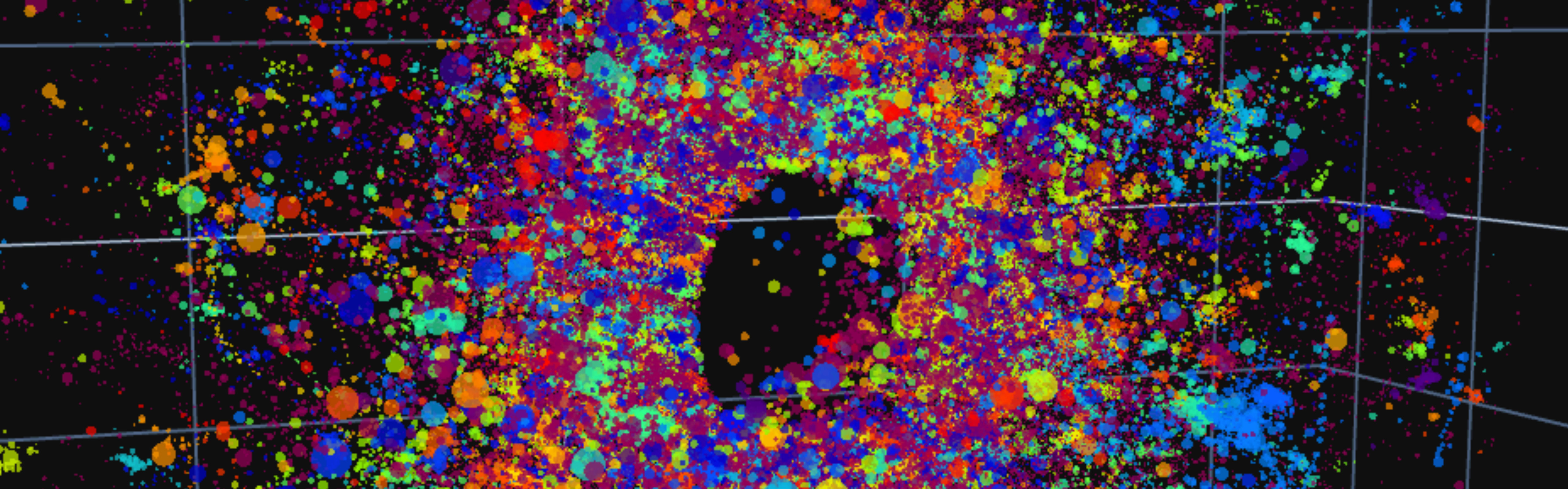


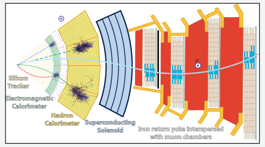
Conceptual Considerations for ML-based Reconstruction

Jan Kieseler

ML in the analysis chain (in CMS)

Particle Flow:

Hits to candidates



Physics Objects

Muon

"Medium"

Electron

"Hard"

b-jet

tau

photon

"fat" jet

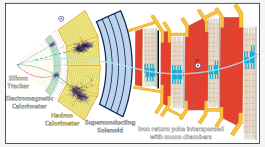
Analysis

ML in the analysis chain (in CMS)

"Very hard"

Particle Flow:

Hits to candidates



Physics Objects

Muon

"Medium"

Electron

"Hard"

b-jet

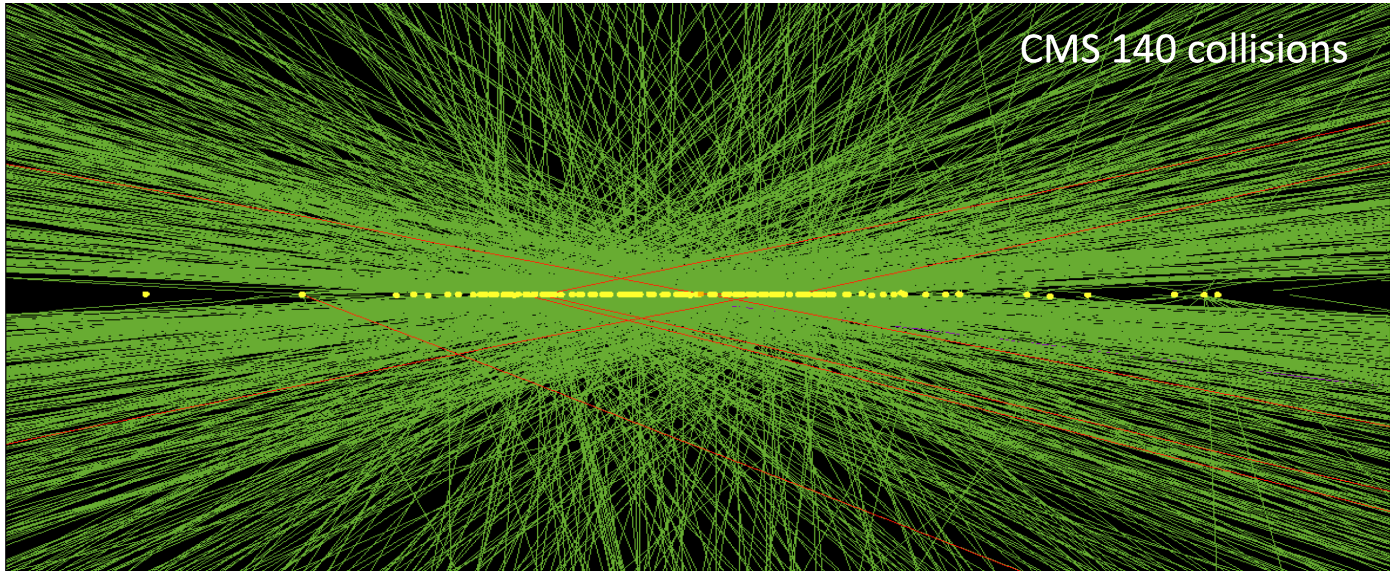
tau

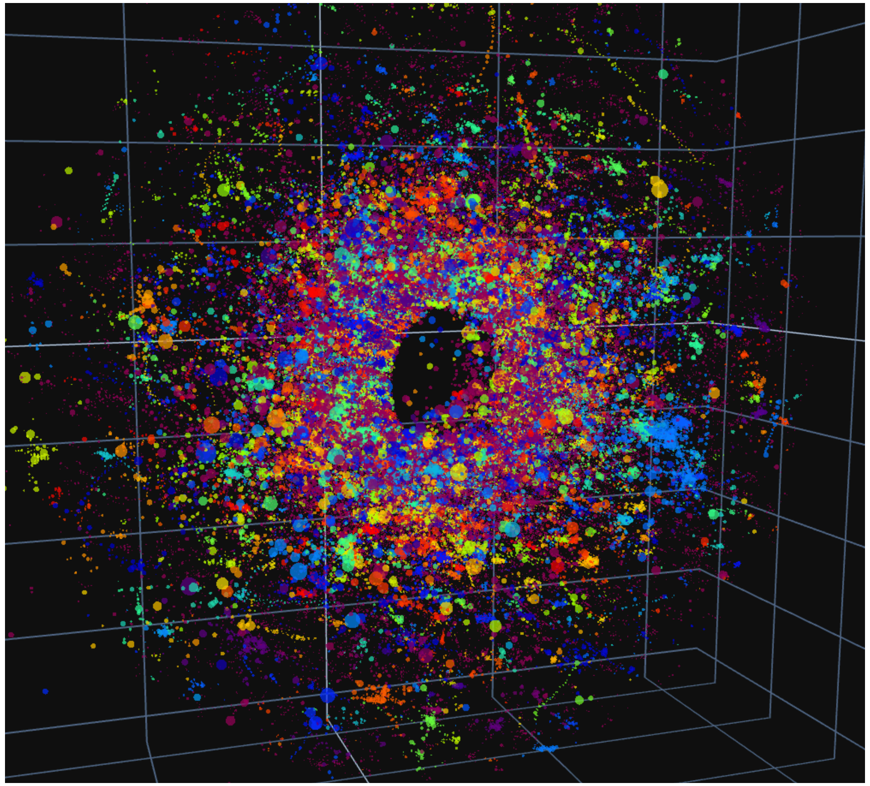
photon

"fat" jet

Analysis

The need for better algorithms

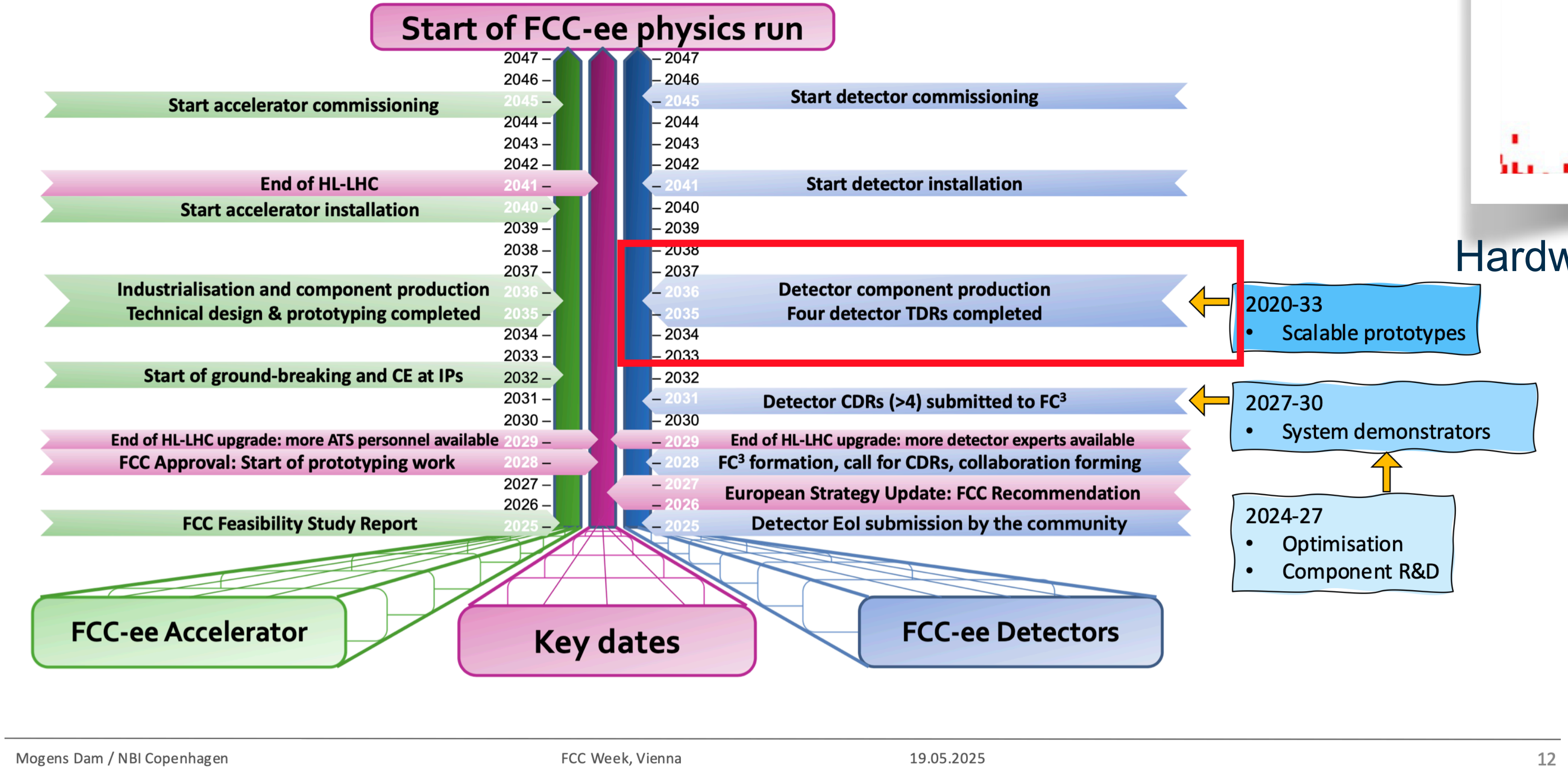


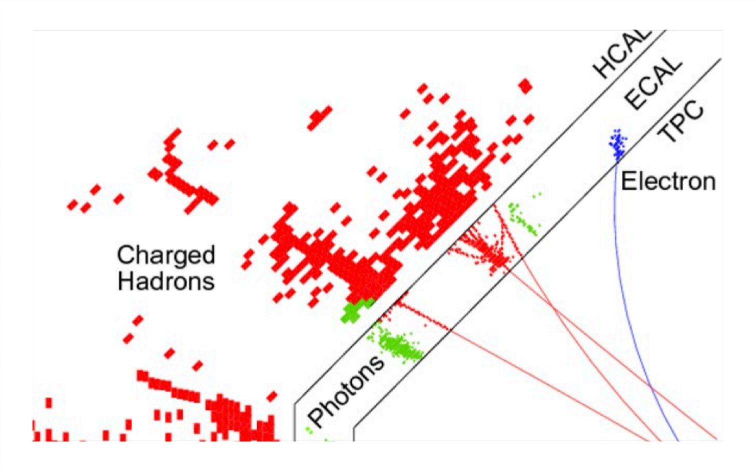


More complex and overlapping patterns to reconstruct to particles: **ML**

Strain on computing ressources: **parallelisable algorithms (ML) on GPUs**

More need for better algorithms: ML approaches are highly flexible





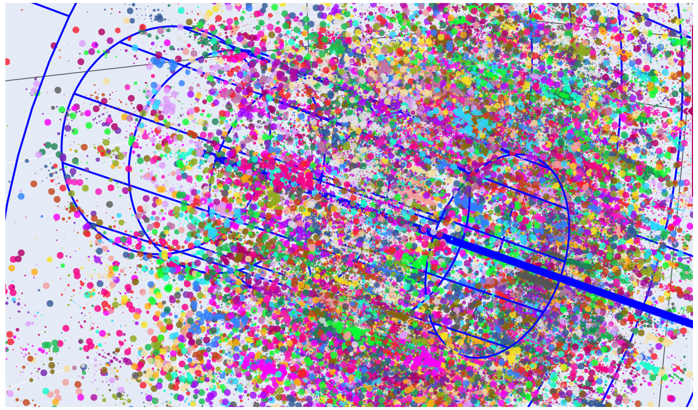
Hardware choices and reconstruction are inevitably intertwined "Co-Design"

CALO5
A new dimension in shower reconstruction

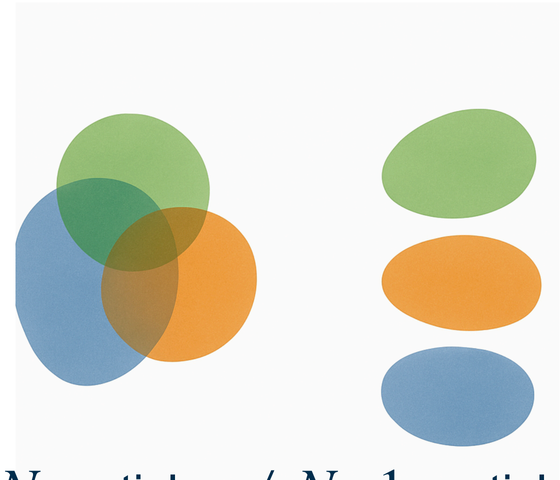
New hardware capabilities

➡ ML means: adaptive algorithms, low turn-around, possibility for (gradient based) detector optimisation

Conceptual Considerations: Particle Flow



High input multiplicities:
Demand on resources

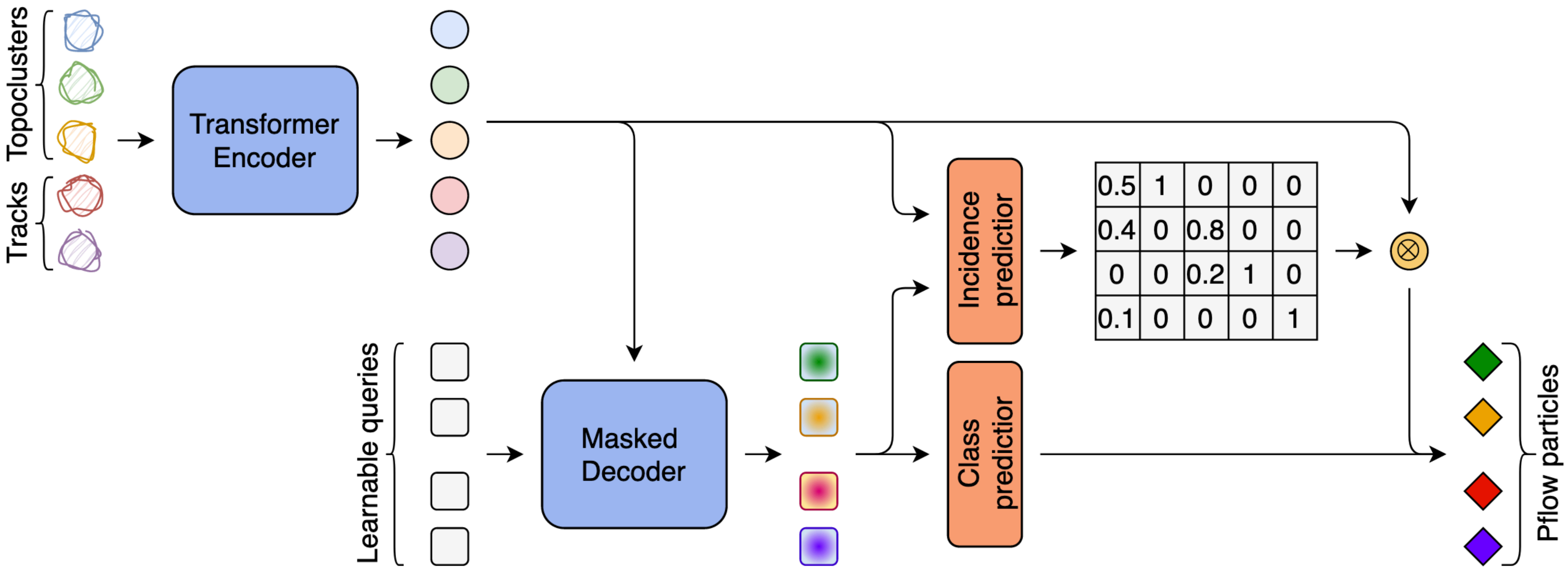


N particles $\neq N \cdot 1$ particle
advanced concepts

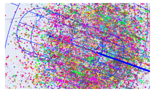
Physics robustness:
Careful target definition

GLOW

arXiv:2508.20092

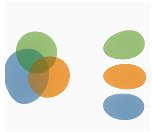


Trained and applied to CLIC detector dataset
O(100) inputs



Targets particles with at least a track or calorimeter deposit

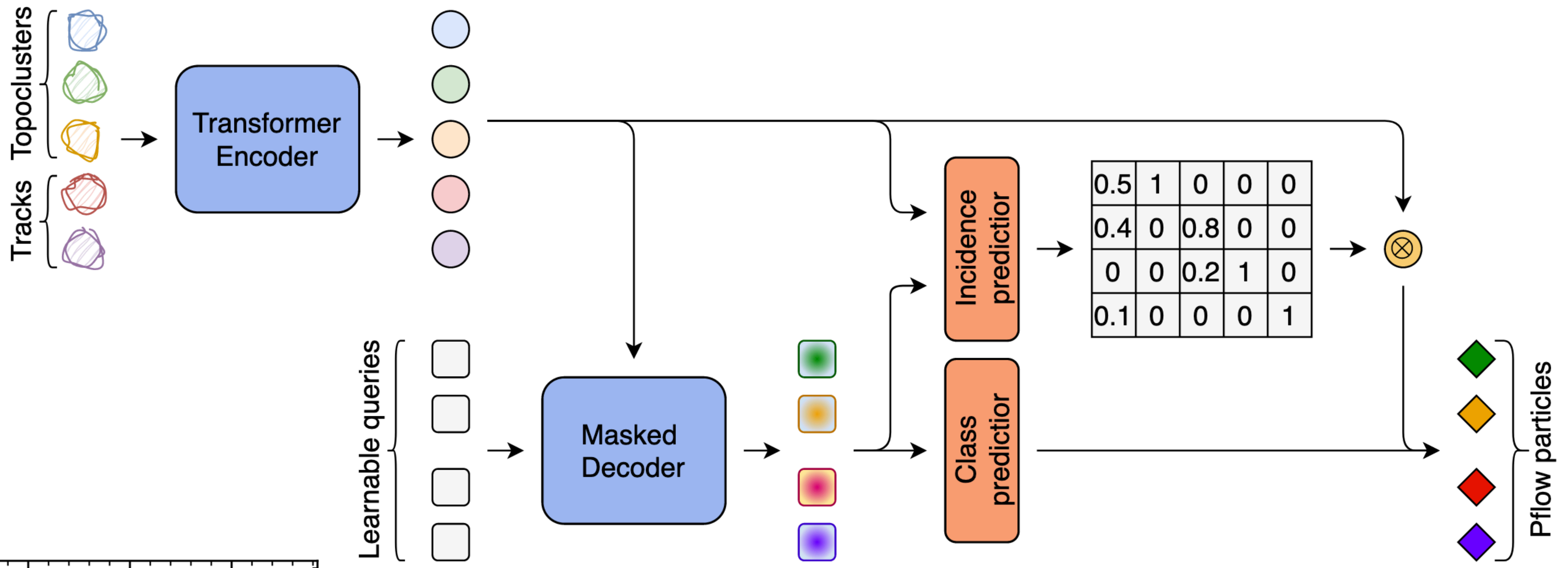
Counts particles globally*

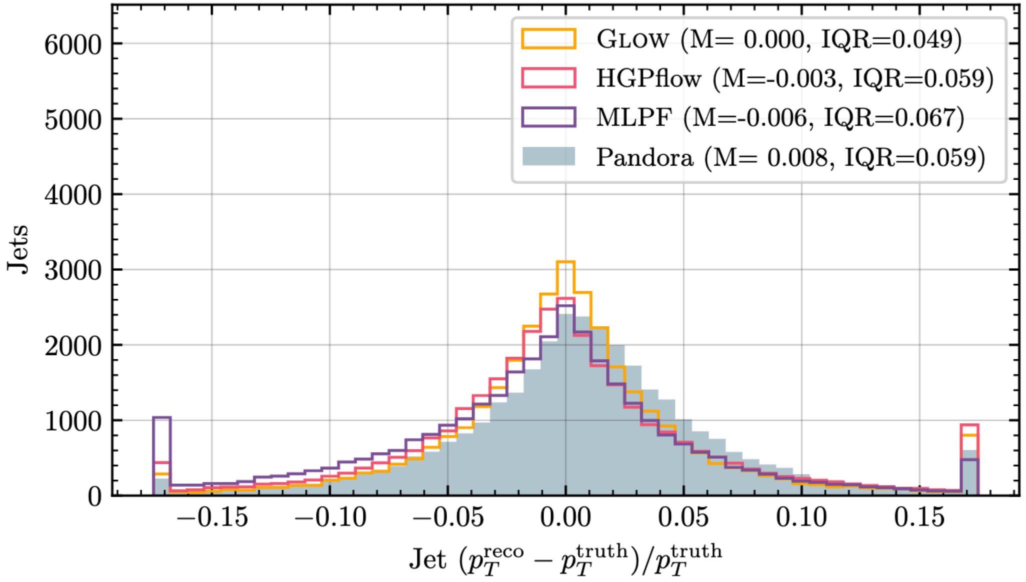


* similar to other HGPFlow variants

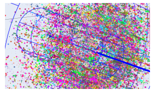
GLOW

arXiv:2508.20092



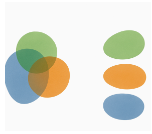


Trained and applied to CLIC detector dataset
O(100) inputs



Targets particles with at least a track or calorimeter deposit

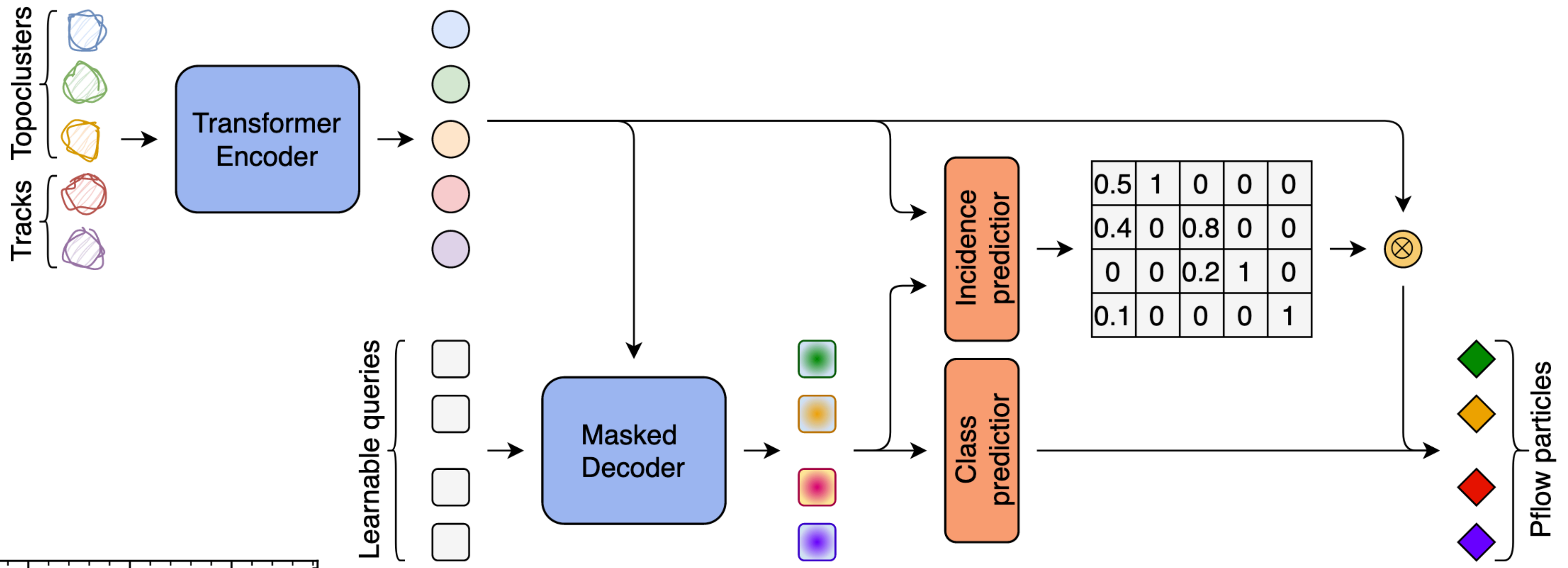
Counts particles globally*

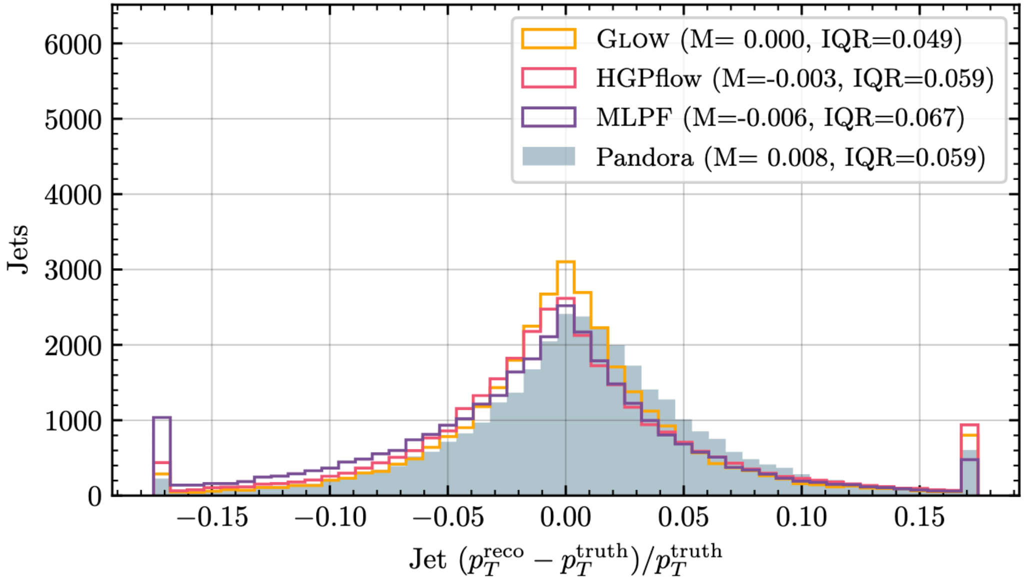


* similar to other HGPFlow variants

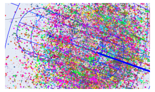
GLOW

arXiv:2508.20092



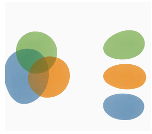


Trained and applied to CLIC detector dataset
O(100) inputs



Targets particles with at least a track or calorimeter deposit

Counts particles globally*



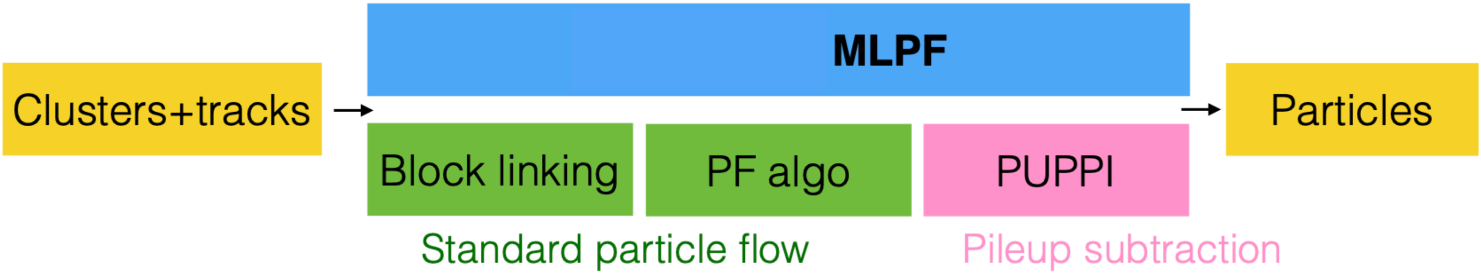
Prior probabilities from the training dataset

For illustration: $P(H_0\,|\,\boldsymbol{x}) = \dfrac{f(\boldsymbol{x}\,|\,H_0)\pi_0}{f(\boldsymbol{x}\,|\,H_0)\pi_0 + f(\boldsymbol{x}\,|\,H_1)\pi_1}$

* similar to other HGPFlow variants

CMS MLPF

CMS-DP-2025-033

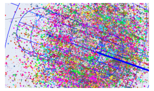


Aims:

Simplified pipeline

Faster runtime

Improved physics performance

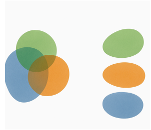


O(1000) inputs

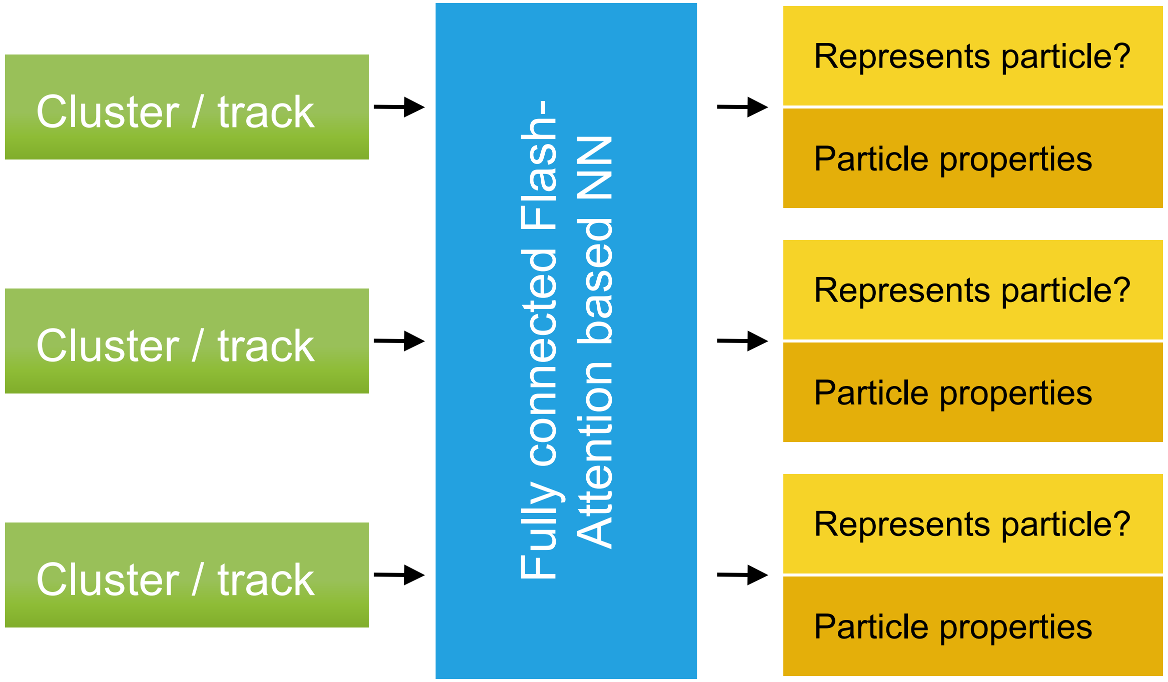
Define targets:

- Charged: by track
- Neutral: merge all in one cluster

- Pre-define one representative point



Fully supervised training on $t\bar{t}$, QCD, $Z \rightarrow \tau\tau$

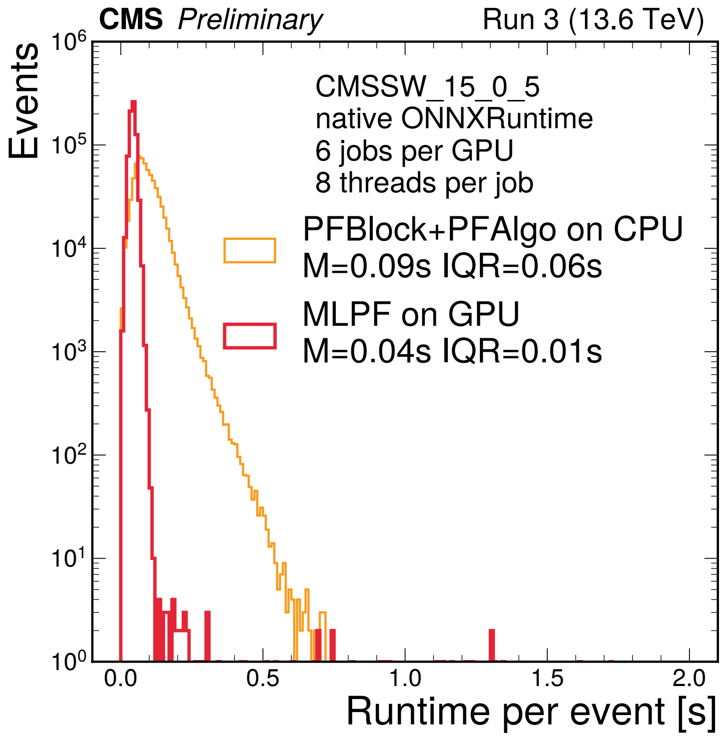


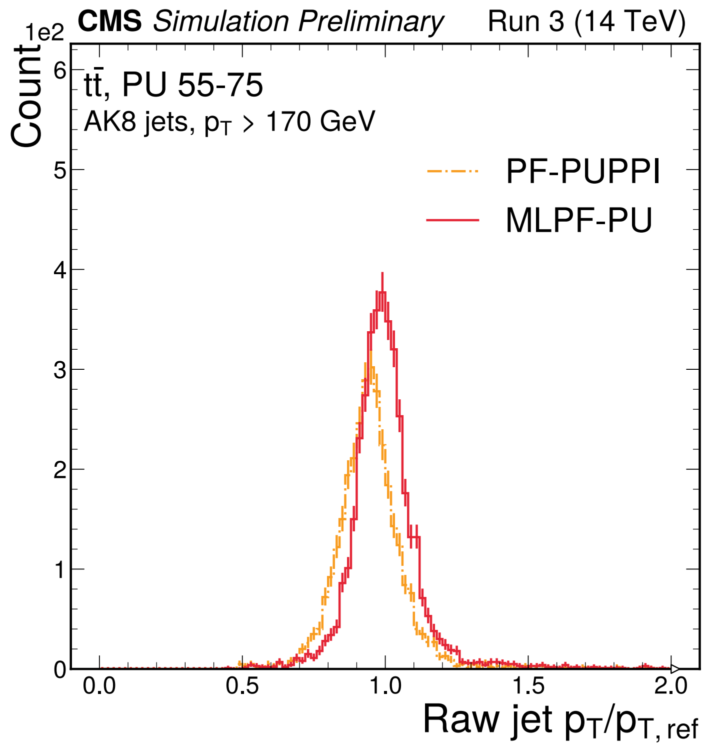
Adapts to different geometries nicely (e.g. FCC det.)

F. Mokhtar, et al, https://arxiv.org/abs/2503.00131

(Follows a simplified object condensation paradigm)

It’s hard but worth it: Improving physics and runtime





CMS-DP-2025-033

- ➡ **Improved jet resolution over standard PF with identical inputs**
- ➡ Fast runtime
- ➡ **Including timing information in ML models is straight forward**

Locality as a must

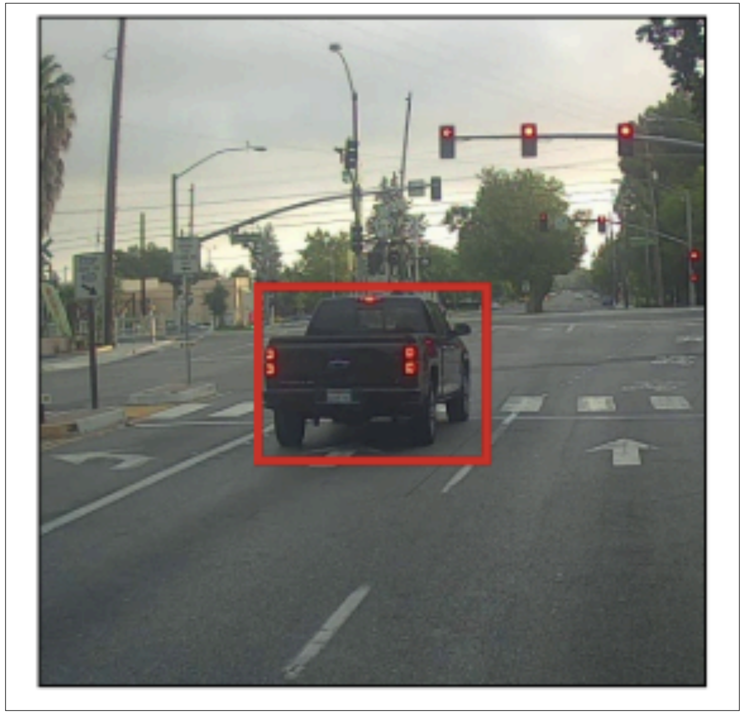
Calibration procedures heavily rely on independence of particles reconstructed in (very) different detector regions

Z/gamma jet balance

Tag & Probe

Event mixing

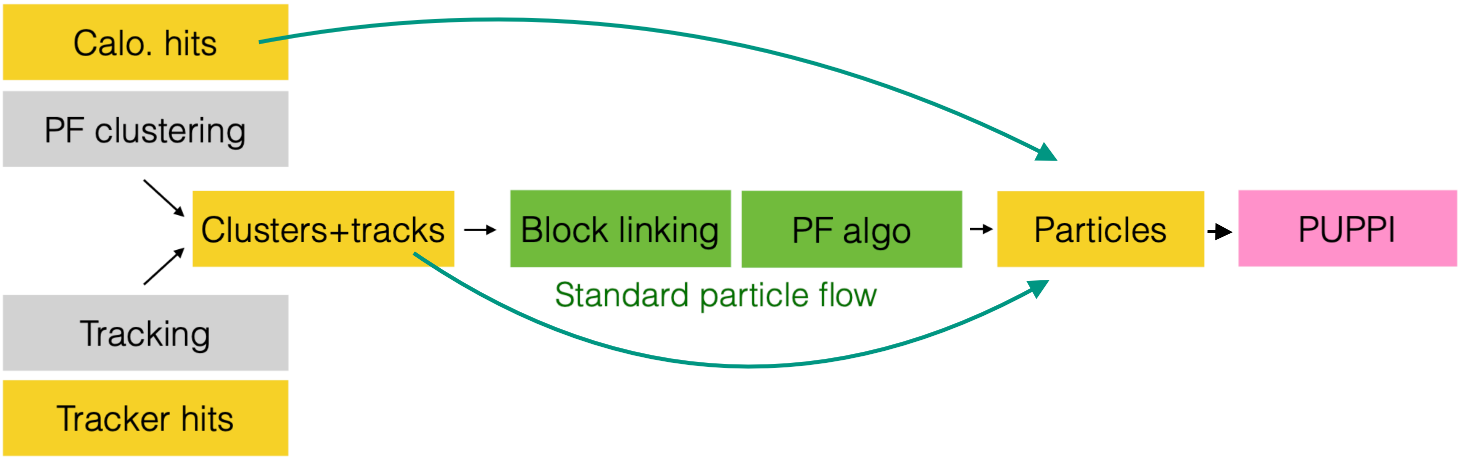
AD bump hunts

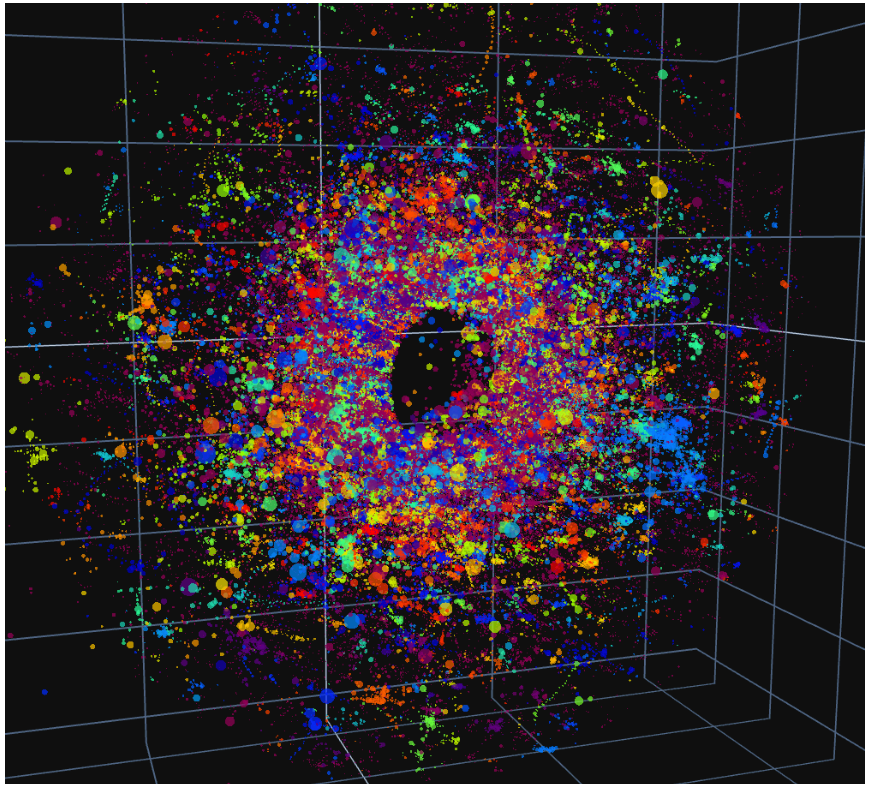


Include Calorimeter Hits: towards “end-to-end”

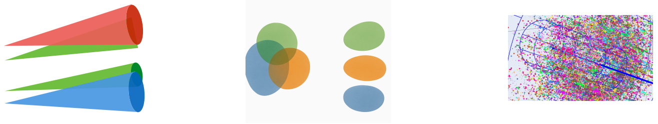
Utilise higher granular inputs from calorimeter hits

First reconstruct all particles, then perform pileup subtraction
(depends on PV definition)





MLPF approach not applicable anymore: target, concepts, and resources



The target matters

With ML calorimeter clustering, a new more granular and general definition is needed for a robust algorithm

Why:

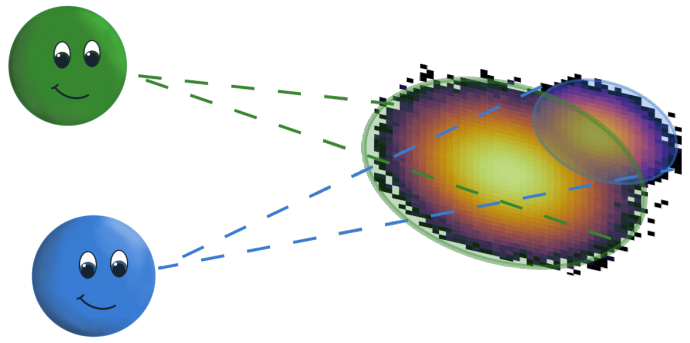
Example from classification:

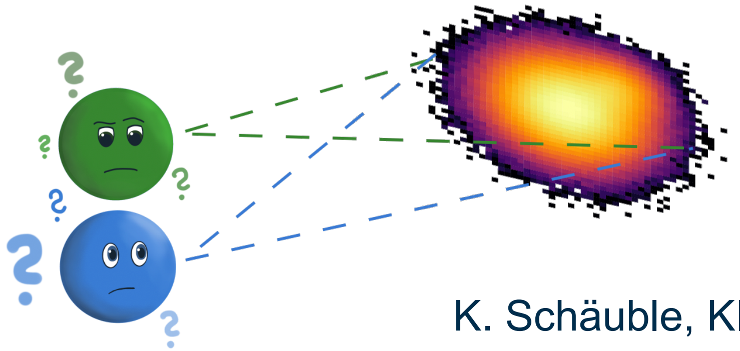
$$P(H_0 \,|\, \boldsymbol{x}) = \frac{f(\boldsymbol{x} \,|\, H_0)\pi_0}{f(\boldsymbol{x} \,|\, H_0)\pi_0 + f(\boldsymbol{x} \,|\, H_1)\pi_1}$$

Prior probabilities from the training dataset

The more the algorithm guesses the more it relies on the priors: the less it generalises to different physics scenarios.

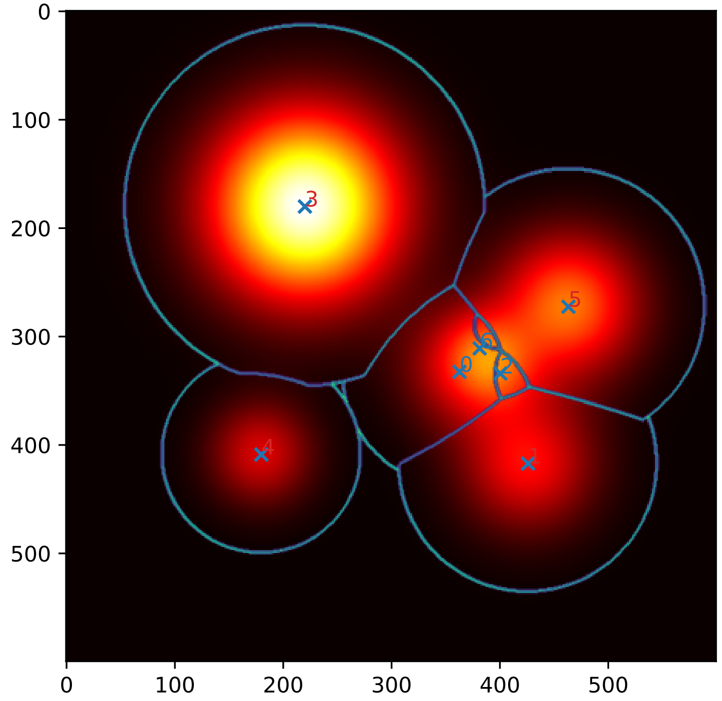
A good truth definition leads to a robust algorithm, it relies on what we consider in principle resolvable by the detector.

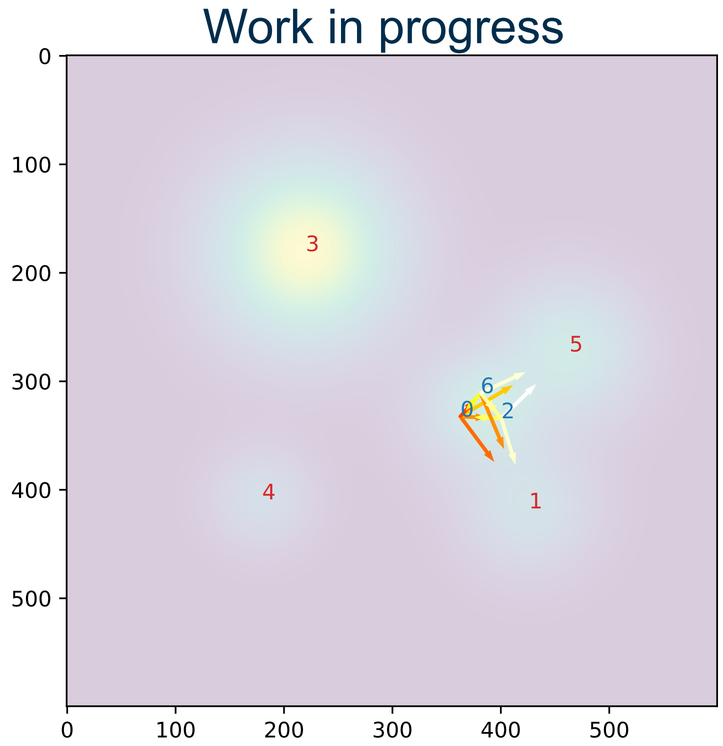




K. Schäuble, KIT

Truth merging





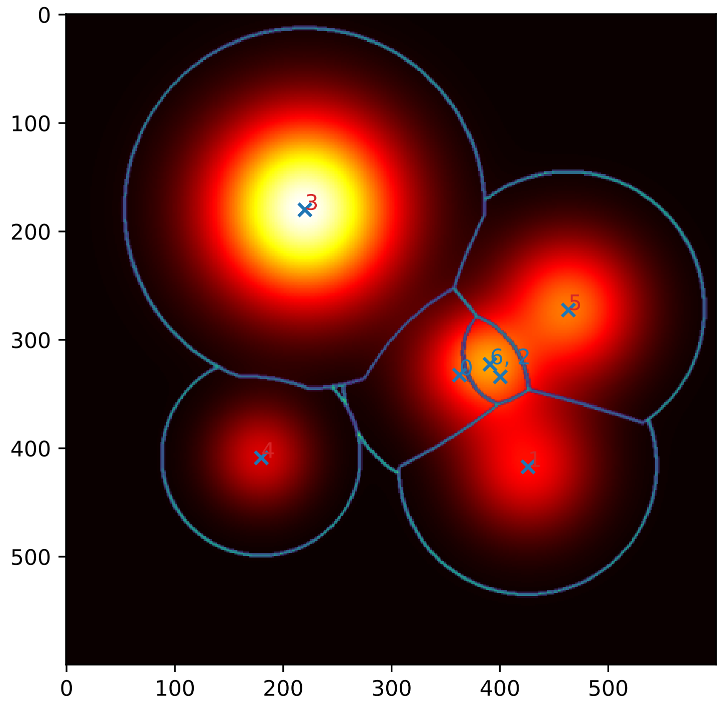
Based on resolvability w.r.t sensors: implement detector geometry in a generic way

Creates a directed graph: avoids over-merging

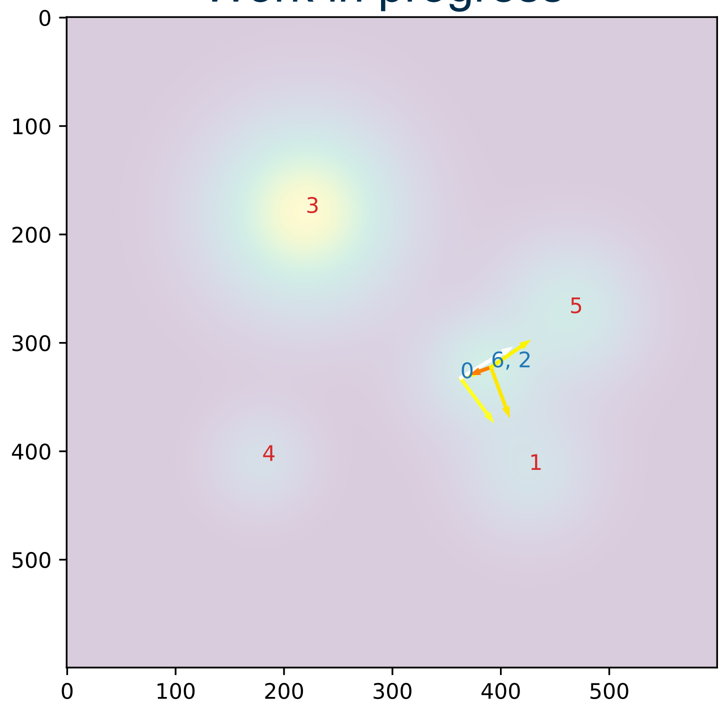
Performs resolvability-ordered merging: adds IRC safety

A. Busamolino, KIT, paper in prep.

Truth merging



Work in progress



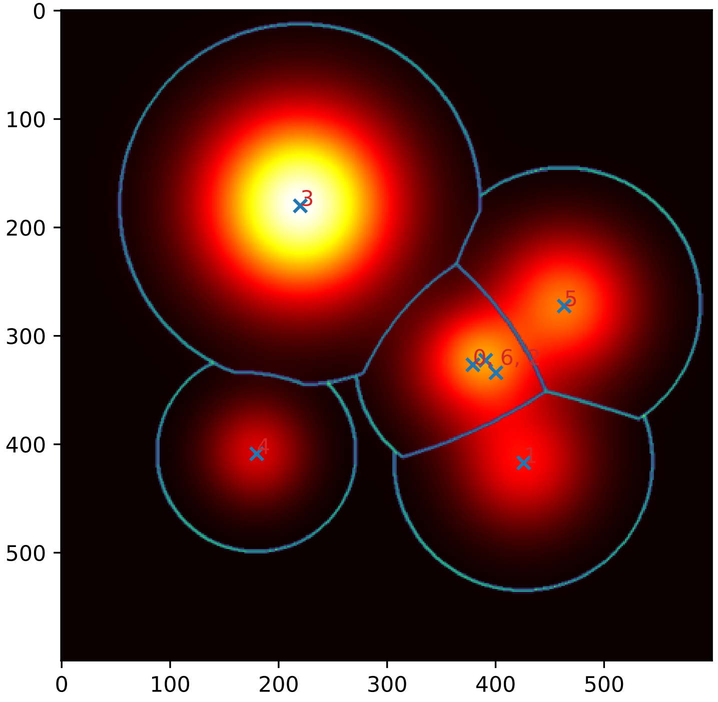
Based on resolvability w.r.t sensors: implement detector geometry in a generic way

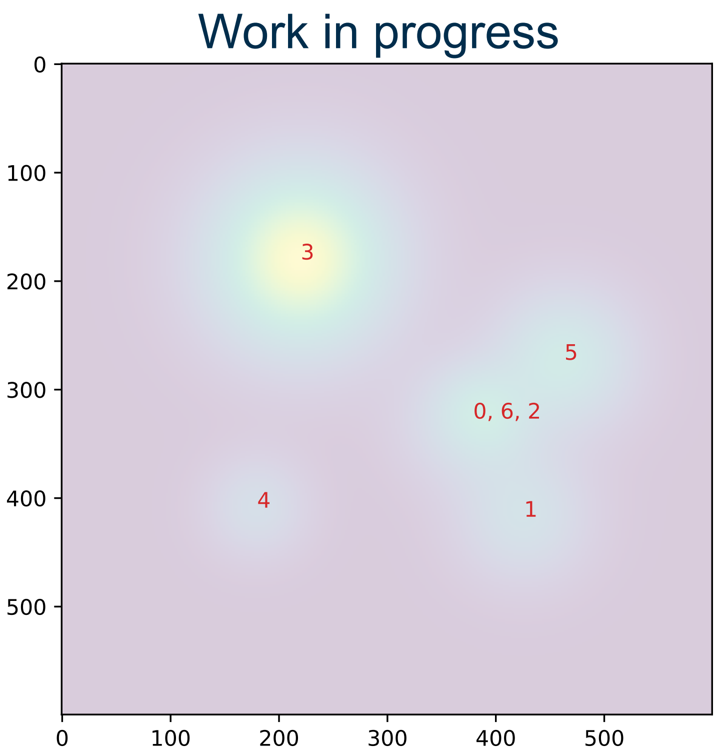
Creates a directed graph: avoids over-merging

Performs resolvability-ordered merging: adds IRC safety

A. Busamolino, KIT, paper in prep.

Truth merging





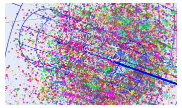
Based on resolvability w.r.t sensors: implement detector geometry in a generic way

Creates a directed graph: avoids over-merging

Performs resolvability-ordered merging: adds IRC safety

A. Busamolino, KIT, paper in prep.

Processing a large number of sensors: GravNet



Qasim, JK, et al arXiv:1902.07987

High input dimensionality

N^2 connections → local connections $N \cdot K$ - what are the ‘best’ connections?

Increase robustness

event topology prior → local particle *density* prior

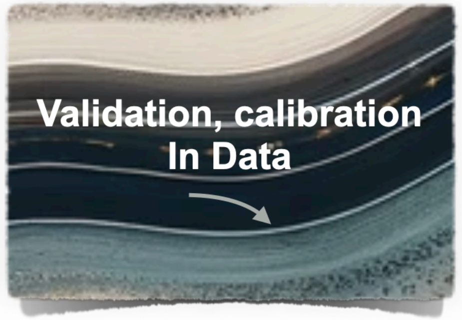
GravNet:

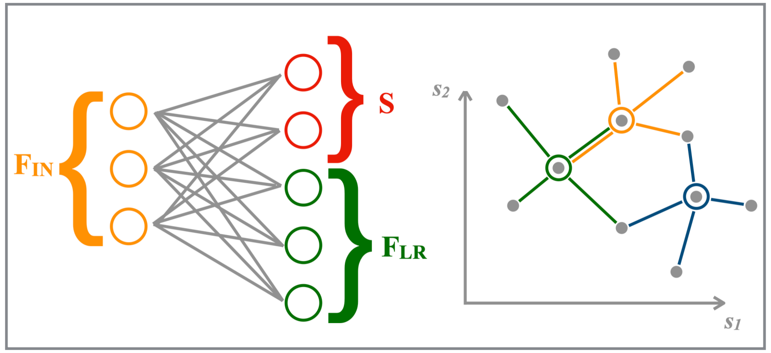
Distance weighted information exchange

$$x_j = \left(\sum_{i \,\epsilon\, N_j} \left(\exp(-d_{ij}^2)\, x_i \right),\ \max_{i \,\epsilon\, N_j} \left(\exp(-d_{ij}^2)\, x_i \right) \right)$$

✓ **explicit locality**

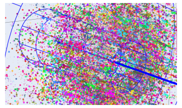
✓ **interpretable**





KIT

Processing a large number of sensors: GravNet



Qasim, JK, et al arXiv:1902.07987

High input dimensionality

N^2 connections → local connections $N \cdot K$ - what are the 'best' connections?

Increase robustness

event topology prior → local particle *density* prior

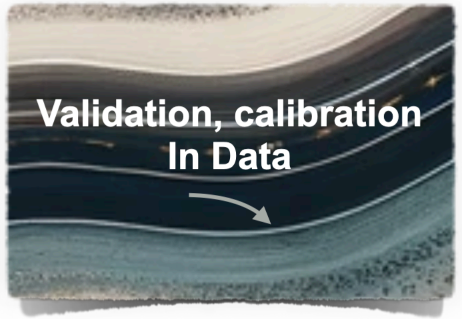
GravNet:

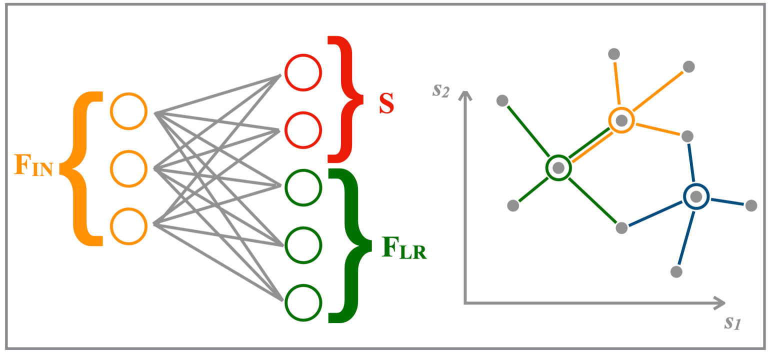
Distance weighted information exchange

$$x_j = \left(\sum_{i \,\epsilon\, N_j} \left(\exp(-d_{ij}^2)\, x_i \right),\ \max_{i \,\epsilon\, N_j} \left(\exp(-d_{ij}^2)\, x_i \right) \right)$$

✓ explicit locality

✓ interpretable



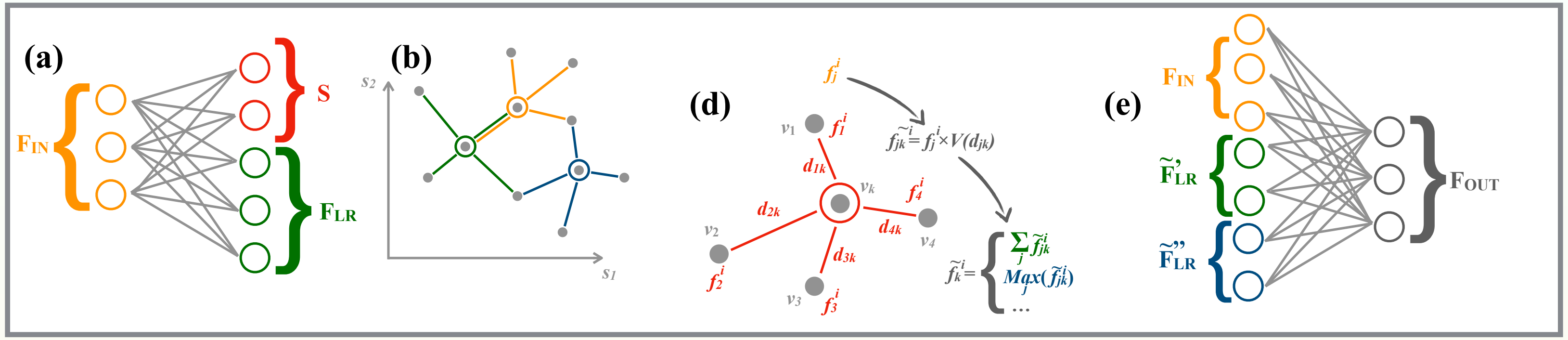


➡Explicitly learns the graph topology

➡Saves 10-100x in terms or resources w.r.t. similar approaches

➡Adaptive explicitly translation equivariant variants being actively developed

GravNet: a faster EdgeConv/DGCNN?



- Instead of generic message passing, use powerful attention

$$y_j = \underset{i \,\epsilon\, N(j)}{\square} a(x_j, x_i) \cdot \mathrm{MLP}(x_i)$$

Does not depend on j

EdgeConv / DGCNN

$$y_j = \underset{i \,\epsilon\, N(j)}{\square} MLP(x_j, x_j - x_i)$$

Per edge MLP:
N x K x F

- Can be rewritten

Per-vertex MLP: N x F

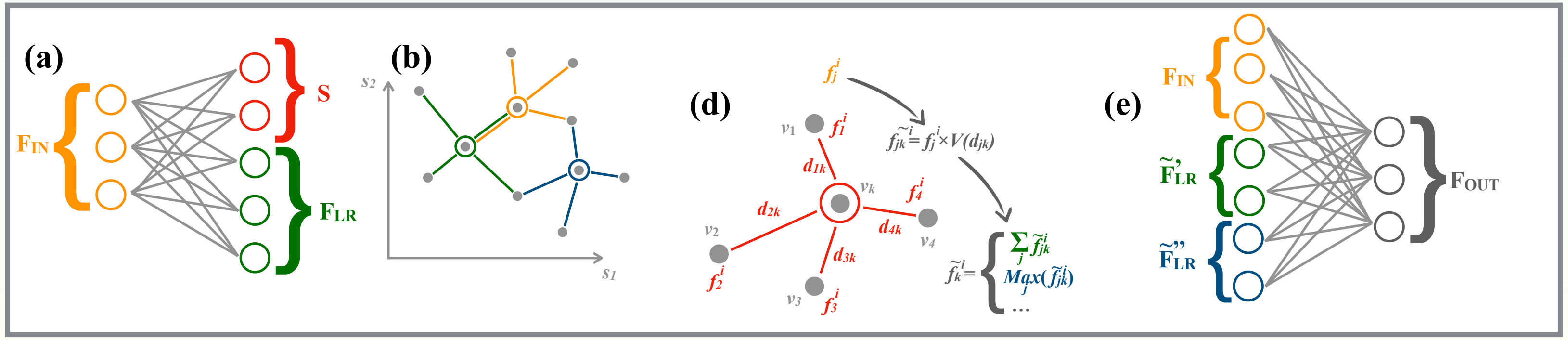
$$\tilde{x}_i = \mathrm{MLP}(x_i)$$

$$y_j = \underset{i \,\epsilon\, N(j)}{\square} a(x_j, x_i) \cdot \tilde{x}_i$$

- Brings a factor of about factor 100 speedup

arXiv:1902.07987

Building the graph topology through attention



- Attention scales the information exchange: higher value → more exchange

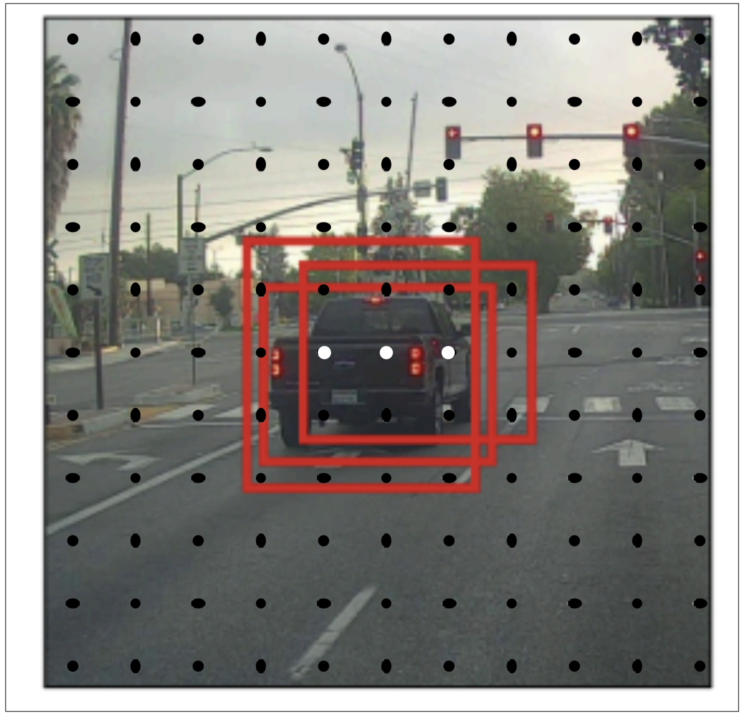
$$y_j = \underset{i \in N(j)}{\square} a(x_j, x_i) \cdot \tilde{x}_i$$

- Learn an embedding space S for each point (low dimensional: fast)
- Chose $a(x_i, x_j)$ such that close-by means a lot of information exchange:

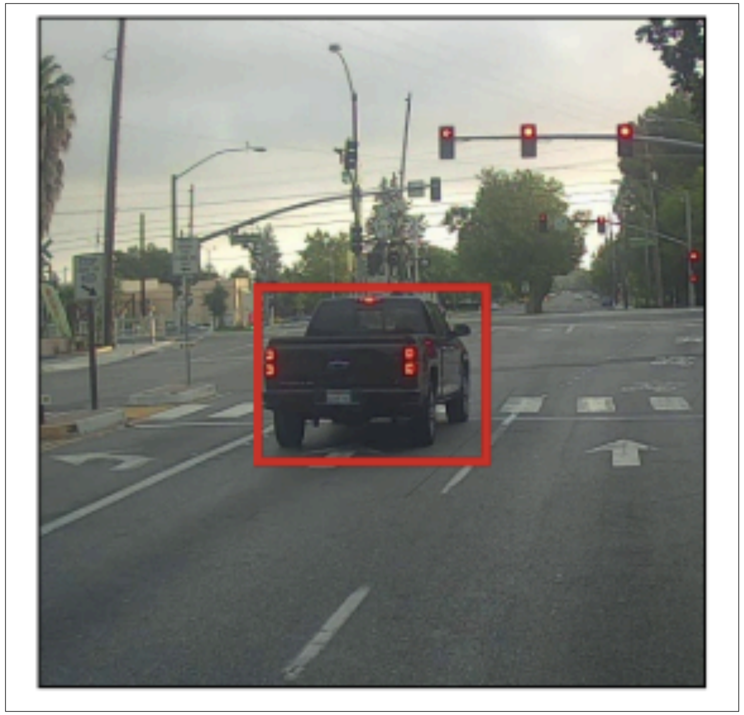
$a(x_i, x_j) = \exp(-10 \cdot d_{ij}^2)$: **distance weighted**

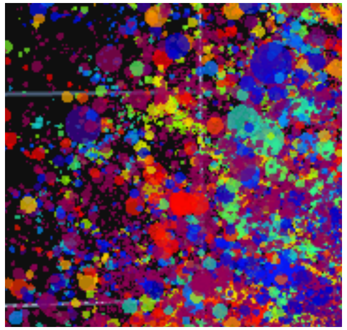
➡The K nearest points are the ones that need to exchange the most information: graph topology build

Multi-object detection: A look at computer vision



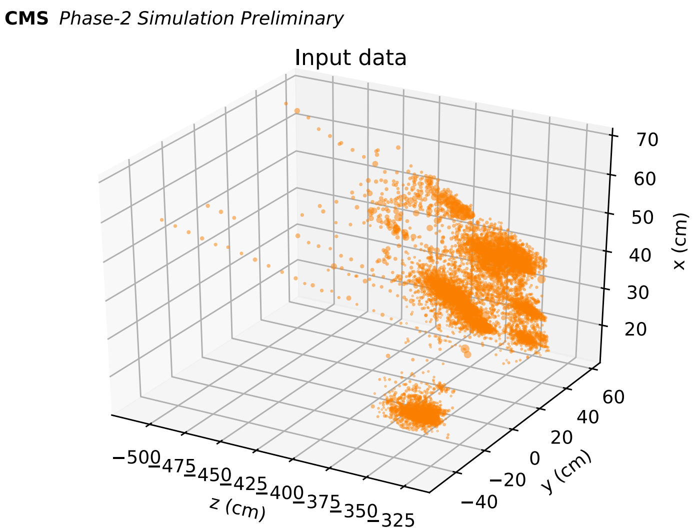
Overlap removal





Multi-Particle Detection: Object Condensation

J.K., EPJC 2020



Hit / track

Position, energy, **time, …**

Hit / track

Hit / track

GravNet

Repr. score β

Particle properties

Cluster Coordinates

Repr. score β

Particle properties

Cluster Coordinates

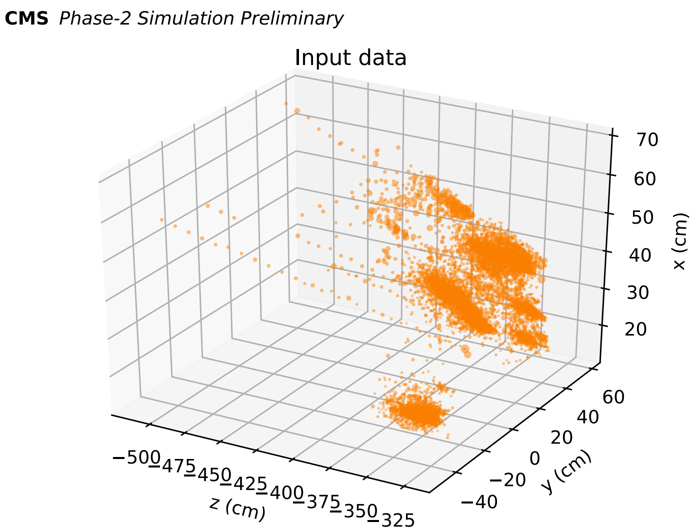
Repr. score β

Particle properties

Cluster Coordinates

Multi-Particle Detection: Object Condensation

J.K., EPJC 2020



At least one point with high $\beta\ \epsilon[0,1]$ per object.
$\rightarrow$ semi-supervised

Noise should have low β

Hit / track → GravNet → Repr. score β / Particle properties / Cluster Coordinates

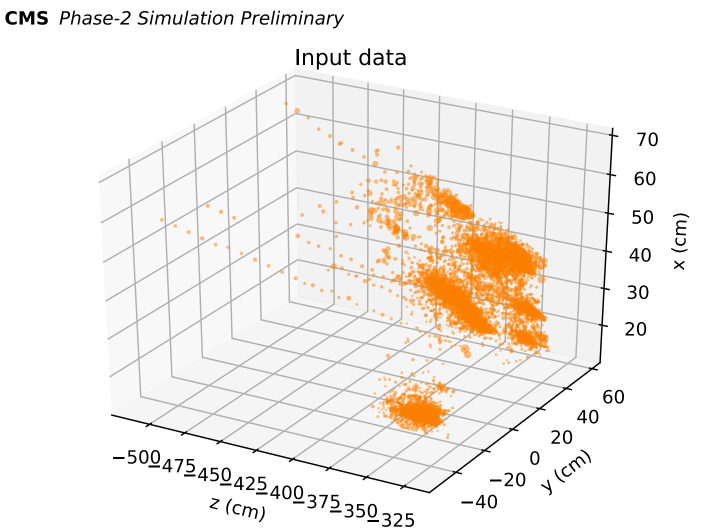
Position, energy, **time, …**

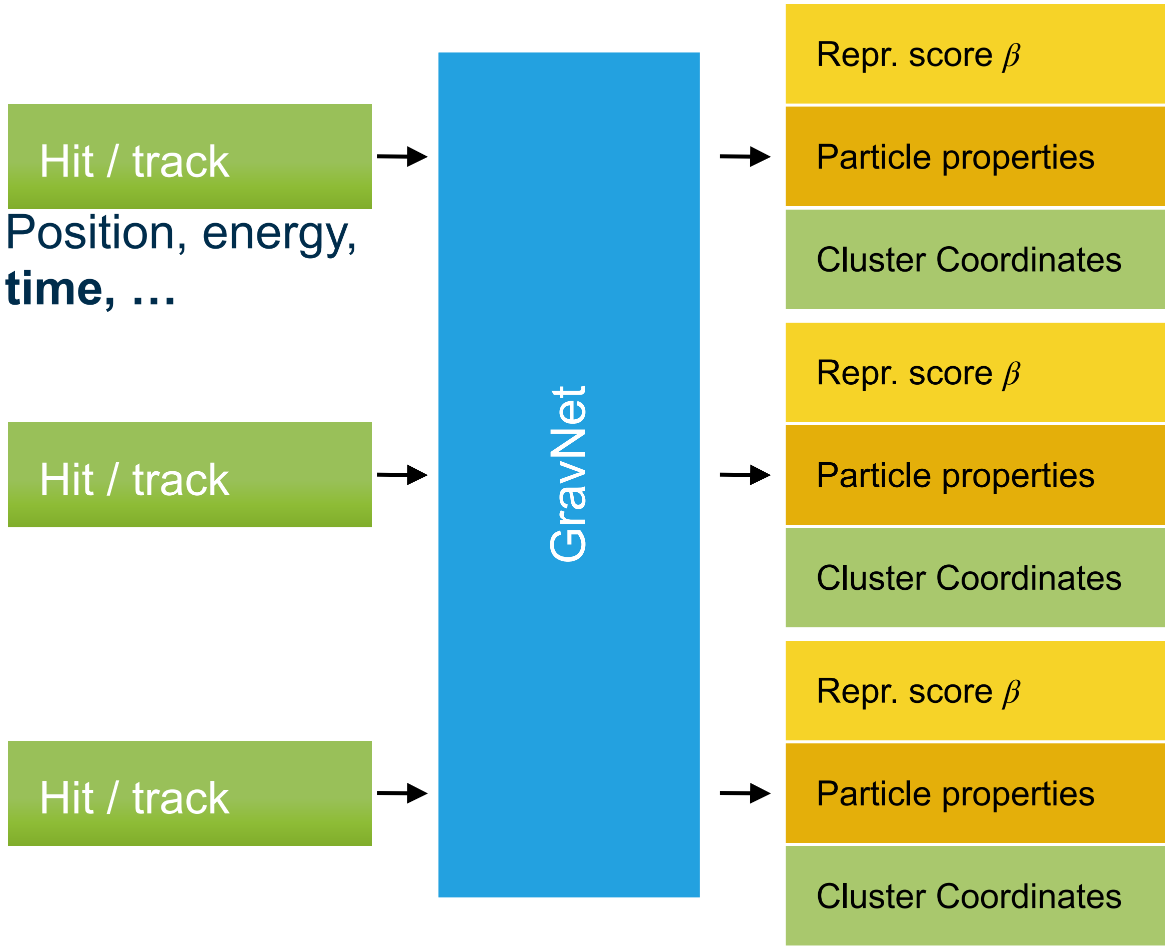
Hit / track → GravNet → Repr. score β / Particle properties / Cluster Coordinates

Hit / track → GravNet → Repr. score β / Particle properties / Cluster Coordinates

Multi-Particle Detection: Object Condensation

J.K., EPJC 2020



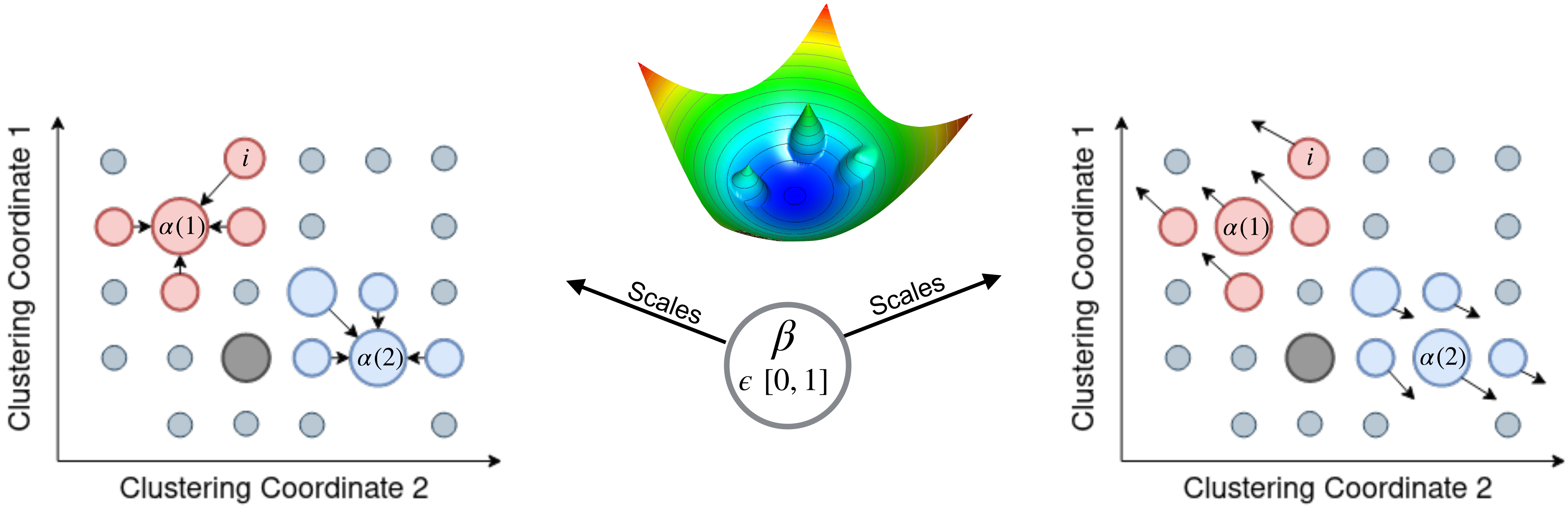


At least one point with high β $\epsilon[0, 1]$ per object.
$\rightarrow$ semi-supervised

Noise should have low β

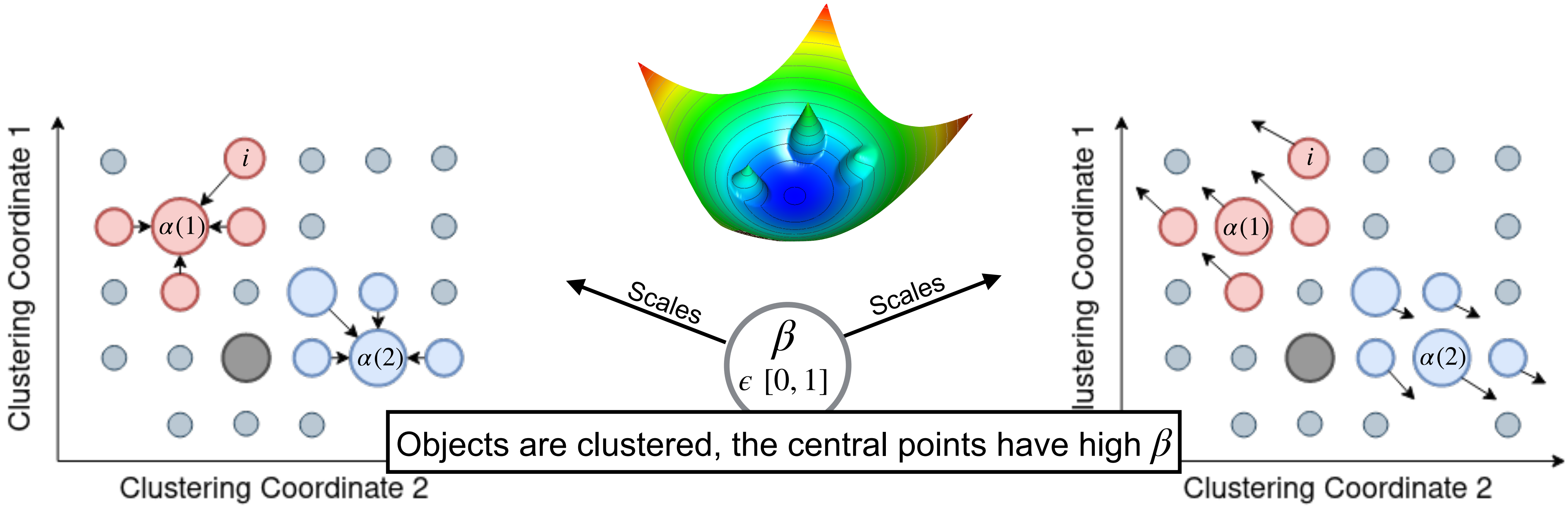
The condensation score β is central

Simultaneous clustering and property prediction



JK, arxiv:2002.03605, EPJC

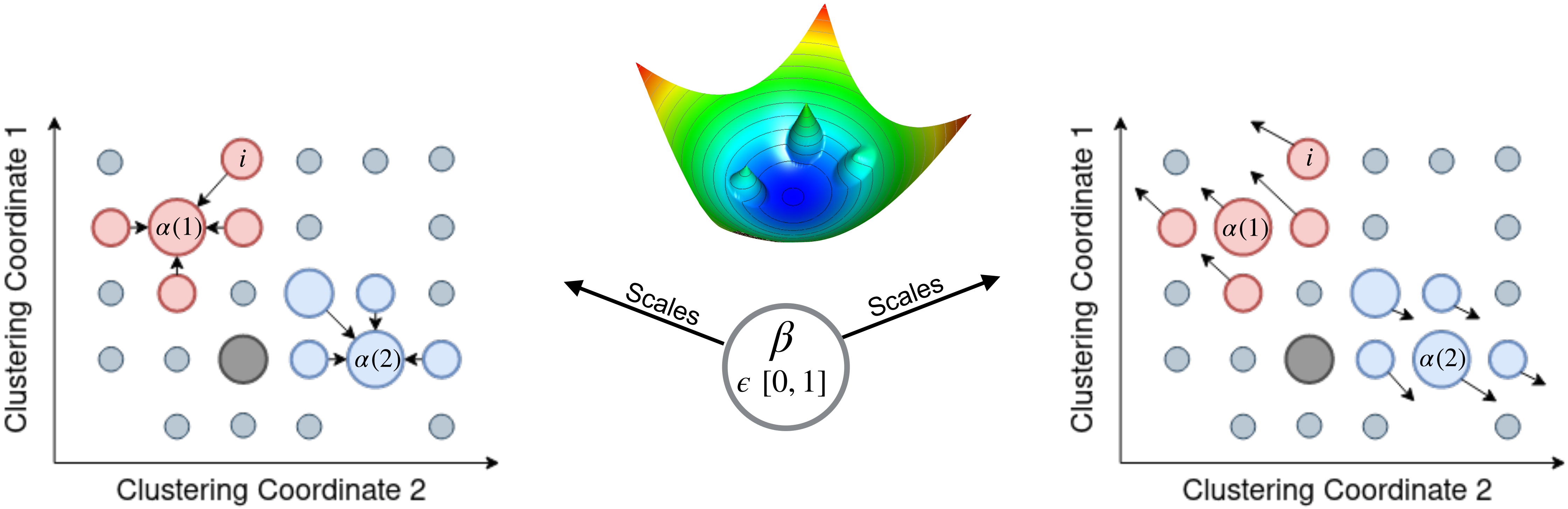
Simultaneous clustering and property prediction



Objects are clustered, the central points have high β

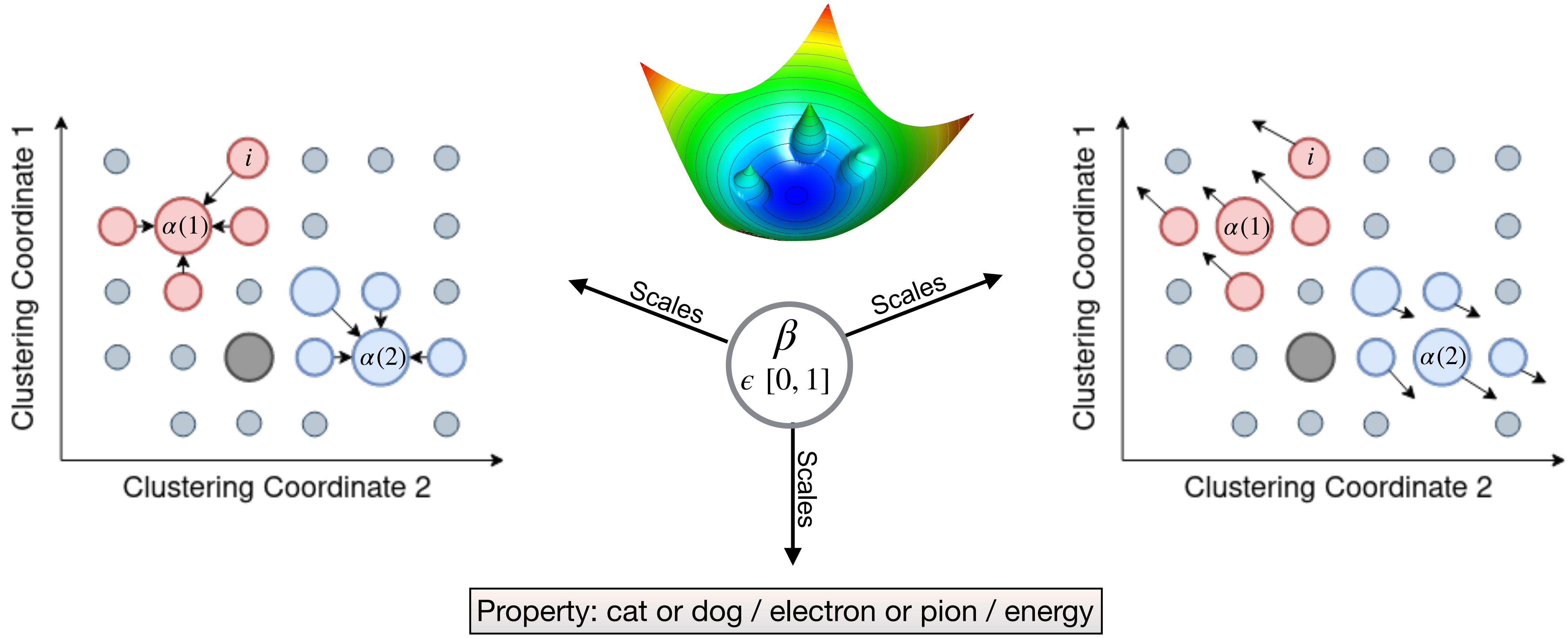
JK, arxiv:2002.03605, EPJC

Simultaneous clustering and property prediction



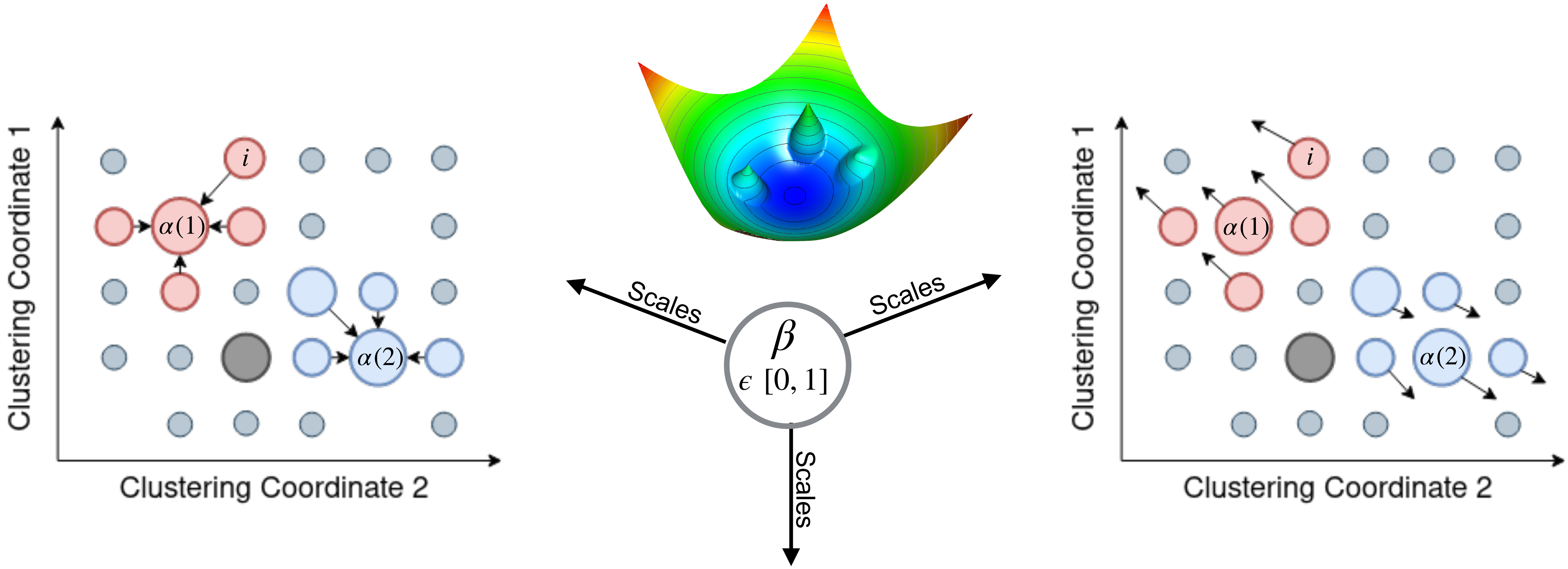
JK, arxiv:2002.03605, EPJC

Simultaneous clustering and property prediction



JK, arxiv:2002.03605, EPJC

Simultaneous clustering and property prediction

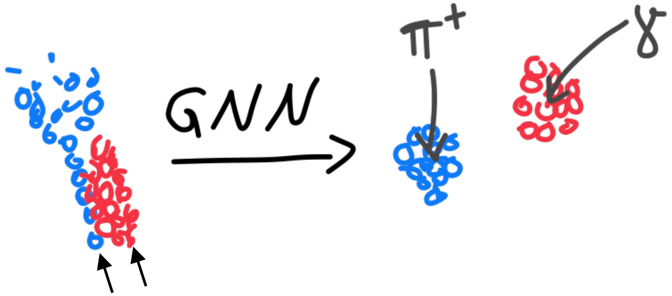


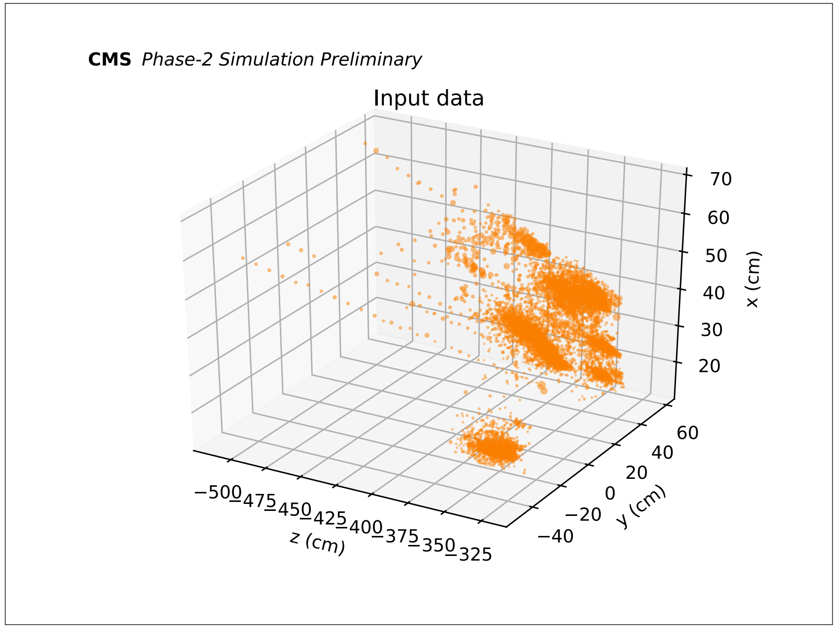
Property: cat or dog / electron or pion / energy

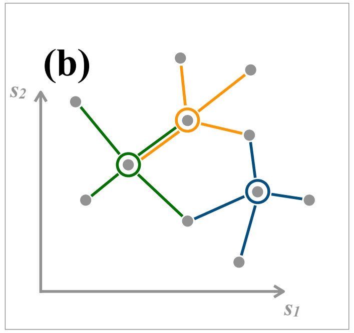
Center of each cluster also carries best property estimate: condensation point

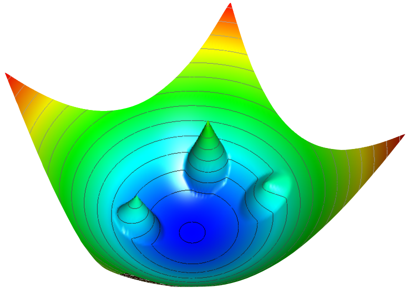
JK, arxiv:2002.03605, EPJC

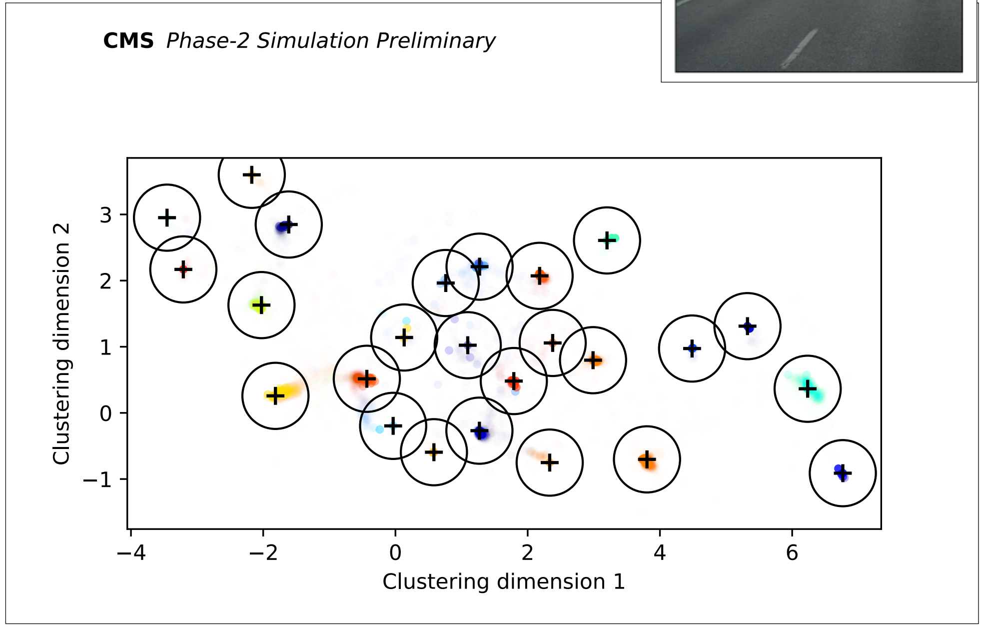
Object Condensation: Transform a complex problem into a simple one









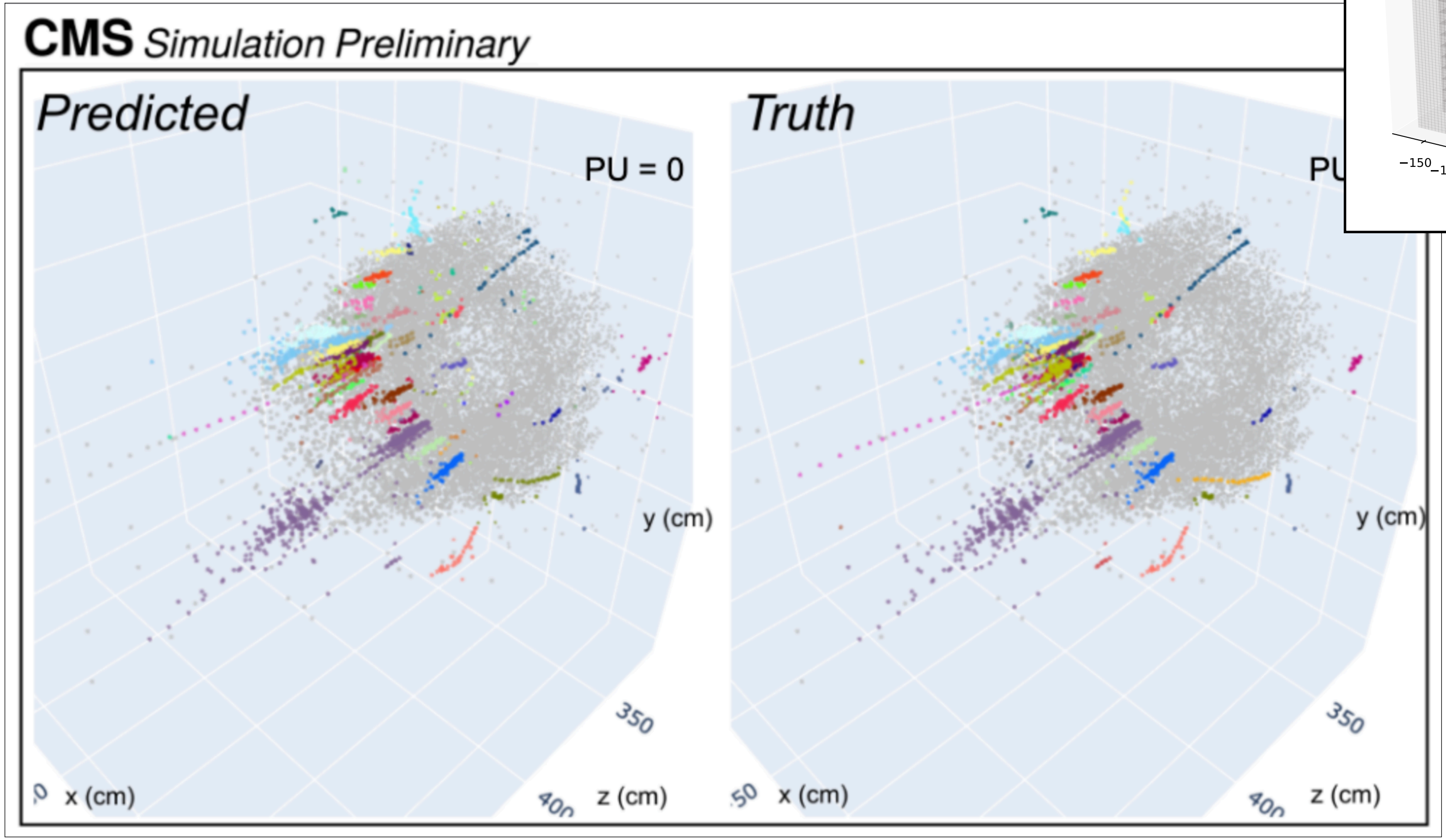


CMS DP-20/004

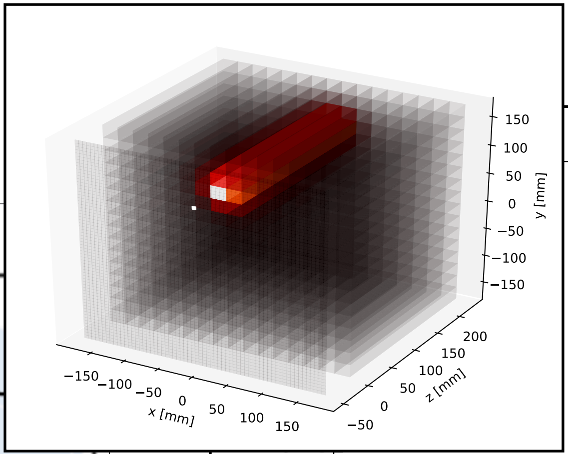
➡ Also loss is explicitly using local densities

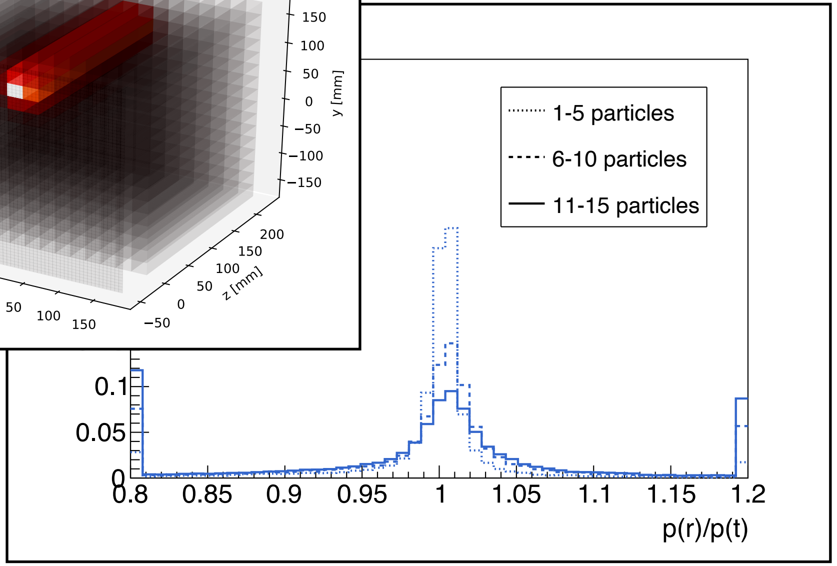
➡ However: still hand-tuned inference clustering/masking, no adaptive dimensionality reduction… there is still work to do!

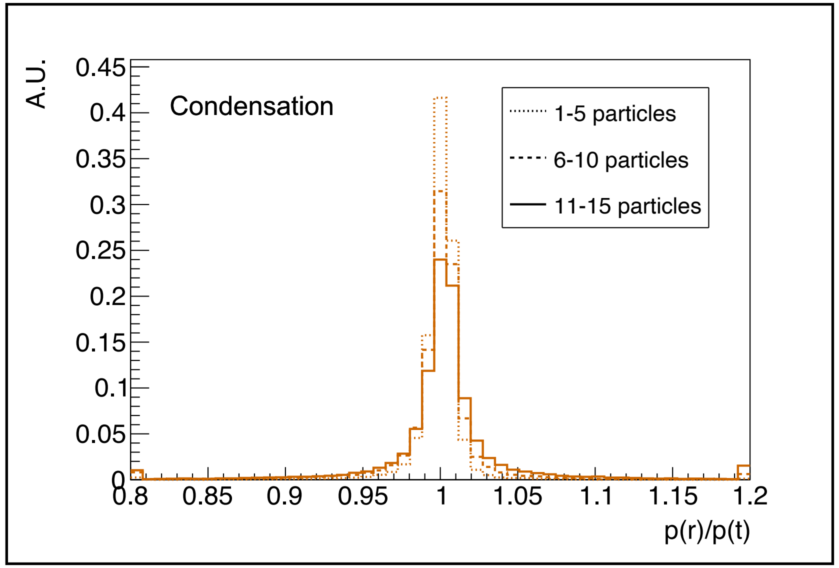
Large gains possible



CMS DP-22-001





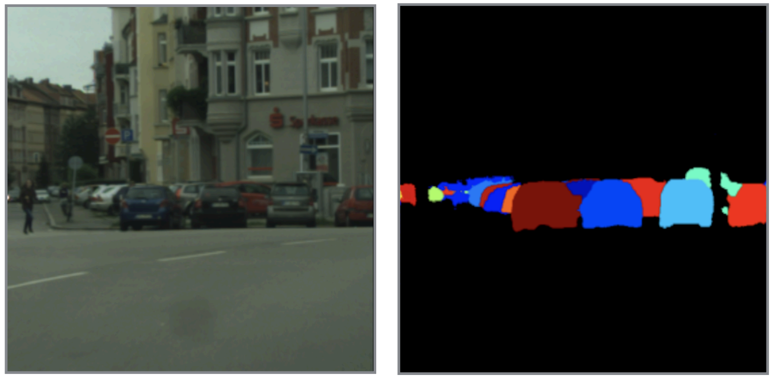


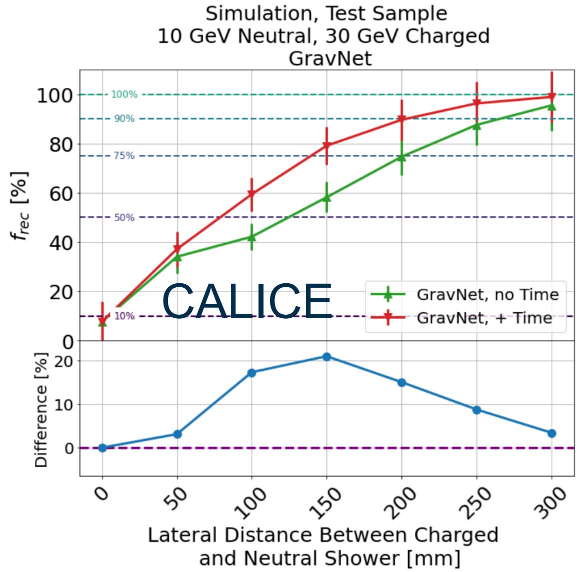
JK, arxiv:2002.03605, EPJC

➡ Very promising results

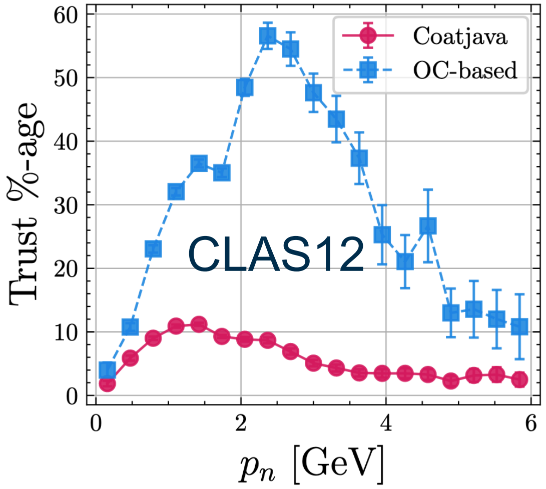
S.R. Qasim, N. Chern. J.K., arXiv:2204.01681,
S. Battarchaya, .. , J.K., et al., arXiv:2203.01189,
S. Qasim, K. Long, J.K., et al., arXiv:2106.01832
I. Iiyama, .. J.K., et al, arXiv:2008.03601

Very versatile



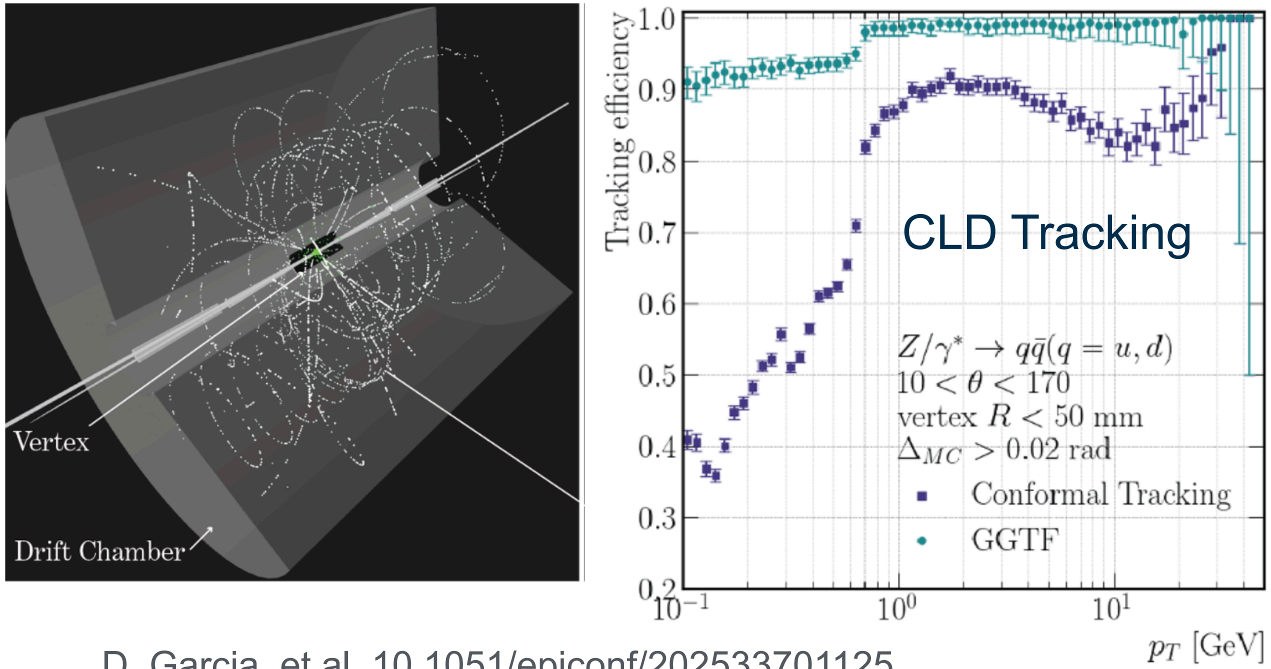


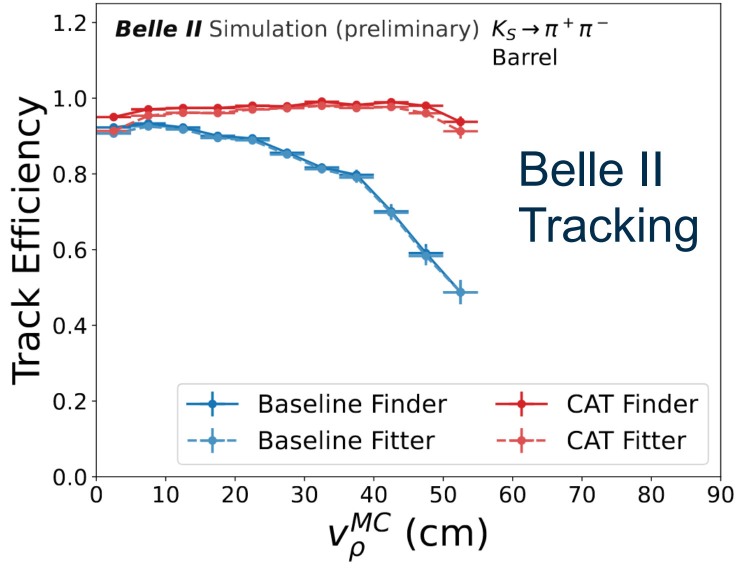
S. Lai, et al, arXiv:2407.00178



G. Matousek, V. Anselm, arxiv:2503.11277

L. Reuter, et al, arXiv:2411.13596, PRD

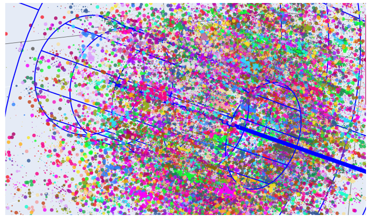




(d) $K_S^0 \to \pi^+\pi^-$, v_ρ^{MC}.

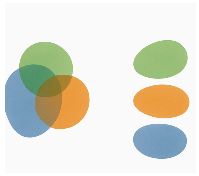
D. Garcia, et al, 10.1051/epjconf/202533701125

From Hits to Physics with ML

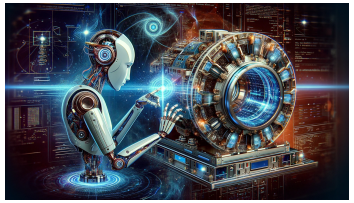


A large territory only being explored since recently

It comes with conceptual and technical challenges that are (being) resolved
Locality, Sampel priors, Robustness



Huge opportunities:



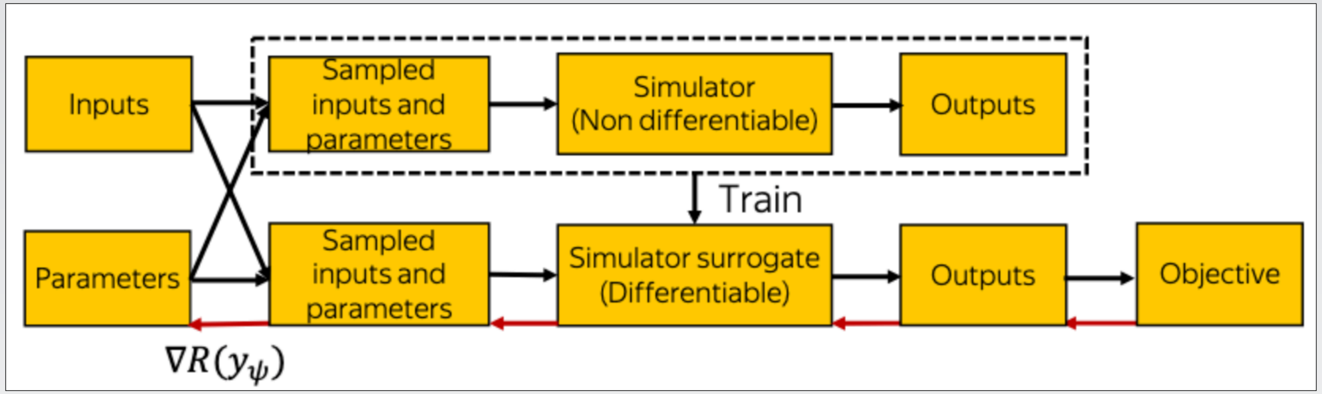
Possibility to tackle computing resource challenges

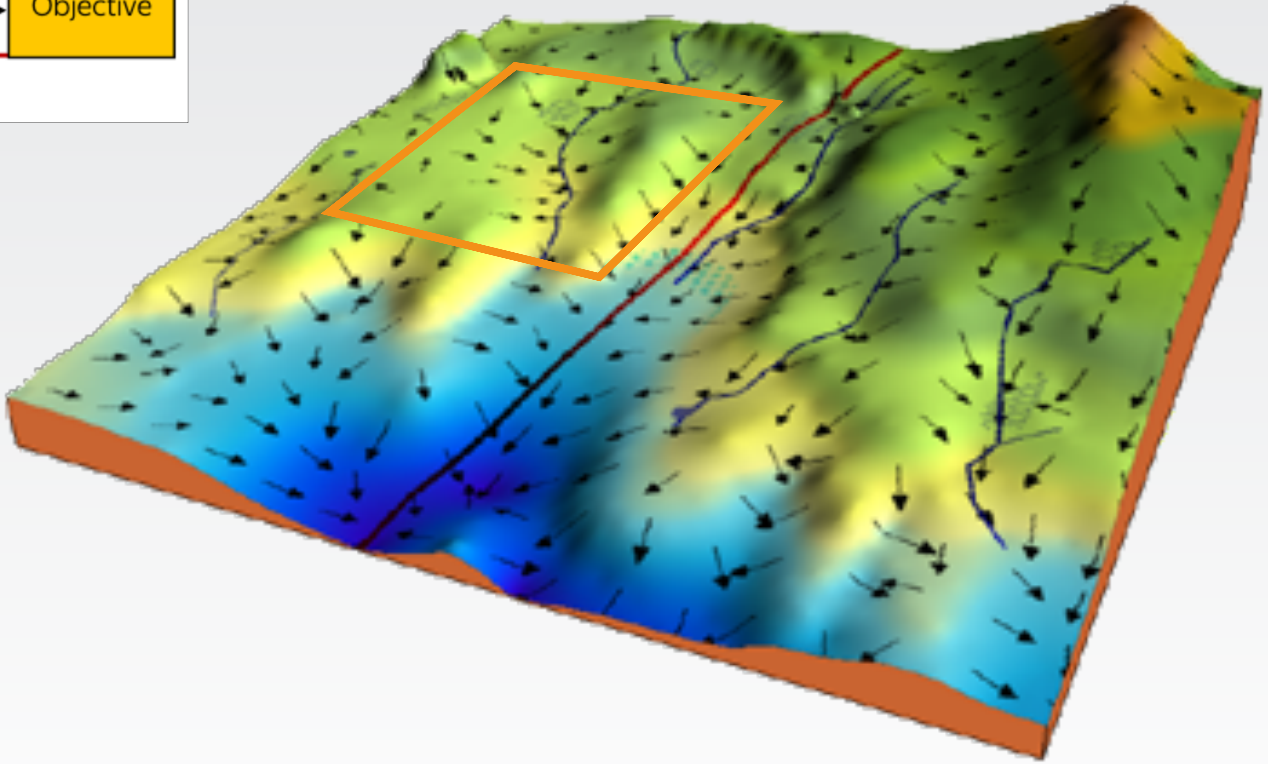
Low development, turn-around, and maintenance cost even on (novel) heterogenous hardware or different detector geometries

All signs point to better physics performance with tremendous lever arm for existing detectors and plenty of opportunities for developing new detectors

BACKUP

A Local Surrogate

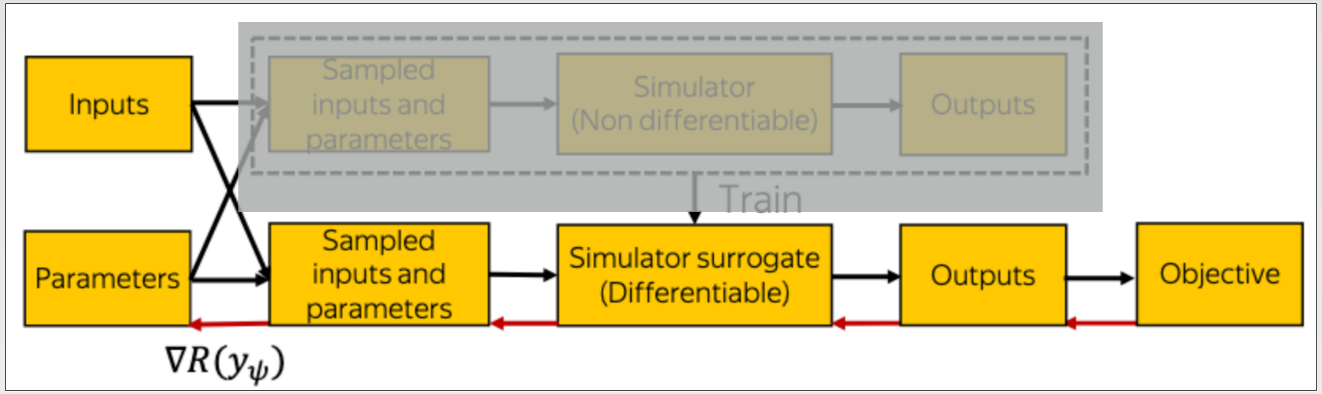


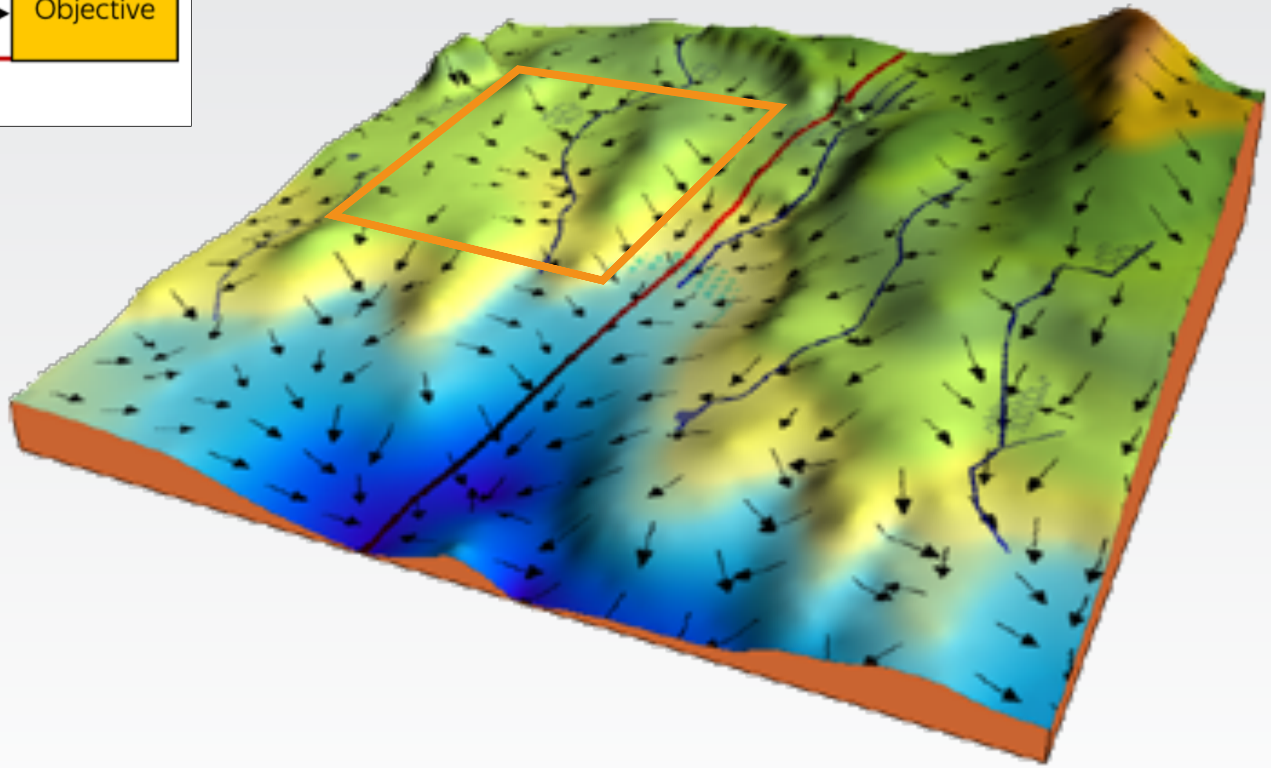


https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate

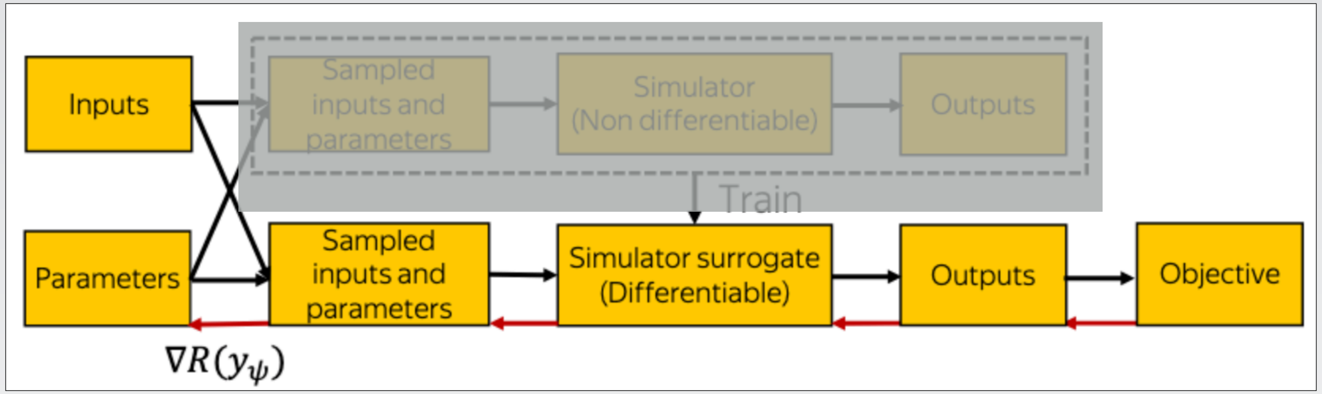


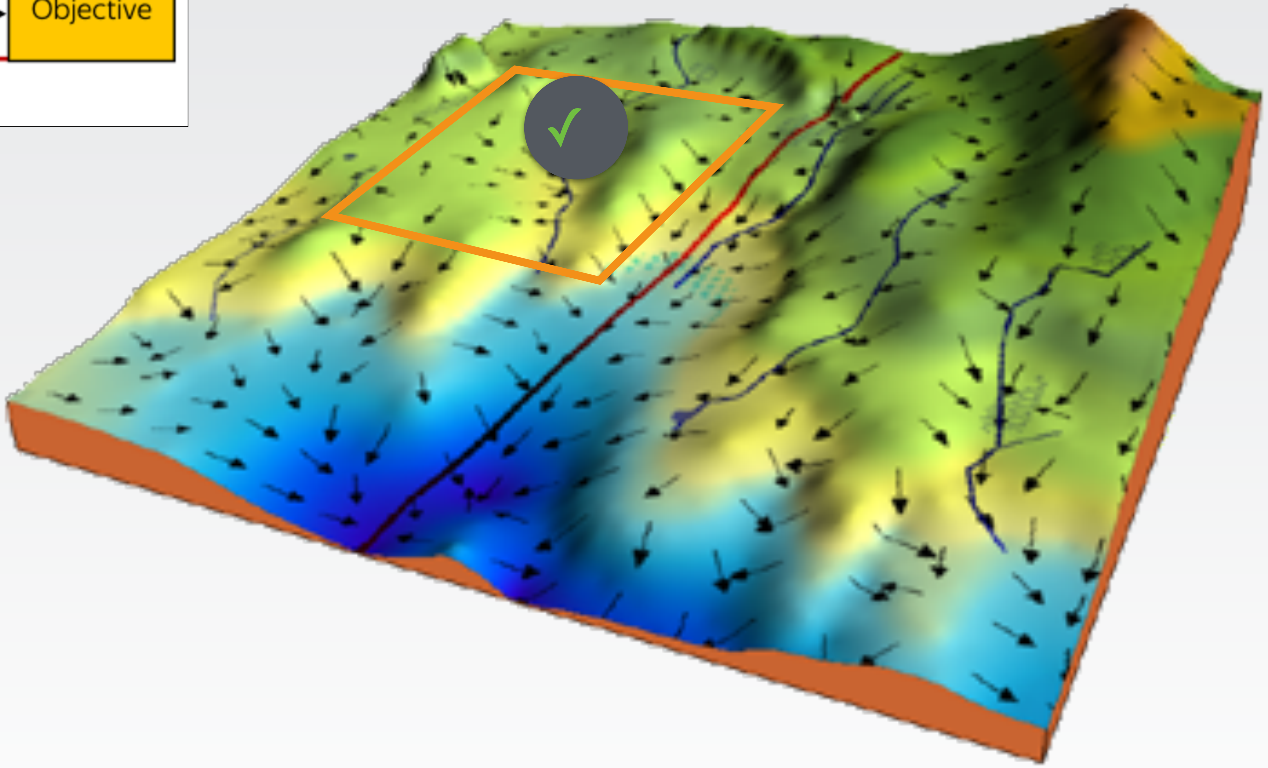


https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate

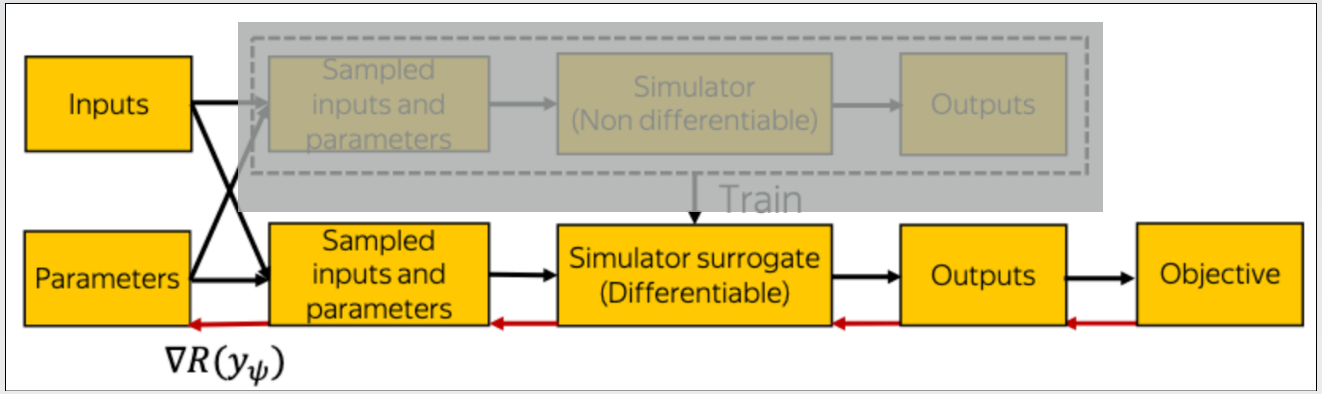


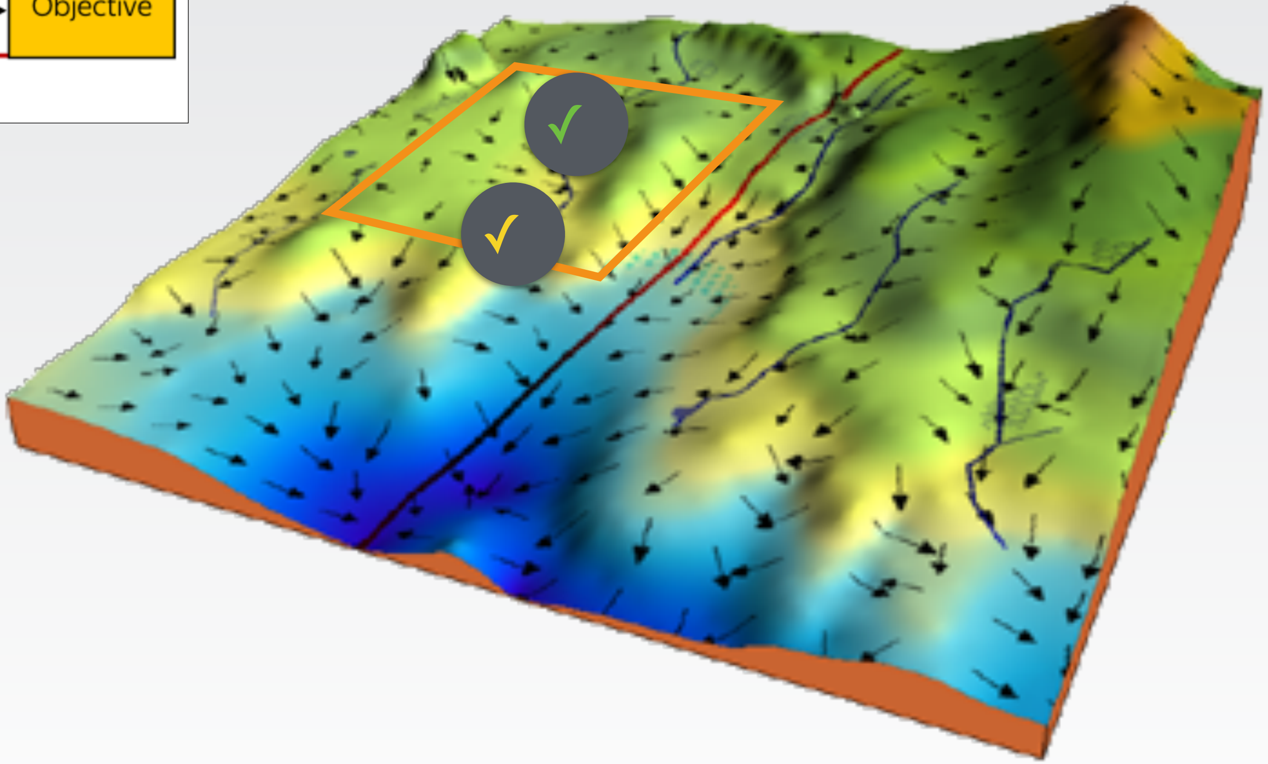


https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate

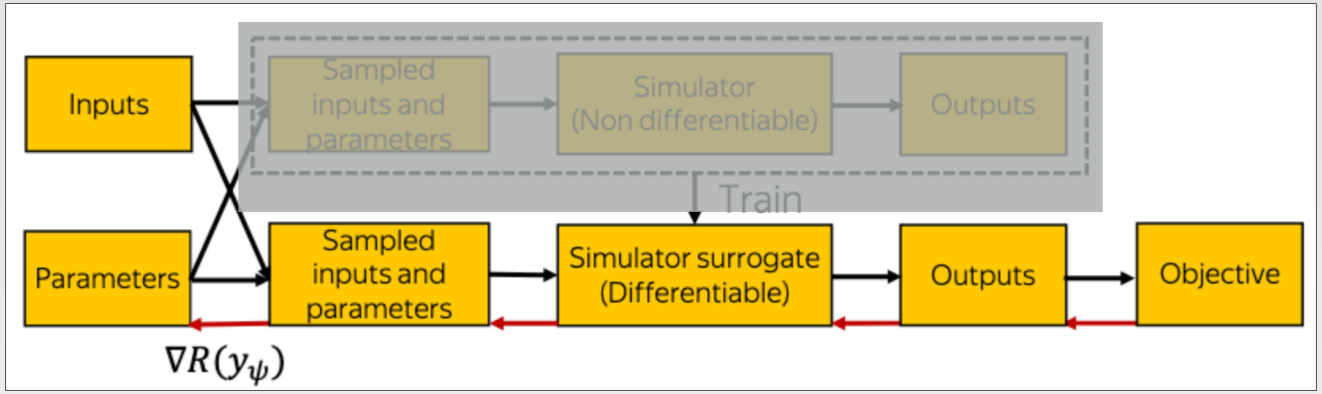


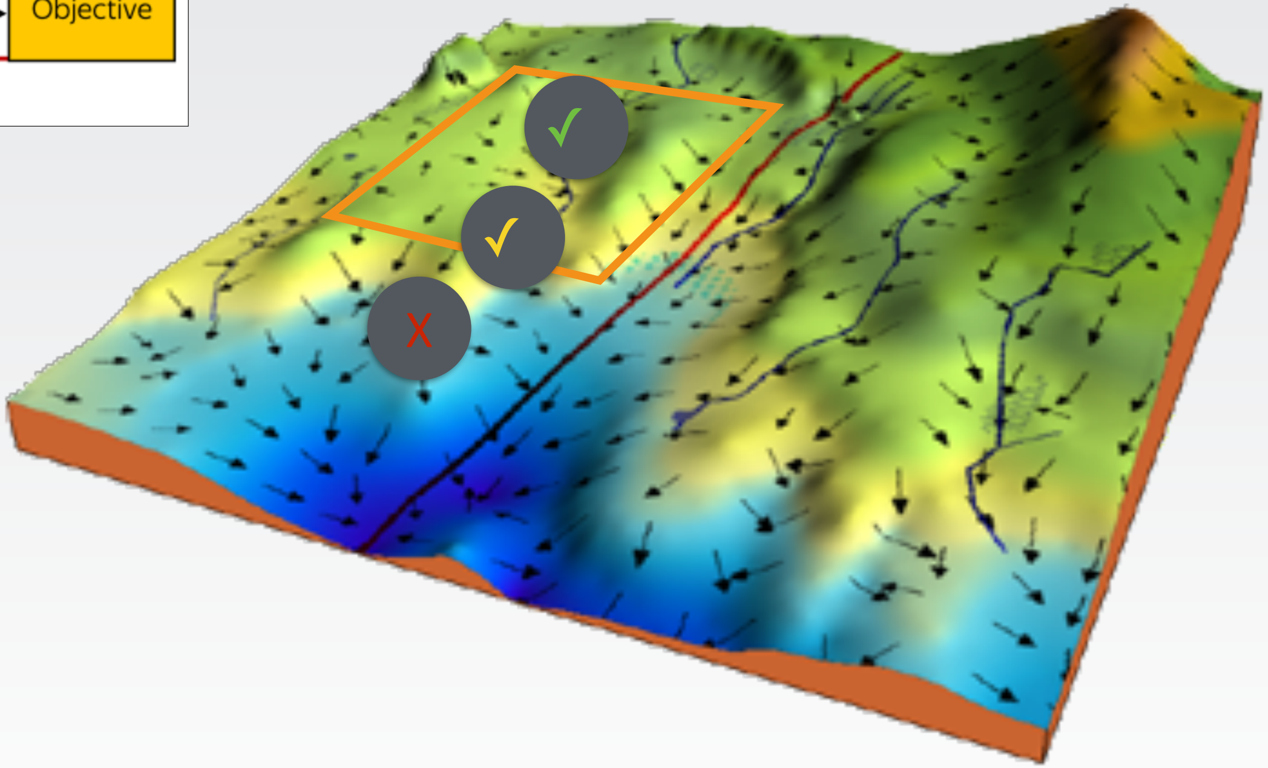


https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate

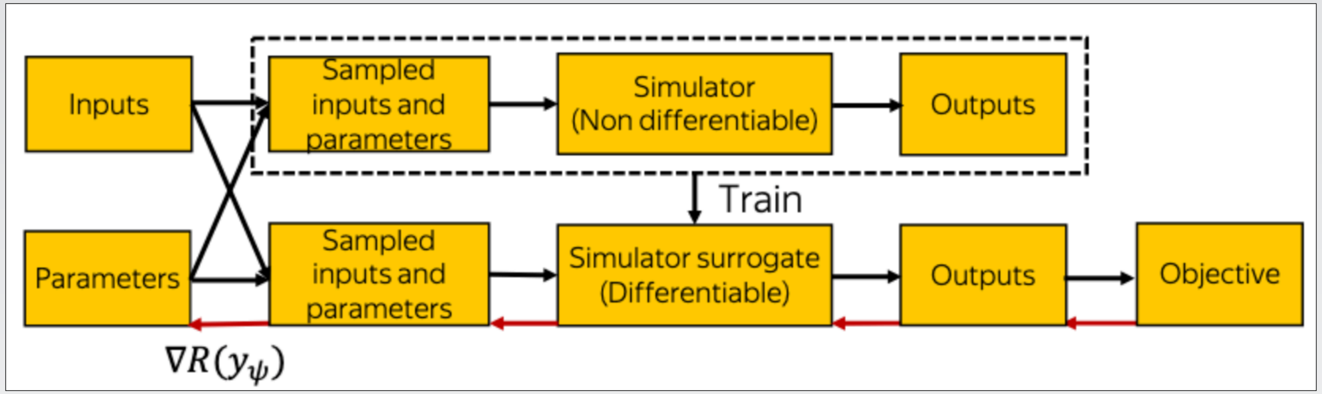


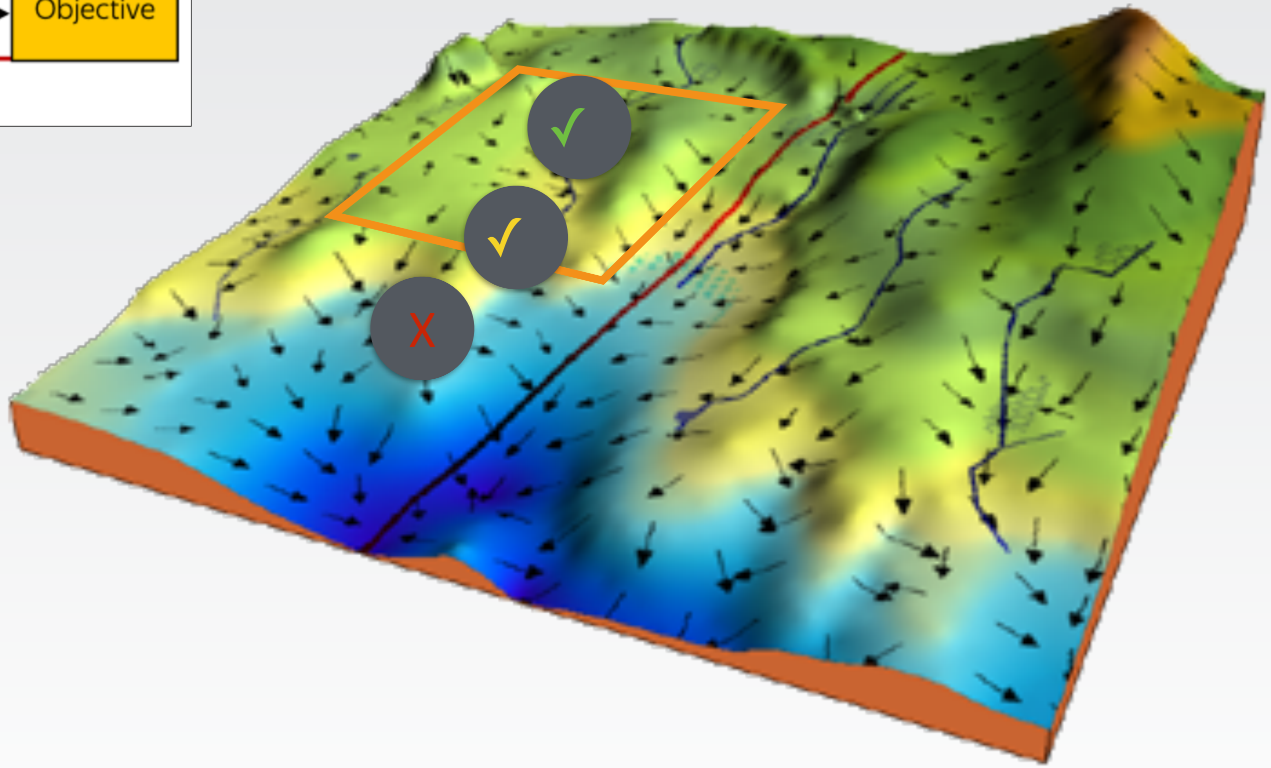


https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate

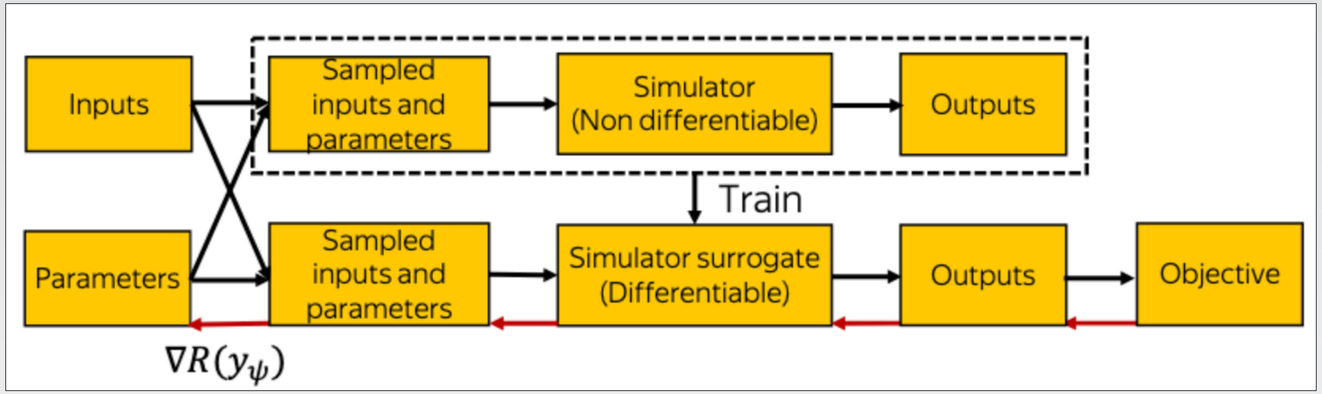


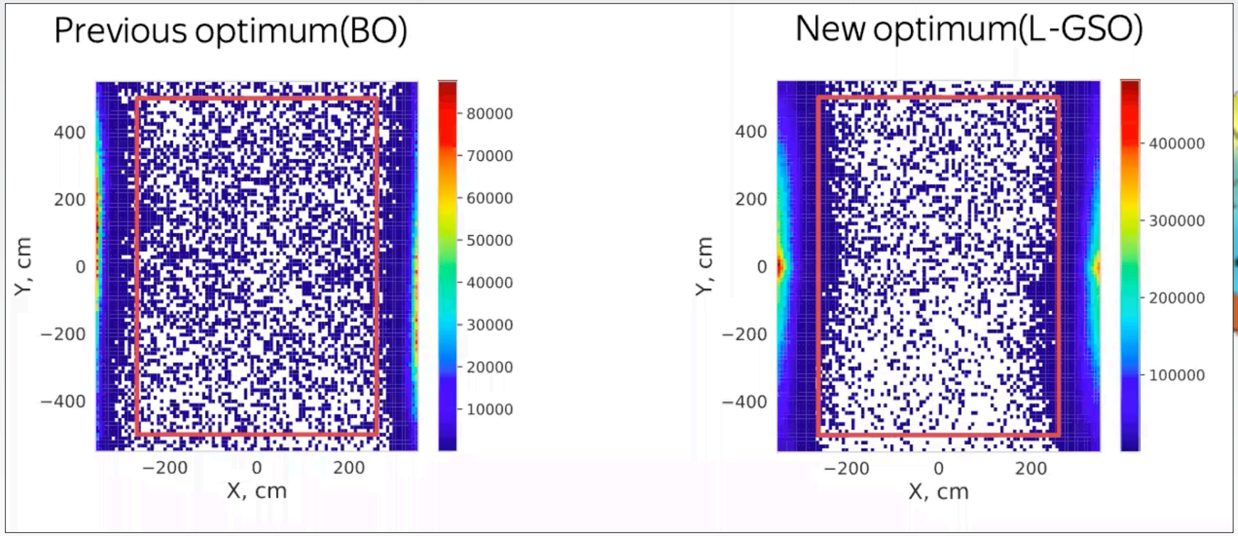


https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

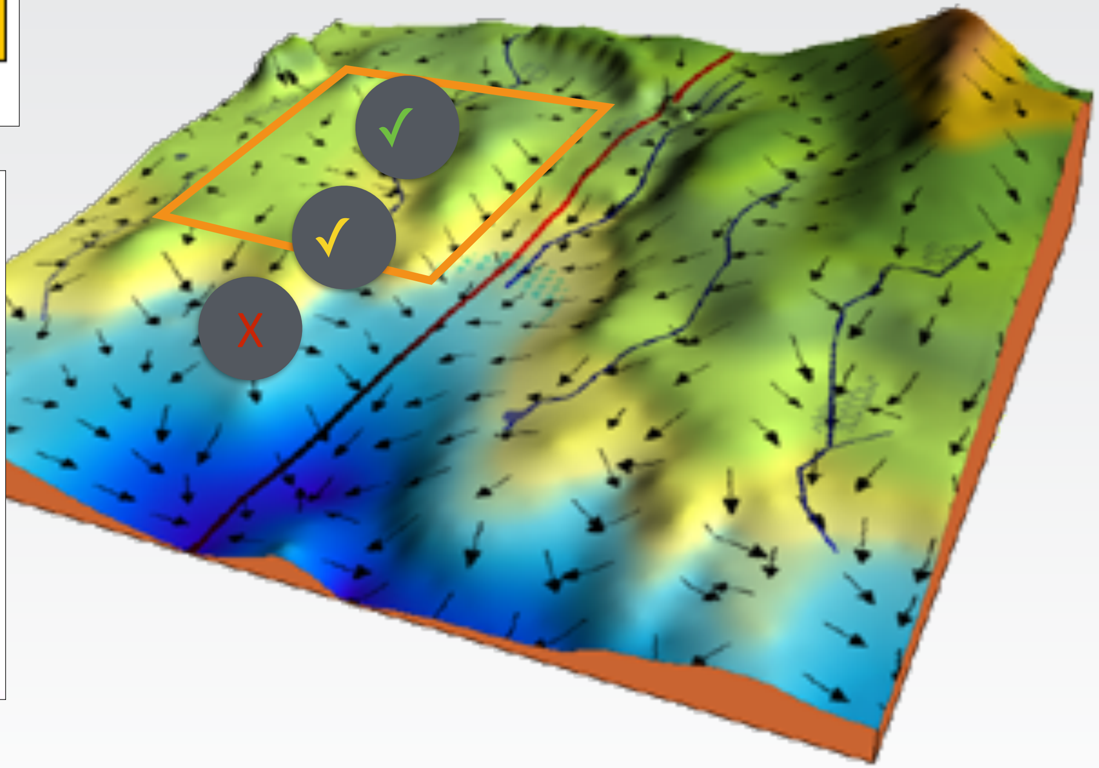
S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate





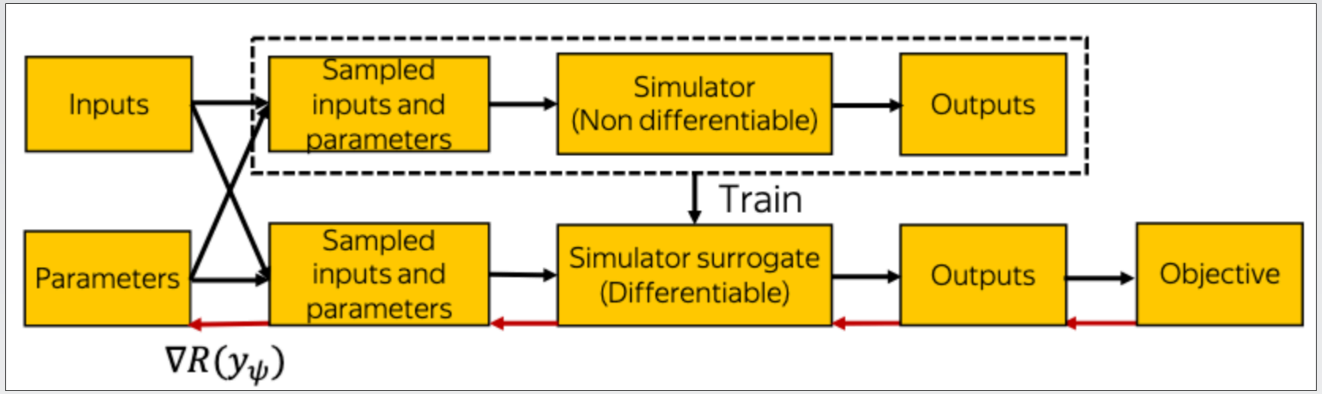
Approx. 30% improvement with a 30% lighter magnet

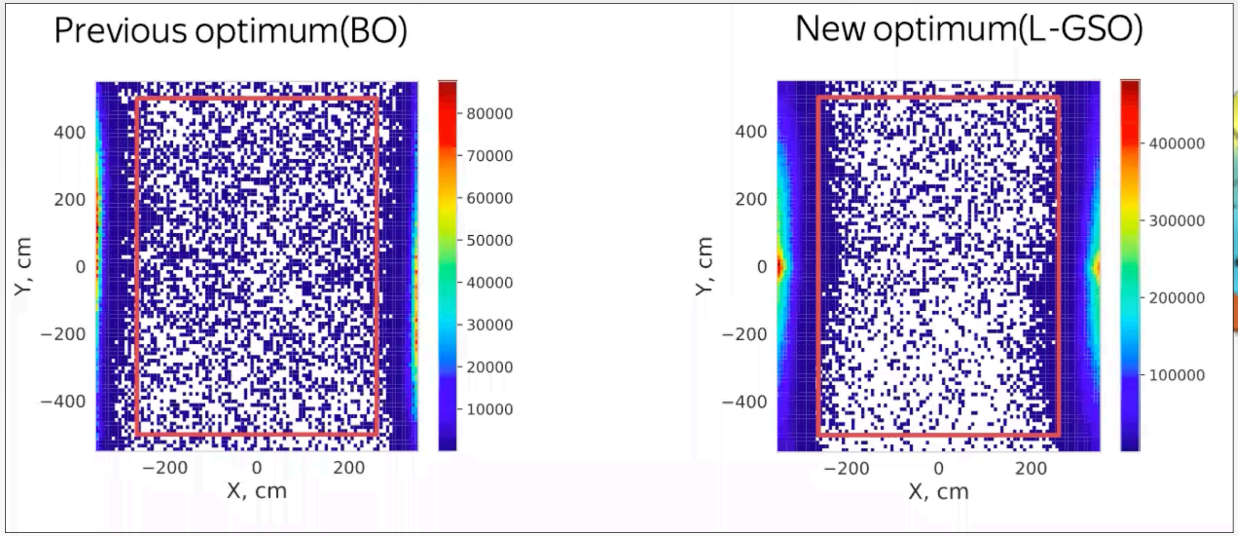


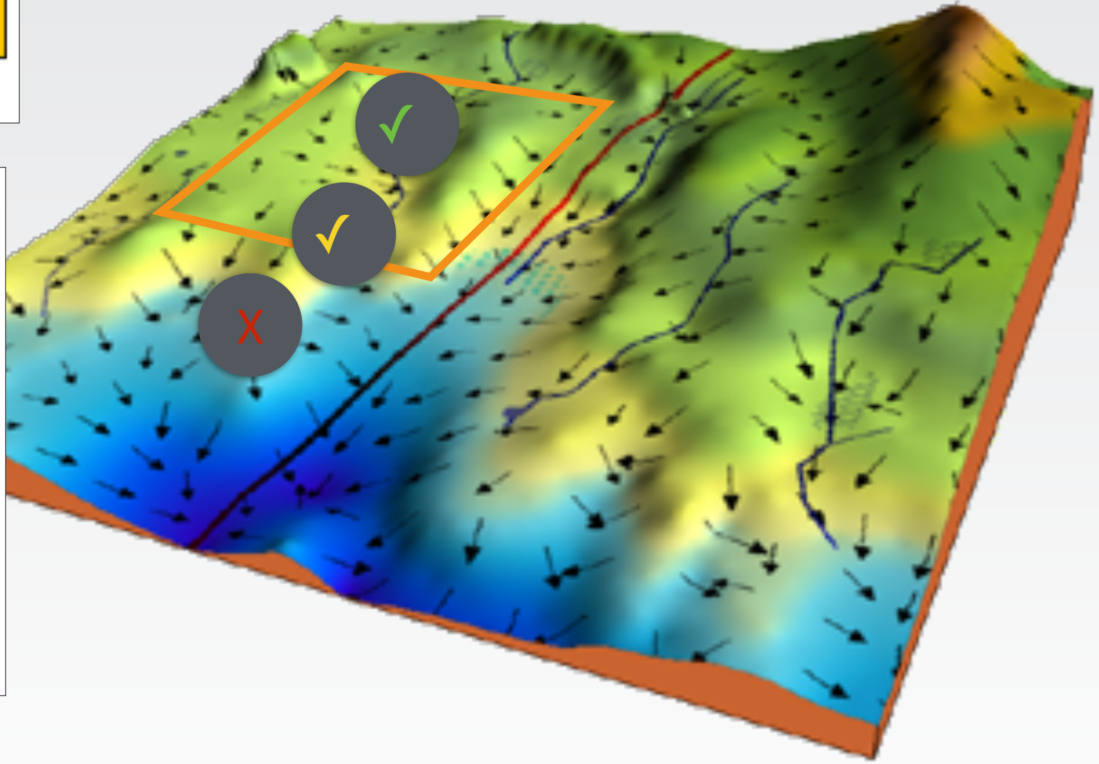
https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html

S. Shirobokov, et al., arXiv:2002.04632

A Local Surrogate







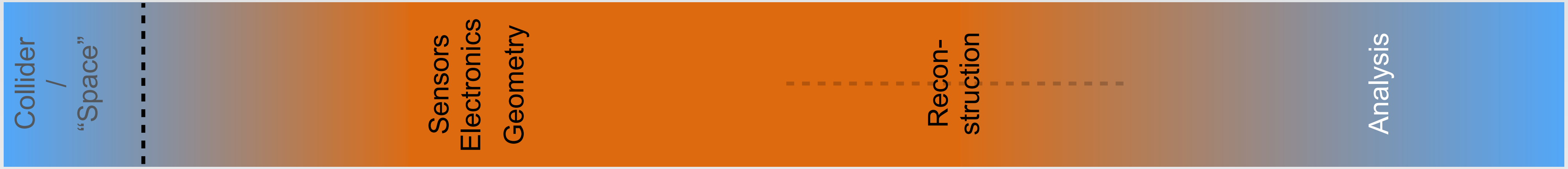
Approx. 30% improvement with a 30% lighter magnet

Local surrogates are a powerful tool to work around non-differentiable parts of the chain

S. Shirobokov, et al., arXiv:2002.04632

latest/gradient_descent.html

Make Wider Steps?

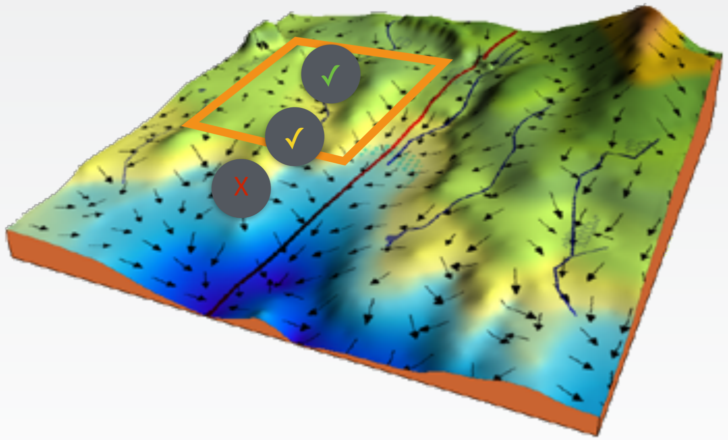


A few quantities

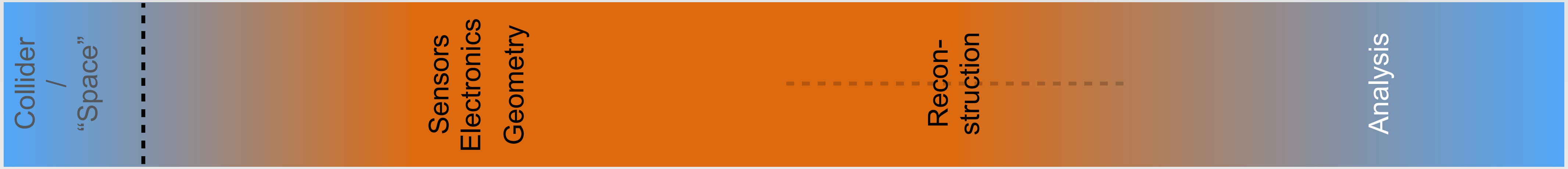
Many pixels, detector elements, hits

A few quantities

Very complex surrogate model



Make Wider Steps?



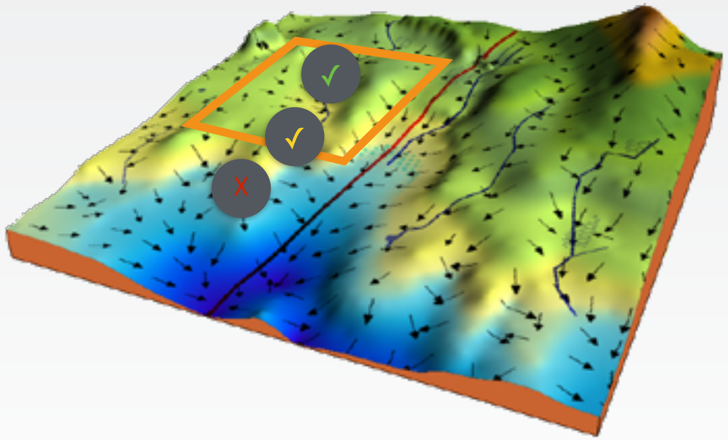
A few quantities

Many pixels, detector elements, hits

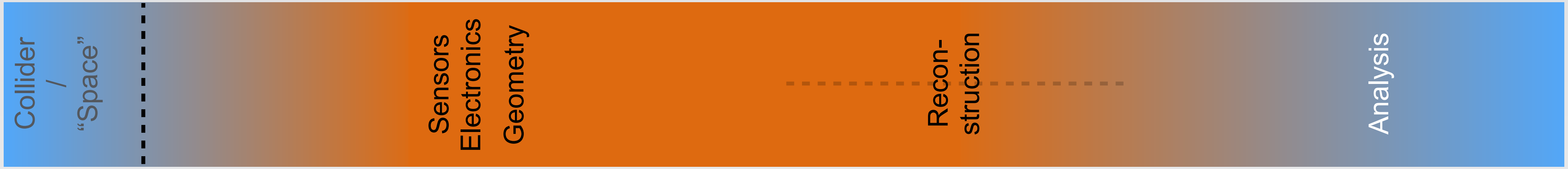
A few quantities

Very complex surrogate model

✓ Much simpler surrogate model



Make Wider Steps?



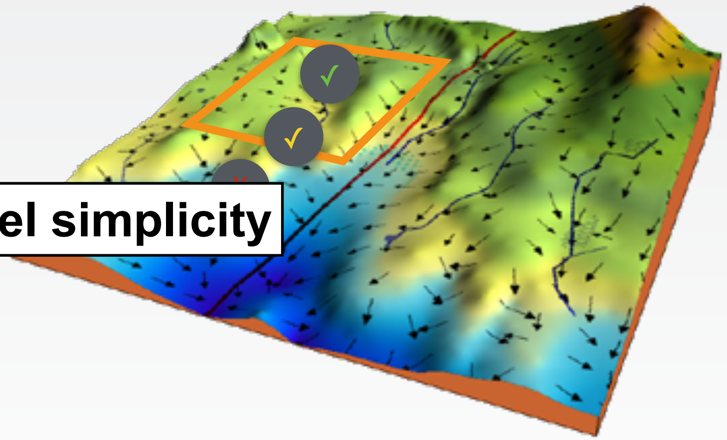
A few quantities

Many pixels, detector elements, hits

A few quantities

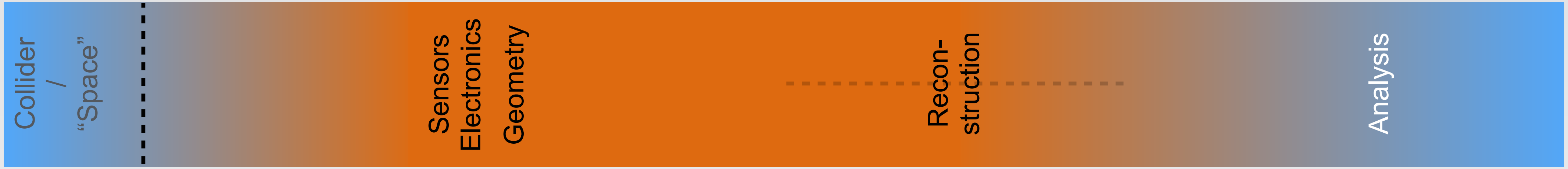
Very complex surrogate model

✓ Much simpler surrogate model



Trade off between computing resources and surrogate model simplicity

Make Wider Steps?



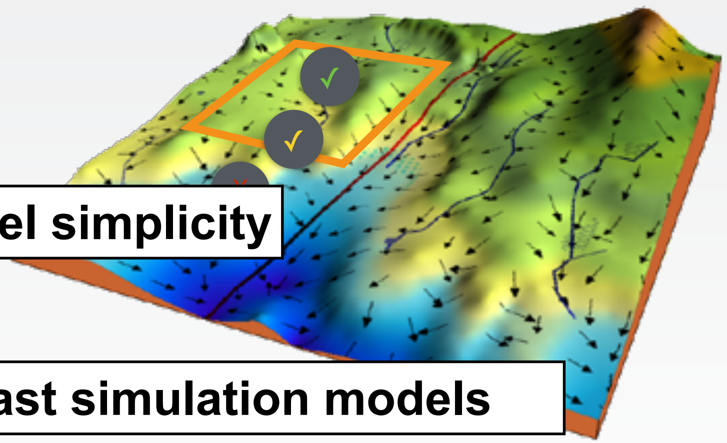
A few quantities

Many pixels, detector elements, hits

A few quantities

Very complex surrogate model

✓ Much simpler surrogate model

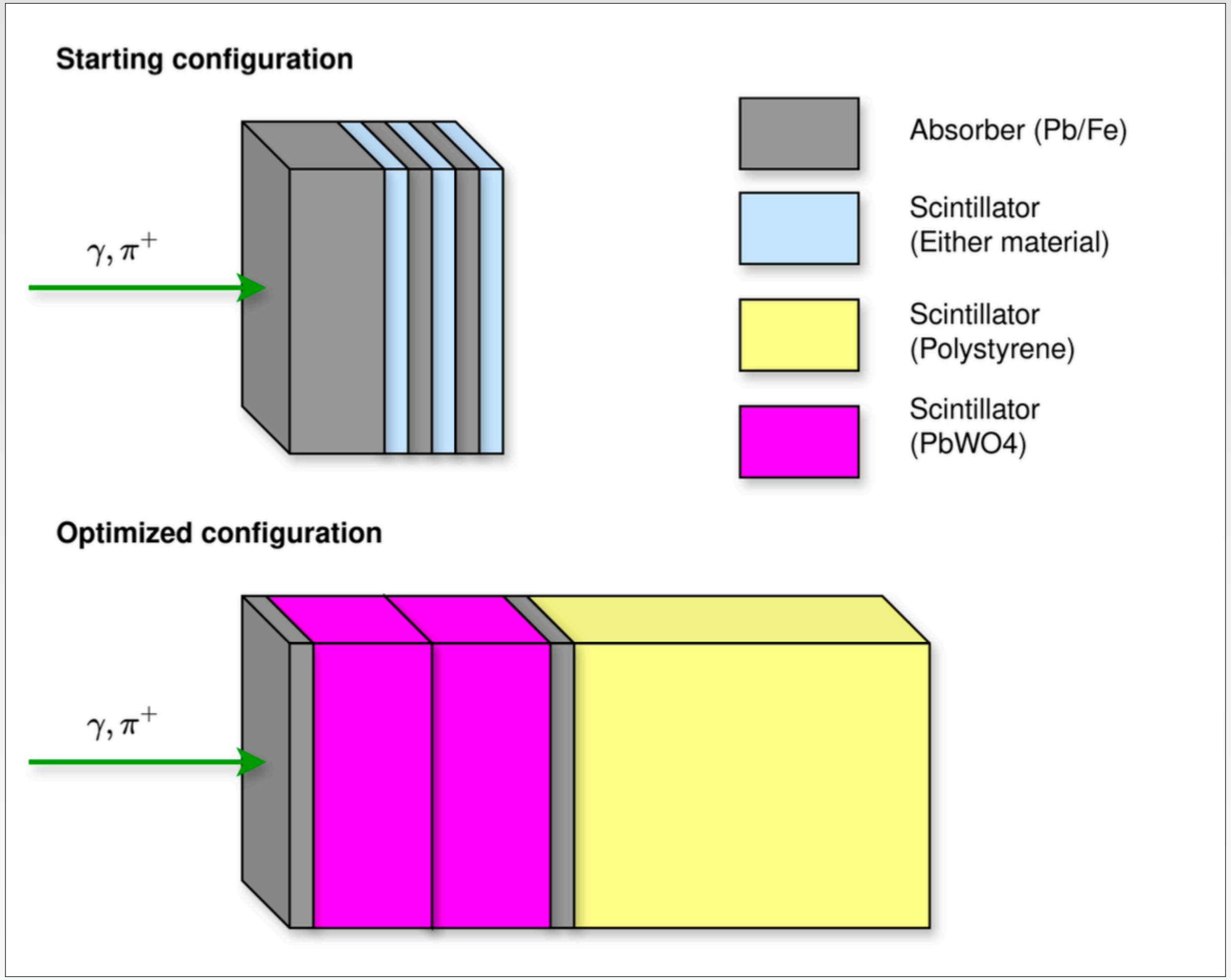


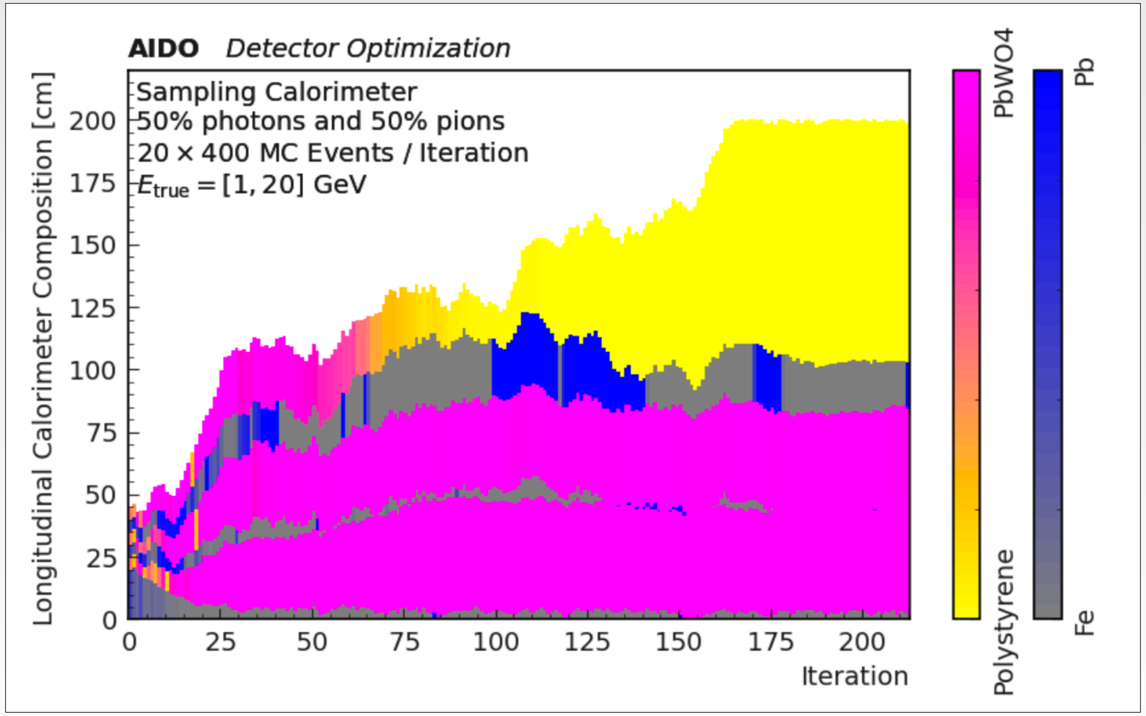
Trade off between computing resources and surrogate model simplicity

These models, applied to existing detectors, are generative (ultra) fast simulation models

Even discrete parameters get optimised

- Add cost and space constraints
- Model cost of materials carefully

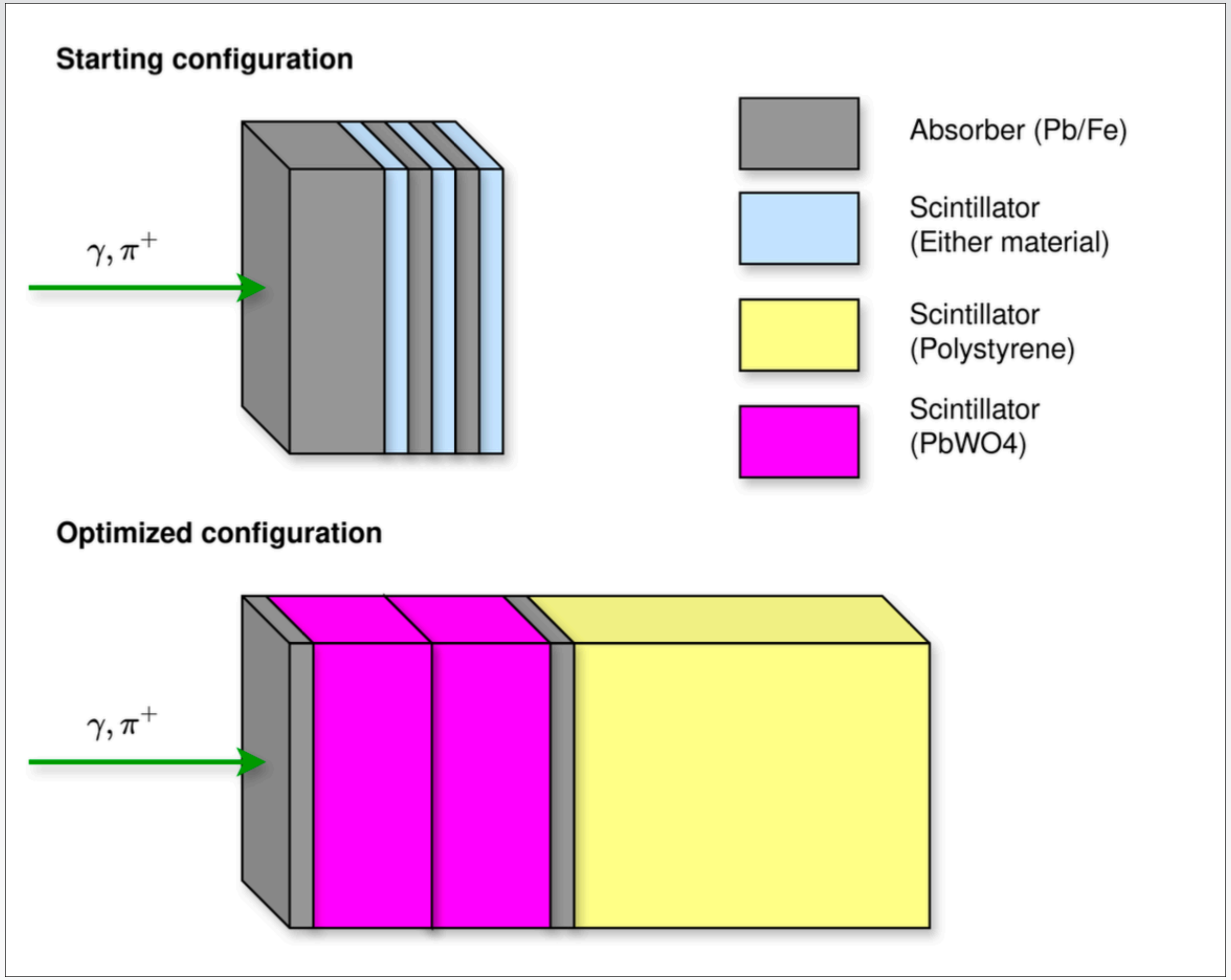




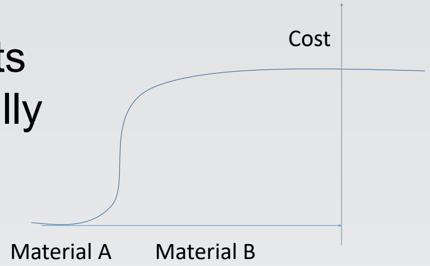
- Best mean resolution of all configurations

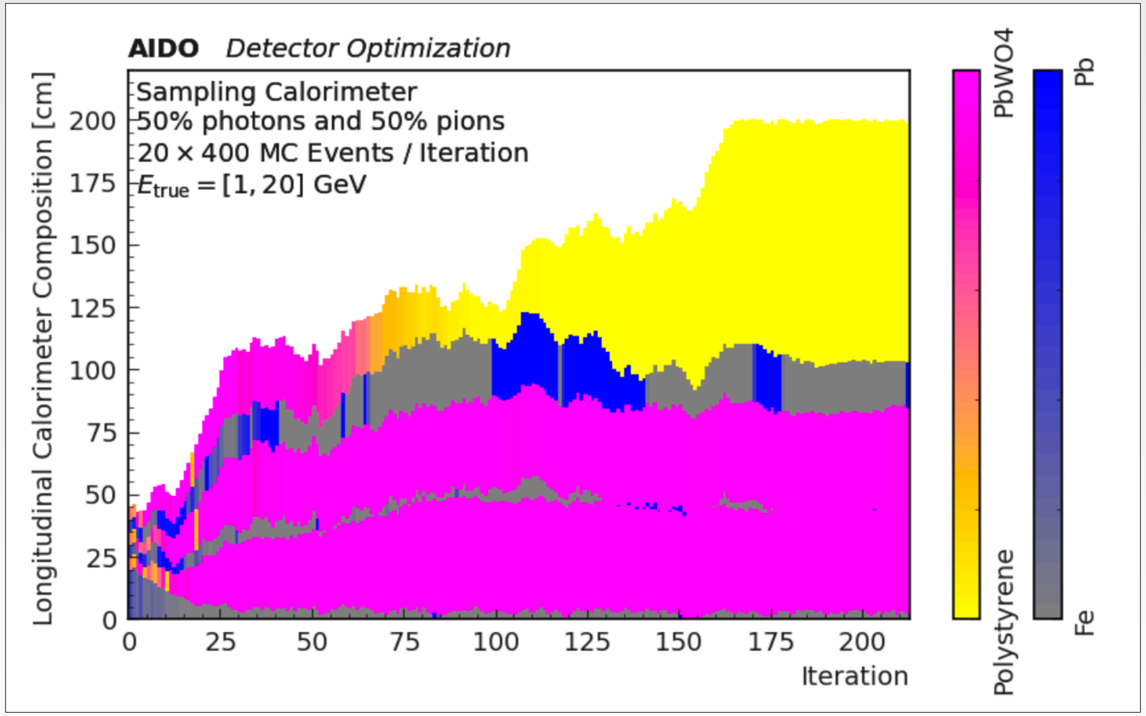
K.Schmidt, et al., arXiv:2502.02152

Even discrete parameters get optimised



- Add cost and space constraints
- Model cost of materials carefully





- Best mean resolution of all configurations

K.Schmidt, et al., arXiv:2502.02152

Sample efficiency

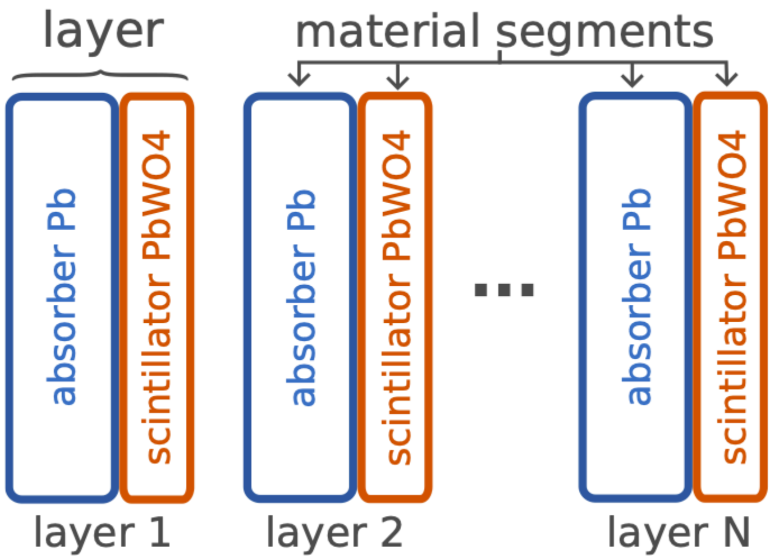
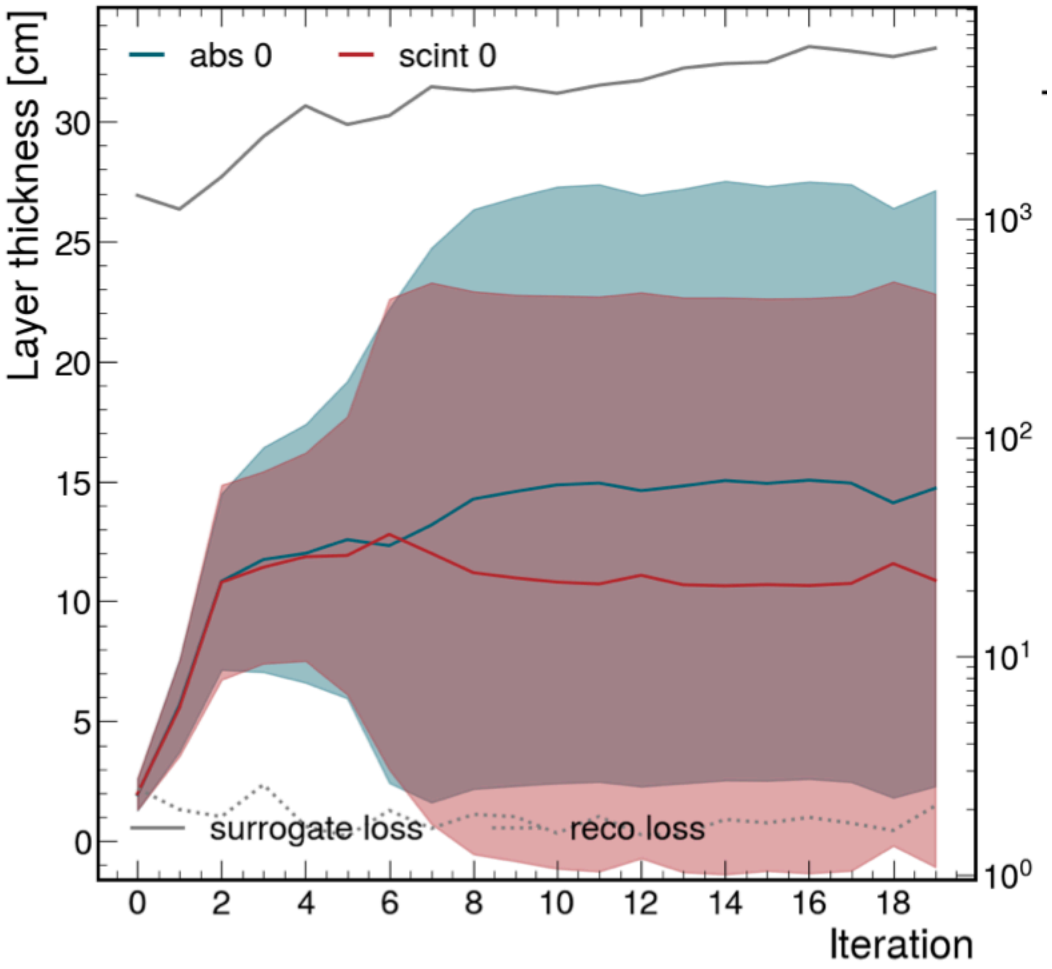


FIG. 1: Calorimeter consisting of layers of interleaved lead absorber and lead-tungstate scintillator segments.



(b) RECO-OPT without TL, 5 events

arXiv:2503.14342