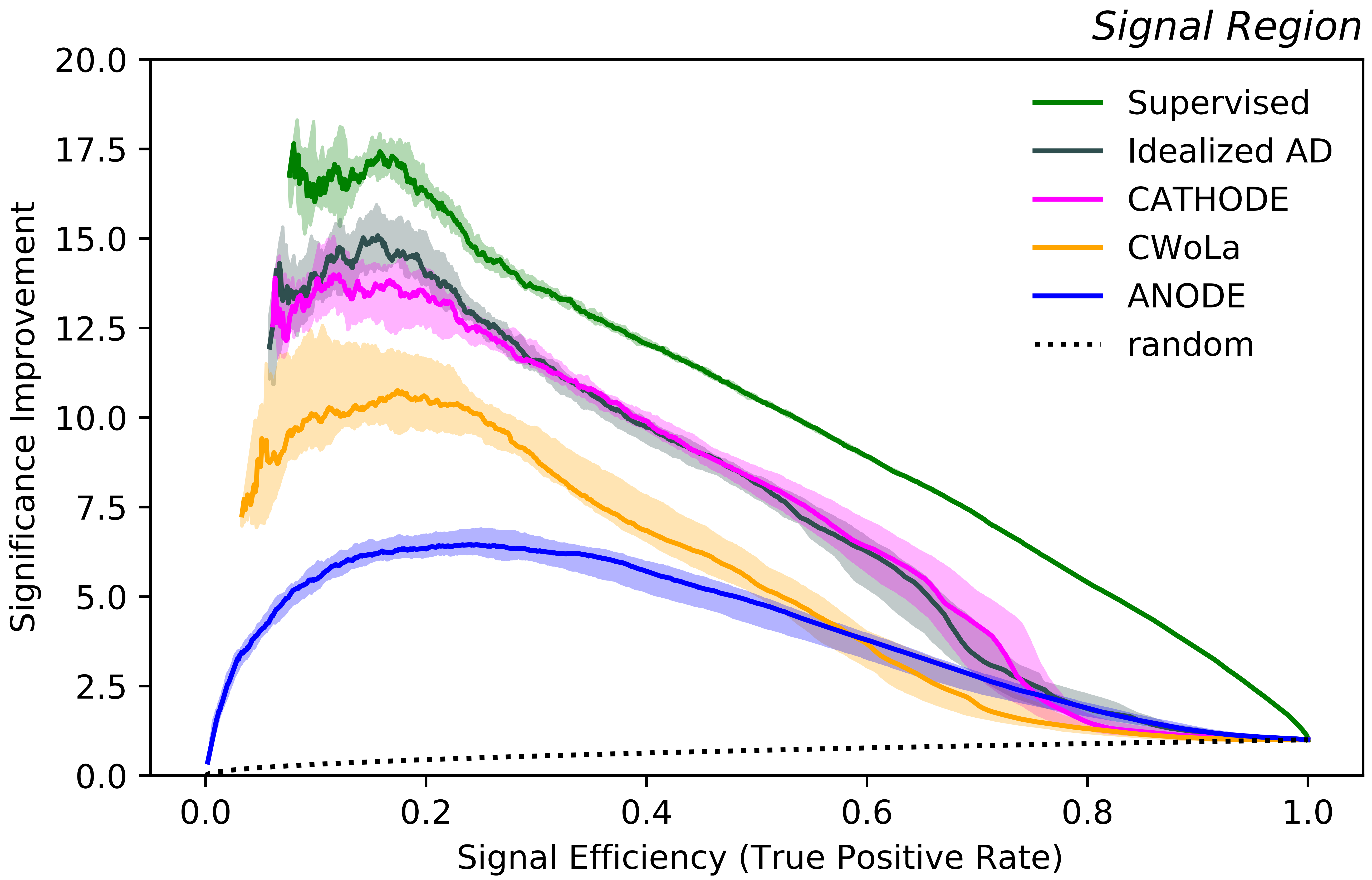
How do they compare?

How to quantify improvement?

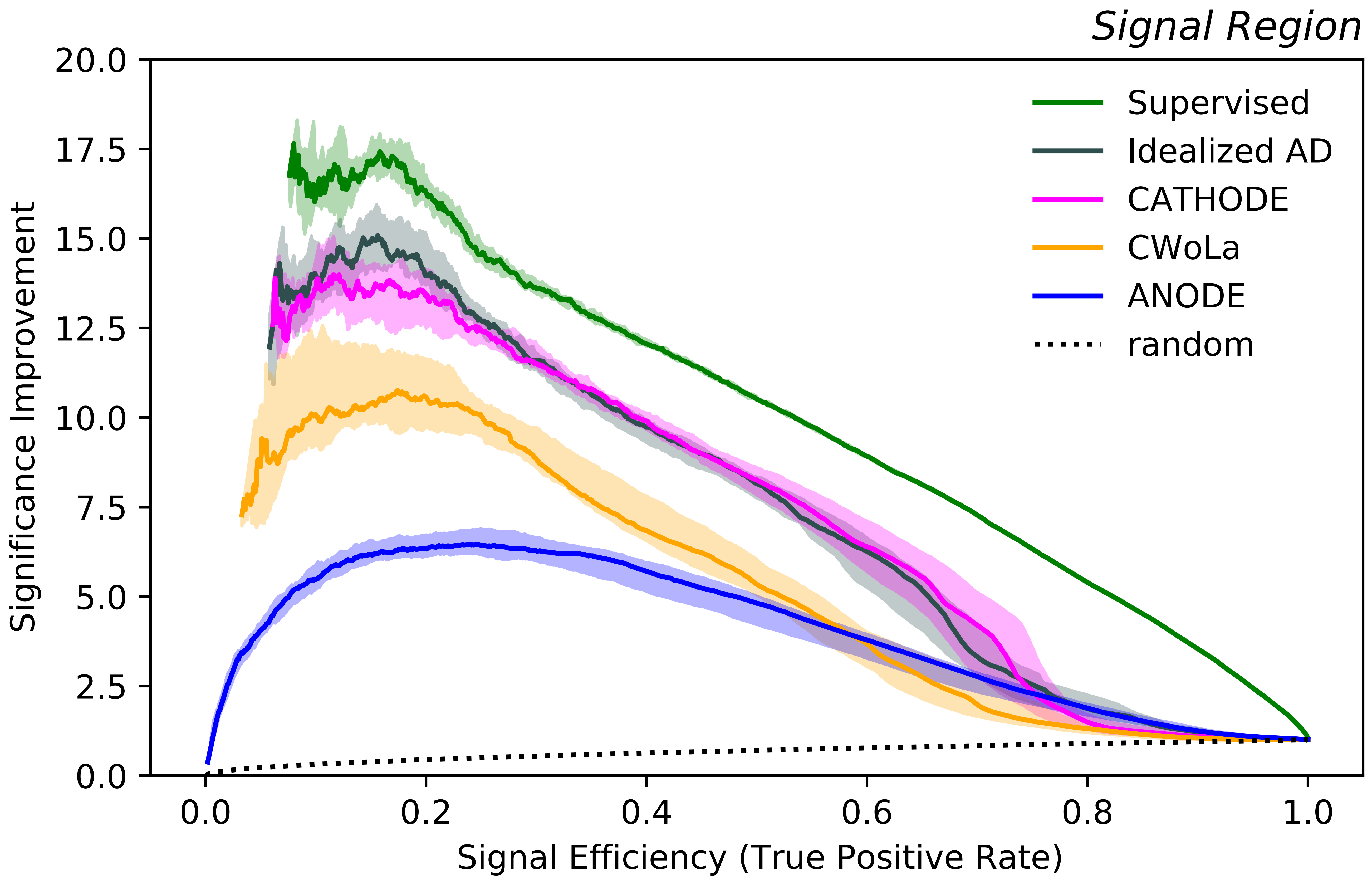


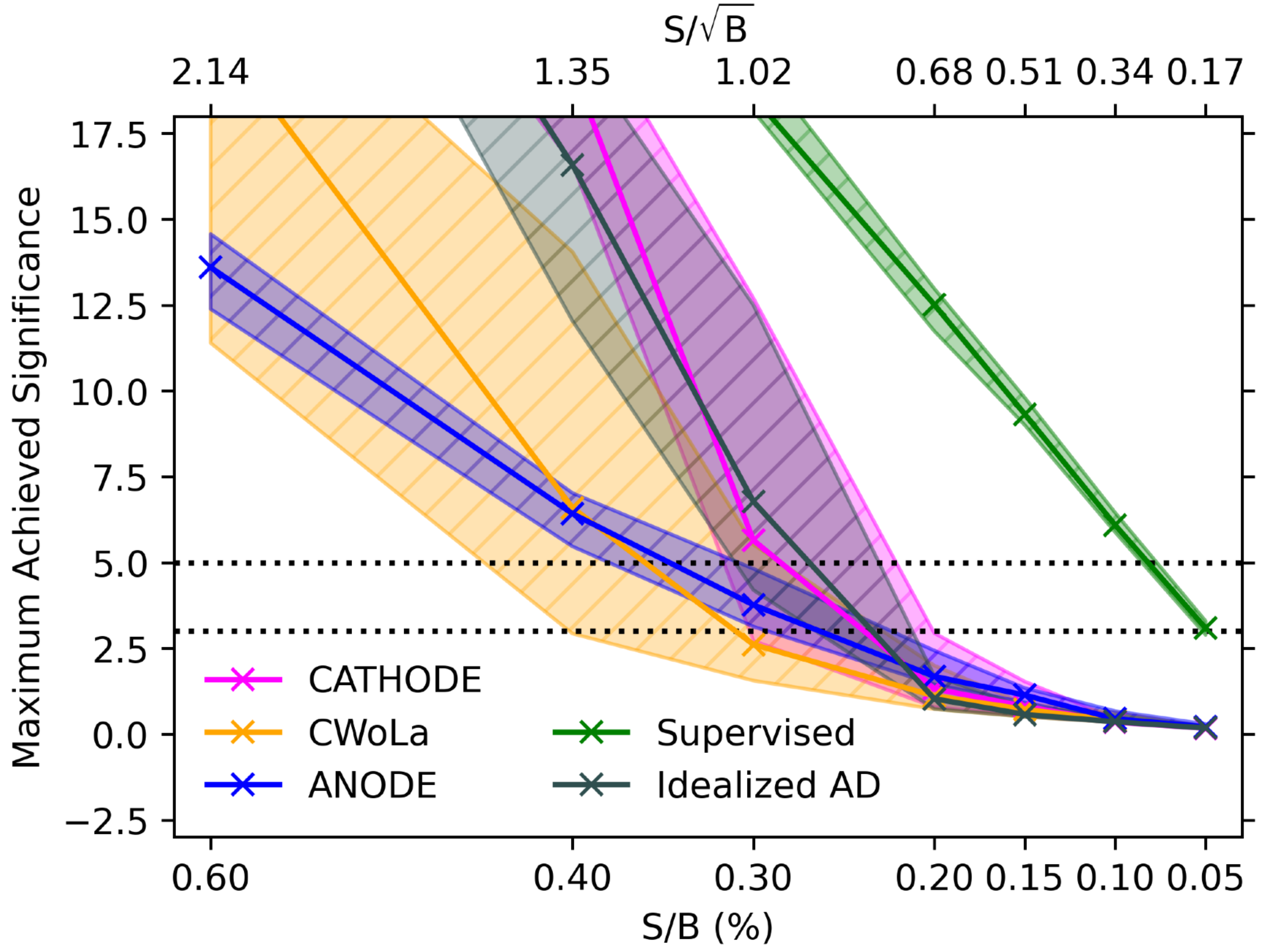
Statistical significance: $\frac{S}{\sqrt{B}}$ $\xrightarrow{\text{AD}}$ $\frac{S \cdot \epsilon_S}{\sqrt{B \cdot \epsilon_B}} = \frac{S}{\sqrt{B}} \cdot \boxed{\frac{\epsilon_S}{\sqrt{\epsilon_B}}}$ ← Improvement factor

Results — Comparison



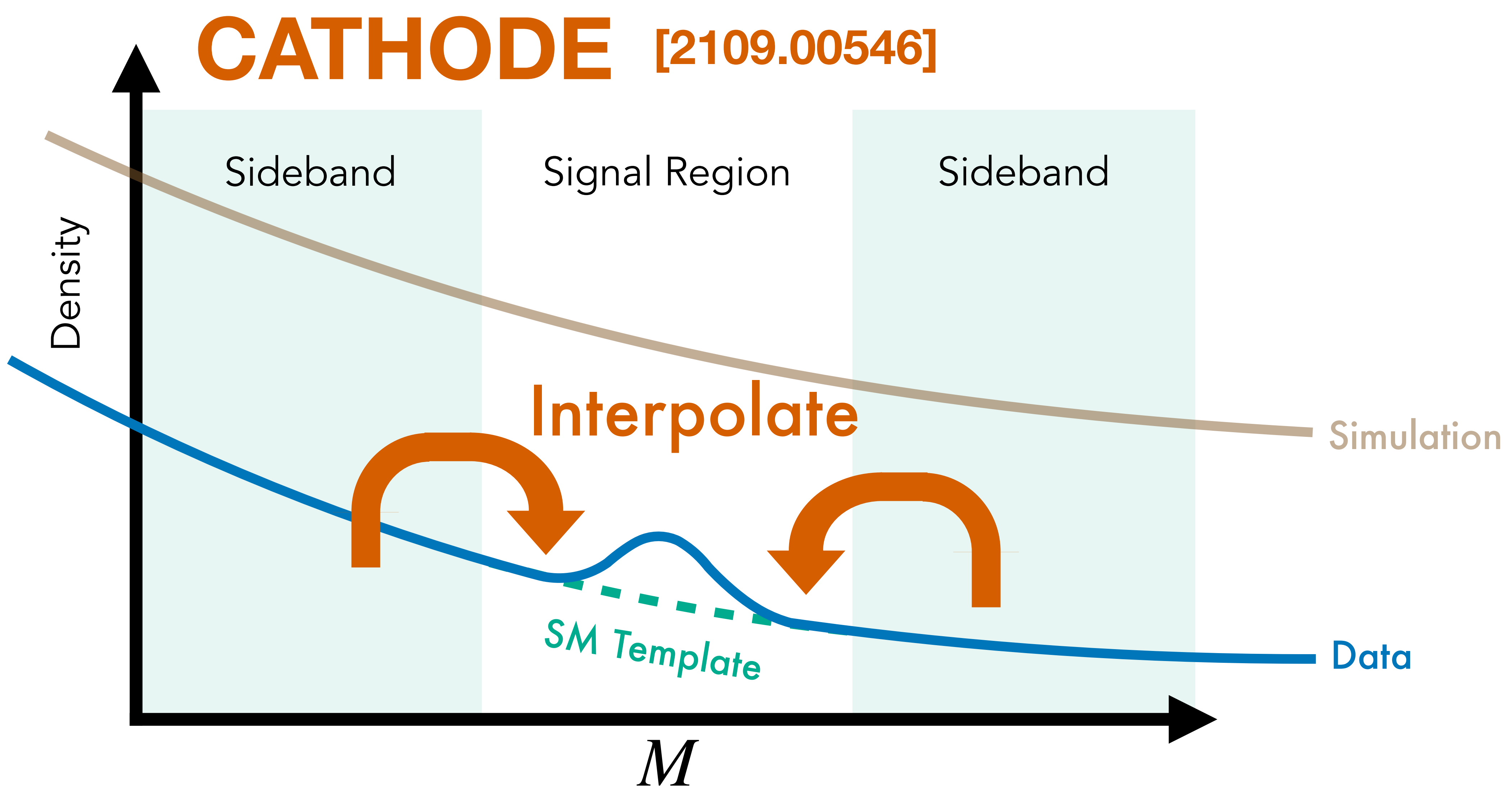
Results — Comparison



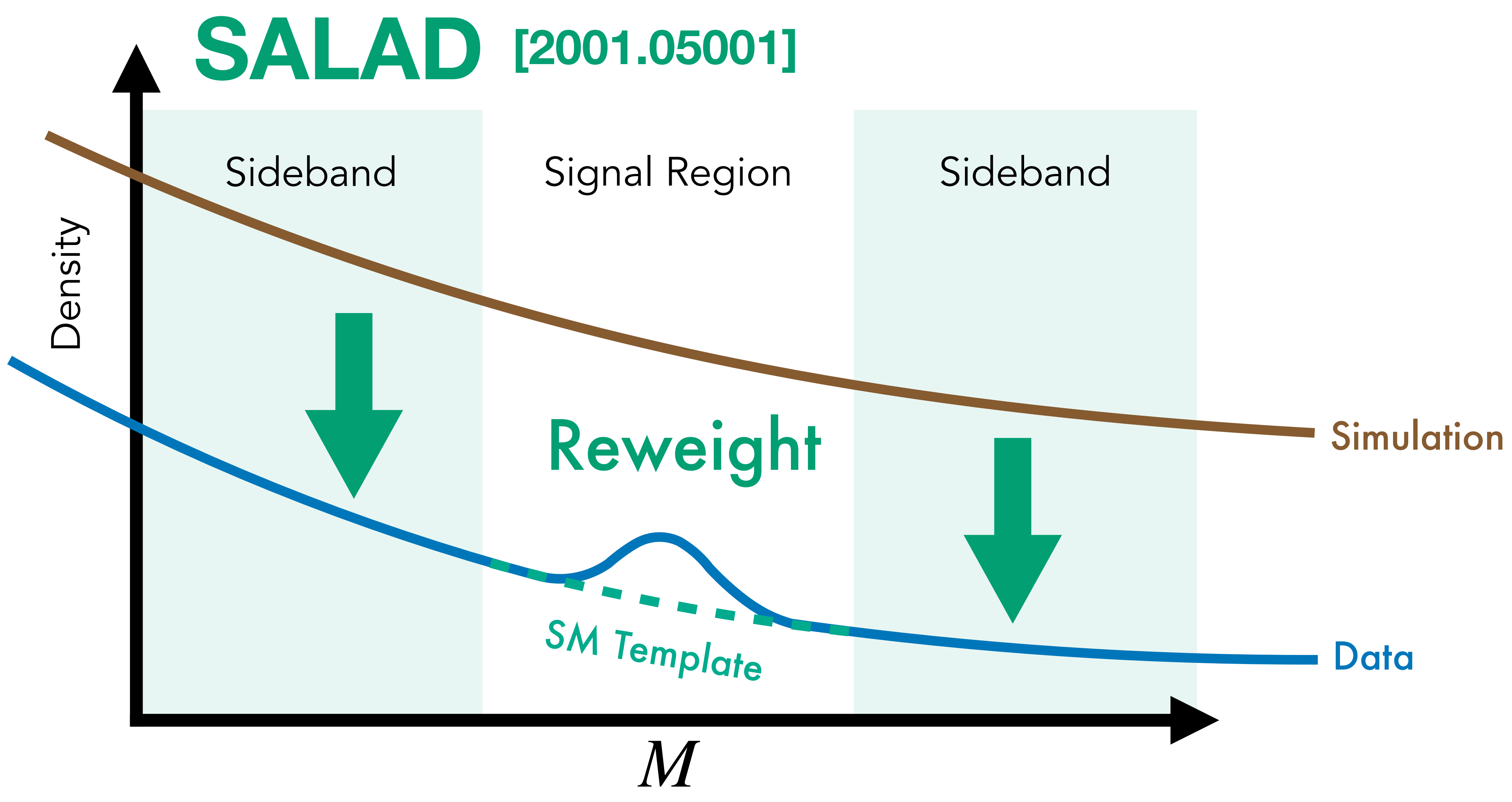


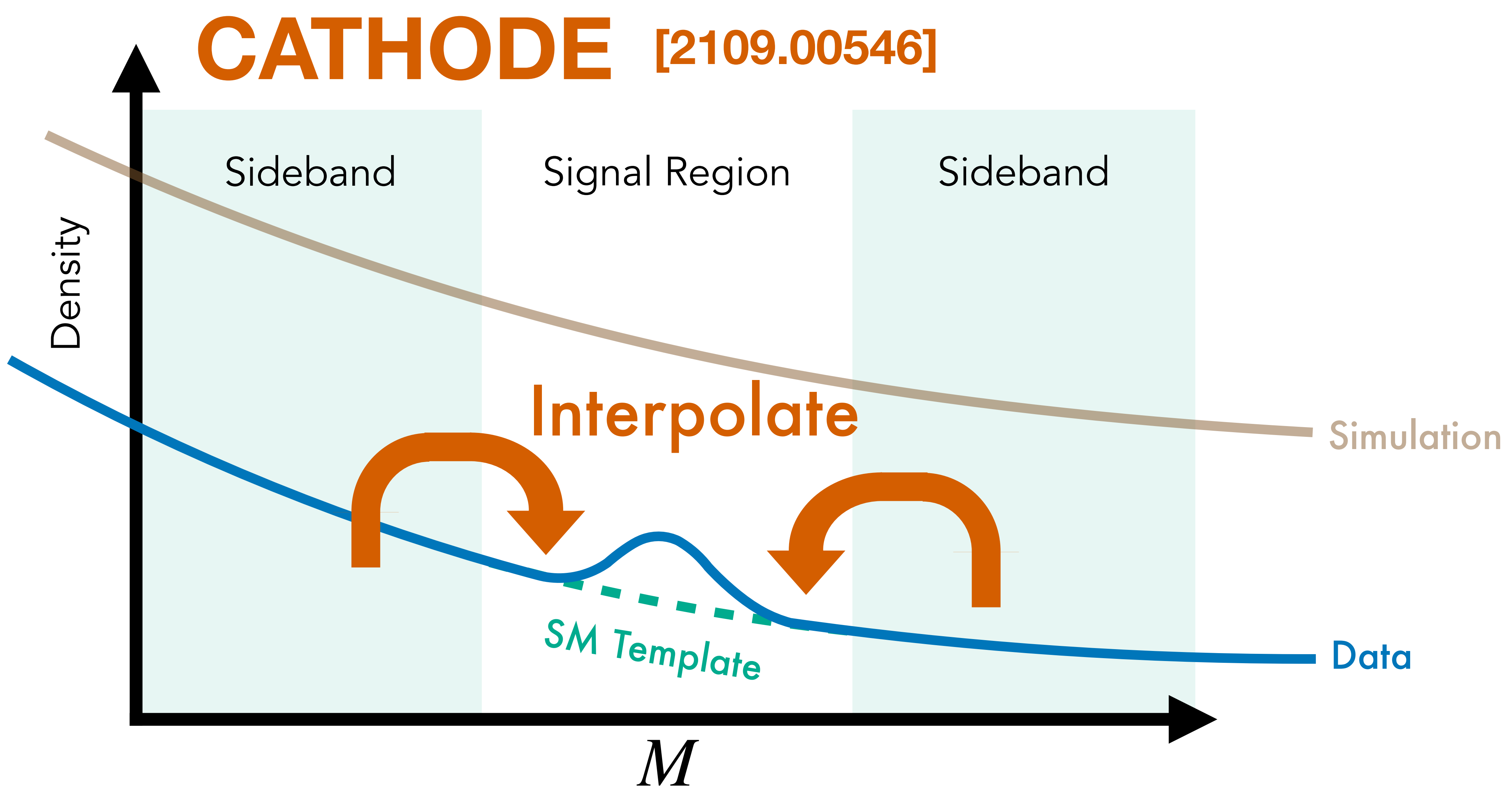
Are there other ways?

ML techniques to construct SM template



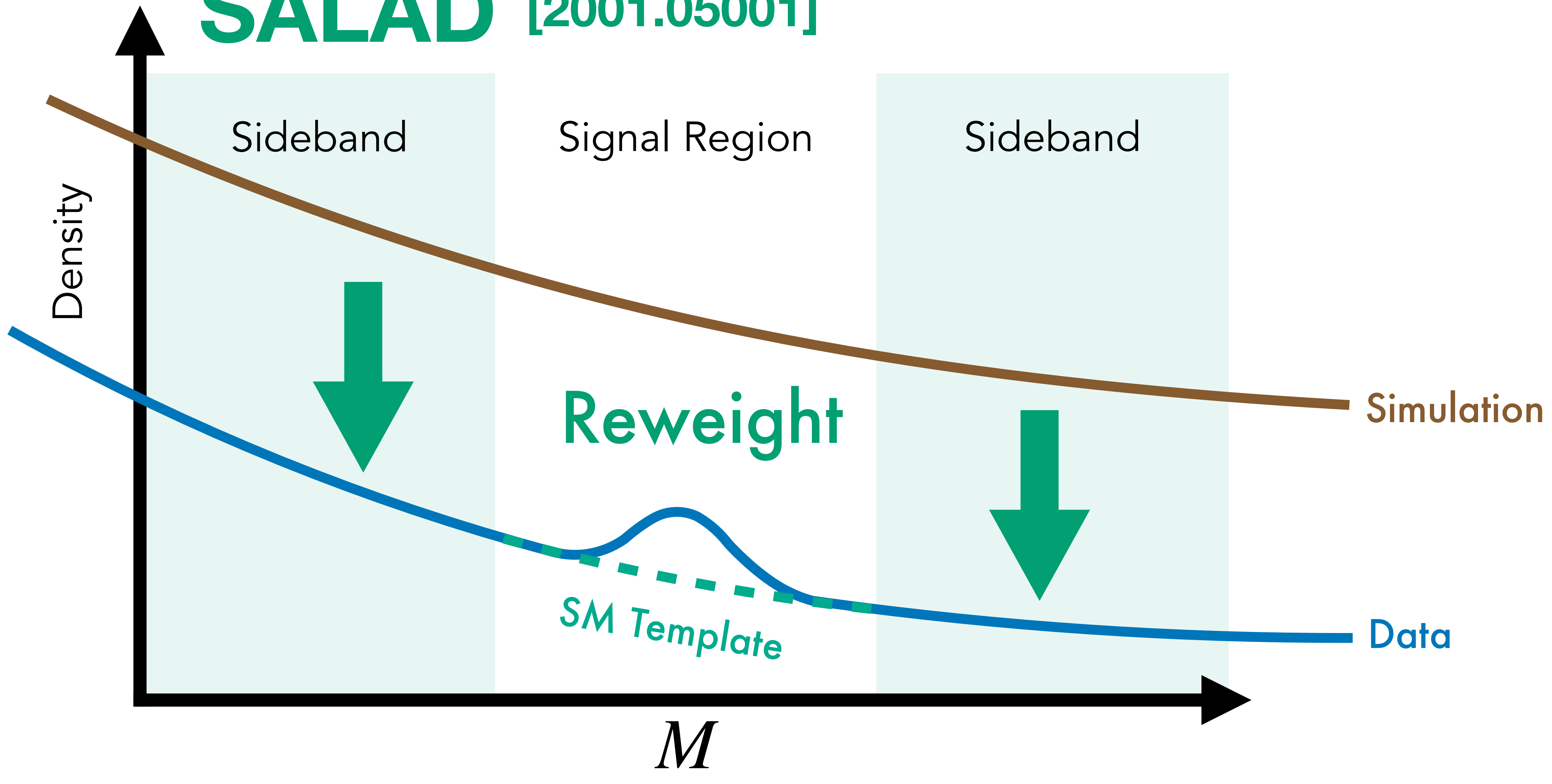
ML techniques to construct SM template



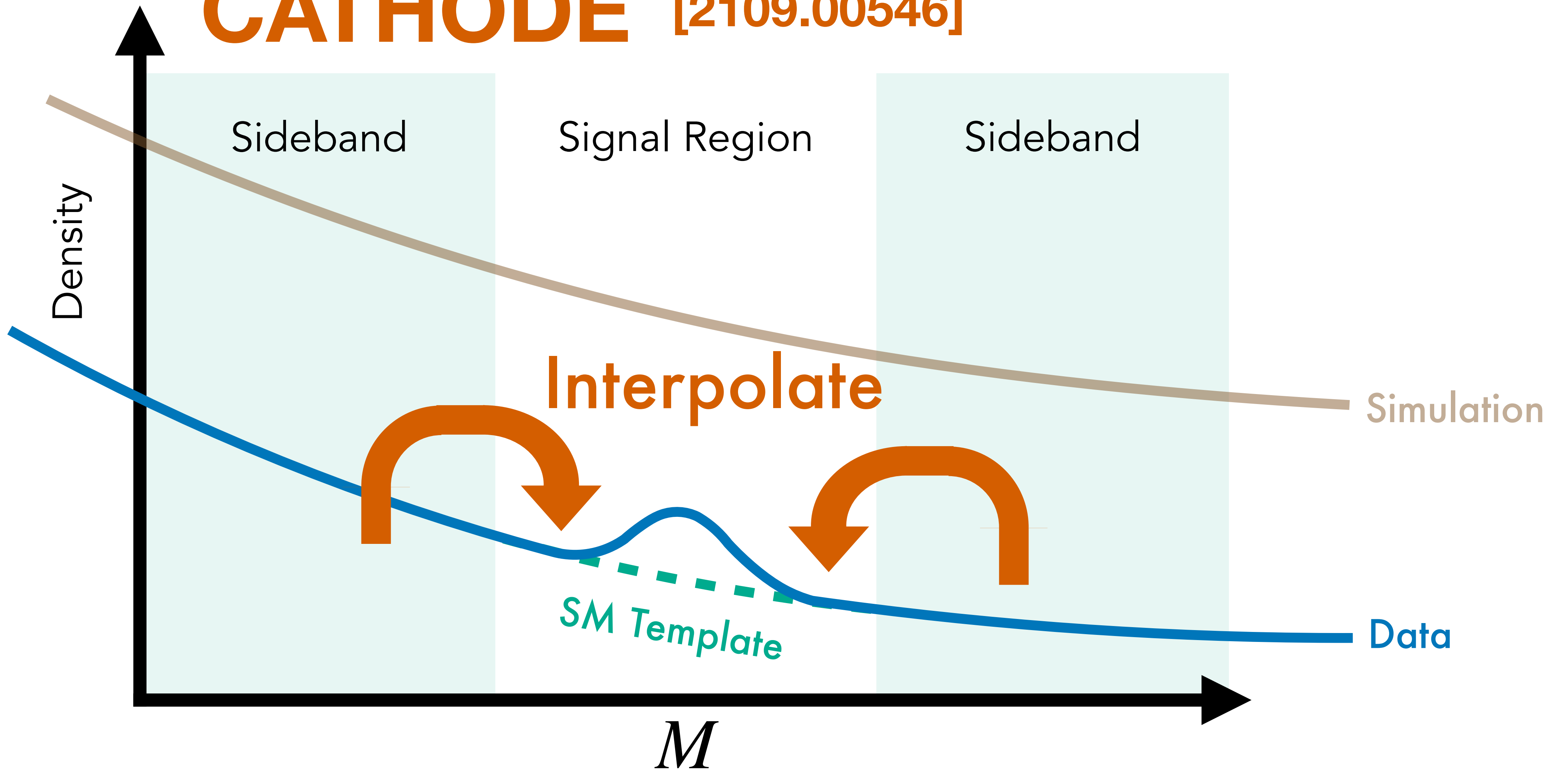


ML techniques to construct SM template

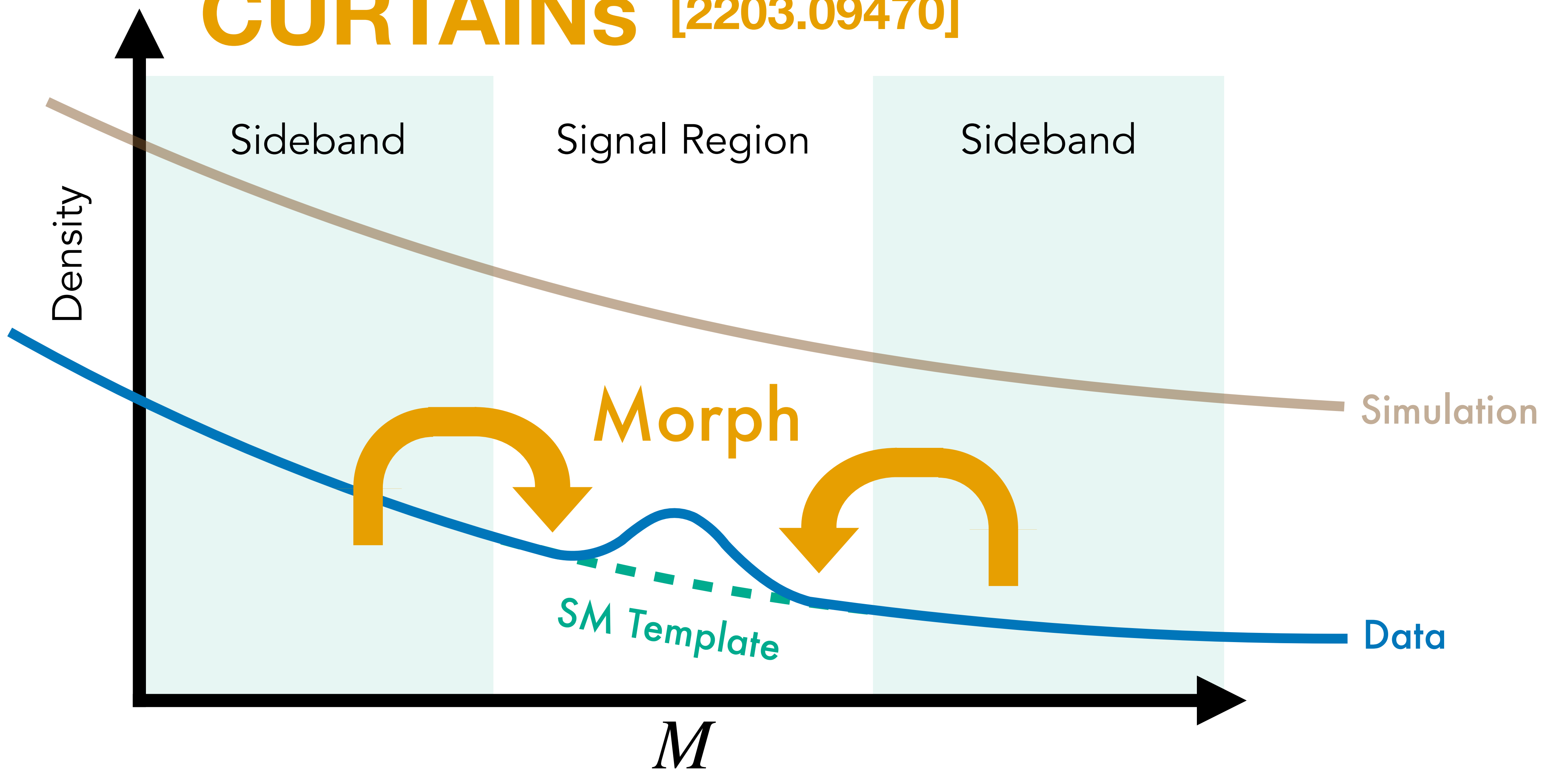
SALAD [2001.05001]



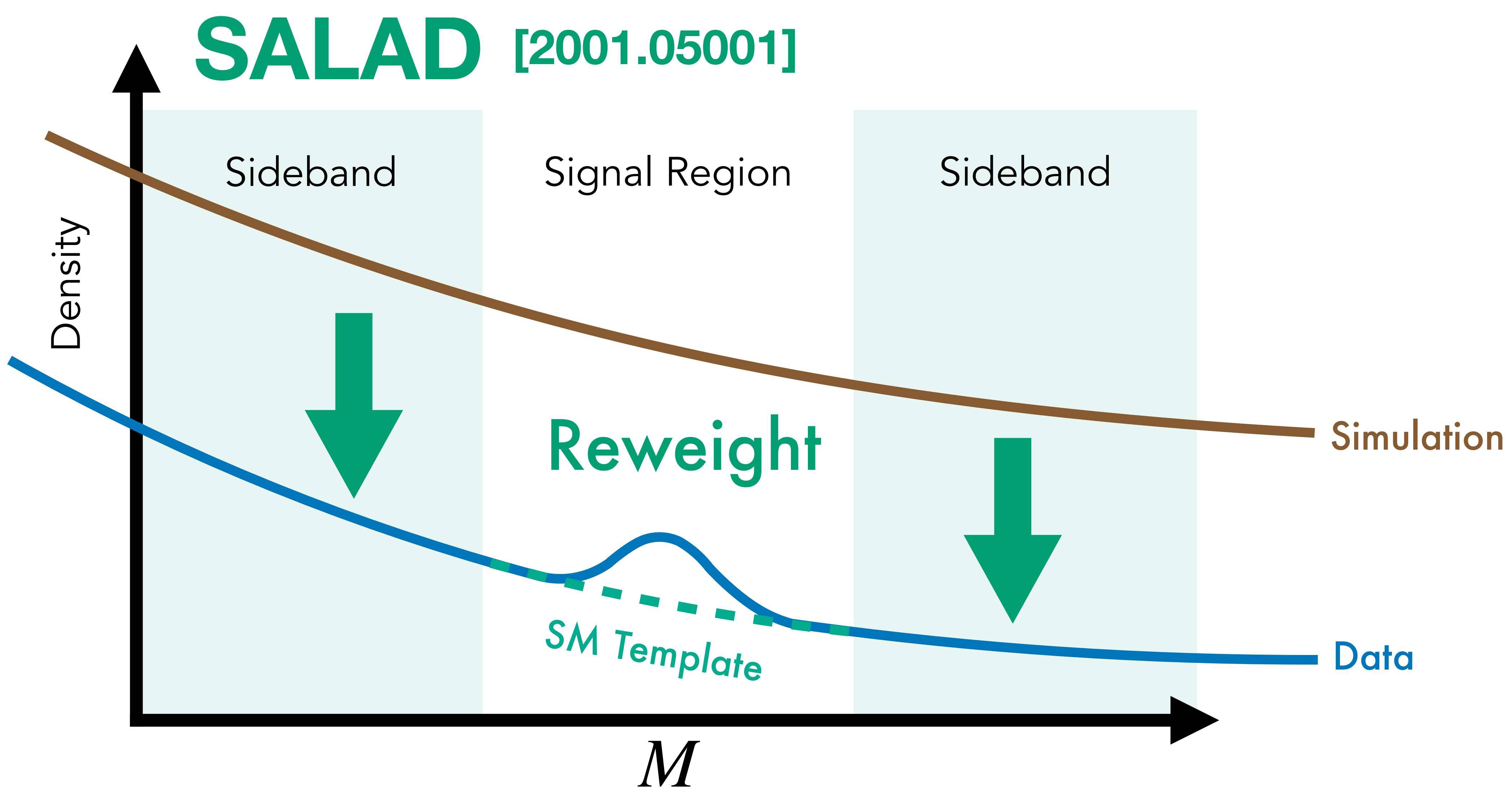
CATHODE [2109.00546]

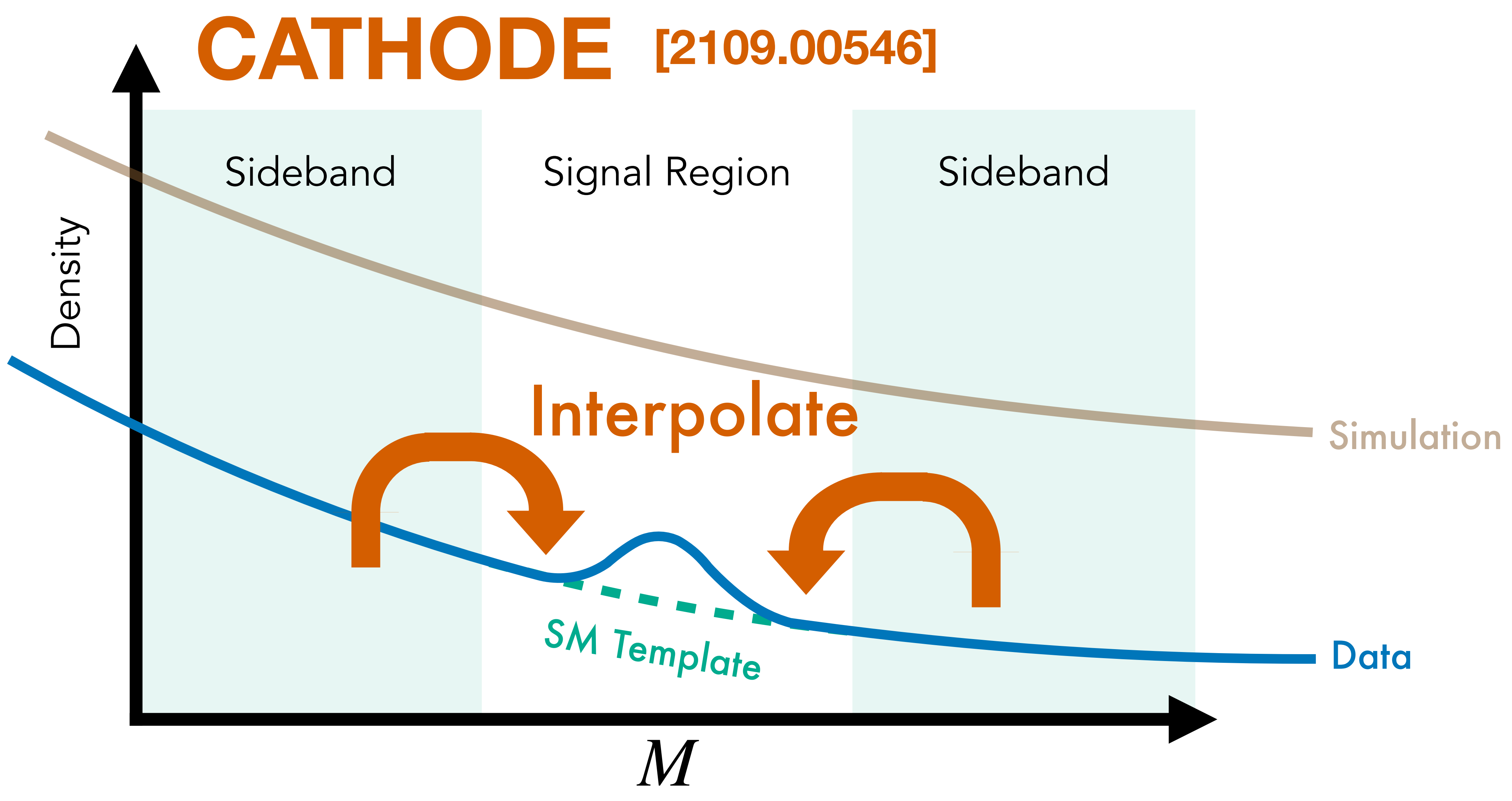


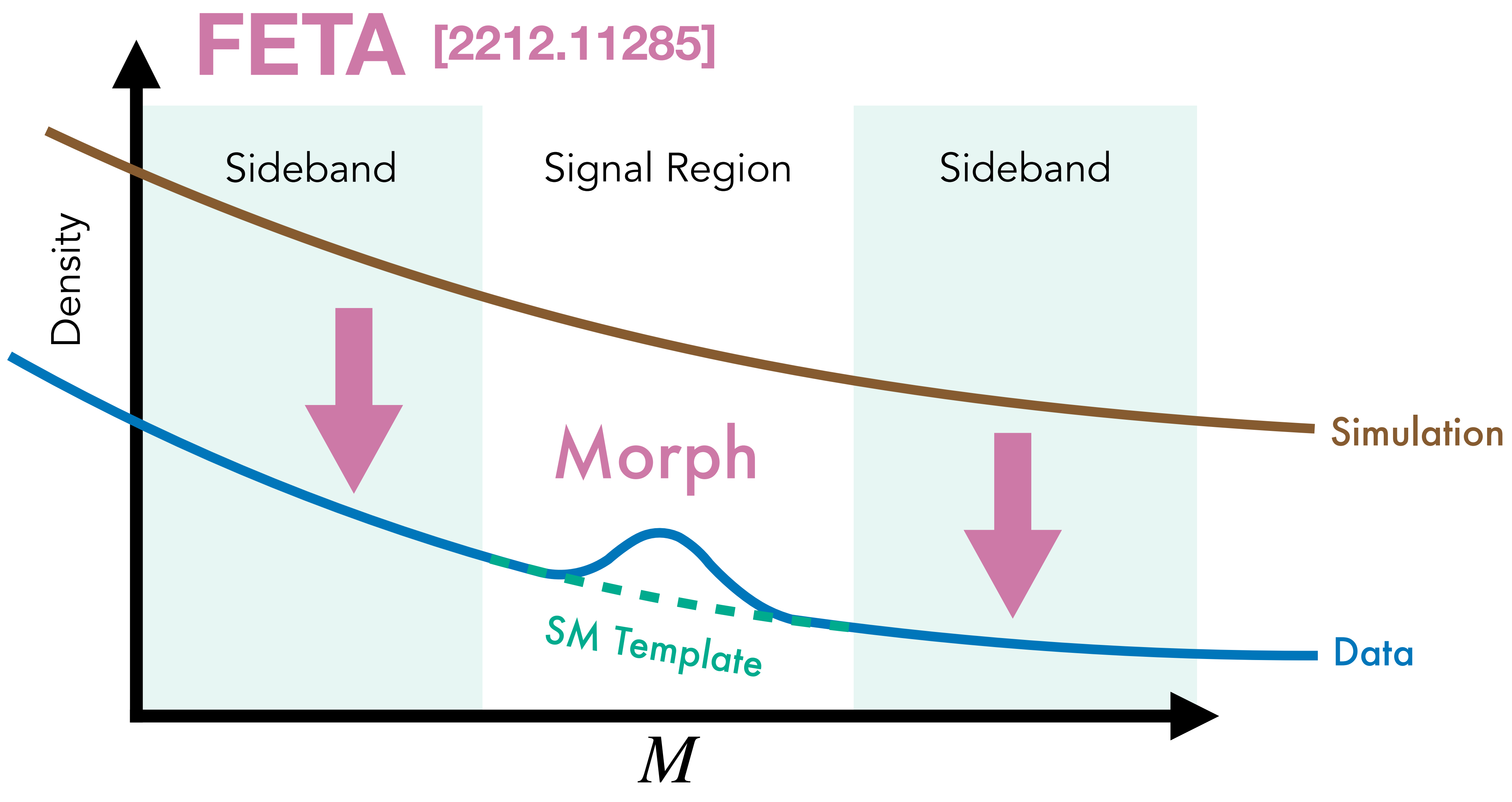
CURTAINs [2203.09470]

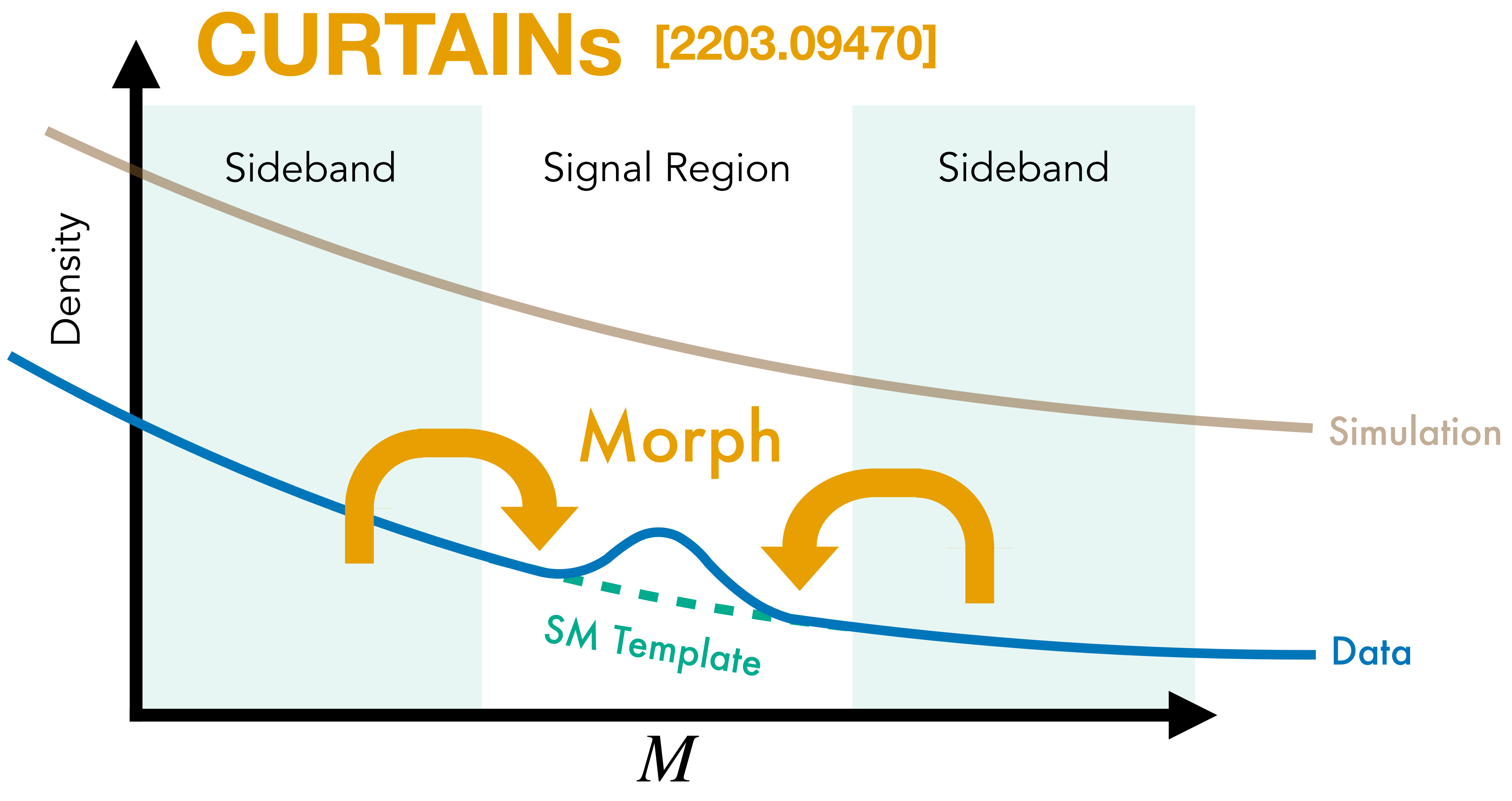


ML techniques to construct SM template

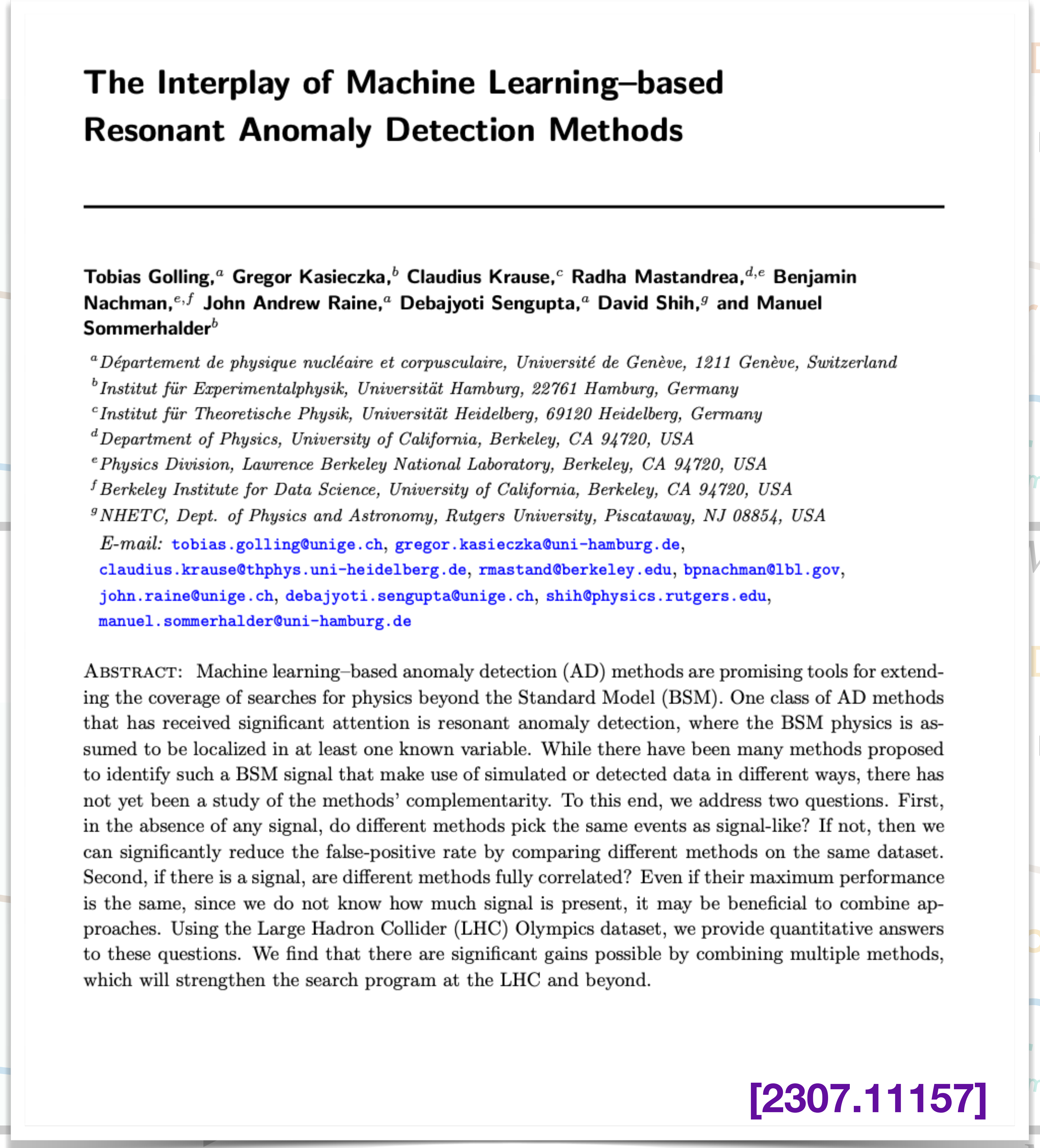








ML techniques to construct SM template



The Interplay of Machine Learning–based Resonant Anomaly Detection Methods

Tobias Golling,[a] Gregor Kasieczka,[b] Claudius Krause,[c] Radha Mastandrea,[d,e] Benjamin Nachman,[e,f] John Andrew Raine,[a] Debajyoti Sengupta,[a] David Shih,[g] and Manuel Sommerhalder[b]

[a] *Département de physique nucléaire et corpusculaire, Université de Genève, 1211 Genève, Switzerland*
[b] *Institut für Experimentalphysik, Universität Hamburg, 22761 Hamburg, Germany*
[c] *Institut für Theoretische Physik, Universität Heidelberg, 69120 Heidelberg, Germany*
[d] *Department of Physics, University of California, Berkeley, CA 94720, USA*
[e] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[f] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[g] *NHETC, Dept. of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA*

E-mail: tobias.golling@unige.ch, gregor.kasieczka@uni-hamburg.de, claudius.krause@thphys.uni-heidelberg.de, rmastand@berkeley.edu, bpnachman@lbl.gov, john.raine@unige.ch, debajyoti.sengupta@unige.ch, shih@physics.rutgers.edu, manuel.sommerhalder@uni-hamburg.de

Abstract: Machine learning–based anomaly detection (AD) methods are promising tools for extending the coverage of searches for physics beyond the Standard Model (BSM). One class of AD methods that has received significant attention is resonant anomaly detection, where the BSM physics is assumed to be localized in at least one known variable. While there have been many methods proposed to identify such a BSM signal that make use of simulated or detected data in different ways, there has not yet been a study of the methods' complementarity. To this end, we address two questions. First, in the absence of any signal, do different methods pick the same events as signal-like? If not, then we can significantly reduce the false-positive rate by comparing different methods on the same dataset. Second, if there is a signal, are different methods fully correlated? Even if their maximum performance is the same, since we do not know how much signal is present, it may be beneficial to combine approaches. Using the Large Hadron Collider (LHC) Olympics dataset, we provide quantitative answers to these questions. We find that there are significant gains possible by combining multiple methods, which will strengthen the search program at the LHC and beyond.

[2307.11157]

Can we do even better?

Example IV

Residual ANODE (R-ANODE)

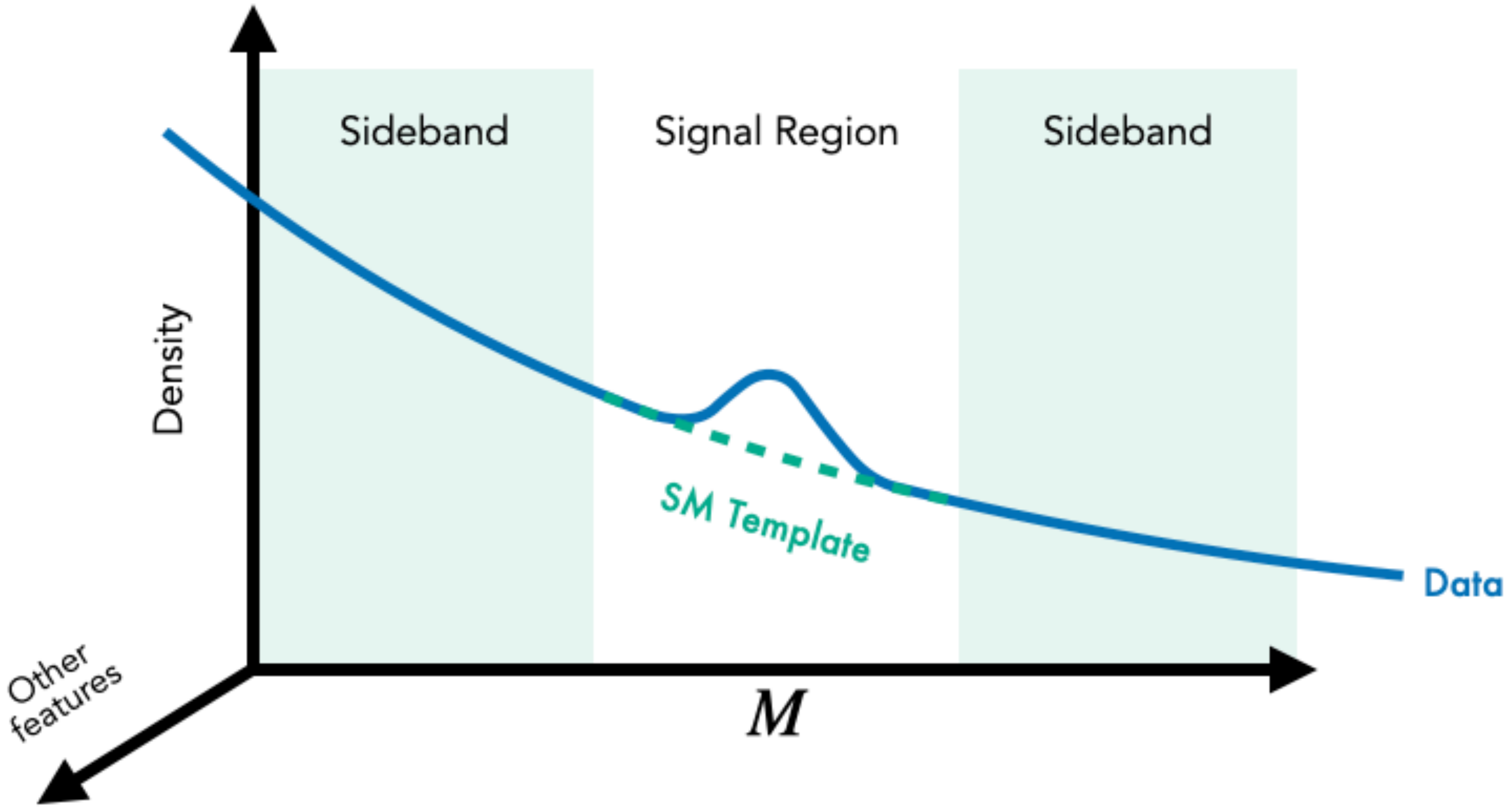
Das, Kasieczka, Shih [2312.11629]

R-ANODE

The ANODE method

$p_{\omega_0}(x|m) \simeq p_{\text{bg}}(x|m)$ Trained in $m \in \text{SB}$

$p_{\omega_1}(x|m) \simeq p_{\text{data}}(x|m)$ Trained in $m \in \text{SR}$



R-ANODE

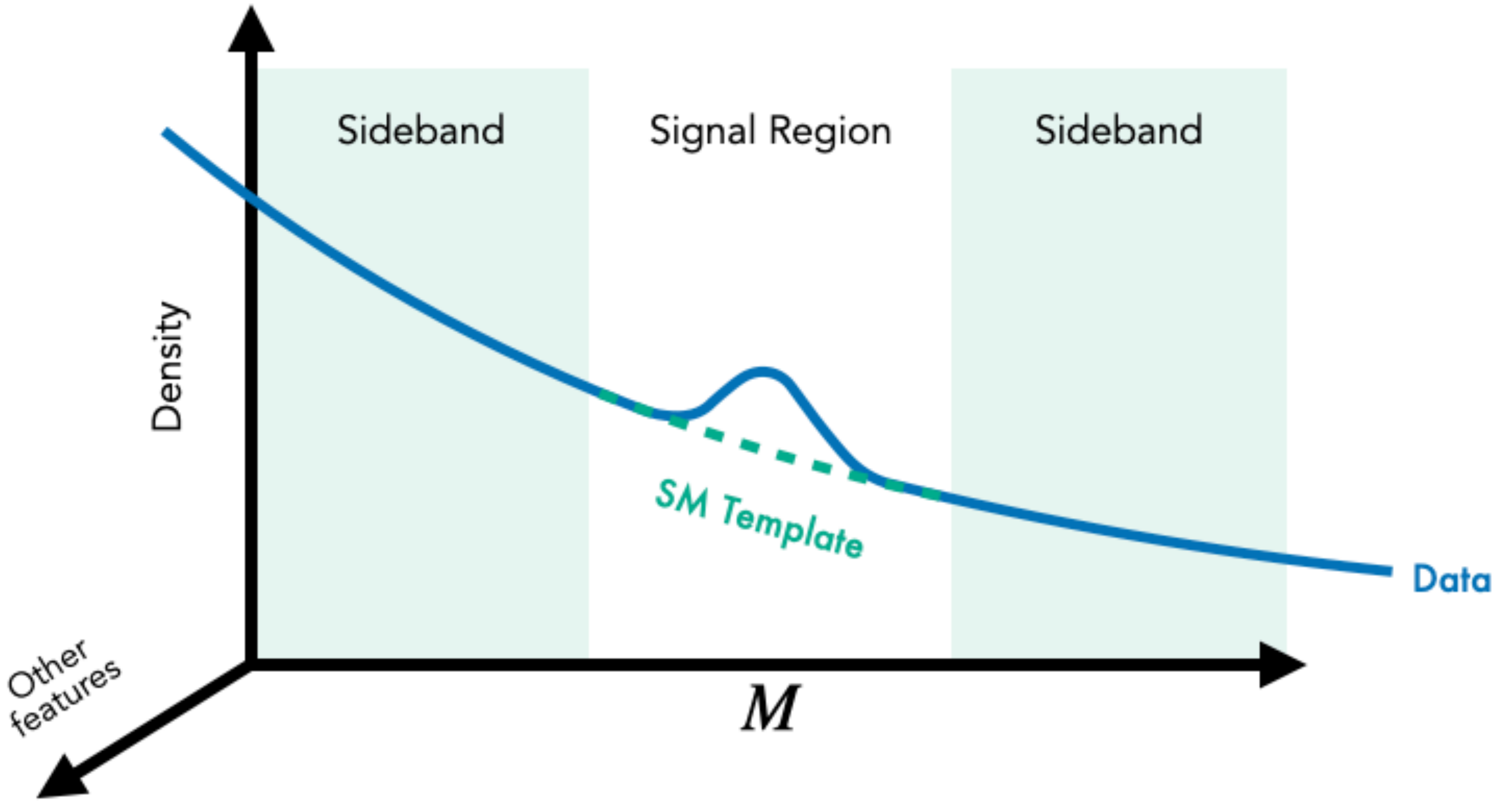
The ANODE method

$$p_{\omega_0}(x\,|\,m) \simeq p_{\text{bg}}(x\,|\,m)$$ Trained in $m \in \text{SB}$

~~$p_{\omega_1}(x\,|\,m) \simeq p_{\text{data}}(x\,|\,m)$ Trained in $m \in \text{SR}$~~

The R-ANODE method

$$p_{\text{data}}(x\,|\,\text{SR}) = w\,p_{\text{sig}}(x\,|\,\text{SR}) + (1-w)\,p_{\text{bg}}(x\,|\,\text{SR})$$



R-ANODE

The ANODE method

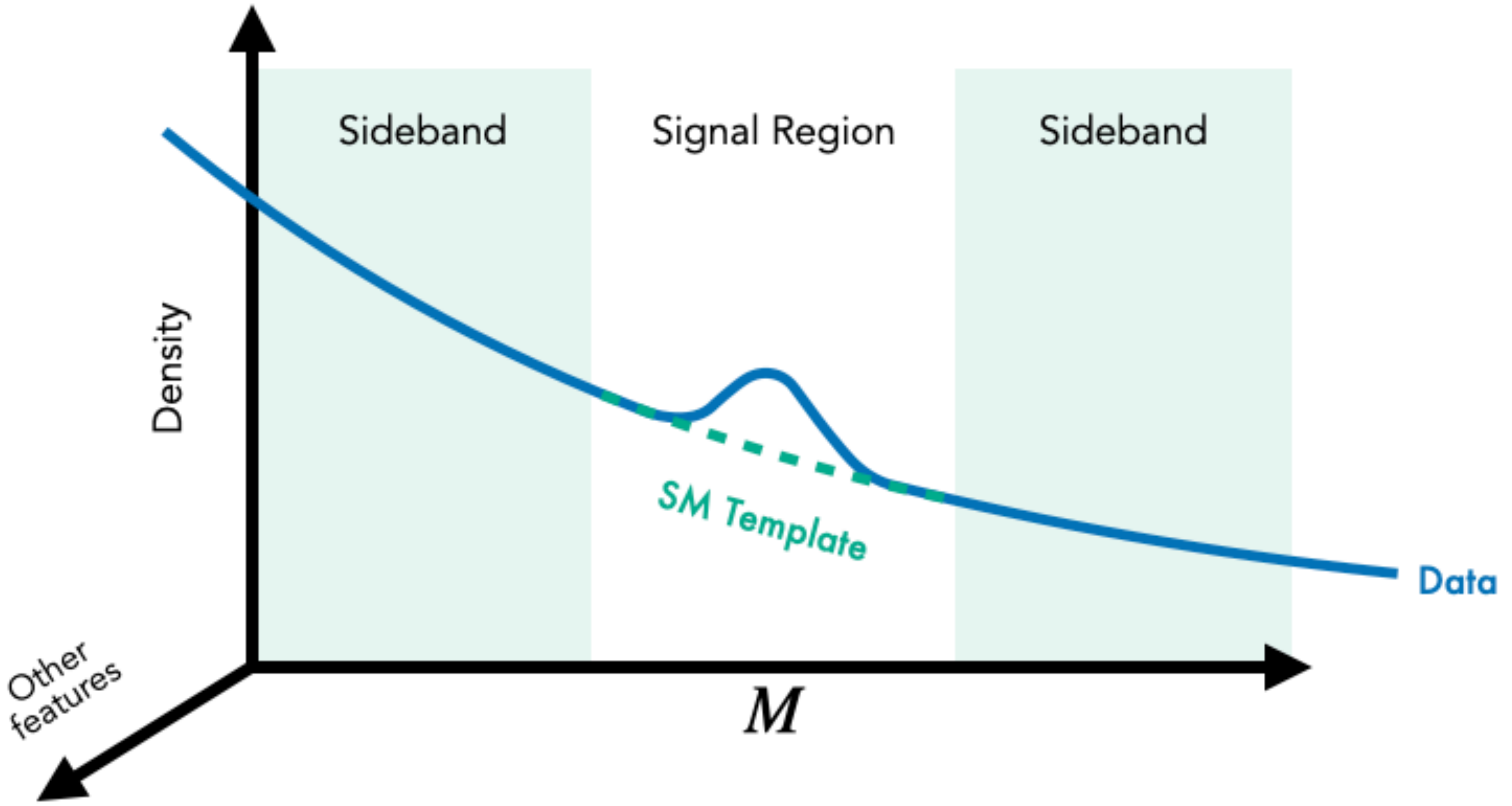
$p_{\omega_0}(x|m) \simeq p_{\text{bg}}(x|m)$ Trained in $m \in \text{SB}$

~~$p_{\omega_1}(x|m) \simeq p_{\text{data}}(x|m)$ Trained in $m \in \text{SR}$~~

The R-ANODE method

$$p_{\text{data}}(x|\text{SR}) = w\,p_{\text{sig}}(x|\text{SR}) + (1-w)\,p_{\text{bg}}(x|\text{SR})$$

NF
$p_{\omega_0}(x|m)$
Trained in $m \in \text{SB}$



R-ANODE

The ANODE method

$p_{\omega_0}(x\,|\,m) \simeq p_{\mathrm{bg}}(x\,|\,m)$ Trained in $m \in \mathrm{SB}$

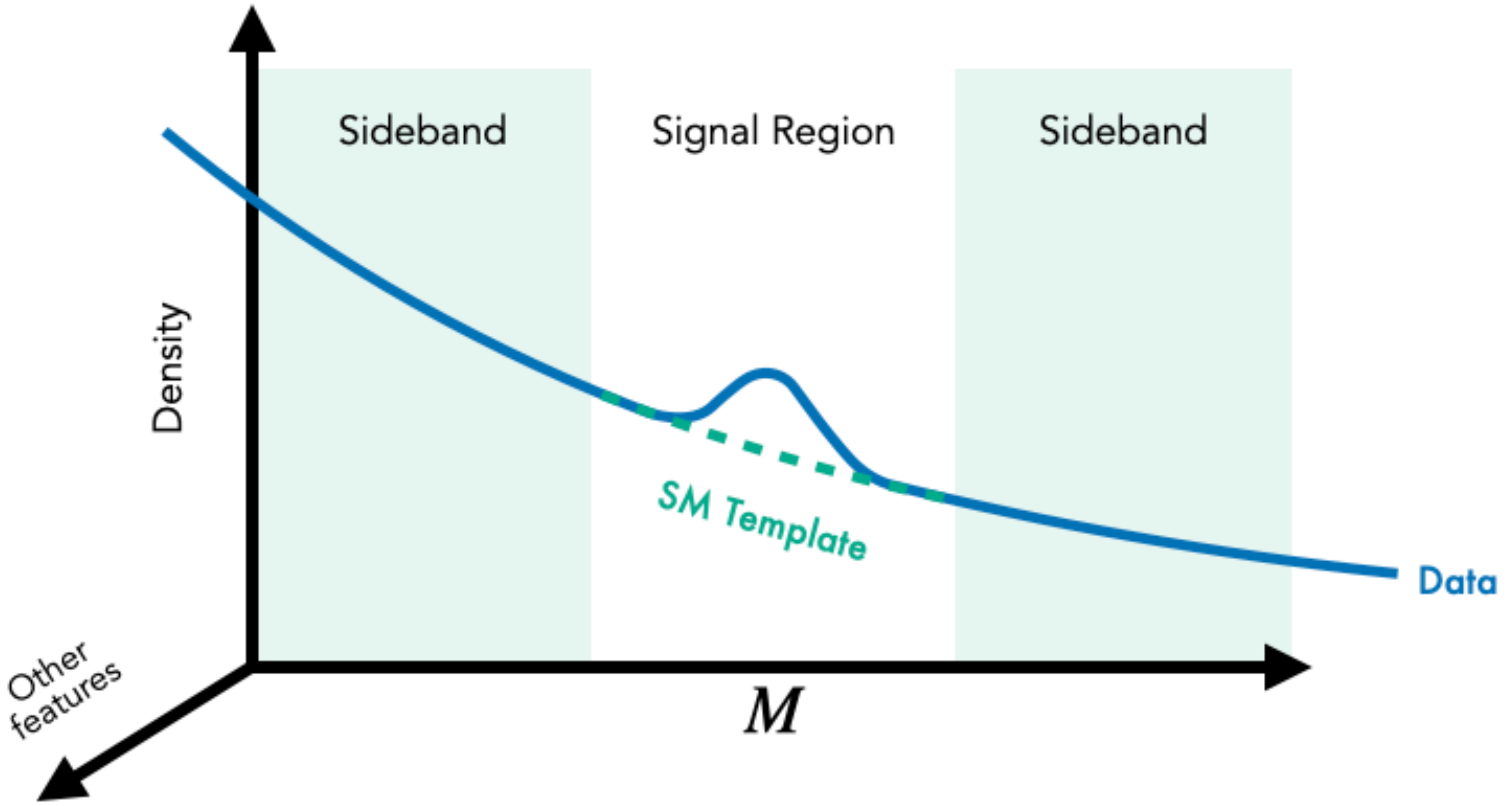
~~$p_{\omega_1}(x\,|\,m) \simeq p_{\mathrm{data}}(x\,|\,m)$ Trained in $m \in \mathrm{SR}$~~

The R-ANODE method

$$p_{\mathrm{data}}(x\,|\,\mathrm{SR}) = w\,p_{\mathrm{sig}}(x\,|\,\mathrm{SR}) + (1-w)\,p_{\mathrm{bg}}(x\,|\,\mathrm{SR})$$

NF
$p_{\omega_s}(x\,|\,m)$
Trained in $m \in \mathrm{SR}$

NF
$p_{\omega_0}(x\,|\,m)$
Trained in $m \in \mathrm{SB}$



R-ANODE

The ANODE method

$p_{\omega_0}(x|m) \simeq p_{\text{bg}}(x|m)$ Trained in $m \in \text{SB}$

~~$p_{\omega_1}(x|m) \simeq p_{\text{data}}(x|m)$ Trained in $m \in \text{SR}$~~

The R-ANODE method

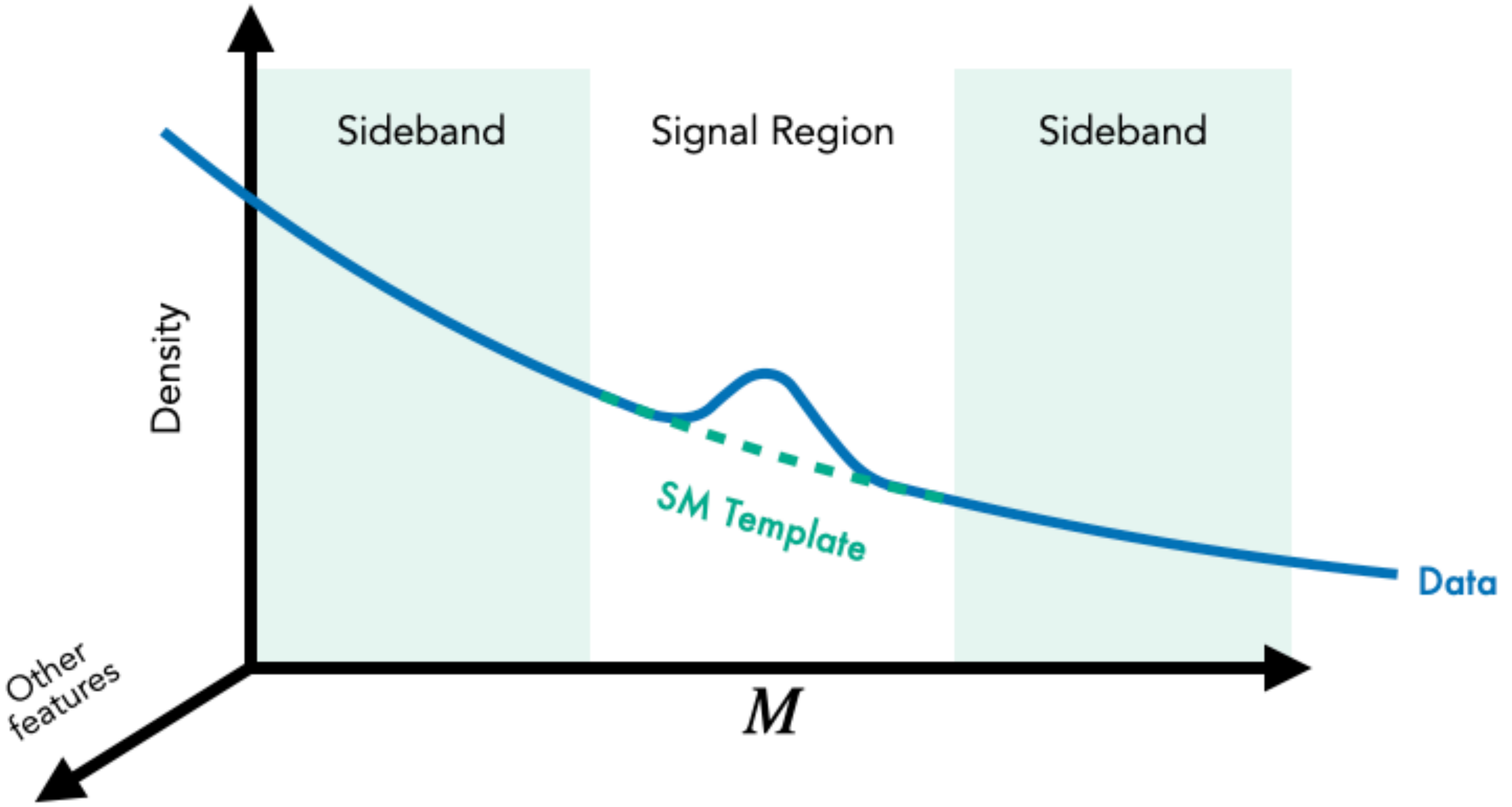
$$p_{\text{data}}(x|\text{SR}) = w\, p_{\text{sig}}(x|\text{SR}) + (1-w)\, p_{\text{bg}}(x|\text{SR})$$

NF $p_{\omega_s}(x|m)$ Trained in $m \in \text{SR}$

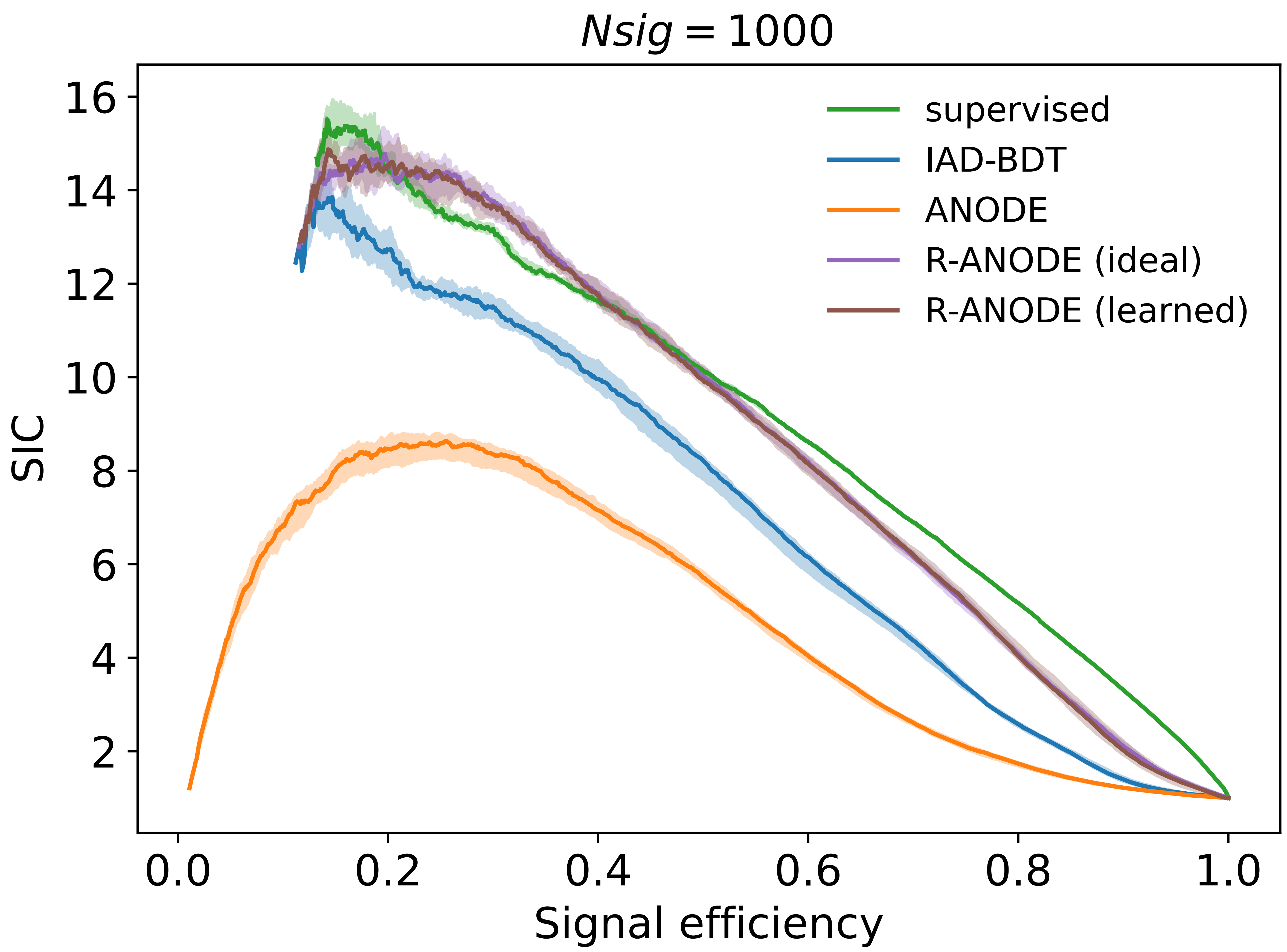
NF $p_{\omega_0}(x|m)$ Trained in $m \in \text{SB}$

R-ANODE Likelihood estimate

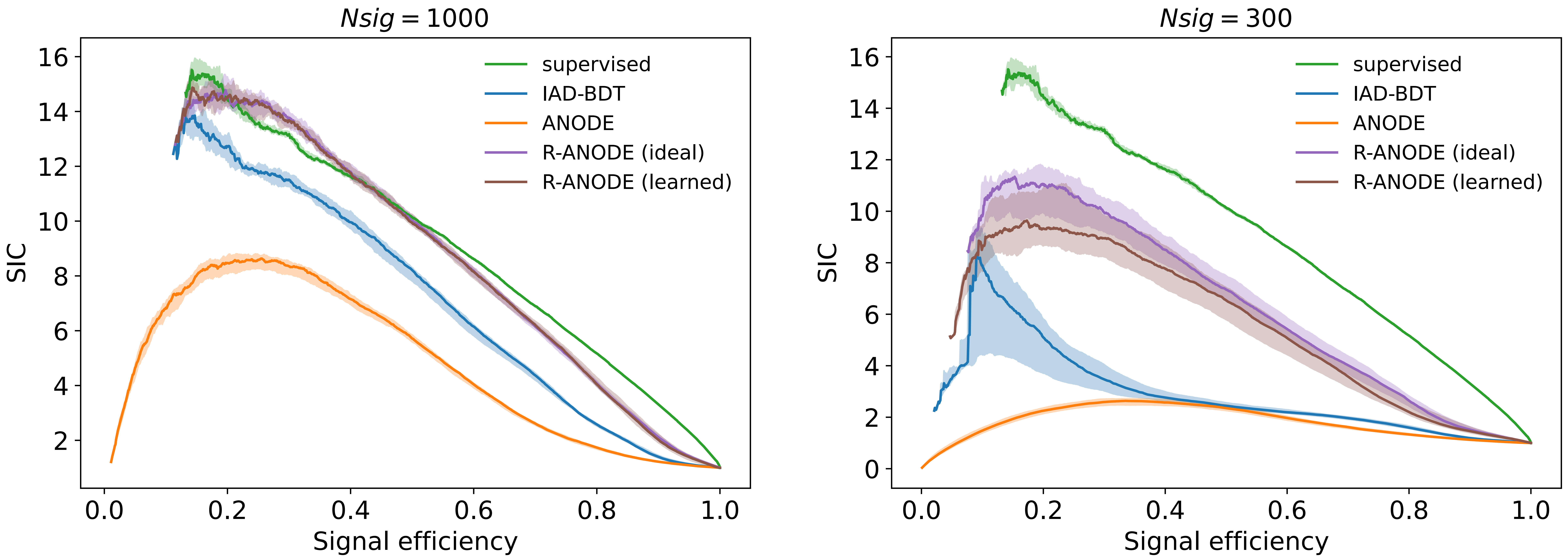
$$R_{\text{R-ANODE}} = \frac{p_{\omega_s}(x|\text{SR})}{p_{\omega_0}(x|\text{SR})} \simeq \frac{p_{\text{sig}}(x|\text{SR})}{p_{\text{bg}}(x|\text{SR})}$$



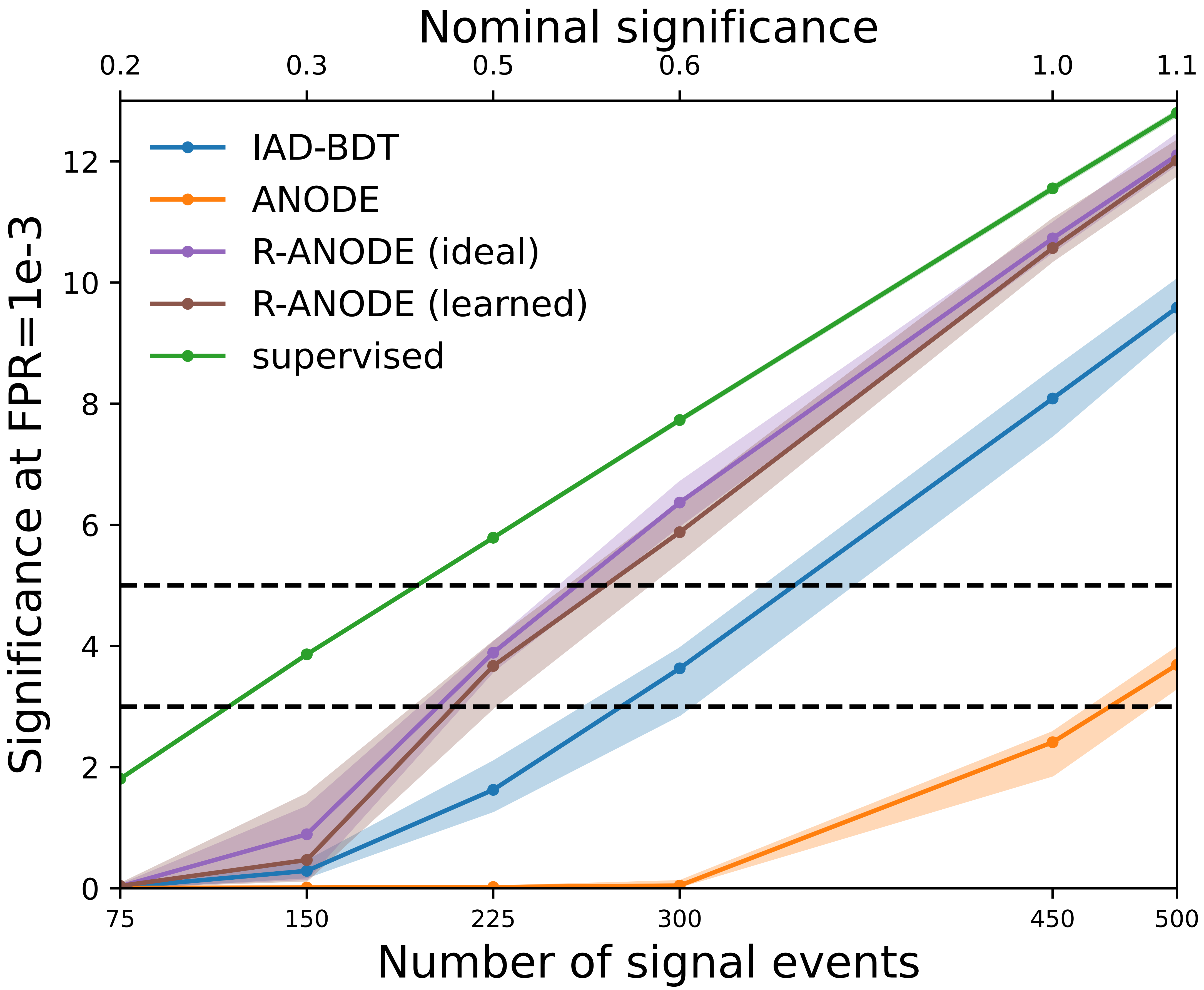
Results — R-ANODE

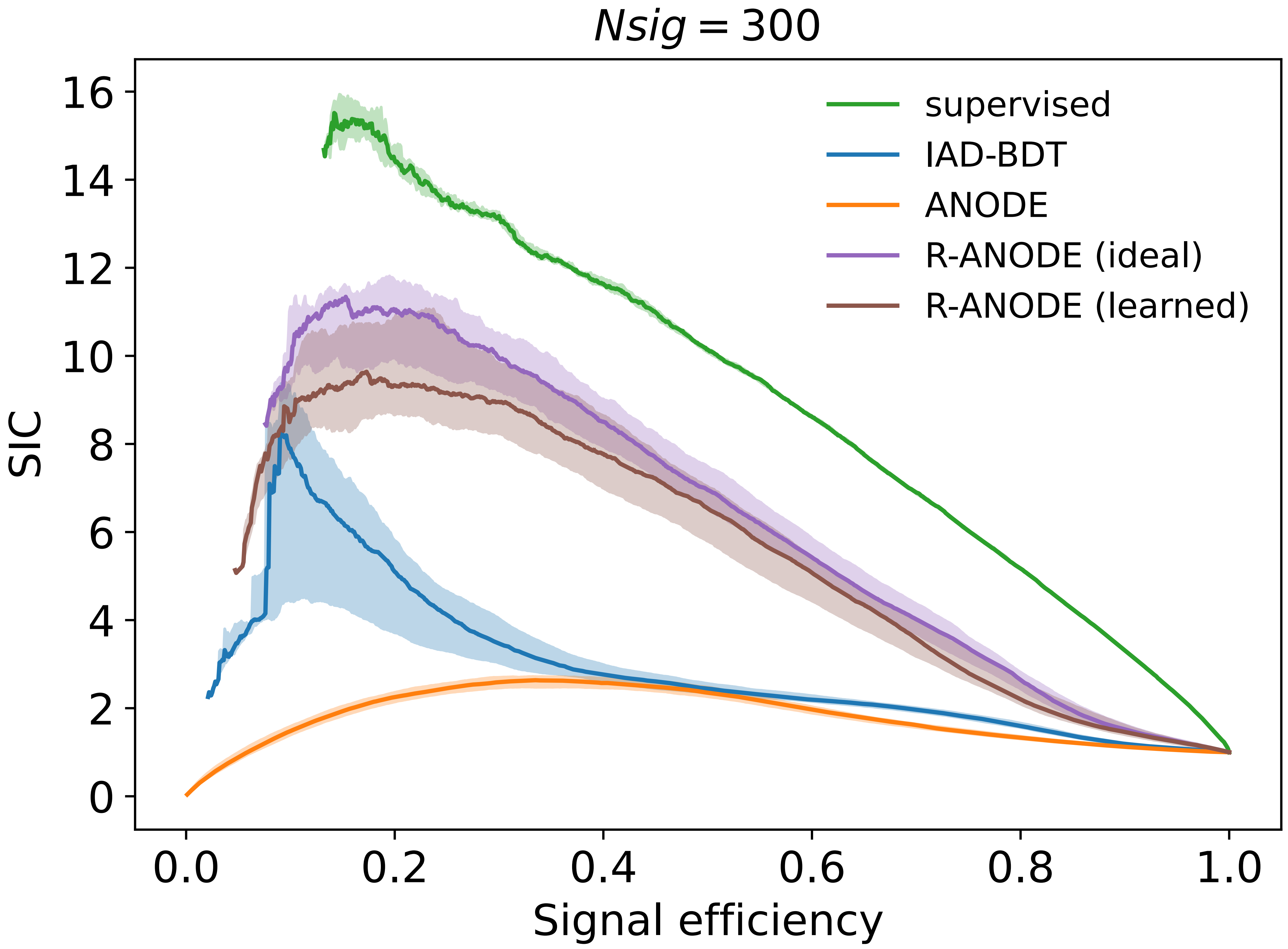


Results — R-ANODE



Results — R-ANODE

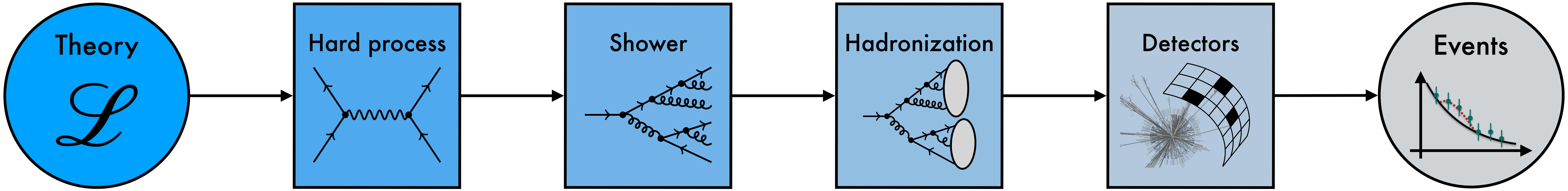




Summary and Outlook

Take-home messages

- ML beneficial in **every step** of the **simulation** and **analysis chain**
- We find both **proof-of-concepts** as well as established use cases (→ **AD, MadNIS,…**)
- Interesting **interplay** between **physics** and **ML**

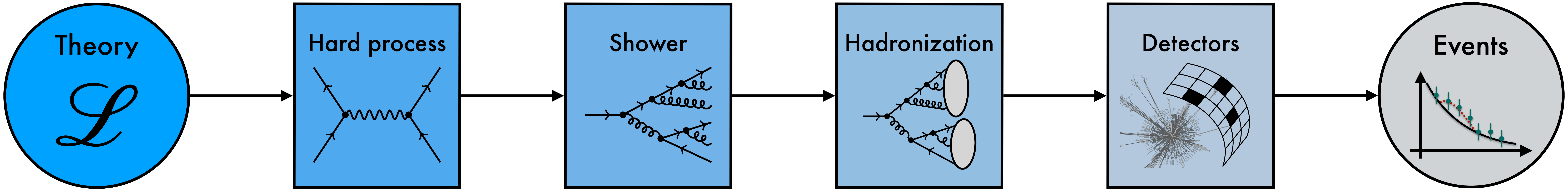


Summary and Outlook

Take-home messages

- ML beneficial in **every step** of the **simulation** and **analysis chain**
- We find both **proof-of-concepts** as well as established use cases (→ **AD, MadNIS,…**)
- Interesting **interplay** between **physics** and **ML**

→ Physics provides **~infinite data** for ML



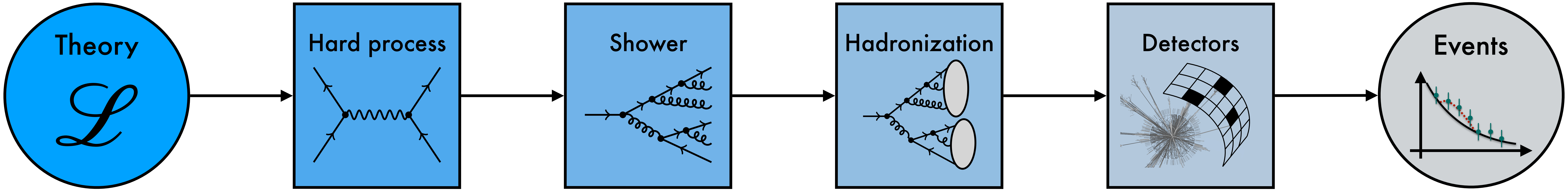
Summary and Outlook

Take-home messages

- ML beneficial in **every step** of the **simulation** and **analysis chain**
- We find both **proof-of-concepts** as well as established use cases (→ **AD, MadNIS,…**)
- Interesting **interplay** between **physics** and **ML**

→ Physics provides **~infinite data** for ML

→ Physics requirements (**precision, symmertries,…**) **different** than industry applications



Summary and Outlook

Take-home messages

- ML beneficial in **every step** of the **simulation** and **analysis chain**
- We find both **proof-of-concepts** as well as established use cases (→ **AD, MadNIS,…**)
- Interesting **interplay** between **physics** and **ML**

  → Physics provides **~infinite data** for ML

  → Physics requirements (**precision, symmertries,…**) **different** than industry applications

Future exercises

- Full integration of ML-based methods into standard tools → **Taggers, MadGraph,….**
- Make everything run on **GPUs** and make it **differentiable**
- Foster deeper collaboration between **theory**, **experiment**, and **ML** community