Karlsruhe Institute of Technology

Simulation of Extensive Air Showers with Deep Neural Networks

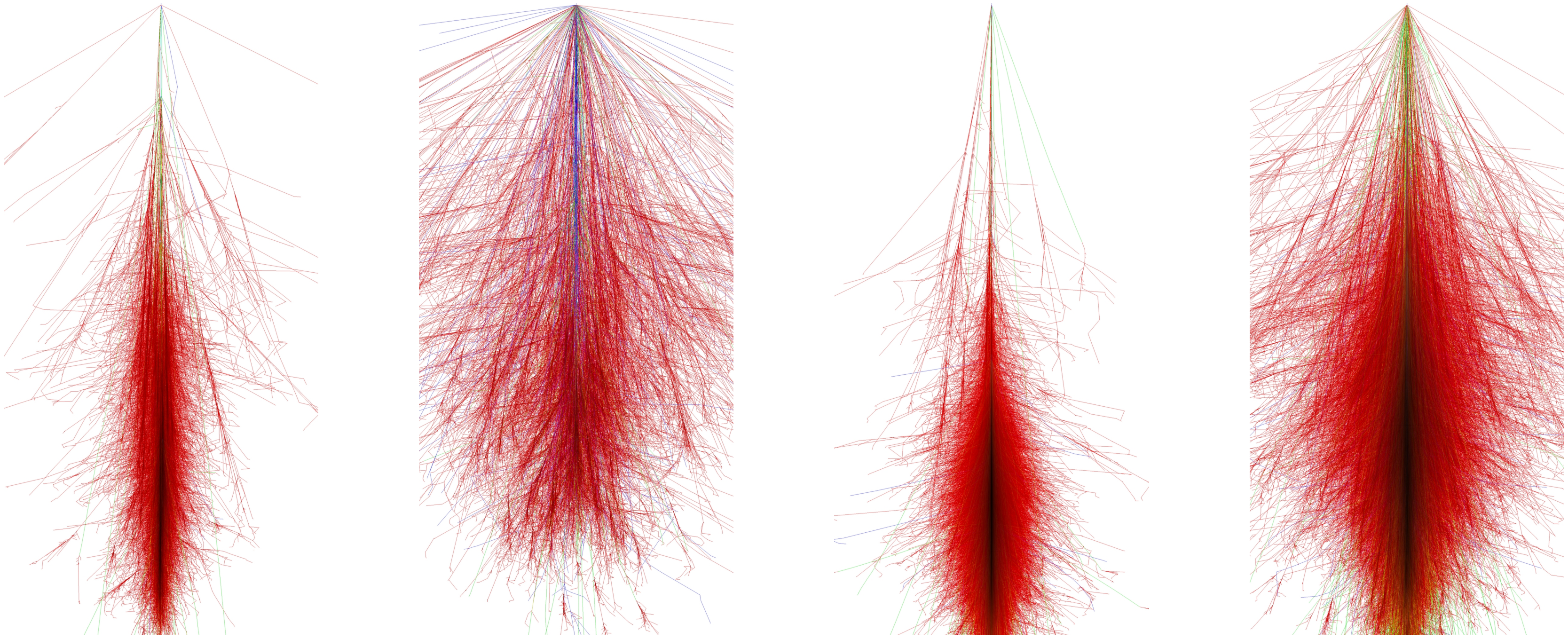
Marcel Köpke, Markus Roth

Big Data Science in Astroparticle Research – Workshop

INSTITUTE FOR NUCLEAR PHYSICS (IKP), FACULTY OF PHYSICS
KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT)

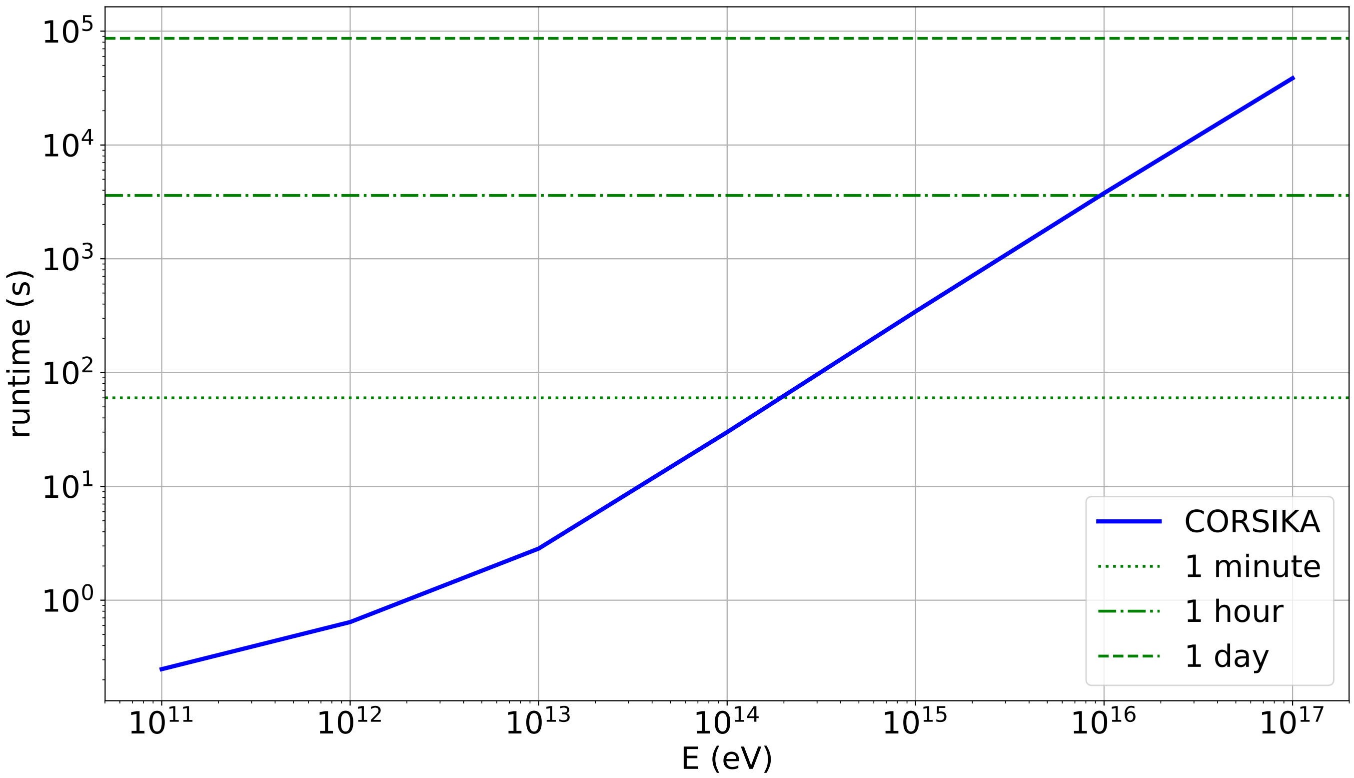
CORSIKA 7 [1]

- Extensive air shower Monte Carlo simulation framework
- Different types of interaction models (EPOS-LHC, QGSJET, SIBYLL, ...)



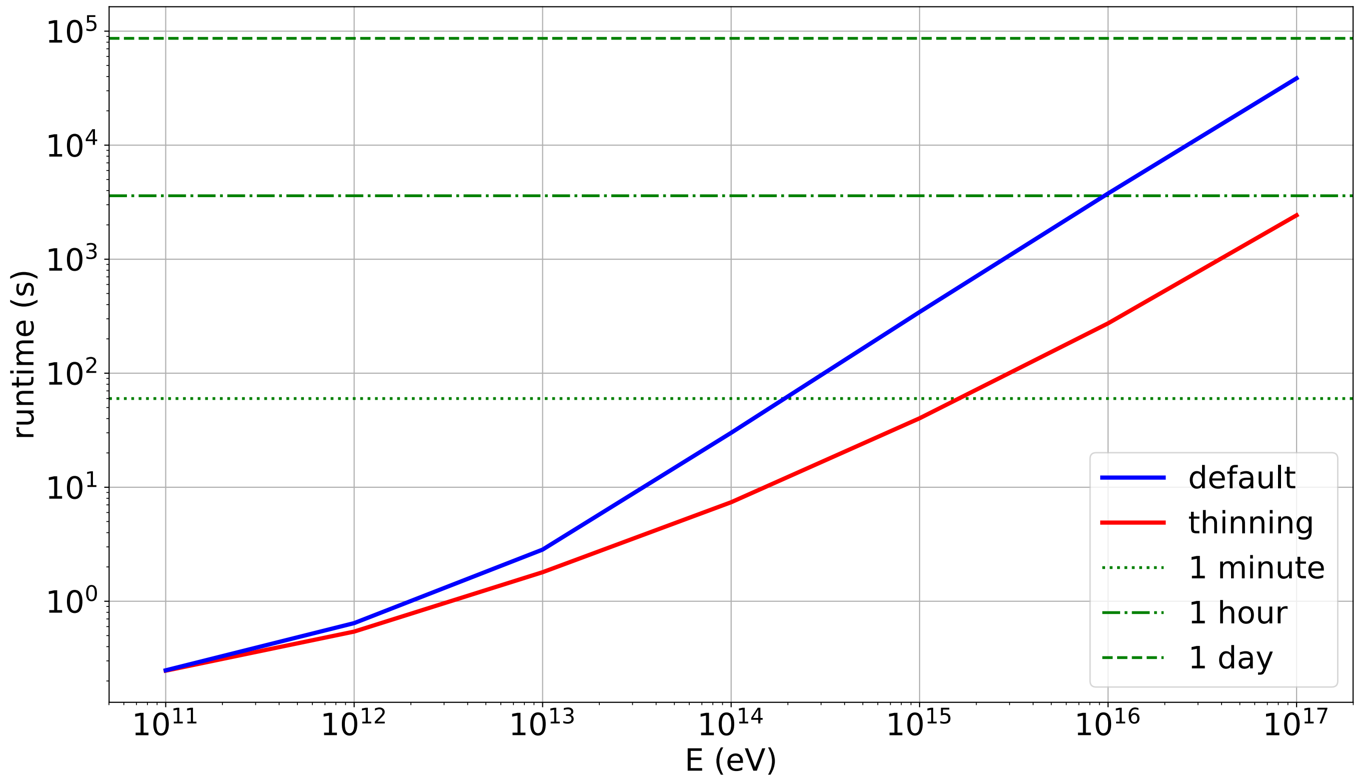
1 TeV Proton | 1 TeV Iron | 10 TeV Proton | 10 TeV Iron

Motivation

- 

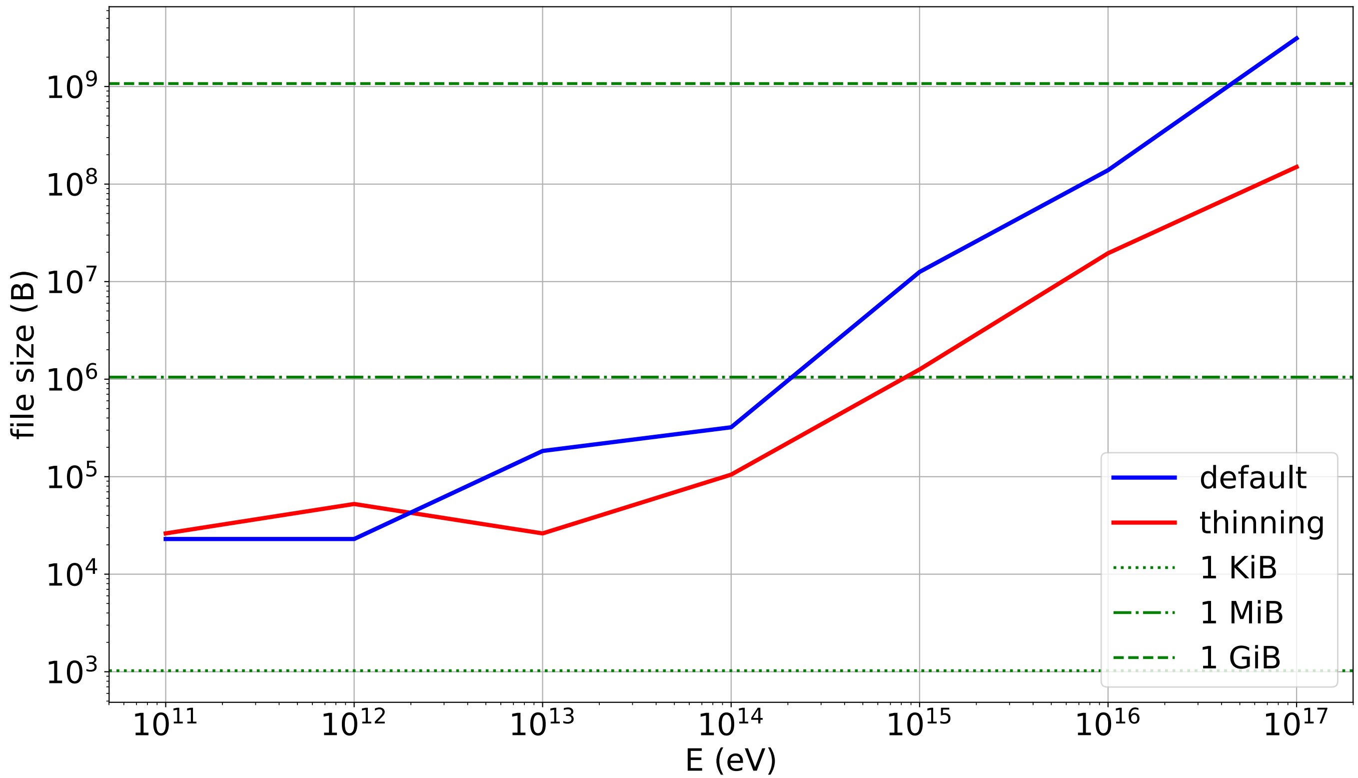
- The time complexity of CORSIKA 7 simulations rises approximately linearly with the primary particle energy

Thinning



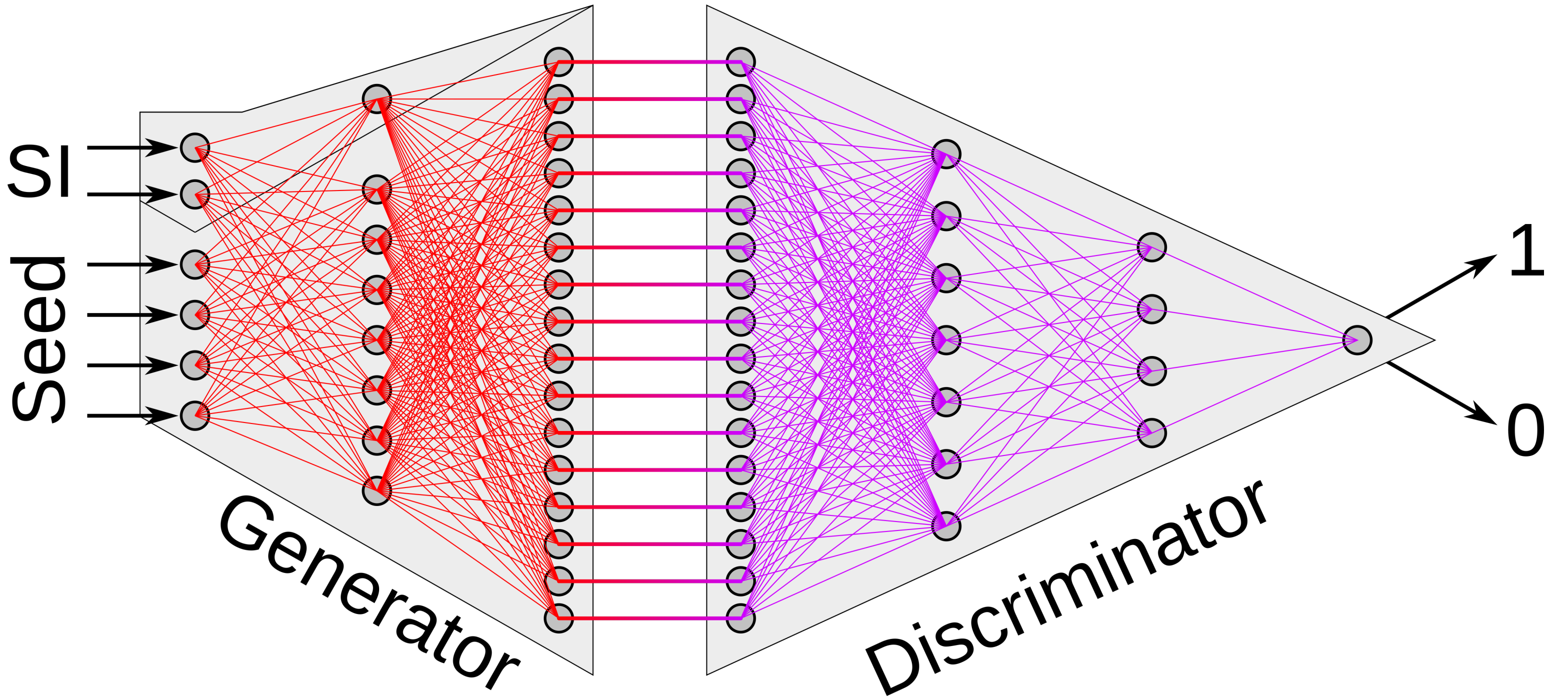
- Reduces (effective) particle content by particle-aggregation
- Preserves shower properties to leading order
- Reduces shower-to-shower fluctuations

Shower Library



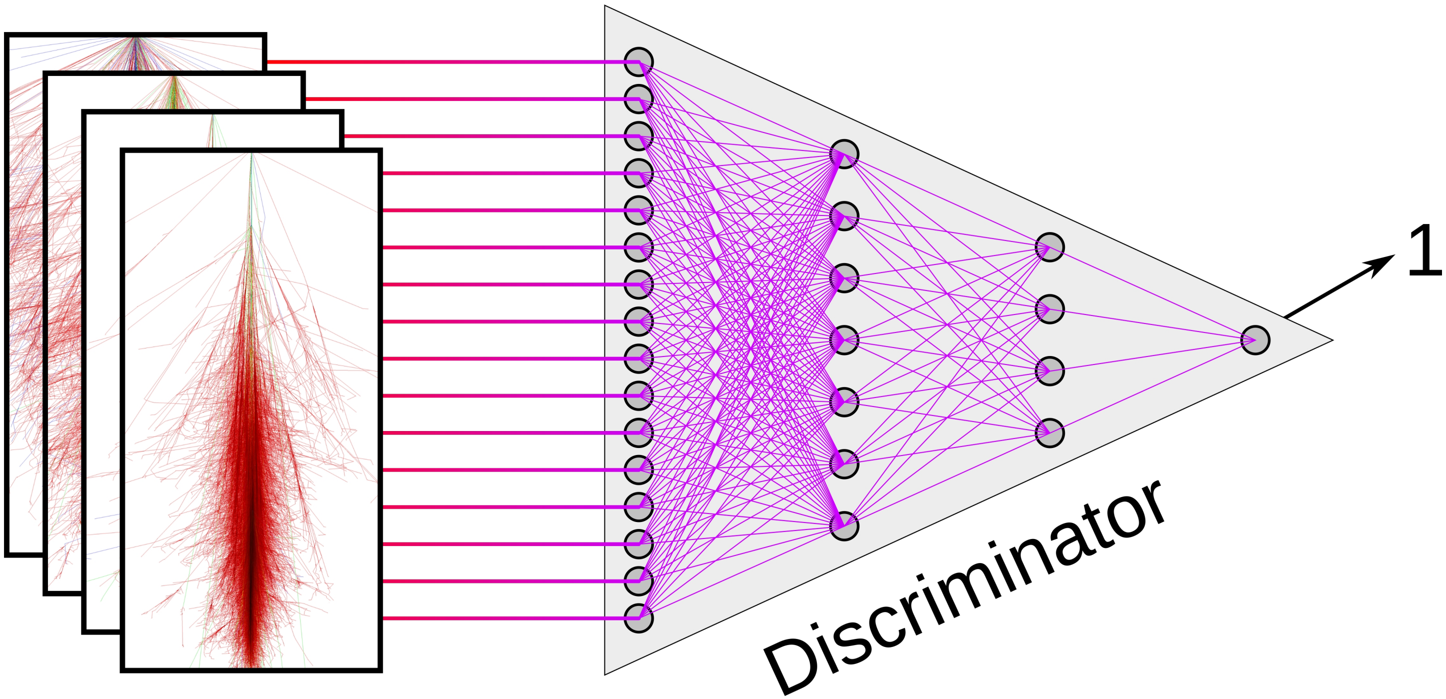
- Shower library required for analyses and model training
- Trained model = effective compression of shower library

Generative Adversarial Neural Network (GAN)



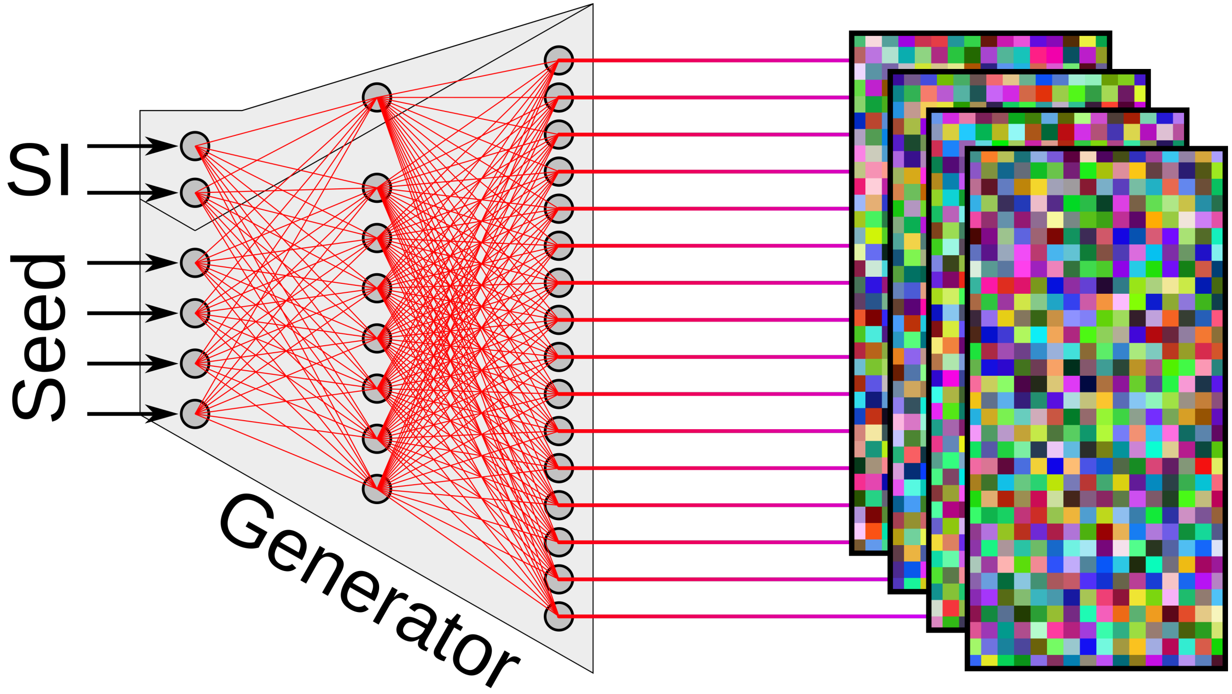
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Discriminator (Part 1)



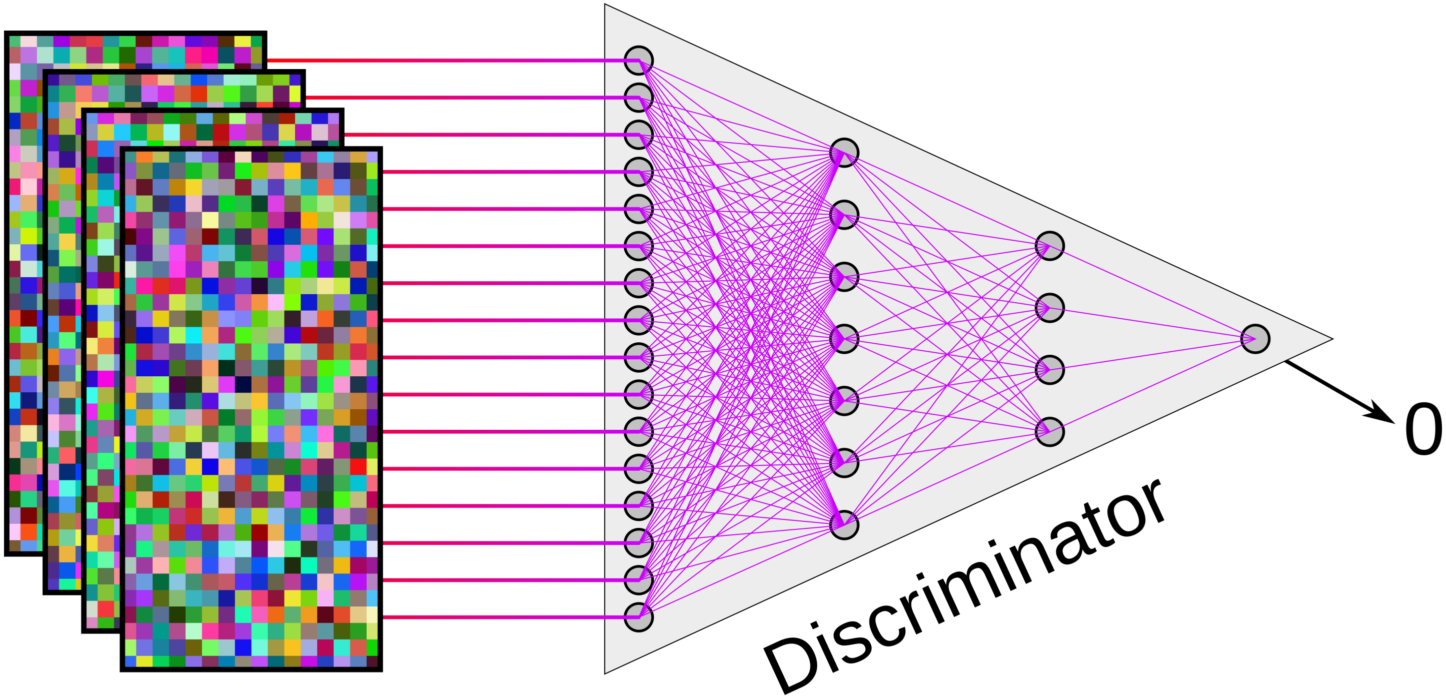
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Sampling



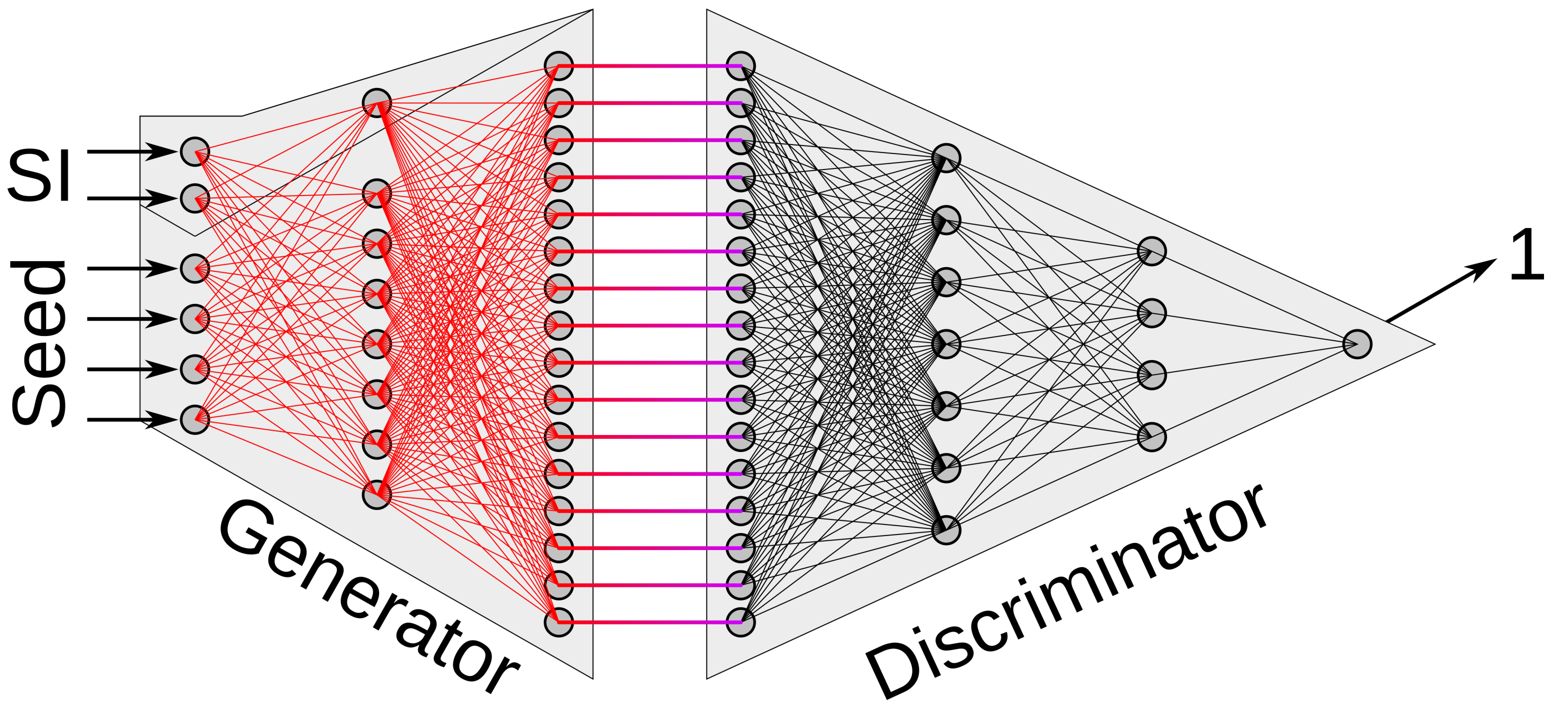
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Discriminator (Part 2)



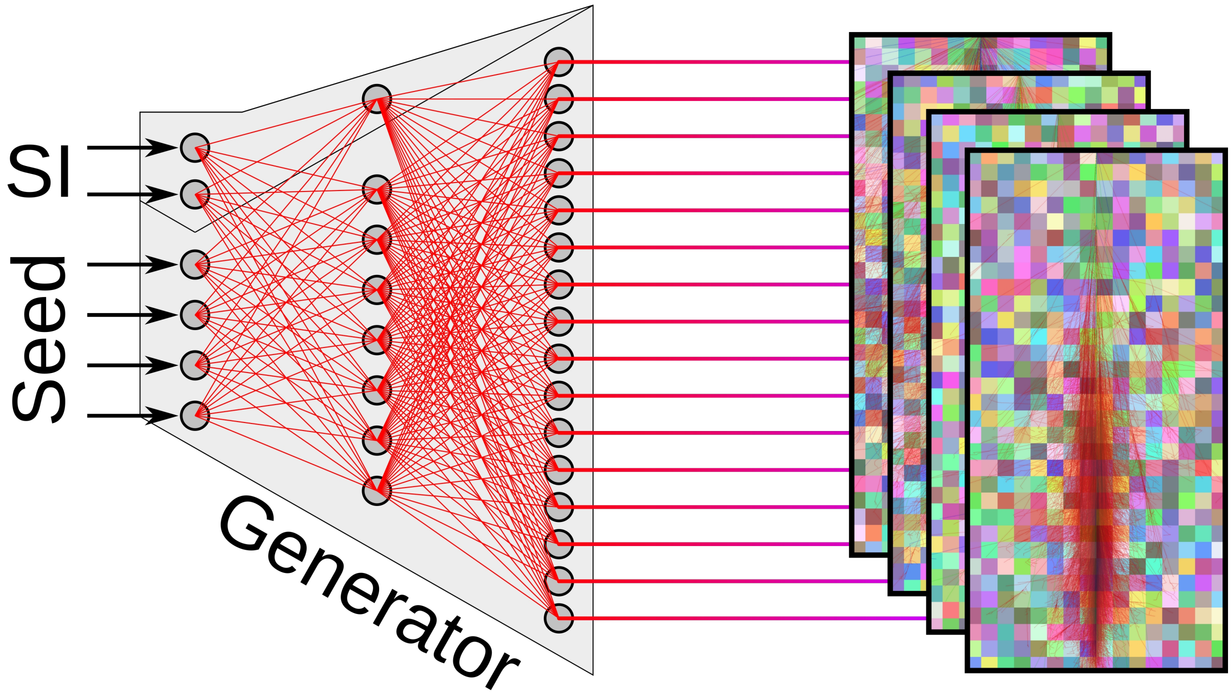
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Generator



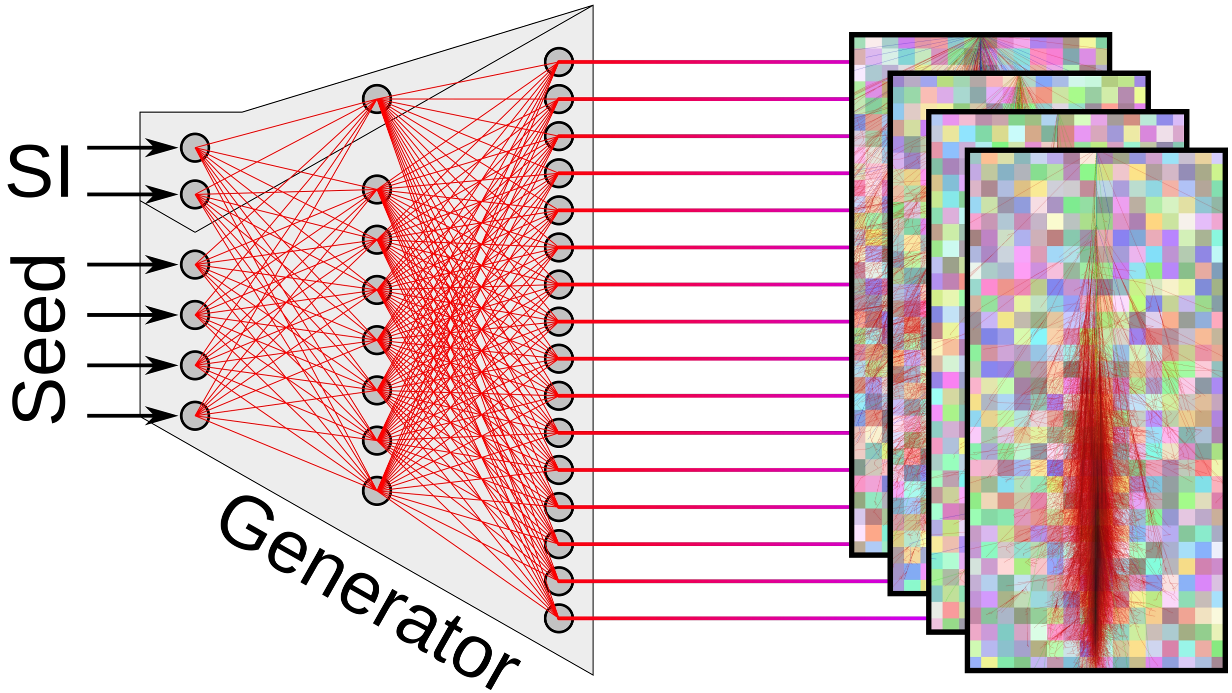
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Result



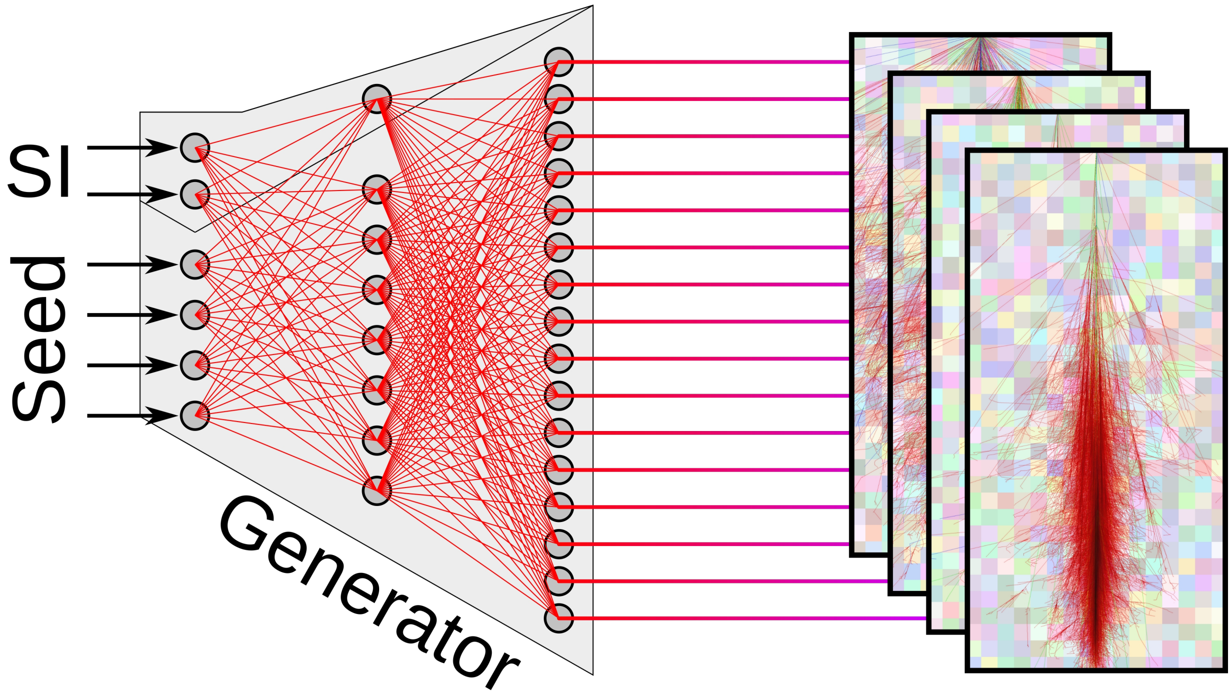
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Result



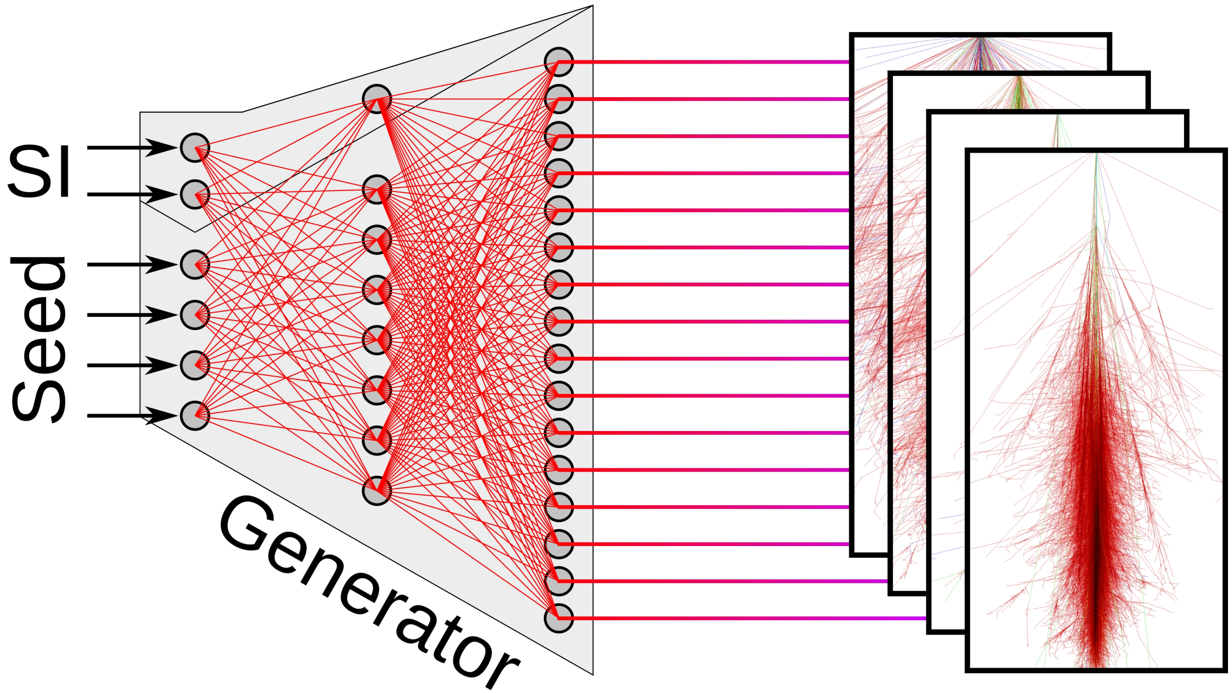
- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Result



- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Training: Result

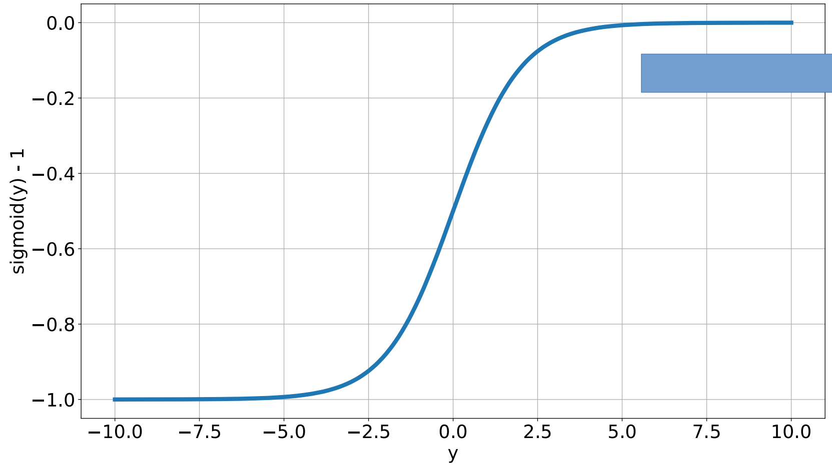


- Train discriminator on real (1) and generated (0) data
- Train generator to outsmart the discriminator

Cross Entropy

- $$\mathrm{CE} = \sum_i -z_i \cdot \log(p_i) + (z_i - 1) \cdot \log(1 - p_i)$$
  with z (true) label and p probability (NN output)

- z = 1: $-\log(\mathrm{sigmoid}(y))$
  $$\Rightarrow \frac{d}{dy}(-\log(\mathrm{sigmoid}(y))) = \mathrm{sigmoid}(y) - 1$$



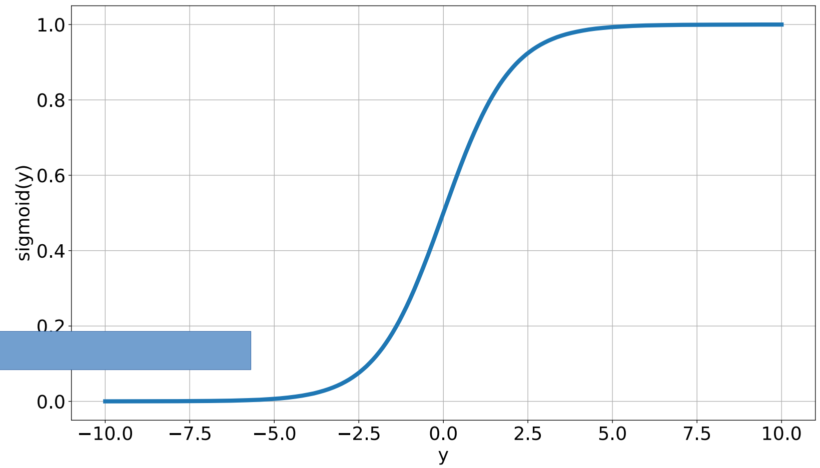
vanishing gradients

Cross Entropy

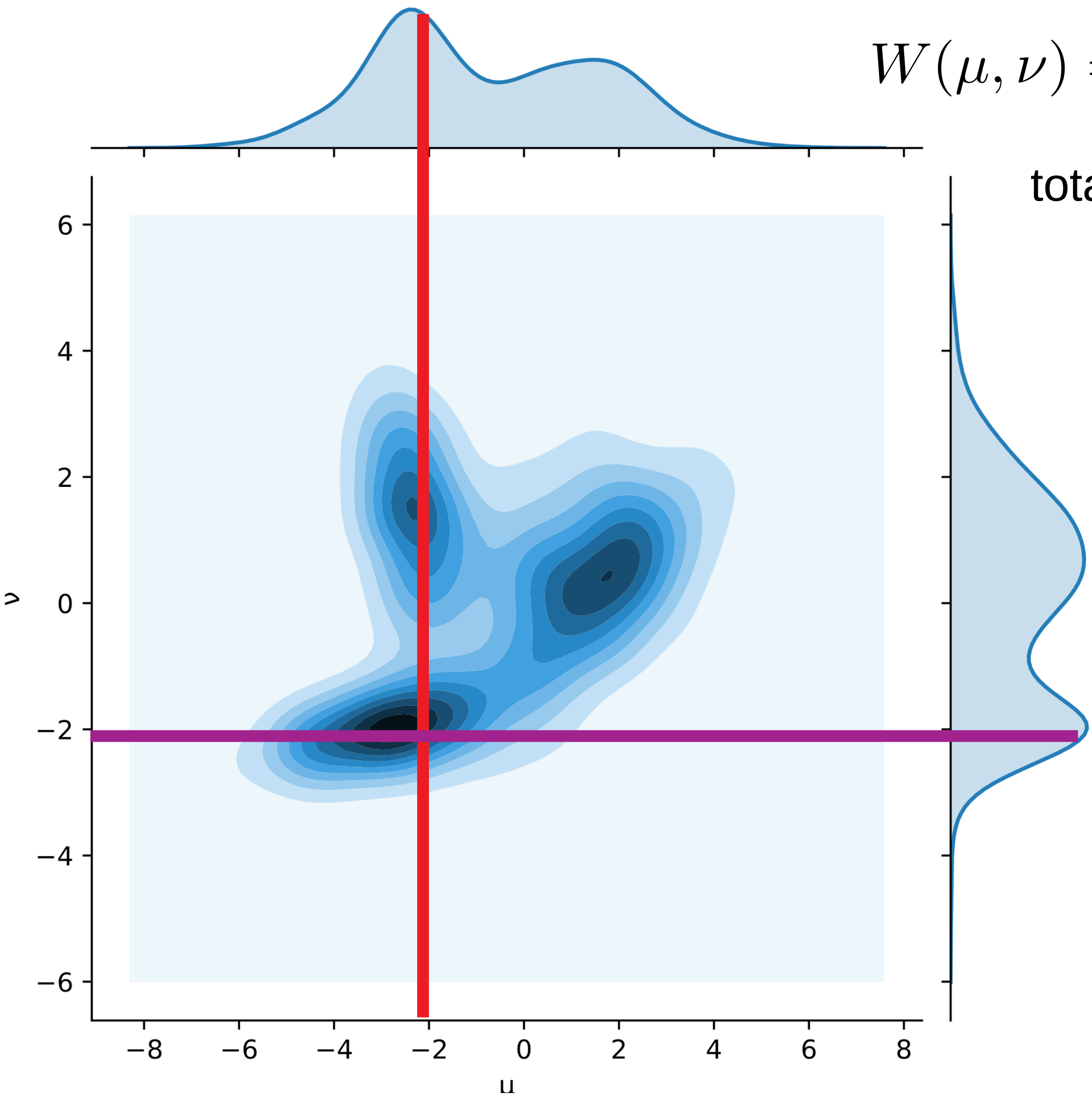
- $$\mathrm{CE} = \sum_i -z_i \cdot \log(p_i) + (z_i - 1) \cdot \log(1 - p_i)$$
  with z (true) label and p probability (NN output)

- z = 0: $-\log(1 - \mathrm{sigmoid}(y))$
  $$\Rightarrow \frac{d}{dy}(-\log(1 - \mathrm{sigmoid}(y))) = \mathrm{sigmoid}(y)$$

vanishing gradients



Wasserstein Distance [2]



$$W(\mu,\nu) = \inf_{\gamma} \int d(x,y)\cdot\gamma(x,y)\ dxdy$$

total (minimal) cost to move all mass

$$\int d(x,y)\cdot\gamma(x,y)dy = c(x)$$

cost to move mass above/below x

$$\int \gamma(x,y)dy = \mu(x)$$

red: total mass pilled up at x

Kantorovich-Rubinstein Duality

- $$W(\mu, \nu) = \sup_{f \in Lip_{\leq 1}} \mathbb{E}_{x\sim\mu}[f(x)] - \mathbb{E}_{y\sim\nu}[f(y)]$$
- f = Neural Network
- Lipschitz continous: $|f(x_1) - f(x_2)| \leq L \cdot \|x_1 - x_2\|$
- Gradient is bounded → Gradient penalty $|\|\nabla f\| - 1| \to 0$

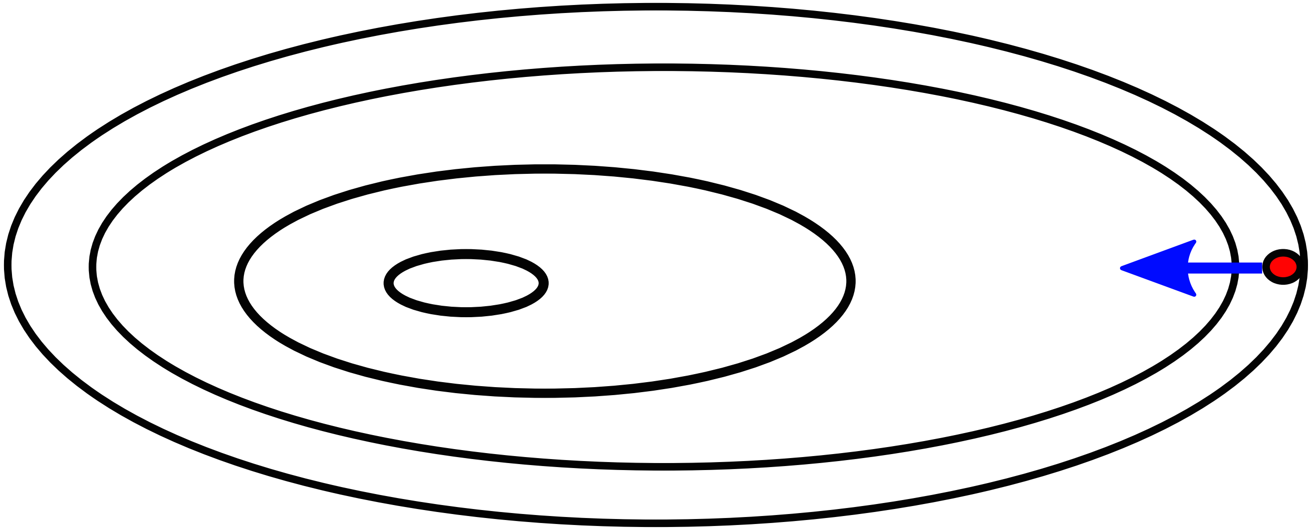
Gradient Penalty

- $$W(\mu,\nu) = \sup_{f \in Lip_{\leq 1}} \mathbb{E}_{x\sim\mu}[f(x)] - \mathbb{E}_{y\sim\nu}[f(y)]$$
- $\mathbb{E}_{x\sim\mu}[f(x)] \to \infty \qquad \mathbb{E}_{y\sim\nu}[f(y)] \to -\infty$
- $f \to a \cdot f$ and $a \to \infty$
- But $|a \cdot f(x) - a \cdot f(y)| \leq L \, ||x - y||$

  $\Rightarrow \; a \cdot \|\nabla f\| \leq L$

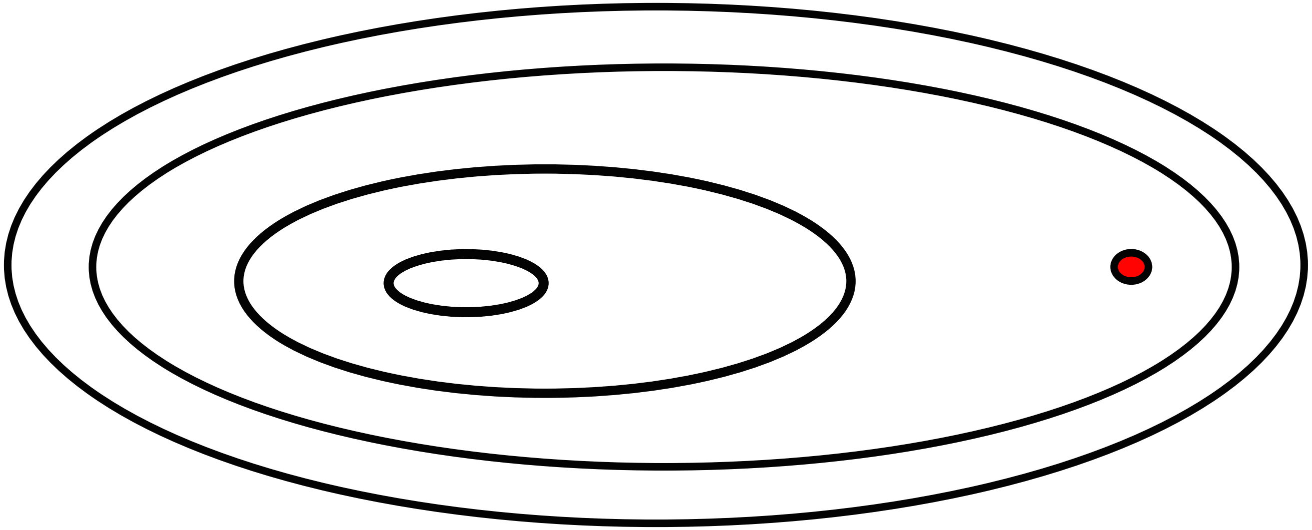
Learning Rate and Momentum

- Ordinary classification:
  - Gradient



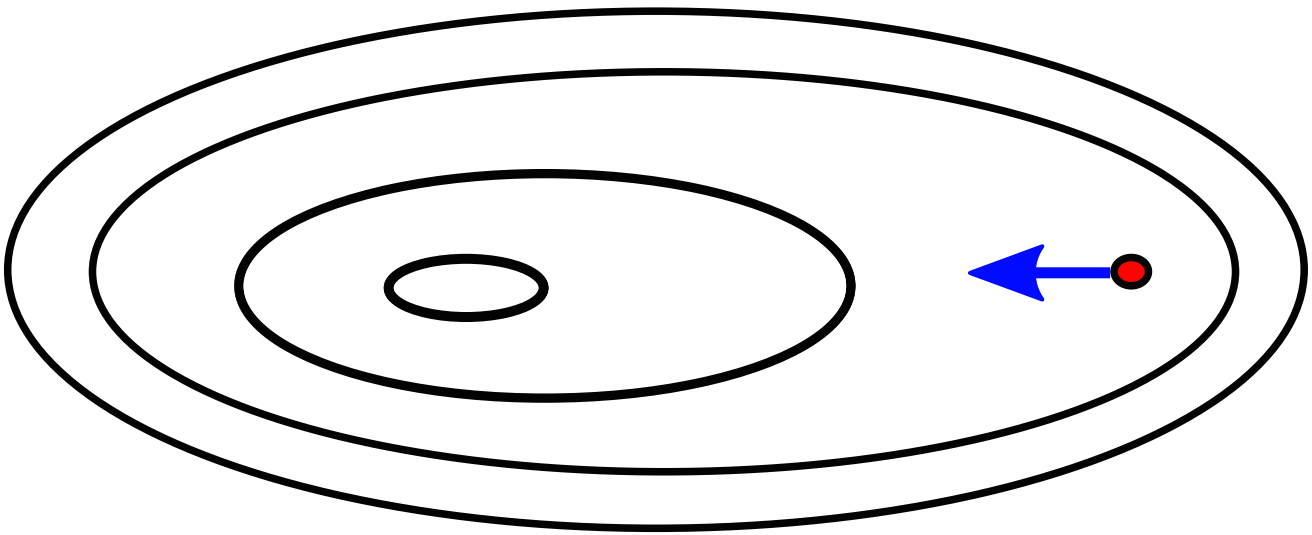
Learning Rate and Momentum

- Ordinary classification:
  - Step



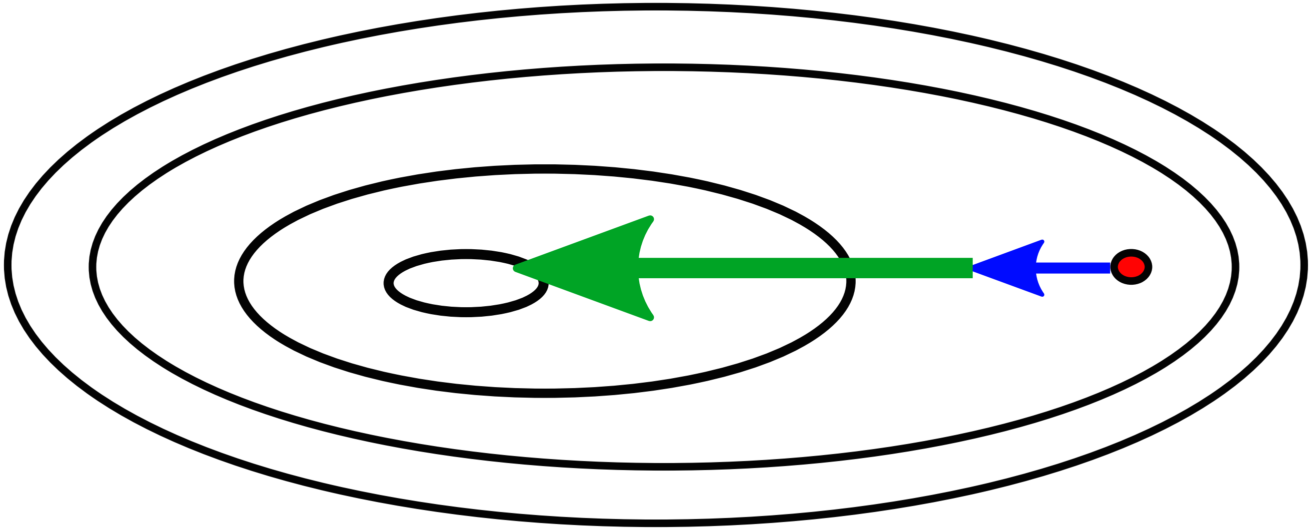
Learning Rate and Momentum

- Ordinary classification:
  - Gradient



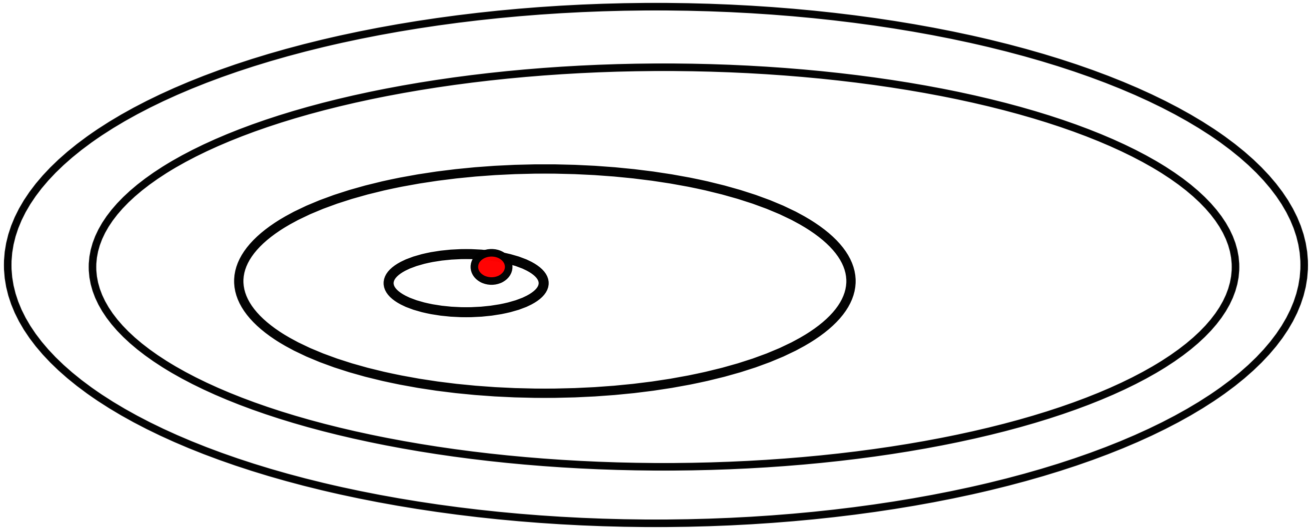
Learning Rate and Momentum

- Ordinary classification:
  - Momentum



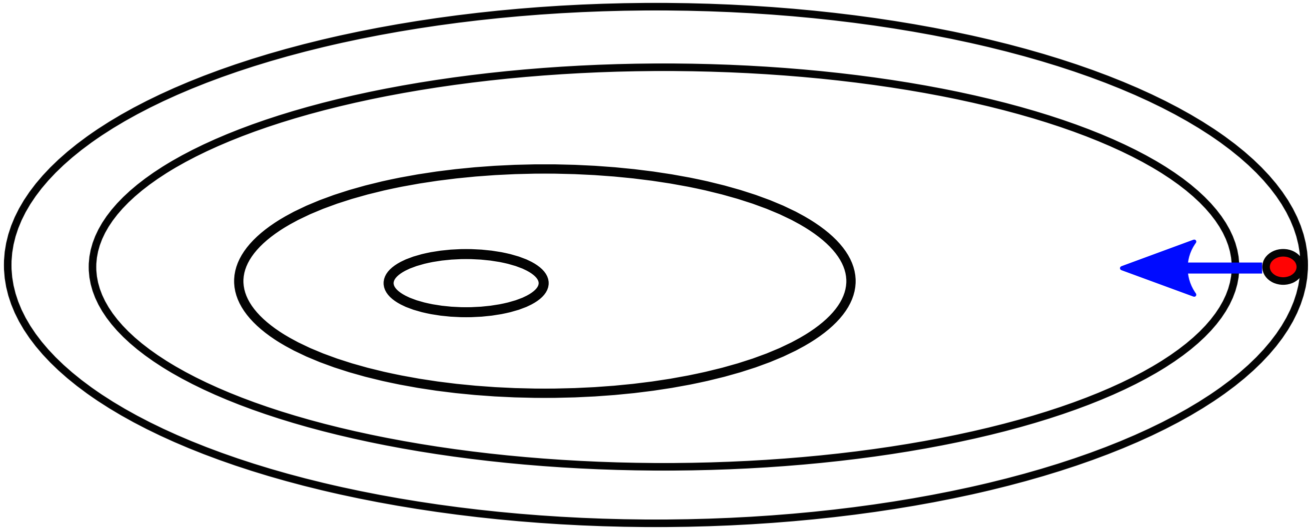
Learning Rate and Momentum

- Ordinary classification:
  - Step



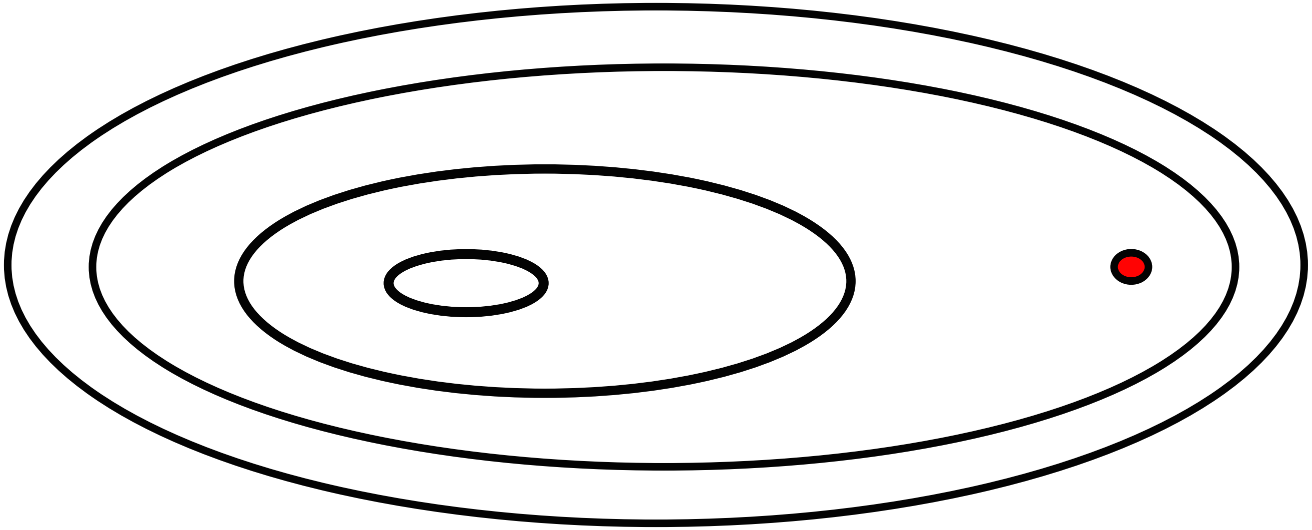
GAN: Learning Rate and Momentum

- Discriminator classification:
  - Gradient



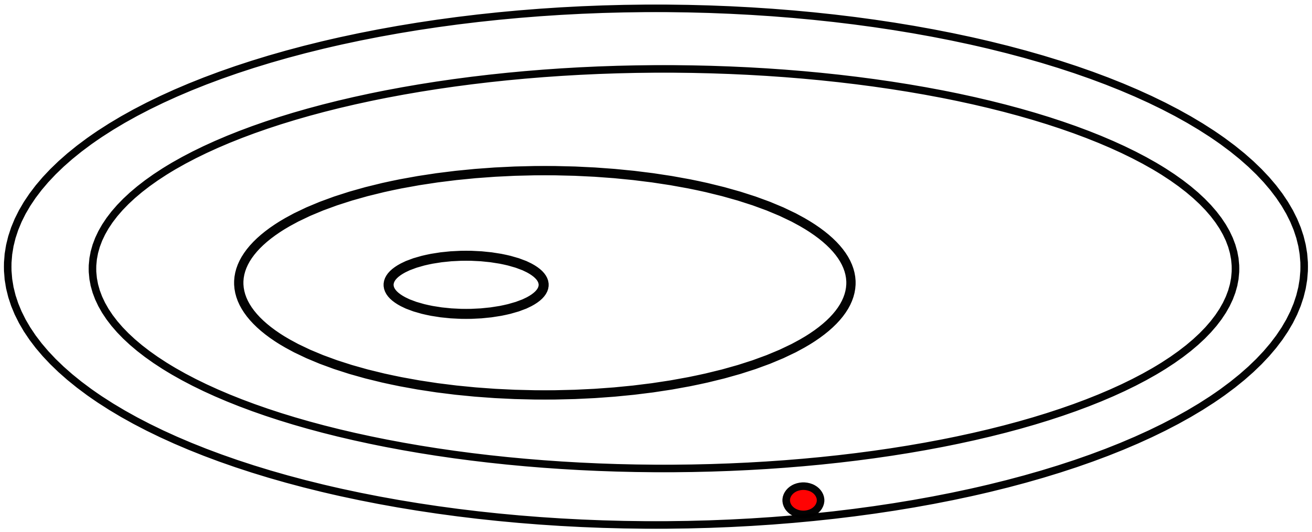
GAN: Learning Rate and Momentum

- Discriminator classification:
  - Step



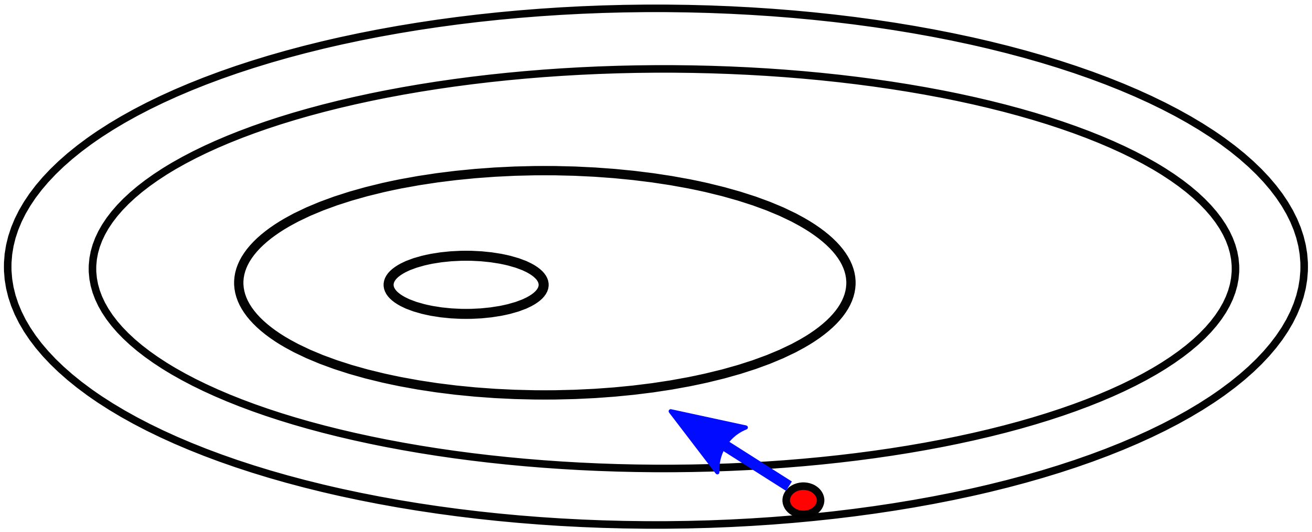
GAN: Learning Rate and Momentum

- Discriminator classification:
  - Generator training



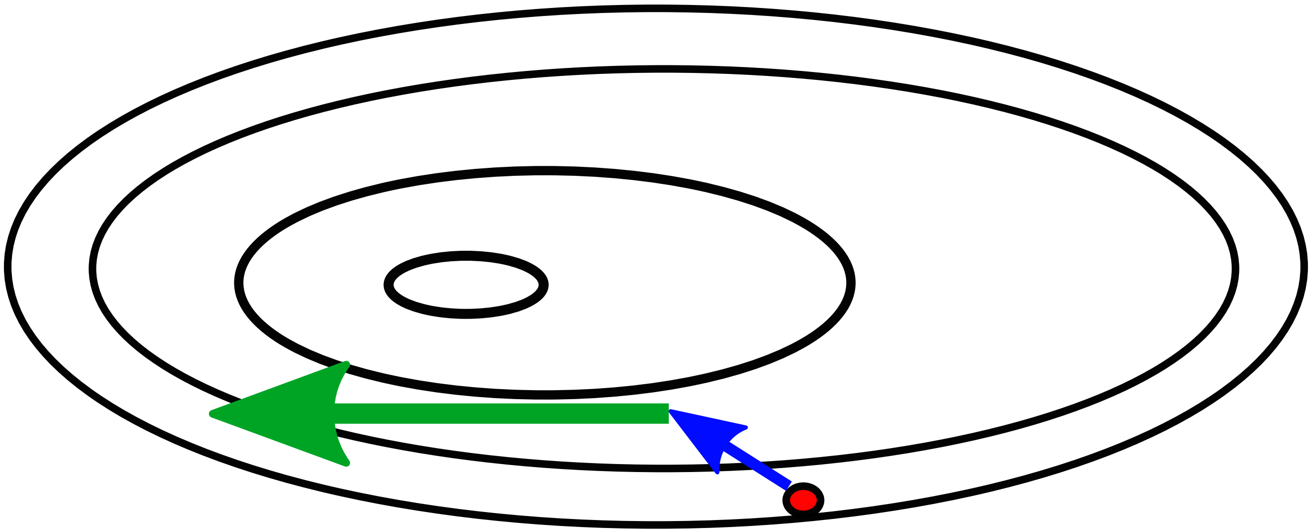
GAN: Learning Rate and Momentum

- Discriminator classification:
  - Gradient



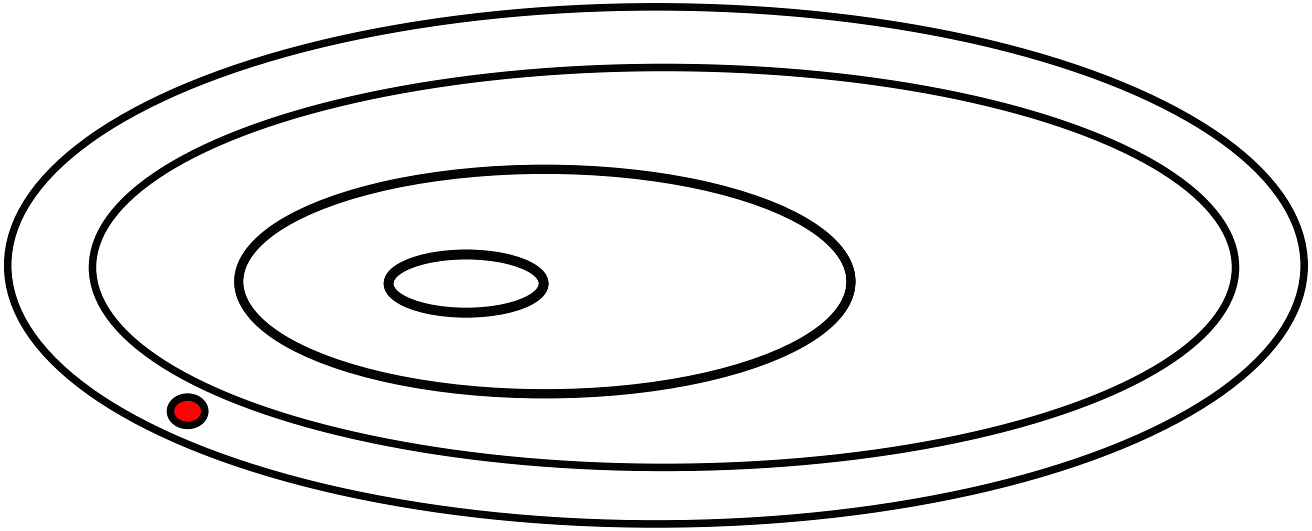
GAN: Learning Rate and Momentum

- Discriminator classification:
  - Momentum



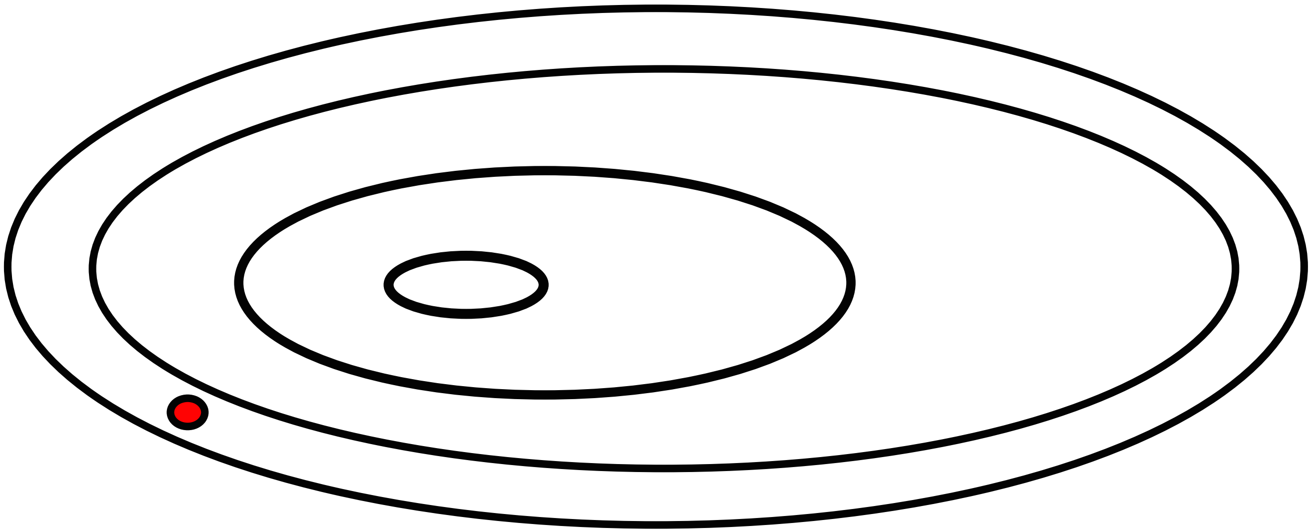
GAN: Learning Rate and Momentum

- Discriminator classification:
  - Step



GAN: Learning Rate and Momentum

- Discriminator classification:
  - Step



- Adam: $\alpha \leq 10^{-4} \quad \beta_1 = 0.5 \quad \beta_2 = 0.9$

Summary

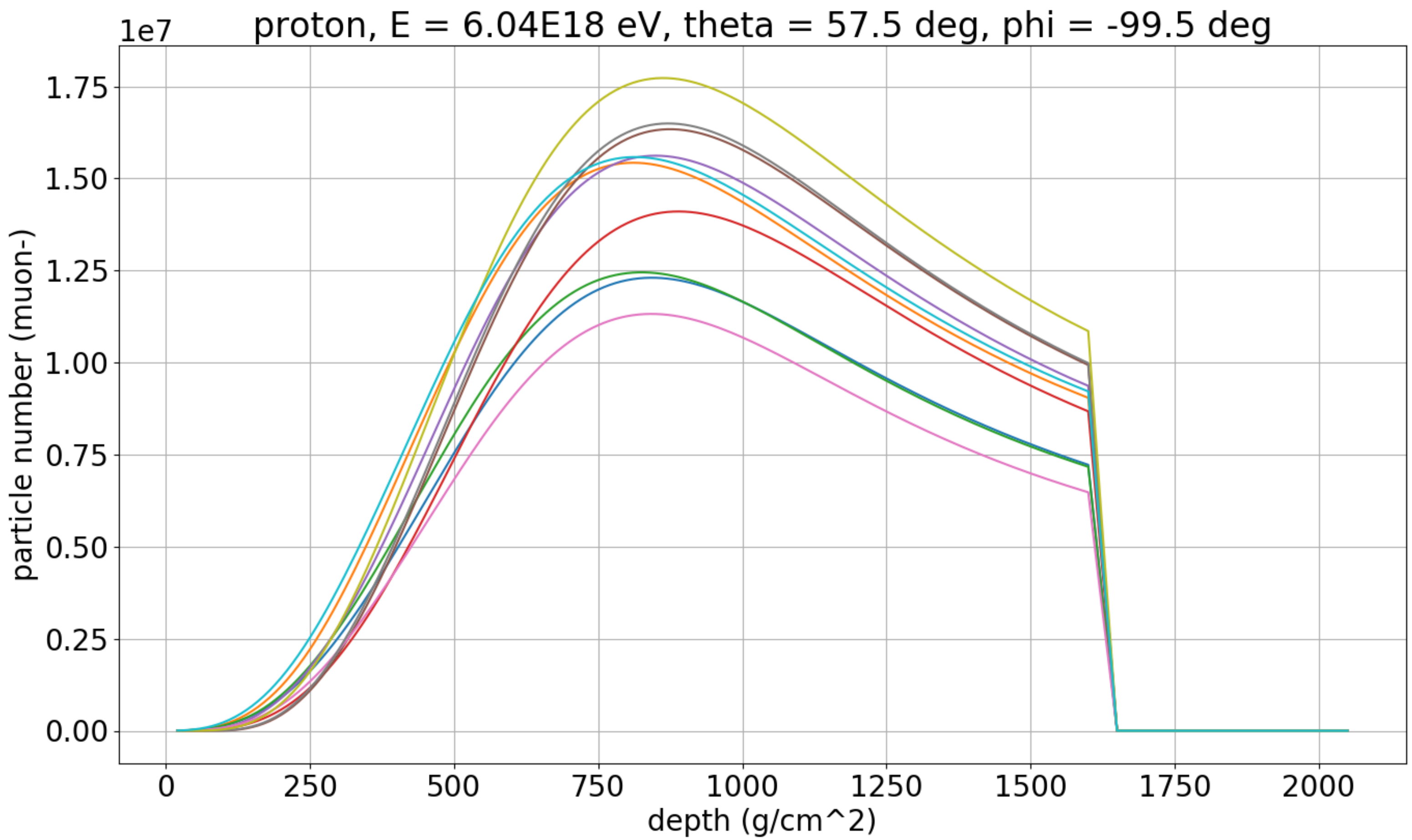
- The Wasserstein distance is a measure of the distance of (data) distributions.
- The implicit usage of Limpschitz continous functions acts as a regulator for the discriminator.
- The regulator removes a malicious scaling degree of freedom.
- The Wasserstein loss is less prone to vanishing gradients.
- Use smaller (than usual) learning rates and less momentum.

Backup

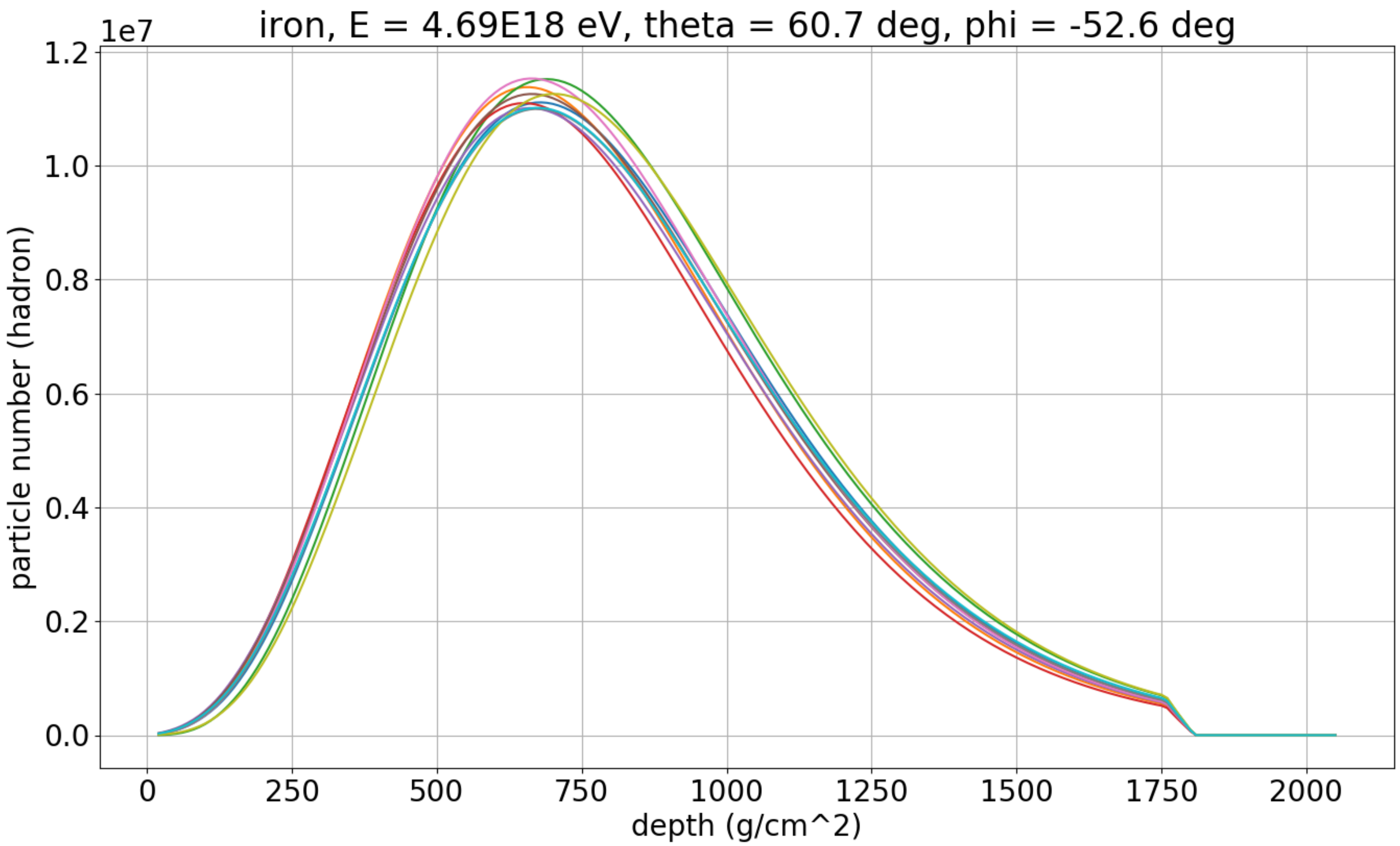
First Test (CONEX)

- CONEX: Hybrid Extenisve Air Shower Simulation
  - first: Monte Carlo until energy threshold (3D)
  - then: cascade equation solver (1D)
  - provides longitudinal profile only
  - runtime: seconds – minutes
- Configuration:
  - E = 1E17 ... 1E19 eV
  - Zenith = 0 ... 65 deg
  - Azimuth = -180 ... 180 deg
- Generated 200k + 300k datapoints

Shower-to-Shower Fluctuations



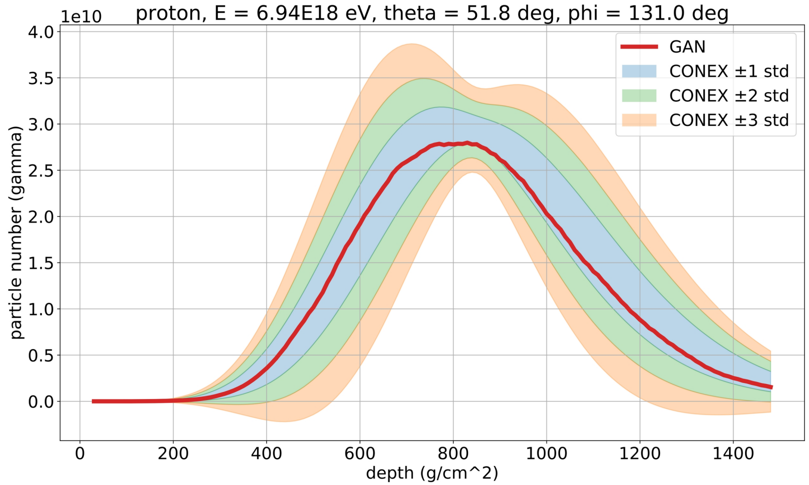
Shower-to-Shower Fluctuations

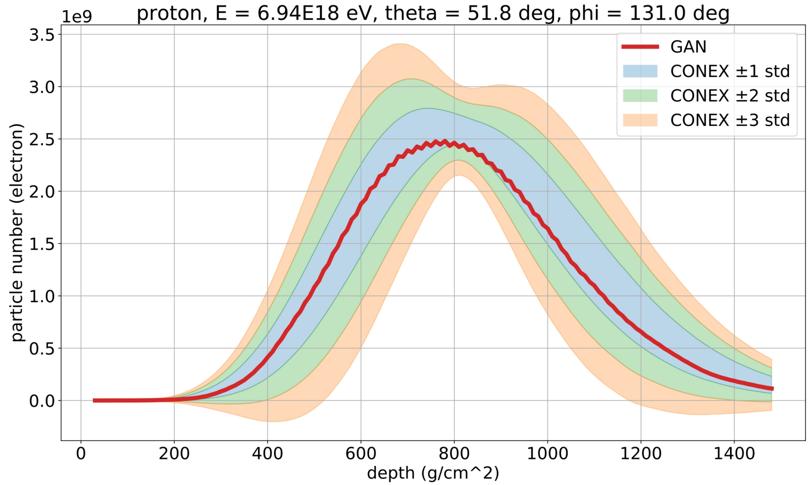


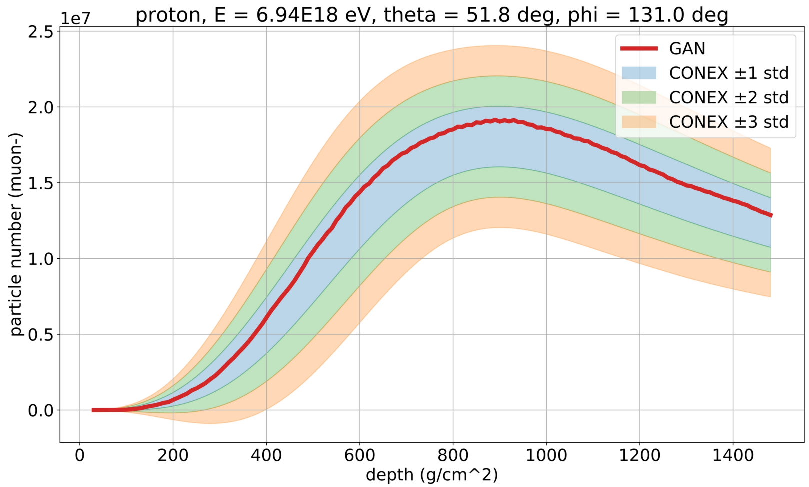
(conditional) WGAN

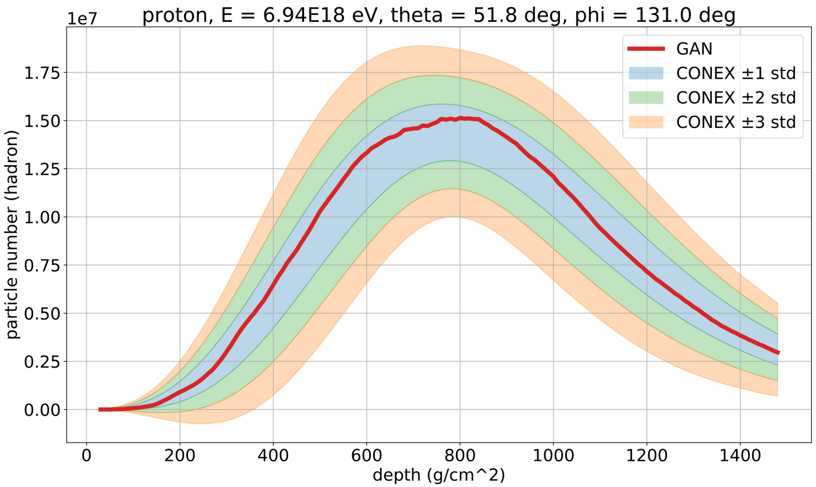
- Generator:
  - 5x Dense
  - 5x TransposeConvolution + Convolution
  - Activation: tanh
- Discriminator:
  - 3x Dense
  - 7x Convolution
  - 2x Dense
  - Activation: tanh
- Trainable parameters: 79.072.457

CONEX vs. GAN

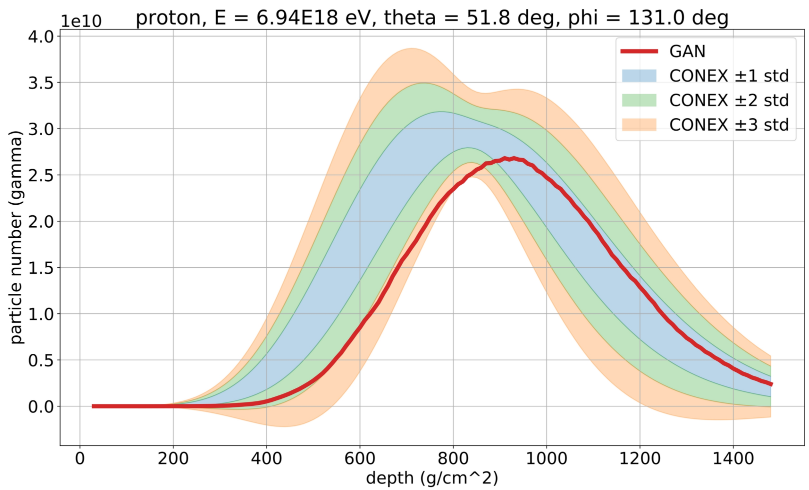

proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e10
particle number (gamma)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std


proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e9
particle number (electron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

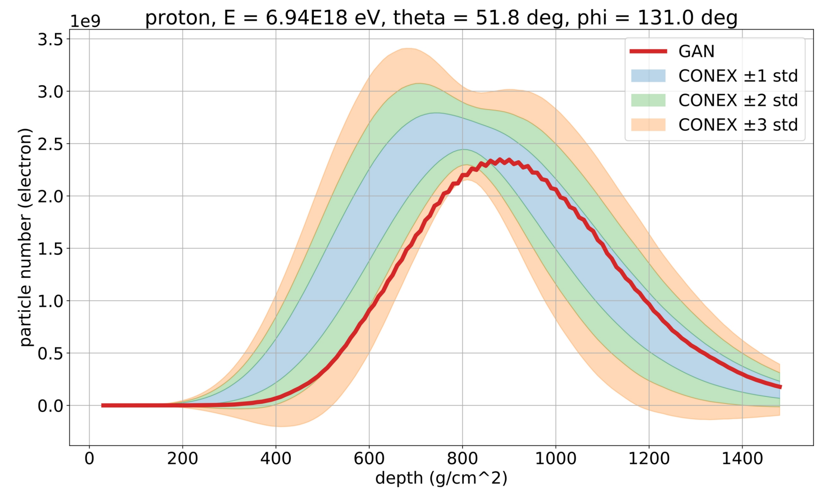

proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e7
particle number (muon-)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std


proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e7
particle number (hadron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

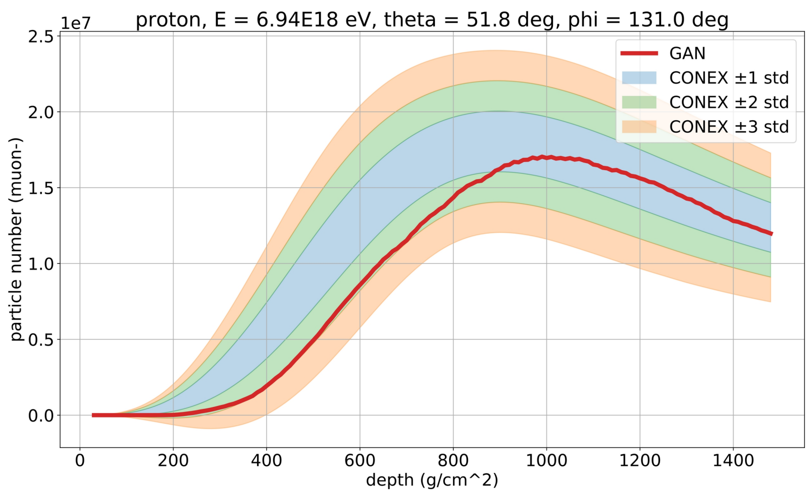
CONEX vs. GAN



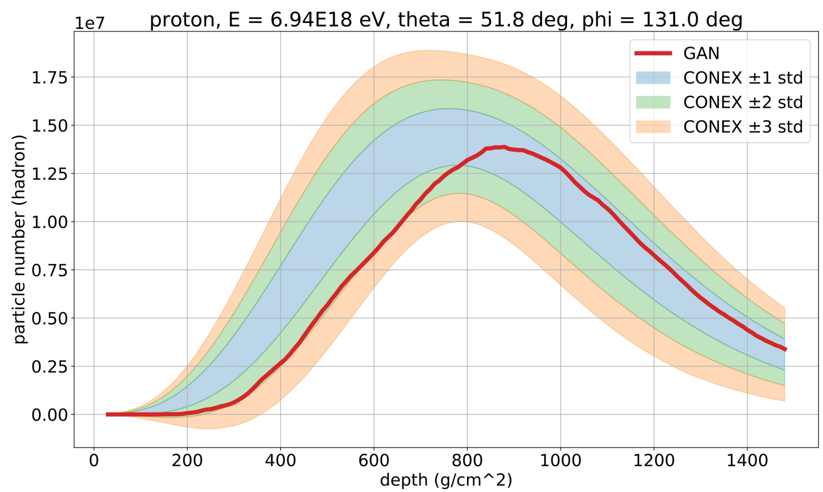
proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e10
particle number (gamma)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std



proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e9
particle number (electron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

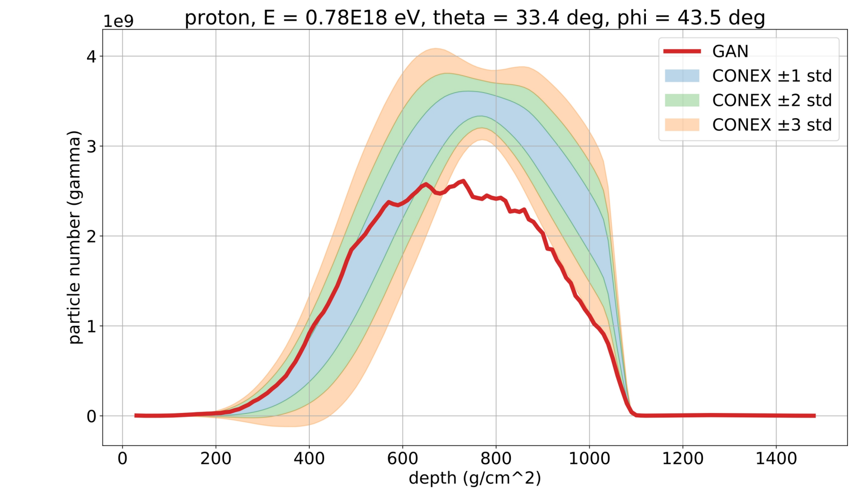


proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e7
particle number (muon-)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

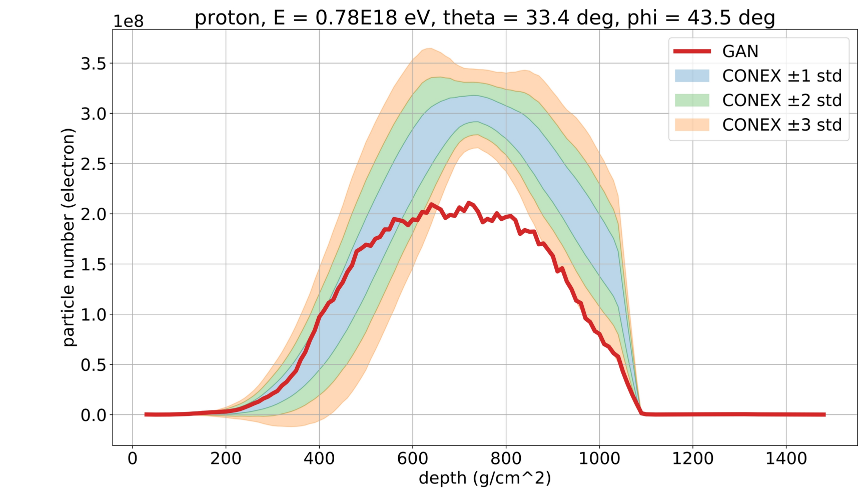


proton, E = 6.94E18 eV, theta = 51.8 deg, phi = 131.0 deg
1e7
particle number (hadron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

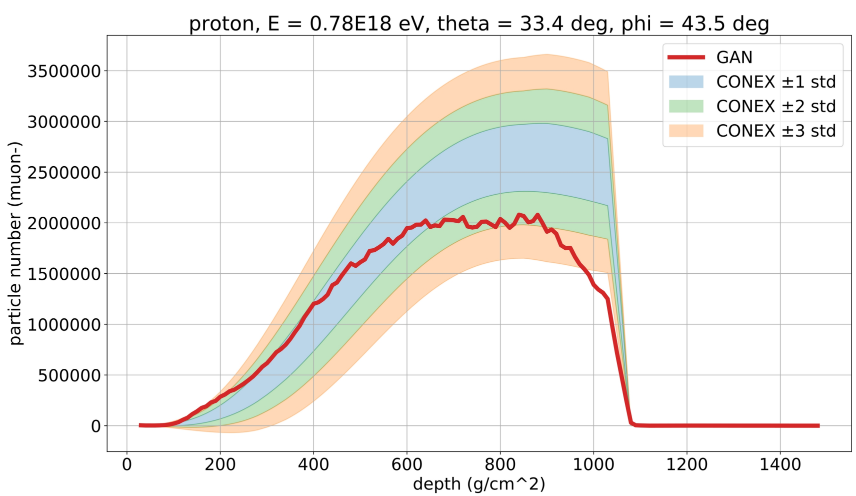
CONEX vs. GAN



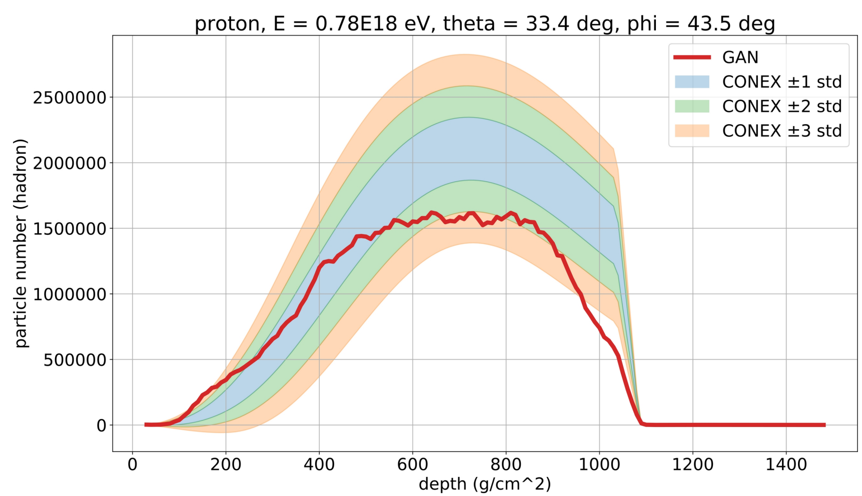
proton, E = 0.78E18 eV, theta = 33.4 deg, phi = 43.5 deg
1e9
particle number (gamma)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std
0 1 2 3 4
0 200 400 600 800 1000 1200 1400



proton, E = 0.78E18 eV, theta = 33.4 deg, phi = 43.5 deg
1e8
particle number (electron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
0 200 400 600 800 1000 1200 1400

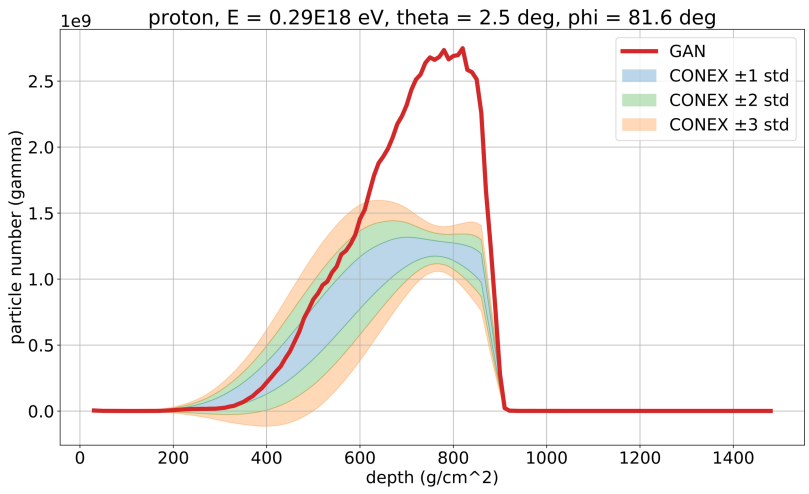


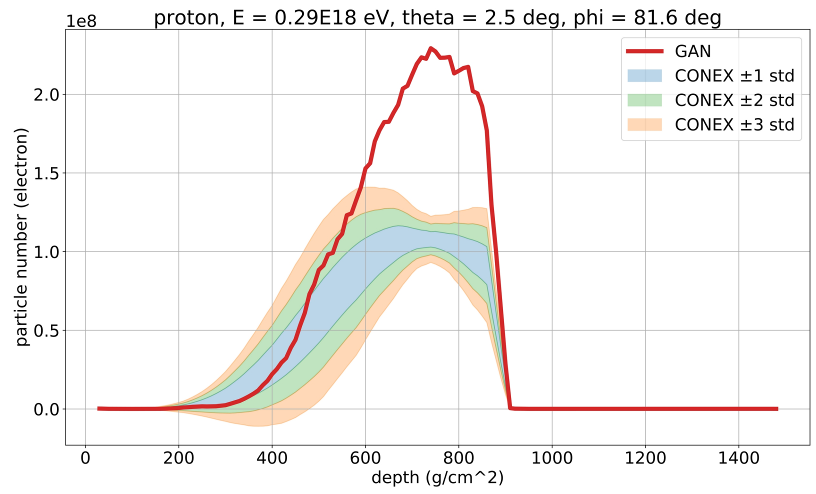
proton, E = 0.78E18 eV, theta = 33.4 deg, phi = 43.5 deg
particle number (muon-)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std
0 500000 1000000 1500000 2000000 2500000 3000000 3500000
0 200 400 600 800 1000 1200 1400

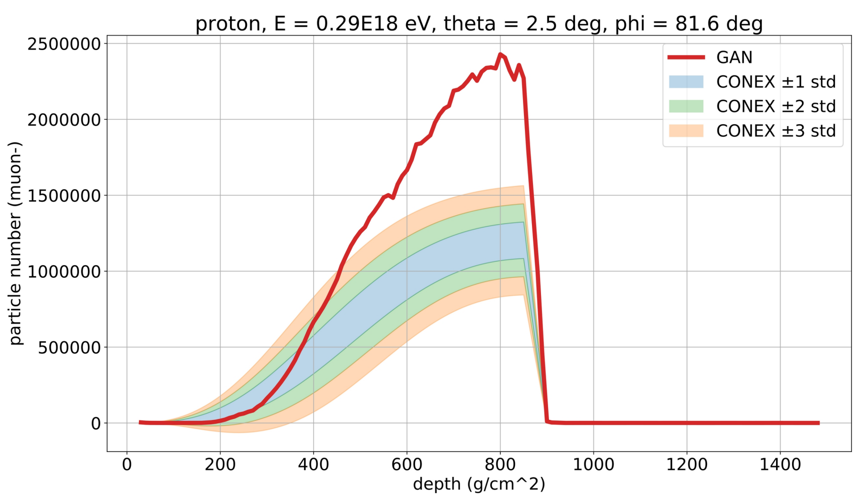


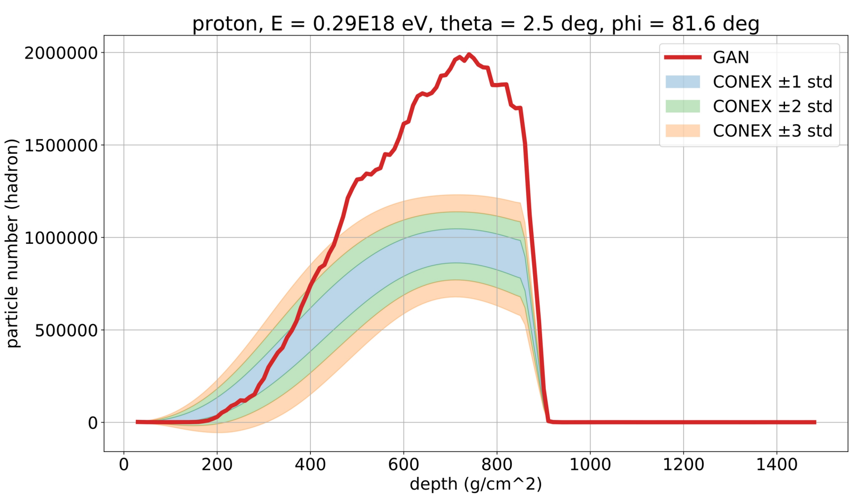
proton, E = 0.78E18 eV, theta = 33.4 deg, phi = 43.5 deg
particle number (hadron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std
0 500000 1000000 1500000 2000000 2500000
0 200 400 600 800 1000 1200 1400

CONEX vs. GAN

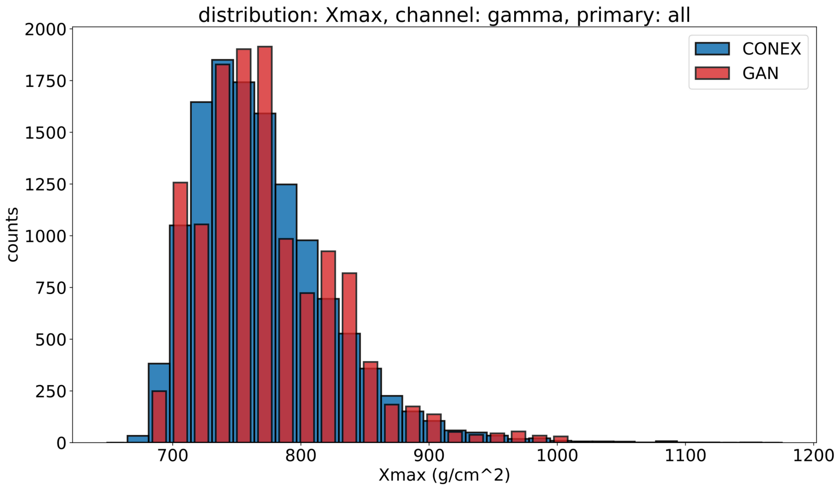

proton, E = 0.29E18 eV, theta = 2.5 deg, phi = 81.6 deg
1e9
particle number (gamma)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

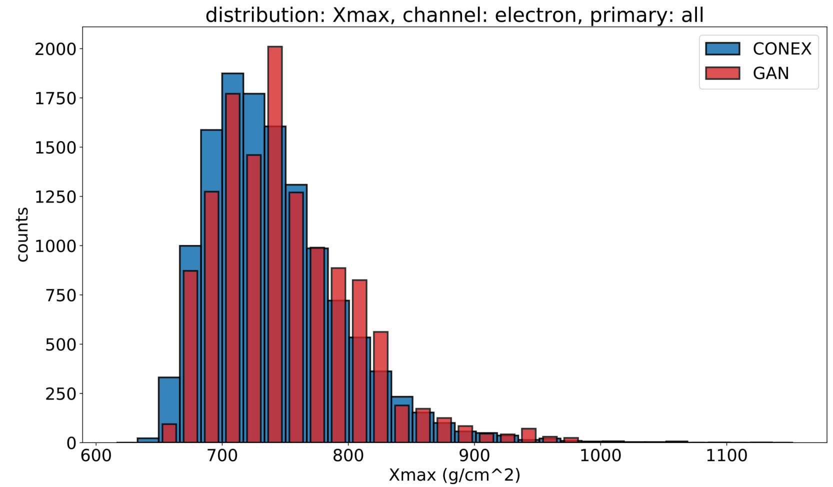

proton, E = 0.29E18 eV, theta = 2.5 deg, phi = 81.6 deg
1e8
particle number (electron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

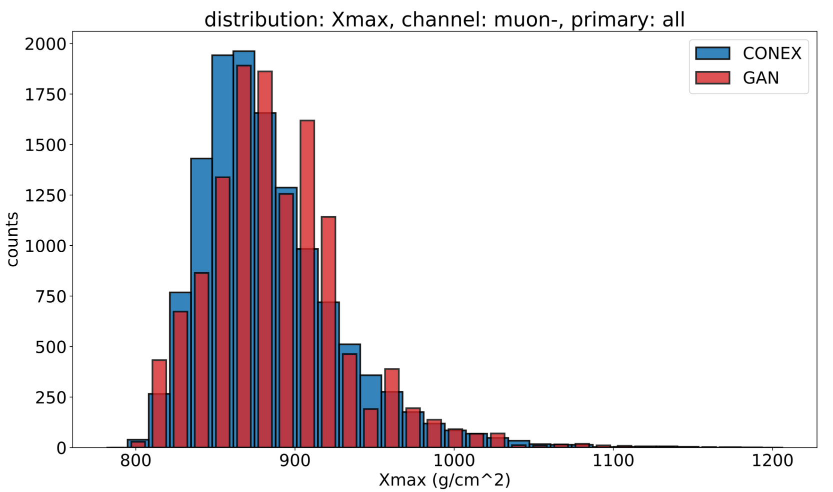

proton, E = 0.29E18 eV, theta = 2.5 deg, phi = 81.6 deg
particle number (muon-)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

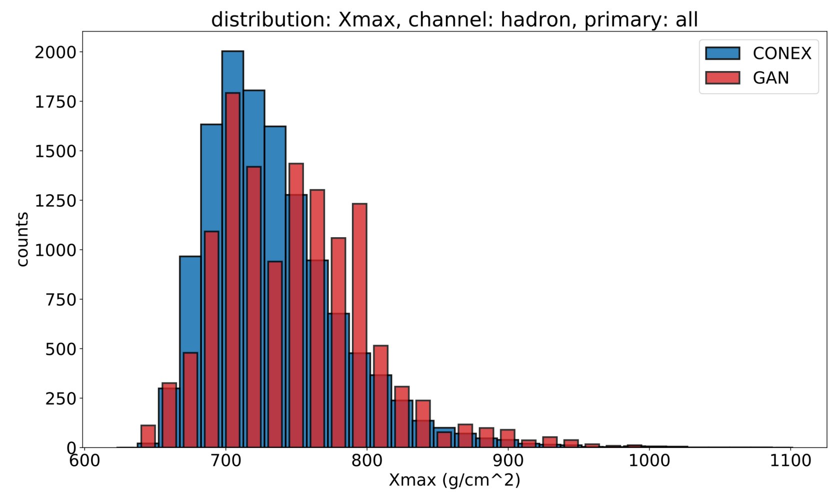

proton, E = 0.29E18 eV, theta = 2.5 deg, phi = 81.6 deg
particle number (hadron)
depth (g/cm^2)
GAN
CONEX ±1 std
CONEX ±2 std
CONEX ±3 std

Xmax Distribution (E > 5E18 eV, theta > 35 deg)







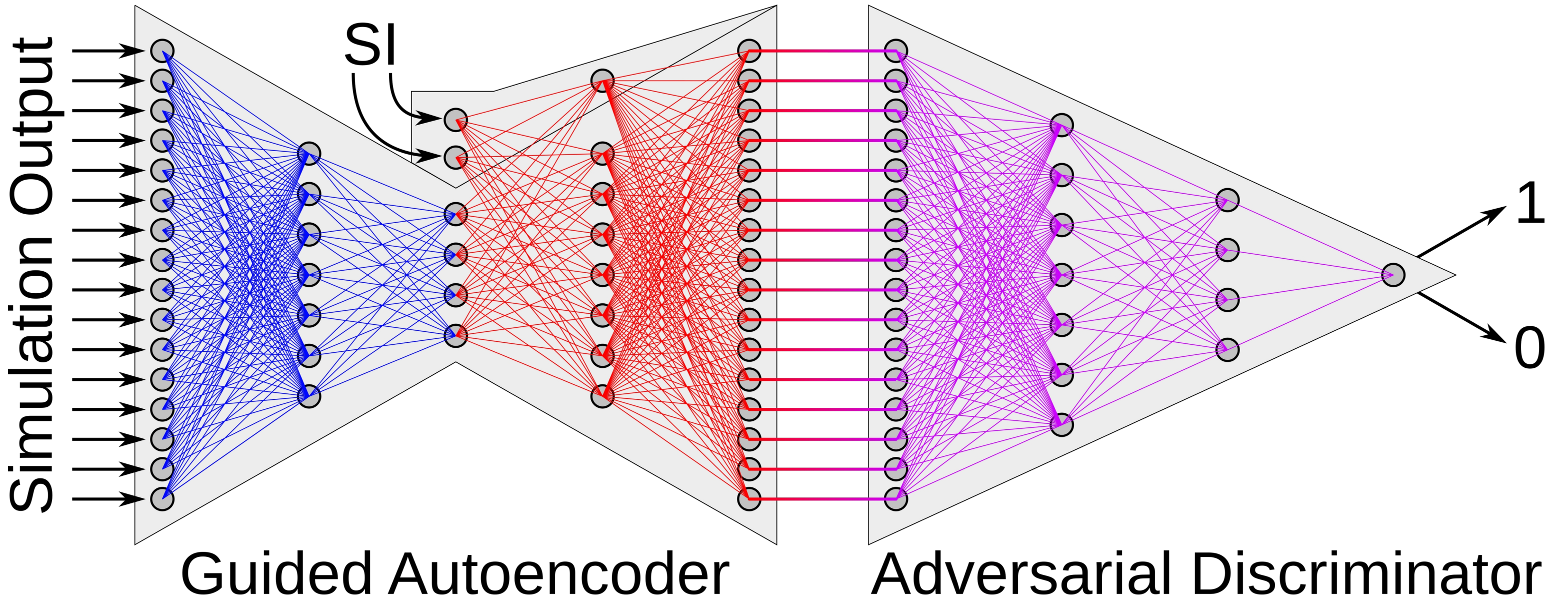


What‘s next?

- Fix low energy behavior (oversampling, architecture, constrainers, …)
- Verify and improve interpolation behavior
- Test adversarial vulnerability
- Template matching/reconstruction
- Refining with data

Fast Implicit Simulation Heuristic (FISH)

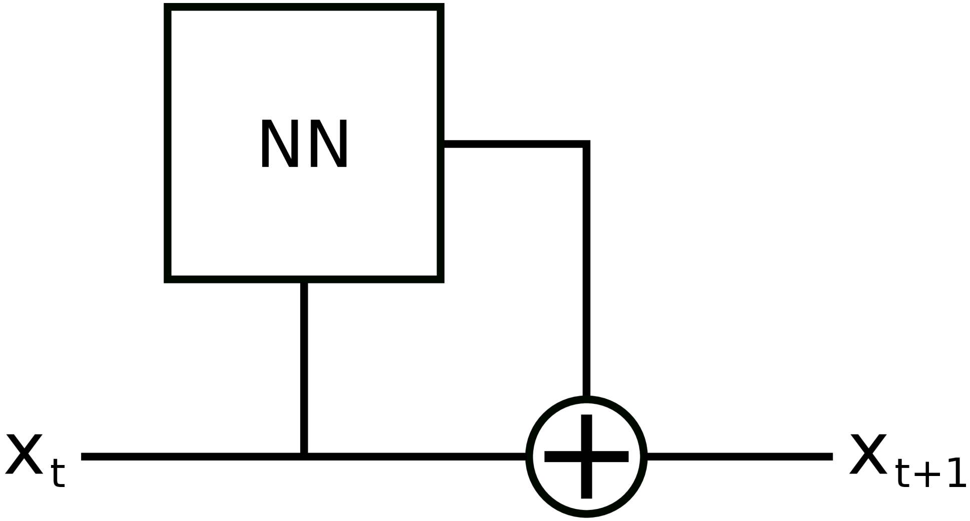
- Autoencoder with Adversarial Metric



- Simulation Input (SI) can be extended with meta-parameters
- Discriminator can be refined with real measurements

Adjustable Accuracy [3]

- ResNet



- Translate to ordinary differential equation (ODE)

$$x_{t+1} = x_t + f(x_t, \theta_t) \;\Rightarrow\; \frac{dx(t)}{dt} = f(x(t), t, \theta)$$

- Solve with standard ODE solver
- Adapt solver accuracy on the fly (training: high, inference: low)

References

- Title picture:
  Photo by Pixabay from Pexels

- Backup picture:
  Photo by Anthony from Pexels

- [1] CORSIKA 7: https://www.ikp.kit.edu/corsika/

- [2] Wasserstein Distance:
  Lambdabadger / CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0)

- [3] „Neural Ordinary Differential Equations“ - Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud – arXiv: 1806.07366