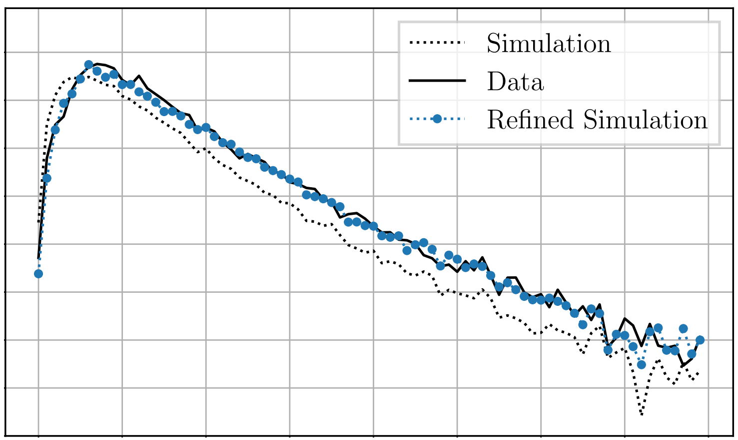
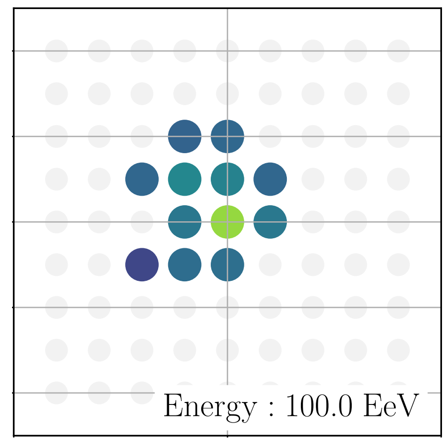
Adversarial training in astroparticle physics

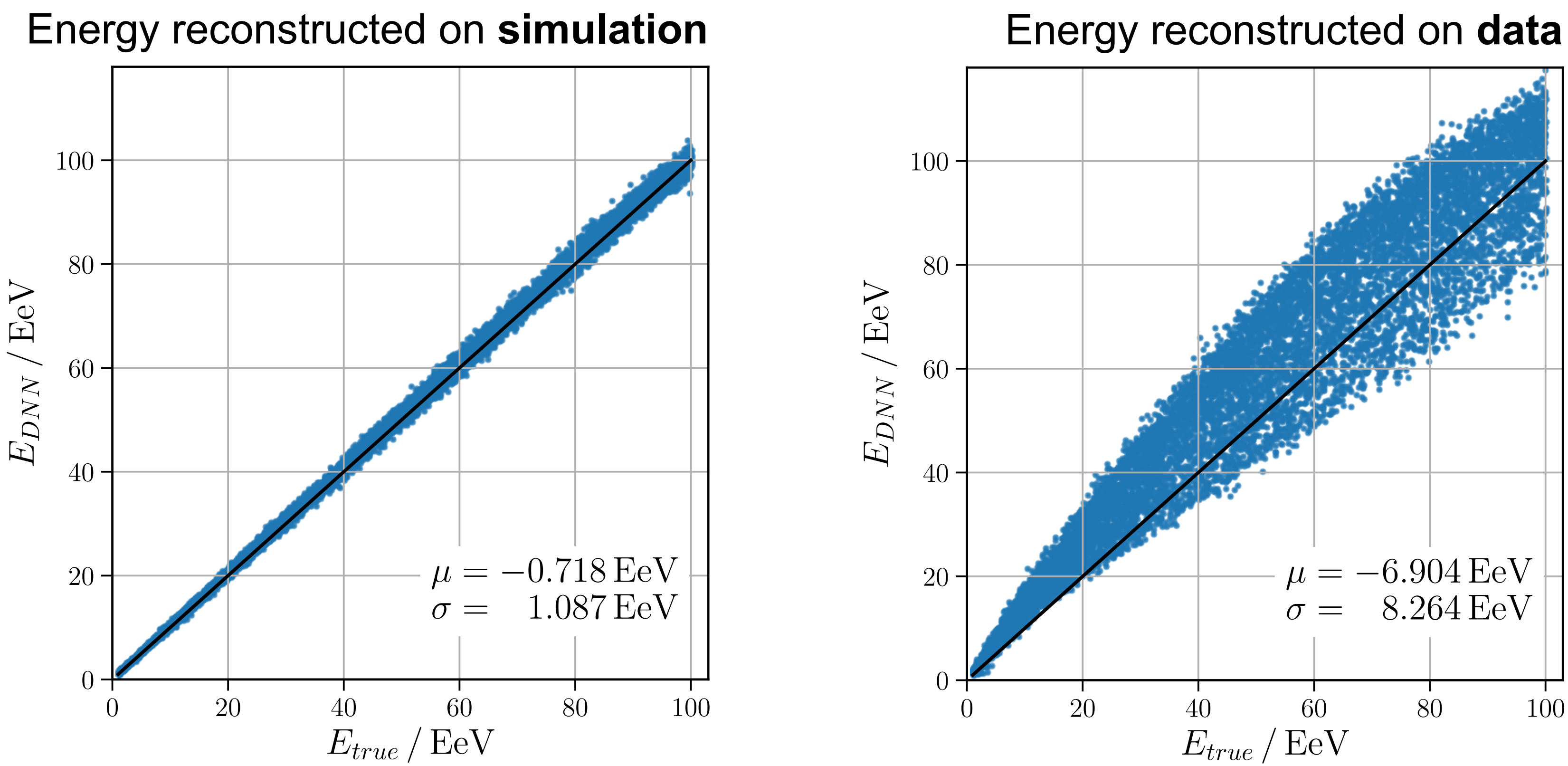
Generating and refining particle detector simulations using the Wasserstein distance in adversarial networks – arXiv:1802.03325





Lukas Geiger, Martin Erdmann, Jonas Glombitza, David Schmidt

Problem: Simulation – data mismatches in supervised training



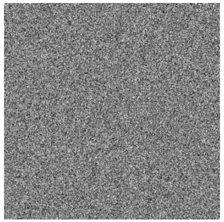
Training with inexact simulations → Bad generalization onto data

RWTH AACHEN UNIVERSITY

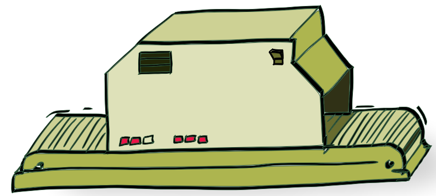
Adversarial training to reduce simulation – data mismatches

Recap – Generative Adversarial Network

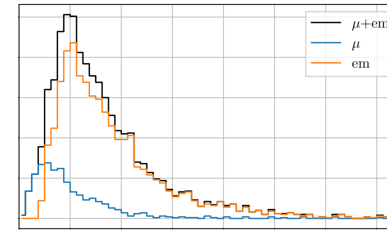
Noise



Generator

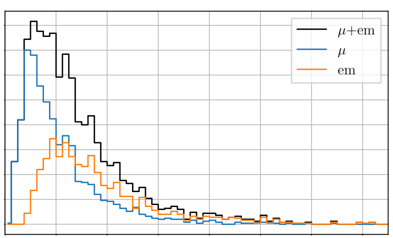


Generated time trace

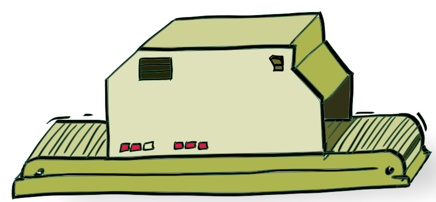


Refining Adversarial Network to improve existing simulations

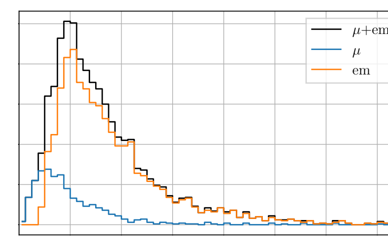
Simulated time trace



Refiner

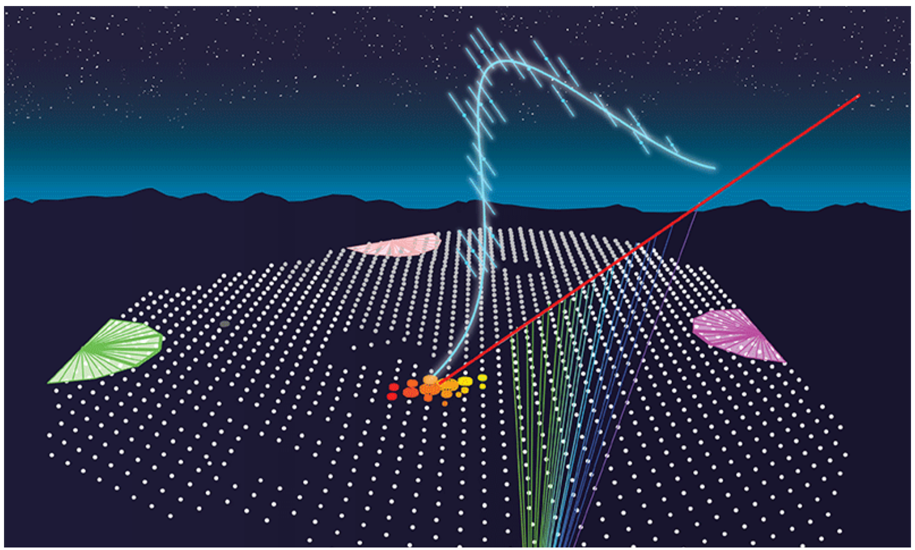


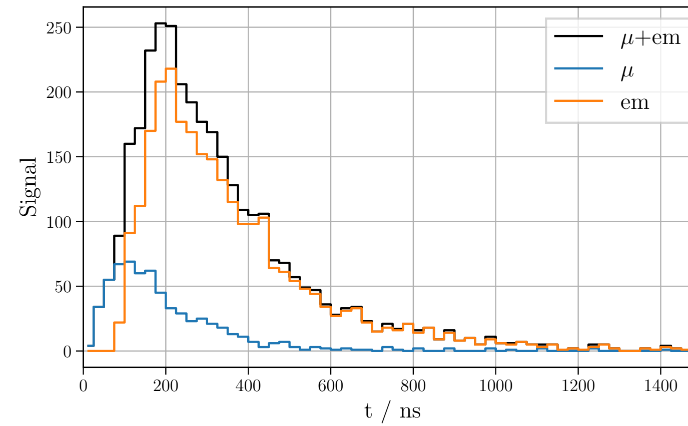
Refined time trace



Simulated data for training and evaluation

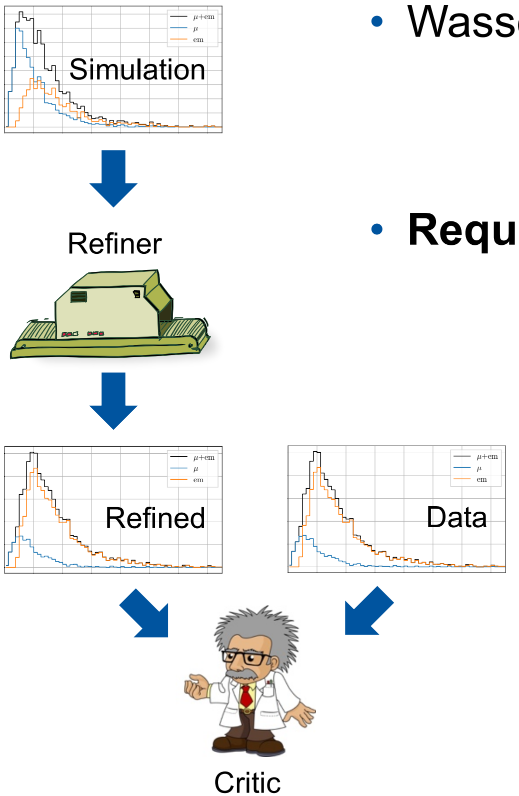
- Proton-induced vertical air showers
- 9 × 9 detectors each with 80 time bins
- 10^5 events with E = (1, ..., 100) EeV following a flat distribution





- **Data** — **30%** electromagnetic, **70%** muonic
- **Simulation** — **70%** electromagnetic, **30%** muonic, 50% detector noise, 50% event-by-event fluctuations

Refining Adversarial Network

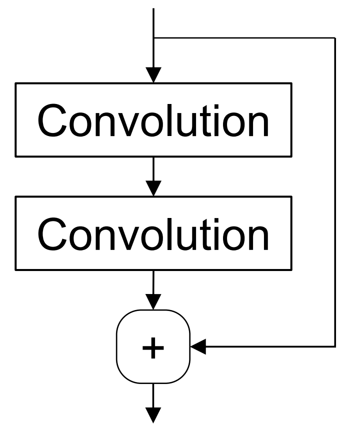


- Wasserstein distance as similarity measure in loss function:

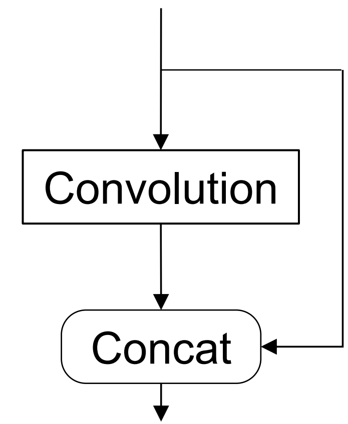
$$W(P_r, P_\theta) = \sup_{f \in Lip_1} \mathbb{E}_{x\sim P_r}[f_w(x)] - \mathbb{E}_{\tilde{x}\sim P_\theta}[f_w(\tilde{x})]$$

- **Requires:** Similar label (energy) distribution of **Simulation** and **Data**

Refiner: ResNet

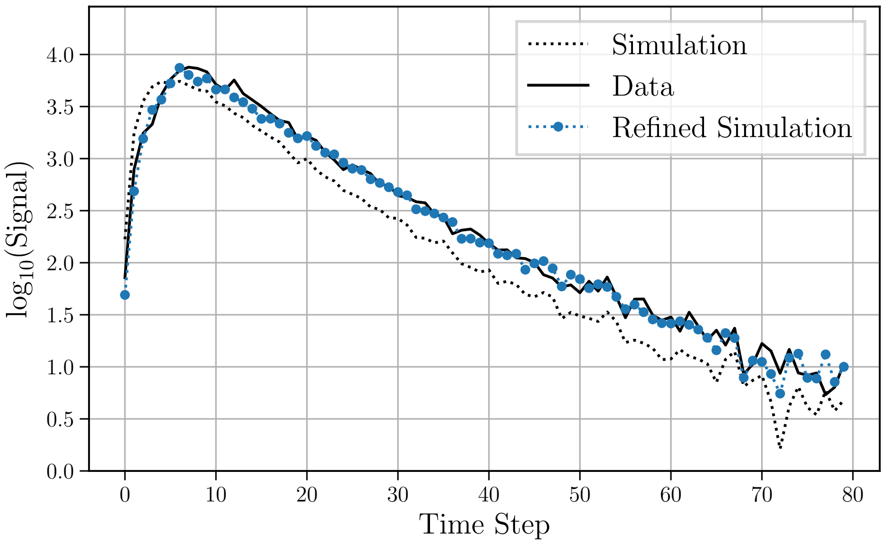


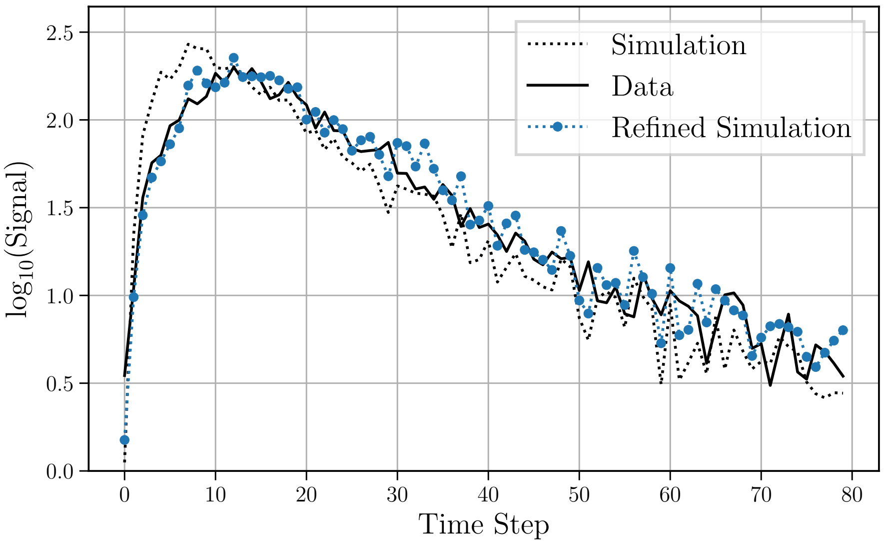
Critic: DenseNet



Refined signal traces

Signal traces of an event with energy E = 69 EeV

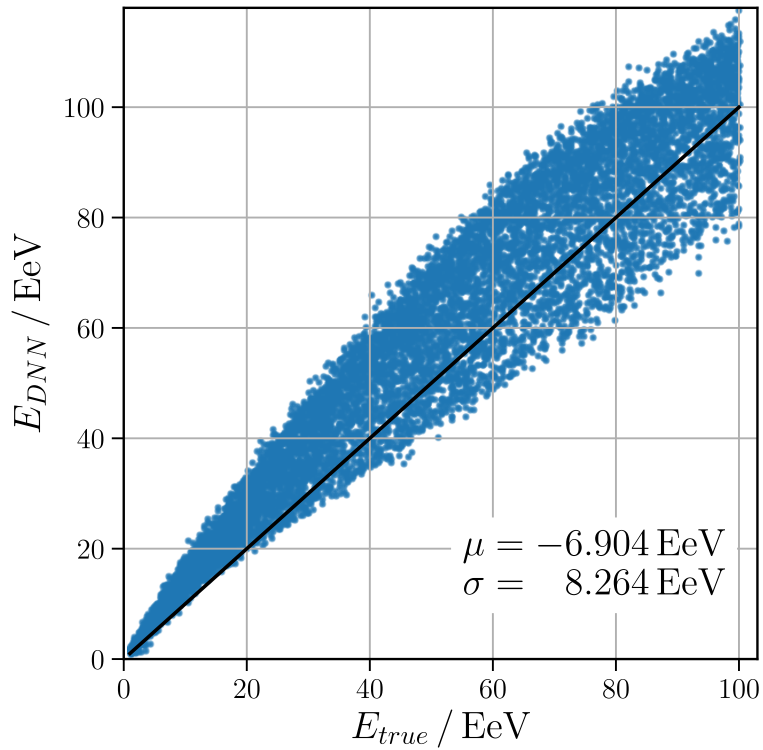




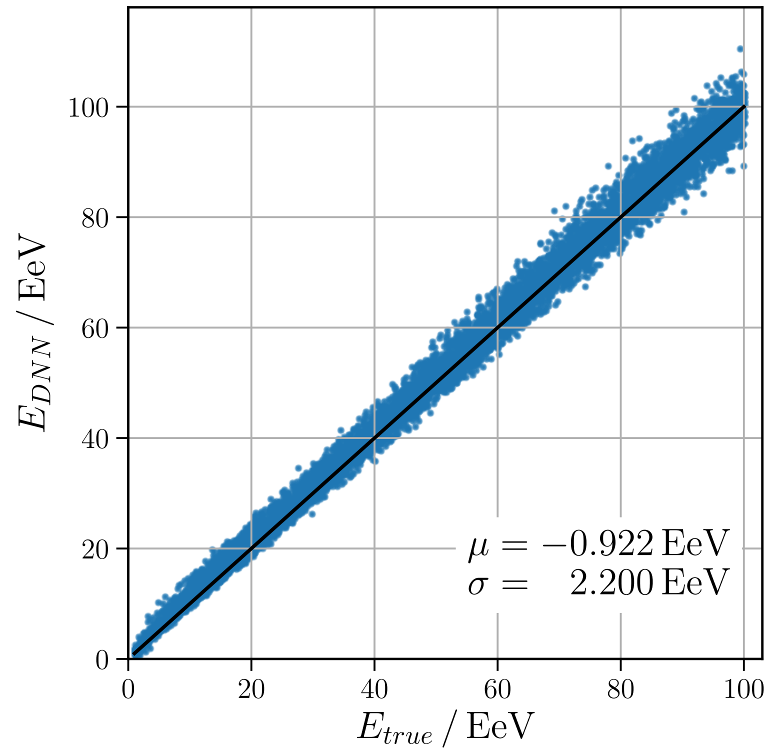
Refiner is able to shift simulation towards data

Supervised training evaluated on data

Trained on **original simulation**



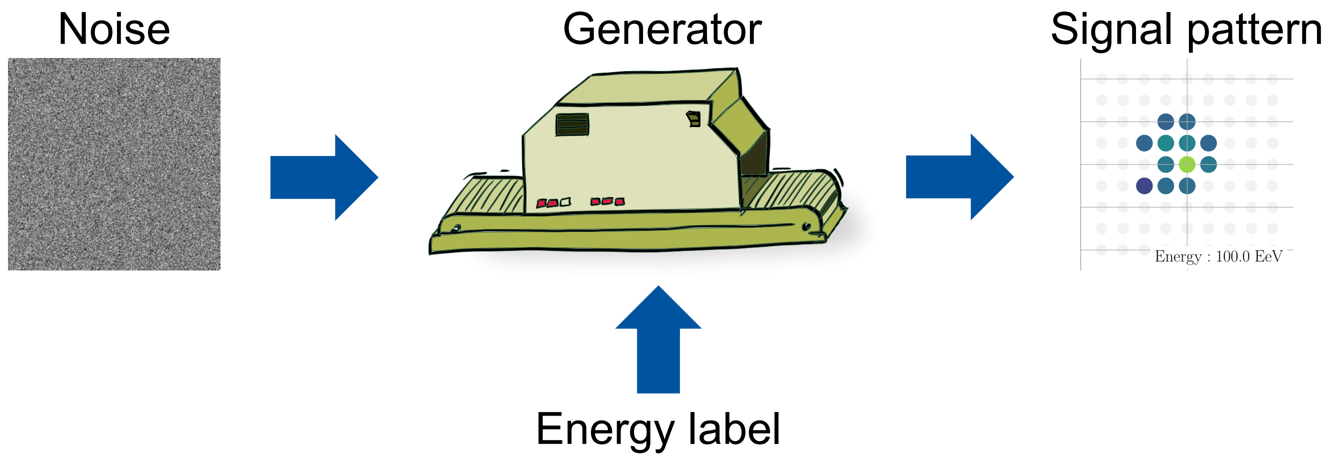
Trained on **refined simulation**



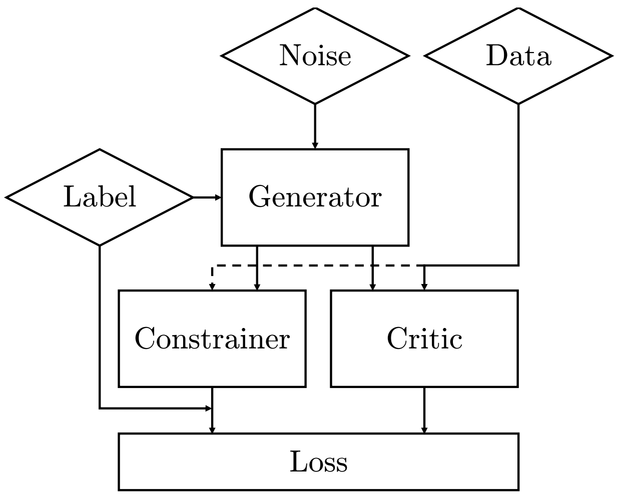
Training on refined simulations is able improve energy reconstruction

Adversarial training to generate detector signal patterns

Conditioned Generative Adversarial Network

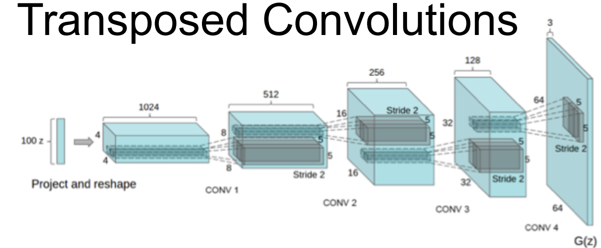


Conditioned Generative Adversarial Network



- Generate sample according to label $\tilde{x} = g_\theta(z, y_{label})$
- Train constrainer a_θ, supervised to minimize
$$[y_{data} - a_{\theta'}(x)]^2$$
- Loss: Extend Wasserstein distance by
$$\kappa[y_{label} - a_{\theta'}(\tilde{x})]^2$$

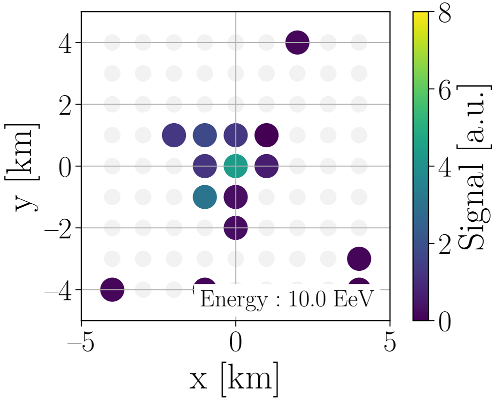
- **Generator:** DCGAN structure based on Transposed Convolutions

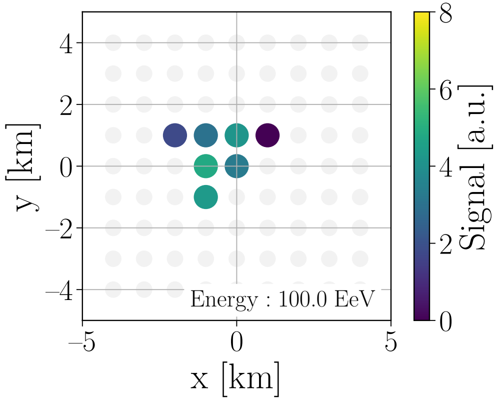


- **Constrainer:** DenseNet architecture

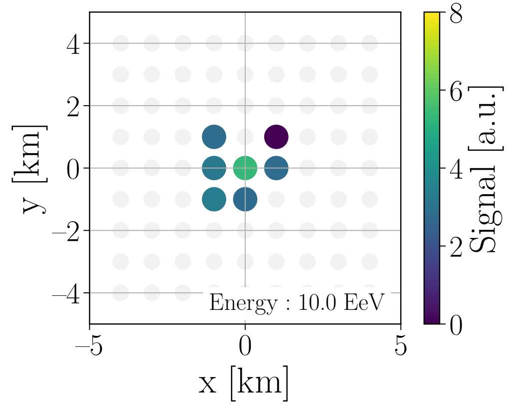
Energy constrained spatial signal pattern

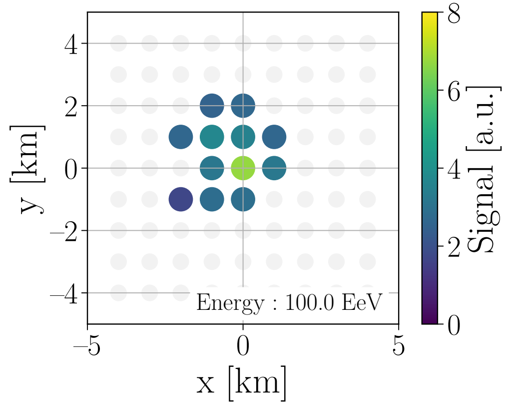
Epoch 1



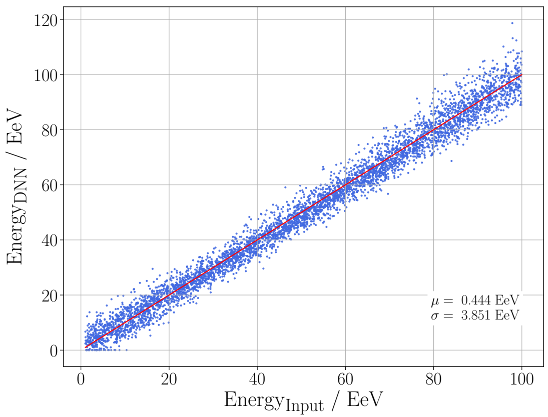


Epoch 95





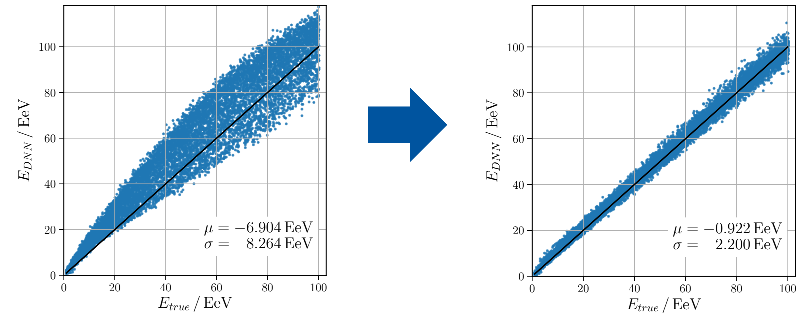
Energy reconstructed by the constrainer network



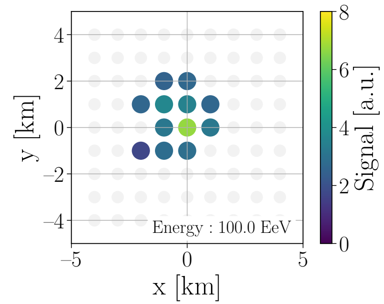
GAN is able to produce realistic shower patterns

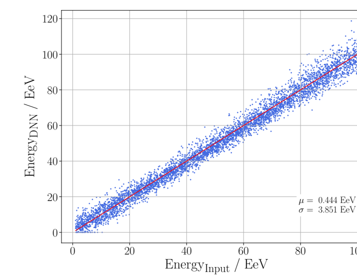
Conclusion

Supervised machine learning is improved after unsupervised refinement of simulation to match data



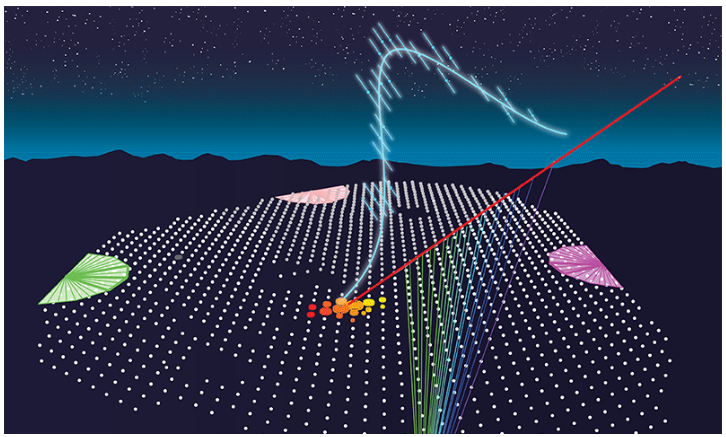
Physics conditioning of WGANs is a potential alternative to expensive simulations

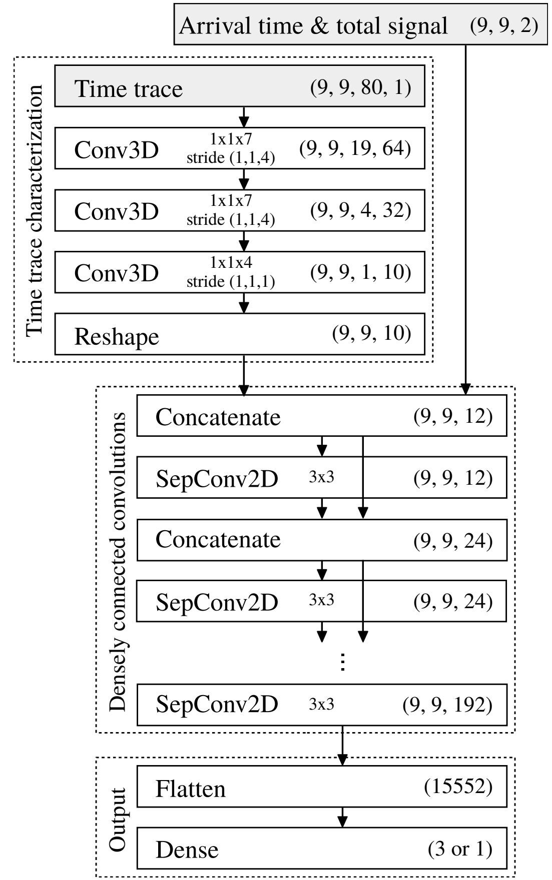




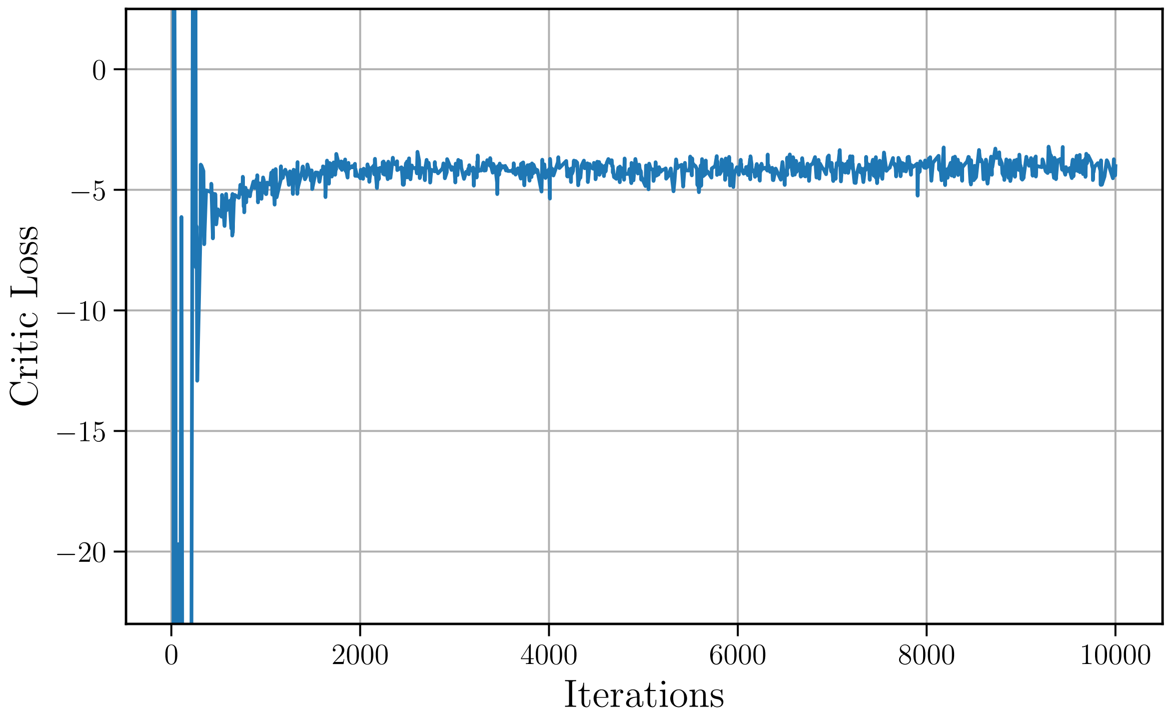
Backup

Network used for energy reconstruction of signal traces – arXiv:1708.00647





Loss of the critic network in the WGAN to refine signal traces



Refiner network as used in the WGAN to refine signal traces

Merge Operation	Operation	Kernel	Feature Maps	Padding	Activation
		$9 \times 9 \times 80 \times 1$ Input			
Addition	Convolution	$1 \times 1 \times 7$	64	same	ReLU
	Convolution	$1 \times 1 \times 7$	64	same	ReLU
Addition	Convolution	$1 \times 1 \times 7$	64	same	ReLU
	Convolution	$1 \times 1 \times 7$	64	same	ReLU
Addition	Convolution	$1 \times 1 \times 7$	64	same	ReLU
	Convolution	$1 \times 1 \times 7$	64	same	ReLU
Addition	Convolution	$1 \times 1 \times 7$	64	same	ReLU
	Convolution	$1 \times 1 \times 7$	64	same	ReLU
	Convolution	$1 \times 1 \times 1$	1	same	ReLU
		$9 \times 9 \times 80 \times 1$ Output			

Backup

Generator network as used in the WGAN to generate signal patterns

Operation	Kernel	Feature Maps	Padding	BN	Activation
Generator $80 + 1$ Input					
Linear	N/A	80		$\times$	ReLU
Transposed Convolution	3×3	64	valid	$\surd$	ReLU
Transposed Convolution	3×3	128	valid	$\surd$	ReLU
Transposed Convolution	3×3	128	valid	$\surd$	ReLU
Transposed Convolution	3×3	256	valid	$\surd$	ReLU
Convolution	3×3	1	same	$\times$	ReLU
Generator $9 \times 9 \times 1$ Output					

Critic network as used in the WGAN to generate signal patterns

Operation	Kernel	Feature Maps	Padding	BN	Activation
Critic $9 \times 9 \times 1$ Input					
Convolution	3×3	64	same	×	LeakyReLU
Convolution	3×3	128	same	×	LeakyReLU
Convolution	3×3	128	same	×	LeakyReLU
Convolution	3×3	256	same	×	LeakyReLU
GlobalMaxPooling				×	
Dropout					
Linear	N/A	100		×	LeakyReLU
Dropout					
Linear	N/A	1		×	
Critic 1 Output					