EVENT GENERATION AND STATISTICAL SAMPLING WITH DEEP GENERATIVE MODELS

Sydney Otten (University of Amsterdam and Radboud University)

3rd IML Machine Learning Workshop

CERN, Geneva, 17.04.2019

Arxiv: 1901.00875

Event Generation and Statistical Sampling with Deep Generative Models and a Density Information Buffer

Sydney Otten,[1,2,*] Sascha Caron,[1,3,†] Wieske de Swart,[1] Melissa van Beekveld,[1] Luc Hendriks,[1] Caspar van Leeuwen,[4] Damian Podareanu,[4] Roberto Ruiz de Austri,[5] and Rob Verheyen[1]

[1] *Institute for Mathematics, Astro- and Particle Physics IMAPP Radboud Universiteit, Nijmegen, The Netherlands*

[2] *GRAPPA, University of Amsterdam, The Netherlands*

[3] *Nikhef, Amsterdam, The Netherlands*

[4] *SURFsara, Amsterdam, The Netherlands*

[5] *Instituto de Fisica Corpuscular, IFIC-UV/CSIC University of Valencia, Spain*

THE BIG PICTURE

WHAT ARE WE DOING? WHY ARE WE DOING THIS?

YES, WE WANT TO PROVIDE AN ALTERNATIVE TO MC GENERATORS

But this requires Monte Carlo! Once trained, the event generation with our ML model is several orders of magnitude faster.

ALLOW FOR MORE “FREEDOM“ FOR GENERATING EVENTS

By enabling targeted event generation and by being able to interpolate between latent space representations

USE THE EVENT GENERATOR AS AN ANOMALY DETECTOR

Train on standard model data, detect anomalous individual events AND overdensities

WE CAN GENERATE BETTER RANDOM NUMBERS

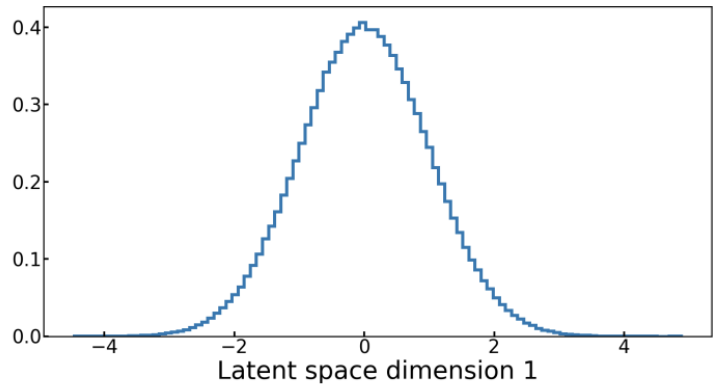
e.g. to improve rejection efficiency for MC integration

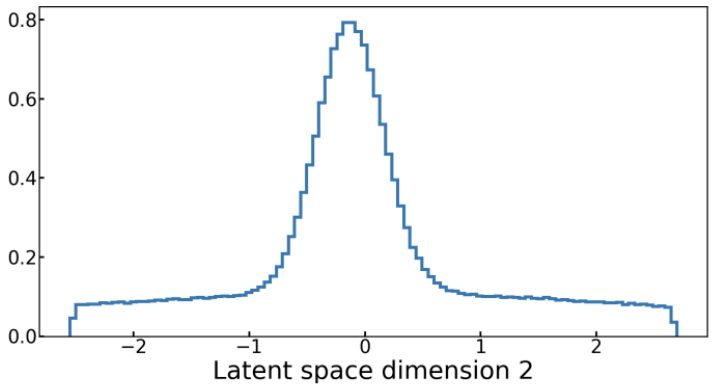
MACHINE LEARNING METHODS

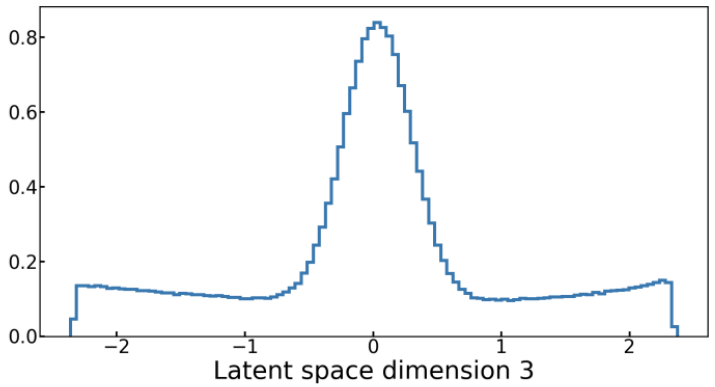
GANs and VAEs

VARIATIONAL AUTOENCODERS

- GANs covered in previous talks
- VAEs consist of two NNs with MSE + KL loss
- We use the Beta-VAE
- In Addition: Density buffer in latent space
- Beta-VAE + Buffer = B-VAE



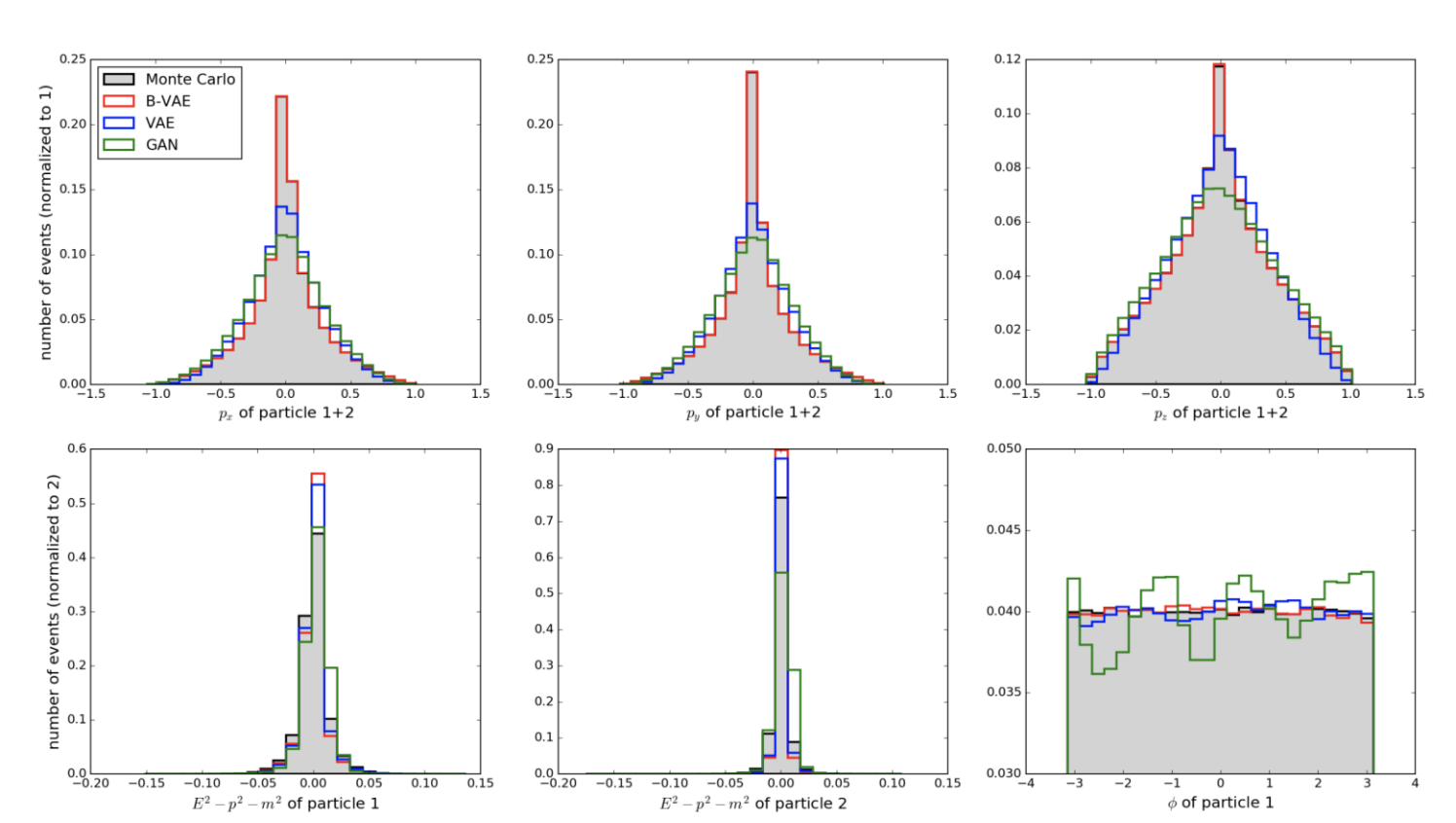




TWO BODY DECAY

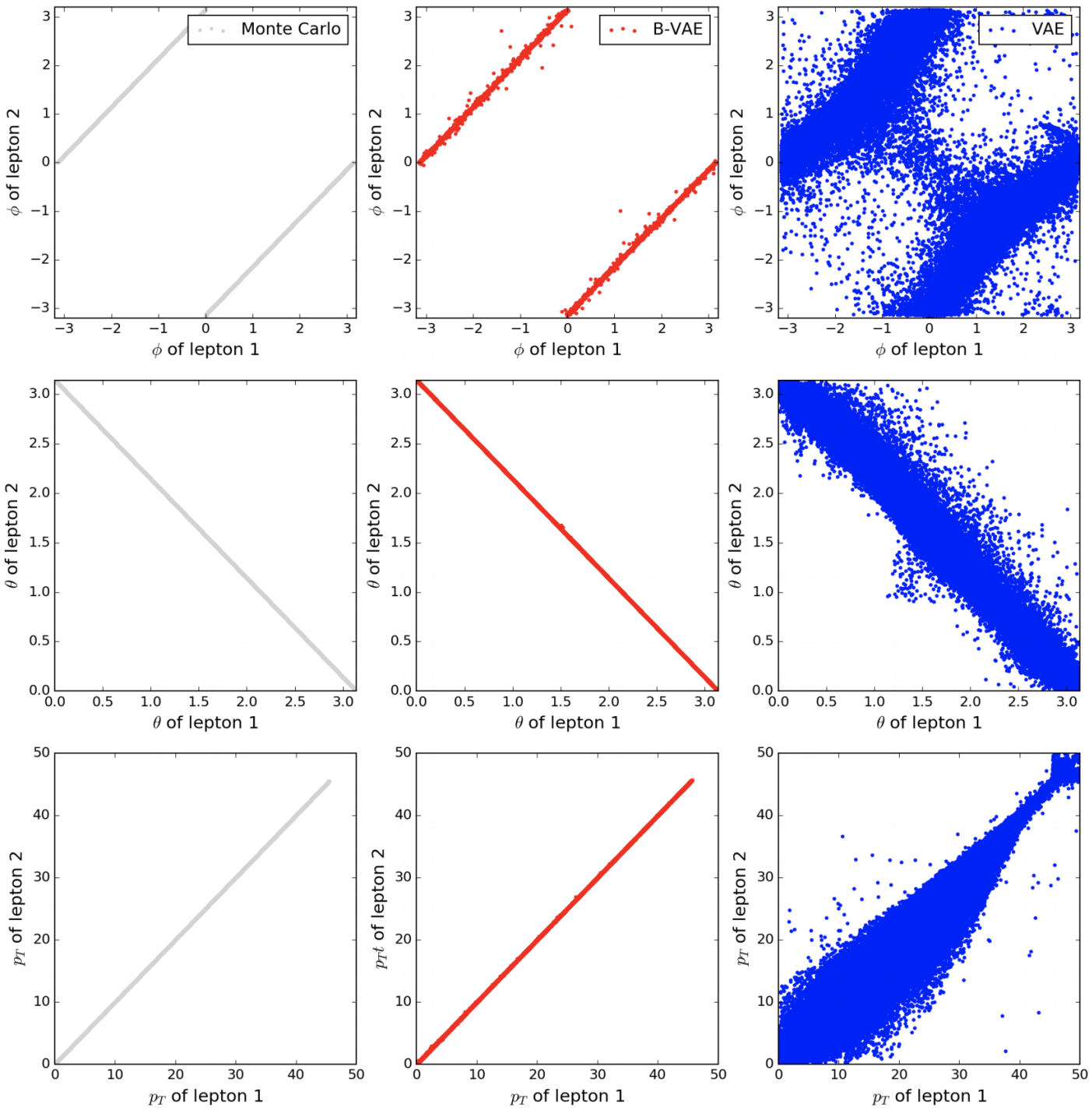
First simple toy model

GANS AND STANDARD VAES DON'T WORK WELL BUT B-VAE DOES



LEPTONIC Z DECAY

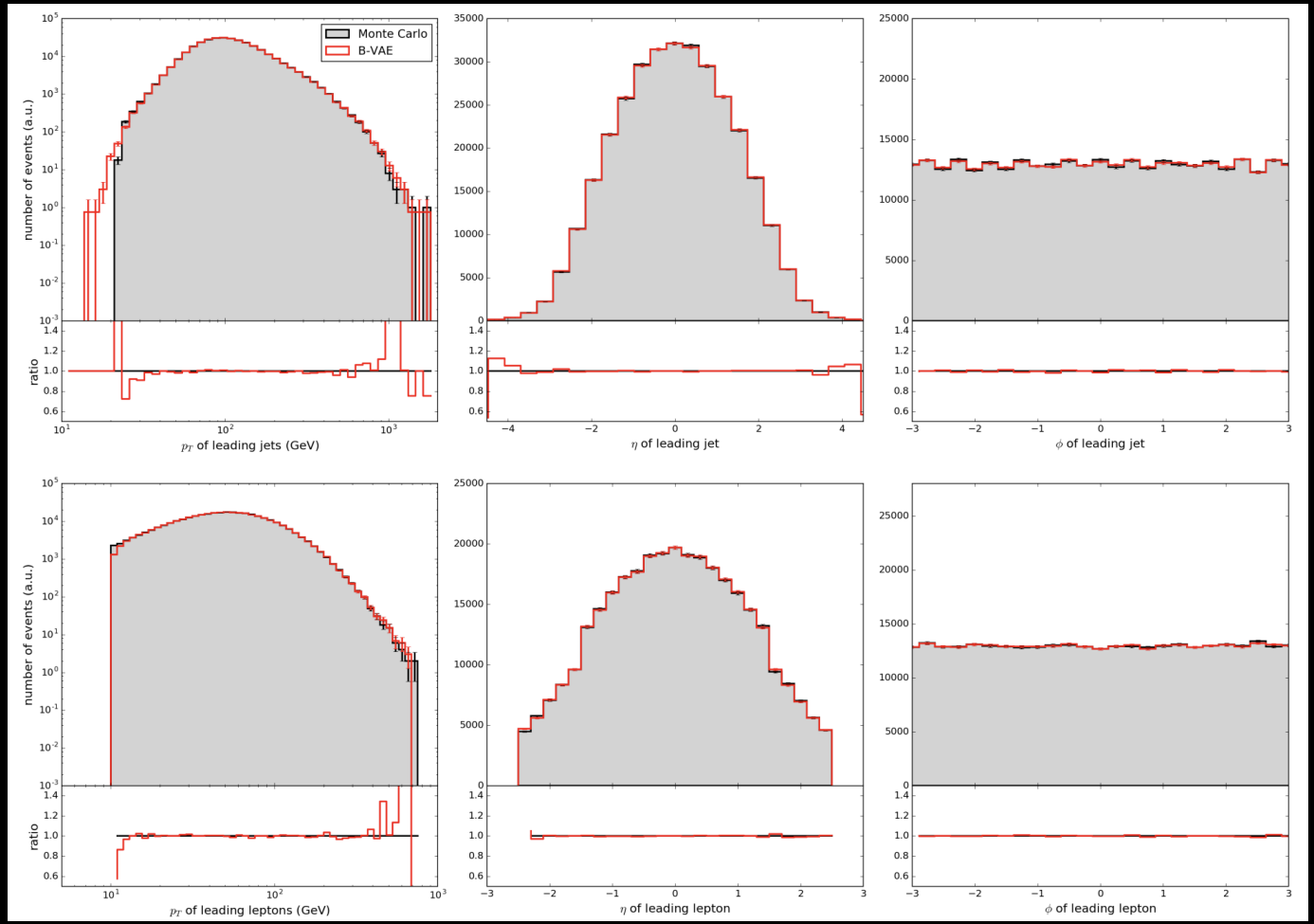
EVENTS ARE PRODUCED BACK TO BACK



TTBAR PRODUCTION

With up to four jets + leptons

ALSO WORKS WELL FOR COMPLICATED PROCESSES

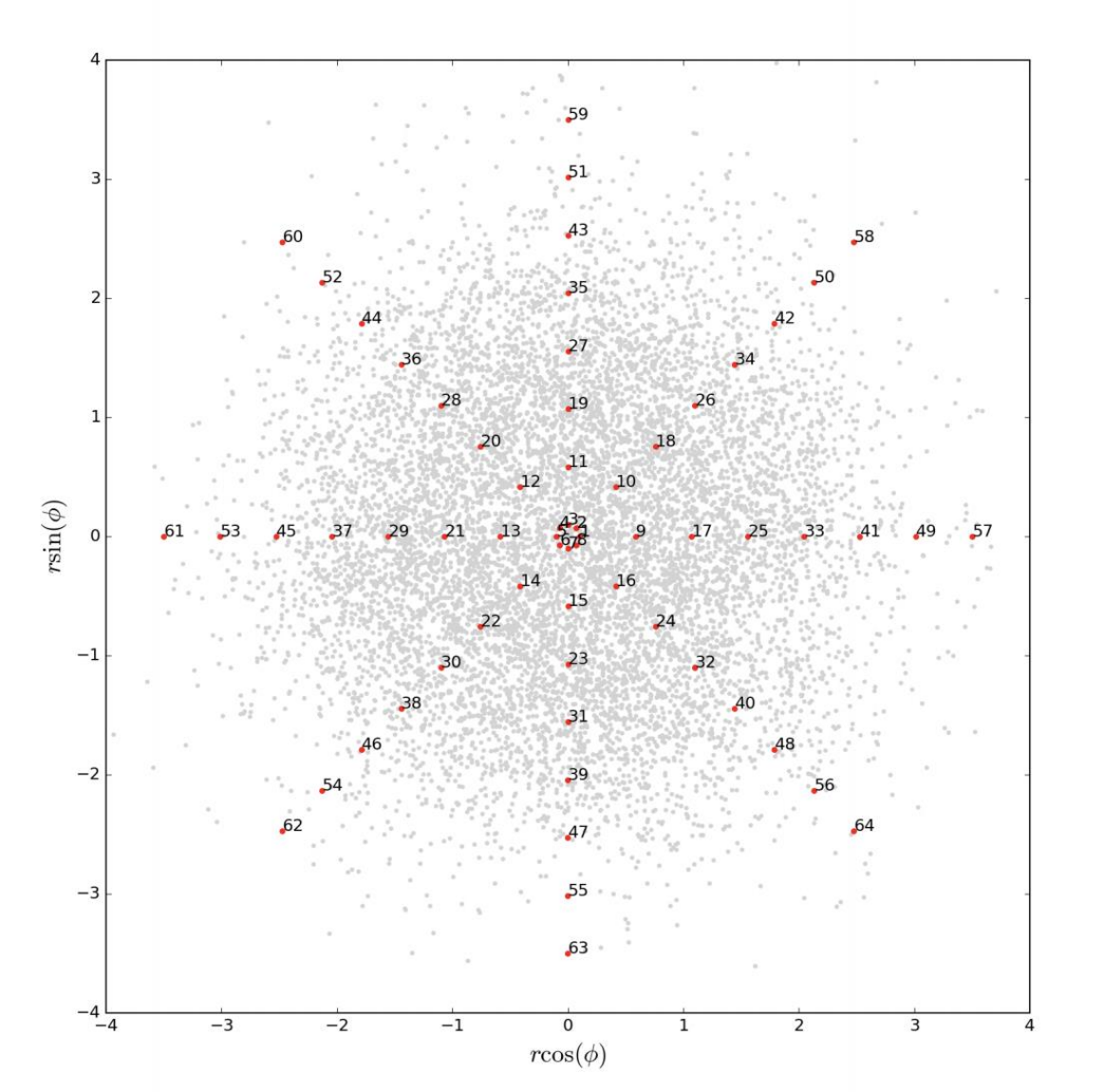


WHY ARE GANS INFERIOR?

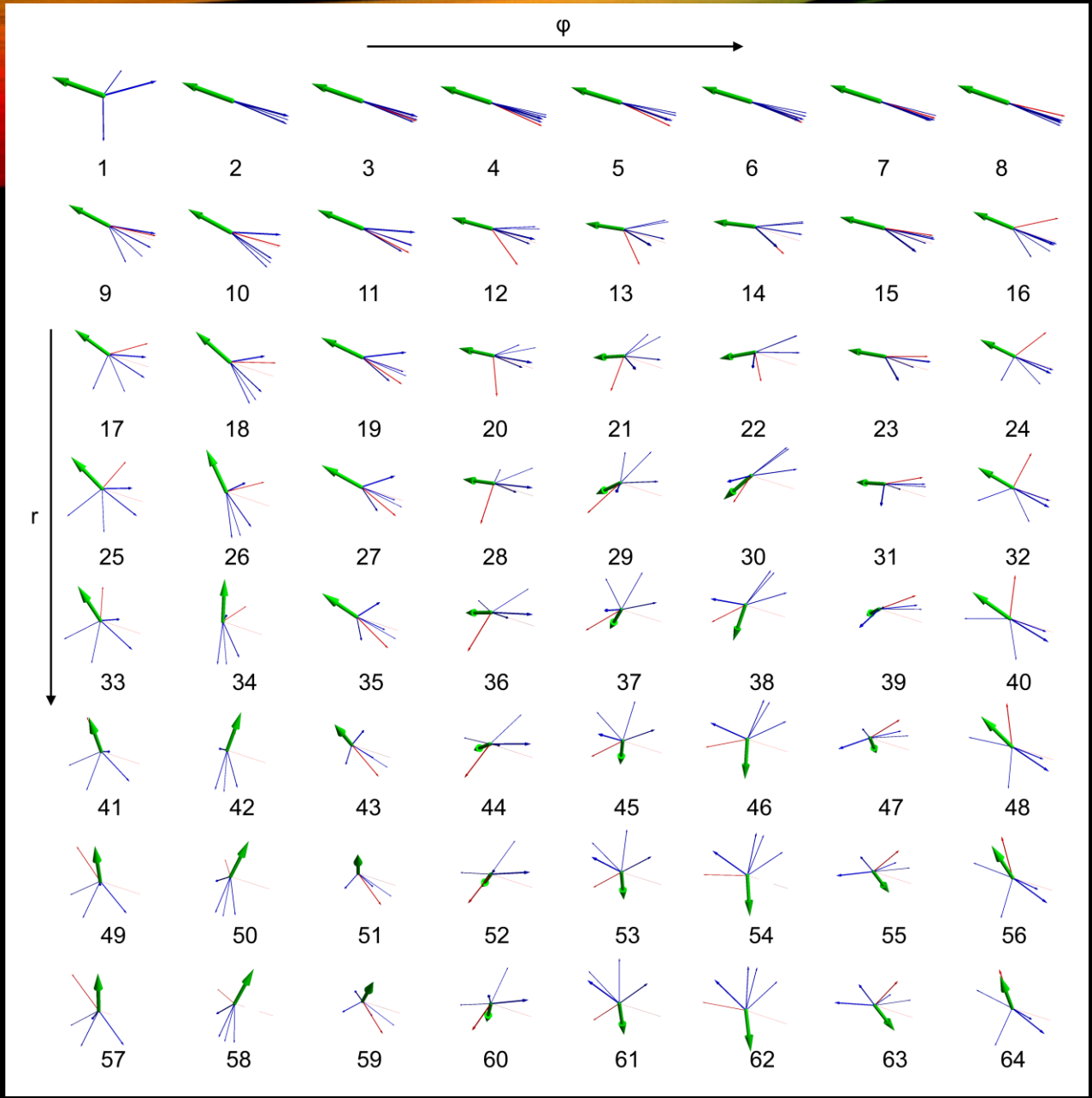
- Unexpected result, focused on GANs when starting project
- GANs generate much better images
- Here: no image data but 4-vectors
- Reason: the discriminator always predicts real or fake for one event at a time
- Therefore: frequency of occurence is not taken care of
- Think of a dataset with 80% cats and 20% dogs
- VAEs place all events in latent space, superior performance demonstrated

EXPLORING LATENT SPACE

With a principal component analysis



SAMPLING IN PCA SPACE



ALLOWS US TO STEER EVENT GENERATION!

CONCLUSION

- Basically we can learn any relevant probability distribution from data
- In particular we can learn to generate complicated events with the correct frequency of occurence
- Has many applications:
  - An 82-dimensional event generator case including many sparse entries worked reasonably well
  - More efficient MC sampling e.g. for integrating matrix elements
  - Learn generator directly from experimental data
  - Create an anomaly detector for new physics
  - Learn the detector response (and its inverse)
  - Applications beyond particle physics

THANK YOU FOR YOUR ATTENTION!