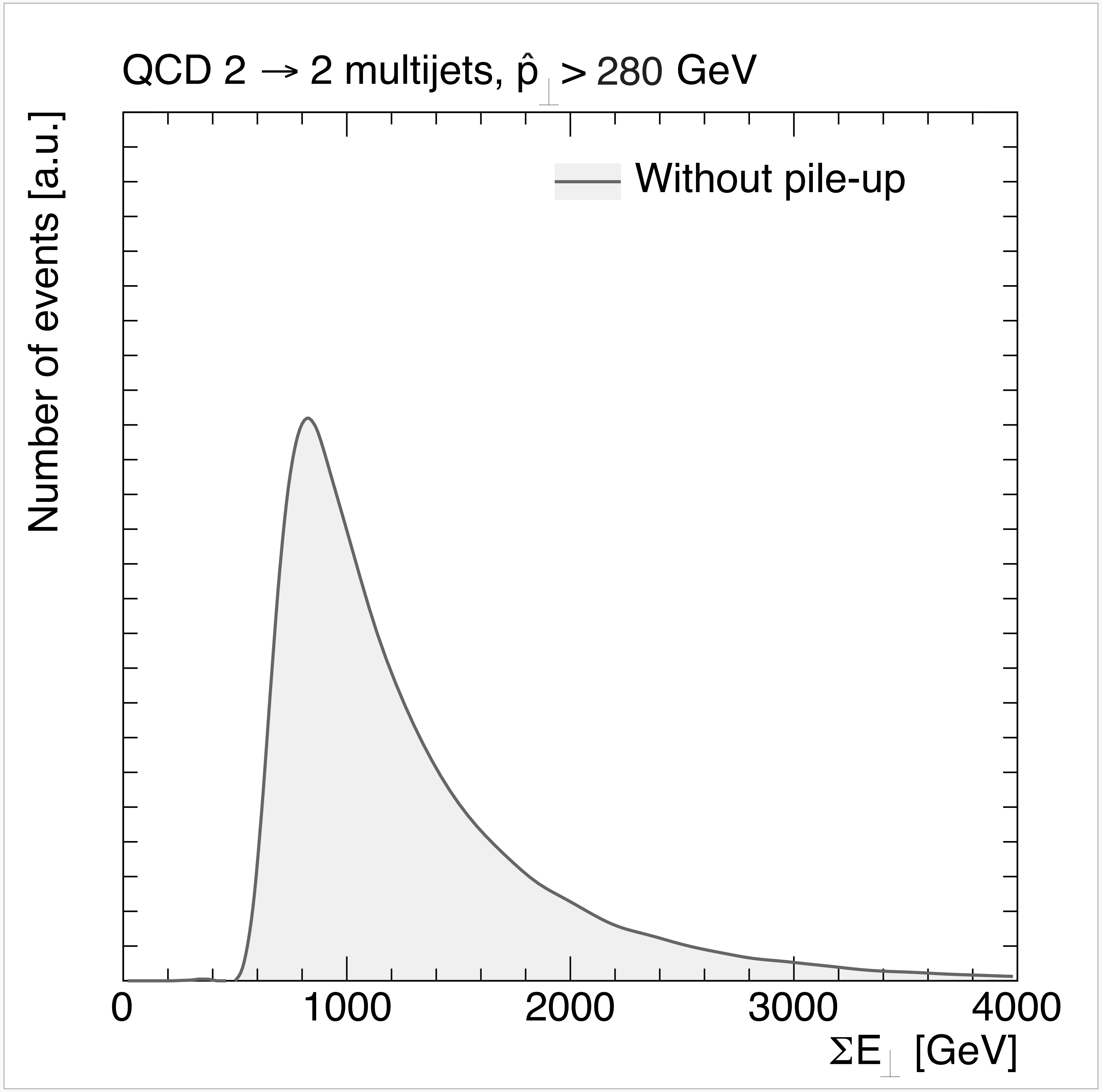
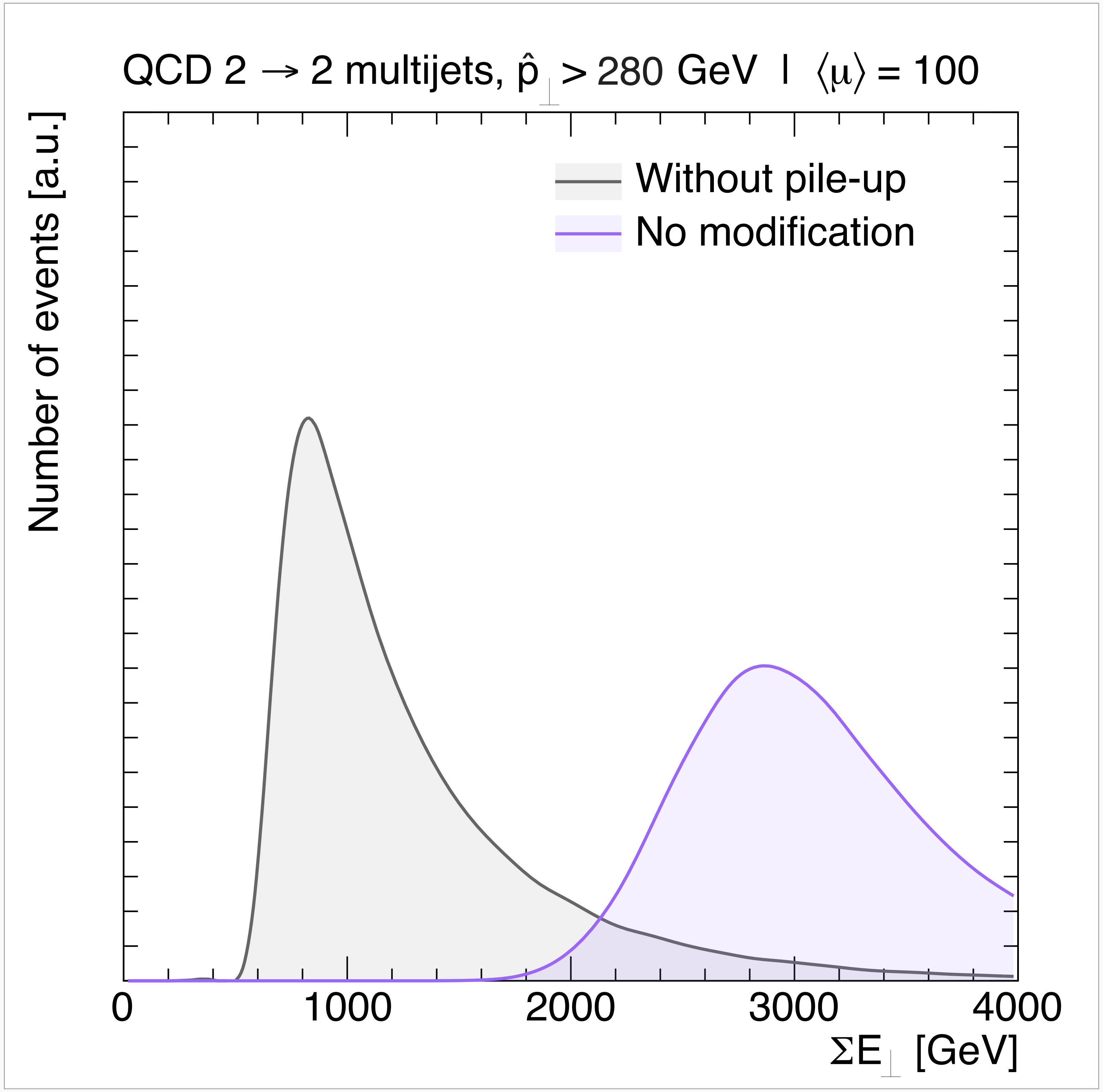
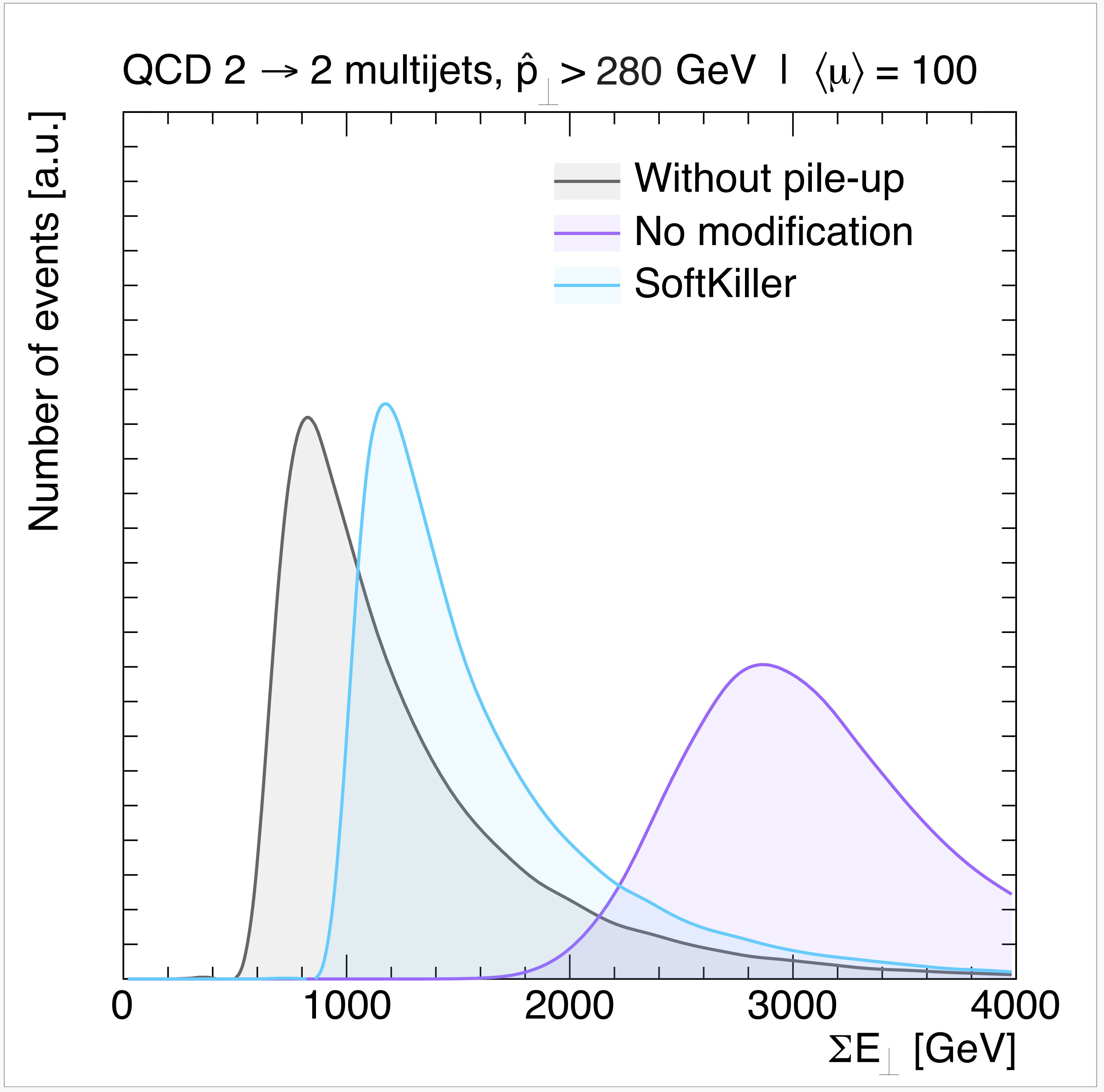
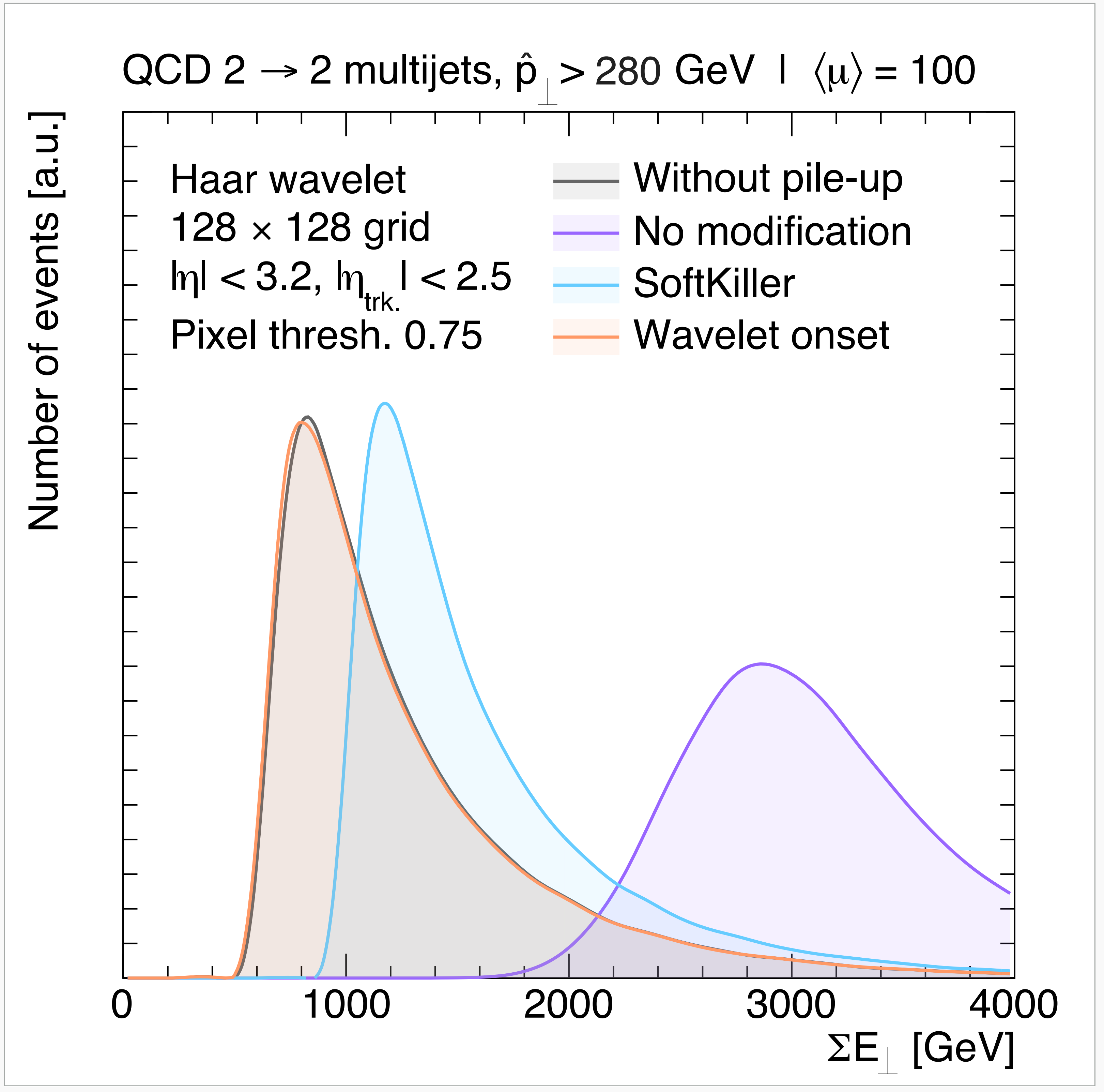

QCD 2 → 2 multijets, p̂⊥ > 280 GeV
Without pile-up
Number of events [a.u.]
0
1000
2000
3000
4000
ΣE⊥ [GeV]



QCD 2 → 2 multijets, p̂⊥ > 280 GeV | ⟨μ⟩ = 100
Without pile-up
No modification
Number of events [a.u.]
0
1000
2000
3000
4000
ΣE⊥ [GeV]



QCD 2 → 2 multijets, p̂⊥ > 280 GeV | ⟨μ⟩ = 100
Number of events [a.u.]
Without pile-up
No modification
SoftKiller
0
1000
2000
3000
4000
ΣE⊥ [GeV]



QCD 2 → 2 multijets, p̂⊥ > 280 GeV | ⟨μ⟩ = 100
Number of events [a.u.]
Haar wavelet
128 × 128 grid
|η| < 3.2, |ηtrk.| < 2.5
Pixel thresh. 0.75
Without pile-up
No modification
SoftKiller
Wavelet onset
0
1000
2000
3000
4000
ΣE⊥ [GeV]

Using wavelets for pile-up mitigation

James Monk · Troels Petersen · **Andreas Søgaard**

University of Copenhagen · University of Edinburgh

BOOST Conference · Zürich
20 July 2016

Outline

1. Wavelet fundamentals
2. Missing- and sum E_T studies
3. Boosted jet studies
4. Summary and outlook

- *Bonus: Learning optimal bases*

Wavelet fundamentals

- Basis functions encoding both *frequency* and *position*
  - “Localised Fourier series”

Wavelet fundamentals

- Basis functions encoding both *frequency* and *position*
  - "Localised Fourier series"
- Angular information at different frequency scales, or 'bands'

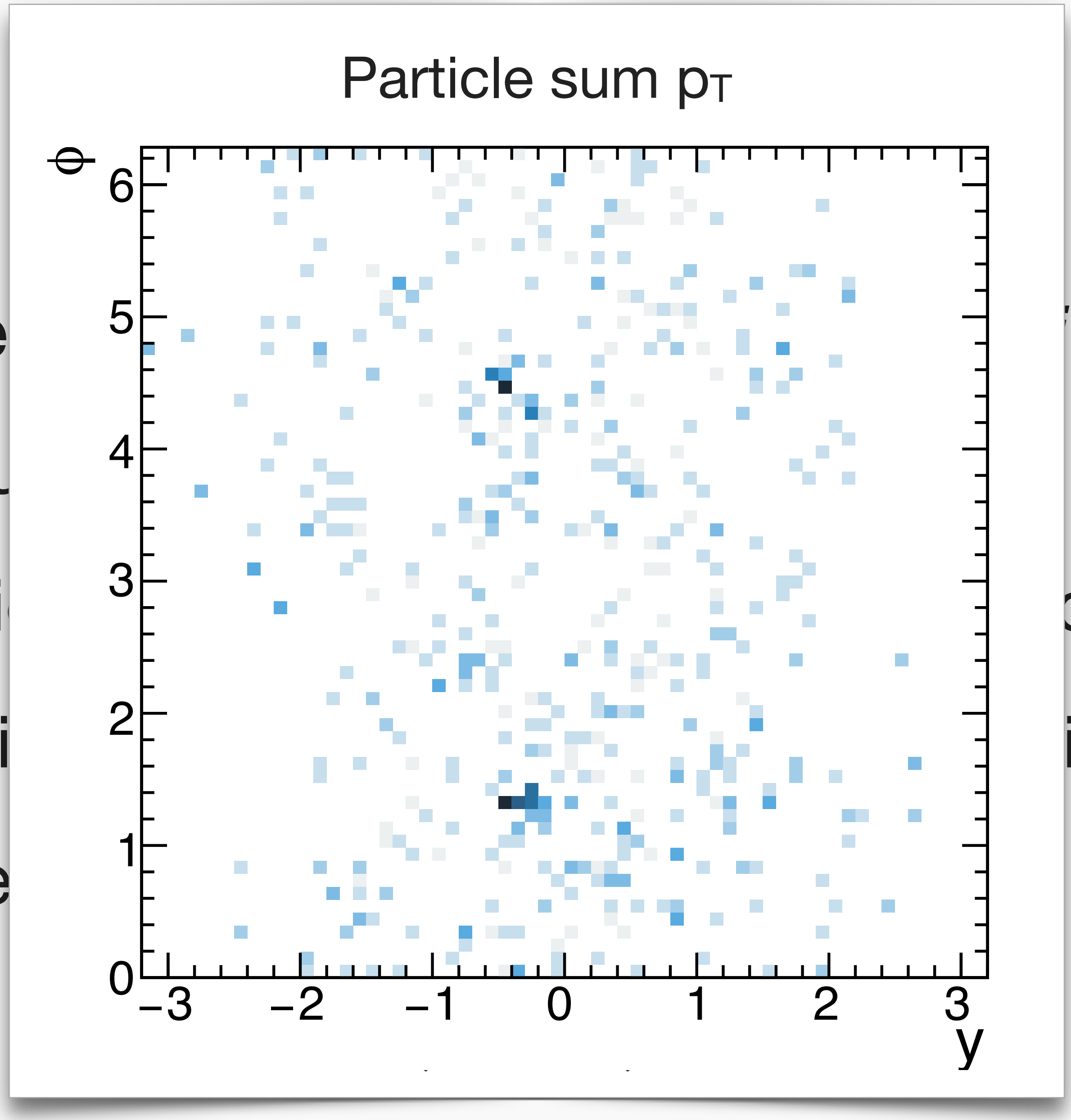
Wavelet fundamentals

- Basis functions encoding both *frequency* and *position*
  - "Localised Fourier series"
- Angular information at different frequency scales, or 'bands'
- Used for de-noising in e.g. imaging and astrophysics

Wavelet fundamentals

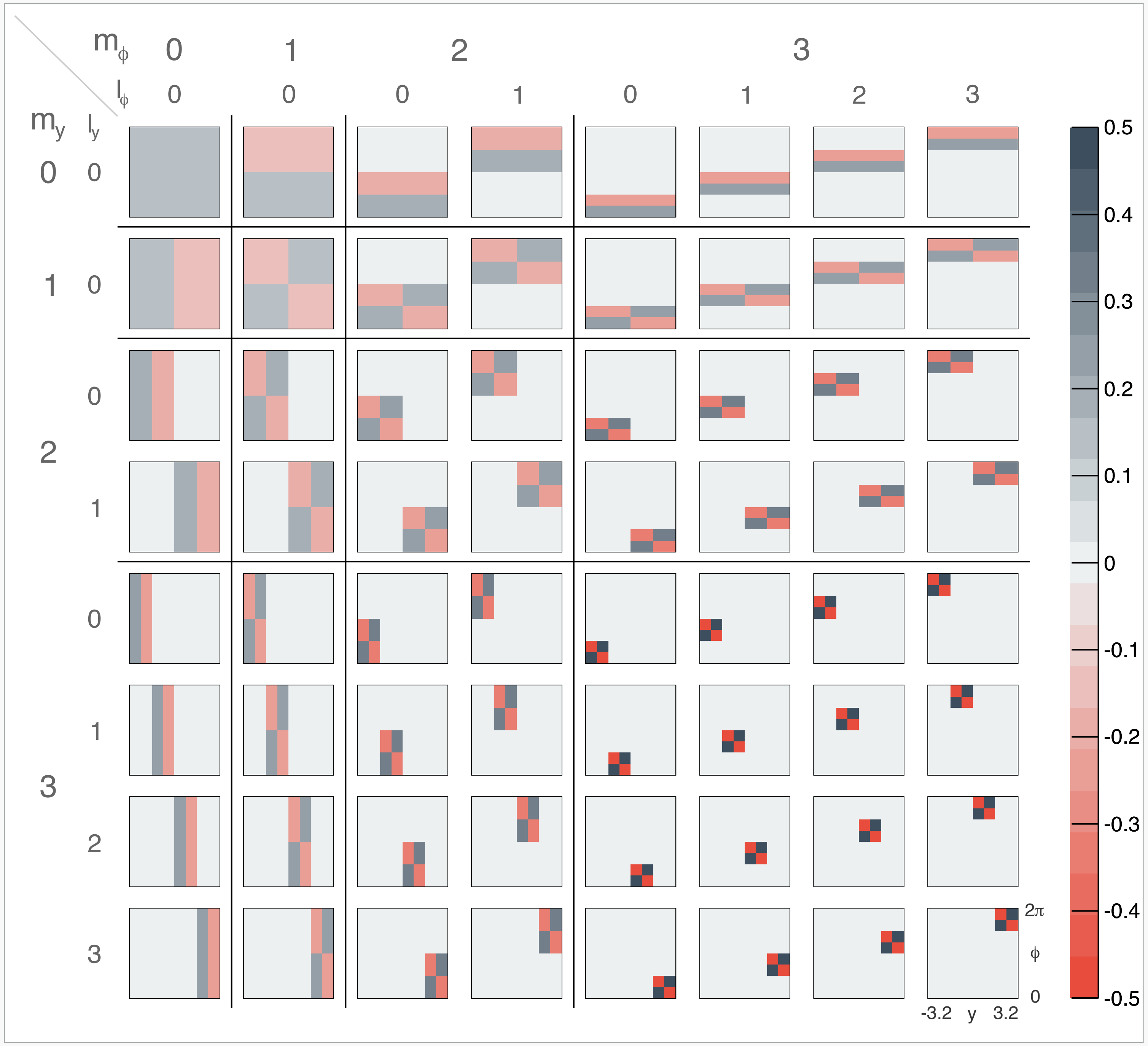
- Basis functions encoding both *frequency* and *position*
  - "Localised Fourier series"
- Angular information at different frequency scales, or 'bands'
- Used for de-noising in e.g. imaging and astrophysics
- For HEP purposes: Discrete 2D (y,ϕ)-input

Wavelet fundamentals

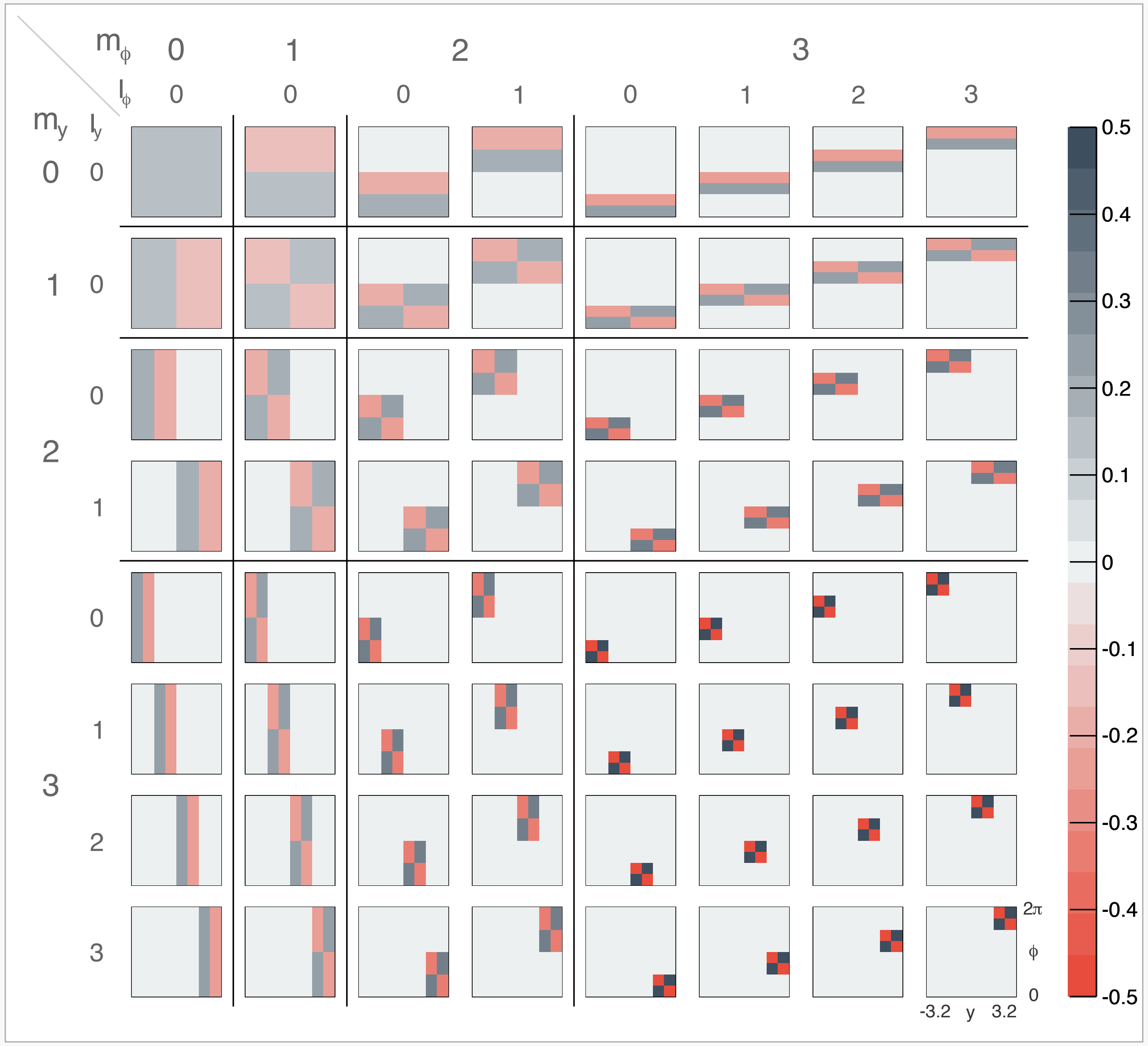


- Basis functions e *tion*
  - “Localised Fou
- Angular informati or ‘bands’
- Used for de-noisi ics
- For HEP purpose

Basis functions · Haar



Basis functions · Haar

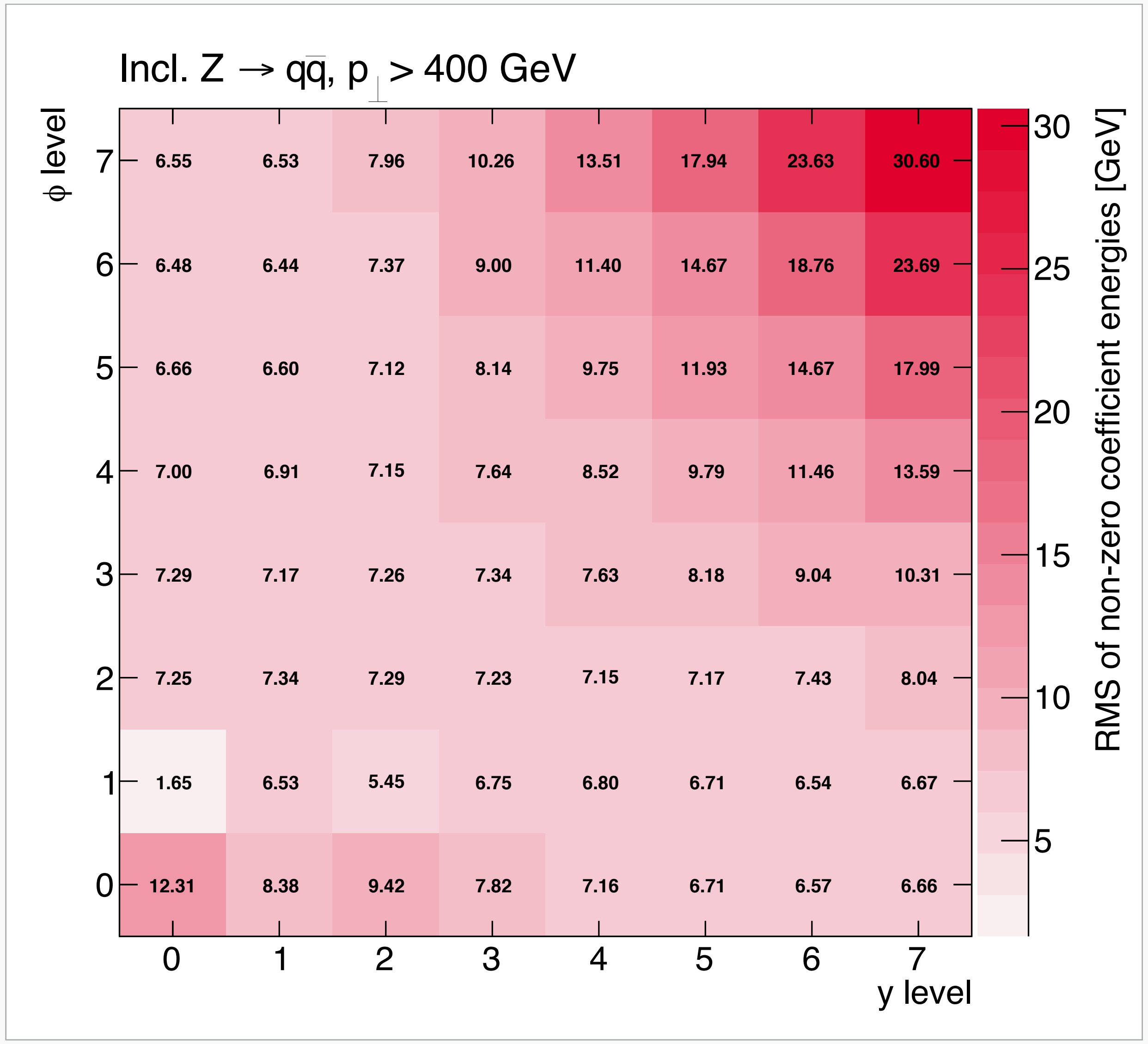


Structure of coefficient energies

- Hard scatter events:
  - Jets characterised by parton showering
  - Should be dominated by small-angle activity

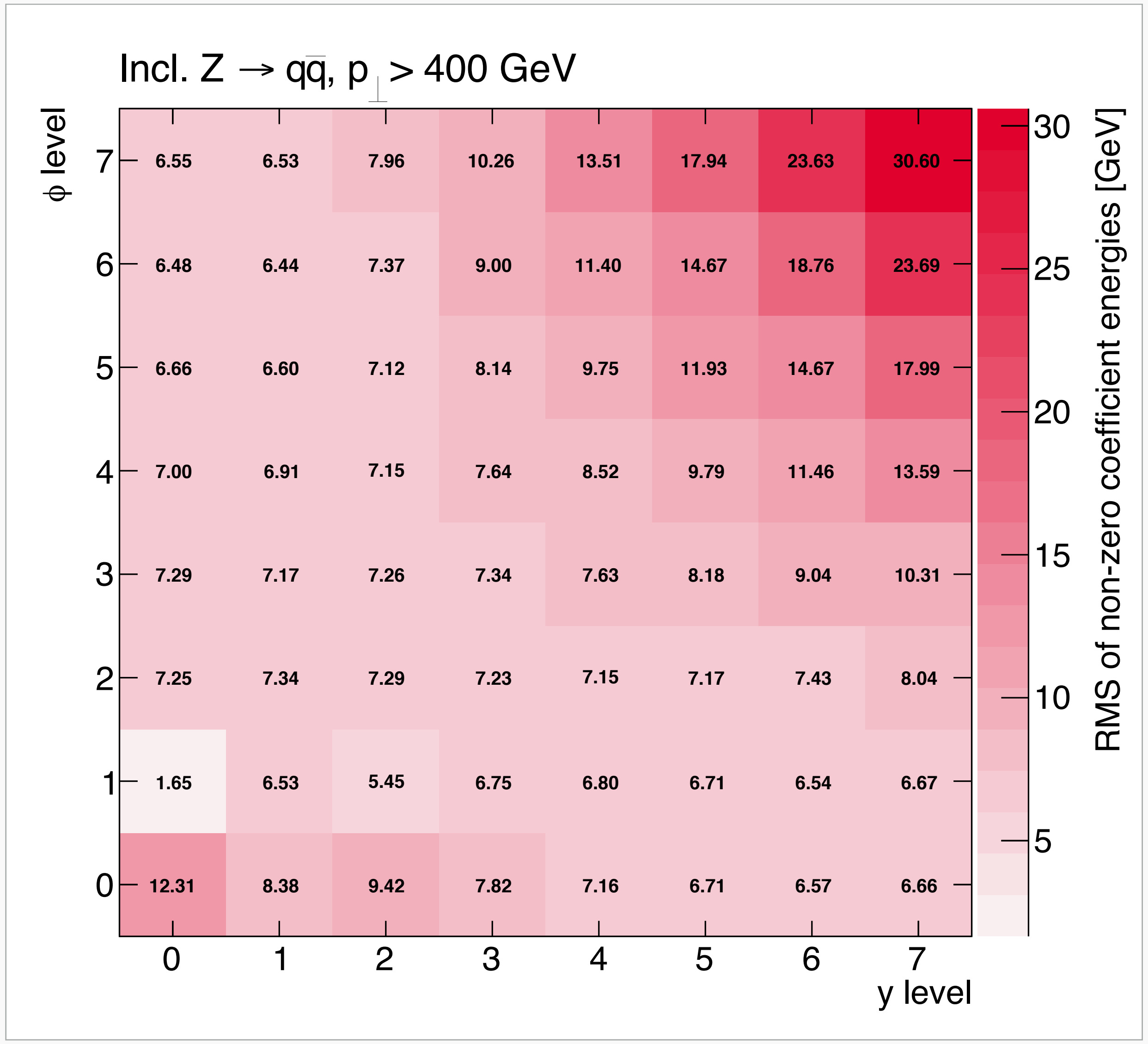
Structure of coefficient energies

- Hard scatter events:
  - Jets characterised by parton showering
  - Should be dominated by small-angle activity



Structure of coefficient energies

- Hard scatter events:
  - Jets characterised by parton showering
  - Should be dominated by small-angle activity

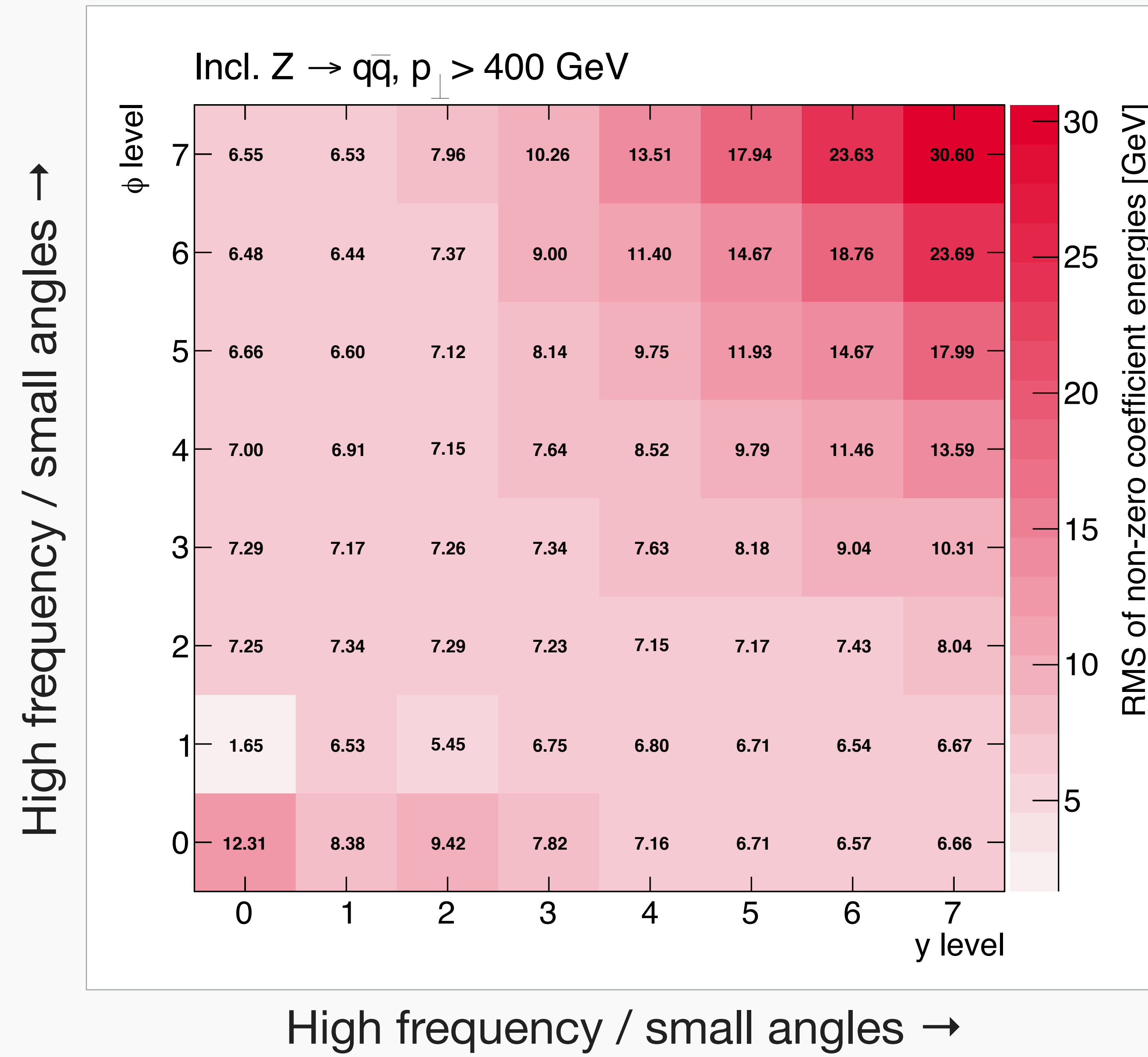


High frequency / small angles →

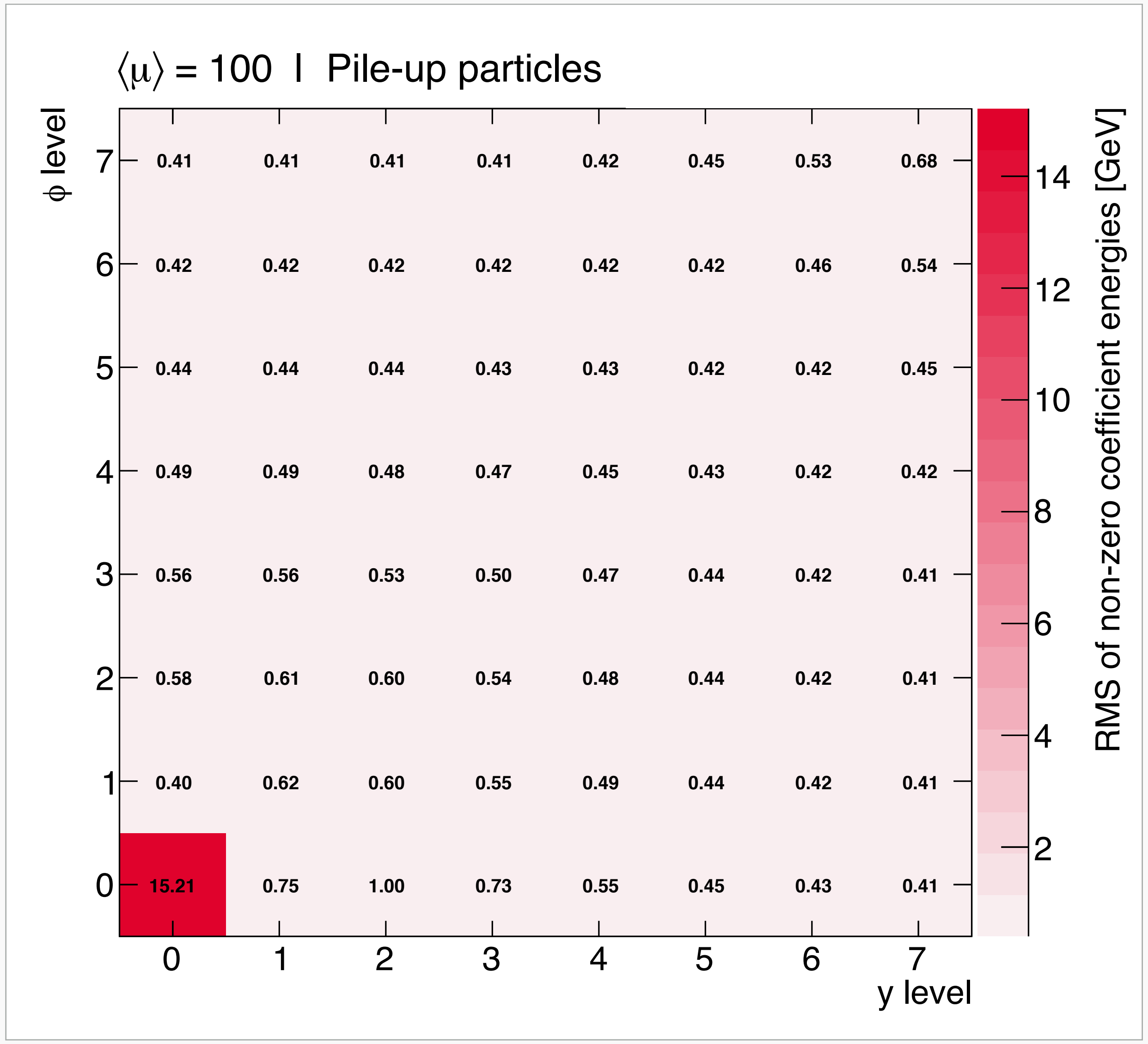
High frequency / small angles →

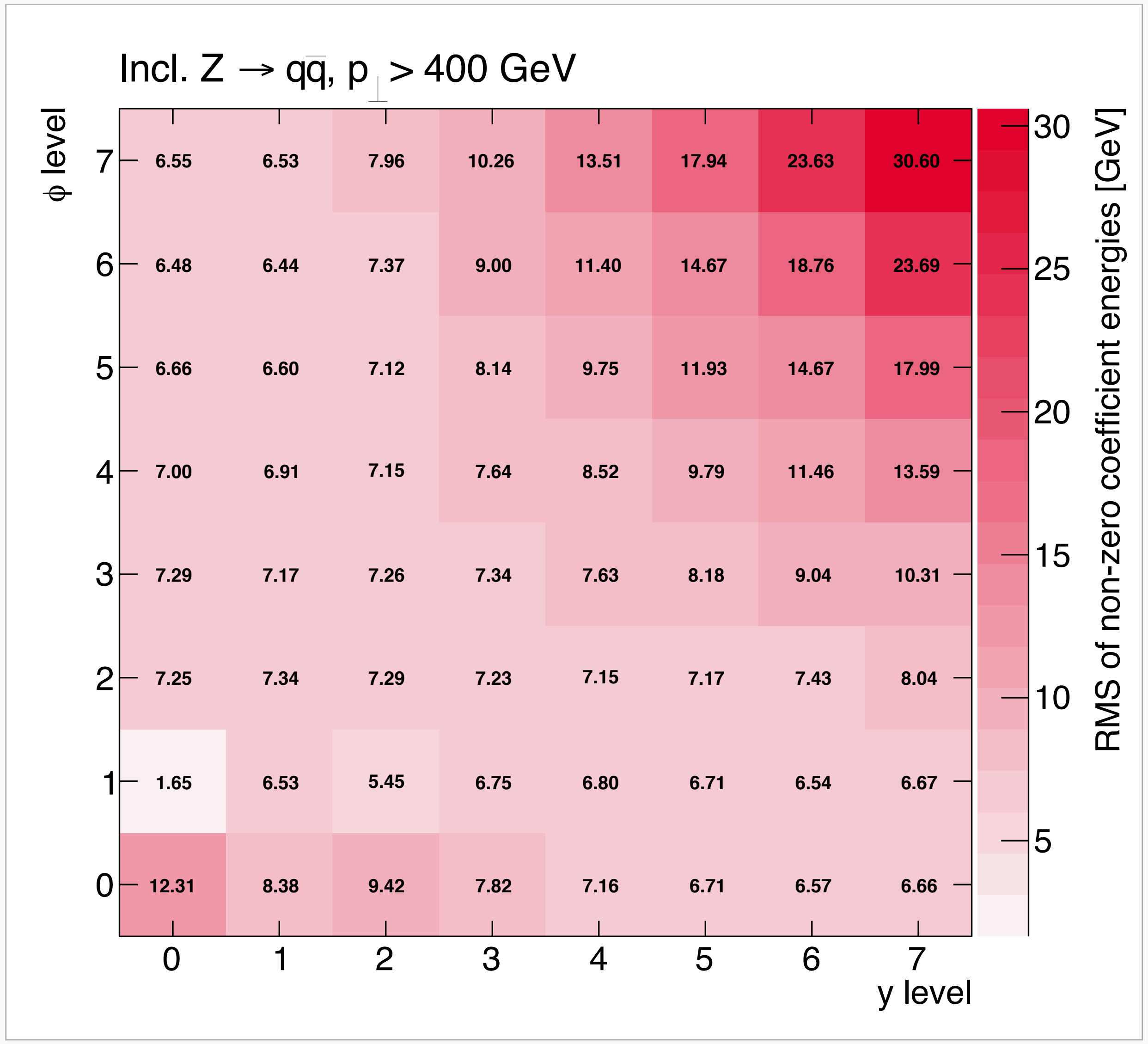
Structure of coefficient energies

- Hard scatter events:
  - Jets characterised by parton showering
  - Should be dominated by small-angle activity
- Pile-up:
  - “White noise”
  - No angular structure: constant activity across frequency bands

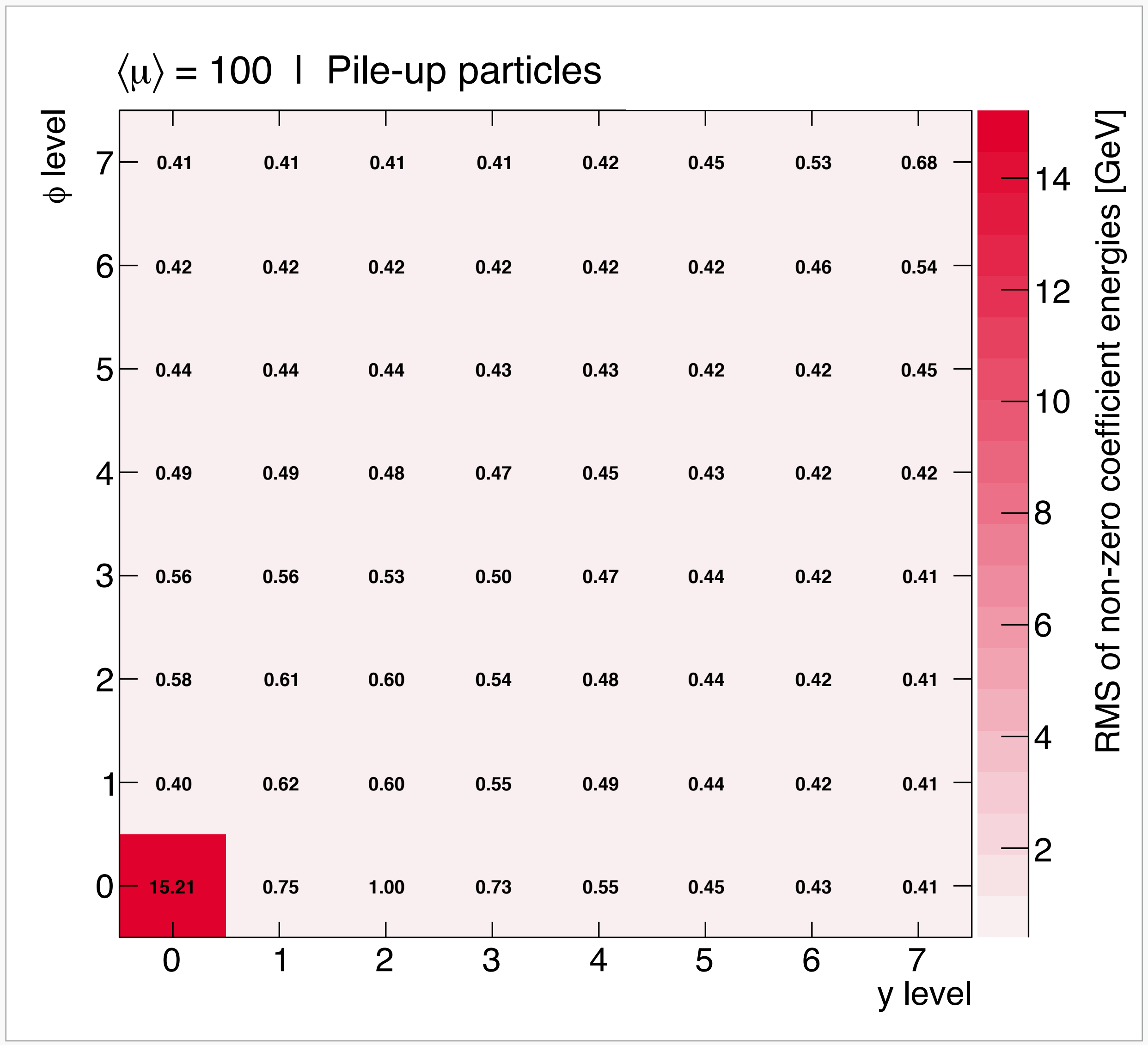


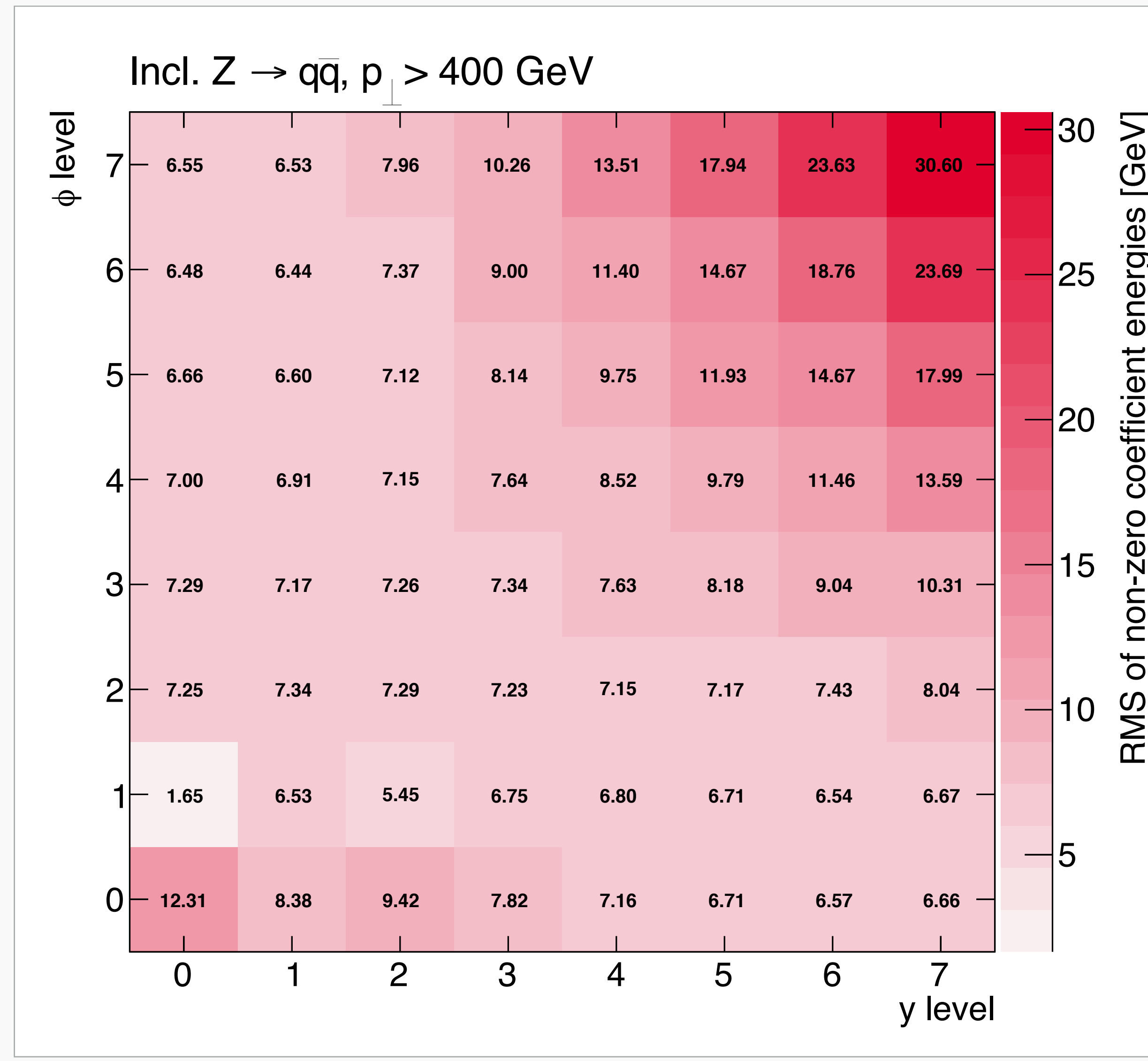
Structure of coefficient energies





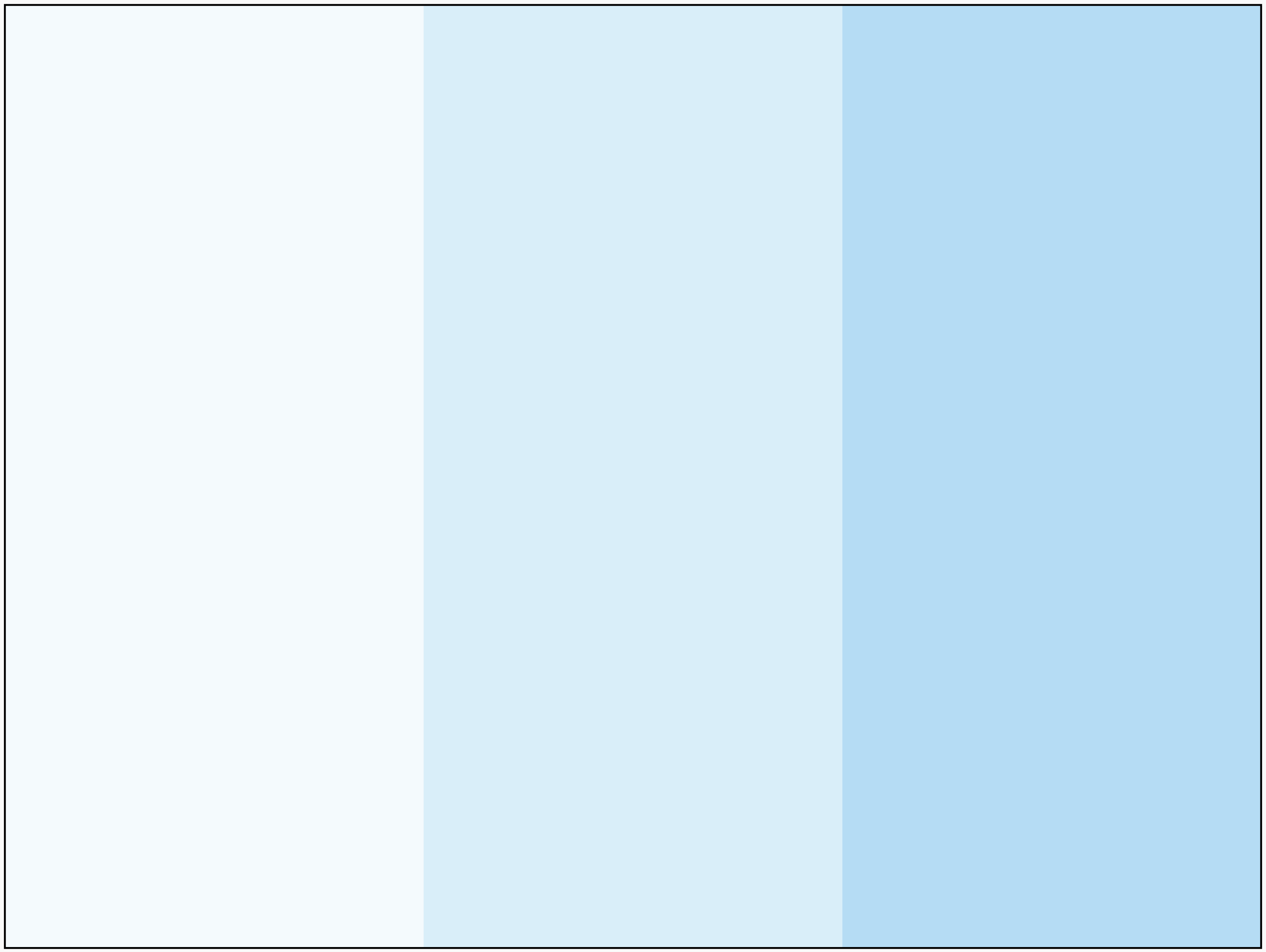
Structure of coefficient energies



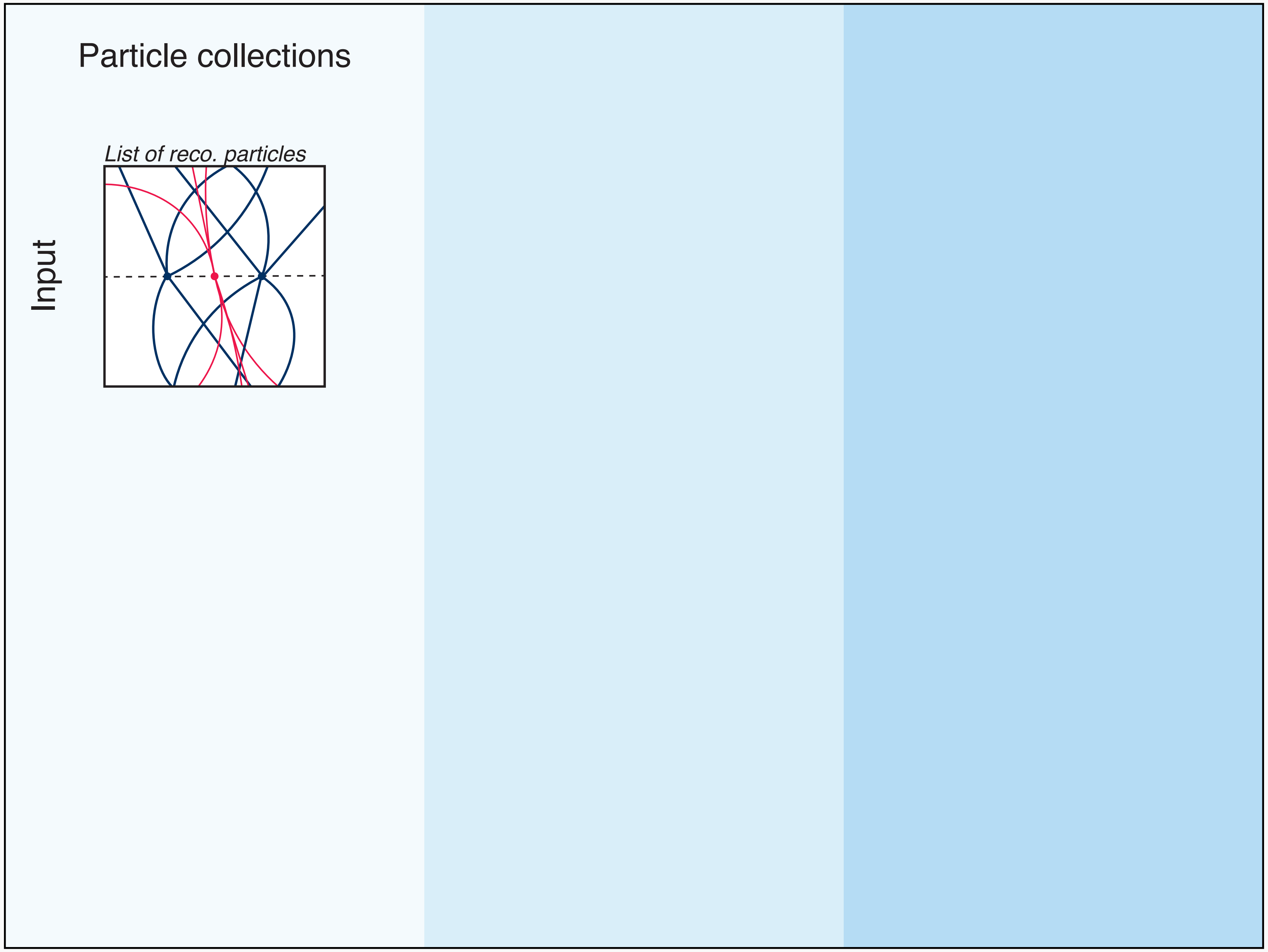


This difference may allow for good separation

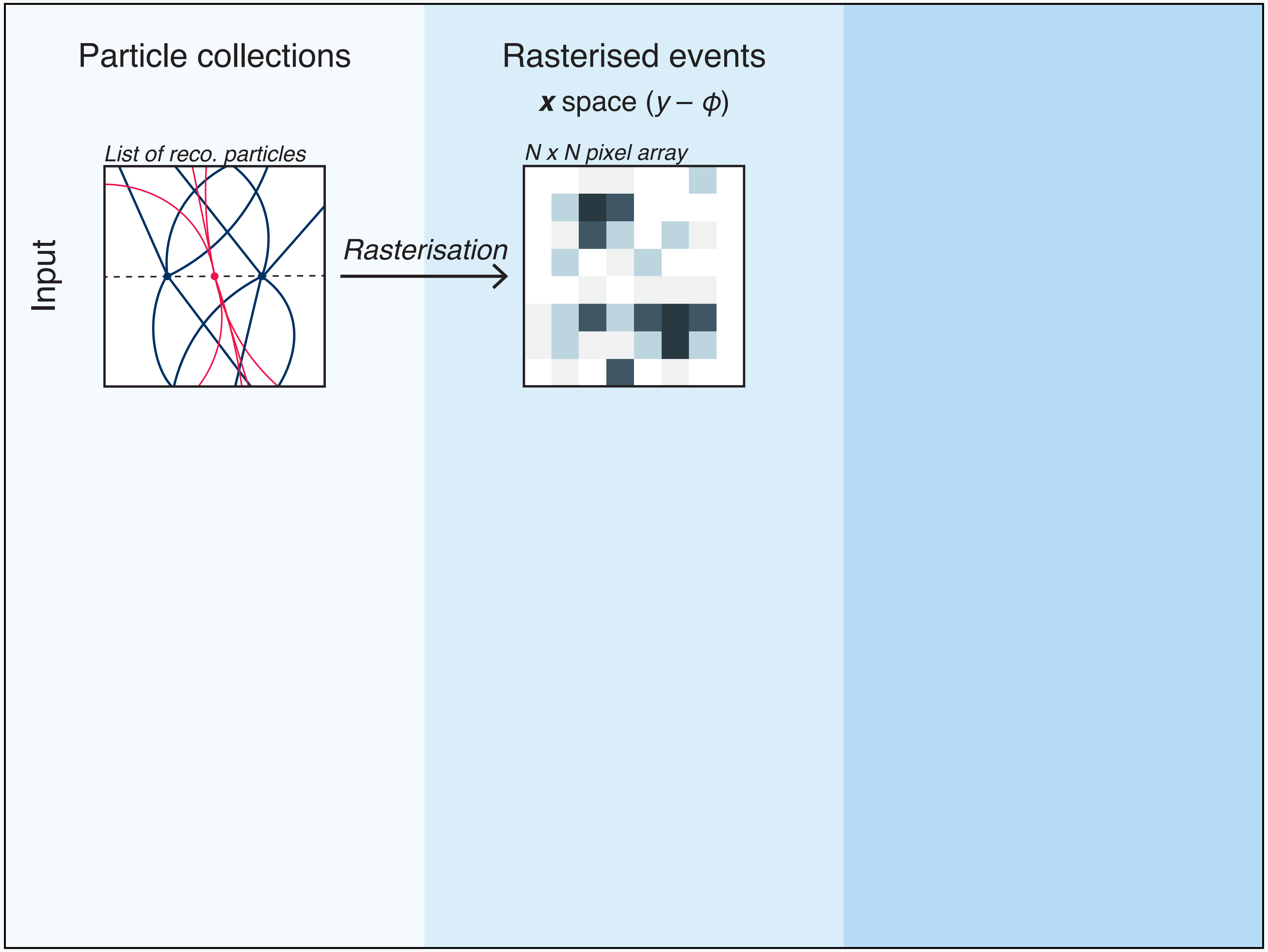
Wavelet analysis



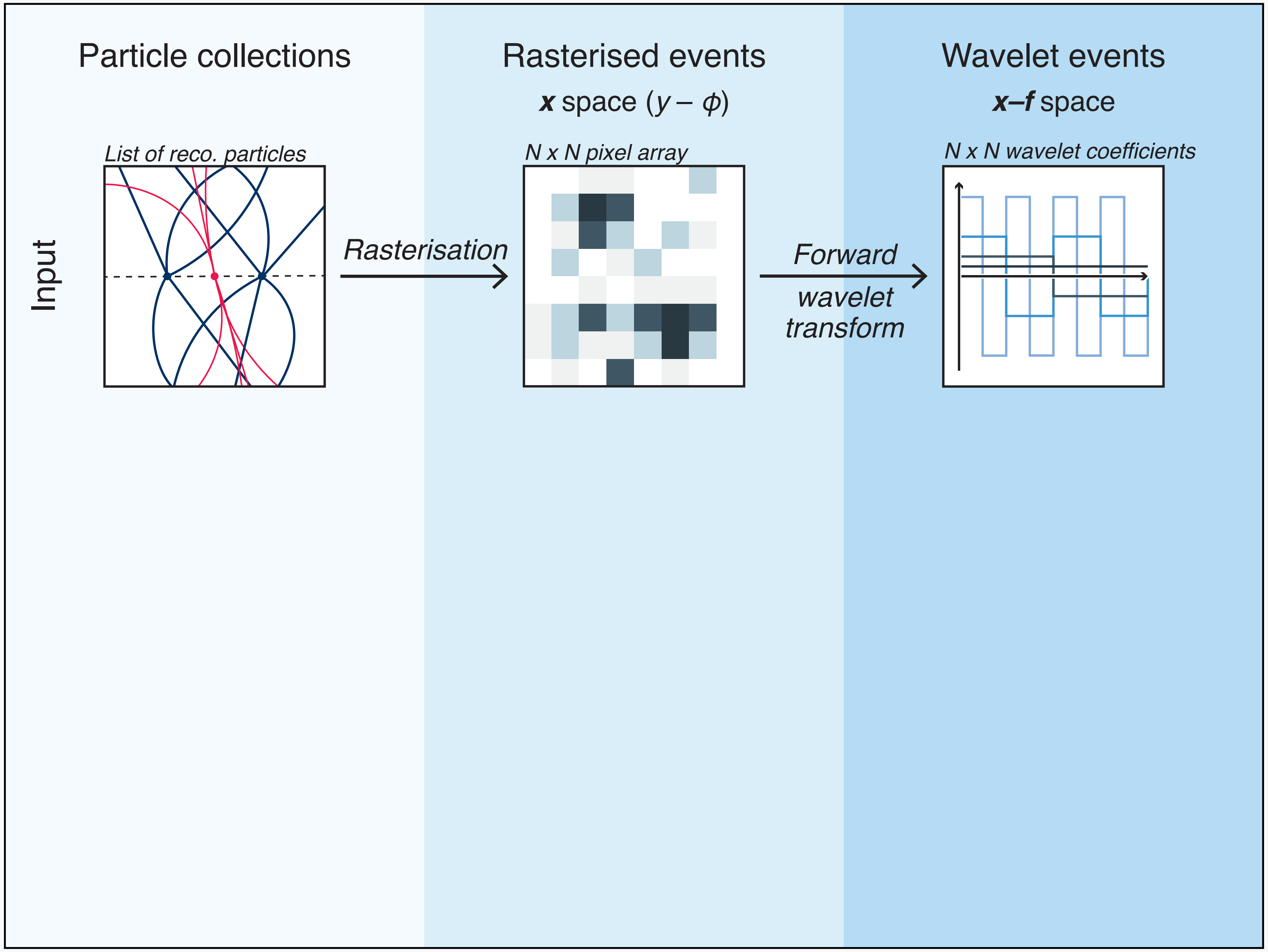
Wavelet analysis



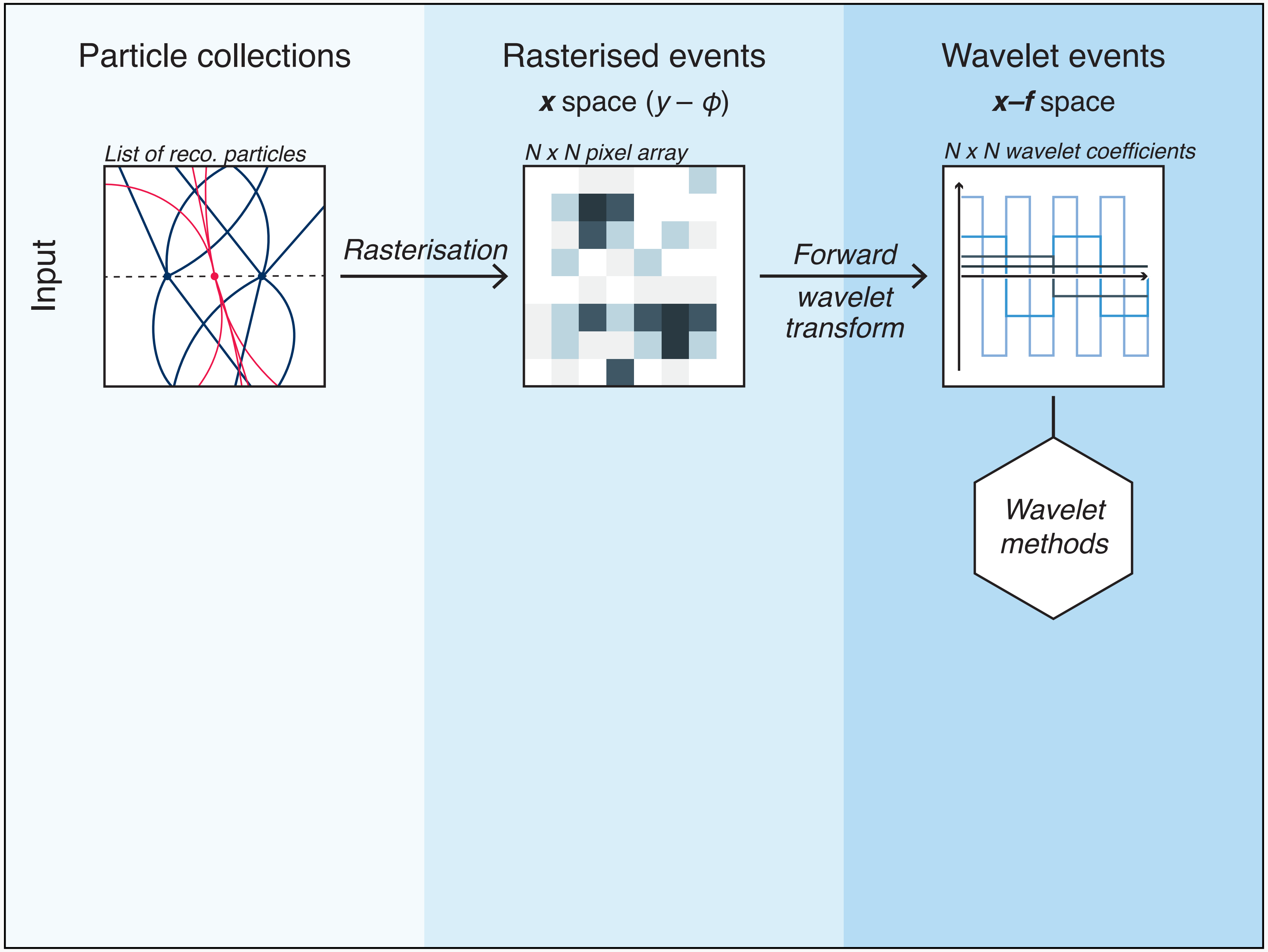
Wavelet analysis



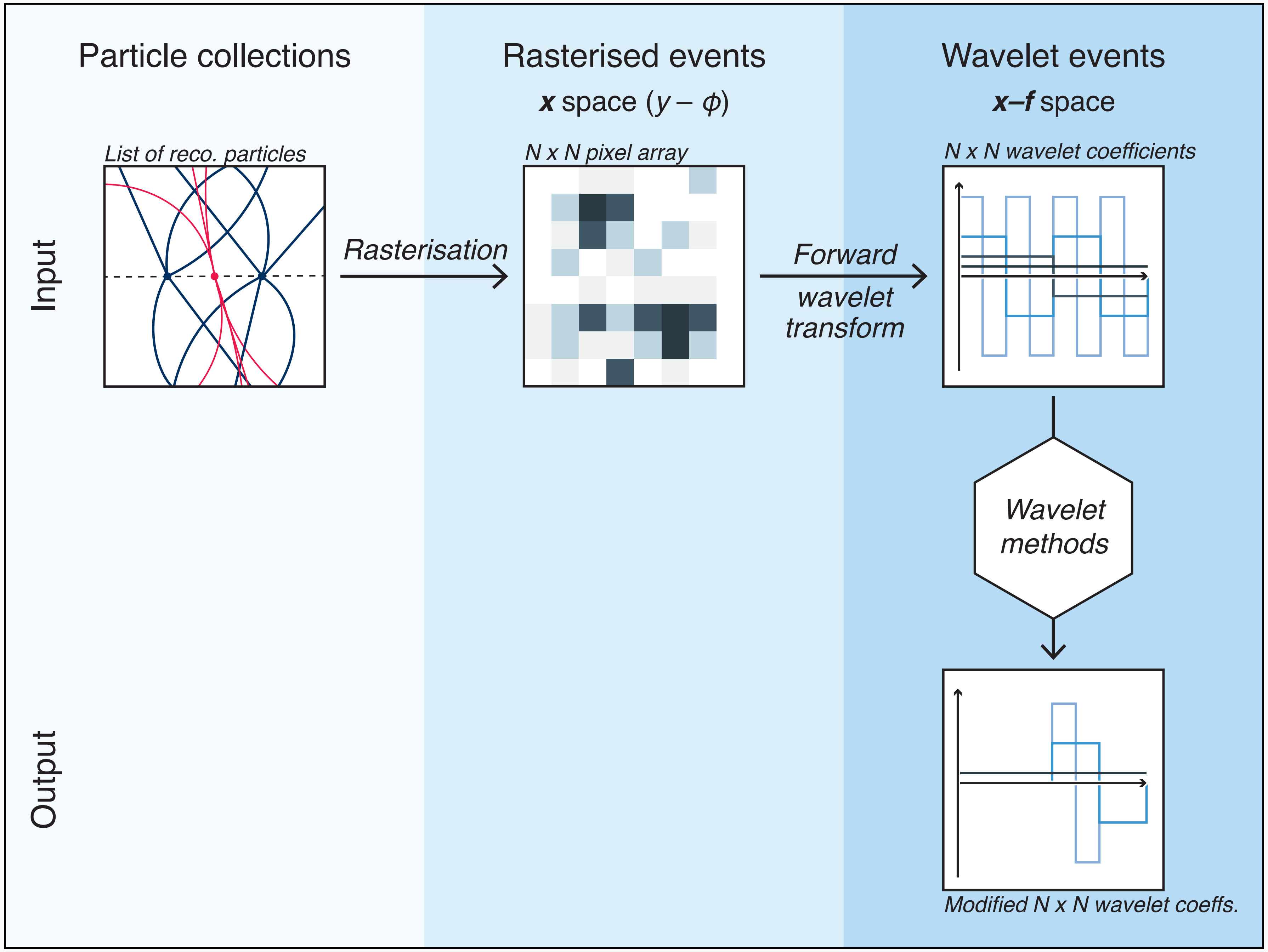
Wavelet analysis



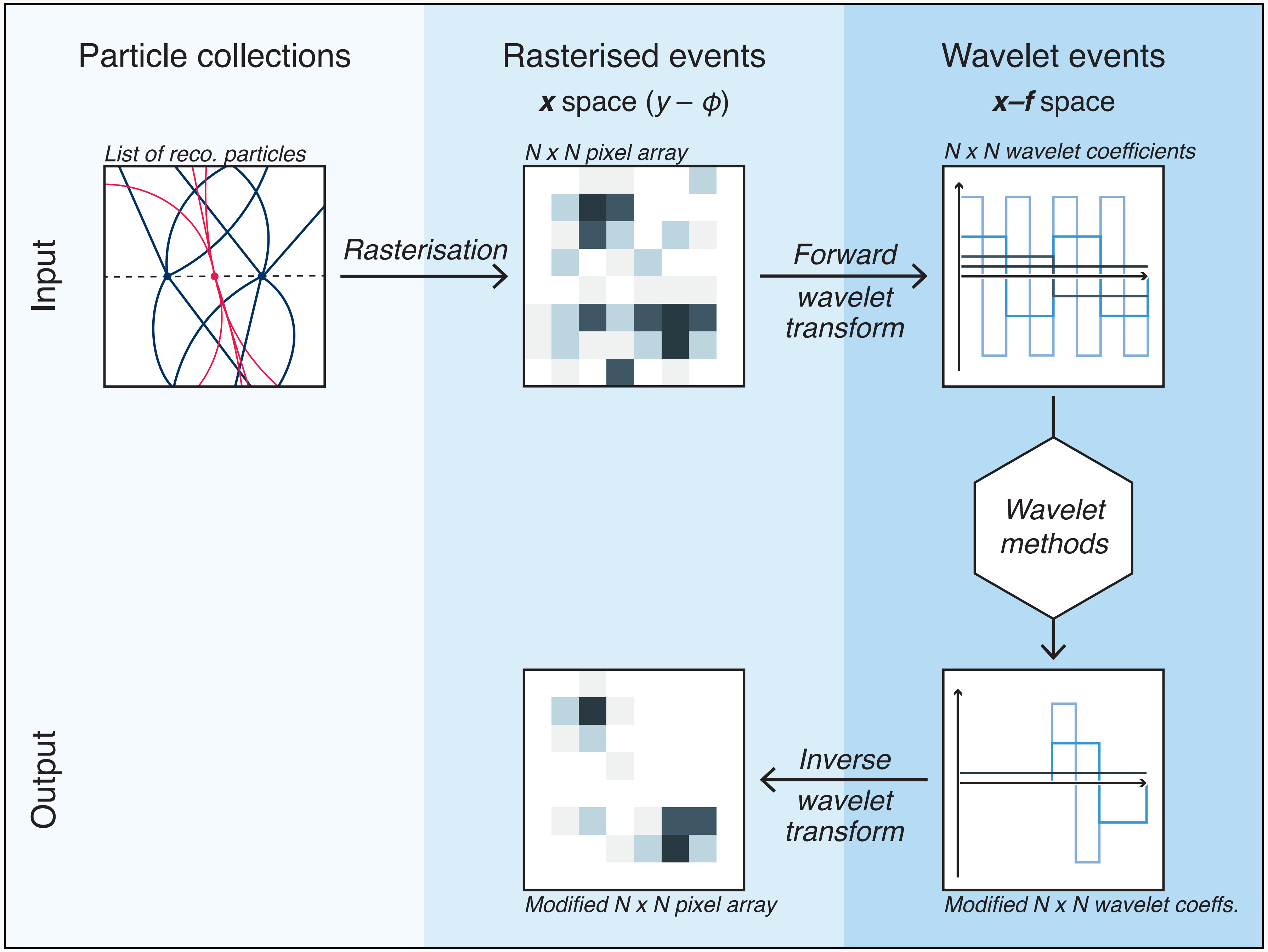
Wavelet analysis



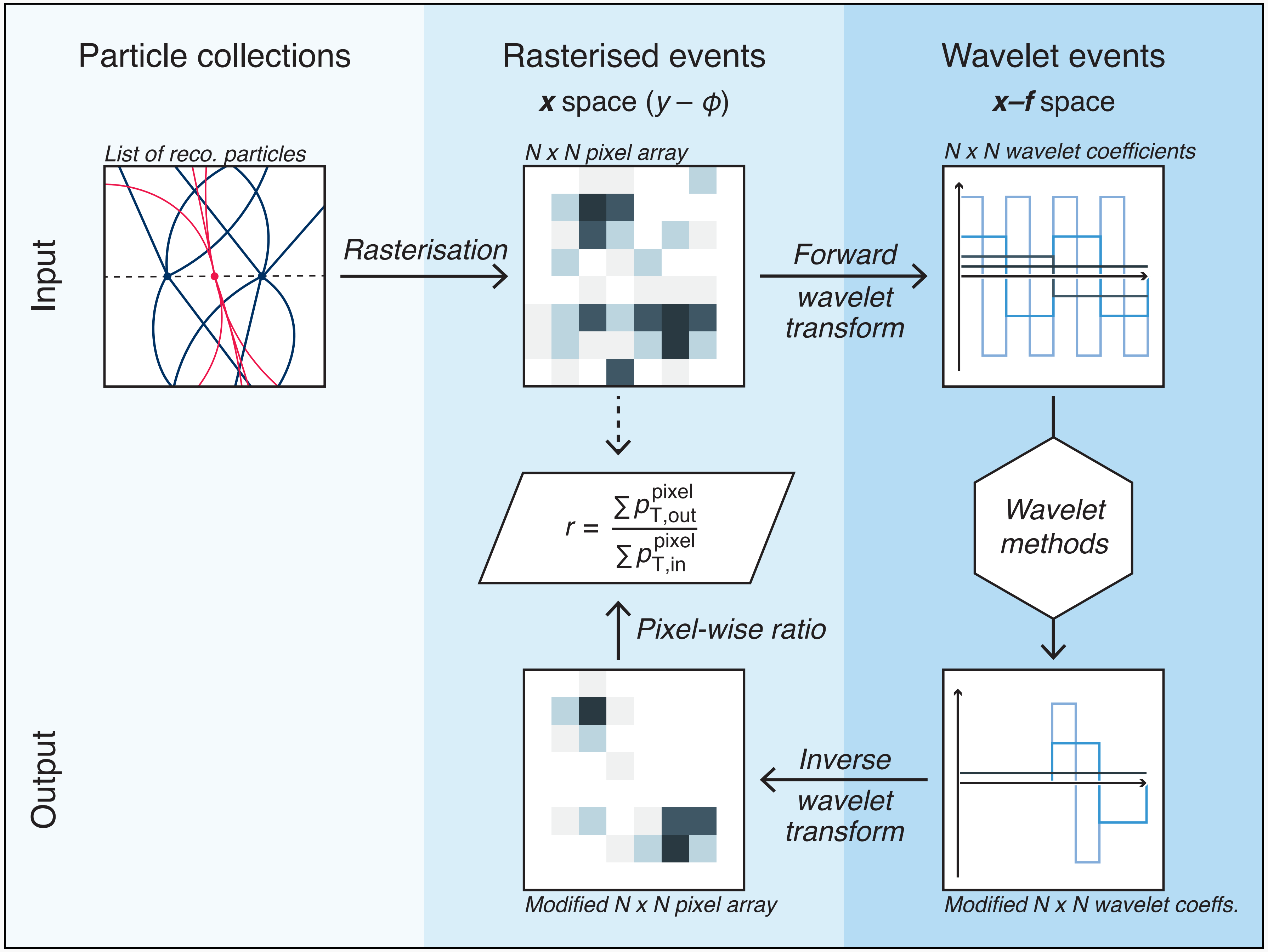
Wavelet analysis



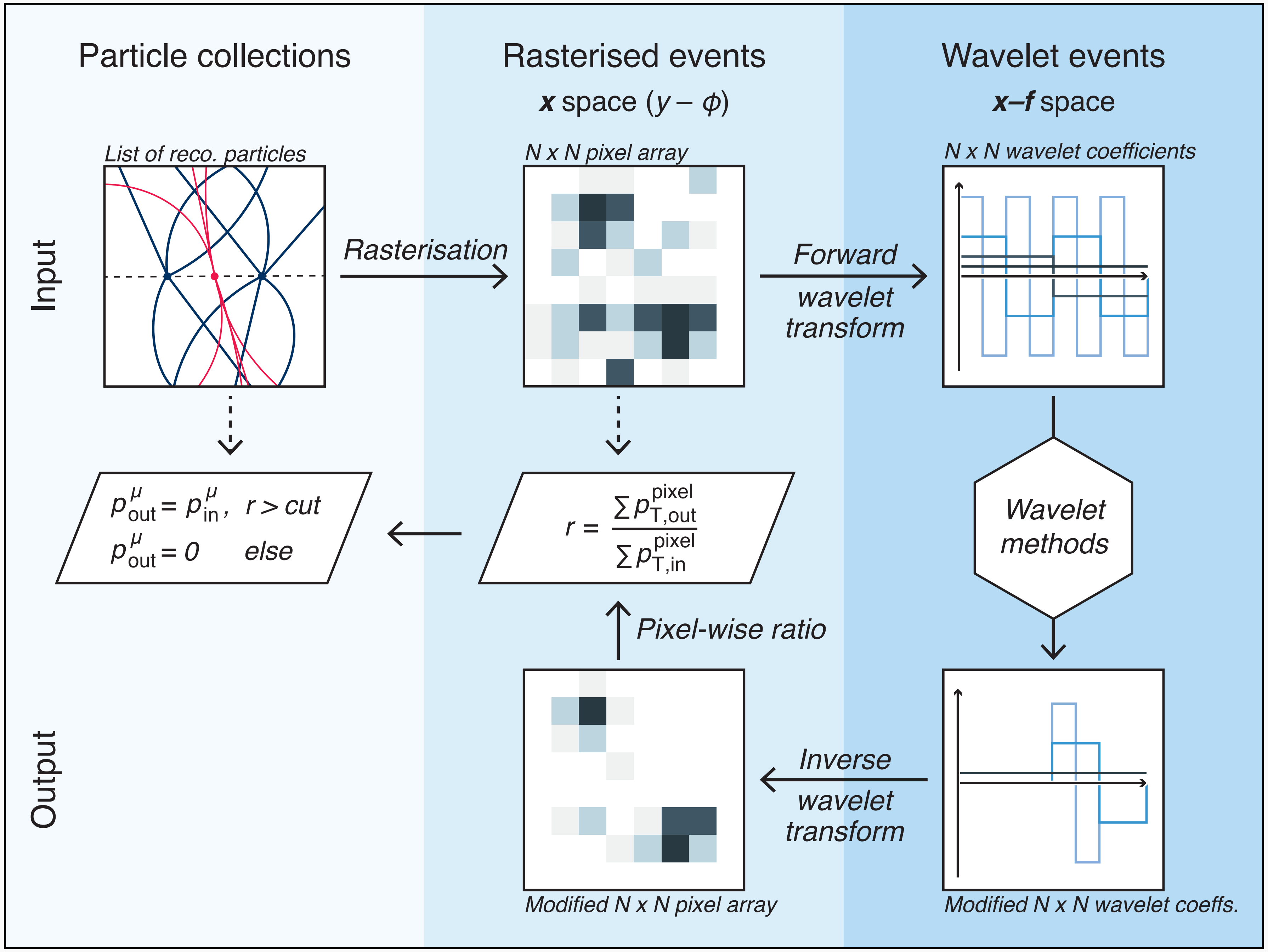
Wavelet analysis



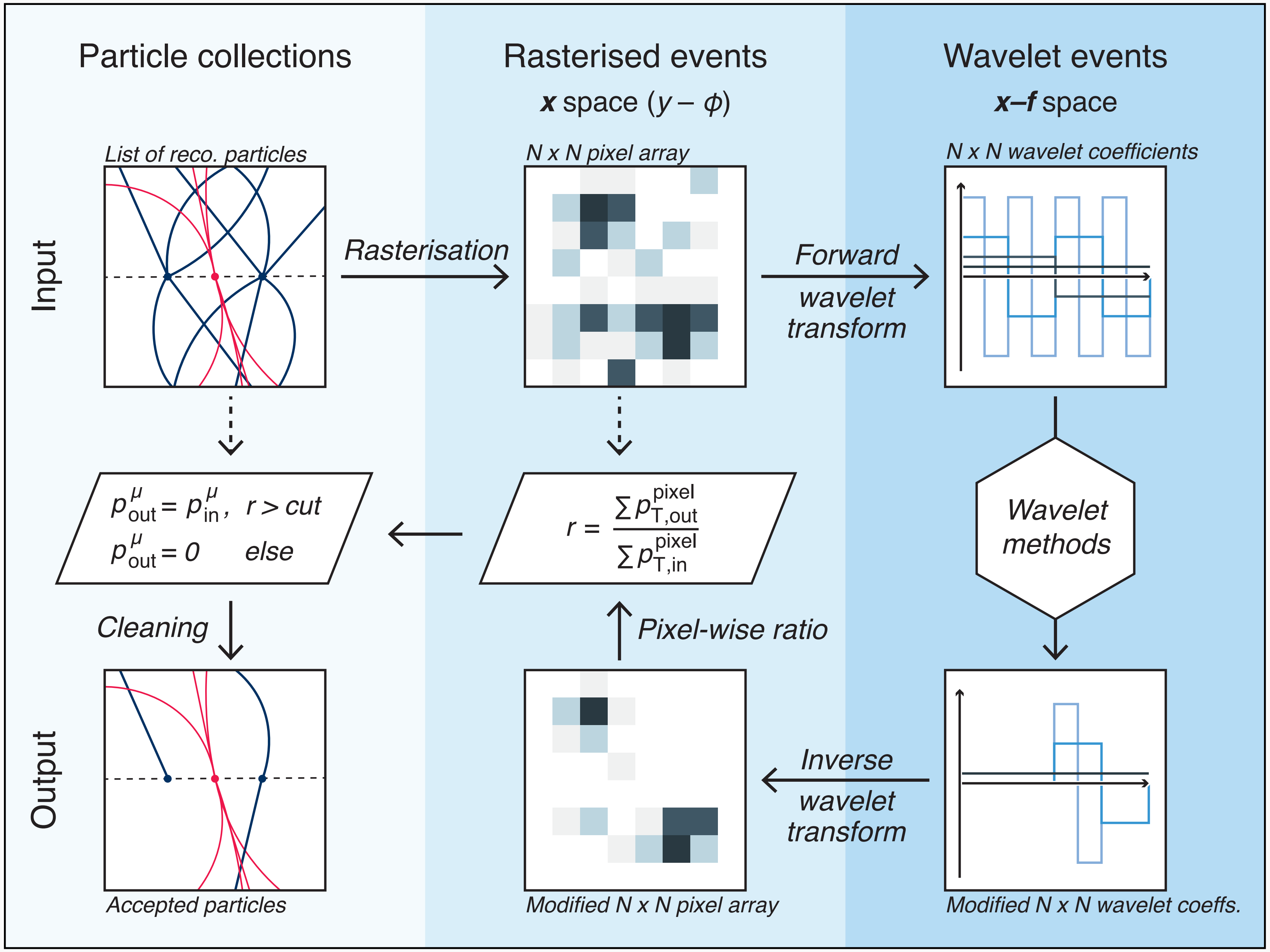
Wavelet analysis



Wavelet analysis



Wavelet analysis



Methods for pile-up mitigation

Methods for pile-up mitigation

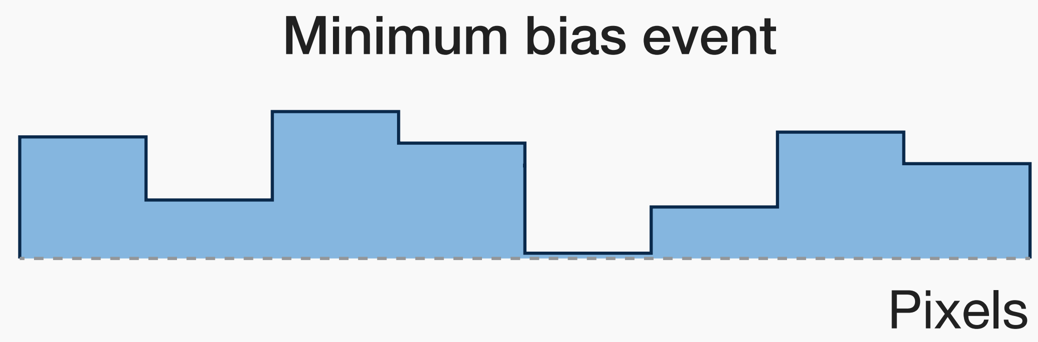
Simplest approach:

Methods for pile-up mitigation

Simplest approach:

- Scale (0,0) coefficient (average energy)

Methods for pile-up mitigation



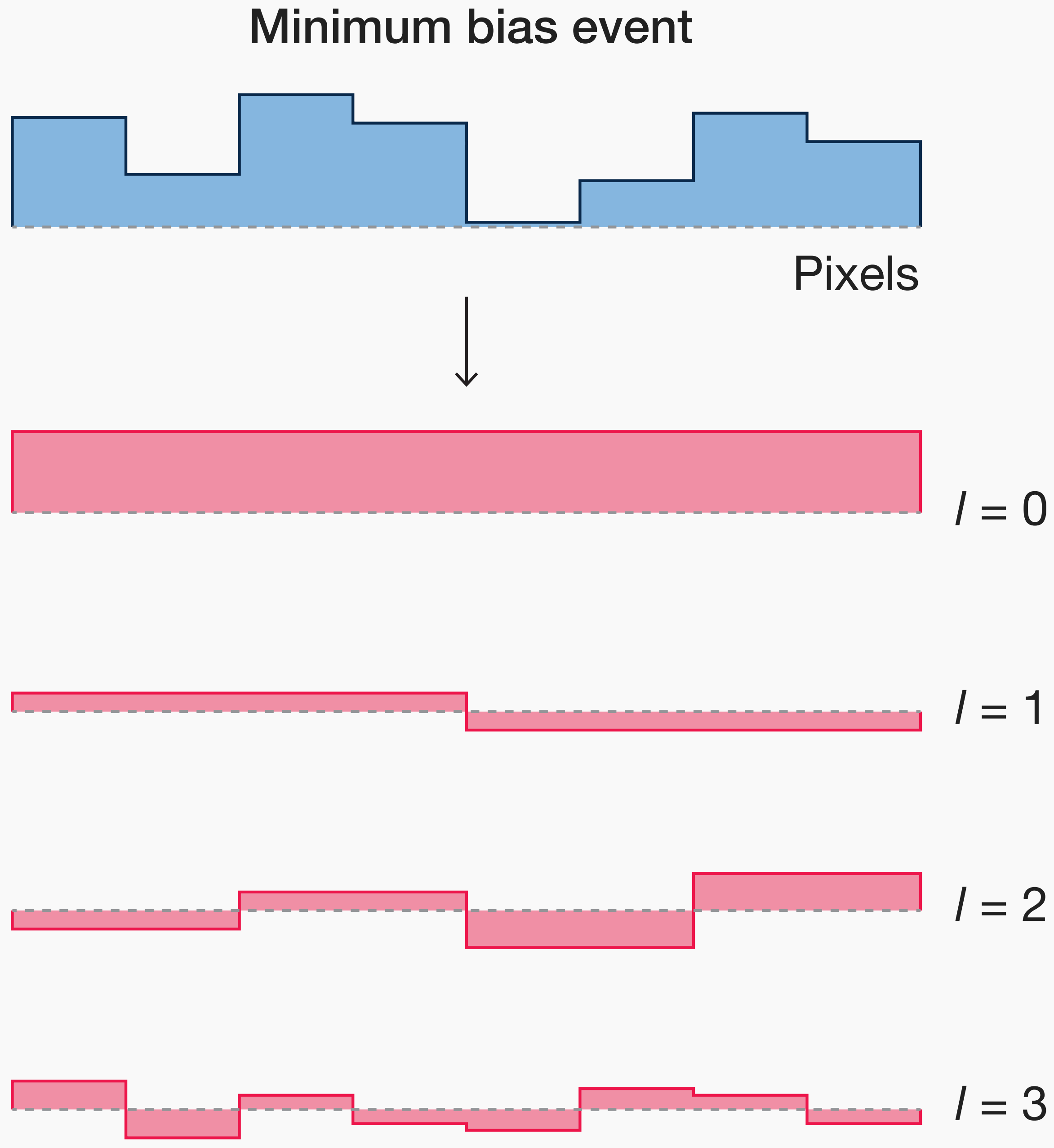
Simplest approach:

- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.

Methods for pile-up mitigation

Simplest approach:

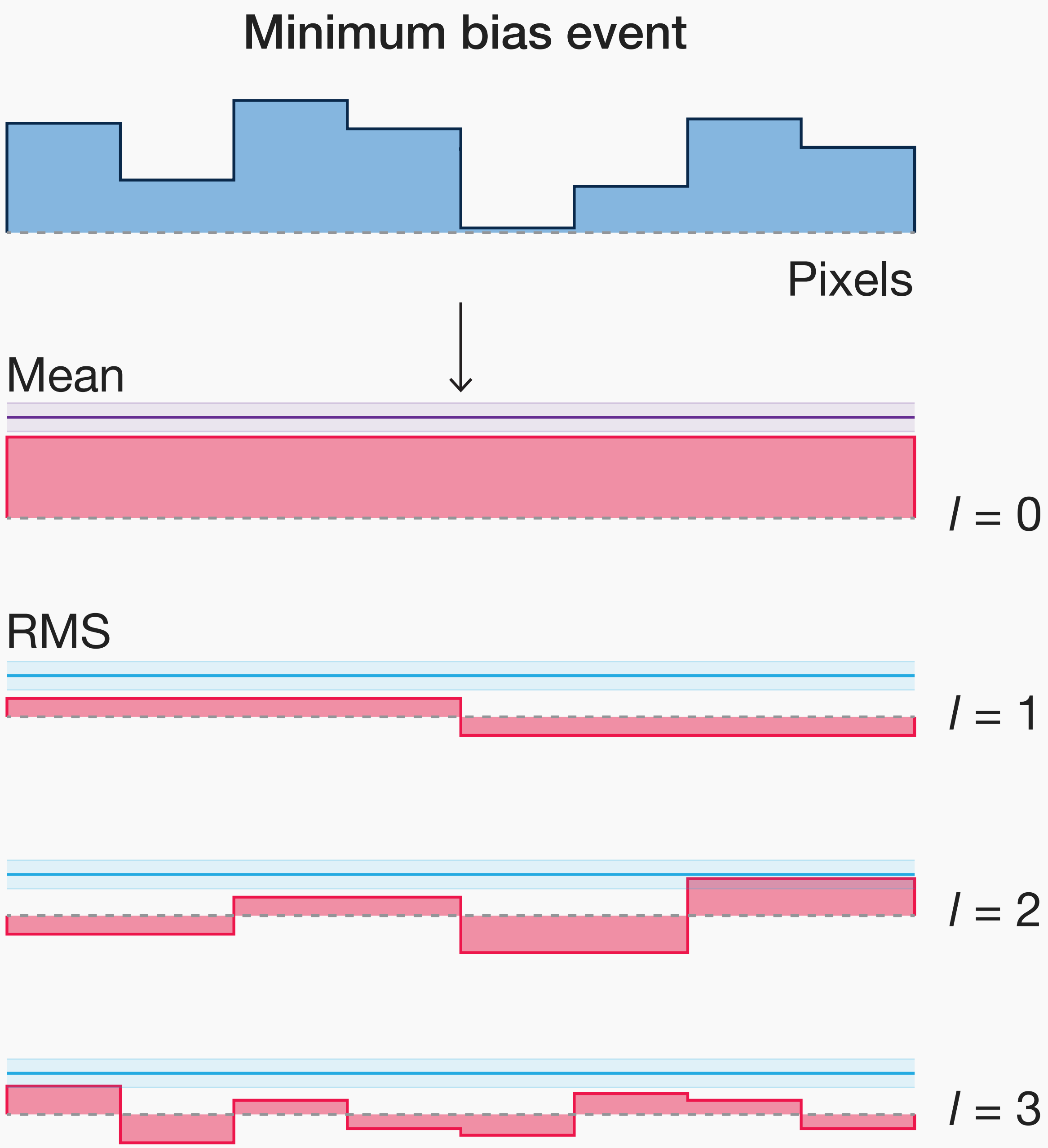
- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.



Methods for pile-up mitigation

Simplest approach:

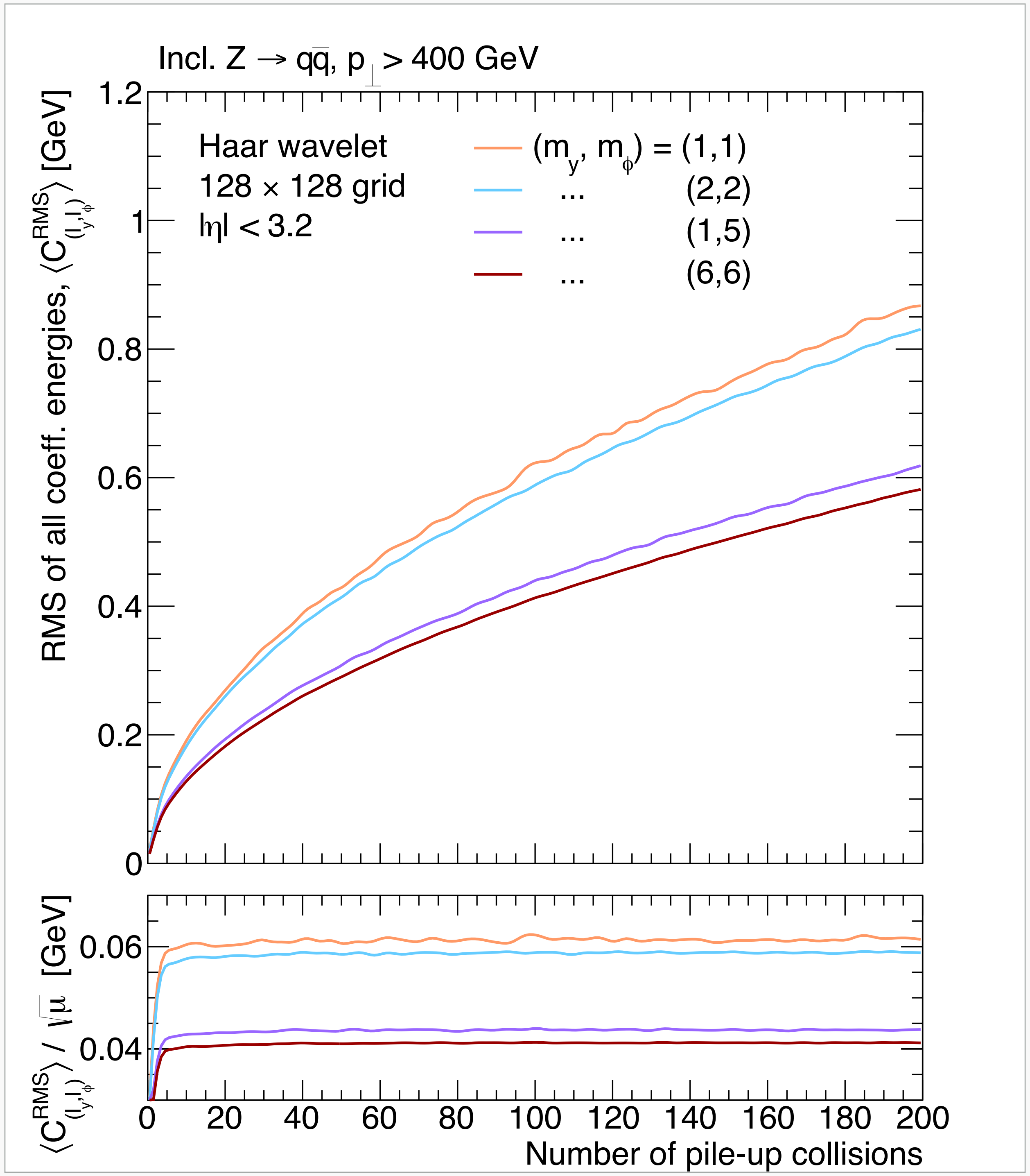
- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.
- E.g. n x $\mathrm{RMS}^{\mathrm{pileup}}$, $m > 0$



Methods for pile-up mitigation

Simplest approach:

- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.
- E.g. n x RMS$^{\text{pileup}}$, $m > 0$
- Per event: Scale with $\sqrt{\mu}$



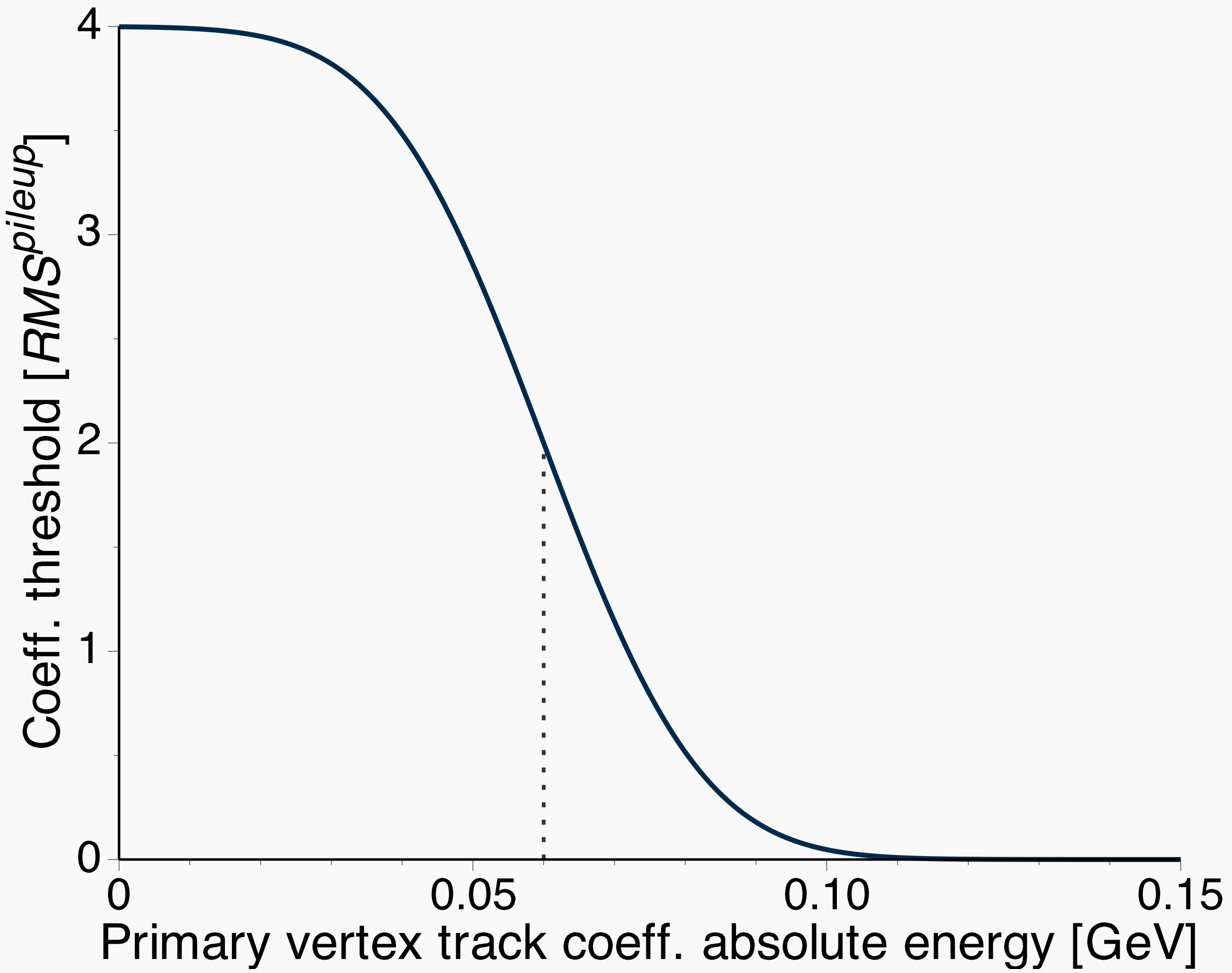
Methods for pile-up mitigation

Simplest approach:

- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.
- E.g. n x $\text{RMS}^{\text{pileup}}$, $m > 0$
- Per event: Scale with $\sqrt{\mu}$

Improvement:

- Use track information



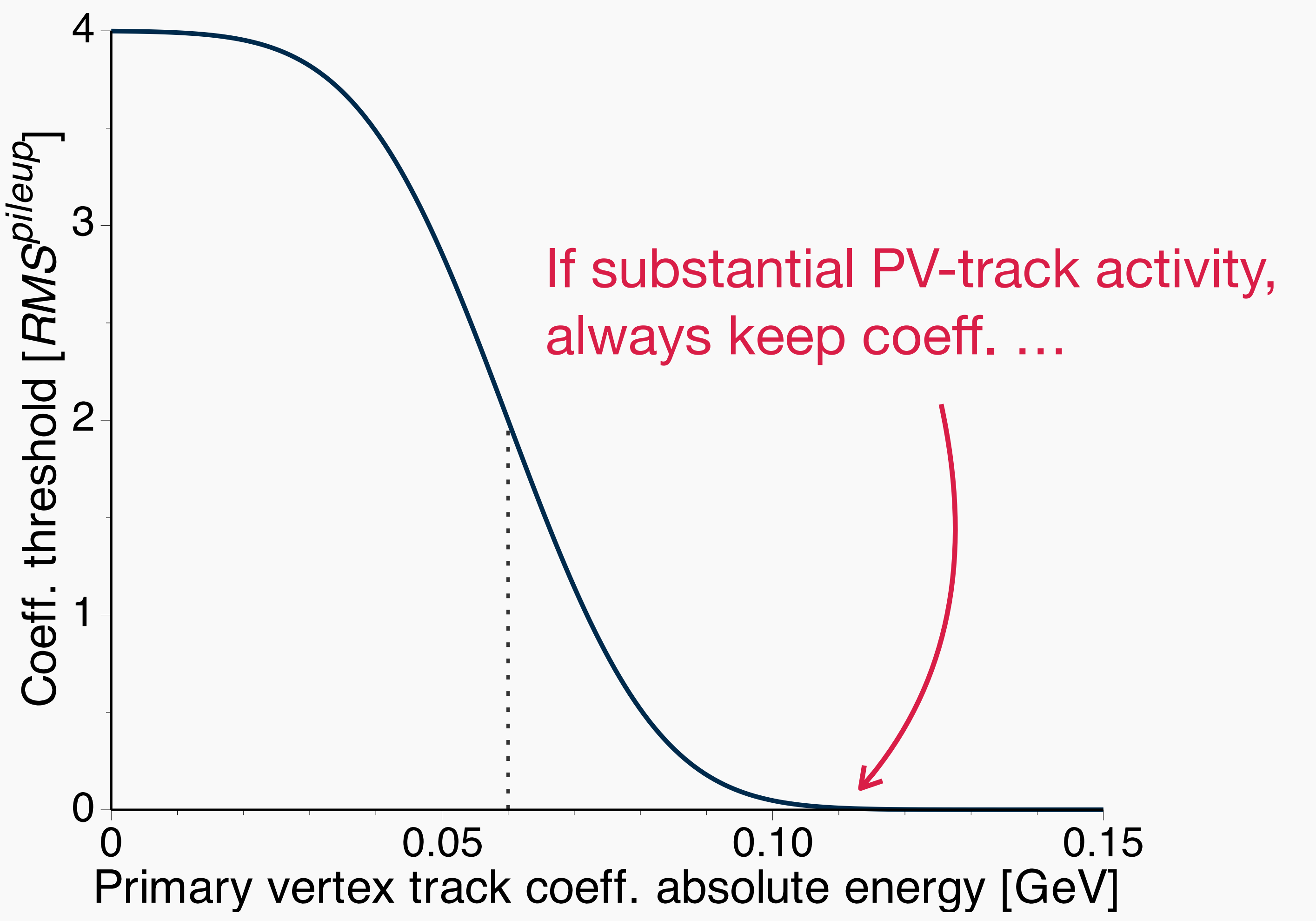
Methods for pile-up mitigation

Simplest approach:

- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.
- E.g. n x $\mathrm{RMS}^{\mathrm{pileup}}$, $m > 0$
- Per event: Scale with $\sqrt{\mu}$

Improvement:

- Use track information



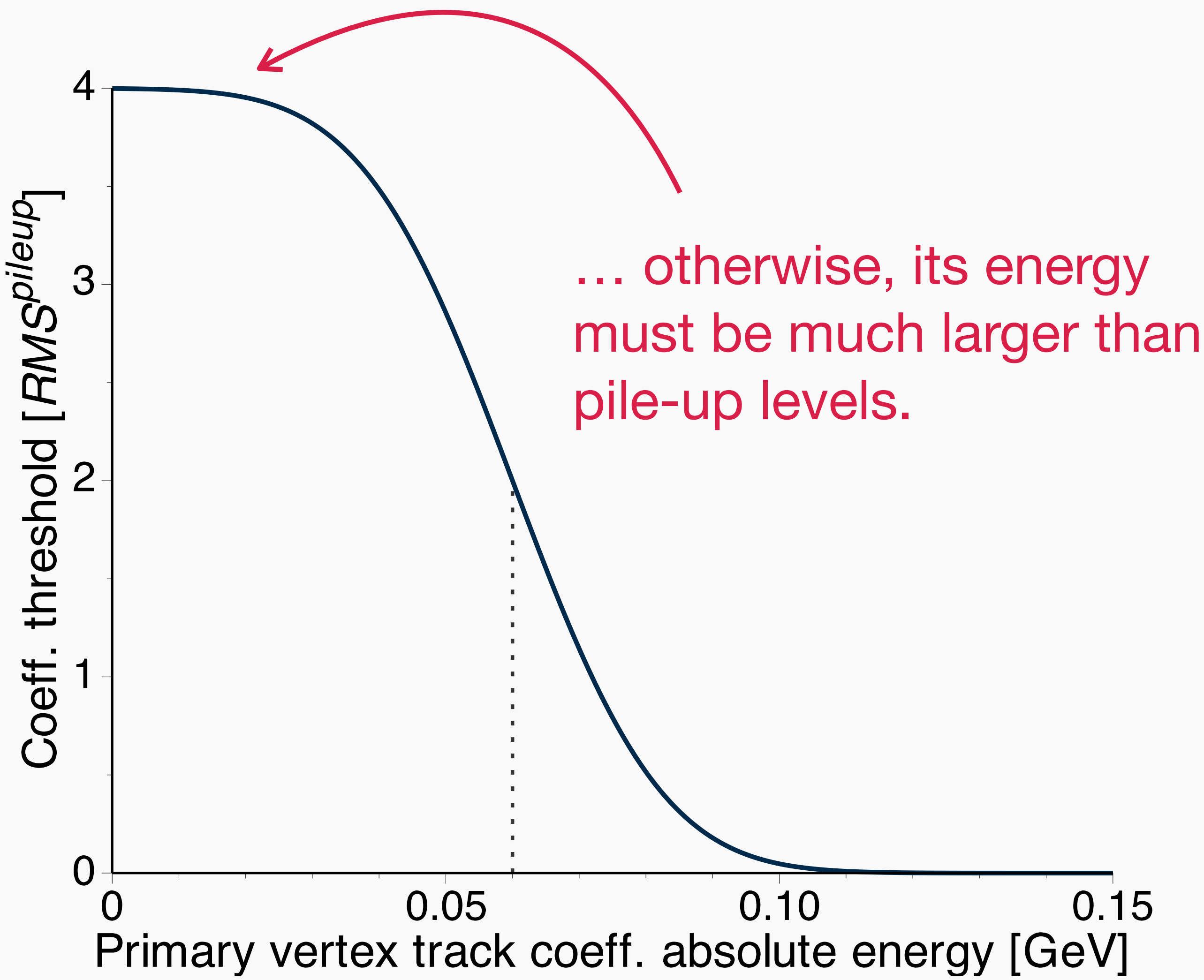
Methods for pile-up mitigation

Simplest approach:

- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.
- E.g. n x RMS^{pileup}, $m > 0$
- Per event: Scale with $\sqrt{\mu}$

Improvement:

- Use track information



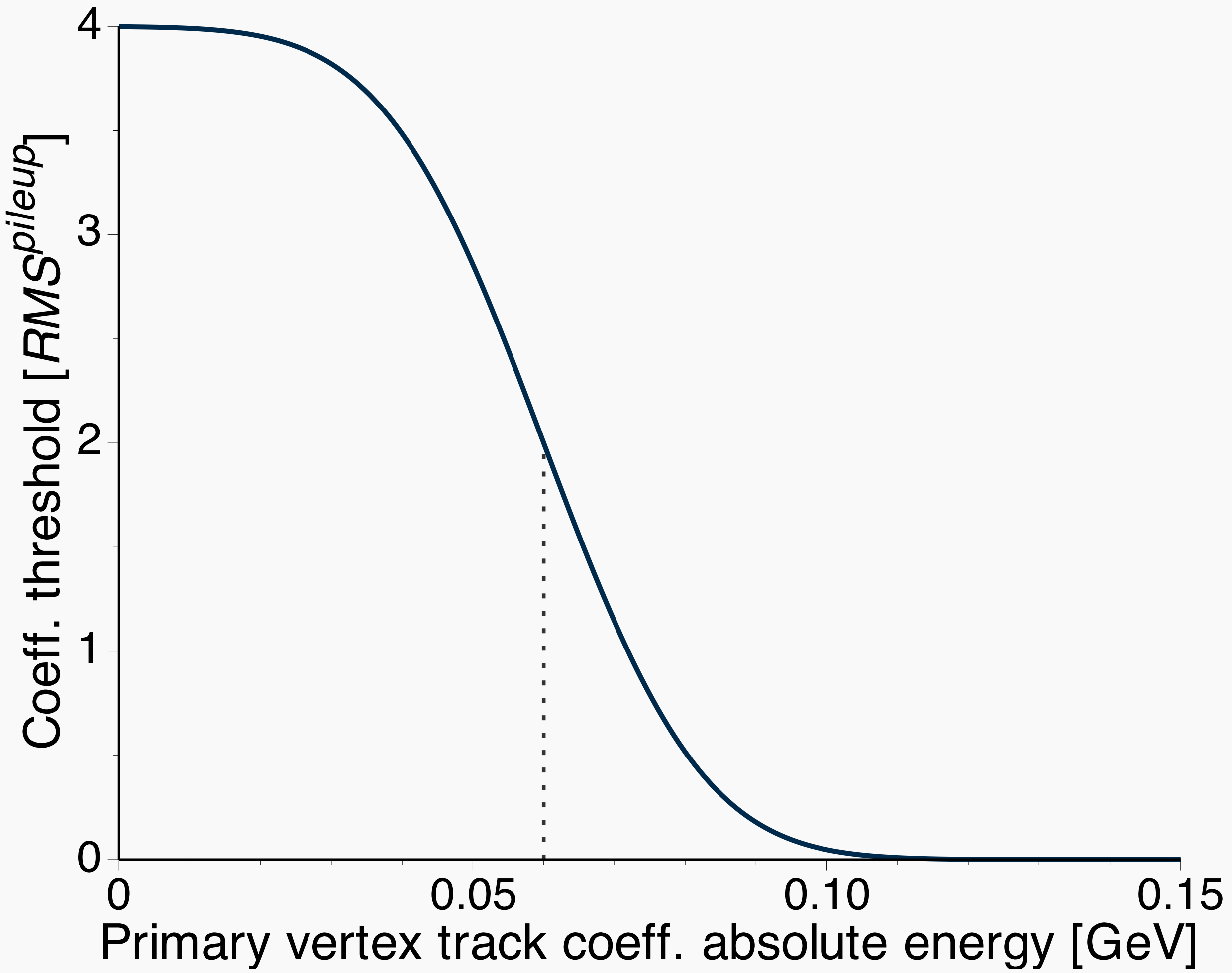
Methods for pile-up mitigation

Simplest approach:

- Scale (0,0) coefficient (average energy)
- Cut on other coeffs.
- E.g. n x $\text{RMS}^{\text{pileup}}$, $m > 0$
- Per event: Scale with $\sqrt{\mu}$

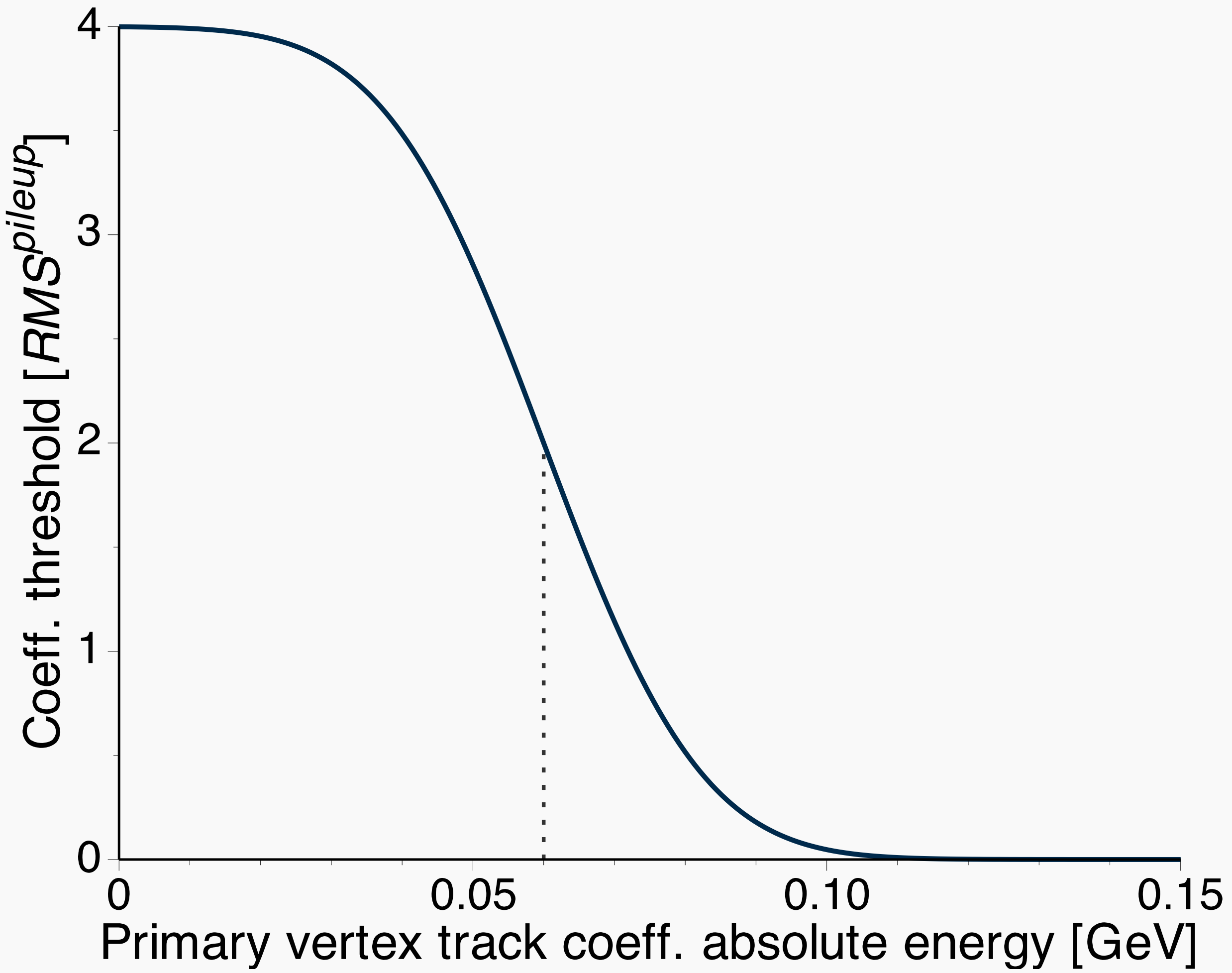
Improvement:

- Use track information
- 'Wavelet onset'



Wavelet setup

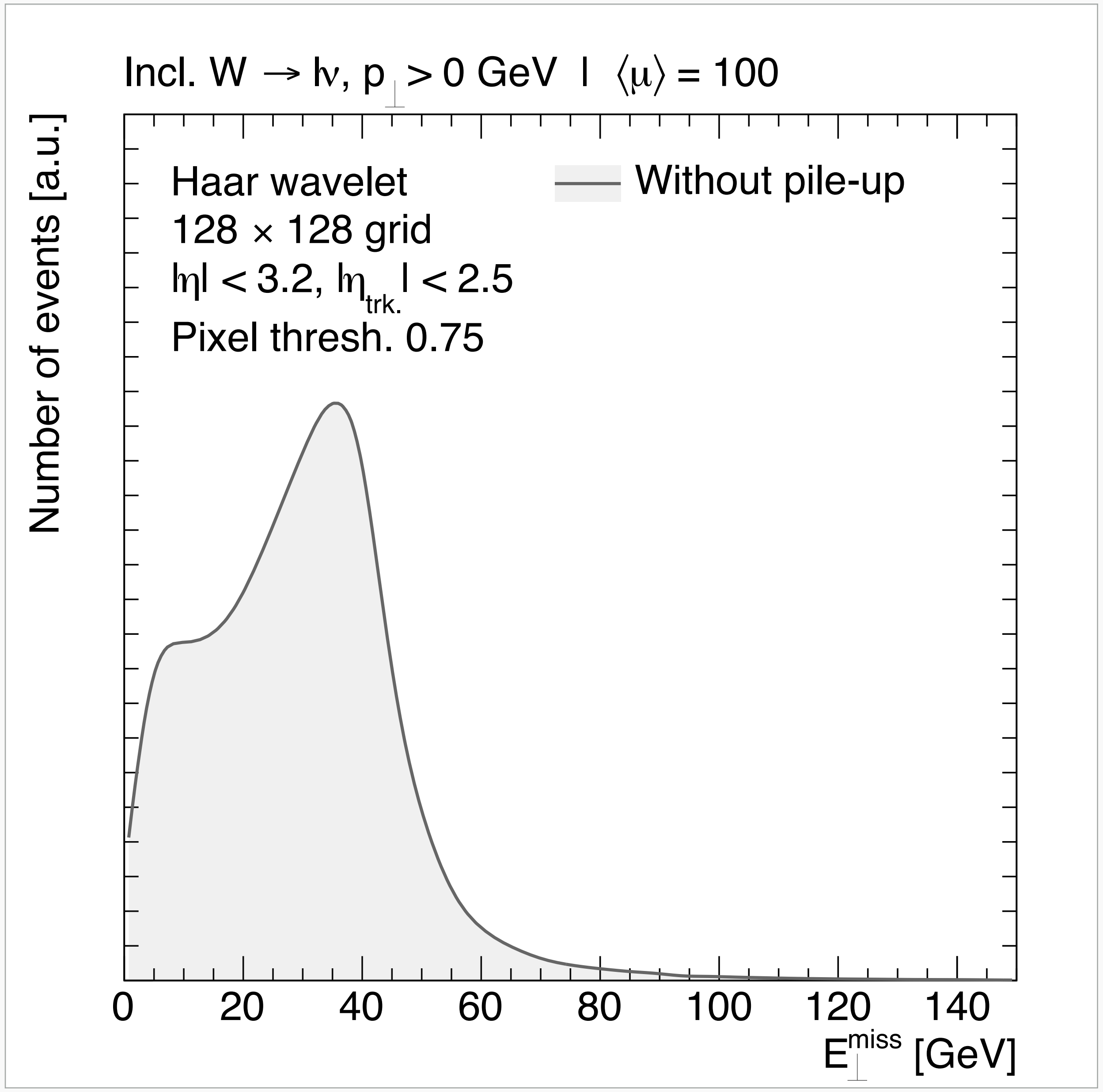
- 128 x 128 pixel grid
- |y| < 3.2 (square)
- Haar wavelet (simplest)
- ‘Wavelet onset’ method with max. of 4 x pileup RMS
- Keep pixels with final-to-initial ratio > 0.75

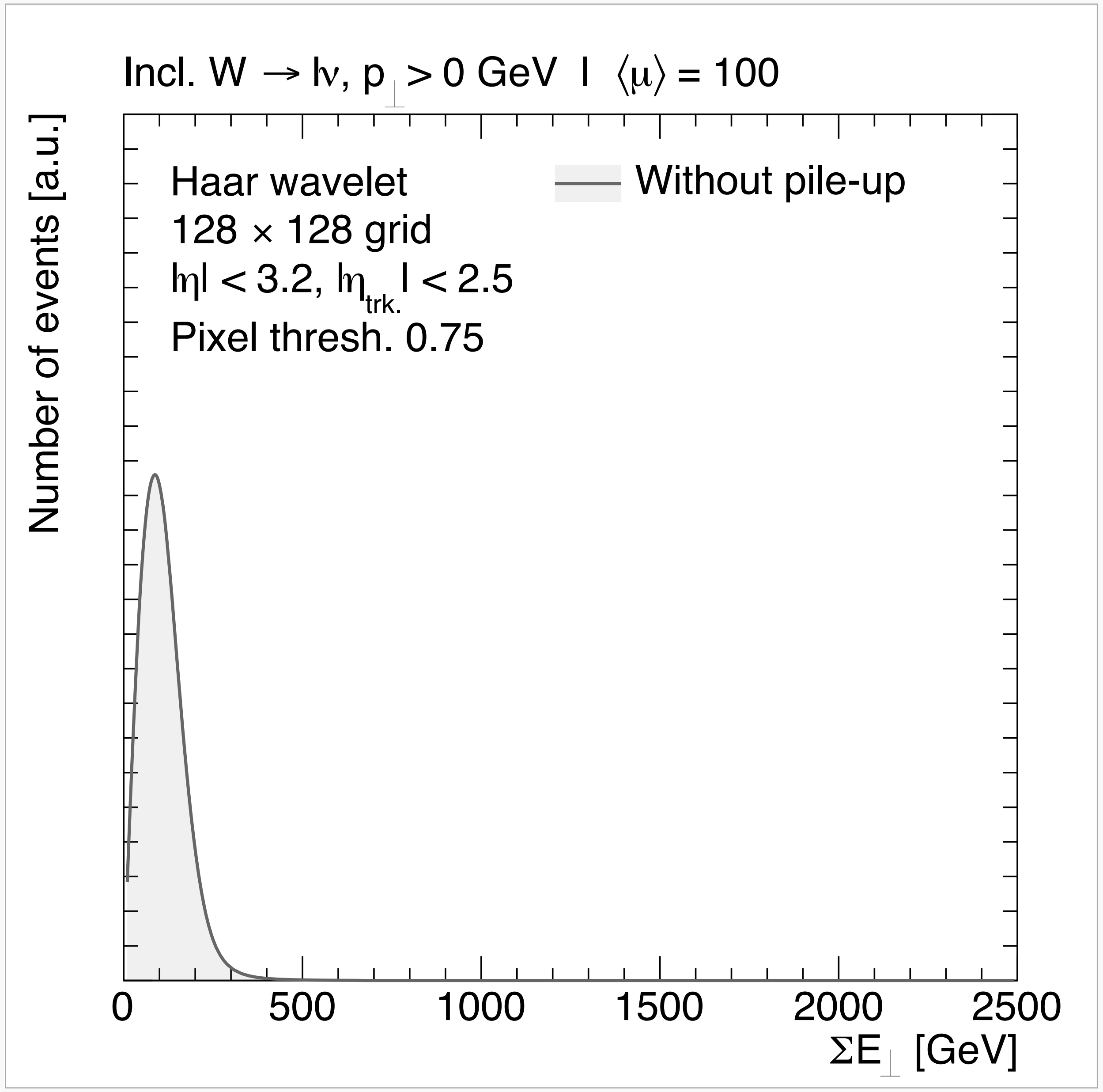


Samples and setup

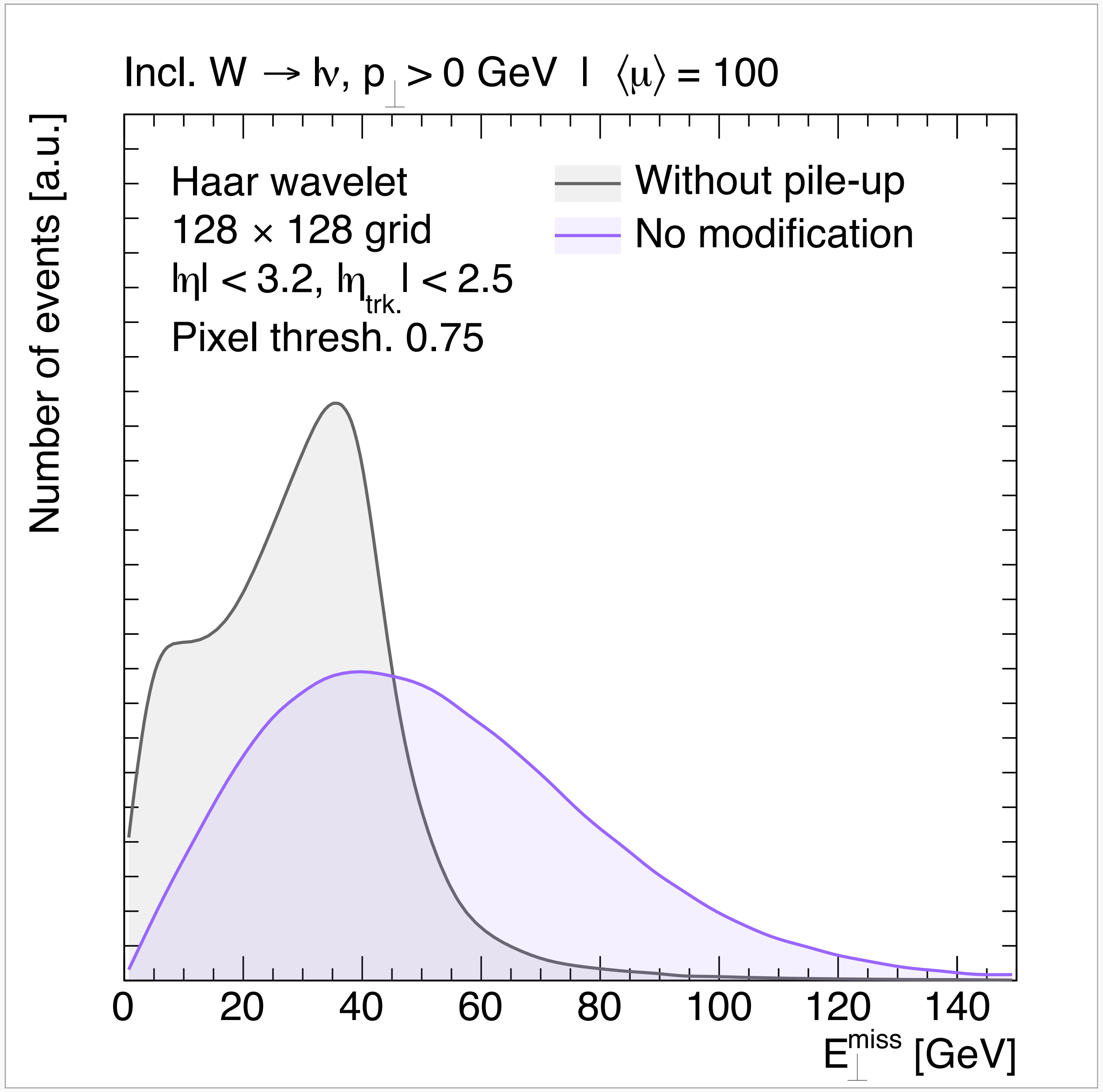
- Selection:
  - Stable, visible final state particles
  - All particles $|\eta| < 3.2$, tracks $|\eta| < 2.5$
  - $p_T > 500$ MeV
  - 100% tracking and vertex matching efficiency
- Signal samples (13 TeV, PYTHIA 8.205, A14-NNPDF23LO tune):
  - $W \rightarrow l\nu$ – no gen. (reco.) level $\hat{p}_T$ (p_T) cut
  - Incl. $Z \rightarrow q\bar{q}$ – $\hat{p}_T$ (p_T) > 280 (400) GeV
  - Incl. QCD $2 \rightarrow 2$ multijets – $\hat{p}_T$ (p_T) > 280 (400) GeV
- Minimum bias samples: PYTHIA 8.205, A2-MSTW (MB) tune:
  - Overlaid using PILEMC.

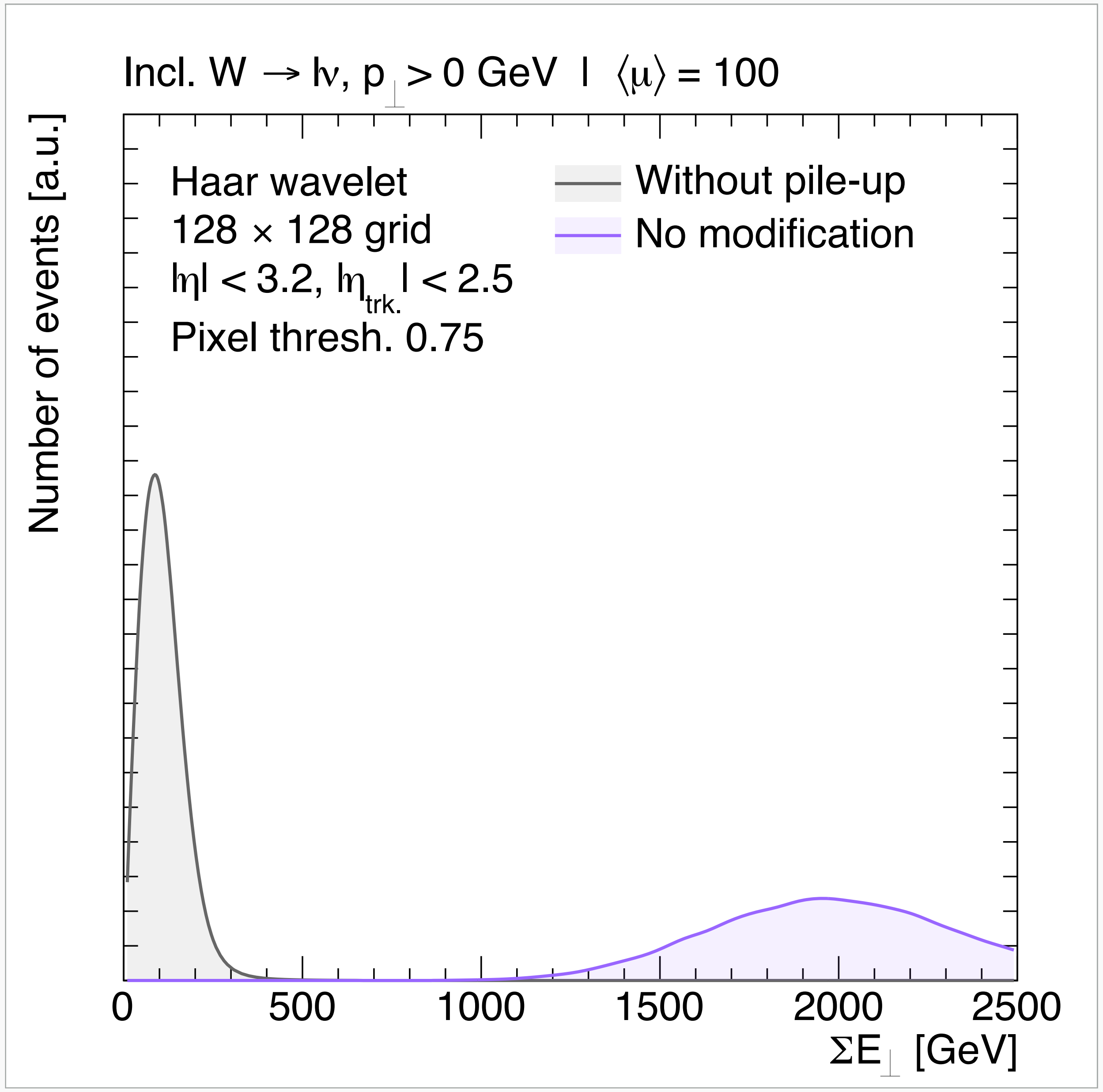
Missing and sum E_T



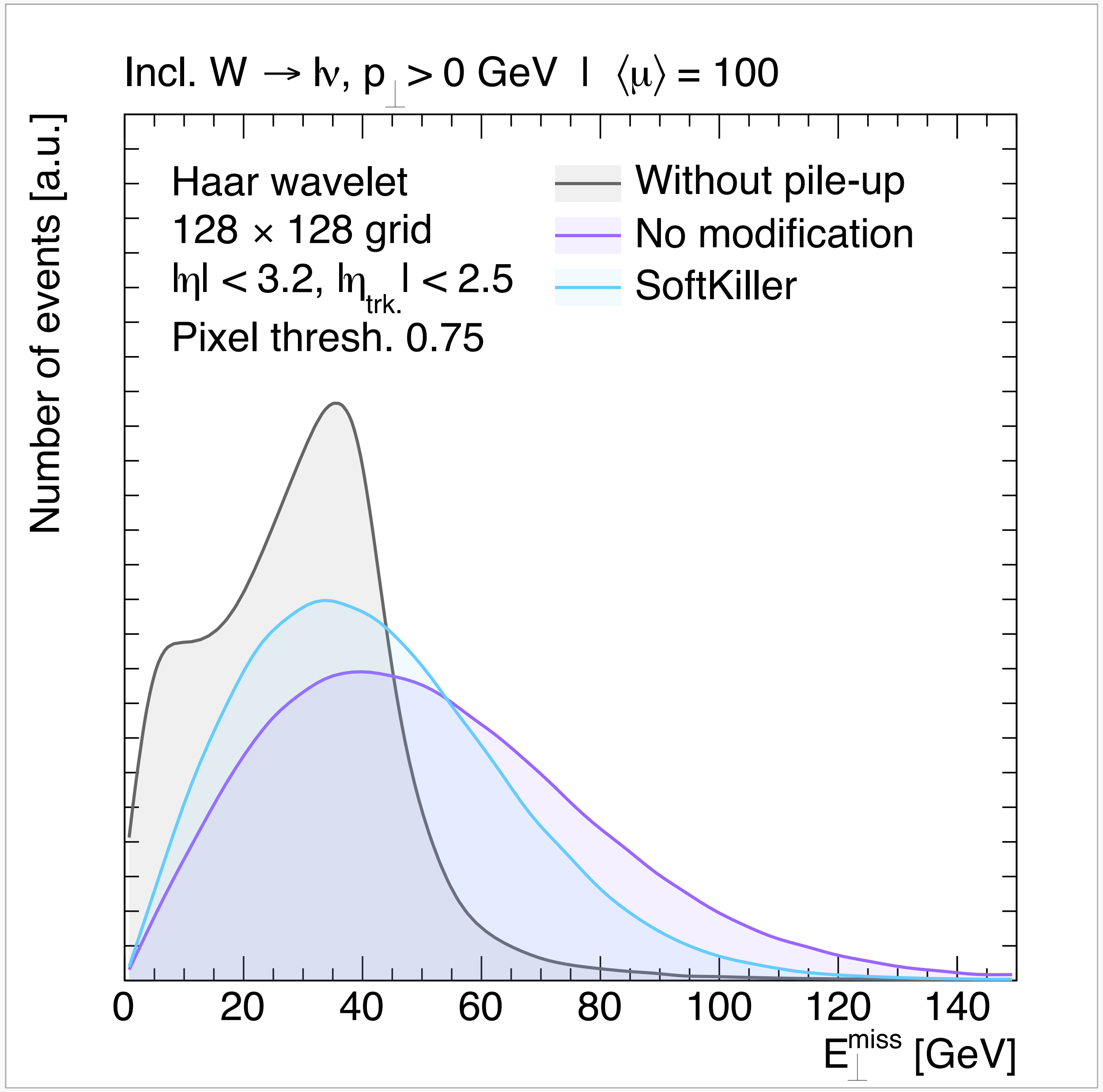


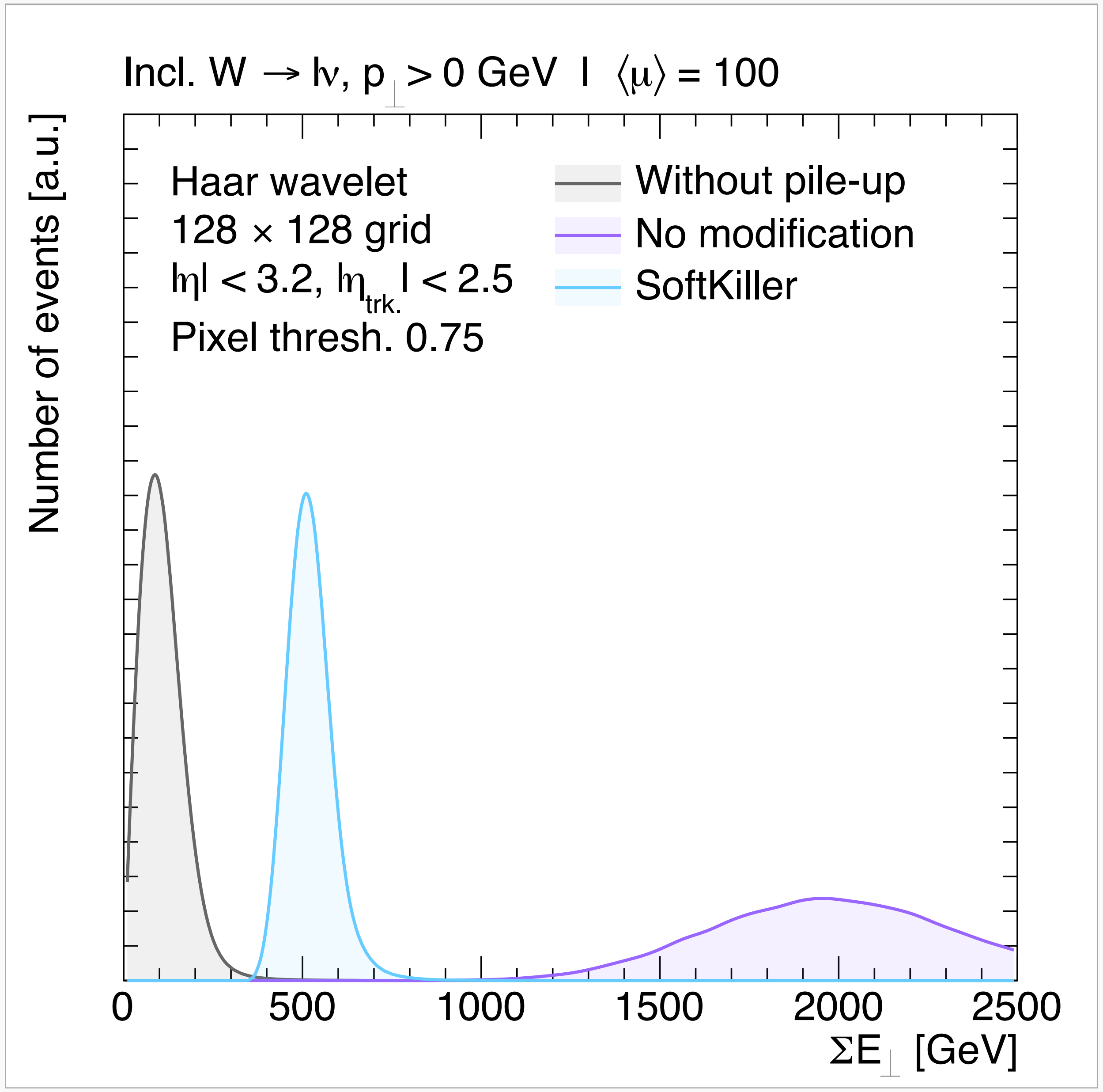
Missing and sum E_T



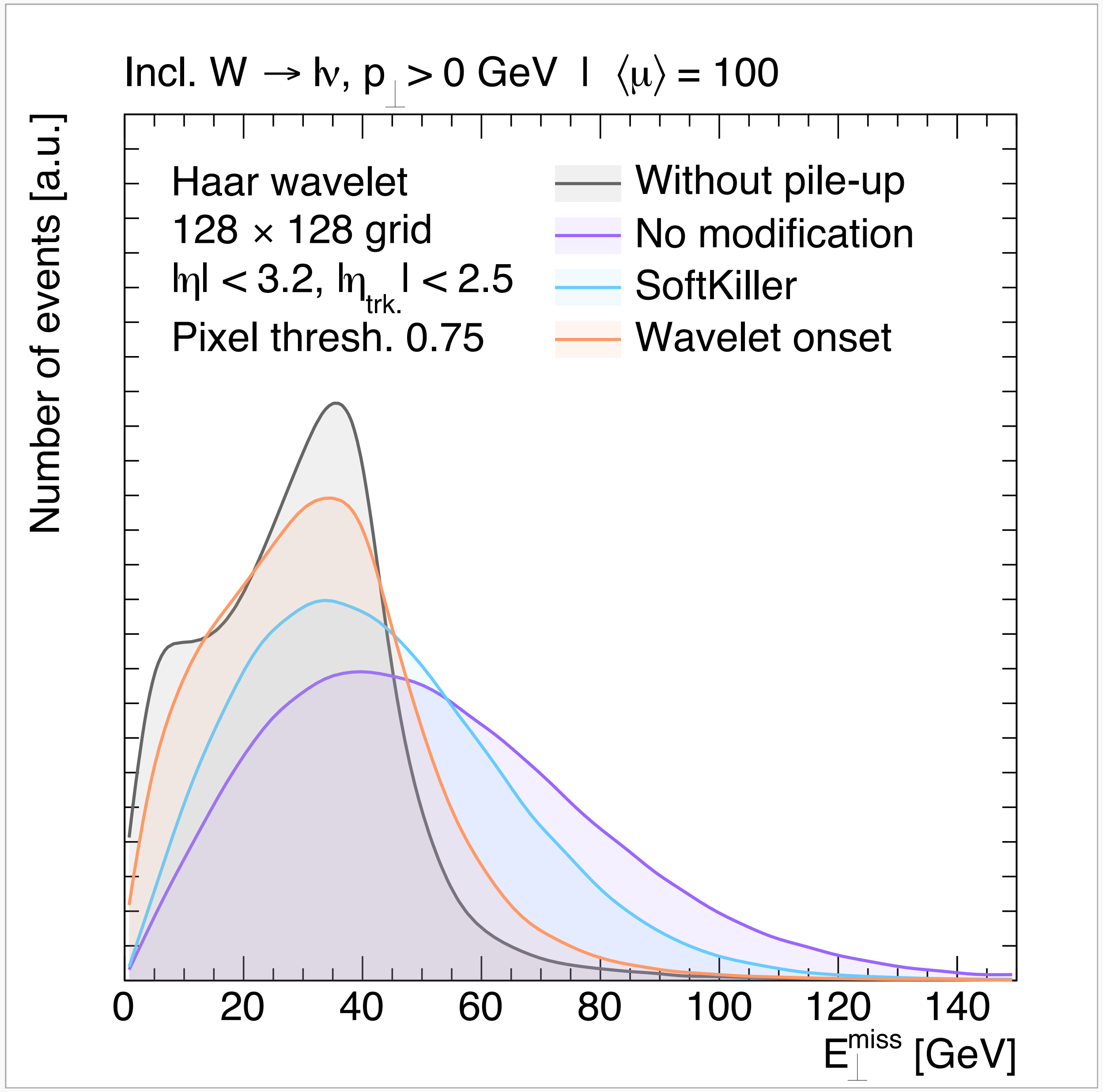


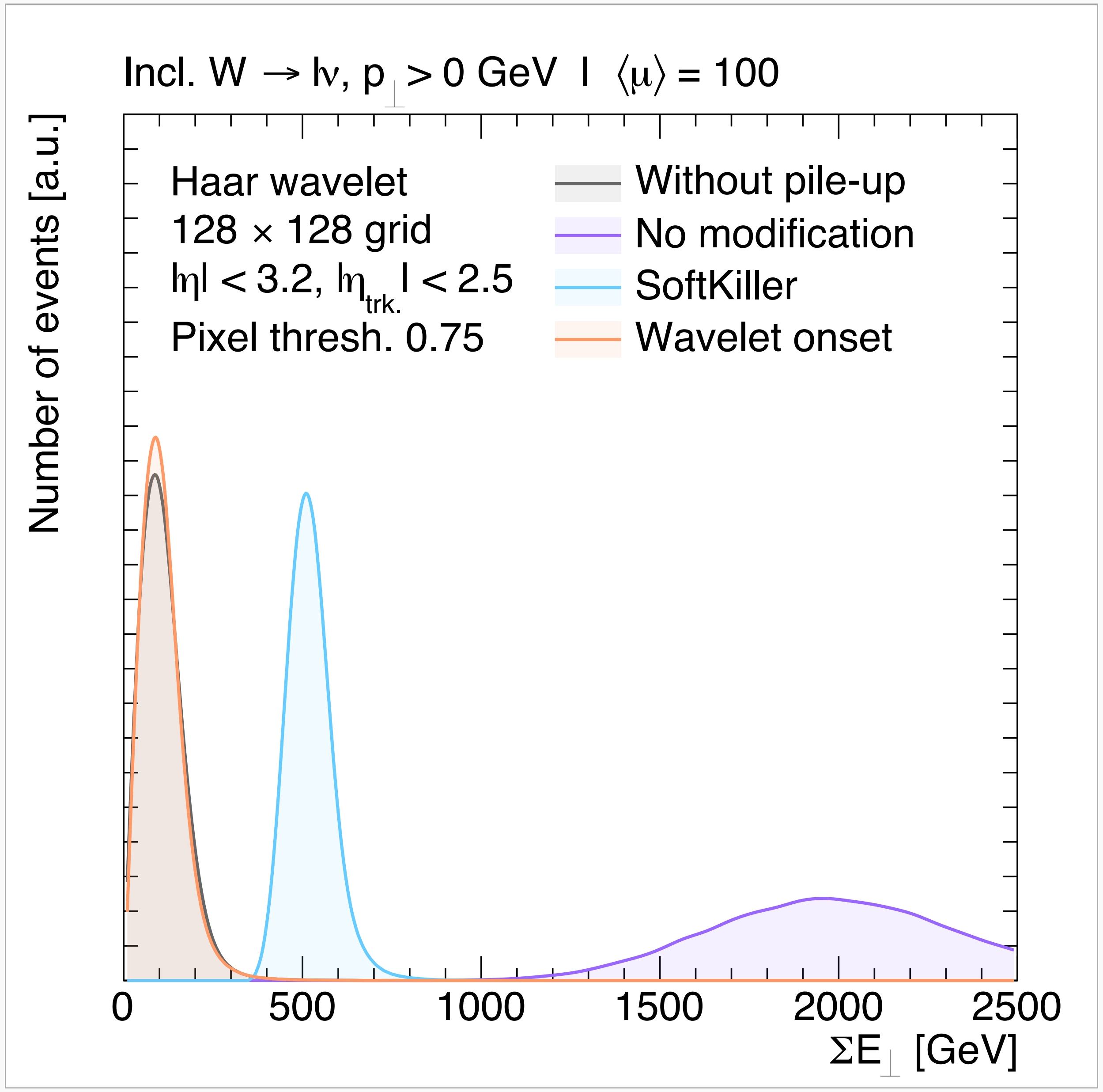
Missing and sum E_T



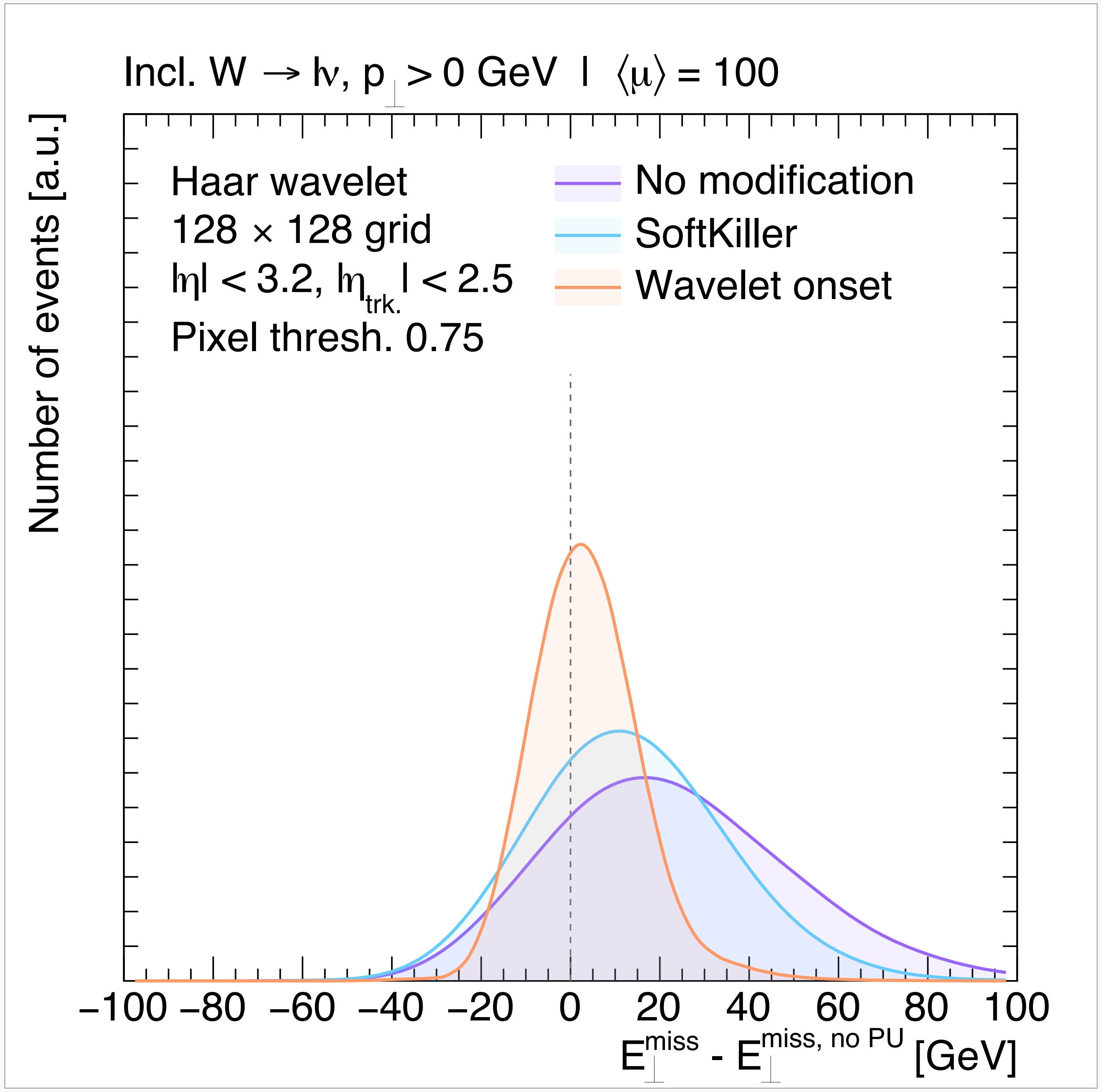


Missing and sum E_T

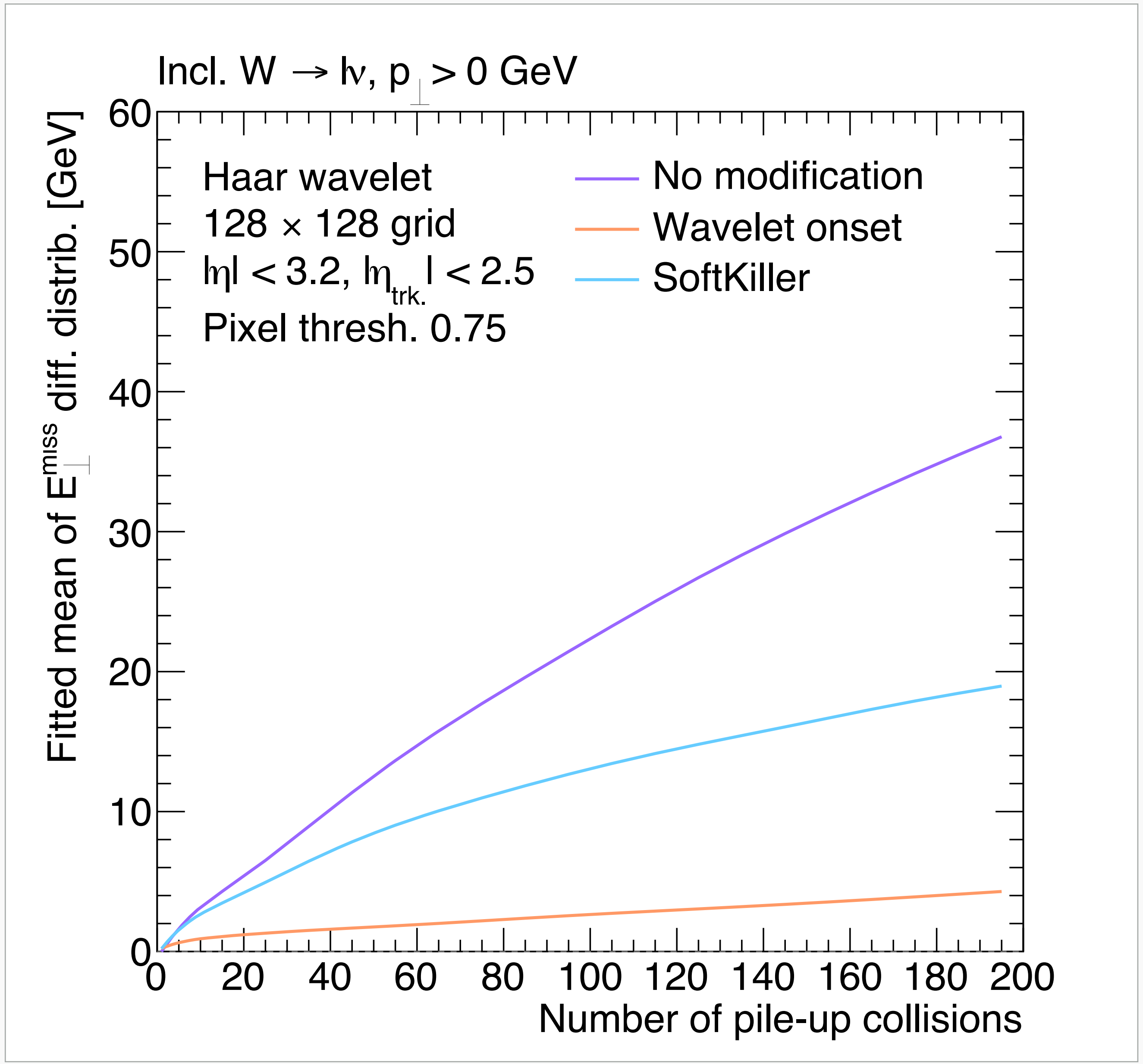


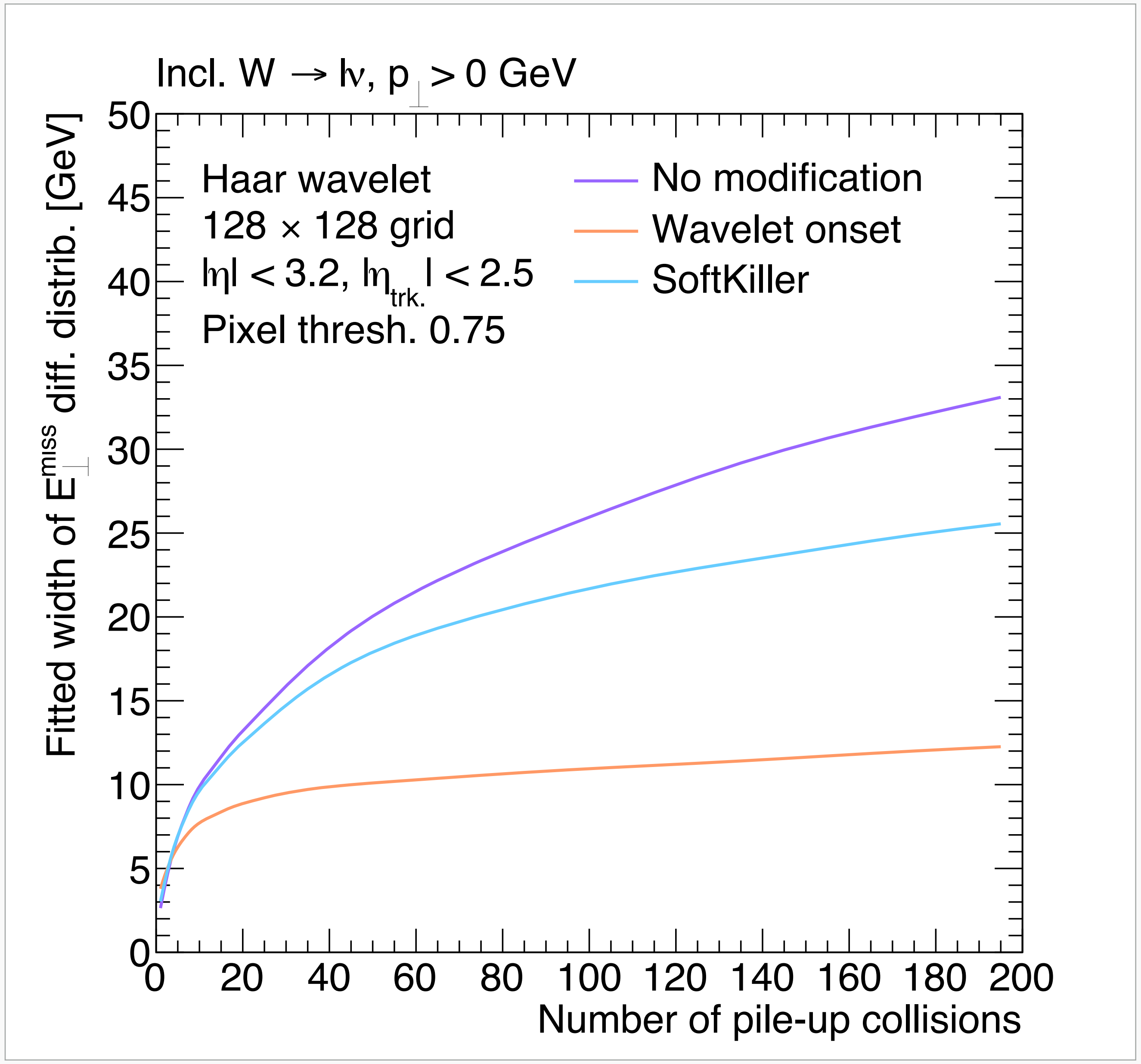


Missing E_T-resolution

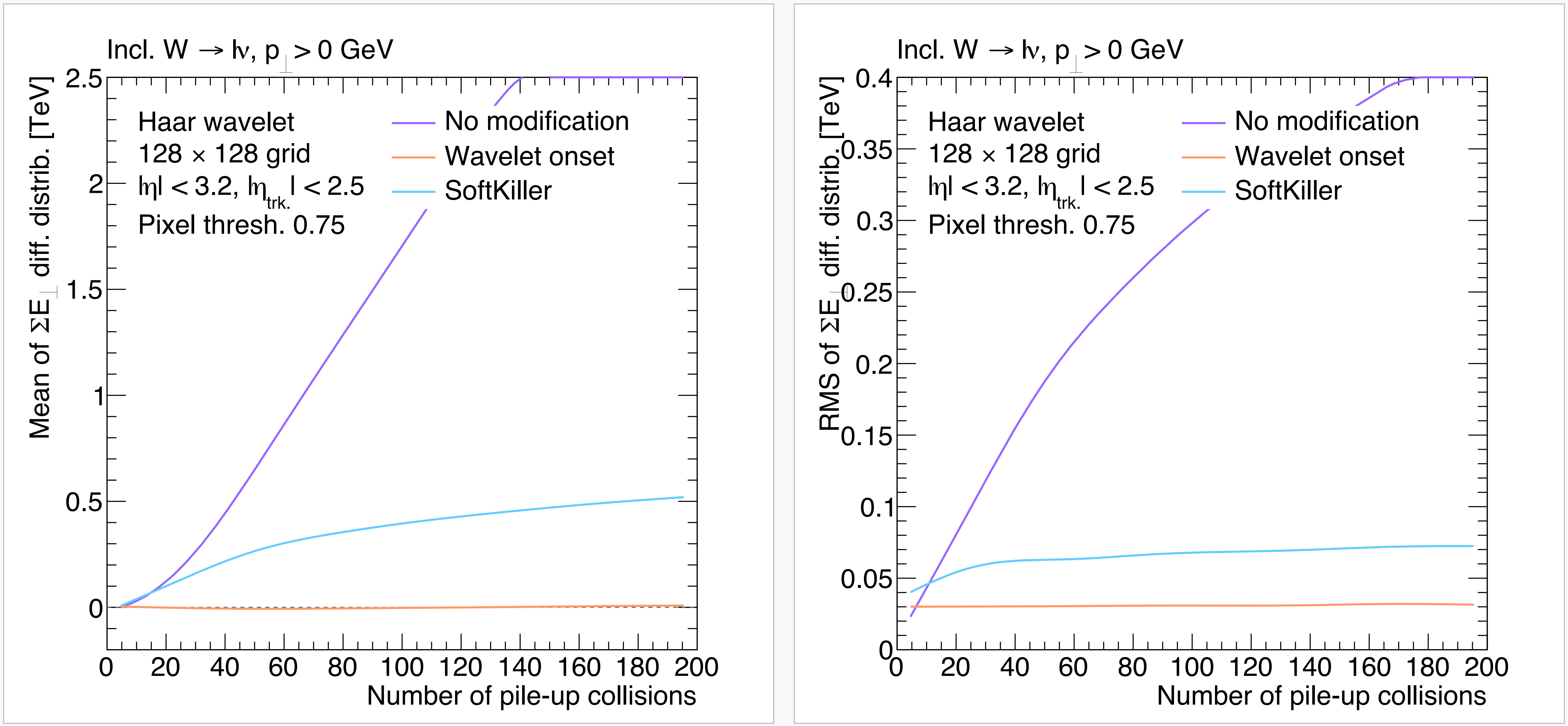


μ-dependence · Missing E_T-resolution

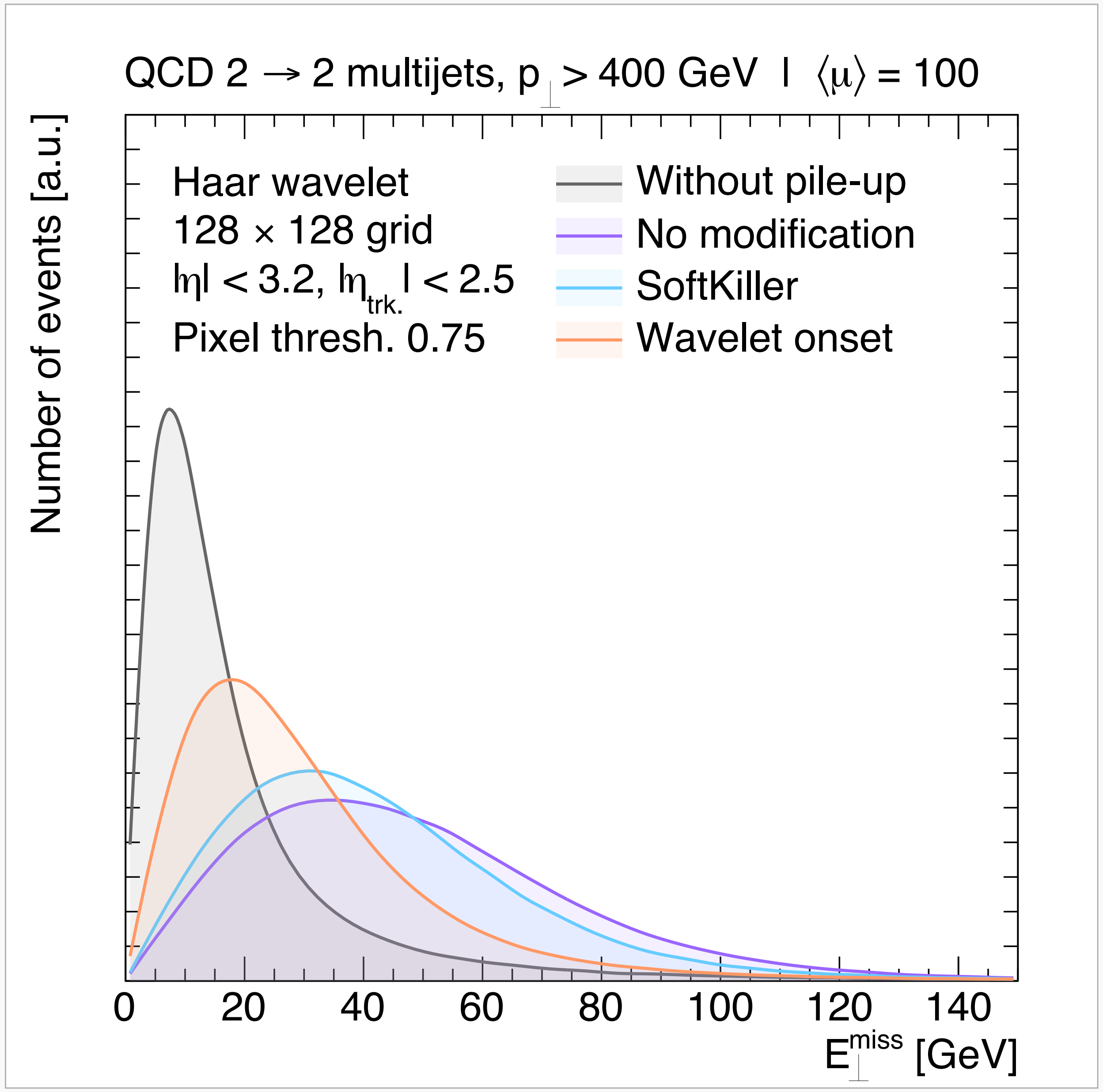


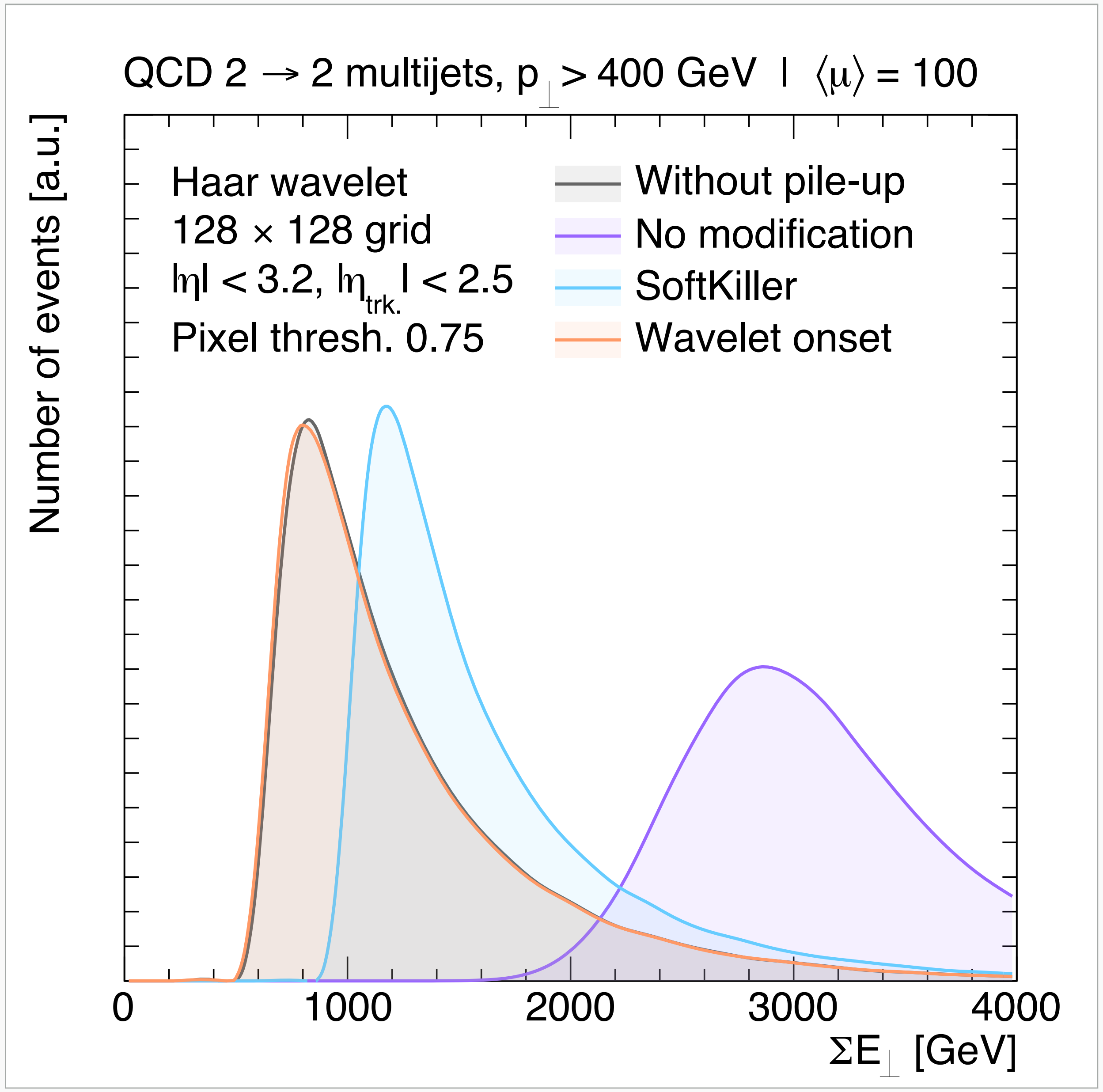


μ-dependence · Sum E_T-resolution



Missing and sum E_T-resolution

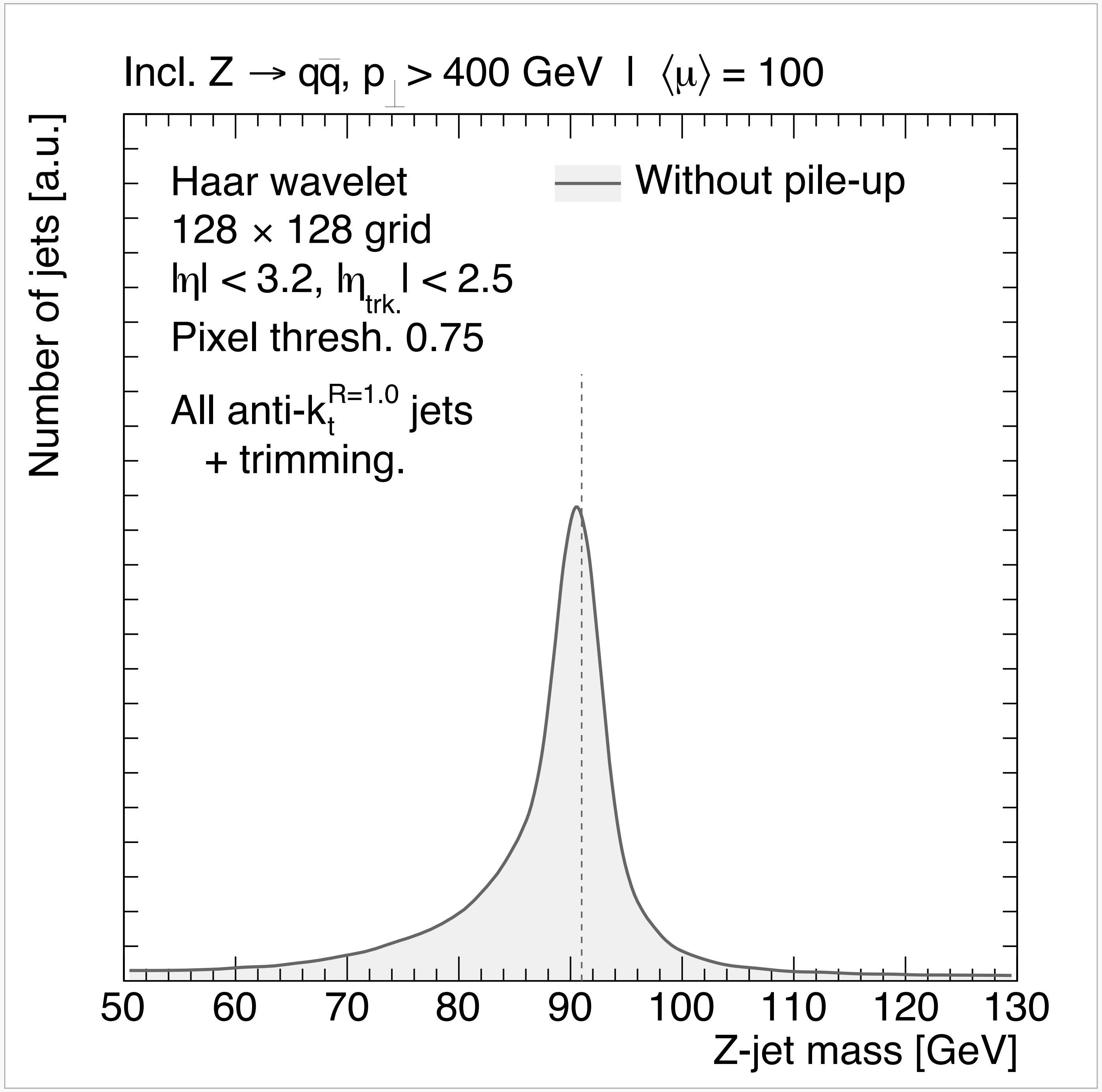




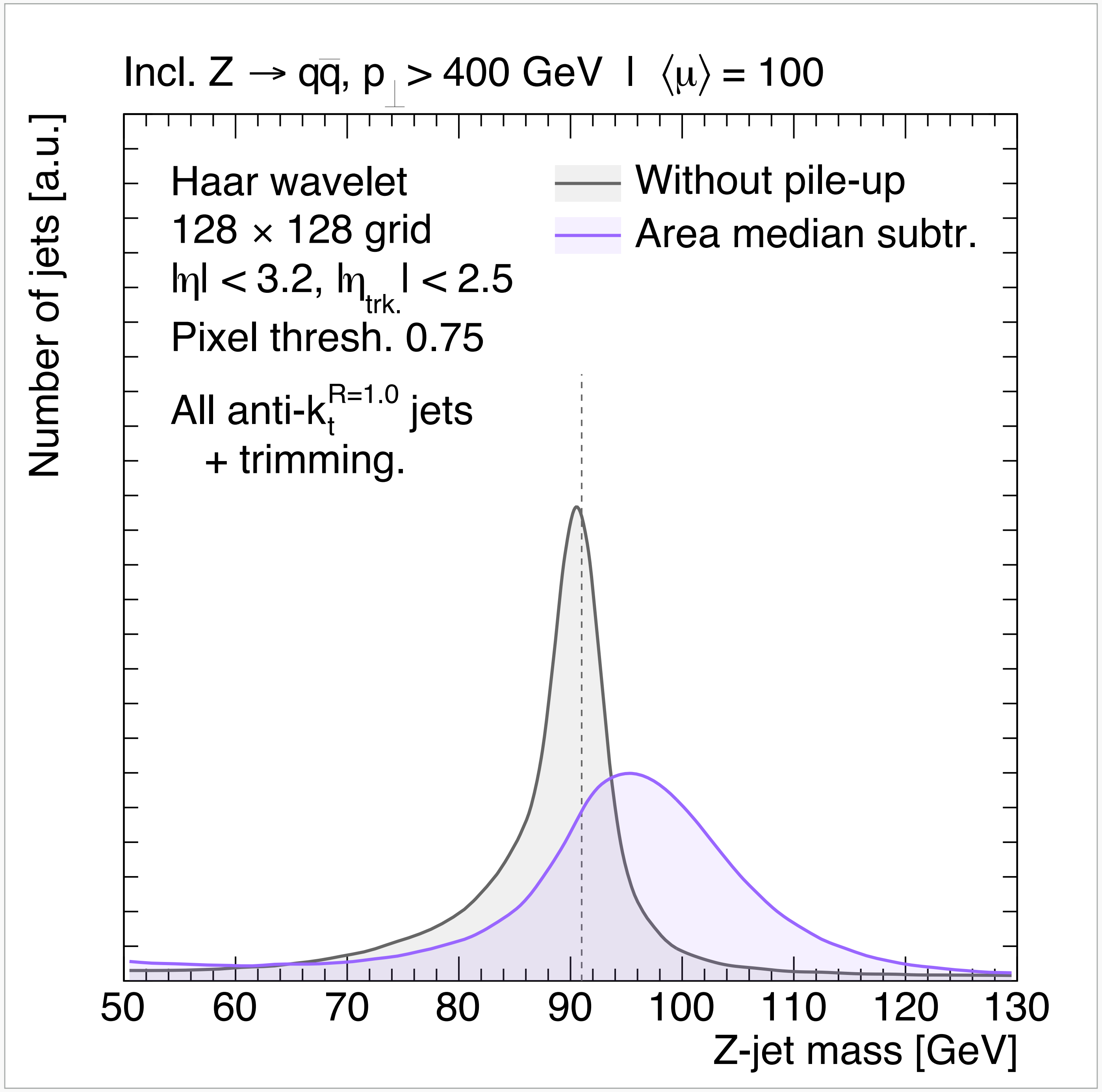
Fat jet studies

- Samples:
  - Incl. $Z \rightarrow q\bar{q}$ and QCD $2 \rightarrow 2$ multijet, $p_T > 400$ GeV
- Jets:
  - Anti-$k_T^{R\,=\,1.0}$ jets clustered with FASTJET
  - Trimming, using $k_T^{R\,=\,0.2}$ jets and p_T-fraction 0.05
  - 'Z-jet': Highest-p_T jet within dR = 0.6 of truth-level Z boson
  - 'Leading jet': Highest-p_T jet
- Comparison:
  - Area median subtraction and SoftKiller

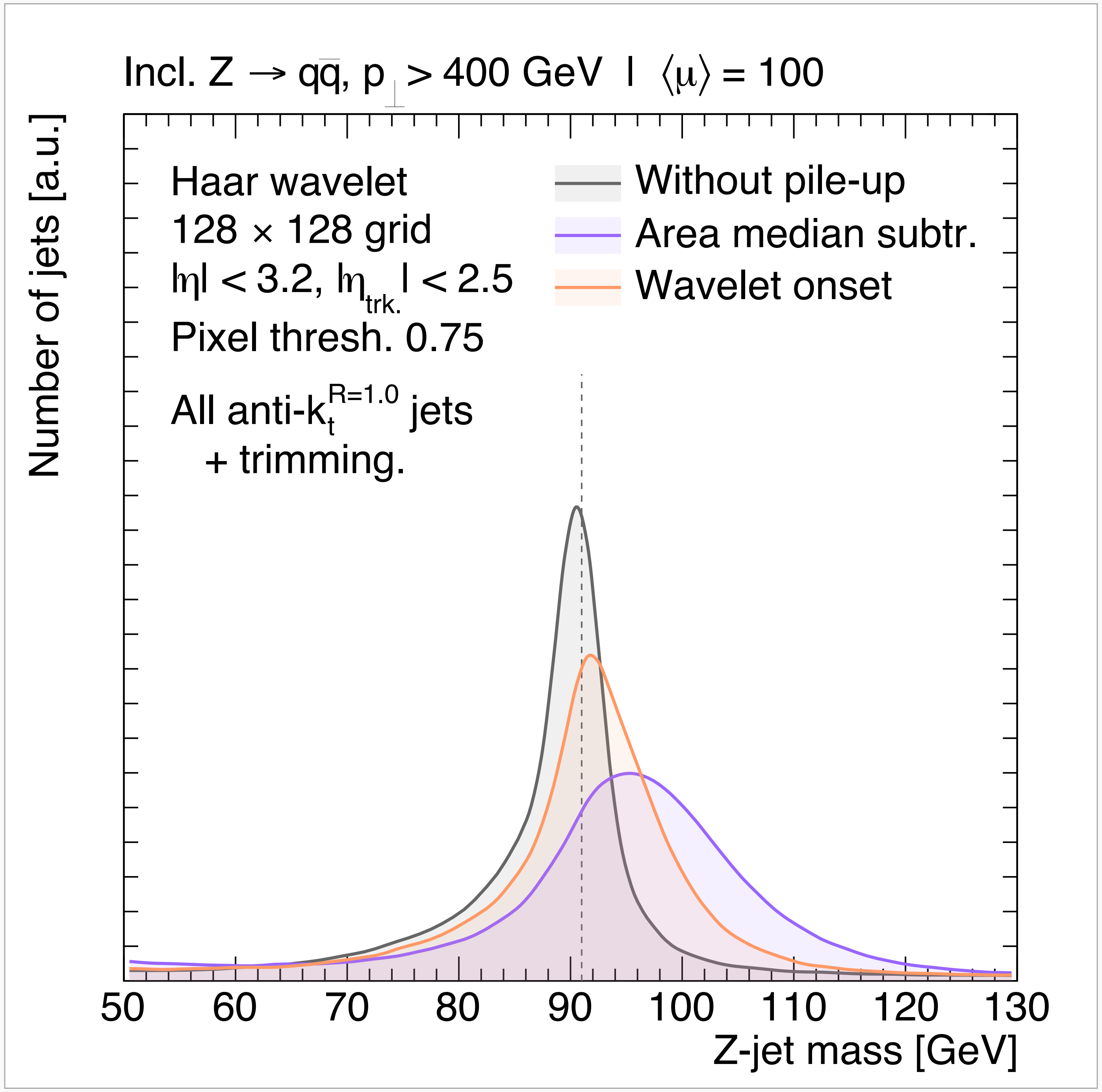
Z-jet mass


Incl. Z → qq̄, p⊥ > 400 GeV | ⟨μ⟩ = 100
Haar wavelet
128 × 128 grid
|η| < 3.2, |ηtrk.| < 2.5
Pixel thresh. 0.75
All anti-kt R=1.0 jets
+ trimming.
Without pile-up
Number of jets [a.u.]
50
60
70
80
90
100
110
120
130
Z-jet mass [GeV]

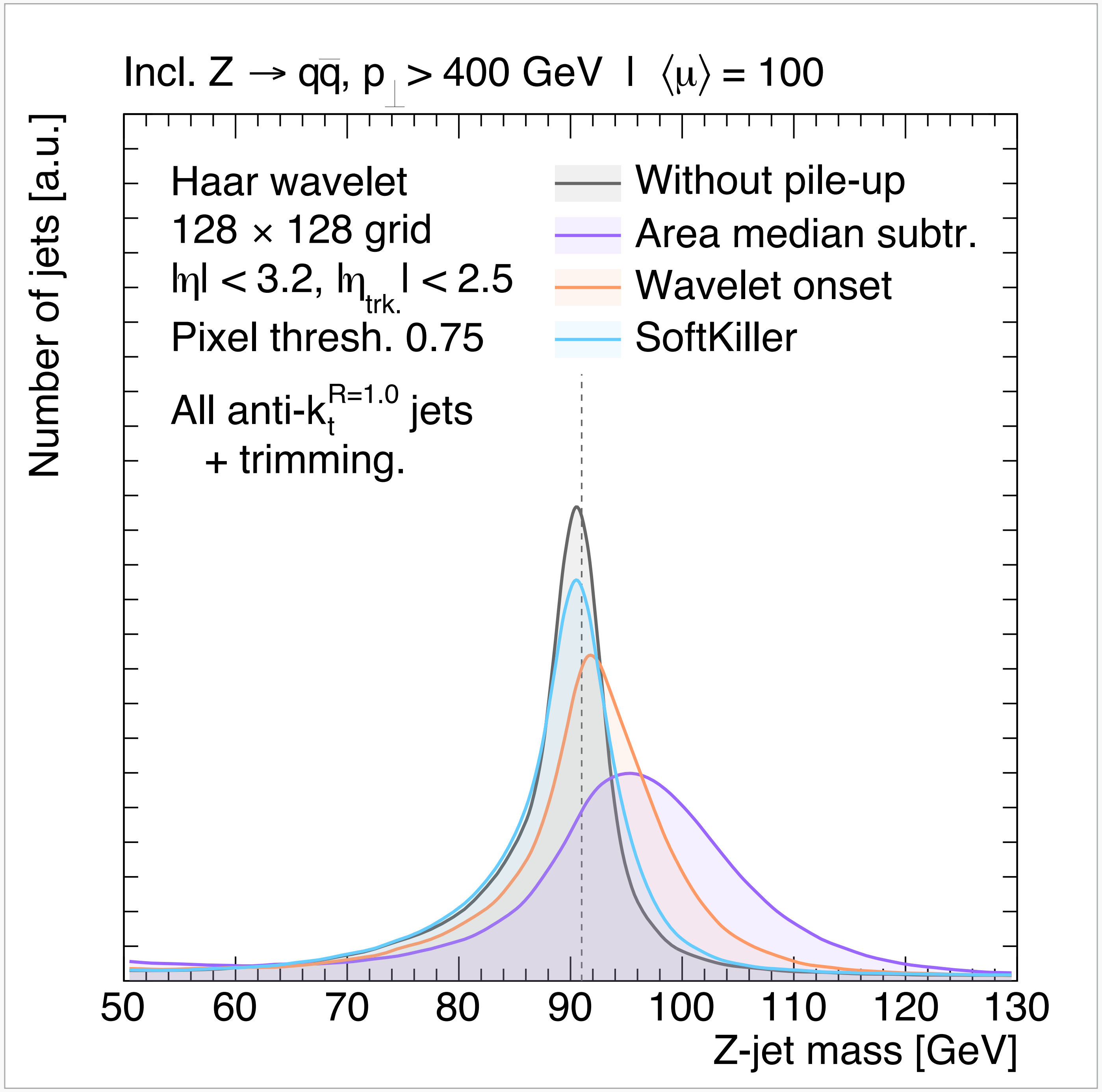
Z-jet mass


Incl. Z → q$\bar{q}$, $p_\perp$ > 400 GeV | ⟨μ⟩ = 100
Haar wavelet
128 × 128 grid
|η| < 3.2, |η$_{trk.}$| < 2.5
Pixel thresh. 0.75
All anti-$k_t^{R=1.0}$ jets
+ trimming.
Without pile-up
Area median subtr.
Number of jets [a.u.]
50
60
70
80
90
100
110
120
130
Z-jet mass [GeV]

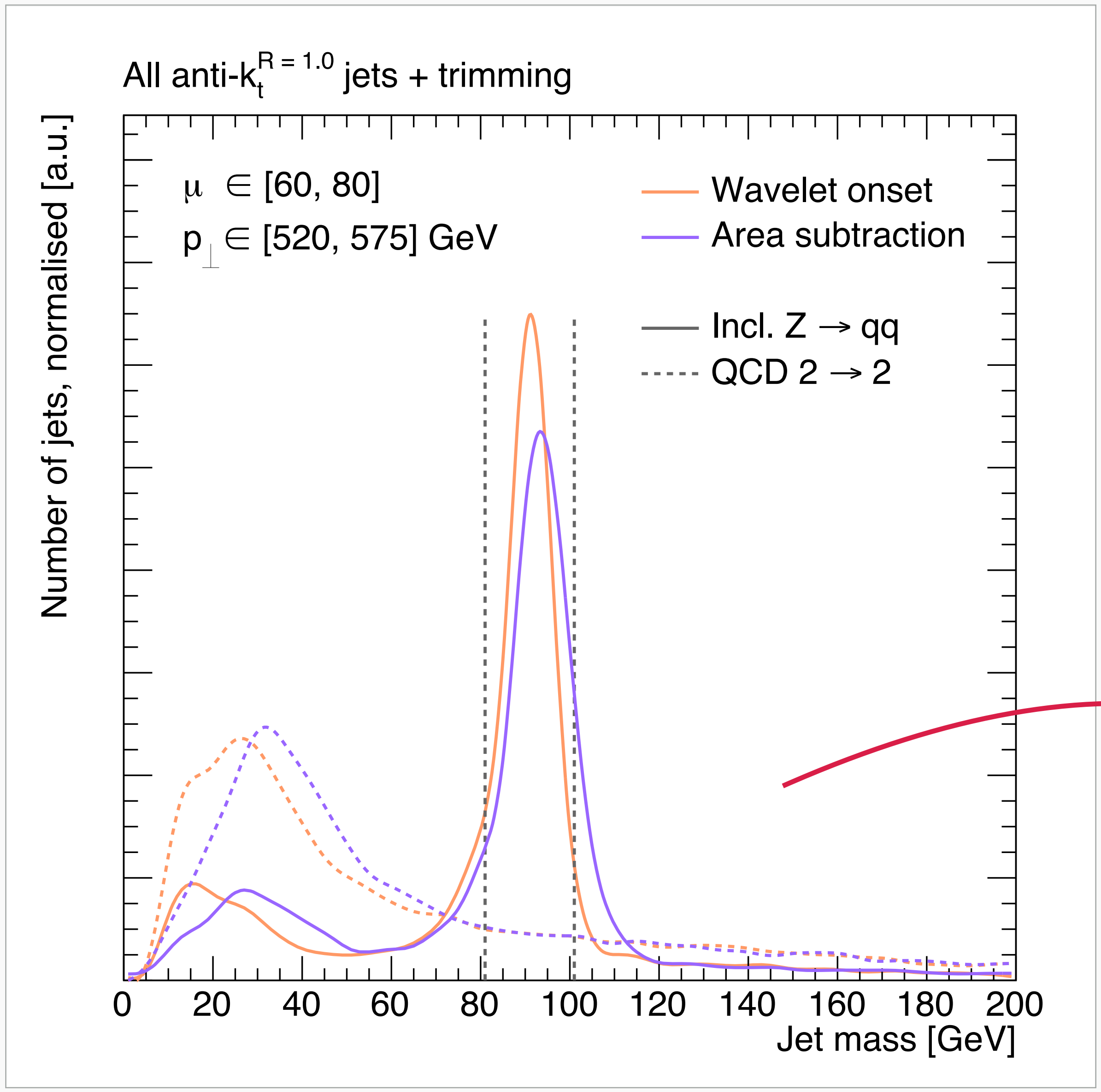
Z-jet mass

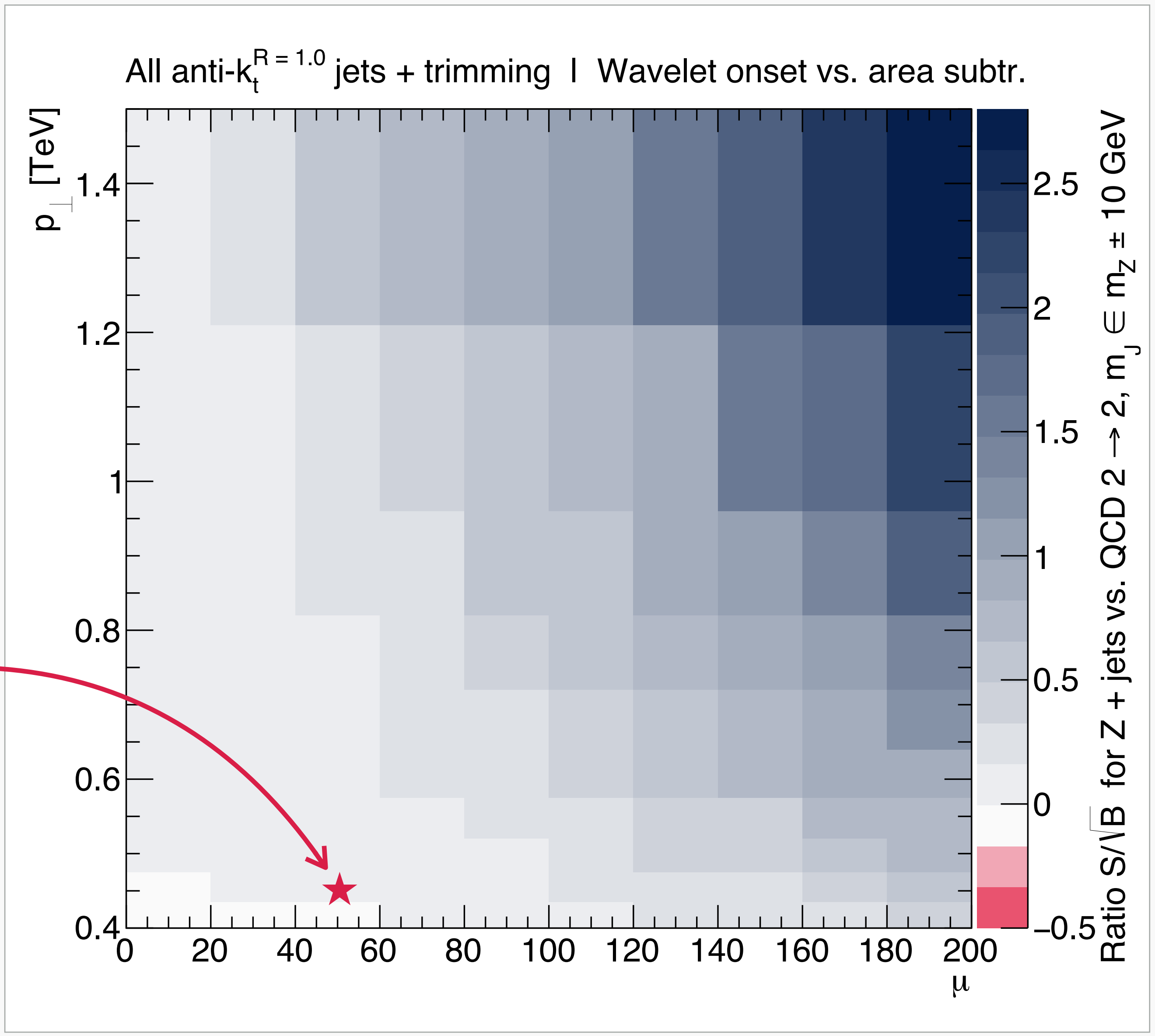

Incl. Z → qq̄, p⊥ > 400 GeV | ⟨μ⟩ = 100
Number of jets [a.u.]
Haar wavelet
128 × 128 grid
|η| < 3.2, |ηtrk.| < 2.5
Pixel thresh. 0.75
All anti-kt R=1.0 jets
+ trimming.
Without pile-up
Area median subtr.
Wavelet onset
50
60
70
80
90
100
110
120
130
Z-jet mass [GeV]

Z-jet mass


Incl. Z → qq̄, p⊥ > 400 GeV | ⟨μ⟩ = 100
Haar wavelet
128 × 128 grid
|η| < 3.2, |ηtrk.| < 2.5
Pixel thresh. 0.75
All anti-kt R=1.0 jets
+ trimming.
Without pile-up
Area median subtr.
Wavelet onset
SoftKiller
Number of jets [a.u.]
50
60
70
80
90
100
110
120
130
Z-jet mass [GeV]

Boson jet sensitivity improvement





Summary and outlook

- Motivated use of wavelets in HEP
- Showed the ability to naturally separate “white noise” pile-up from hard scatter events with small-angle structure
- Substantial potential is seen in improving measurement of both global and local inclusive observables
- Hope to implement similar methods in LHC analyses, ideally using a combination of the observables discussed today

Thank you.

Questions

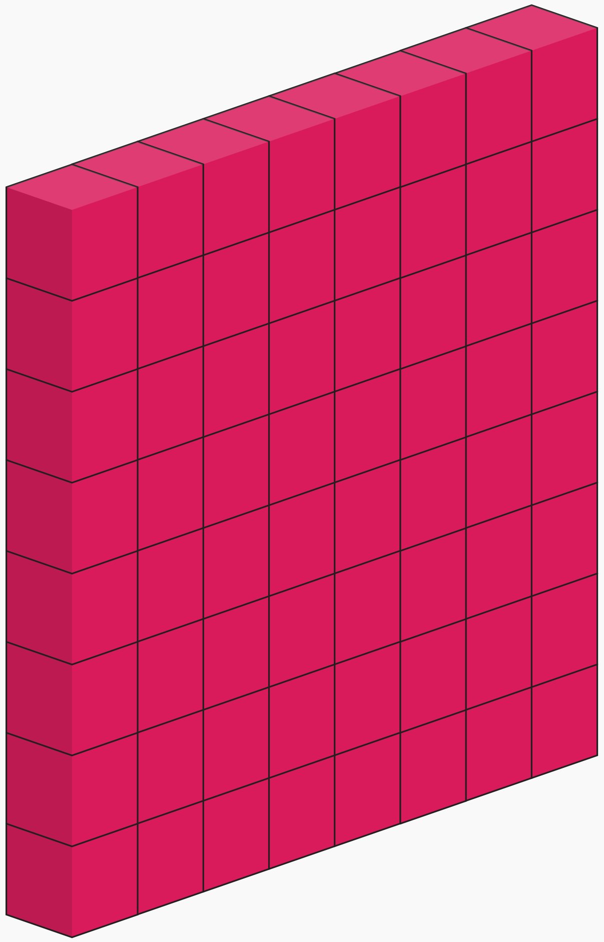
- How does it compare to PuPPI?
  - *Dunno.*
- Have you tried PuPPI?
  - *Nope.*
- Any thoughts on systematics?
  - *Nope.*

Backup

Learning optimal bases

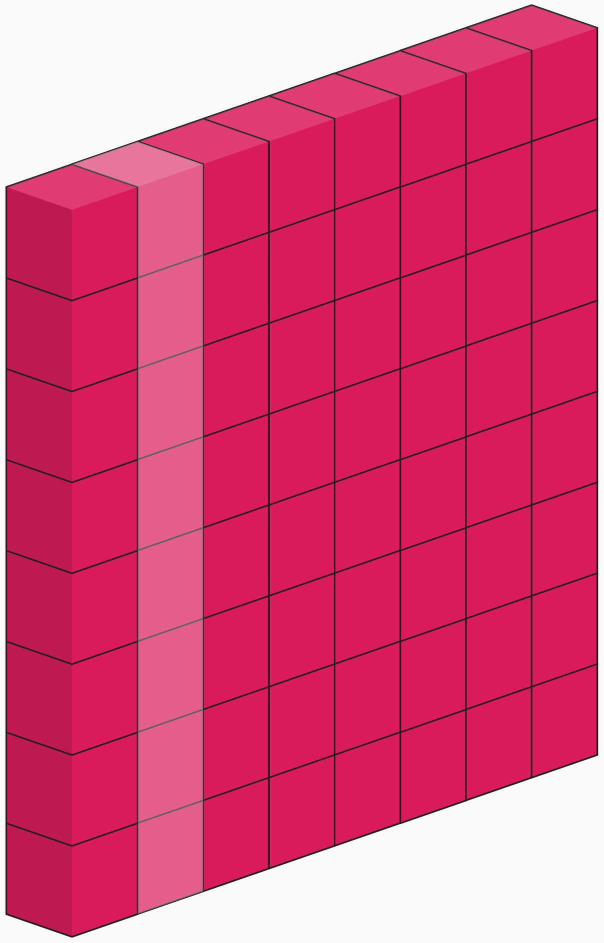
- Wavelet decomposition can be formulated as a deep neural network with a non-trivial architecture.
- Such a NN with 64 x 64 input in principle has 4.4×10^{7} weight coefficients.
- As a wavelet analysis realisation, the NN weight matrices are highly constrained.
- This means that the actual number of coefficients is N = 2, 4, 6, … i.e. the *filter coefficients* of the corresponding wavelet basis.

Wavelet/NN architecture



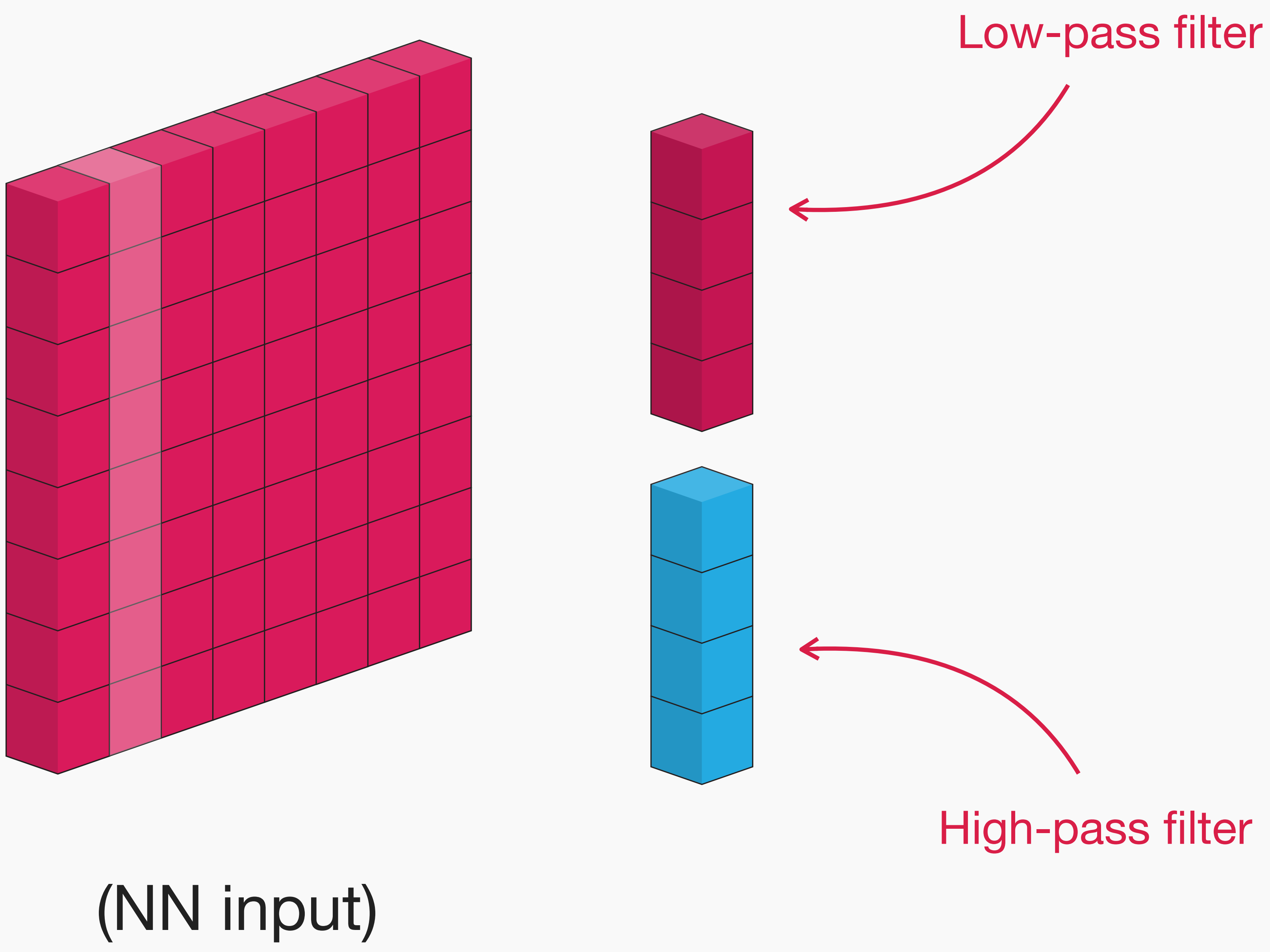
(NN input)

Wavelet/NN architecture

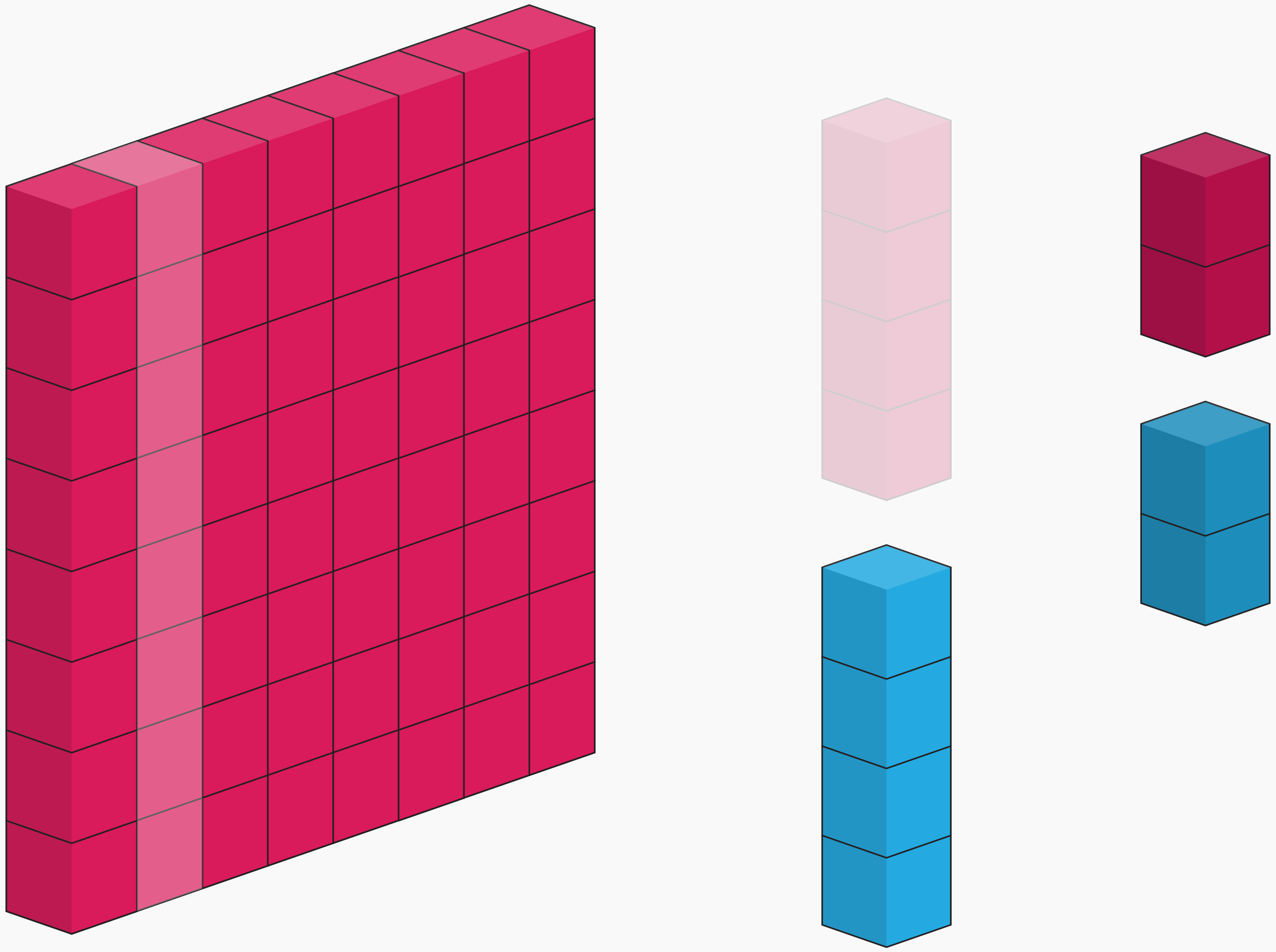


(NN input)

Wavelet/NN architecture

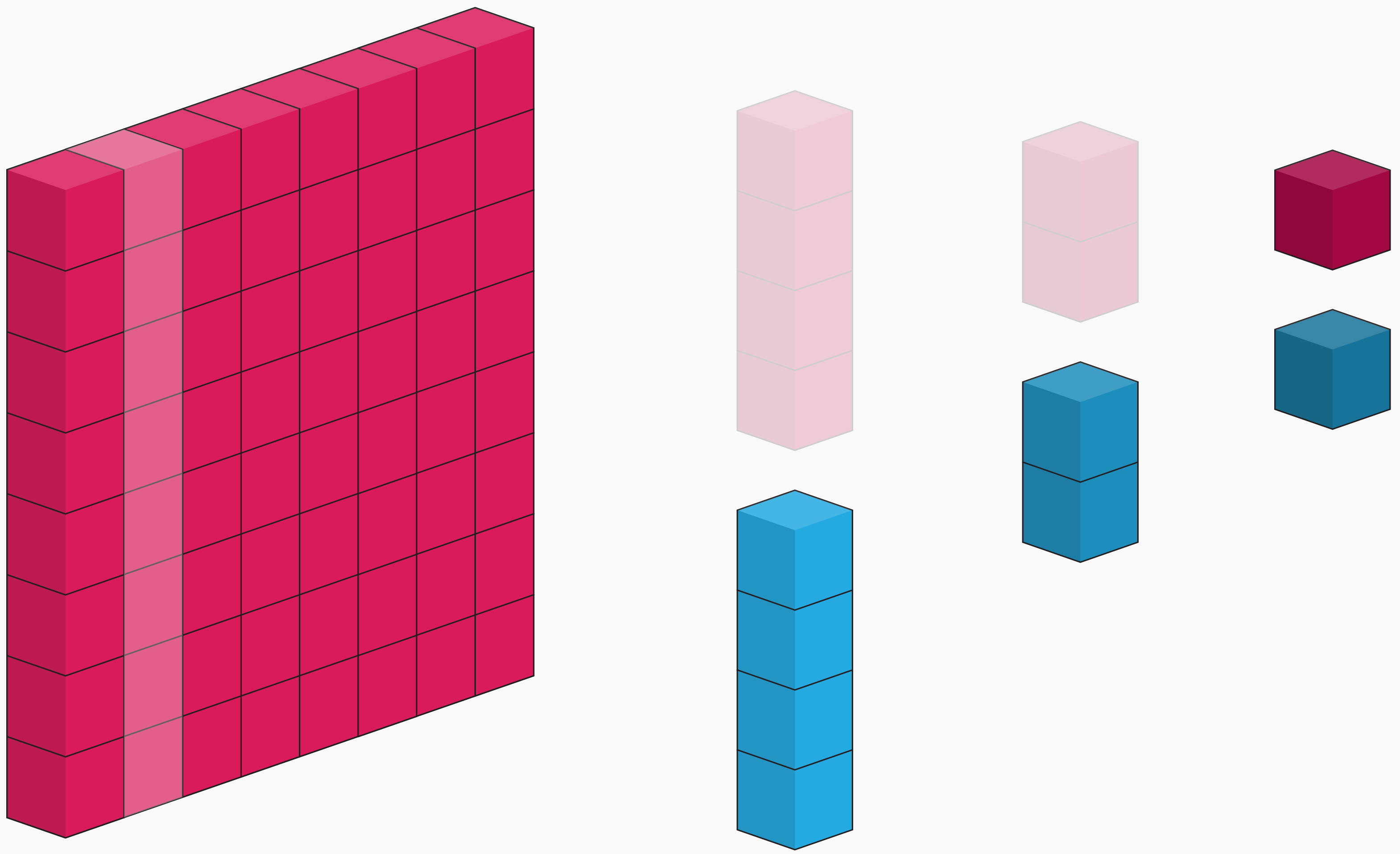


Wavelet/NN architecture



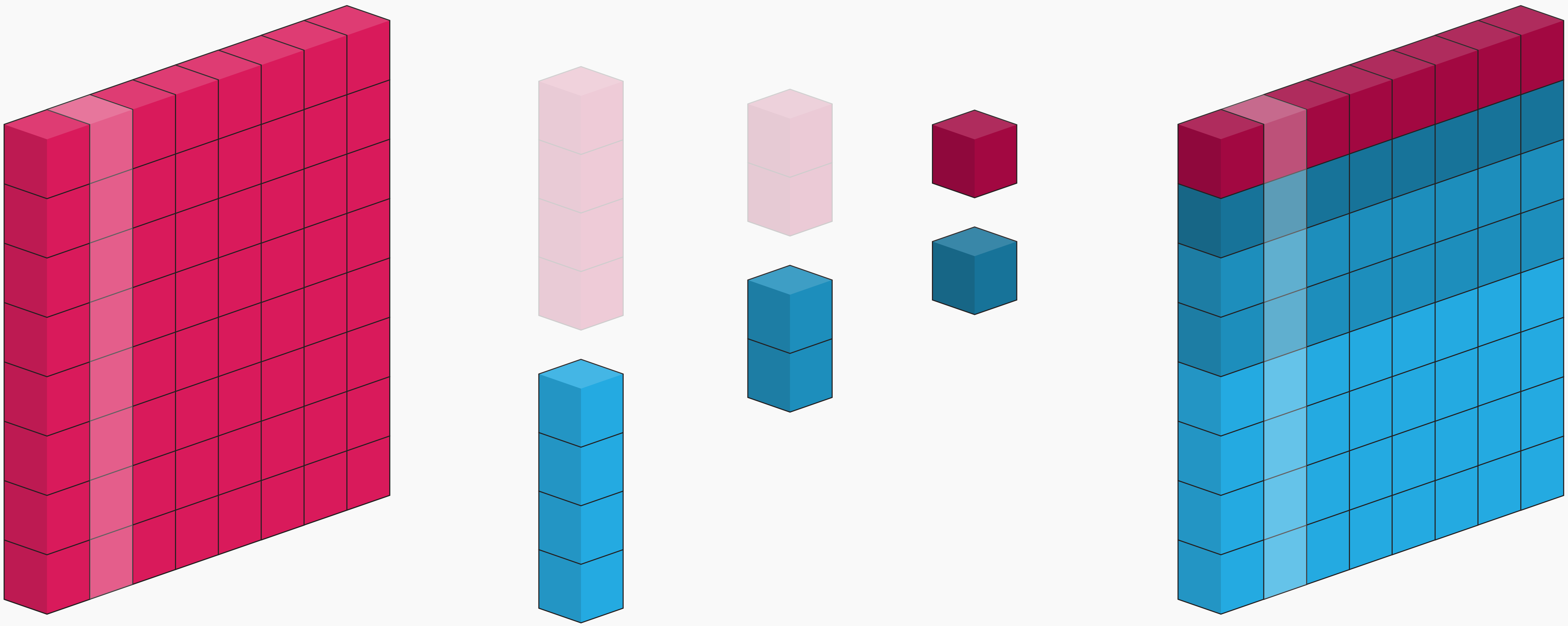
(NN input)

Wavelet/NN architecture



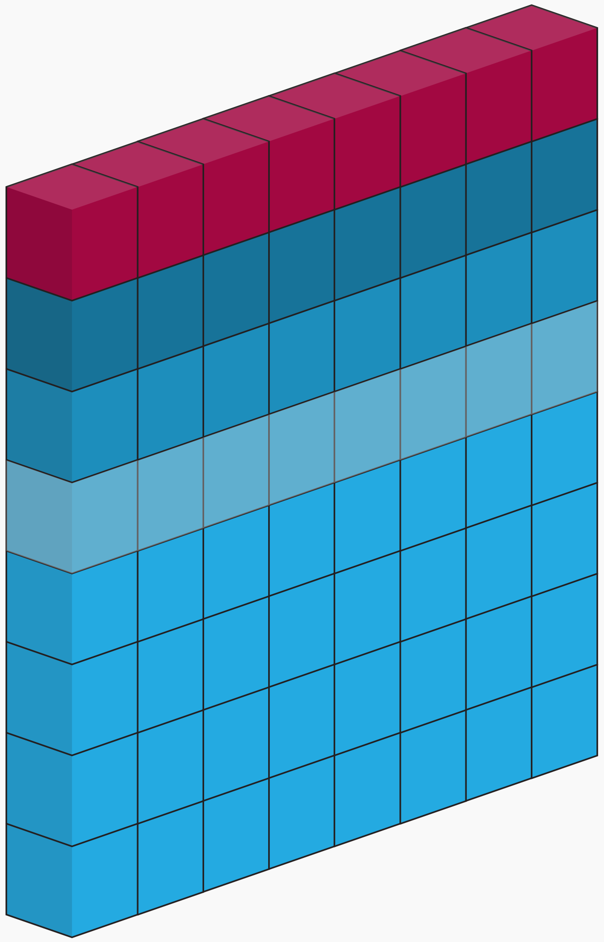
(NN input)

Wavelet/NN architecture

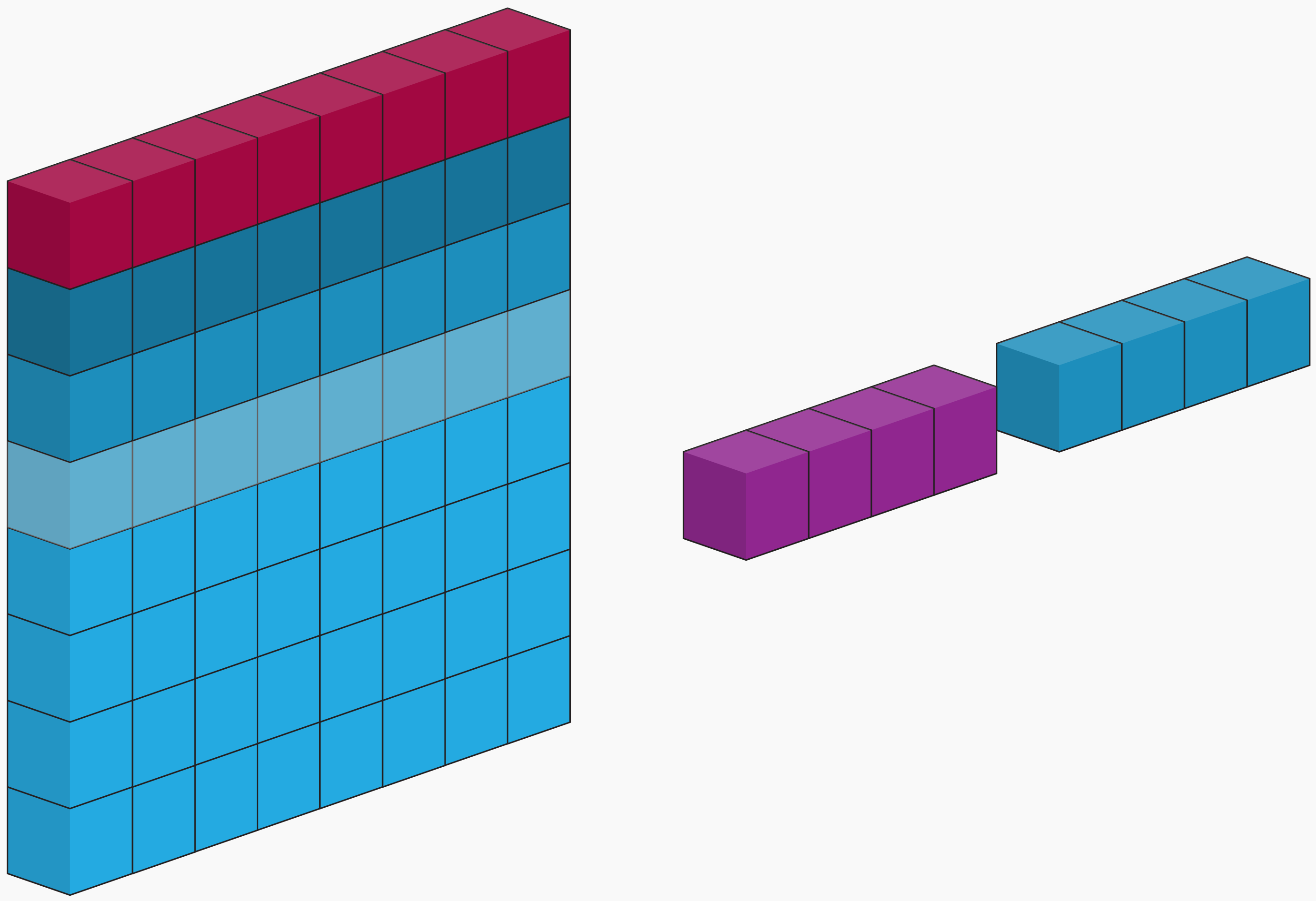


(NN input)

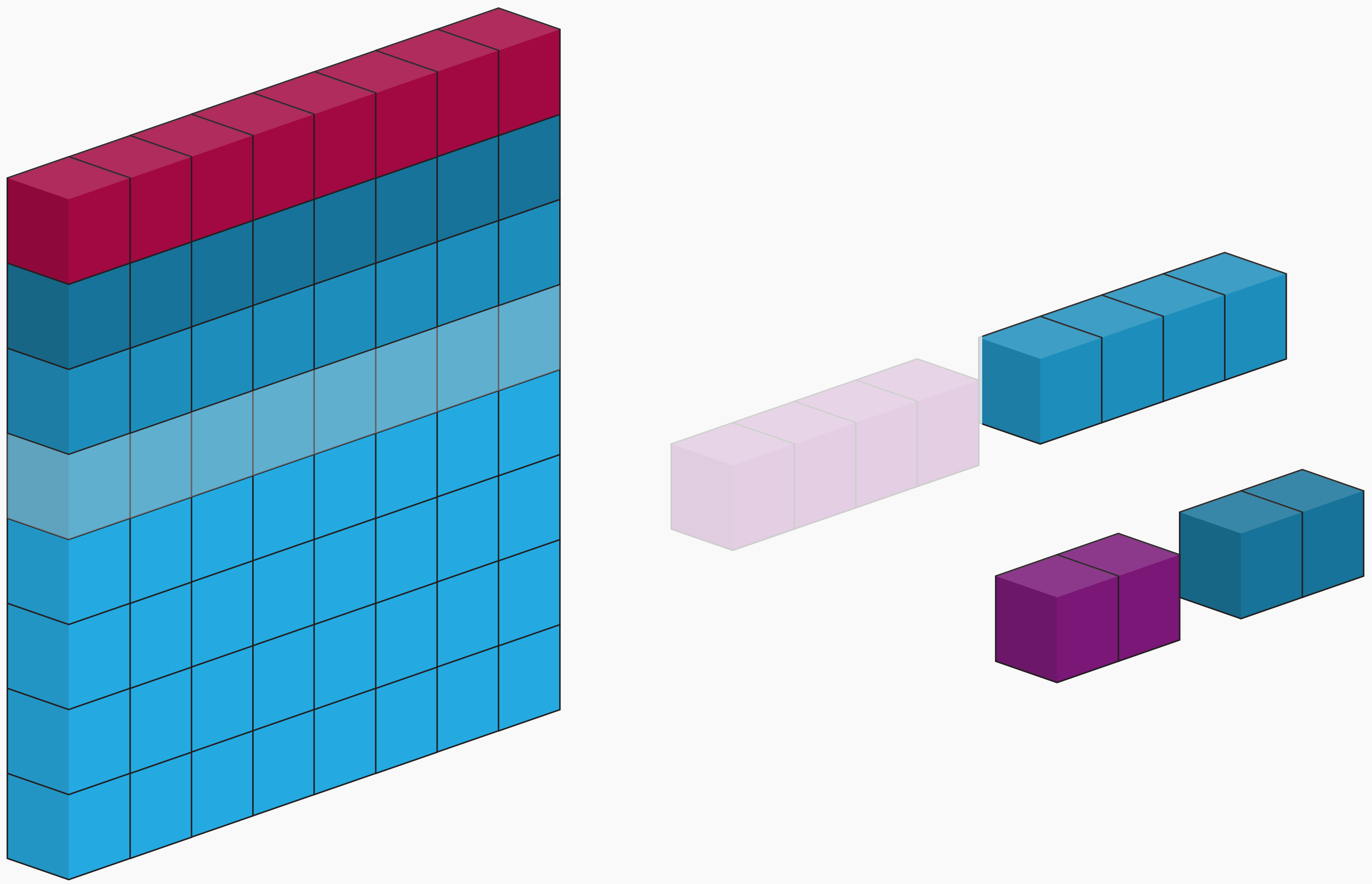
Wavelet/NN architecture



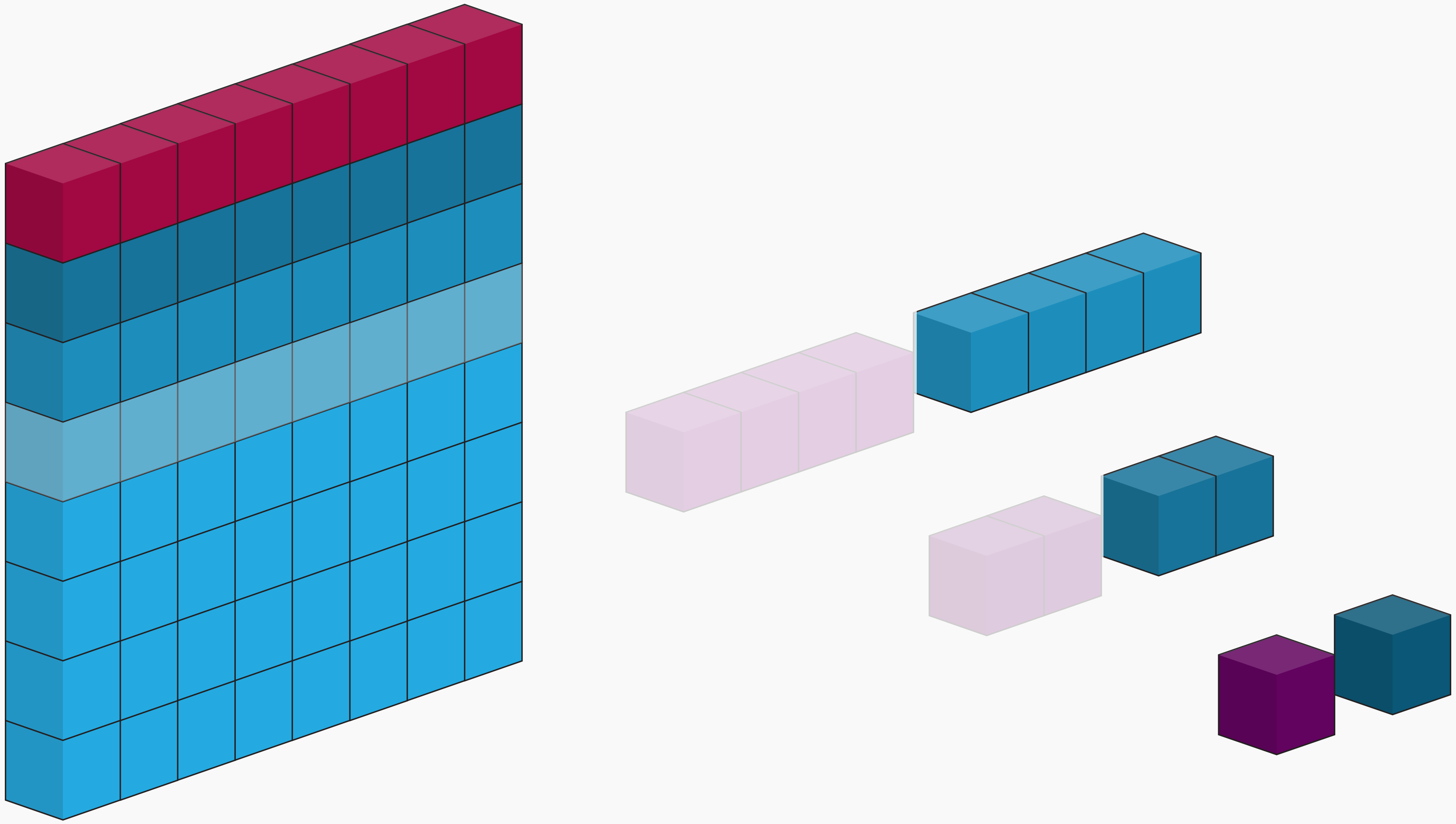
Wavelet/NN architecture



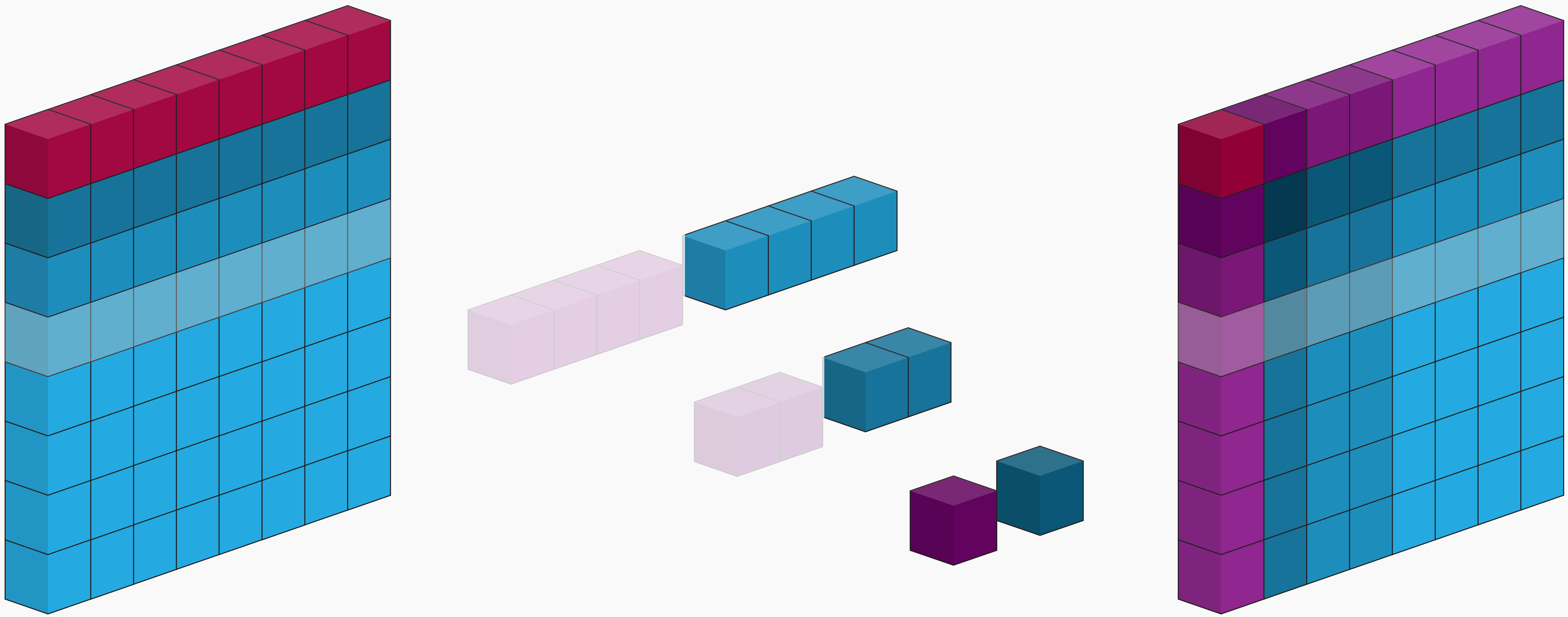
Wavelet/NN architecture



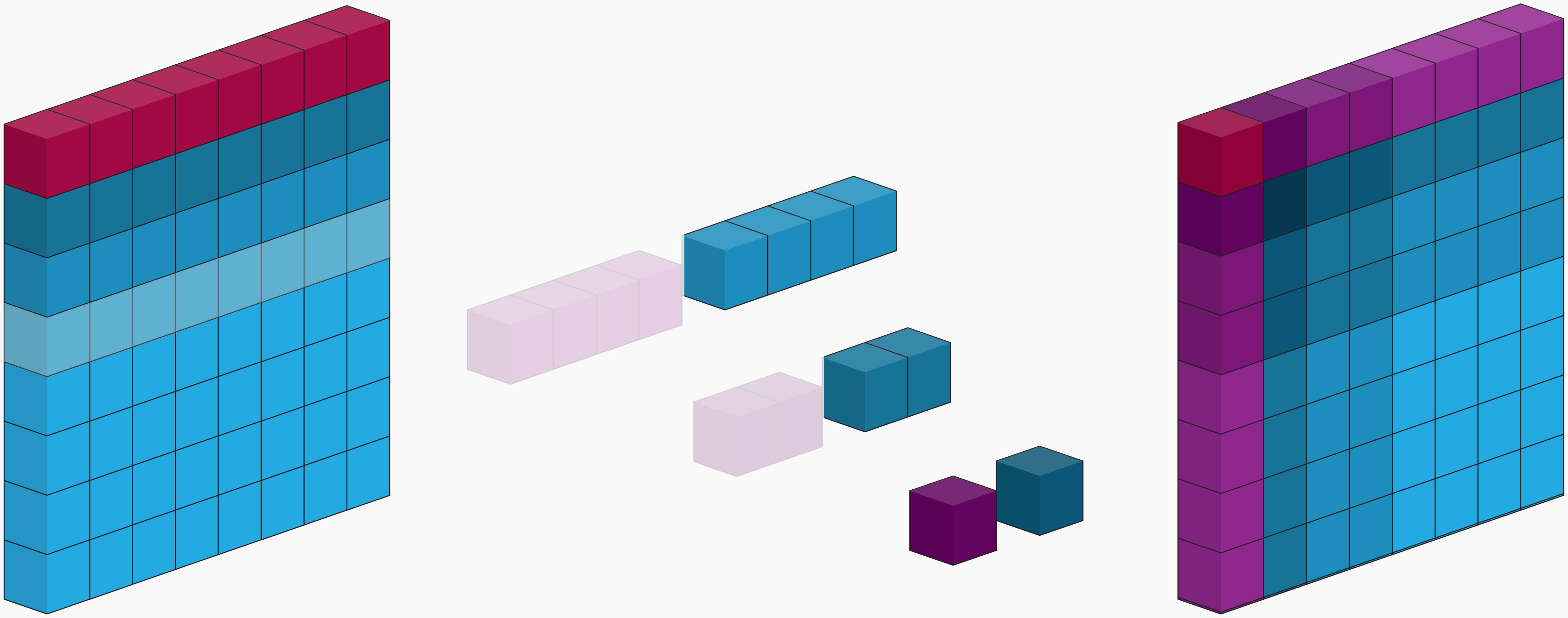
Wavelet/NN architecture



Wavelet/NN architecture



Wavelet/NN architecture



(NN output)

How to measure ‘optimality’

- One can then use gradient descend with back-propagation to learn filter coefficients which are optimal wrt. some criterion.
- One such measure is *sparsity*, i.e. the ability of a basis to efficiently encode information contained in a certain type of input, e.g. jet events.
- Sparsity can be quantified by the *Gini coefficient*
- In addition, we need to impose five constraints on the filter coefficients, to ensure that they result in an actual wavelet basis.
- These are realised as quadratic *regularisation* terms in the combined *cost function*.

Cost function

- Regularisation:

$$R_i = \left(f_i\left[\{a_k\}\right] - d_i\right)^2 \quad \Longrightarrow \quad R = \sum_{i=1}^{5} R_i$$

where a_k are the filter coefficients and d_i is some number.

- Sparsity (Gini coefficient):

$$G\left[\{c_i\}\right] = 1 - 2\sum_{i=1}^{N_c} \frac{c_i}{\|c_i\|_1} \frac{N_c - i + \frac{1}{2}}{N_c}$$

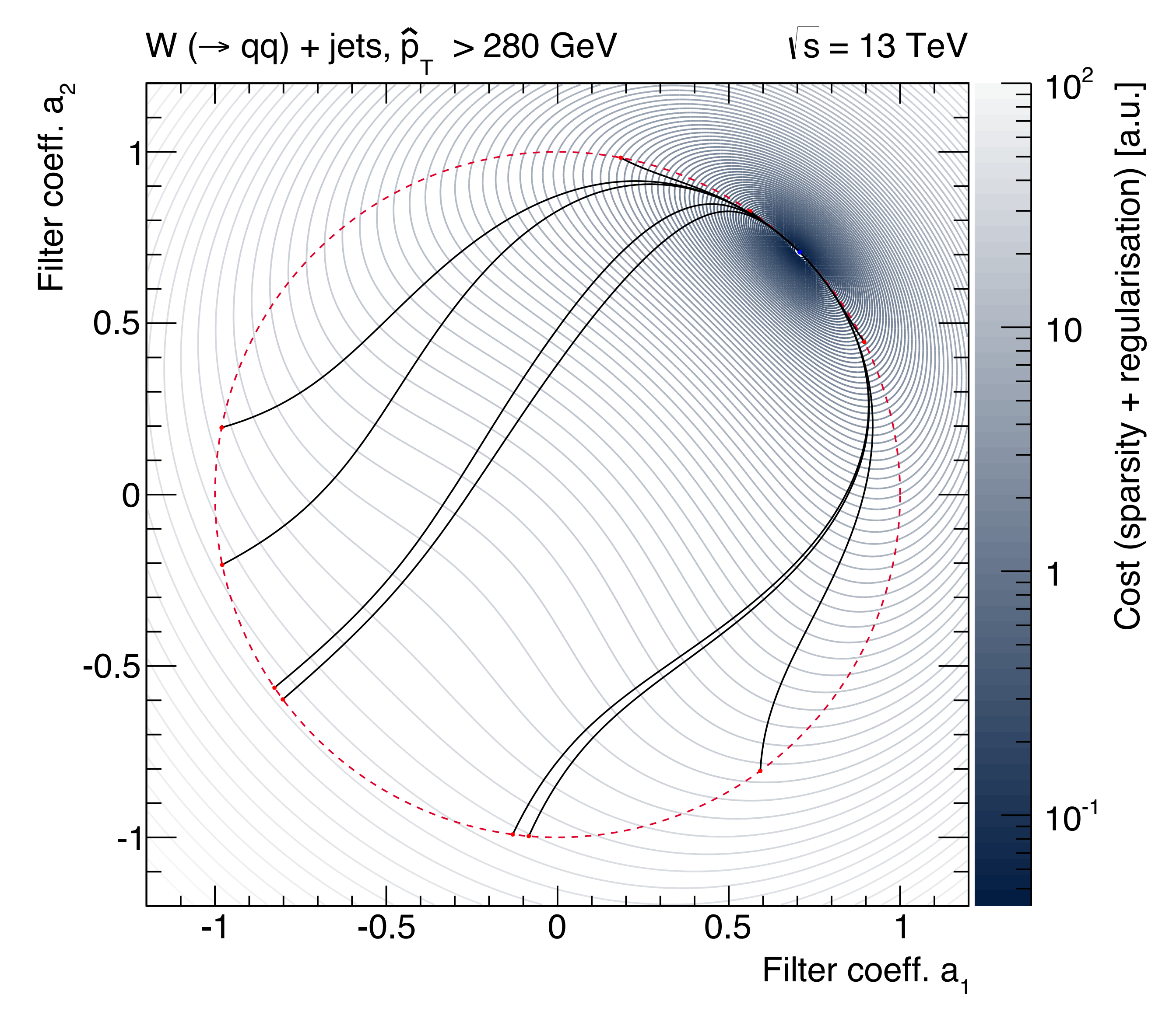
where c_i are the wavelet (not filter) coefficients for a single input, sorted such that $c_i \leq c_{i+1}$ and N_c are the number of such coefficients.

- Combined cost:

$$C = \left(G\left[\{c_i\}\right]\right)^2 + \lambda R, \qquad \lambda \in [0, \infty)$$

Cost map and gradient descend paths

- Loop over e.g. incl. W → qq events, we can optimise the combined sparsity + regularisation cost function.
- Plot shows 10 random initialisations in unit circle (red dots; one of the five constraints) and the learning paths (black lines) going towards a single, global minimum of $(1,1)/\sqrt{2}$: **Haar**
- More than one possible minimum for N > 2 filter coeffs.

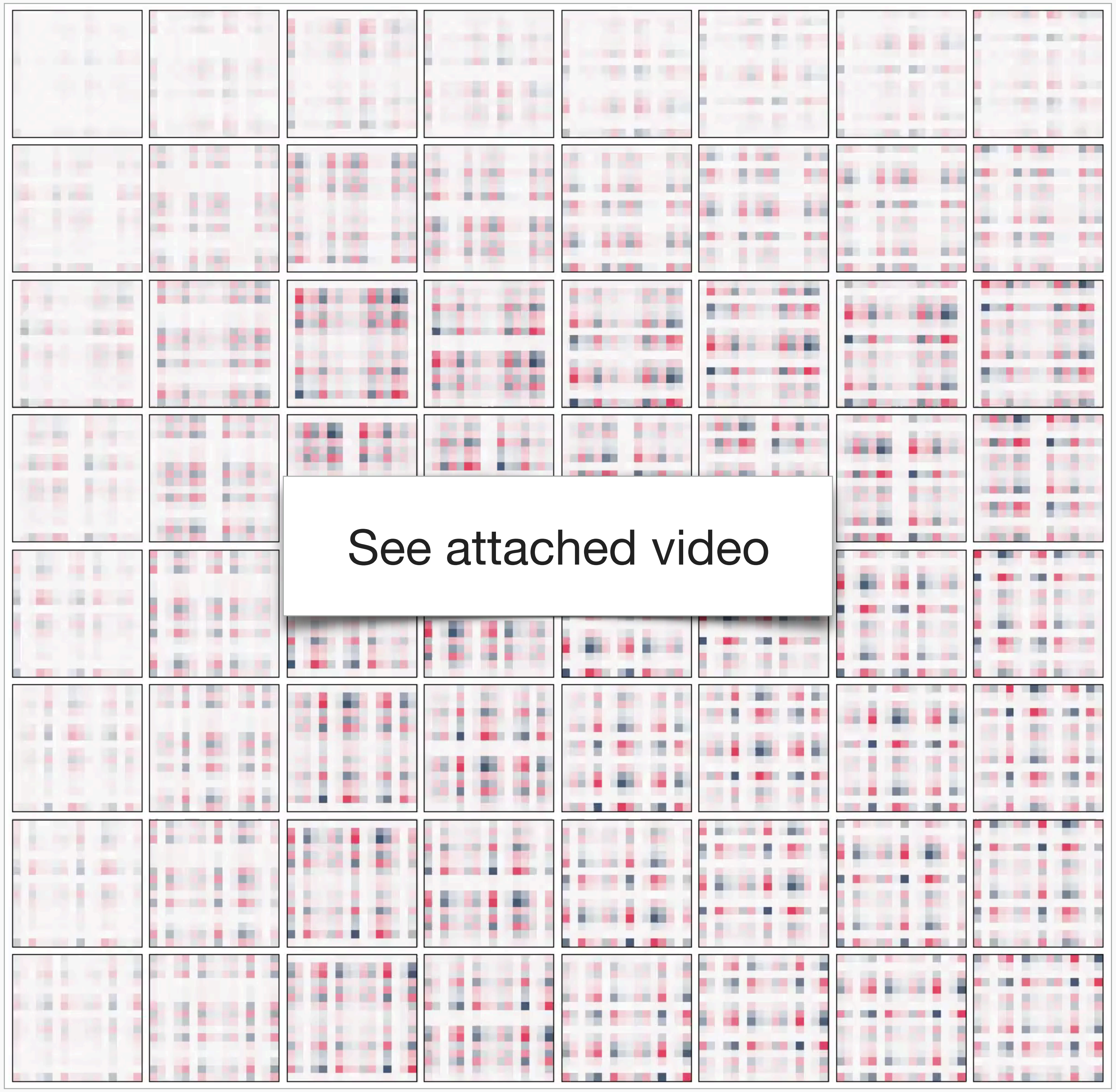


 | andreas.sogaard@ed.ac.uk

Visualising learning

- **16** filter coeffs. = 16-dim. optimisation problem
- Incl. $W \rightarrow qq$, $\hat{p}_T >$ 280 GeV, no pile-up
- 25’000 events.
- Optimisation from random initialisation

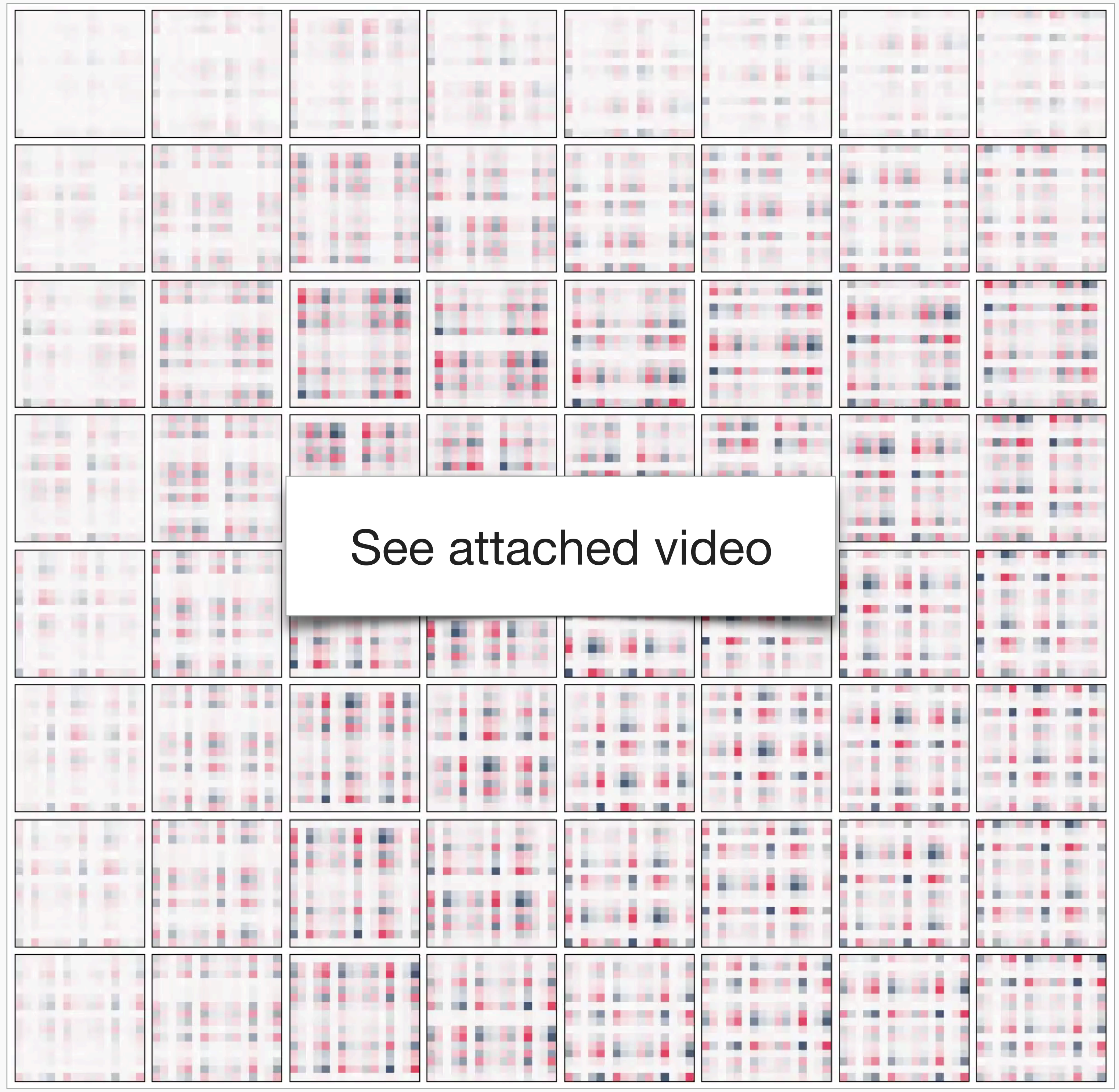
1. part: Regularisation
    - Normalisation
    - Orthogonality
    - Self-similarity
2. part: Sparsity



Visualising learning

- **16** filter coeffs. = 16-dim. optimisation problem
- Incl. $W \to qq$, $\hat{p}_T >$ 280 GeV, no pile-up
- 25’000 events.
- Optimisation from random initialisation

1. part: Regularisation
    - Normalisation
    - Orthogonality
    - Self-similarity
2. part: Sparsity



Use

1. *Learn* the best bases for various specialised tasks, like pile-up mitigation,
2. *Learn* observables for discriminating e.g. vector boson and 'QCD' jets, based on differences in their angular structure
3. Probe the splitting functions directly,
4. Possibly use these results to *learn* parton showering (and hadronisation) from data, with minimal theory input
5. Perhaps even develop a jet clustering algorithm based on this type of learning — however, this is perhaps best done using recurrent neural networks like LSTM

References

- J. Monk, *Wavelet Analysis of Boosted Bosons*, BOOST 2014
  - [https://indico.cern.ch/event/302395/contributions/692393/attachments/571643/787362/JMonk_Boost2014.pdf]
- J. Monk, *Wavelet Analysis: Event De-noising, Shower Evolution and Jet Substructure Without Jets* (2014)
  - [arxiv.org/pdf/1405.5008.pdf]
- NewWave, C++ library for performing wavelet analyses
  - [newwave.hepforge.org]
- A. Søgaard, *Boosted Bosons and Wavelets* (2015), MSc thesis
  - [cds.cern.ch/record/2055290]