UNIVERSITÉ DE GENÈVE

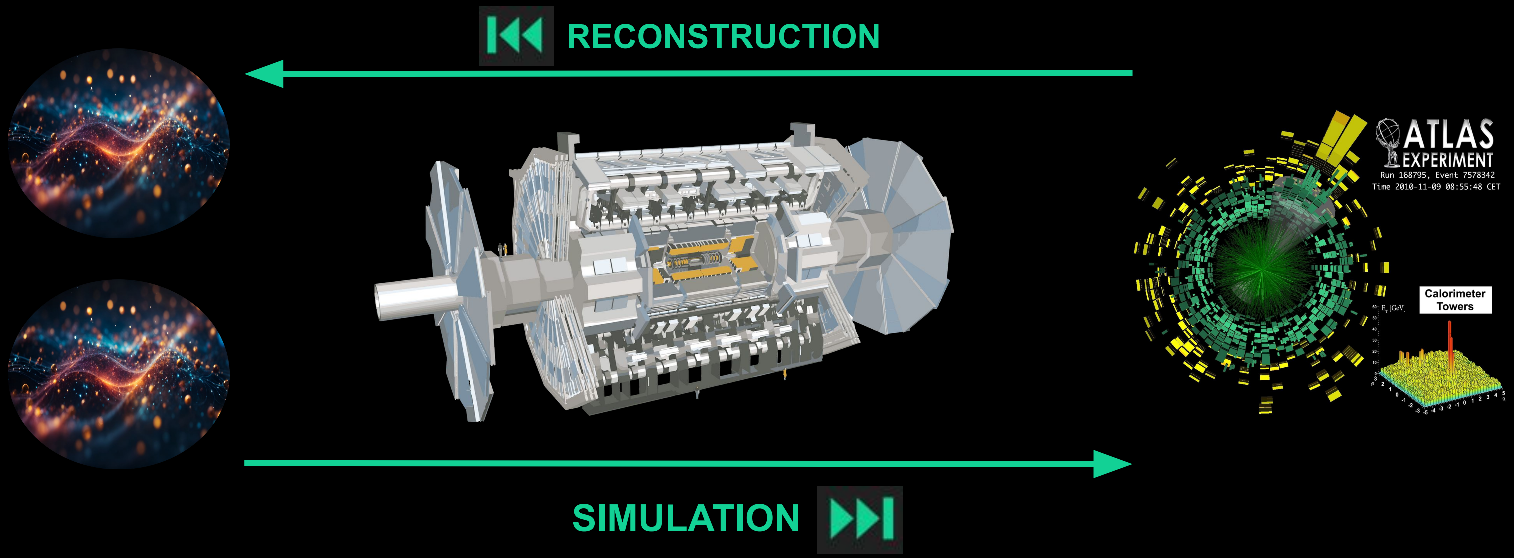
KIT

ML enhanced optimal detector design

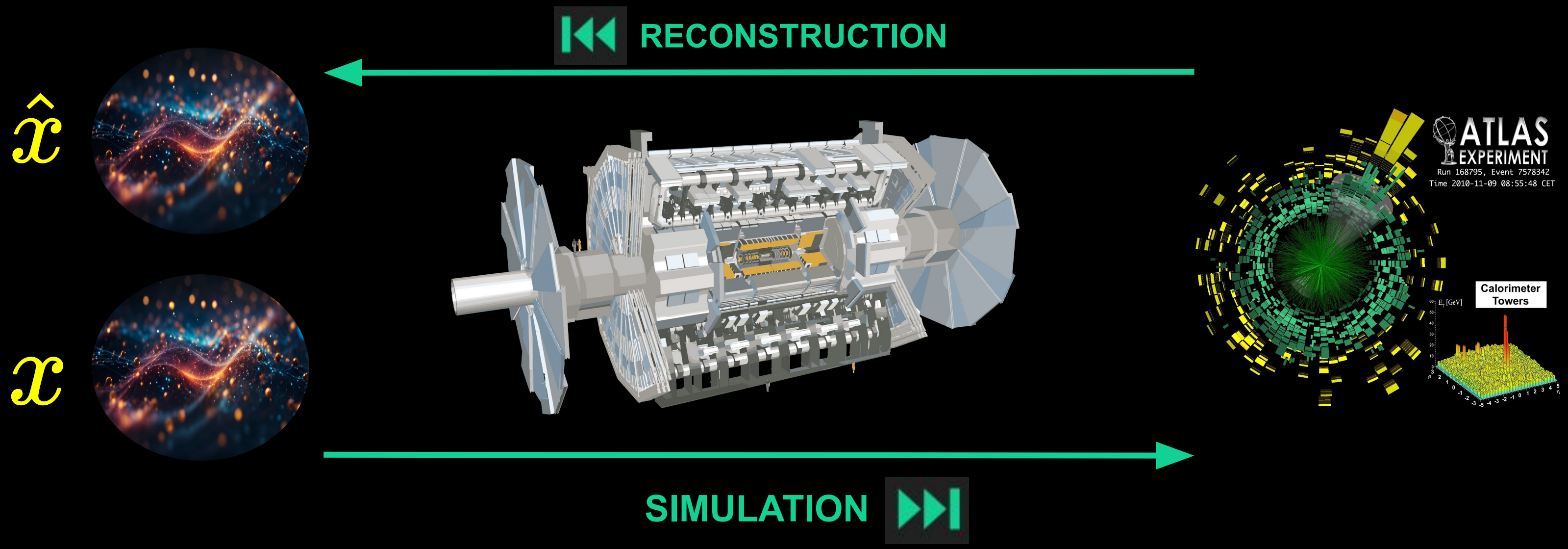
François Fleuret, Tobias Golling, Jan Kieseler, Stephen Mulligan, Atul Sinha, Kinga Anna Wozniak

CHIPP AI/ML & computing workshop
19.06.2024

The stage: problem and solution idea



The stage: problem and solution idea



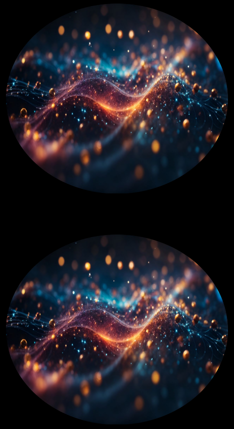
➔ We want x and x^ to be ‘close’

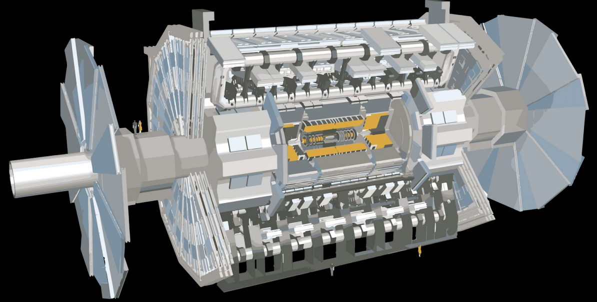
The stage: problem and solution idea

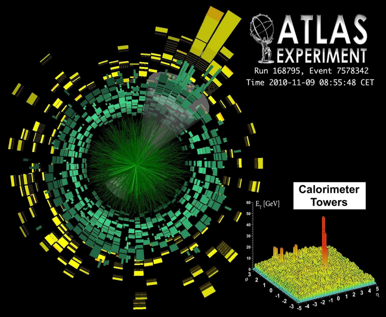
RECONSTRUCTION

$\hat{x}$

x







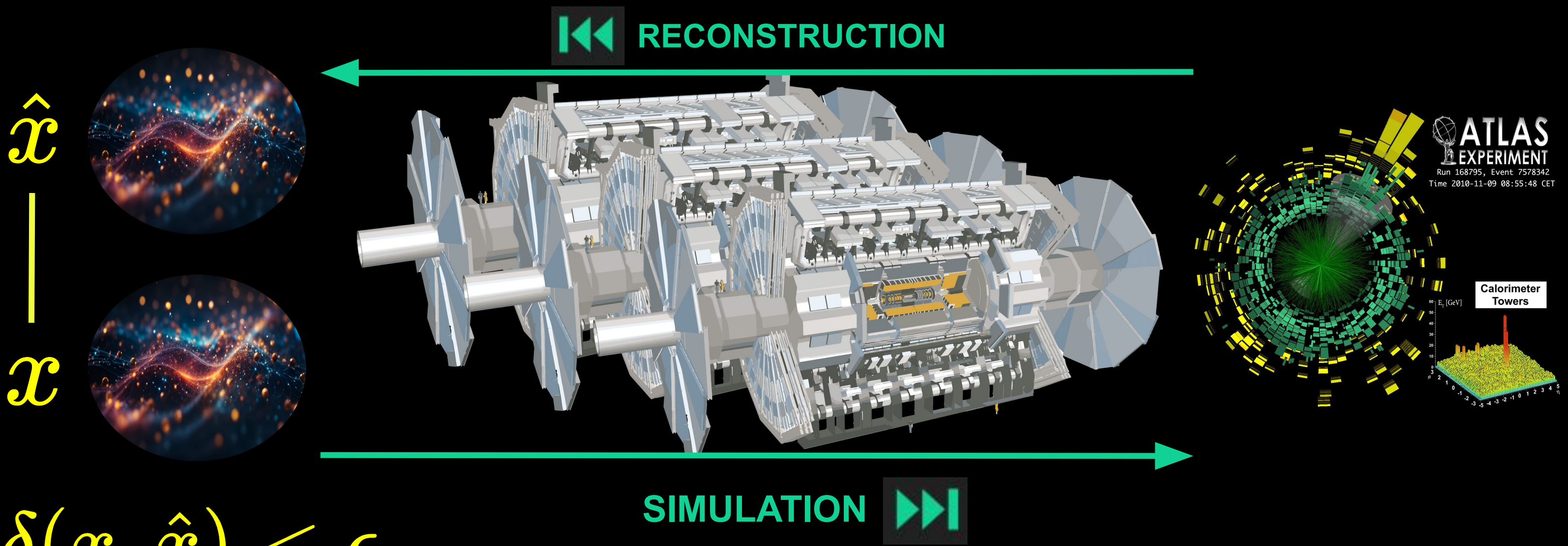
SIMULATION

$$\delta(x, \hat{x}) < \epsilon$$

➔ We want x and x^ to be ‘close’

➔ Developing detector with physics task in mind → impacts ‘closeness’

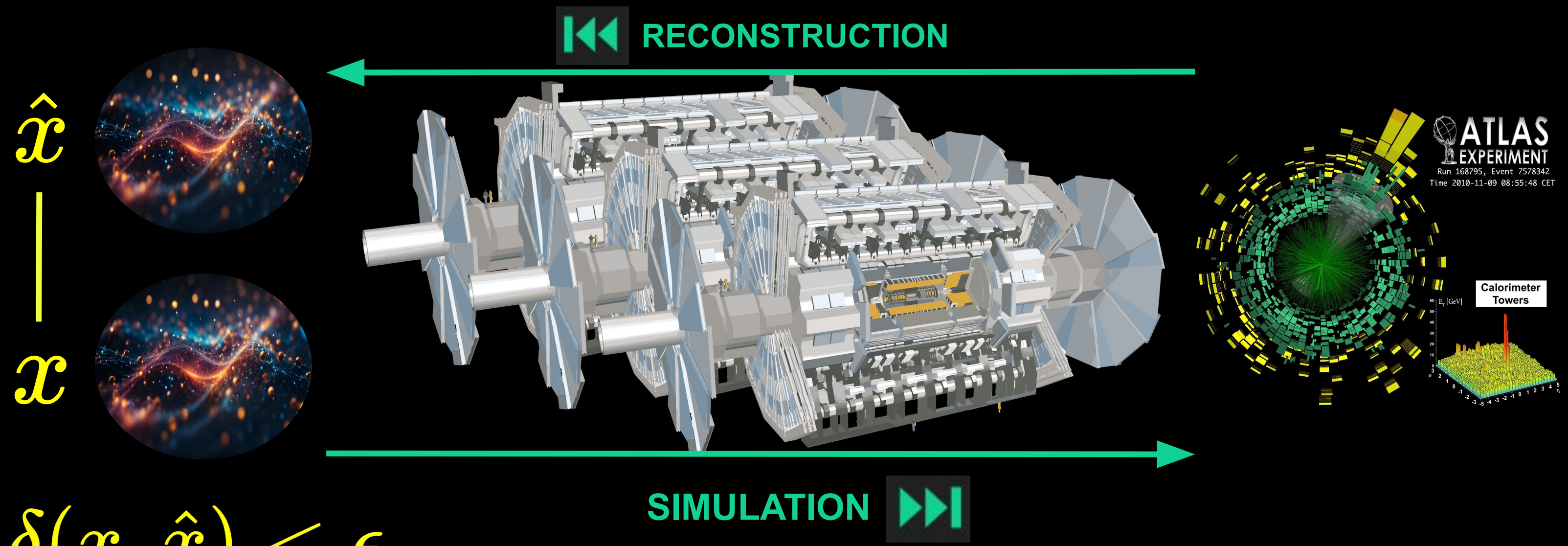
The stage: problem and solution idea



$$\delta(x, \hat{x}) < \epsilon$$

- ➔ We want x and x^ to be ‘close’
- ➔ Developing detector with physics task in mind → impacts ‘closeness’
- ➔ Different options for detector design → giving different δ

The stage: problem and solution idea



$$\delta(x, \hat{x}) < \epsilon$$

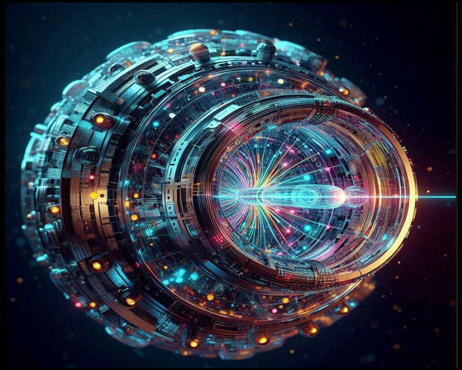
- ➔ We want x and x^ to be ‘close’
- ➔ Developing detector with physics task in mind → impacts ‘closeness’
- ➔ Different options for detector design → giving different δ
- ➔ resource demanding & not differentiable wrt design

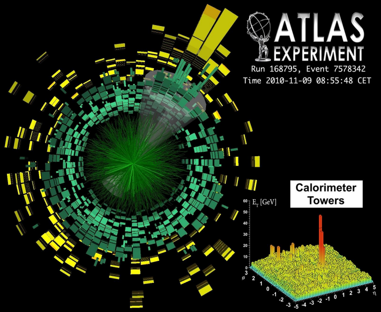
The stage: problem and solution idea

RECONSTRUCTION

$\hat{x}$

x





SIMULATION

$$\delta(x, \hat{x}) < \epsilon$$

☆ plus maybe missing out on some unconventional design→innovation

- ➔ We want x and x^ to be ‘close’
- ➔ Developing detector with physics task in mind → impacts ‘closeness’
- ➔ Different options for detector design → giving different δ
- ➔ resource demanding & not differentiable wrt design

Fast assessment of design with local surrogate model

To assess detector design i for task j, simulation + reconstruction has to be developed

RECONSTRUCTION

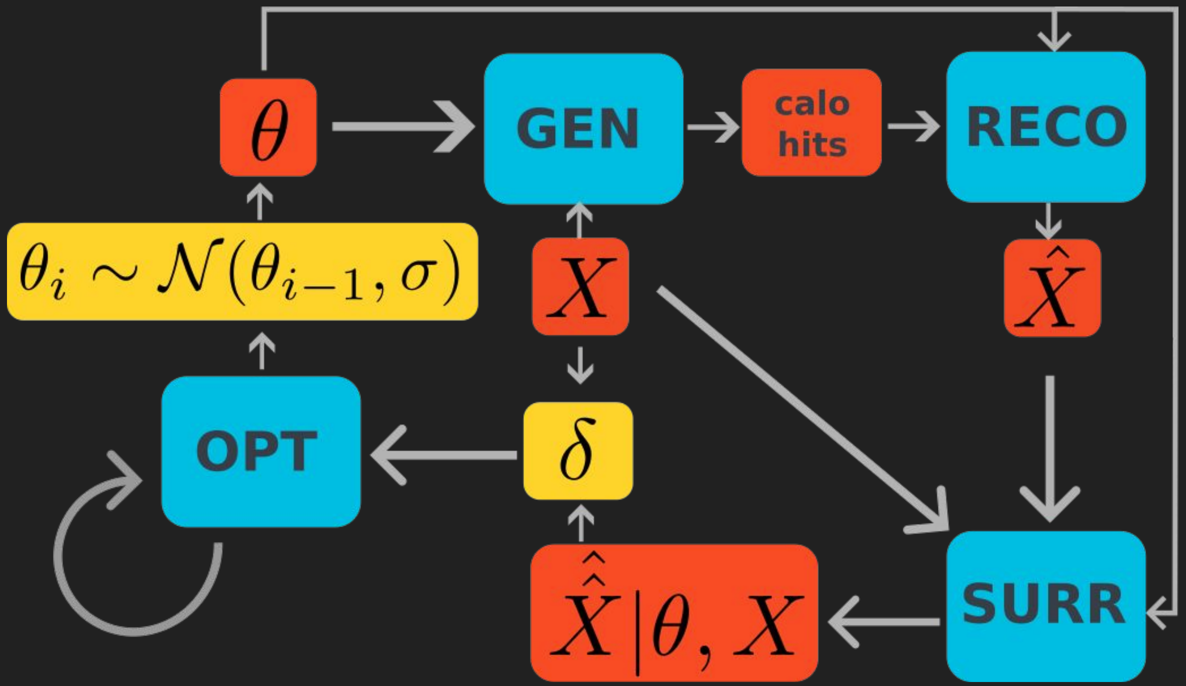
$$\arg\min_{\theta} \delta(X, D_\theta(X))$$

SIMULATION

- ➔ resource demanding
- ➔ not differentiable wrt design
- ➔ High complexity (pixels, hits, ...)

- For a set of designs $\{D^{(i)}\}$, each parametrized by a $\theta^{(i)}$, we want to find design that minimizes δ between the true and reconstructed x
- Need differentiability wrt θ
- Idea: Replace Simulation + Reco by a local ML Surrogate
  - Surrogate conditioned on local θ
  - Once trained, can provide a direction for optimization

Local surrogate setup



Gives access to p(δ| θ, x)→a scalar which encompasses sim+reco process

Which δ to pick?

Mutual information (MI): proxy for δ truth vs measurement

Information theoretic metric: $\mathrm{MI}(A,B) \equiv H(A) - H(A|B)$

= average reduction in uncertainty about A when observing B=b

= average amount of information that B conveys about A

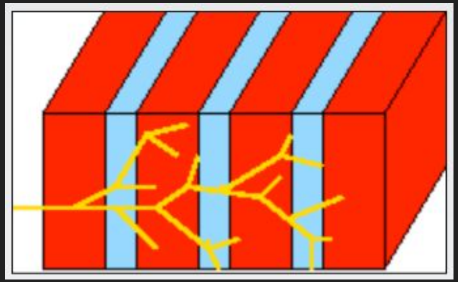
$$\min \delta(x, \hat{x}) \sim \max \mathrm{MI}(x, \hat{x})$$

Strengths of MI:

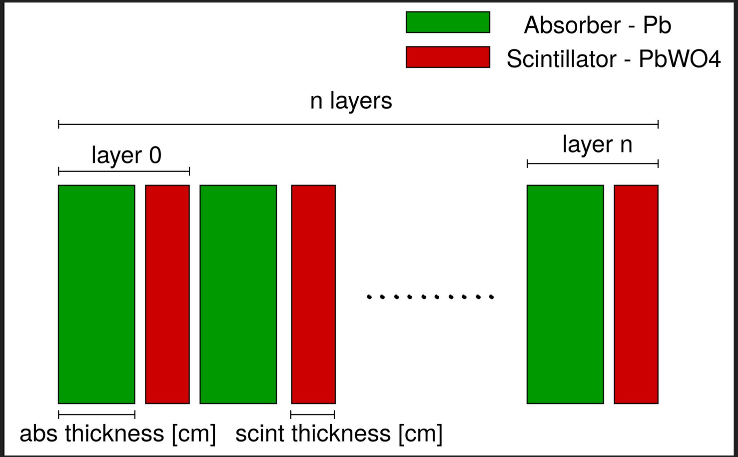
- Tells us if information is conserved during sim+reco process
- Captures non-linear dependencies
- X is multidimensional
  - →MI able to cover large part of phase space
  - →MI able to cover multiple tasks at once → δ can be multi-task

Drawback: Must be recomputed for each design and inputs

Data & Studies



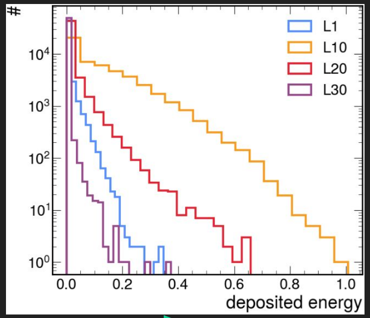
- Geant4 simulation of calorimeter
- Study single particle shot orthogonally at detector (50K)
- Two types of layers: absorber (cheap, lost info) & scintillator (expensive, yields info)
- Particles: photons and hadrons
- Recovering energy deposited
- θ = layer count, layer thickness
- Task: energy resolution and particle ID



Mutual information: # layers vs energy

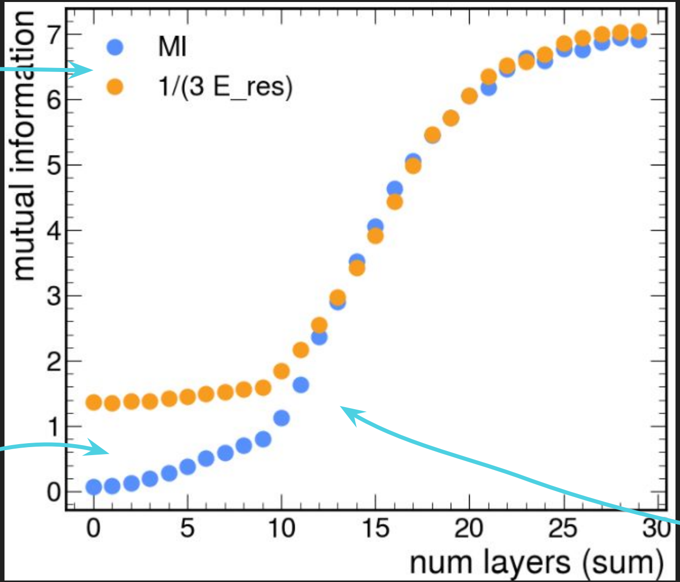
- MI score for set of designs
- Design θ = number of layers in detector
- Tasks: true energy of particle

Y: energy deposited in each layer



X: true particle energy

MI ∝ inverse of energy resolution



Plateau after majority of energy absorbed

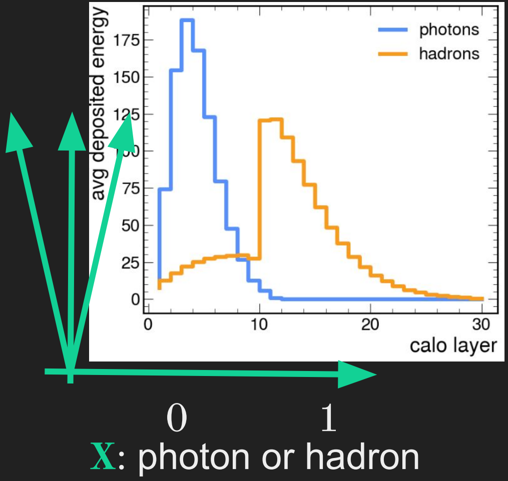
Low MI → Energy containment

Material change in layer 10: lead -> brass

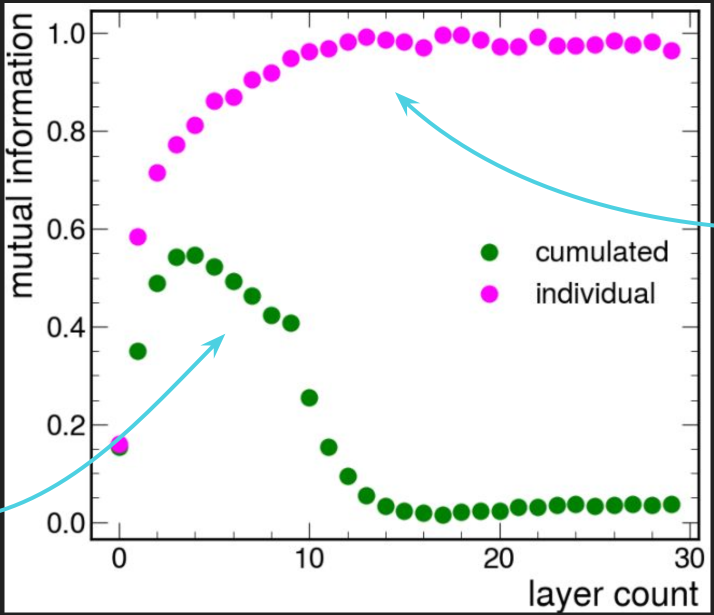
Mutual information: # layers vs particle ID

- MI score for set of designs
- Design θ = number of layers in detector
- Tasks: ID of particle

K x Y: energy deposited in each layer



Accumulation dilutes positional information of deposit (early/late)

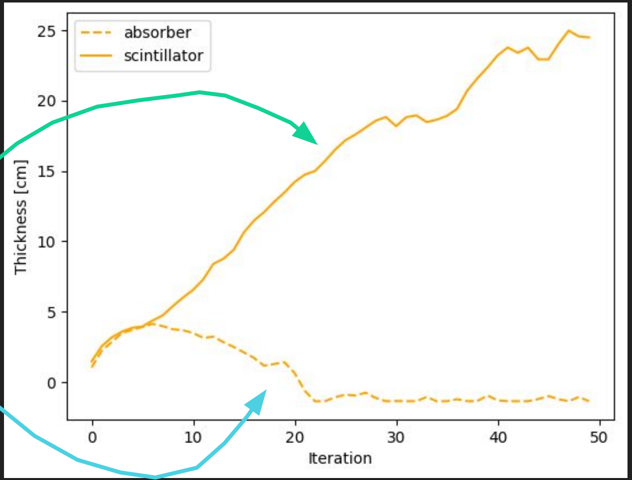


Individual contribution informs pid discrimination until deposit saturation

Surrogate: scintillator and absorber thickness

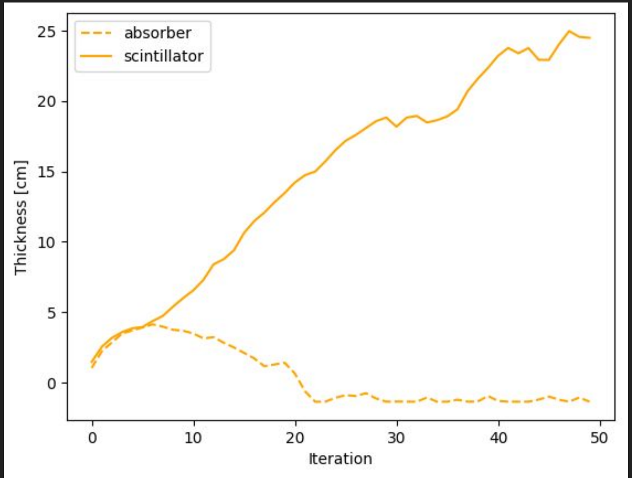
- Design θ = thicknesses of absorber and scintillator
- Metric δ = energy accuracy

Thickness Evolution:
increase scintillator
decrease absorber

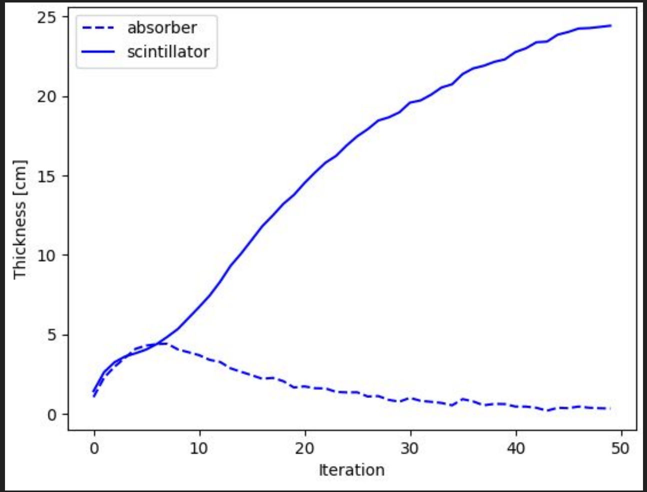


Surrogate: scintillator and absorber thickness

- Design θ = thicknesses of absorber and scintillator
- Metric δ = energy accuracy



re-initialized bumpy model w/o transfer learning



refined smooth transfer learning model

Transfer Learning Hypothesis:

Surrogate is able to map out local θ landscape across design tests

Conclusion

Problem:

development of detector involves simulation + reconstruction

→resource demanding

Solution:

- Surrogate model which maps out local design parameter space and allows for detector optimization
- Mutual information as a viable metric to encompass high-dimensional complex

First promising results applying solution to typical HEP tasks

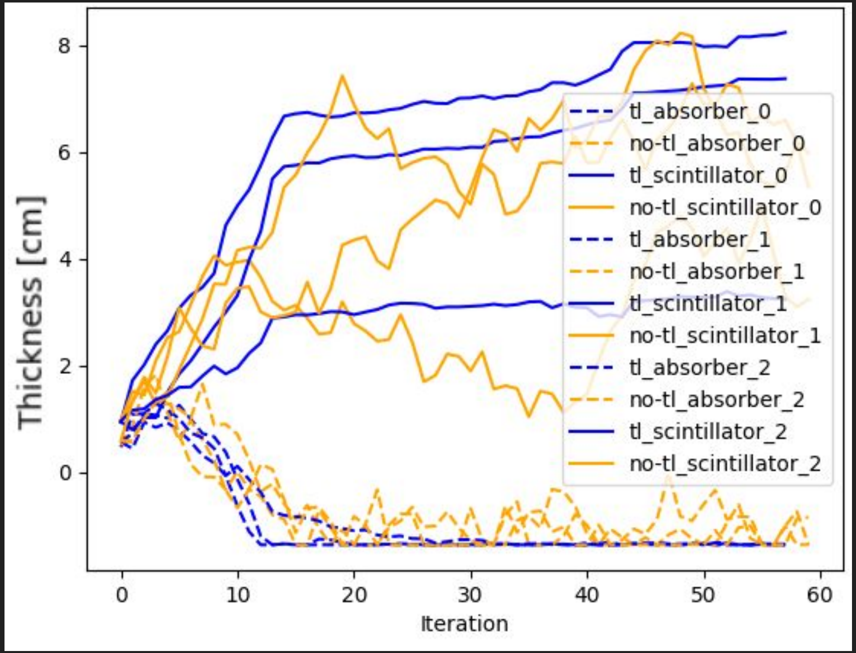
Backup

Surrogate: Scintillator and Absorber Thickness

Applicable to more complex multi-layer problems

Absorbers minimized

- - - - - - - - -



Scintillators maximized with decreasing intensity

Smooth refined model evolution

Mutual Information - The Theory

$$\mathrm{MI}(X,Y) \equiv H(X) - H(X|Y)$$

H … entropy

Equivalent to Kullback-Leibler divergence between the joint distribution and the product of the marginals

$$D_{KL}(P(X,Y)||P(X) \otimes P(Y))$$

Mutual Information Estimation with DNNs

Mine: Mutual Information Neural Estimation

Donsker-Varadhan dual representation of D_{KL}

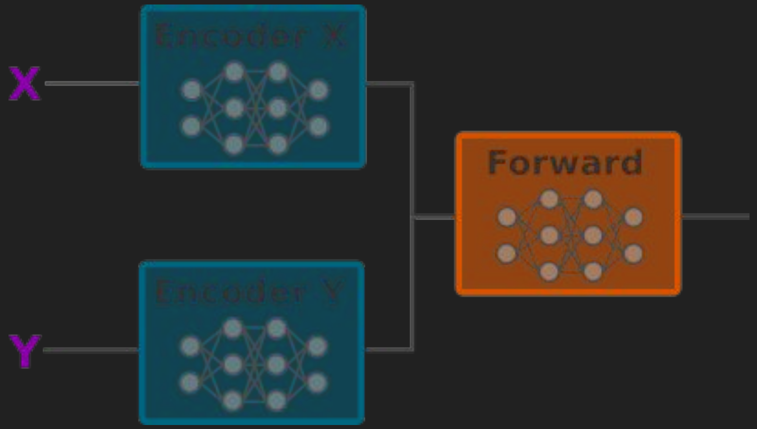
$$D_{\mathrm{KL}}(U \,||\, V) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_U[T] - \log(\mathbb{E}_V[e^T])$$

For a class of functions T for which the expectations are finite

=> Chose T to be parametrized by a neural network T_θ

$$D_{\mathrm{KL}}(P(X,Y) \,||\, P(X) \otimes P(Y)) = \sup_{\theta\in\Theta} \mathbb{E}_{P_{XY}}[T_\theta] - \log(\mathbb{E}_{P(X)\otimes P(Y)}[e^{T_\theta}])$$

MI Model - architecture



X, Y … input random variables

Π … permutation

$$P(X,Y) \sim \mathrm{NN}(X,Y)$$
$$P(X) \otimes P(Y) \sim \mathrm{NN}(X, \Pi(y \in Y))$$