How to GAN for LHC

Tilman Plehn

Universität Heidelberg

CERN 3/2020

GAN basics

MC crucial for LHC physics

– goal: data-to-data with fundamental physics input only

– MC challenges

  higher-order precision in bulk
  coverage of tails
  inversion/unfolding to access fundamental QCD

– neural network benefits

  best available interpolation
  structured latent space
  lightning speed, once trained
  inversion solved
  training on MC and/or data, anything goes

– GANs the coolest kid on the block

  generator trying to produce best events
  discriminator trying to catch generator
  $\longrightarrow$ competing towards (Nash) equilibrium

GAN algorithm

GANning LHC events

- training: true events $\{x_T\}$ following $p_T(x)$
  output: generated events $\{r\} \to \{x_G\}$ following $p_G(x)$
- discriminator constructing $D(x)$ [$D(x) = 1, 0$ true/generator]

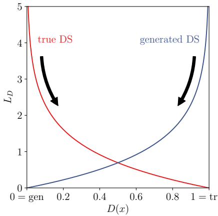
$$L_D = \big\langle -\log D(x) \big\rangle_{x\sim P_T} + \big\langle -\log(1 - D(x)) \big\rangle_{x\sim P_G} \to 2\log 0.5$$

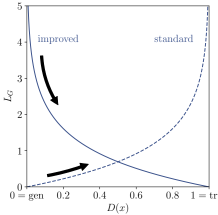
- generator giving events [D needed]

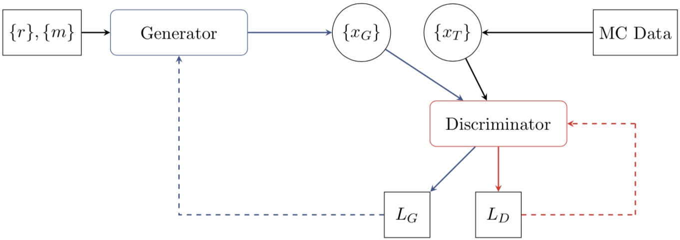
$$L_G = \big\langle -\log D(x) \big\rangle_{x\sim P_G} \to -\log 0.5$$

- loss function evaluated over batch
- noise reduction/stabilization: gradient penalty

$\Rightarrow$ statistically independent copy of training events







GANs at LHC

Phase space networks

- MC integration [Bendavit (2017)]
- NNVegas [Klimek (2018)]

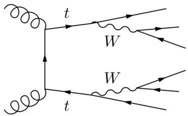
Existing GAN studies

- Jet Images [de Oliveira (2017), Carazza (2019)]
- Particle shower in Calorimeters [Paganini (2017), Musella (2018), Erdmann (2018), Ghosh (2018)]
- Event generation [Otten(2019), Hashemi (2019), Di Sipio (2019), Butter (2019), Martinez (2019), Alanazi (2020)]
- Unfolding [Datta (2018), Bellagente (2019)]
- Templates for QCD factorization [Lin (2019)]
- EFT models [Erbin (2018)]
- Event subtraction [Butter (2019)]

MC generators

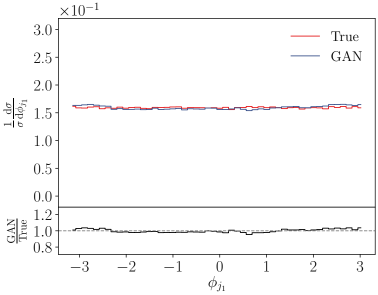
- neural importance sampling [Bothmann (2020)]
- i-flow in SHERPA [Gao (2020)]

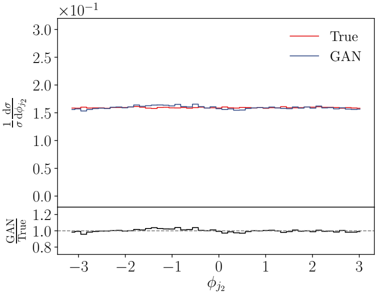
1– How to GAN LHC Events



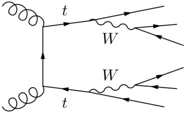
Idea: replace ME for hard process [Butter, TP, Winterhalder]

- medium-complex final state $t\bar{t} \to 6$ jets
  $t/\bar{t}$ and $W^{\pm}$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to$ 12 dof [constants hard to learn]
- flat observables flat [phase space coverage okay]





1– How to GAN LHC Events



Idea: replace ME for hard process [Butter, TP, Winterhalder]

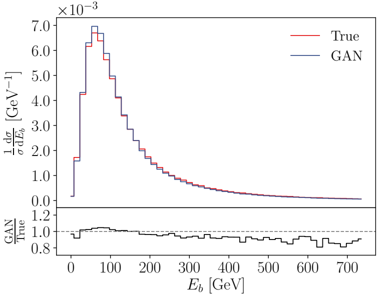
– medium-complex final state $t\bar{t} \to 6$ jets

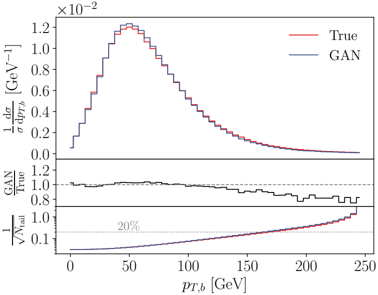
$t/\bar{t}$ and $W^\pm$ on-shell with BW $6 \times 4 = 18$ dof

on-shell external states $\to$ 12 dof [constants hard to learn]

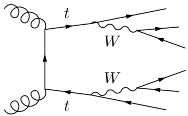
– flat observables flat [phase space coverage okay]

– direct observables with tails [statistical error indicated]





1– How to GAN LHC Events



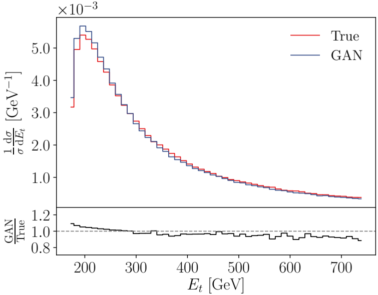
Idea: replace ME for hard process [Butter, TP, Winterhalder]

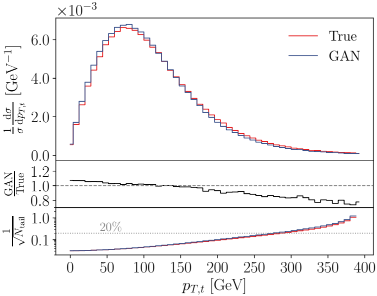
– medium-complex final state $t\bar{t} \to 6$ jets
  $t/\bar{t}$ and $W^\pm$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to 12$ dof [constants hard to learn]

– flat observables flat [phase space coverage okay]

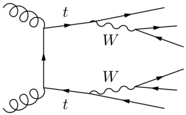
– direct observables with tails [statistical error indicated]

– constructed observables similar





1– How to GAN LHC Events



Idea: replace ME for hard process [Butter, TP, Winterhalder]

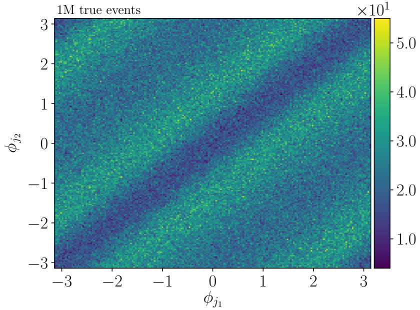
– medium-complex final state $t\bar{t} \to 6$ jets
  $t/\bar{t}$ and $W^{\pm}$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to$ 12 dof [constants hard to learn]

– flat observables flat [phase space coverage okay]

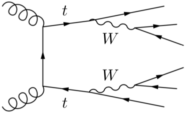
– direct observables with tails [statistical error indicated]

– constructed observables similar

– improved resolution [1M training events]

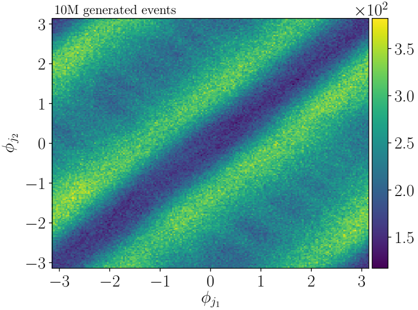


1– How to GAN LHC Events

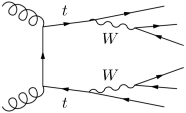


Idea: replace ME for hard process [Butter, TP, Winterhalder]

- medium-complex final state $t\bar{t} \to 6$ jets
  $t/\bar{t}$ and $W^\pm$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to 12$ dof [constants hard to learn]
- flat observables flat [phase space coverage okay]
- direct observables with tails [statistical error indicated]
- constructed observables similar
- improved resolution [10M generated events]

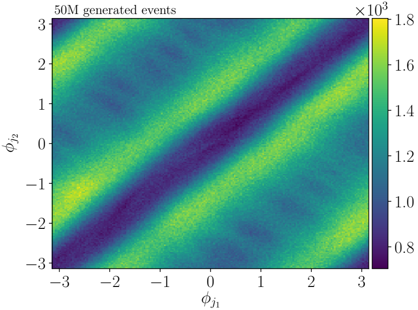


1– How to GAN LHC Events



Idea: replace ME for hard process [Butter, TP, Winterhalder]

- medium-complex final state $t\bar{t} \to 6$ jets
  $t/\bar{t}$ and $W^{\pm}$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to 12$ dof [constants hard to learn]
- flat observables flat [phase space coverage okay]
- direct observables with tails [statistical error indicated]
- constructed observables similar
- improved resolution [50M generated events]
- GAN concept working



Intermediate resonances

Narrow phase space structures

– VEGAS: adaptive sampling

– MC: phase space mapping [BW $\to$ flat, multi-channel]

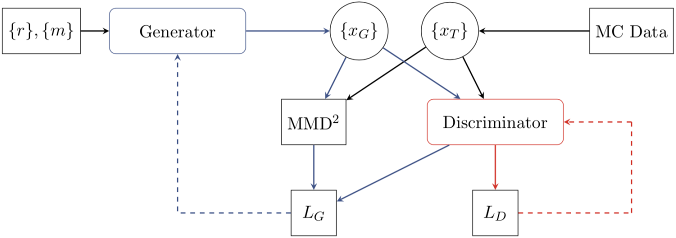
– generally 1D features

phase space boundaries
kinematic cuts
invariant masses [top, W]

– batch-wise comparison of distributions, MMD loss with kernel k

$$\text{MMD}^2 = \left\langle k(x,x')\right\rangle_{x,x'\sim P_T} + \left\langle k(y,y')\right\rangle_{y,y'\sim P_G} - 2\left\langle k(x,y)\right\rangle_{x\sim P_T,y\sim P_G}$$

$$L_G \to L_G + \lambda_G \,\text{MMD}^2 \,,$$



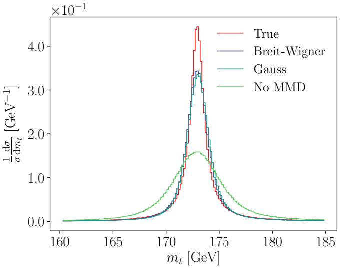
Event quality

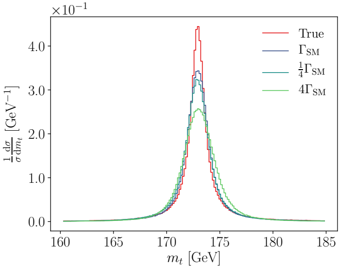
Study of MMD loss [coming again, later]

– input: function of 4-momenta, rough resolution

not: value of intermediate mass
not: value of width [dynamic feature]

– minor impact of kernel function and width





Event quality

Study of MMD loss [coming again, later]

- input: function of 4-momenta, rough resolution

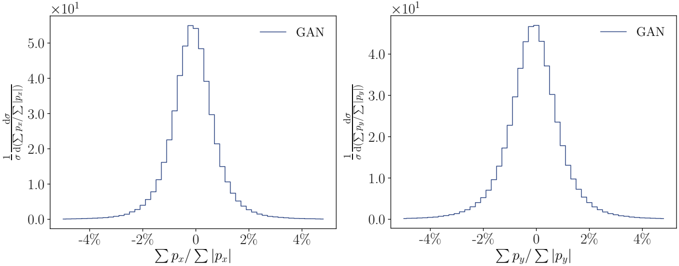
  not: value of intermediate mass

  not: value of width [dynamic feature]

- minor impact of kernel function and width

Challenges

- momentum conservation to per-cent



Event quality

Study of MMD loss [coming again, later]

– input: function of 4-momenta, rough resolution

not: value of intermediate mass
not: value of width [dynamic feature]

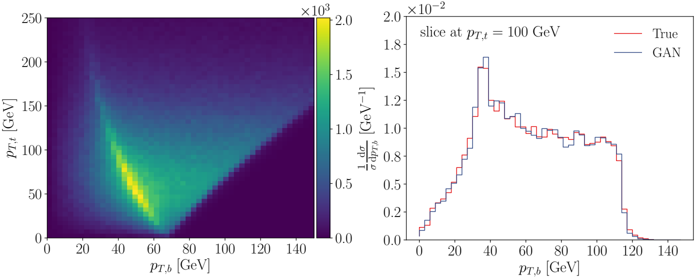
– minor impact of kernel function and width

Challenges

– momentum conservation to per-cent

– 2D correlations

– Case convincing?



2– How to GAN event subtraction

Idea: subtract event samples without bins [Butter, TP, Winterhalder]

– statistical uncertainty

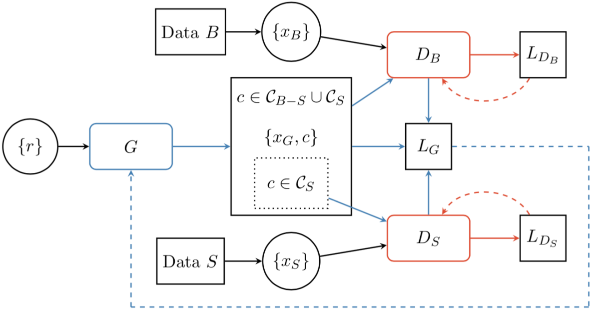
$$\Delta_{B-S} = \Delta_{n_B N_B - n_S N_S} = \sqrt{\Delta^2_{n_B N_B} + \Delta^2_{n_S N_S}} = \sqrt{n_B^2 N_B + n_S^2 N_S} > \max(B, S)$$

– applications in LHC physics

soft-collinar subtraction, multi-jet merging
on-shell subtraction
background/signal subtraction

– GAN setup

1. differential, steep class label
2. sample normalization



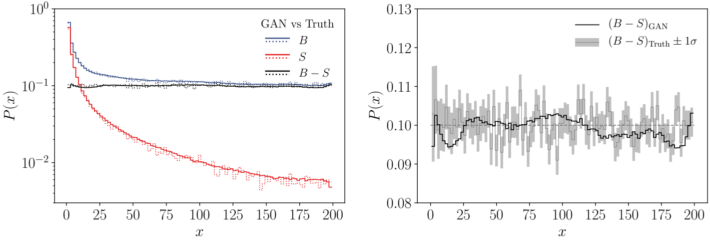
Subtracted events

How to beat statistics by subtracting

1– 1D toy example

$$P_B(x) = \frac{1}{x} + 0.1 \qquad P_S(x) = \frac{1}{x} \qquad \Rightarrow \qquad P_{B-S} = 0.1$$

– statistical fluctuations reduced (sic!)



Subtracted events

How to beat statistics by subtracting

1– 1D toy example

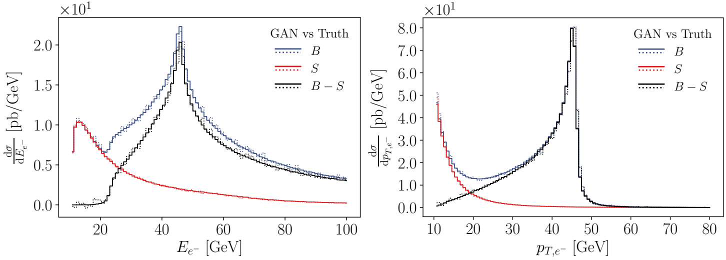
$$P_B(x) = \frac{1}{x} + 0.1 \qquad P_S(x) = \frac{1}{x} \qquad \Rightarrow \qquad P_{B-S} = 0.1$$

– statistical fluctuations reduced (sic!)

2– event-based background subtraction [weird notation, sorry]

$$pp \to e^+e^- \quad \text{(B)} \qquad pp \to \gamma \to e^+e^- \quad \text{(S)}$$

– Z-pole remaining



Subtracted events

How to beat statistics by subtracting

1– 1D toy example

$$P_B(x) = \frac{1}{x} + 0.1 \qquad P_S(x) = \frac{1}{x} \qquad \Rightarrow \qquad P_{B-S} = 0.1$$

– statistical fluctuations reduced (sic!)

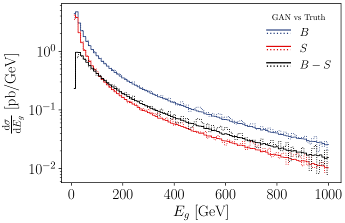
2– event-based background subtraction [weird notation, sorry]

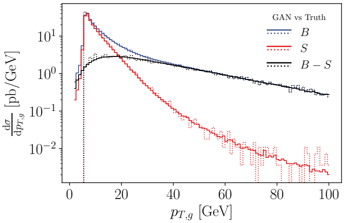
$$pp \to e^+e^- \quad \text{(B)} \qquad pp \to \gamma \to e^+e^- \quad \text{(S)}$$

– Z-pole remaining

3– collinear subtraction [assumed non-local]

$$pp \to Zg \qquad \text{(B: matrix element, S: collinear approximation)}$$





– Proper new idea, any applications?

3– How to GAN away detector effects

Open problem of publishing kinematic information [e.g. global SMEFT analyses]

– total rates without necessary information
  STXS model-dependent
  unfolded distributions extremely convenient [$t\bar{t}$ results]

– challenges in unfolding

  non-invertible detector simulation
  model dependence
  flexibility/reliability [training on some event set]

– benefits from unfolding actual data

  access to hard matrix element/first-principles QCD
  matrix element method

General: how to invert Markov processes [Bellagente, Butter, Kasiczka, TP, Winterhalder]

– detector simulation typical Markov process

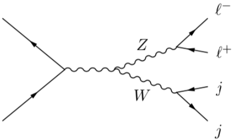
– inversion possible, in principle [entangled convolutions]

– GAN task

$$\text{partons} \xrightarrow{\text{DELPHES}} \text{detector} \xrightarrow{\text{GAN}} \text{partons}$$

$\Rightarrow$ Full unfolded phase space

Standard GAN

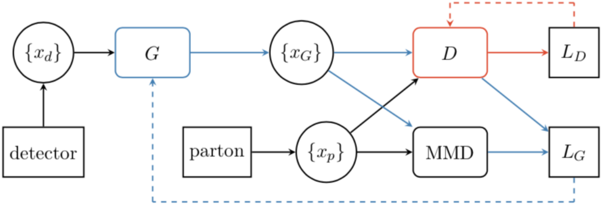


Reconstructing the parton level

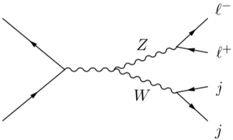
– $pp \to ZW \to (\ell\ell)\ (jj)$

– broad jj mass peak
narrow $\ell\ell$ mass peak
modified $2 \to 2$ kinematics
fun phase space boundaries

– GAN same as event generation [with MMD]



Standard GAN



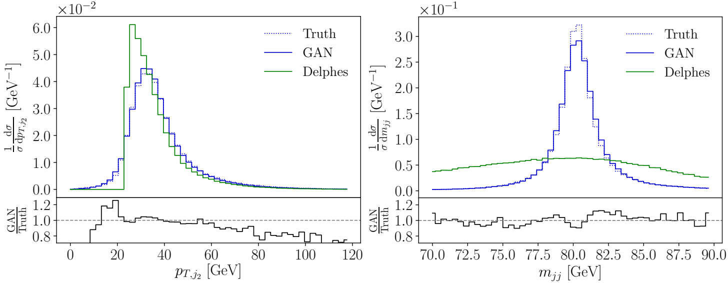
Reconstructing the parton level

– $pp \to ZW \to (\ell\ell)\ (jj)$

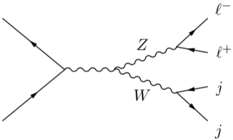
– broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
  fun phase space boundaries

– GAN same as event generation [with MMD]

– full inversion fine



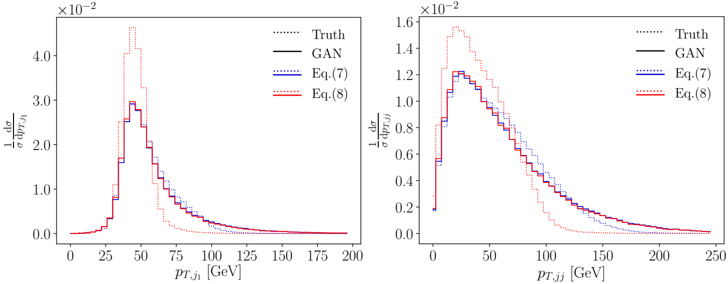
Standard GAN



Reconstructing the parton level

- $pp \to ZW \to (\ell\ell)\ (jj)$
- broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
  fun phase space boundaries
- GAN same as event generation [with MMD]
- full inversion fine
- serious problem: kinematics cuts in test data [88%, 38% events]

$$p_{T,j_1} = 30 \ldots 100 \text{ GeV} \tag{7}$$
$$p_{T,j_1} = 30 \ldots 60 \text{ GeV} \quad \text{and} \quad p_{T,j_2} = 30 \ldots 50 \text{ GeV} \tag{8}$$

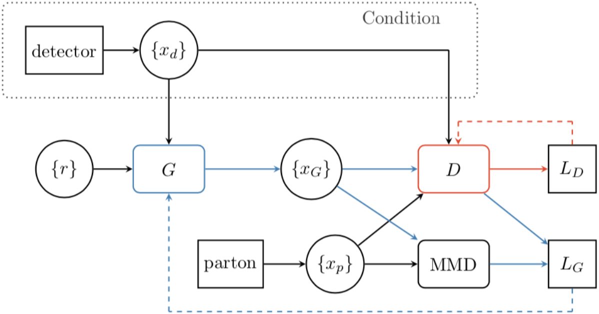


Fully conditional GAN

Conditional GAN

– map random numbers to parton level

  hadron level as condition [matched event pairs]

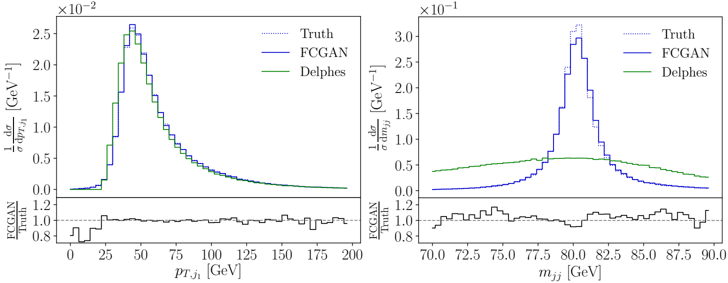


Fully conditional GAN

Conditional GAN

– map random numbers to parton level
  hadron level as condition [matched event pairs]

– full inversion also fine



Fully conditional GAN

Conditional GAN

– map random numbers to parton level

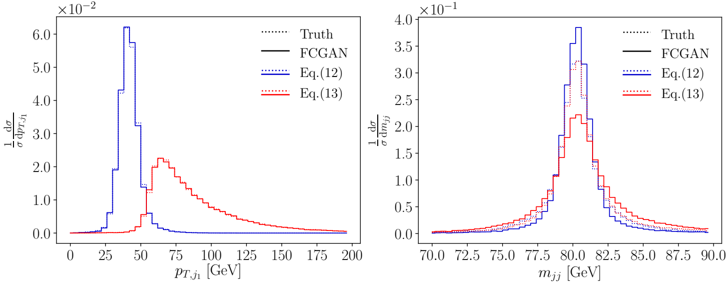
  hadron level as condition [matched event pairs]

– full inversion also fine

– tougher cuts challenging m_{jj} [14%, 39% events, no interpolation, MMD not conditional]

$$p_{T,j_1} = 30 \dots 50 \text{ GeV} \quad p_{T,j_2} = 30 \dots 40 \text{ GeV} \quad p_{T,\ell-} = 20 \dots 50 \text{ GeV} \qquad (12)$$

$$p_{T,j_1} > 60 \text{ GeV} \qquad (13)$$



Fully conditional GAN

Conditional GAN

– map random numbers to parton level
  hadron level as condition [matched event pairs]

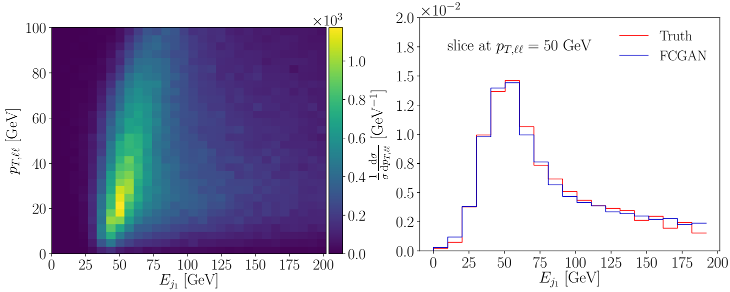
– full inversion also fine

– tougher cuts challenging m_{jj} [14%, 39% events, no interpolation, MMD not conditional]

$$p_{T,j_1} = 30 \dots 50 \text{ GeV} \quad p_{T,j_2} = 30 \dots 40 \text{ GeV} \quad p_{T,\ell^-} = 20 \dots 50 \text{ GeV} \tag{12}$$

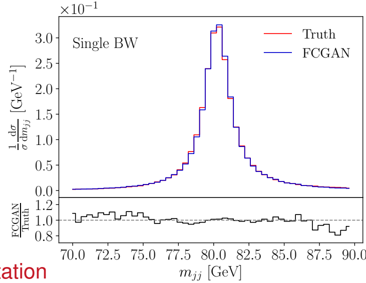
$$p_{T,j_1} > 60 \text{ GeV} \tag{13}$$

– pretty pictures in 2D



$\Rightarrow$ 1.FCGAN unfolding works!

Dynamic MMD



Technical side-remark: dynamic MMD

– minimal input

functional form of correlation m_{jj}
kernel shape (irrelevant) and resolution

– Adaptive resolution?

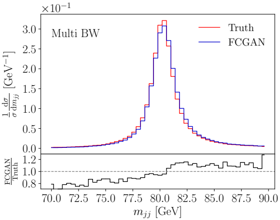
Technical side-remark: dynamic MMD implementation

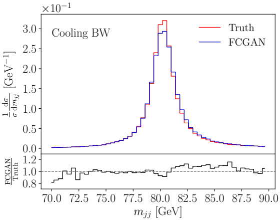
– multiple fixed-width kernels

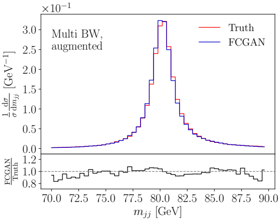
– multiple kernels for conditional input

– cooling kernel [from SD of generator m_{jj}]

$\Rightarrow$ Technical implementation still open...







BSM Injection

Different training (MC) and actual data... [not in v1, thank you to Ben Nachman]

...or model dependence of unfolding

...or localization in latent space

– train: SM events
  test: 10% events with W' in s-channel $\Rightarrow$ Any guesses?

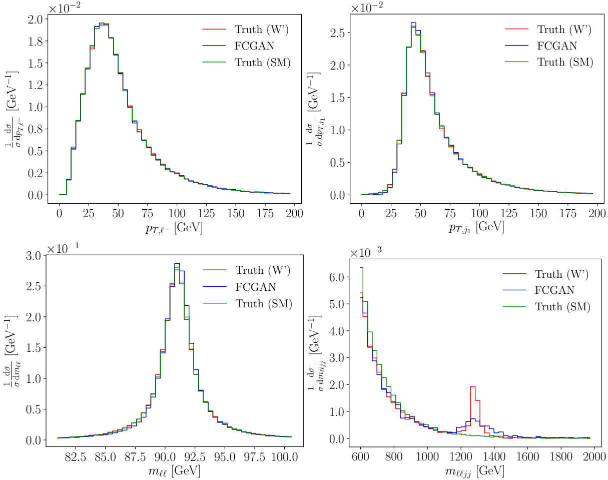
BSM Injection

Different training (MC) and actual data... [not in v1, thank you to Ben Nachman]

...or model dependence of unfolding

...or localization in latent space

– train: SM events
  test: 10% events with W' in s-channel ⇒ Any guesses?



4– Superresolution GANs (preview)

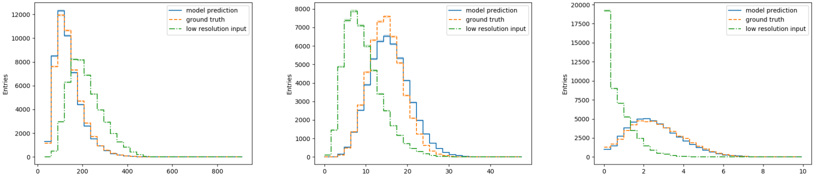
Getting inspired [Blecher, Butter, Keilbach, TP + Irvine]

– take high-resolution calorimeter images
  down-sample to 1/8th 1D resolution
  GAN inversion

– works because the GAN learn structure [showers are QCD]

– start from low-resolution calorimeter images
  GAN high-resolution images

– energy of constituents no.1,10,30



⇒ GANs are kind of magic

Outlook

Machine learning a great tool box

LHC physics really is big data
imagine classification was a starting point
jet classification largely established

generative networks good for theory
advantage 1: NN interpolation
advantage 2: latent space structures
advantage 3: training on MC and/or data

Any ideas?