III. Physikalisches Institut

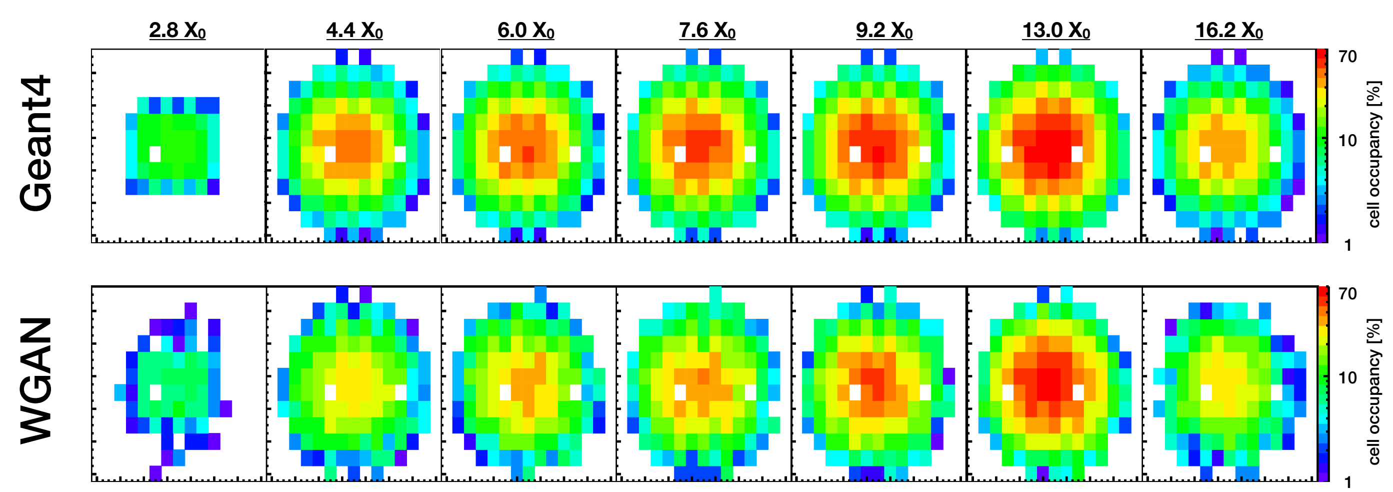
RWTH AACHEN UNIVERSITY

Bundesministerium für Bildung und Forschung

CERN

Big Data Science in Astroparticle Research - Workshop

Precise simulation of electromagnetic calorimeter showers using a Wasserstein Generative Adversarial Network



Martin Erdmann, Jonas Glombitza, Thorben Quast*

19.02.2019

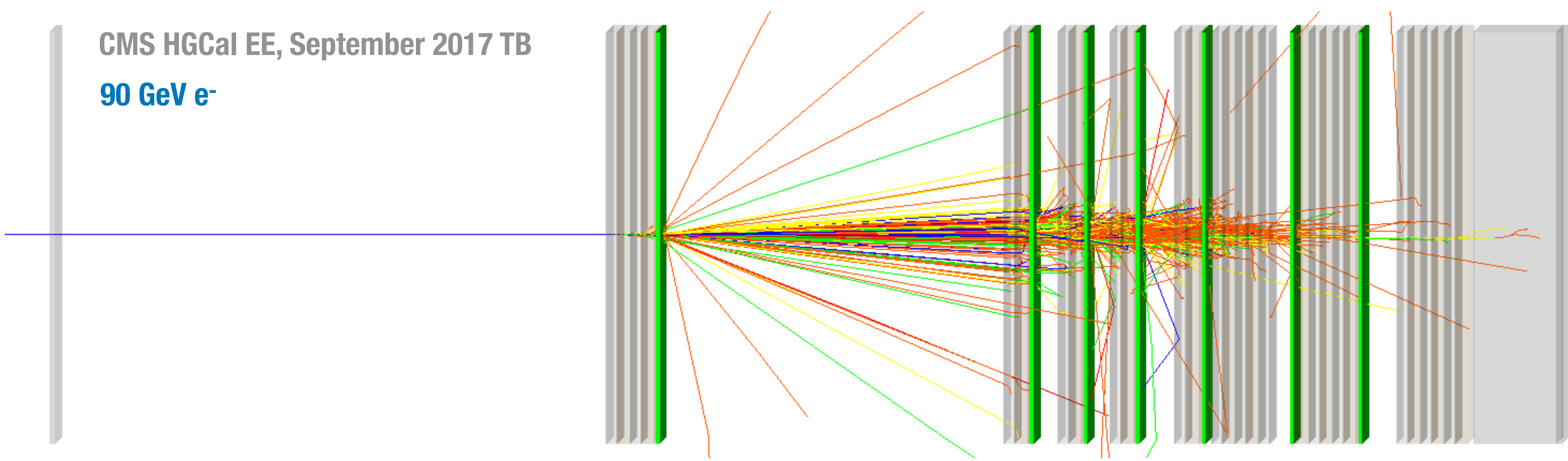
CMS

Calorimeter simulation nowadays

- Computationally expensive: simulation of particles interacting with material.

Geant 4

- electromagnetic & hadronic physics, lists with increasing/decreasing accuracy.



Near future: Simulation with generative models ?

- Computationally expensive: simulation of particles interacting with material.

Geant 4

- electromagnetic & hadronic physics, lists with increasing/decreasing accuracy.

- Grand goal: replace simulation steps by *ultra* ***fast, accurate*** generative methods.

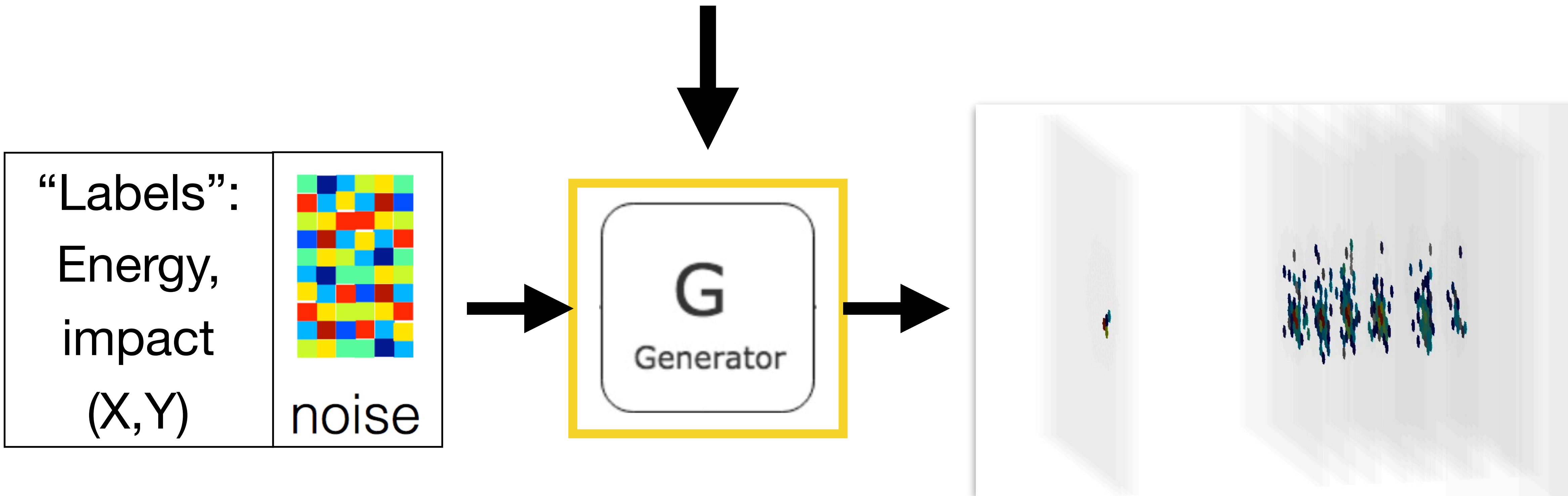
➡ **Step 1: Focus on simulation of particles showers in calorimeters.**

Proof-of-principle already demonstrated:

- *arXiv:1701.05927v2, arXiv:1705.02355v2, arXiv:1711.08813v1, S. Vallecorsa @ ACAT2017, arXiv:1802.03325v1 , …*

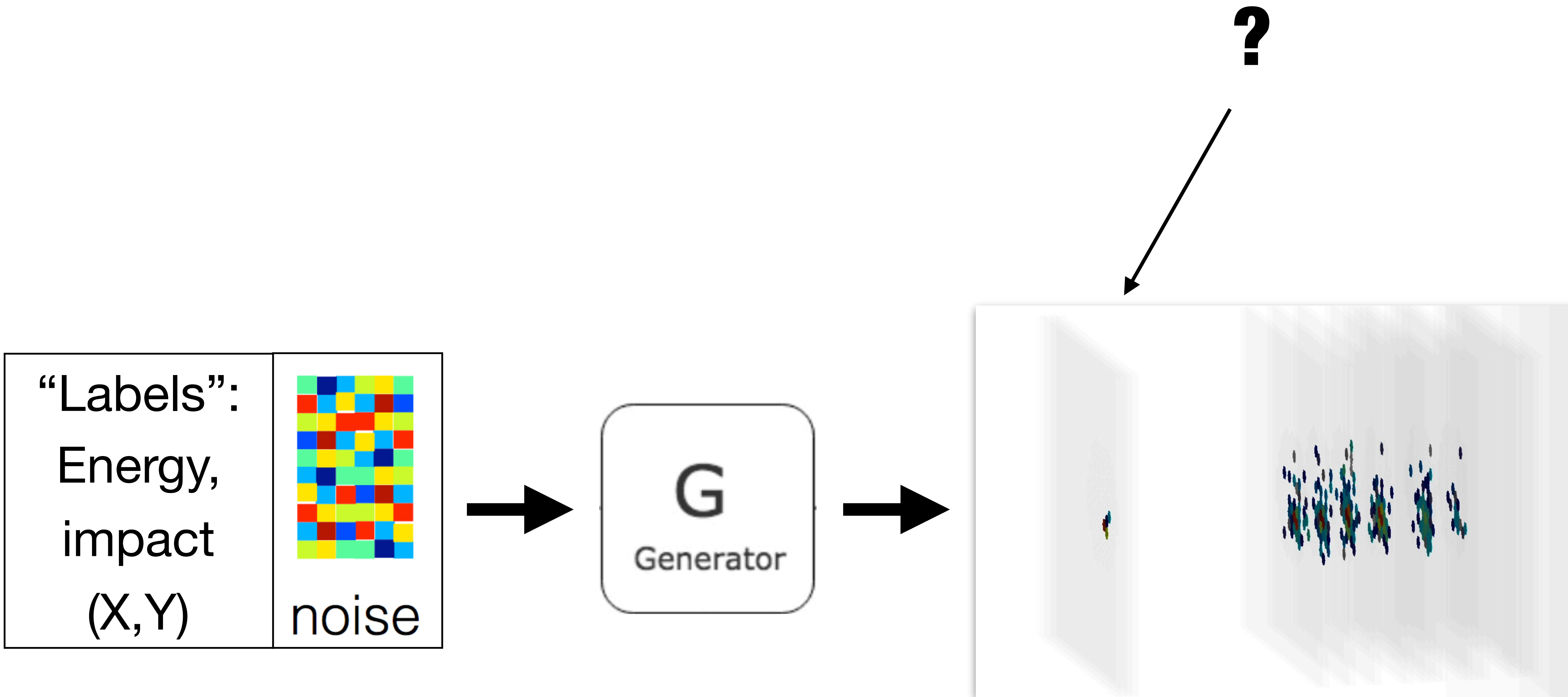
Goal formulation

We want G: **Deep neural network**
= Function, $O(10^5)$ free parameters



Thorben Quast
quast@physik.rwth-aachen.de

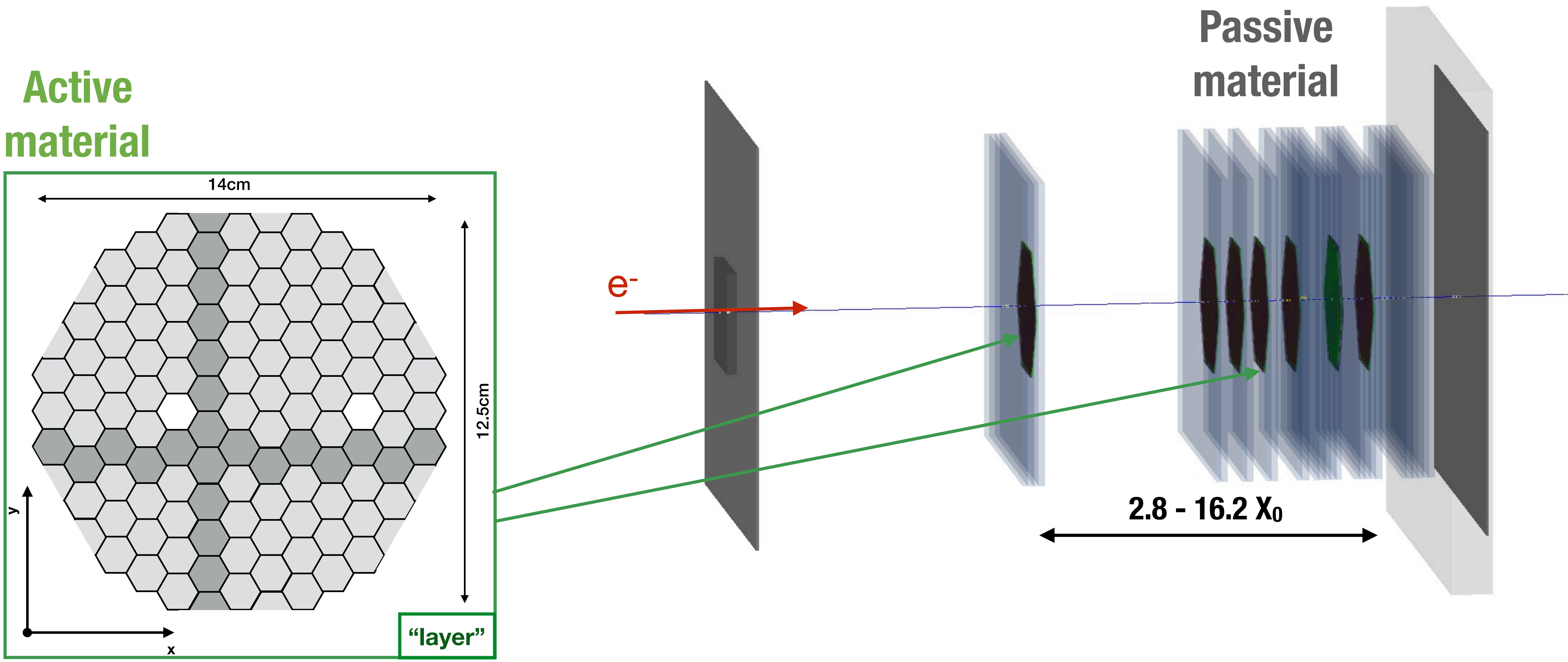
Goal formulation



Assumed calorimeter: HGCAL prototype (2017)

<u>HGCAL = Sampling calorimeter</u>

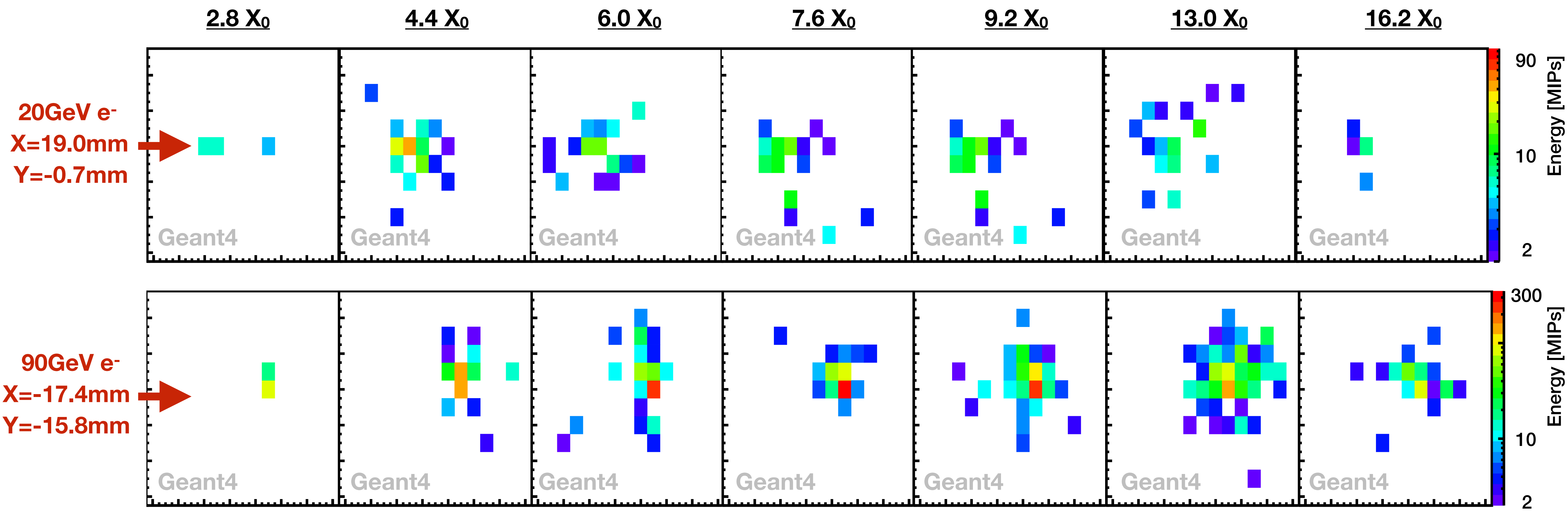
- **7** sensitive silicon **layers**.
- **Hexagonal** pixels with ~1cm in diameter, 128 per layer.



Thorben Quast
quast@physik.rwth-aachen.de

Exemplary Geant4 shower images

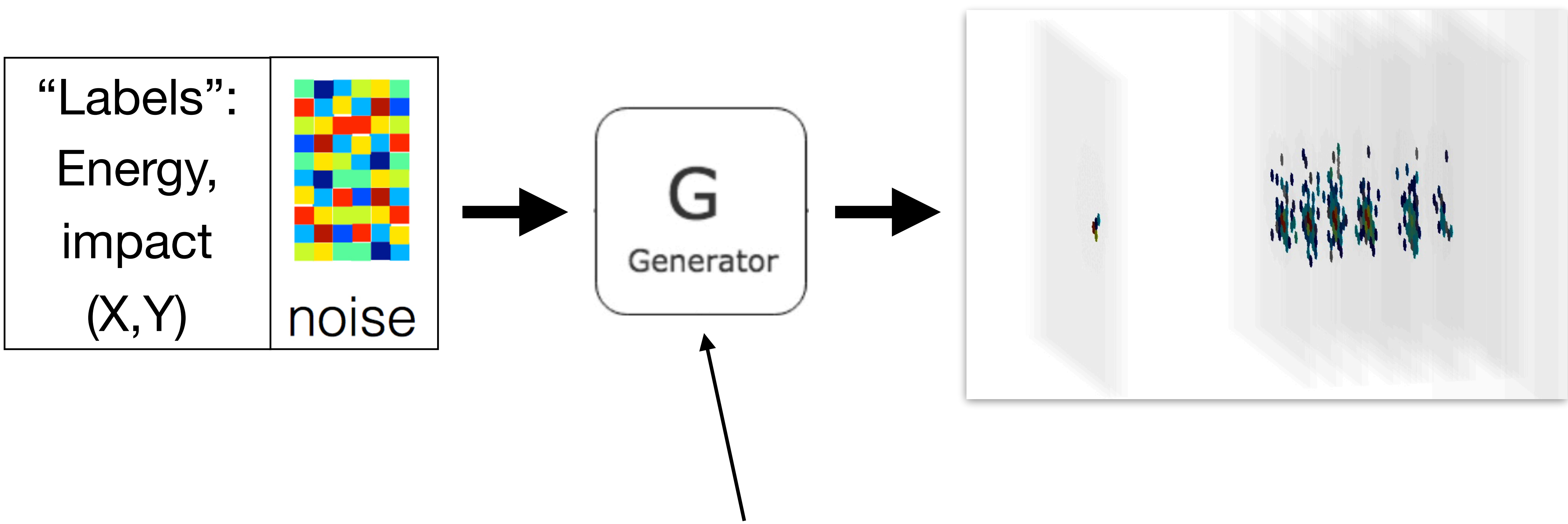
1 shower image:
12 x 15 x 7 tensor, intensity <—> energy



Sample:

- **20, 32, 50, 80 & 90 GeV electrons, O(100k)** showers each.
- Additional **70 GeV electron** sample not used in the training.

Goal formulation



How to train?

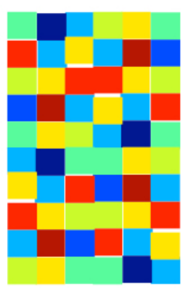
Thorben Quast
quast@physik.rwth-aachen.de

Concept of Generative Adversarial Neural Networks

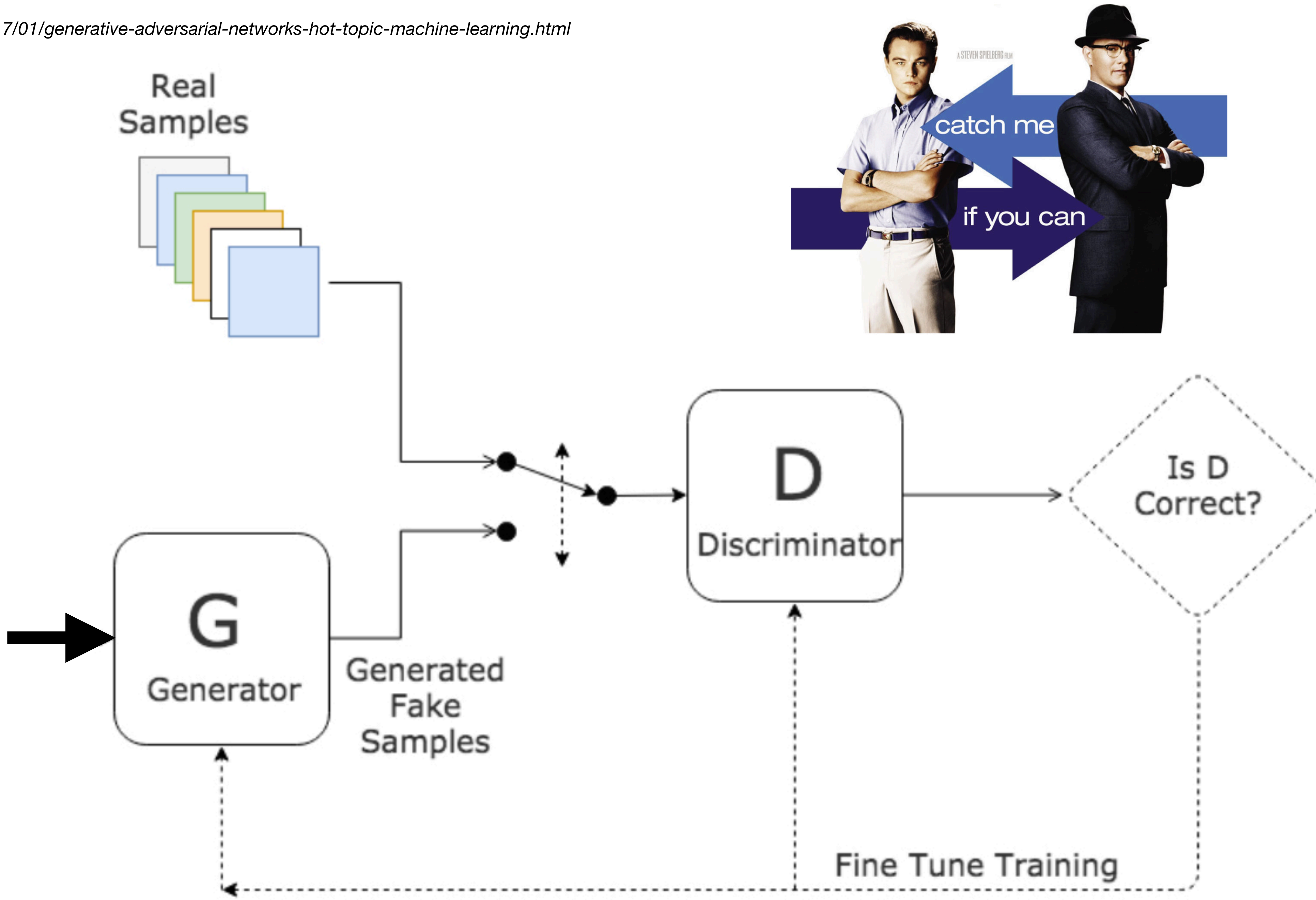
modified from: https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html

GAN

"Labels":
Energy,
impact
(X,Y)



noise



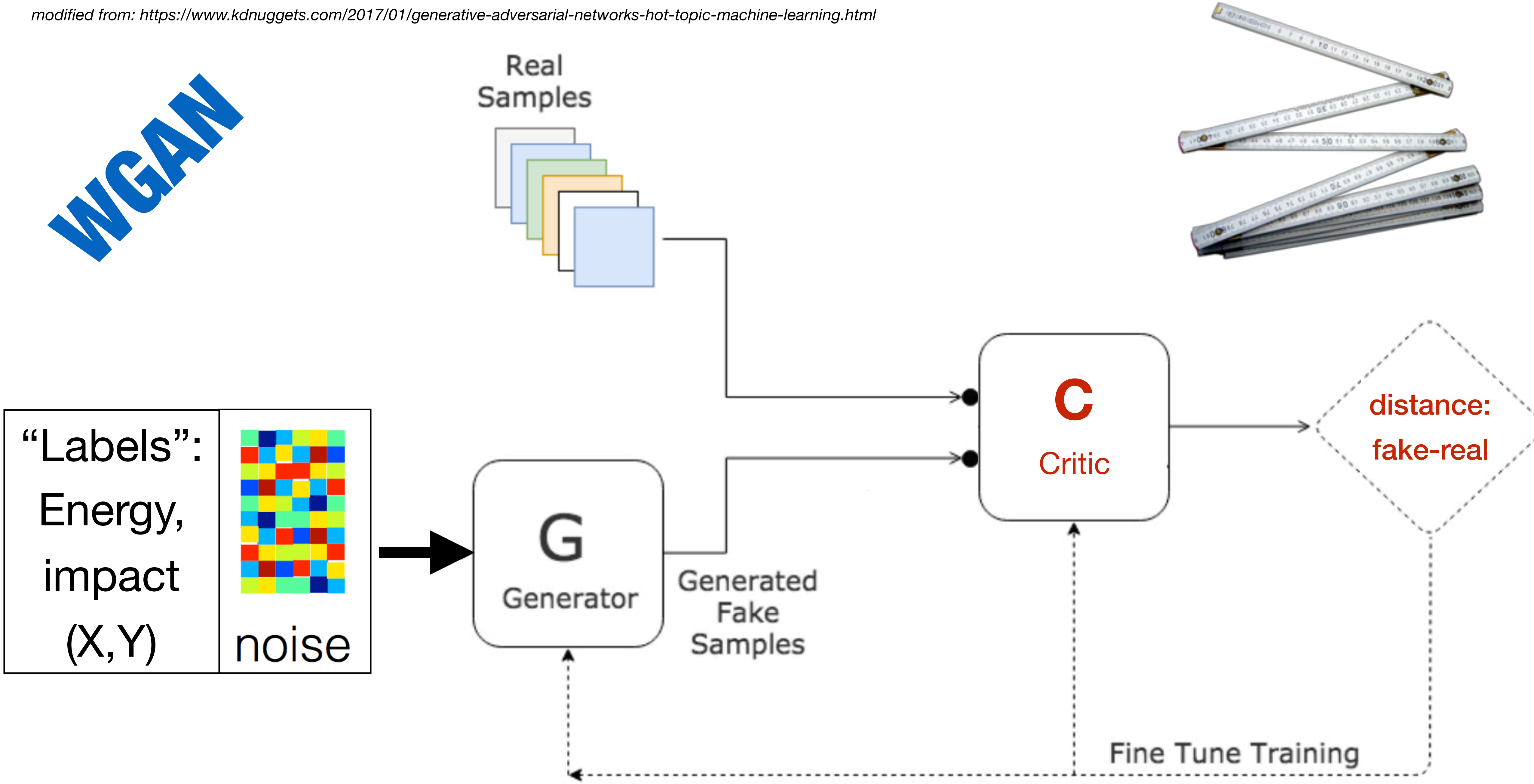
Ian J. Goodfellow's (2014)

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Concept of Generative Adversarial Neural Networks

modified from: https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html

WGAN



EARTH MOVER DISTANCE

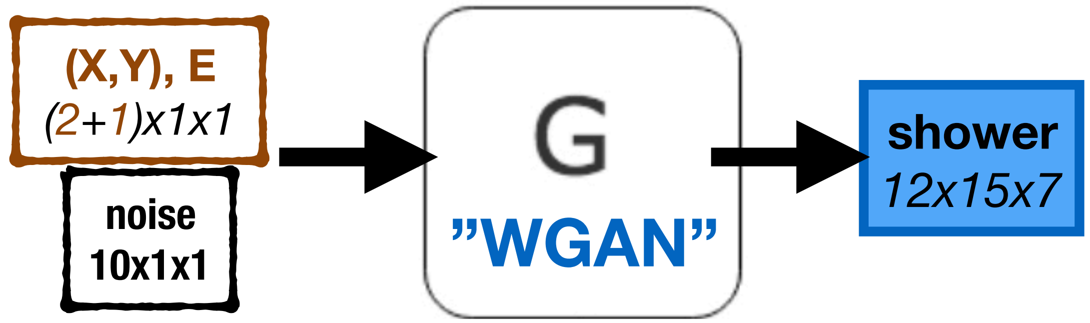
➡*arXiv:1704.00028v3*

$$L = \underbrace{\mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}\left[D(\tilde{\boldsymbol{x}})\right] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathbb{P}_r}\left[D(\boldsymbol{x})\right]}_{\text{Original critic loss}} + \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{\boldsymbol{x}}\sim\mathbb{P}_{\hat{\boldsymbol{x}}}}\left[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1)^2\right]}_{\text{Our gradient penalty}}.$$

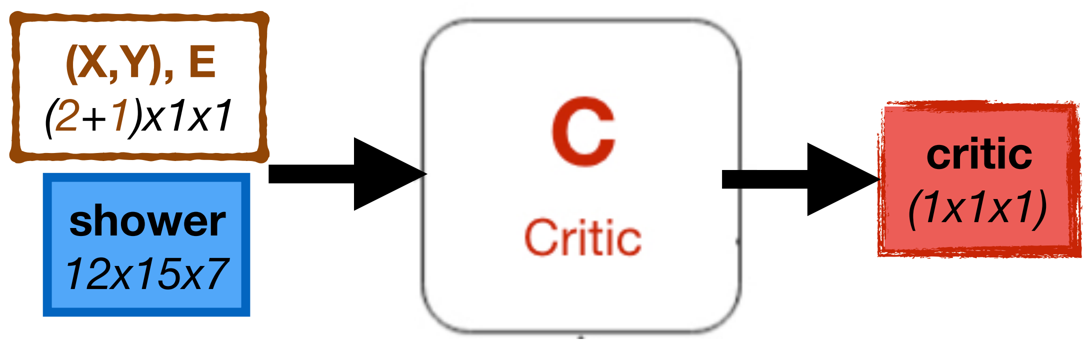
Thorben Quast
quast@physik.rwth-aachen.de

Training strategy using WGANs

- **Generator network (WGAN)** maps (noise, labels) to generated showers.



- **Critic network (C)** estimates the ***Earth Mover*** distance btw. generated & real showers.



Figures of merit for training:

Critic loss:

$$c_{loss} = -C(\text{showers}_{Geant4}, \text{labels}_{Geant4}) + C(\text{showers}_{gen}, \text{labels}_{gen}) + \lambda \times \textbf{gradient penalty},$$

$\lambda := 5$

Generator loss w.r.t. critic:

$$g_{loss,\, c} = -\, C(\text{showers}_{gen}, \text{labels}_{gen})$$

Adding labels

- **2 constrainer networks** for energy- (**E**) and position regression (**P**) on shower images.

Energy regression network **E**

shower *12x15x7* $\xrightarrow{E}$ **energy** *1x1x1*

Position regression network **P**

shower *12x15x7* $\xrightarrow{P}$ **(X,Y)** *(2x1x1)*

- **E** and **P** trained using Geant 4 showers - no bias from generated showers.

Energy and position regression losses:

$\mathbf{e}_{\mathbf{loss,\ Geant4}} = (\mathbf{E}(\mathrm{showers}_{\mathrm{Geant4}}) - E_{\mathrm{Geant4}})^2$, $\mathbf{p}_{\mathbf{loss,\ Geant4}} = (\mathbf{P}(\mathrm{showers}_{\mathrm{Geant4}}) - \mathrm{pos.}_{\mathrm{Geant4}})^2$

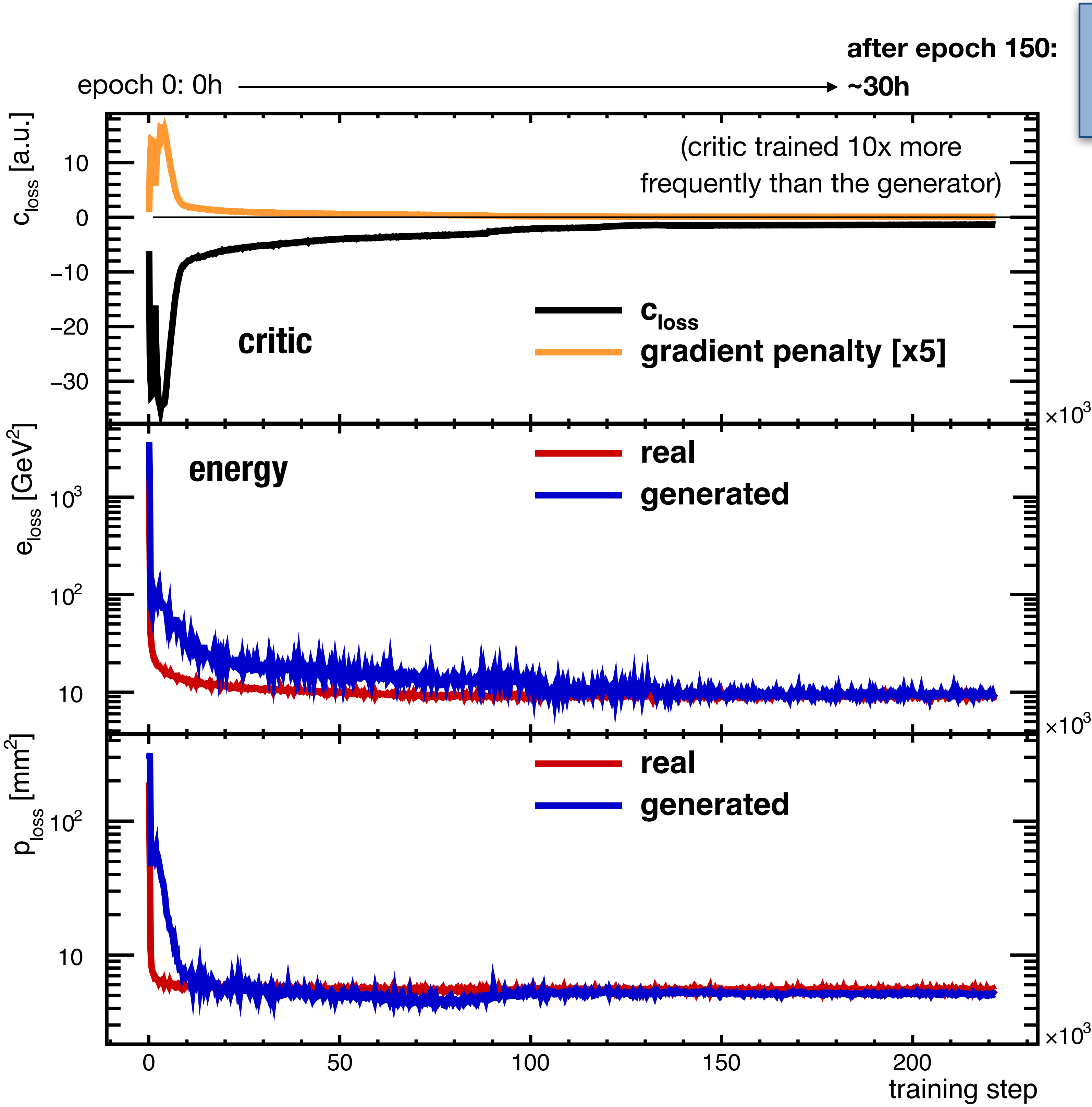
- **Generator is additionally trained to minimise the regression errors.**

➡**Total generator loss combines generator related losses.**

$$\mathbf{g}_{\mathbf{loss,\ tot}} = \mathbf{g}_{\mathbf{loss,\ c}} + \kappa_e \times |\mathbf{e}_{\mathbf{loss,\ Geant4}} - \mathbf{e}_{\mathbf{loss,\ gen}}| + \kappa_p \times |\mathbf{p}_{\mathbf{loss,\ Geant4}} - \mathbf{p}_{\mathbf{loss,\ gen}}|,$$

$$\kappa_e := \kappa_p := 0.01$$

System of networks trained for one day



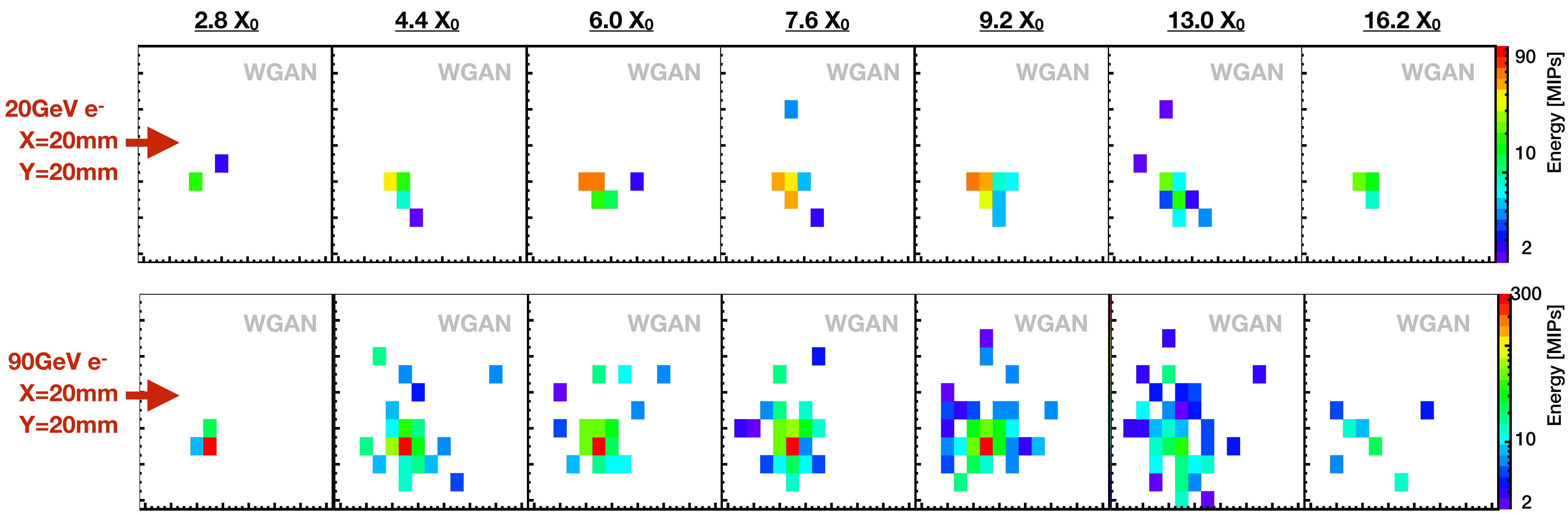
Software: Tensorflow v1.5.
Hardware: NVIDIA GTX1080 GPU.

✓ **Critic loss** converging to 0.
✓ Vanishing gradient penalty.

✓ Energy regression loss **converging** fast.
✓ Loss on generated images **converging.**

✓ Position regression loss **converging.**
✓ Loss on generated images **converging.**

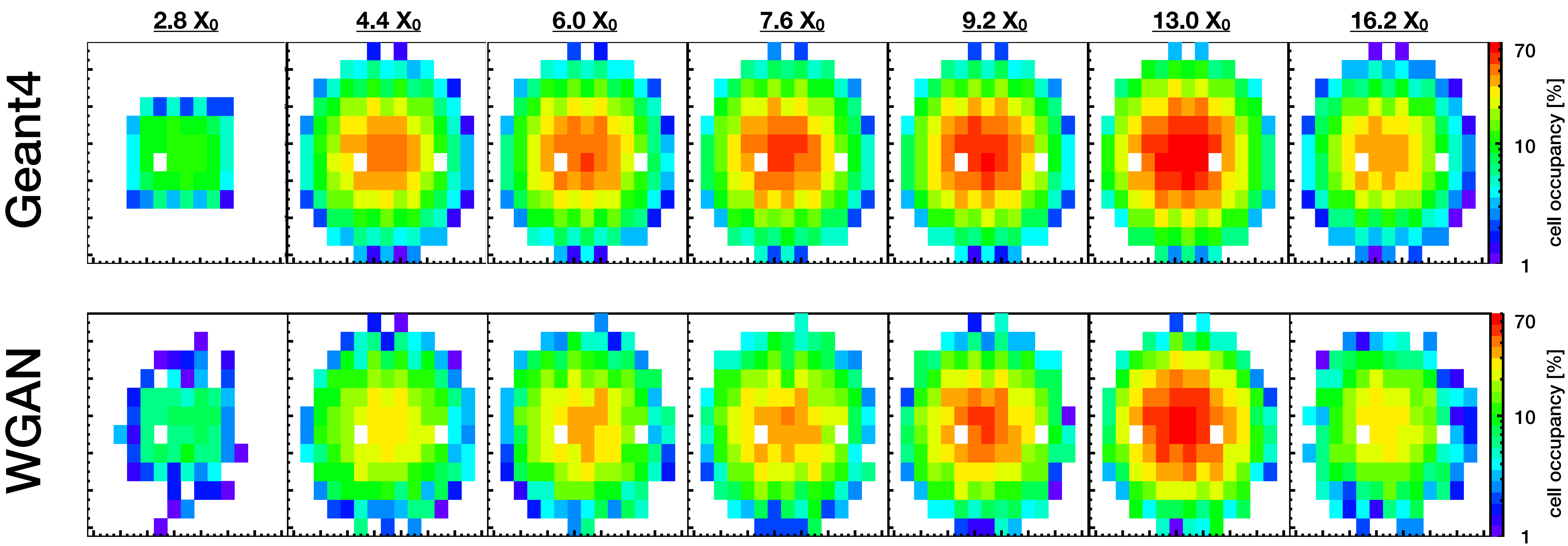
Generated electron showers look reasonable



Side note:

- Reasonable shower images are already obtained **after a few training epochs.**

WGAN has learnt: Pixel occupancy



note: masking of regions outside the acceptance in the WGAN

✓Radial development.

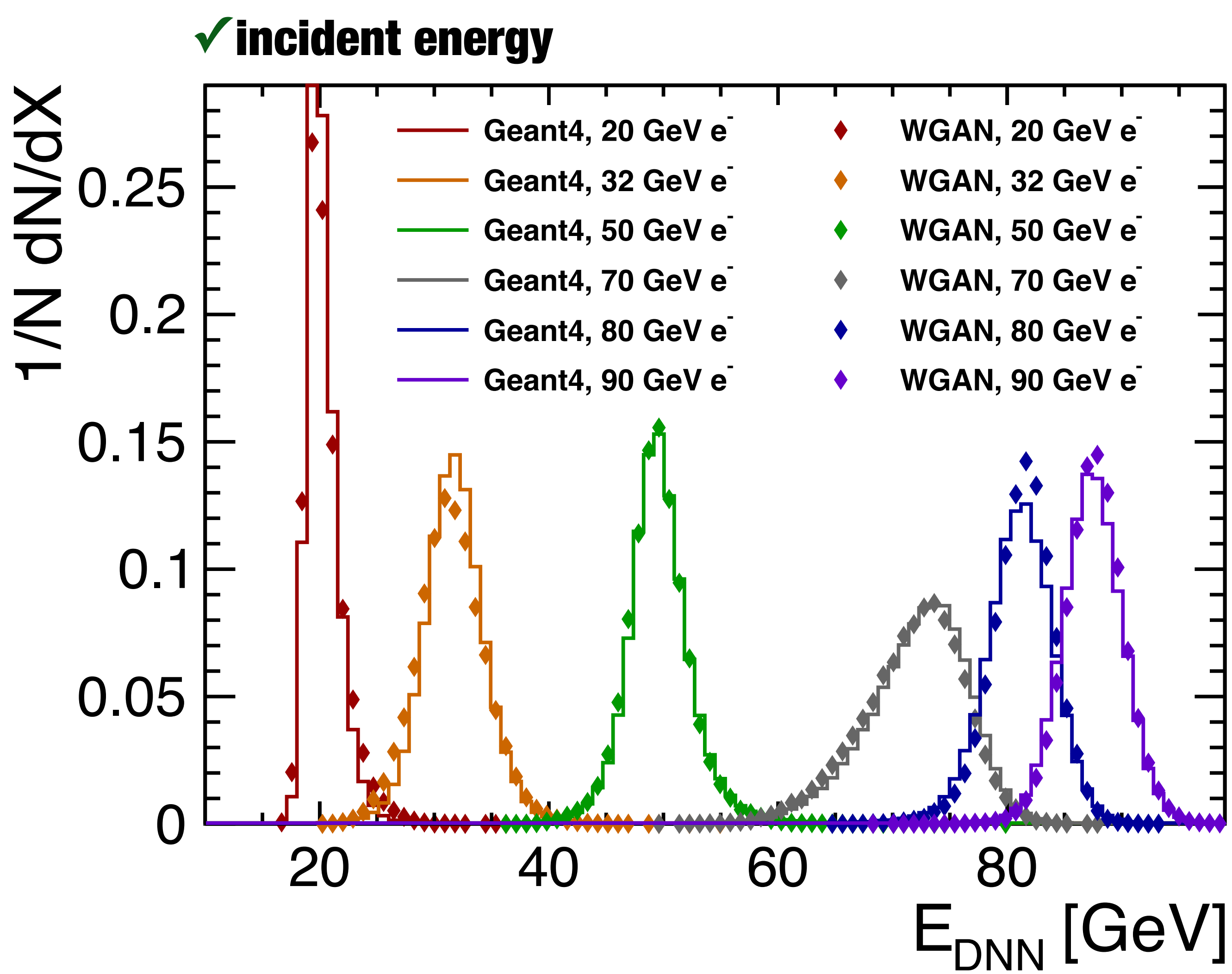
x WGAN: Overall scale slightly underestimated.

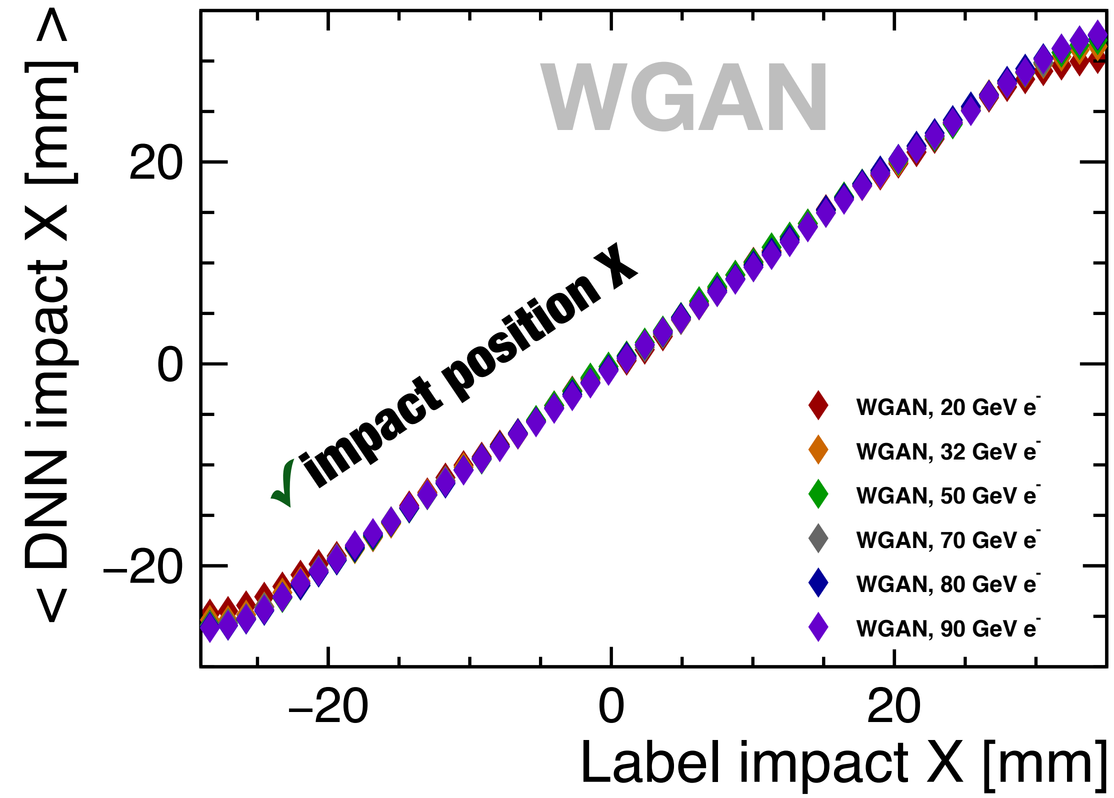
Generated events: Dependence on labels

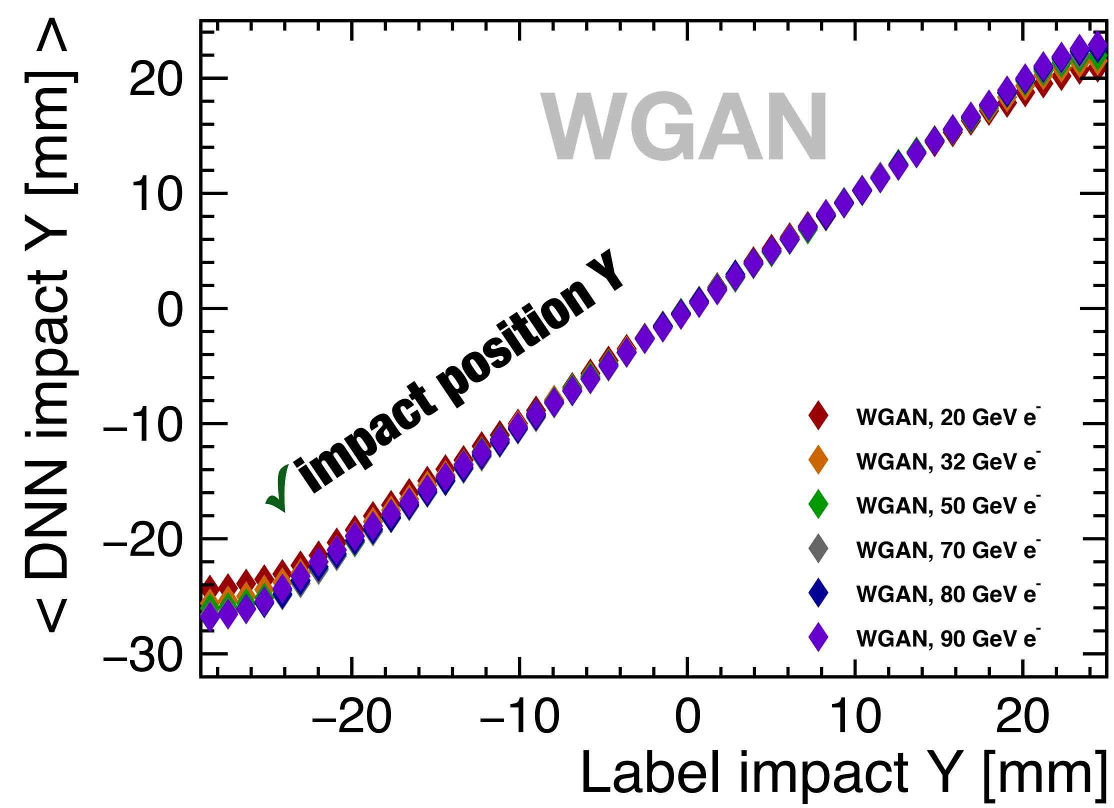
If WGAN has learnt to respect labels:

Reconstructed quantities of generated showers **correlate with true label.**

Note: **70GeV sample** not used in training.

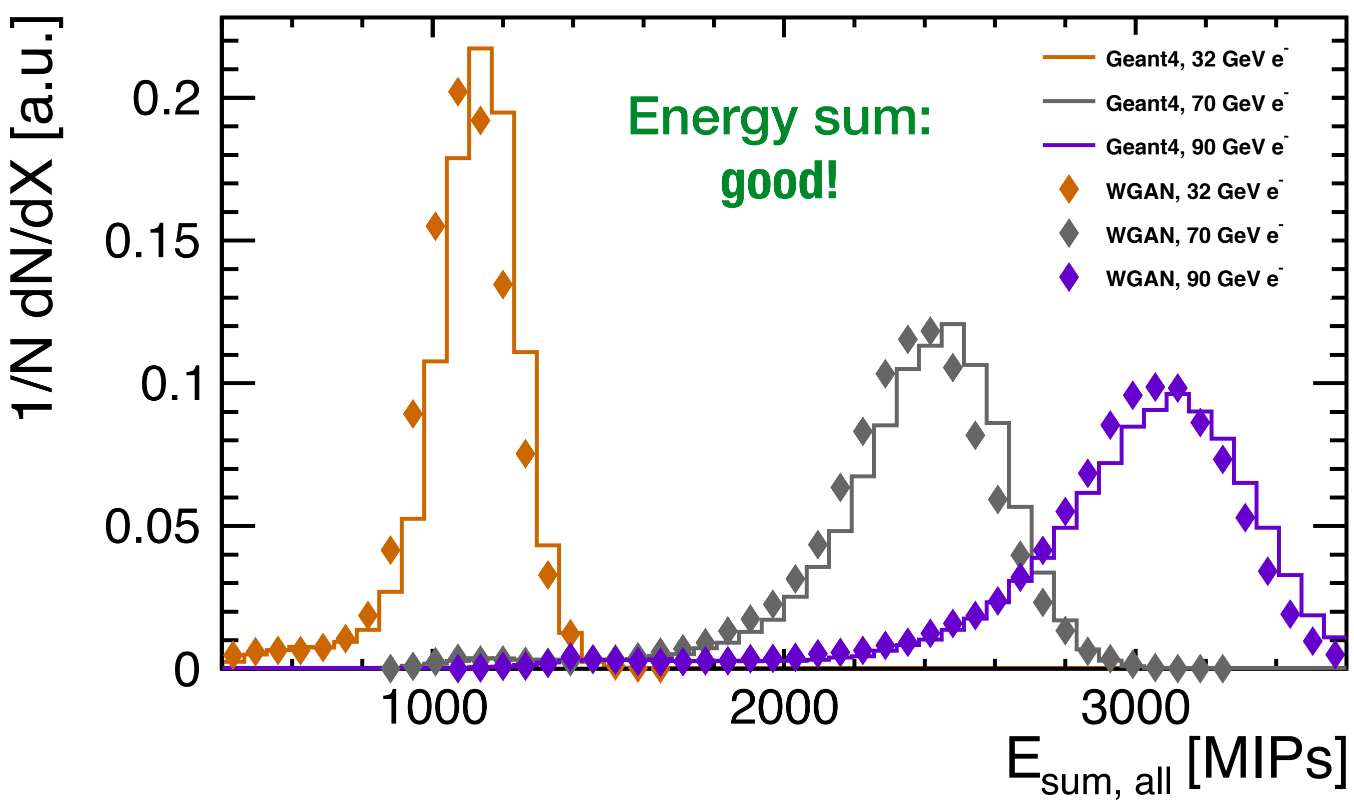


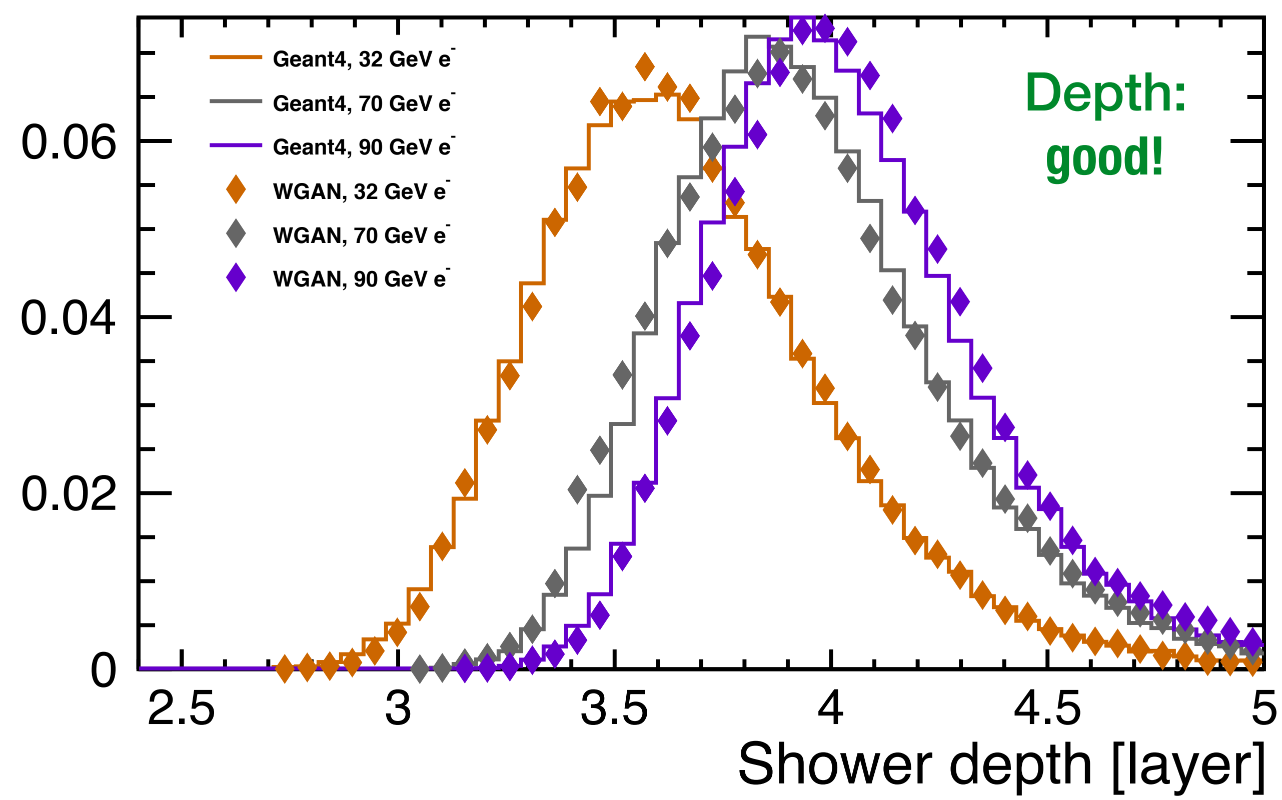




III. Physikalisches Institut RWTH AACHEN UNIVERSITY

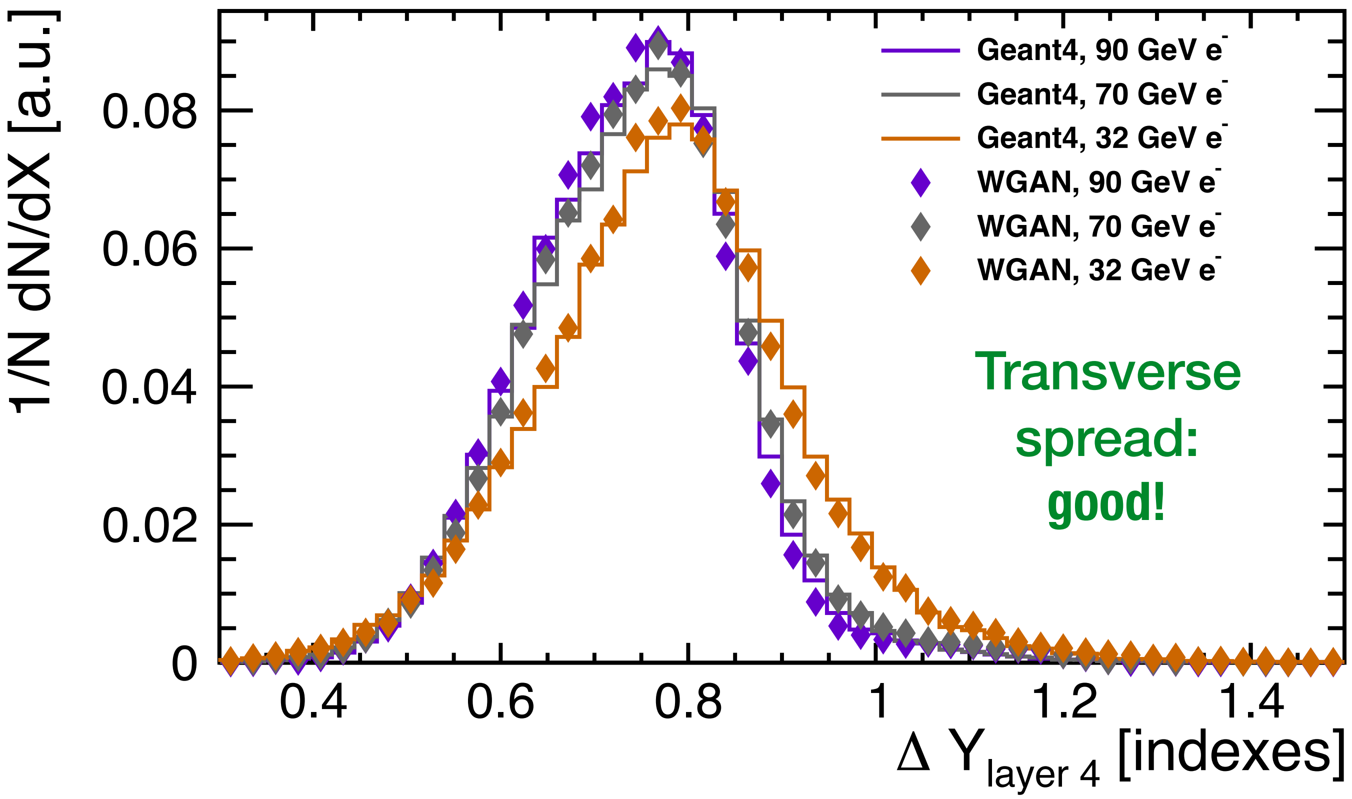
Distributions of 1D observables: Good

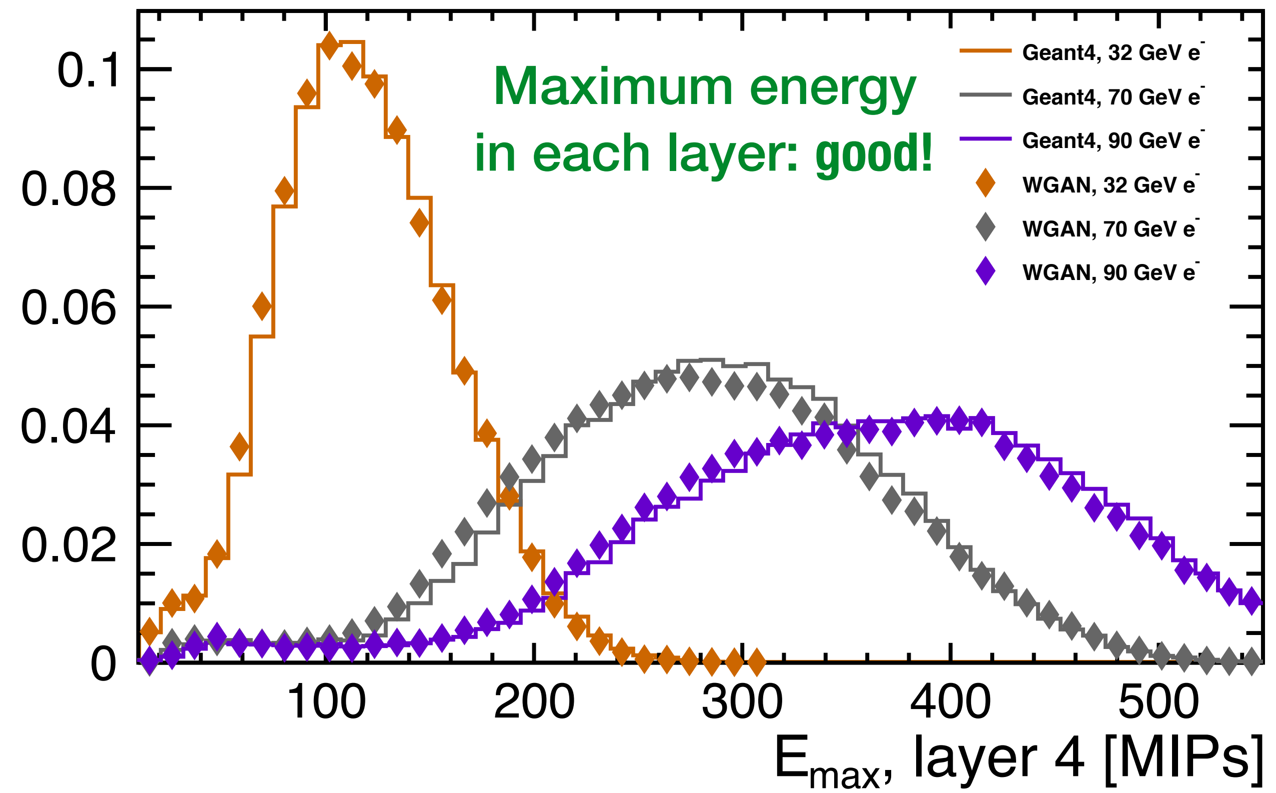




Note: **70GeV sample** not used in training.

Distributions of 1D observables: Good

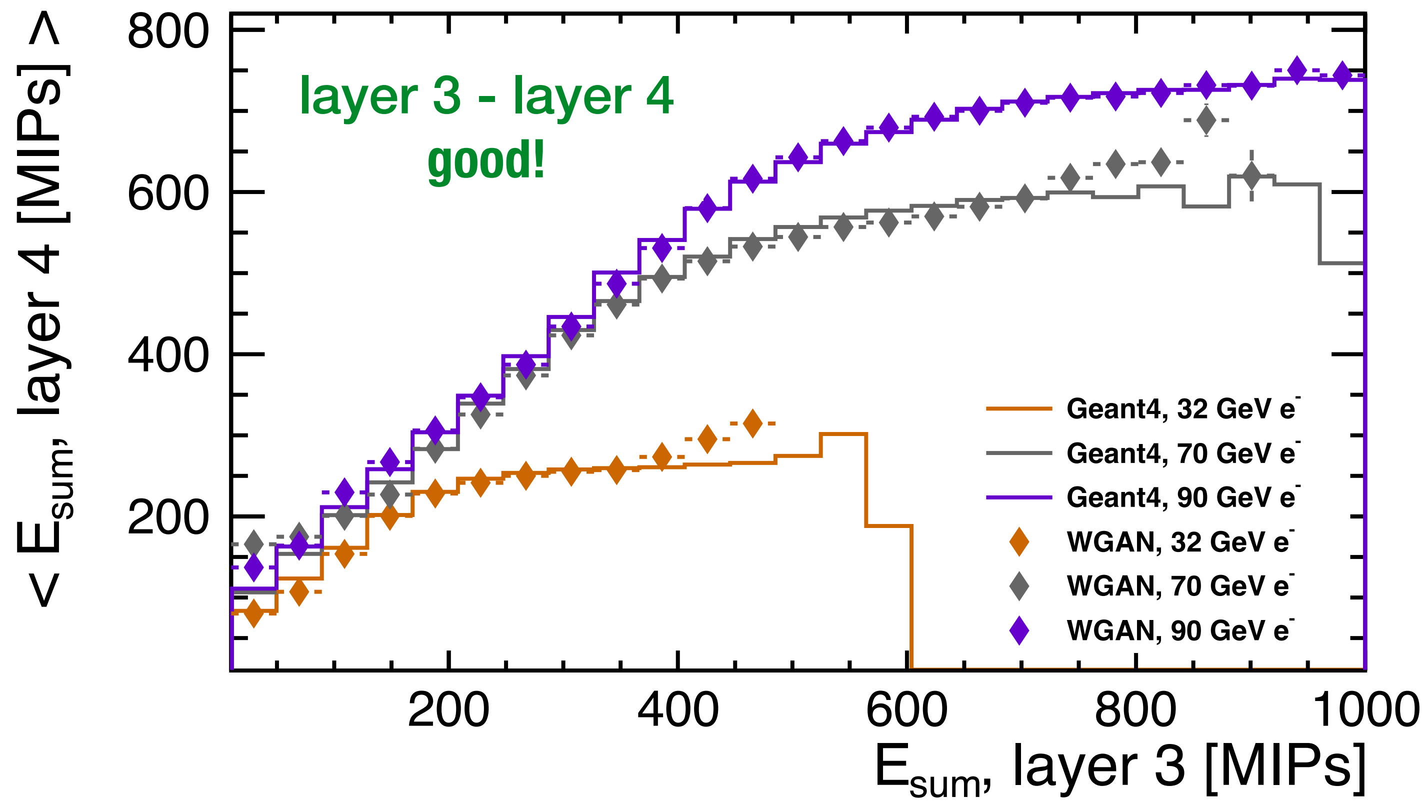


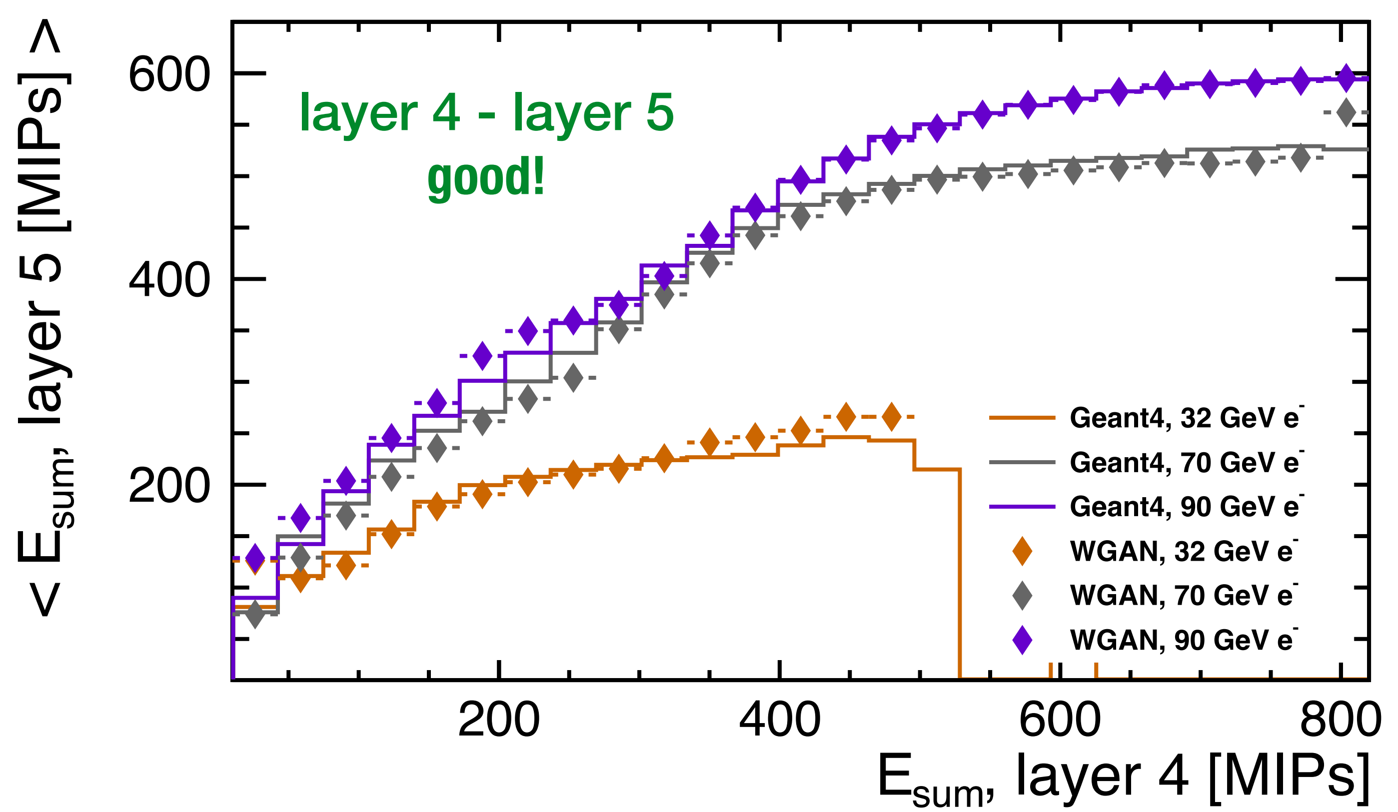


Note: **70GeV sample** not used in training.

Correlation between layers: Good

- Summed energy in one layer
  <—>
  sum in previous layer.



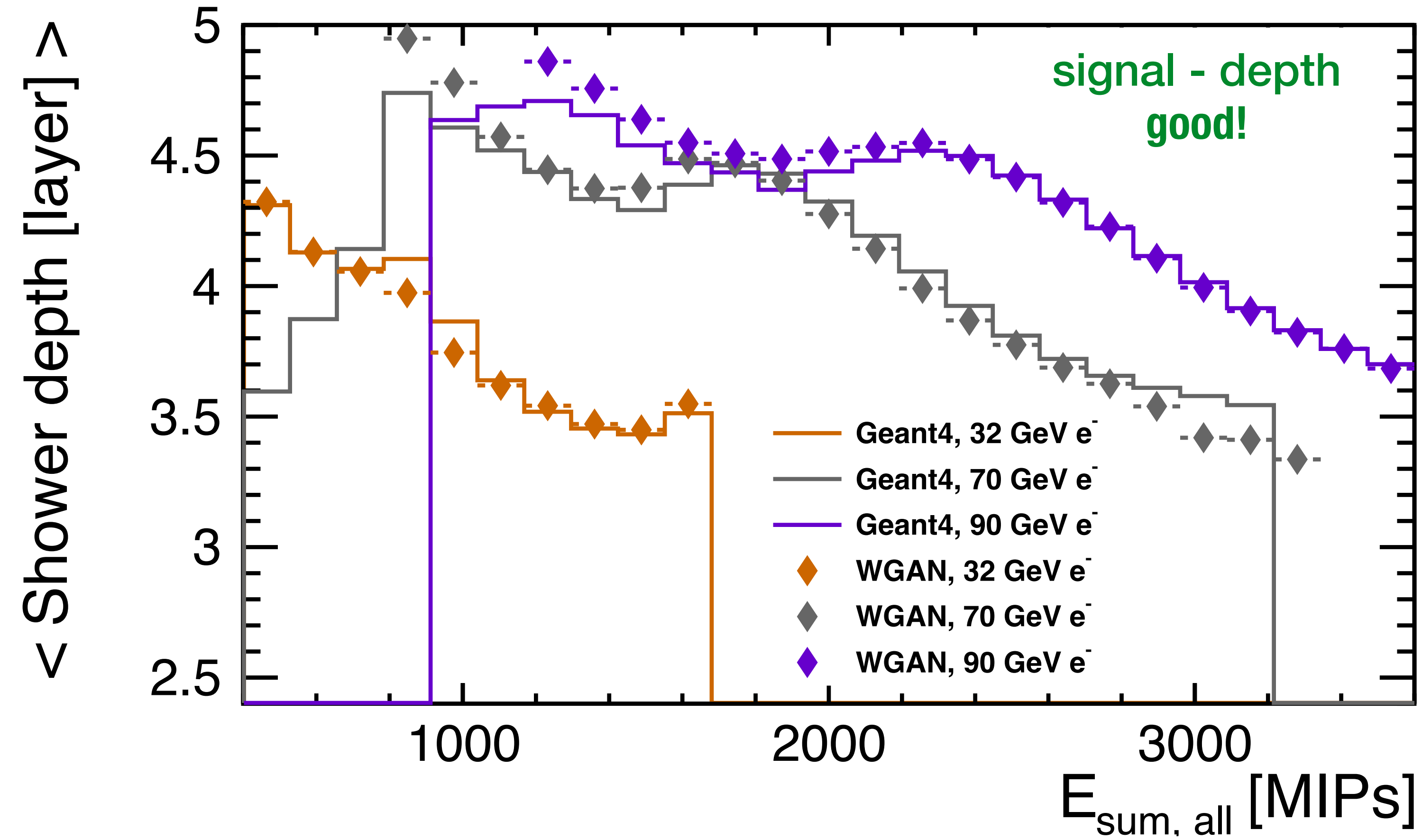


Note: **70GeV sample** not used in training.

Correlation of depth and signal sum: Good

- Specific sampling configuration:

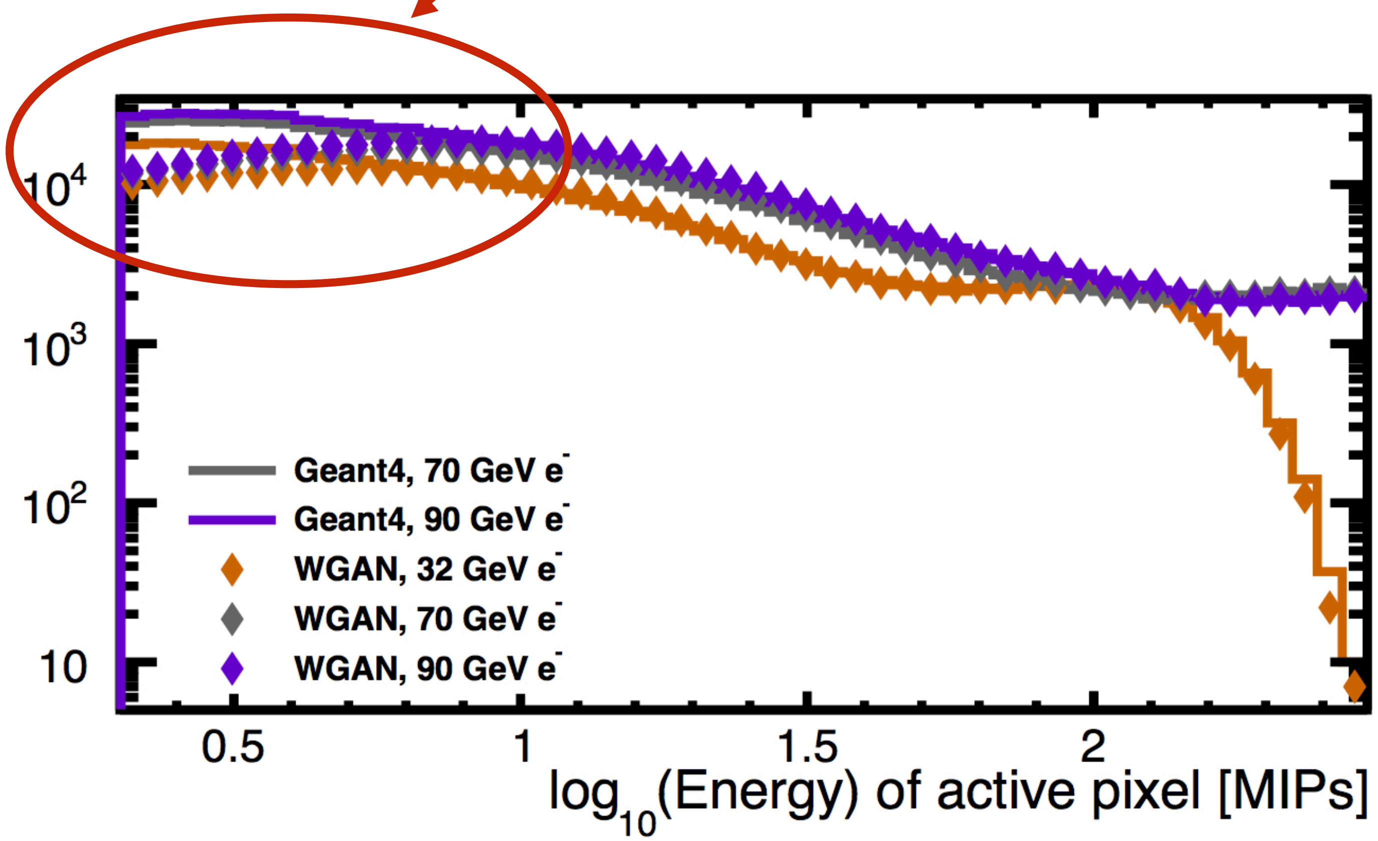
↑ shower depth
<—>
↓ summed signal



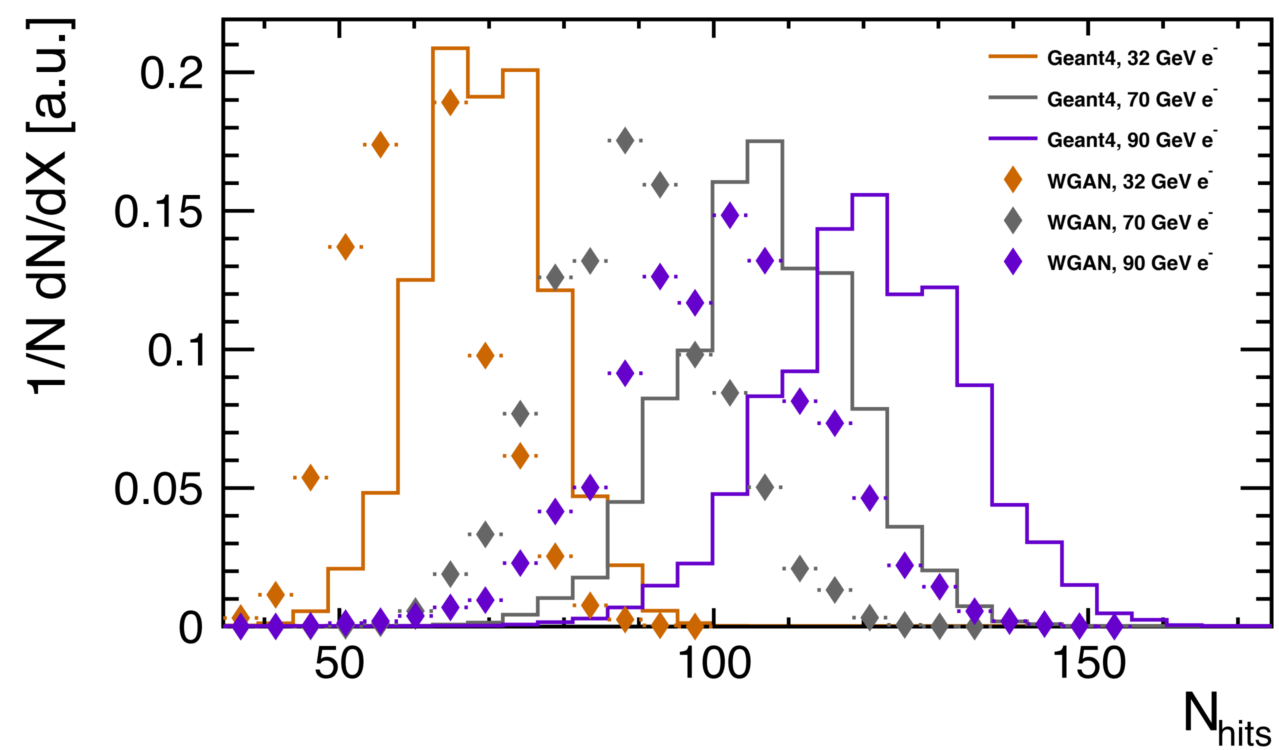
Note: **70GeV sample** not used in training.

Discrepancy at low energy densities

Underrepresented: Energy densities below 10MIPs/pixel.



➡ **N_{hits}: too low**



- ρ_E<10MIPs/pixel: **Only ~10% contribution** to the total shower signal.

O(x1000) faster calorimeter simulations possible

- Typical 20-90GeV e^- shower generated within **0.5-2 seconds** using **Geant 4**.

Different hardware setups, fixed generator network architectures

Method	Computing Setup	20 GeV e^-	Speed-up	90 GeV e^-	Speed-up	
Geant 4	*any*	O(500ms)	-	O(2000ms)	-	Slow
WGAN	Intel® Xeon® CPU E5-1620	52 ms	**x10**	52 ms	**x40**	
WGAN	NVIDIA® Quadro® K2000	3.6 ms	**x140**	3.6 ms	**x560**	
WGAN	NVIDIA® GTX™ 1080	0.3 ms	**x1660**	0.3 ms	**x6660**	Fast

WGAN evaluation: **No energy dependence.**

Summary: Calorimeter WGAN

- Generative models: promising **fast simulation tools** for particles' passage through matter.

This study:

- **Wasserstein GAN** concept instead of traditional GANs.
- **Conditioning** impact **position** & **incident** energy shower generating electrons.
- (**CMS HGCal prototype** as real-life calorimeter assumed.)

Key observations:

➡Many **reconstructed quantities** & key **correlations** of generated showers appear in many aspects **surprisingly close** to Geant 4 simulation.

➡Discrepancy for low energy densities.

➡Here: Inference step **O(1000)x faster** than Geant 4 .

➡Article in CSBS ➡Preprint on arXiv

Additional material

Wasserstein GANs

- Concept of Wasserstein loss (Arjovsky et al. 2017) is used.

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma} \big[\|x - y\| \big] \xleftrightarrow[\text{cool duality}]{} W(\mathbb{P}_r, \mathbb{P}_\theta) = \sup_{\|f\|_L \le 1} \mathbb{E}_{x\sim\mathbb{P}_r}[f(x)] - \mathbb{E}_{x\sim\mathbb{P}_\theta}[f(x)]$$

Mathematically motivated approach. Relevant for the application is this.

Proposition 1. *Let $\mathbb{P}_r$ and $\mathbb{P}_g$ be two distributions in $\mathcal{X}$, a compact metric space. Then, there is a 1-Lipschitz function f^* which is the optimal solution of $\max_{\|f\|_L \le 1} \mathbb{E}_{y\sim\mathbb{P}_r}[f(y)] - \mathbb{E}_{x\sim\mathbb{P}_g}[f(x)]$. Let π be the optimal coupling between $\mathbb{P}_r$ and $\mathbb{P}_g$, defined as the minimizer of: $W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\pi\in\Pi(\mathbb{P}_r,\mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\pi}[\|x-y\|]$ where $\Pi(\mathbb{P}_r, \mathbb{P}_g)$ is the set of joint distributions $\pi(x,y)$ whose marginals are $\mathbb{P}_r$ and $\mathbb{P}_g$, respectively. Then, if f^* is differentiable[‡], $\pi(x = y) = 0$[§], and $x_t = tx + (1-t)y$ with $0 \le t \le 1$, it holds that $\mathbb{P}_{(x,y)\sim\pi}\left[\nabla f^*(x_t) = \frac{y - x_t}{\|y - x_t\|}\right] = 1$.*

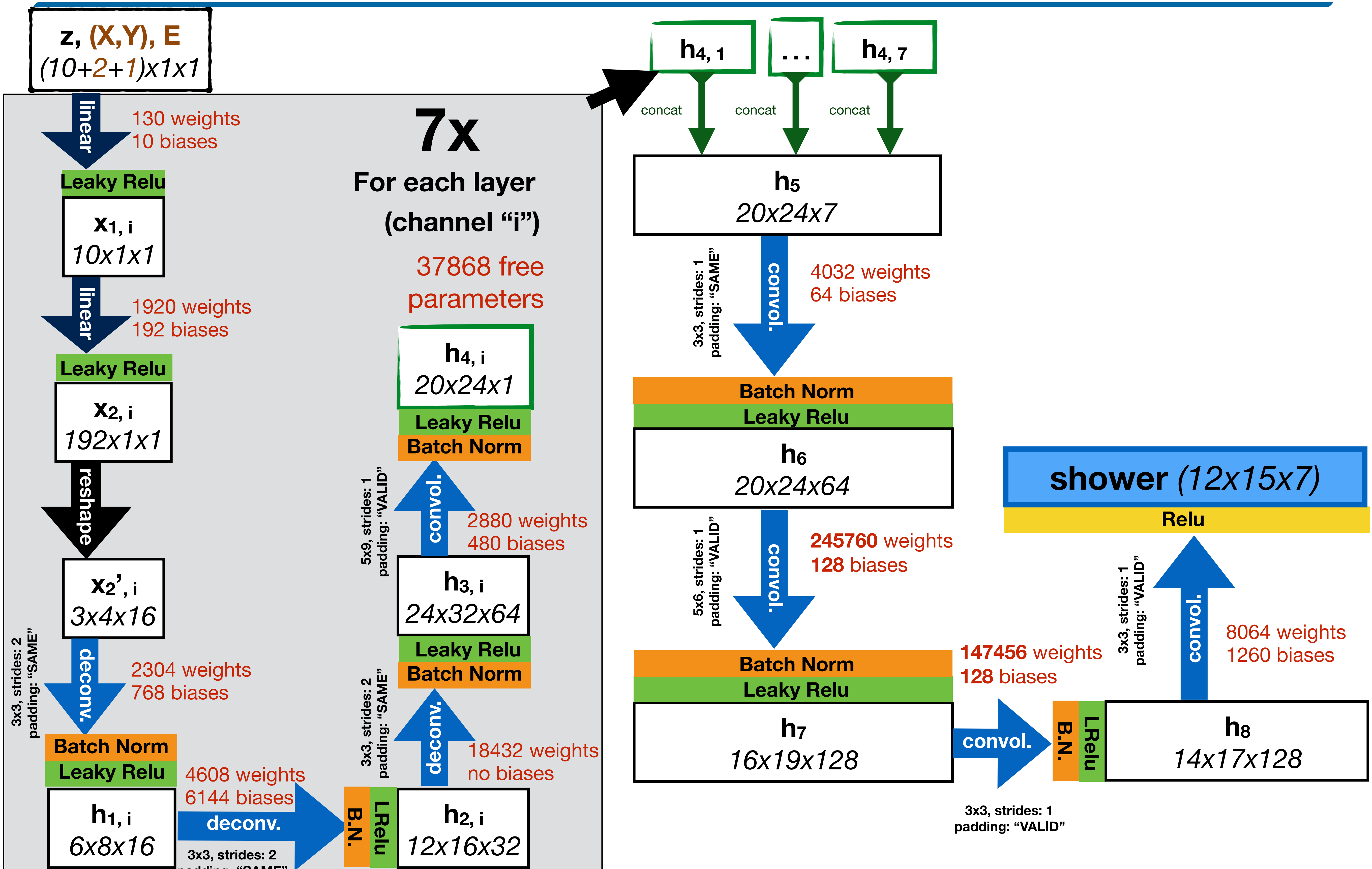
Corollary 1. *f^* has gradient norm 1 almost everywhere under $\mathbb{P}_r$ and $\mathbb{P}_g$.*

$$\boxed{L = \underbrace{\mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}[D(\tilde{\boldsymbol{x}})] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathbb{P}_r}[D(\boldsymbol{x})]}_{\text{Original critic loss}} + \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{\boldsymbol{x}}\sim\mathbb{P}_{\hat{\boldsymbol{x}}}}\left[(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1)^2\right]}_{\text{Our gradient penalty}}.}$$

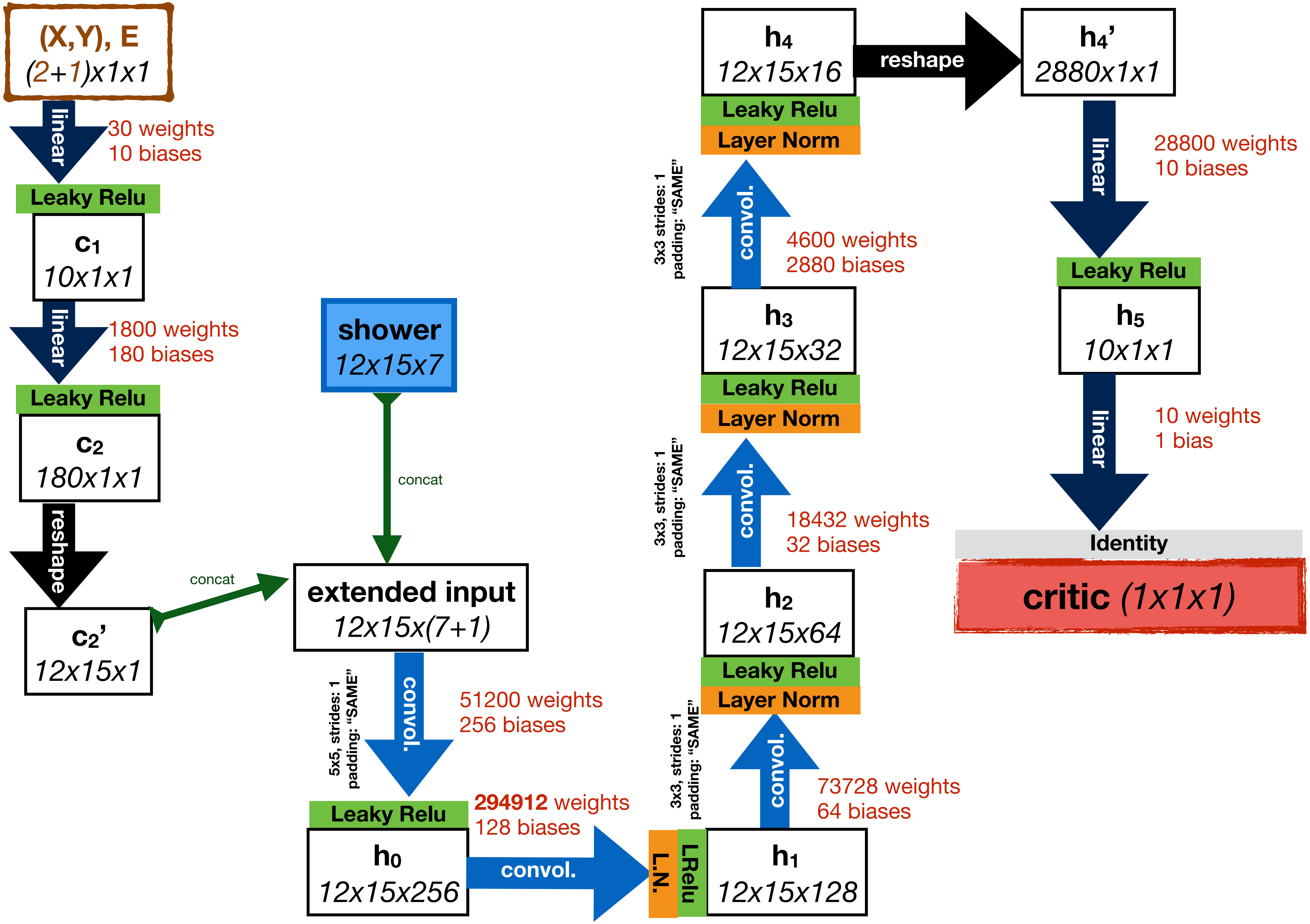
➡ **Critic *D(x)* instead of a discriminator network.**

➡ L is a direct measure for the convergence of the training.

Generator network with ~672k free parameters



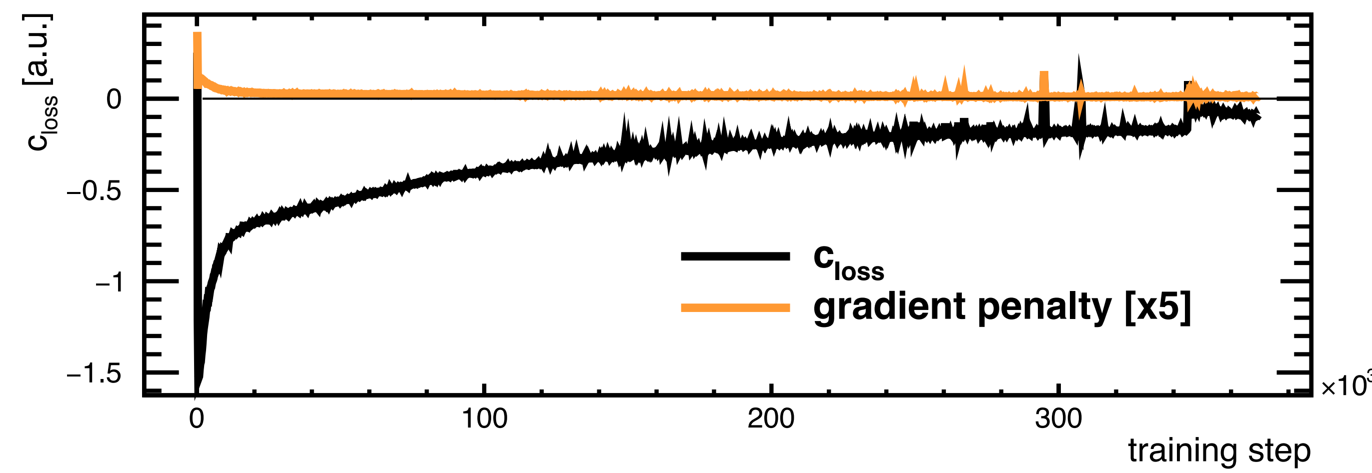
Critic network with ~477k free parameters



Logarithmic intensity: energy’ = log(1+energy)

✓Training converges.

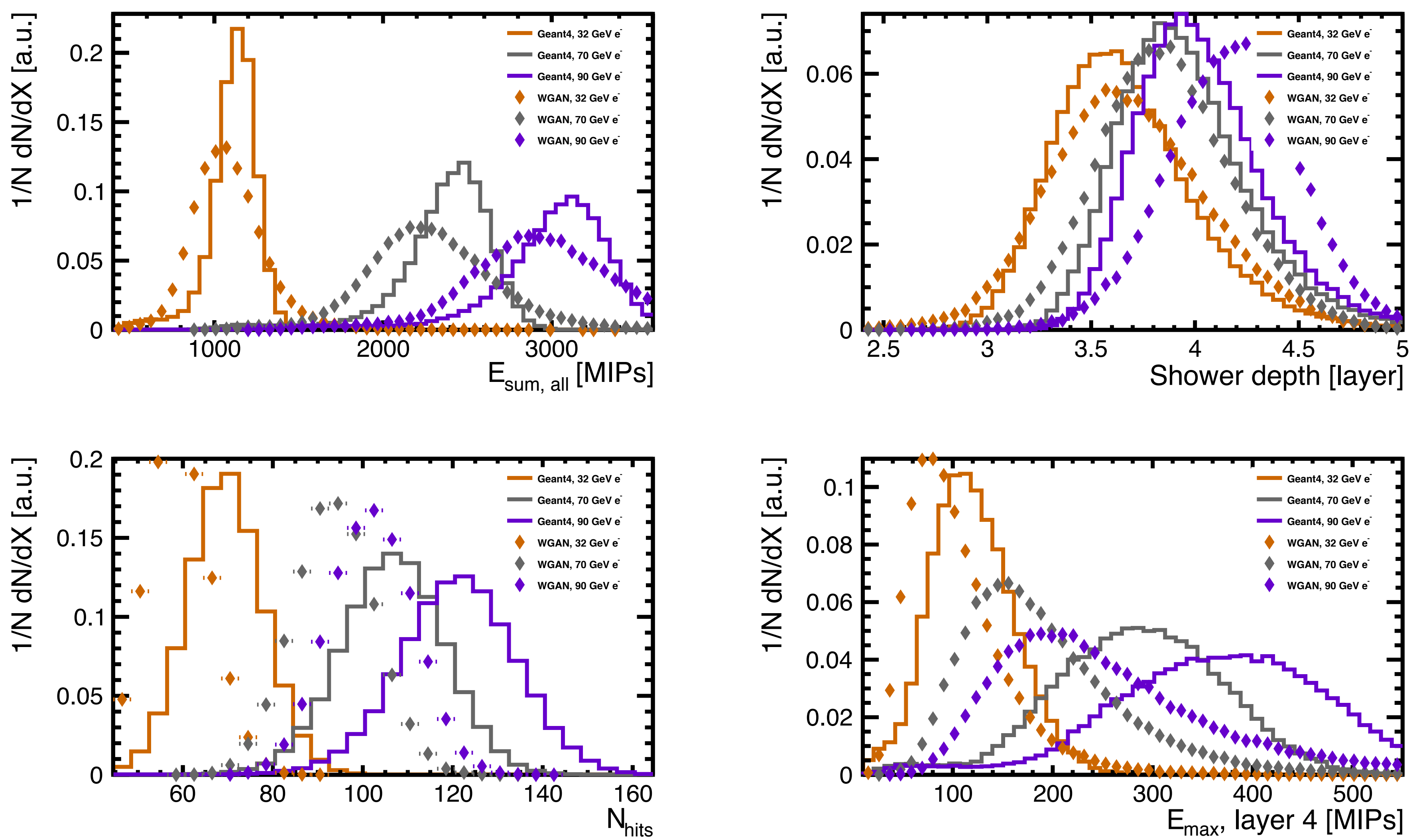
- …less smoothly.



- E_{max} mis-modeled!

- …all others, too.

➡No improvement.



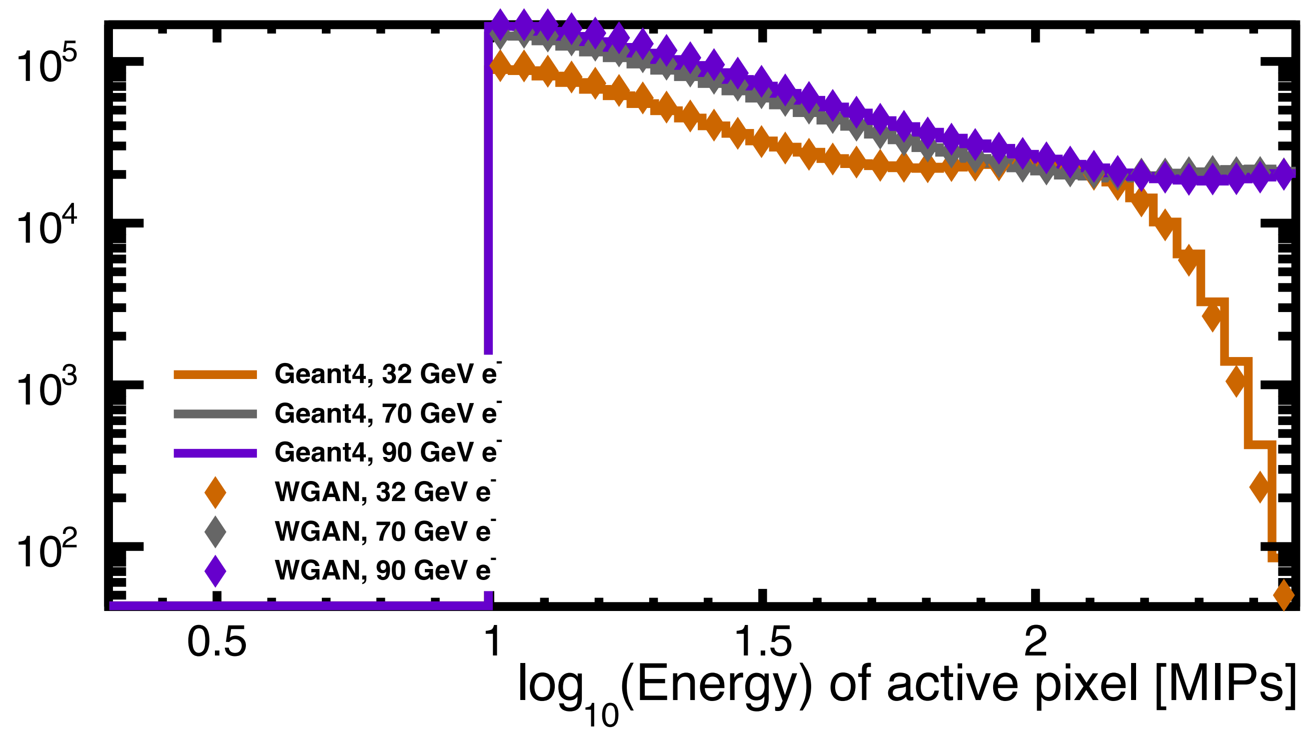
Thorben Quast
quast@physik.rwth-aachen.de

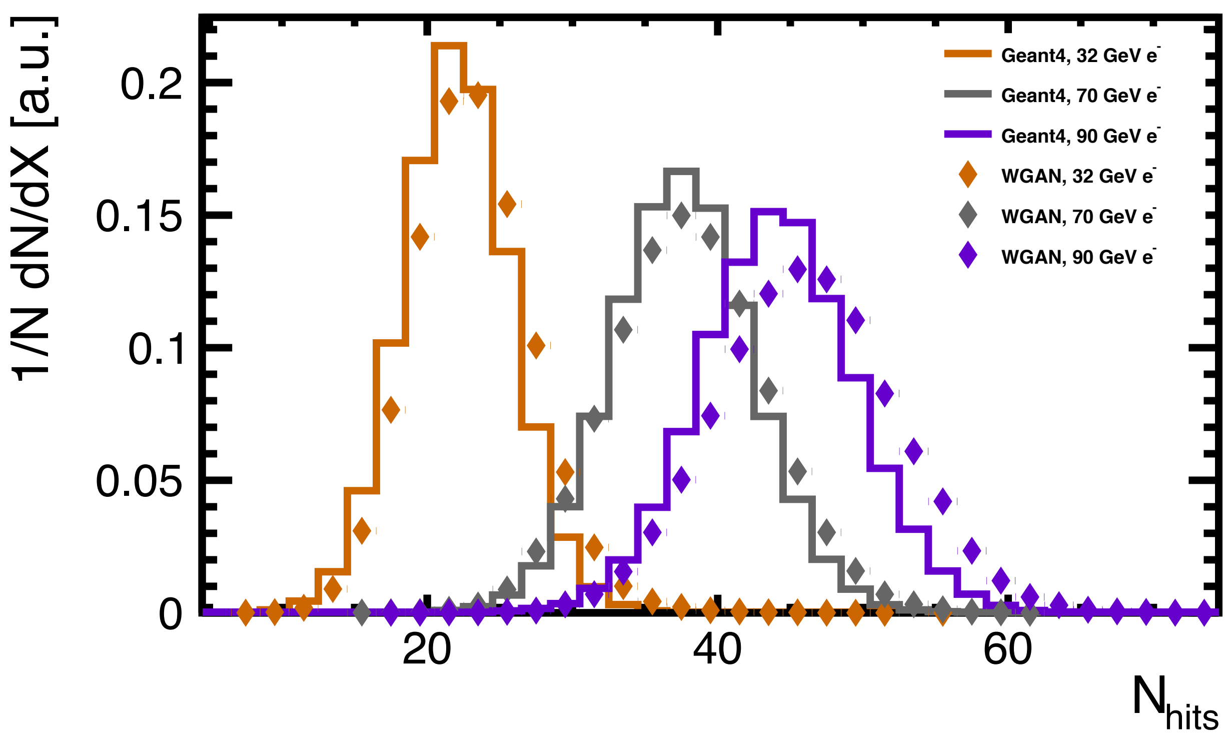
Supplementary benchmark:

10 MIP pixel cut at evaluation

Supplementary benchmark: N_{hits}

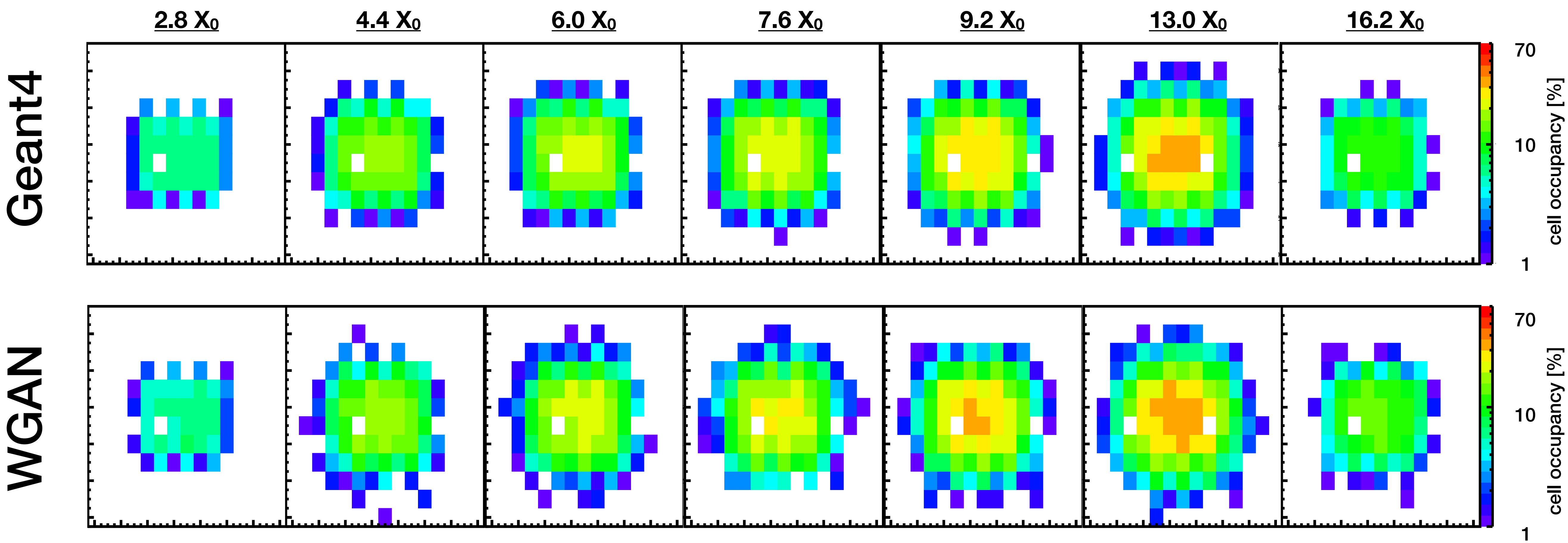
10 MIP cut per pixel





Supplementary benchmark: N_{hits}

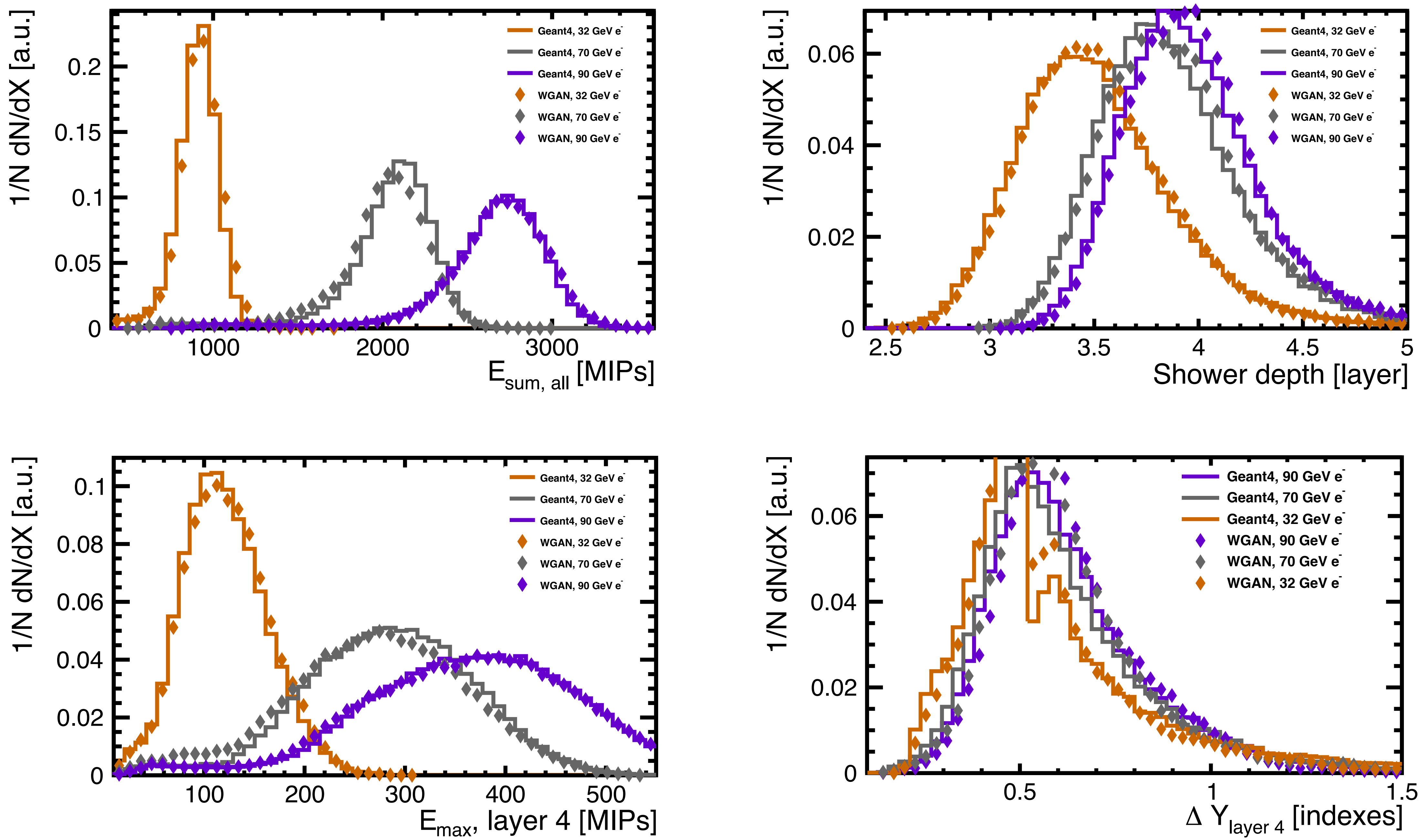
10 MIP cut per pixel



Thorben Quast
quast@physik.rwth-aachen.de

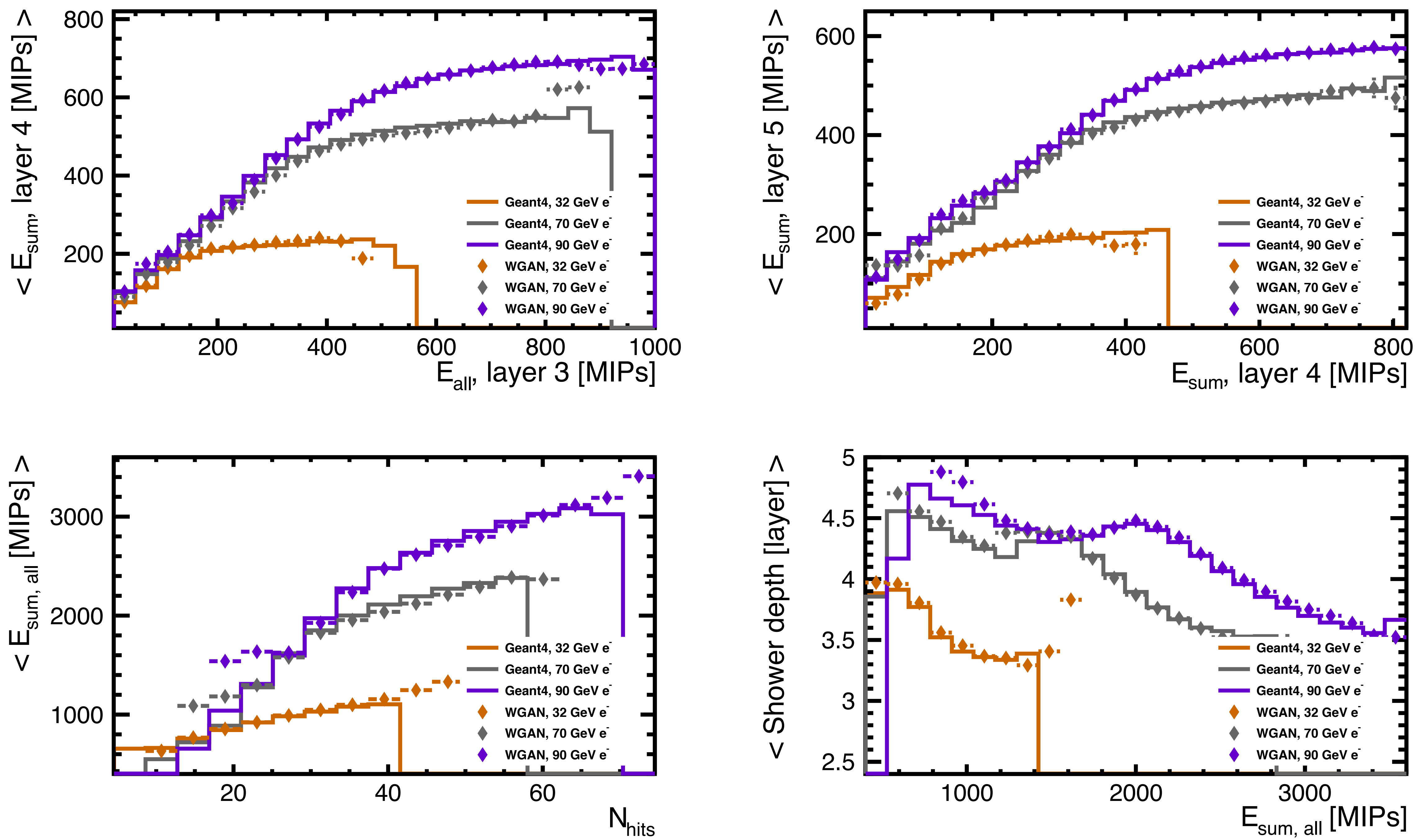
Supplementary benchmark: Observables

10 MIP cut per pixel



Supplementary benchmark: Correlations

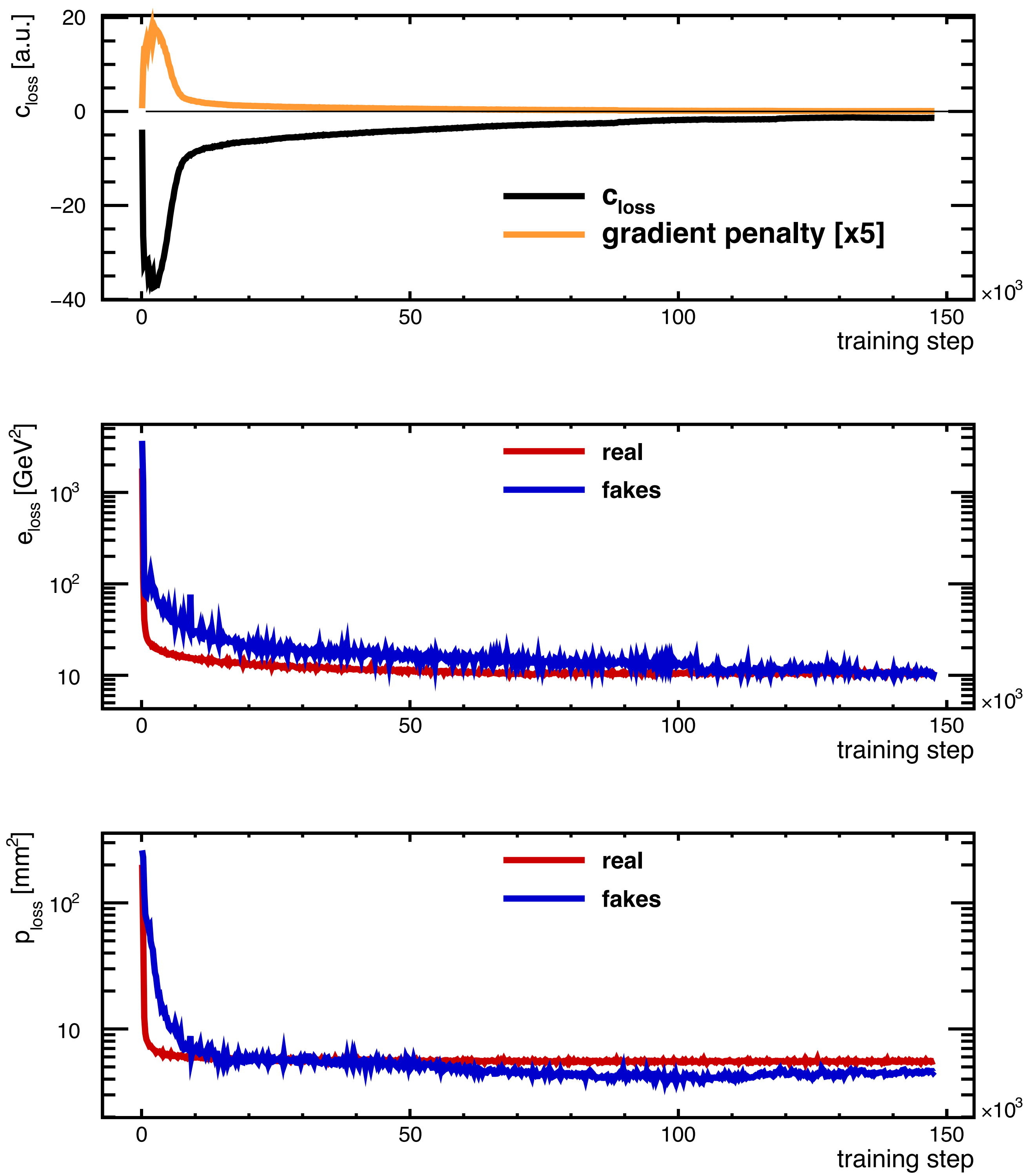
10 MIP cut per pixel

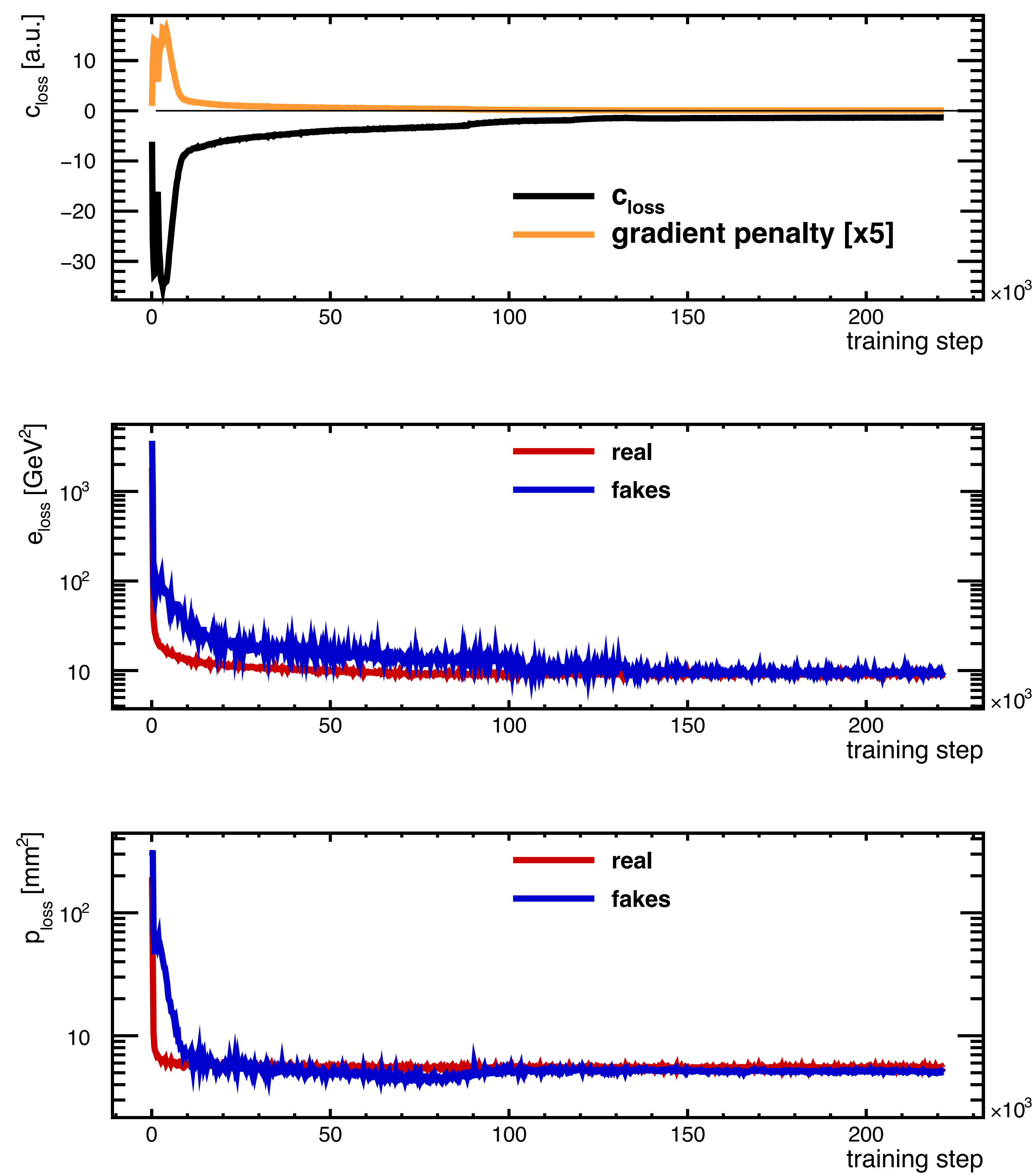


Do dead areas need to be masked?

Comparison: Costs

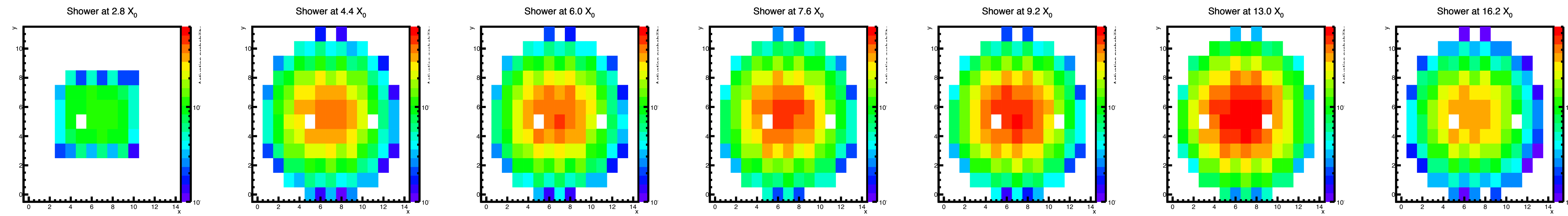
no masking of dead cells vs. **with masking of dead cells**



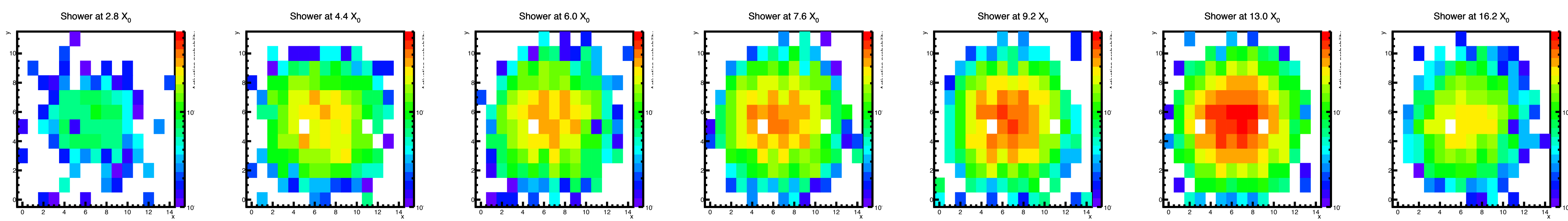


Comparison: Occupancy

90 GeV e- Geant4

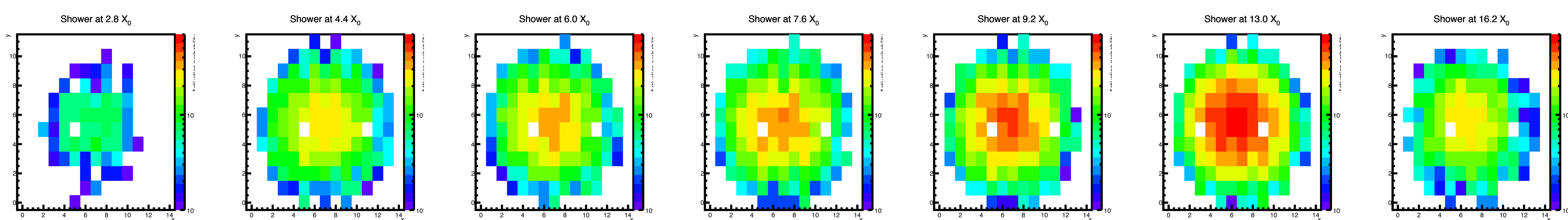


WGAN: no masking of dead cells



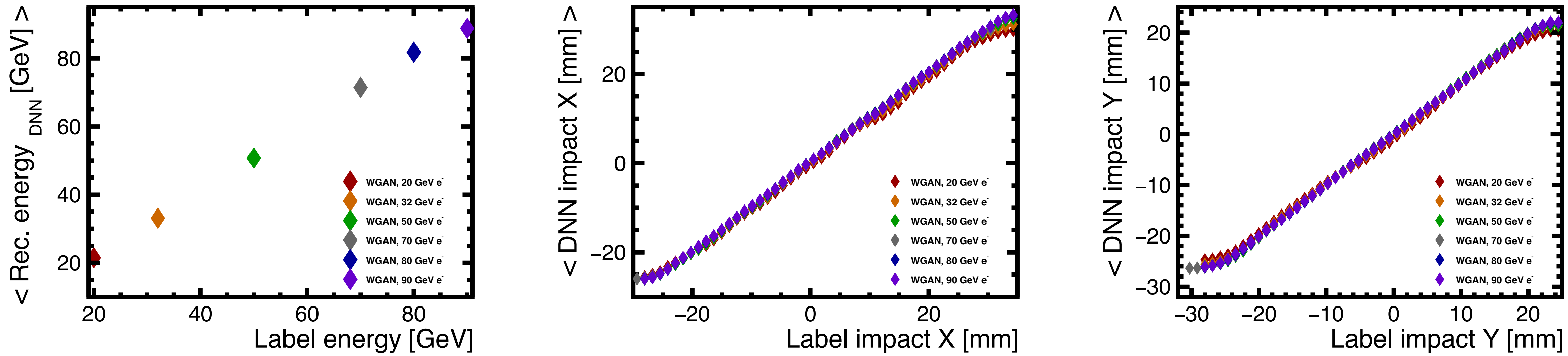
vs.

WGAN: with masking of dead cells



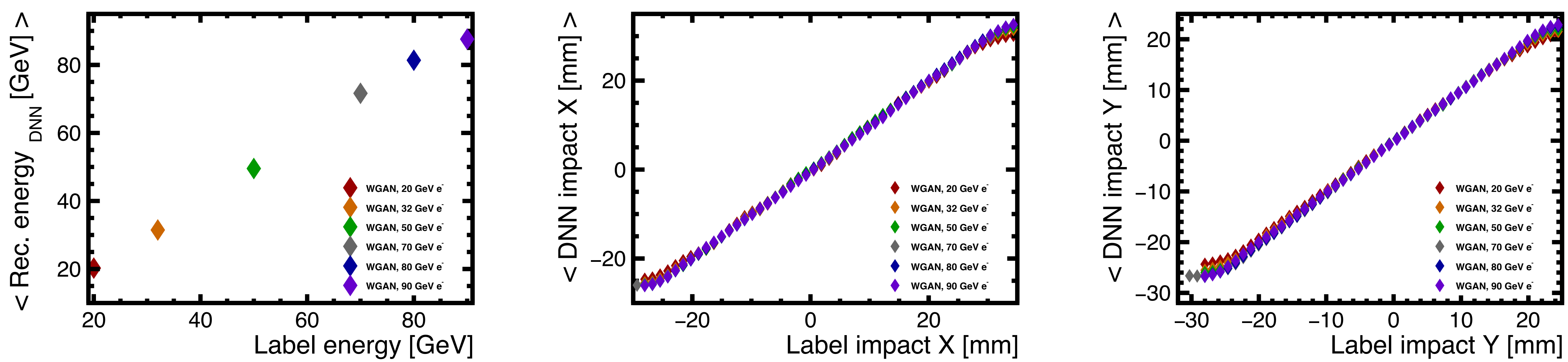
Comparison: Label dependence

no masking of dead cells

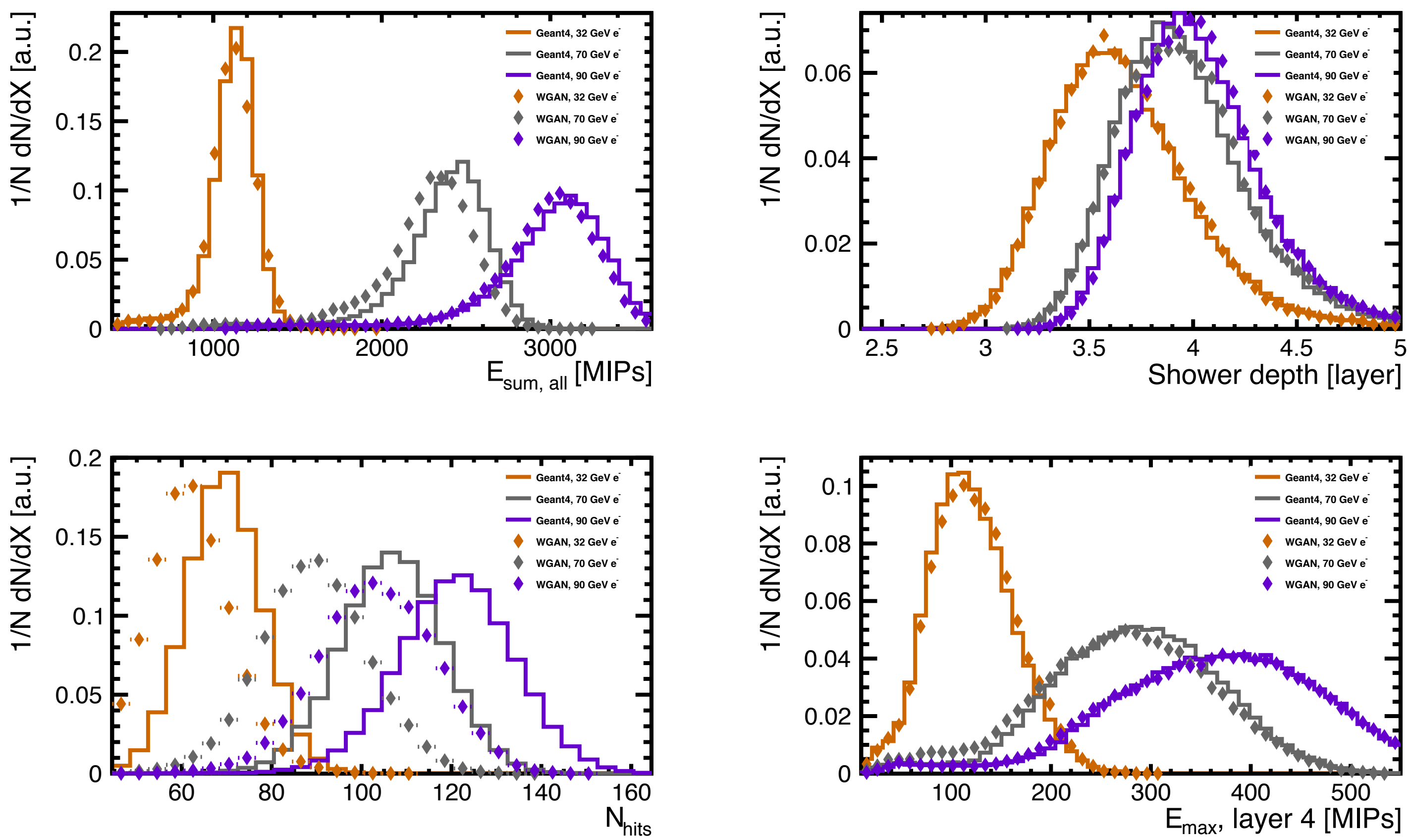


vs.

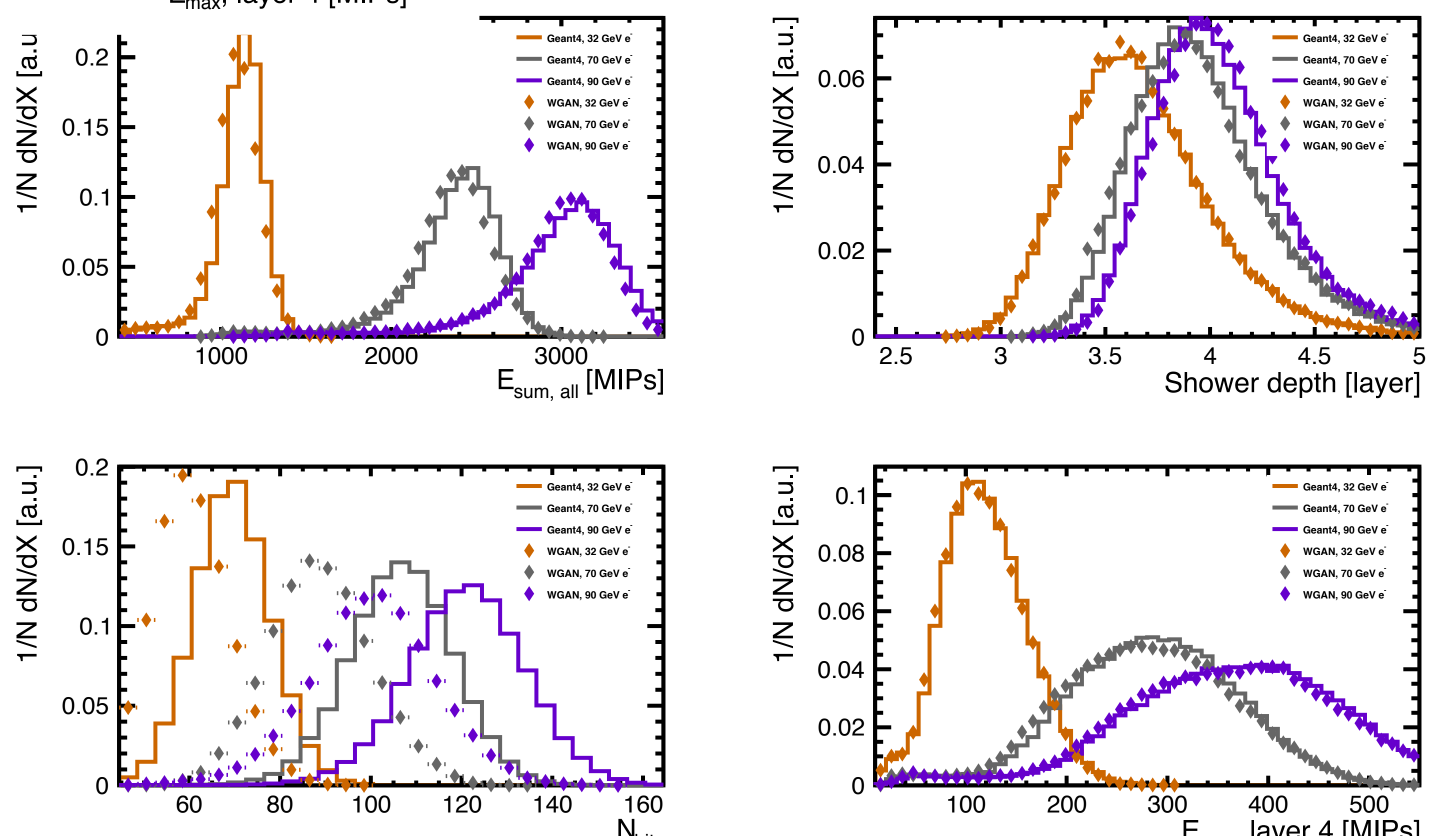
with masking of dead cells



Comparison: Observables

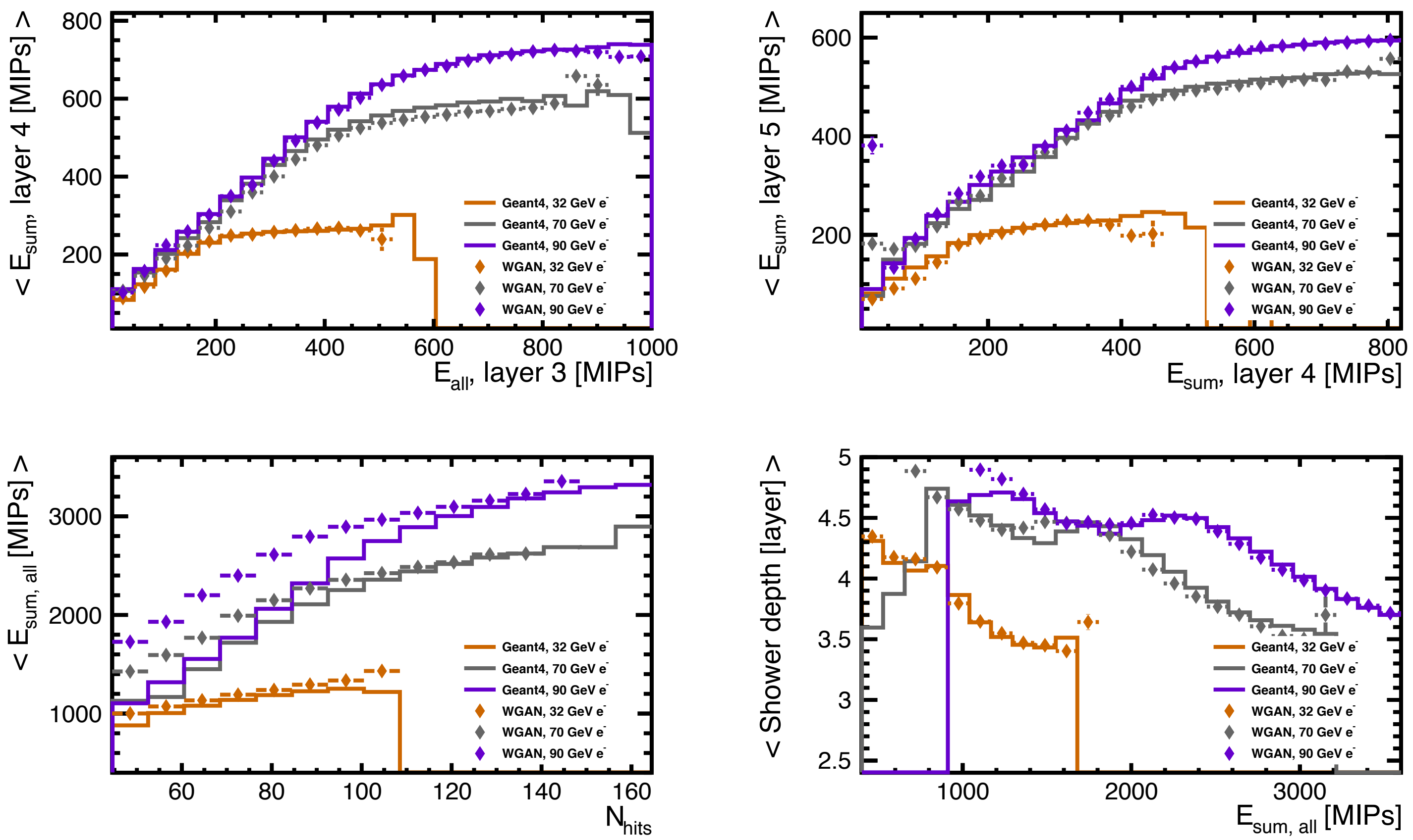


no masking of dead cells

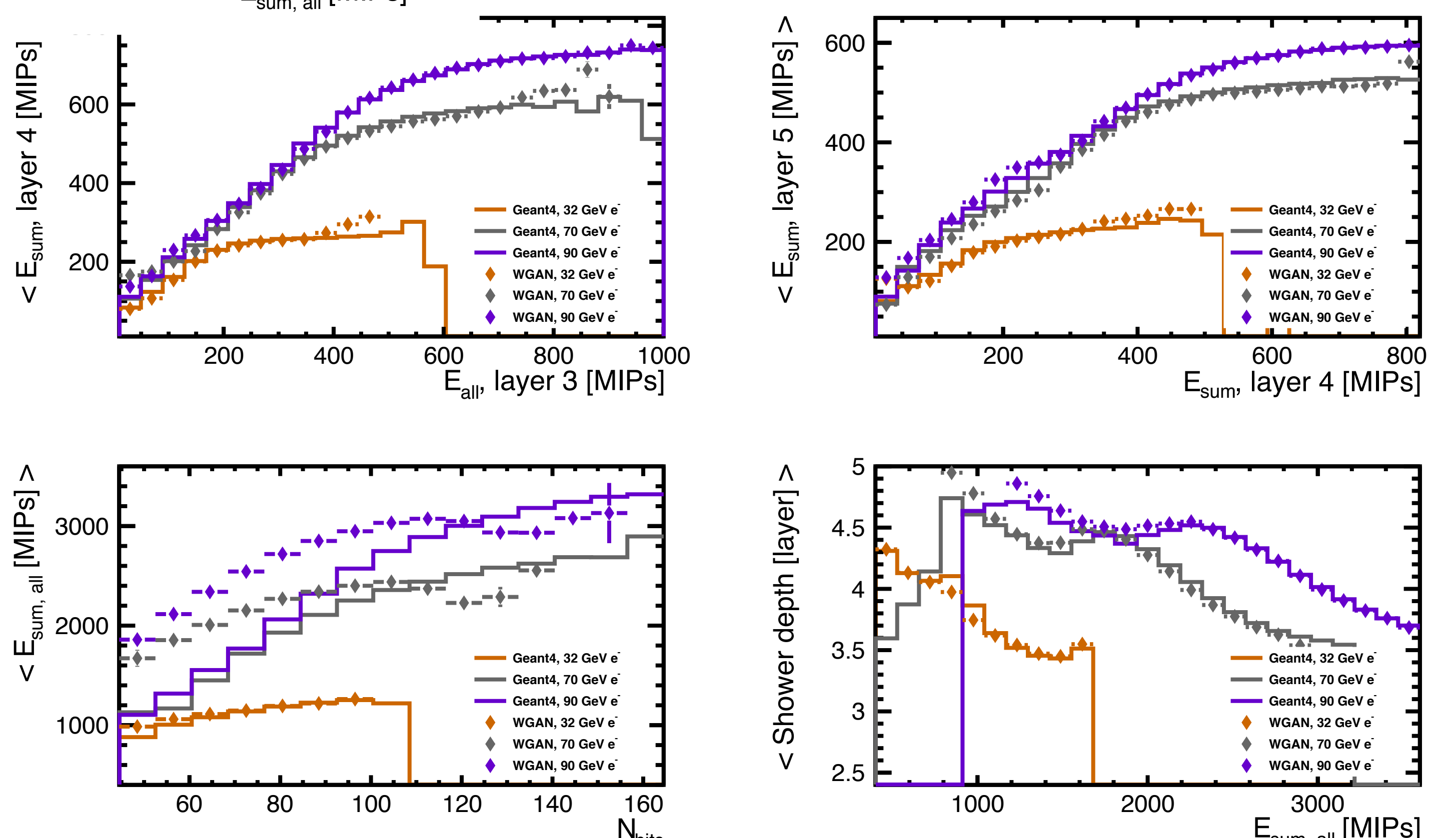


with masking of dead cells

Comparison: Correlations



no masking of dead cells



with masking of dead cells