Forward & Inverse LHC Simulations with NNs

Tilman Plehn

Universität Heidelberg

Nanjing Normal University, 10/2021

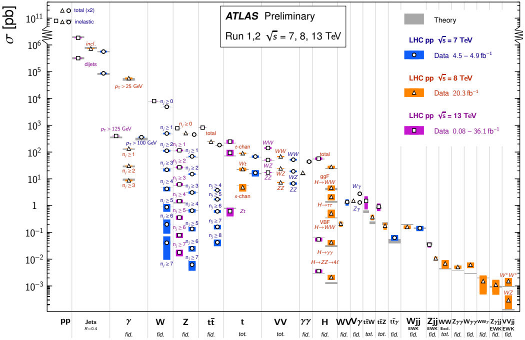
LHC goals

Fundamental questions

- particle nature of dark matter?
- origin of the Higgs mechanism? [hierarchy problem?]
- matter-antimatter asymmetry? [CP-symmetry]
- Standard Model all there is?

Rate measurements

- many processes
- vastly different rates
- high precision
- predicted by theory



LHC goals

Fundamental questions

– particle nature of dark matter?

– origin of the Higgs mechanism? [hierarchy problem?]

– matter-antimatter asymmetry? [CP-symmetry]

– Standard Model all there is?

Rate measurements

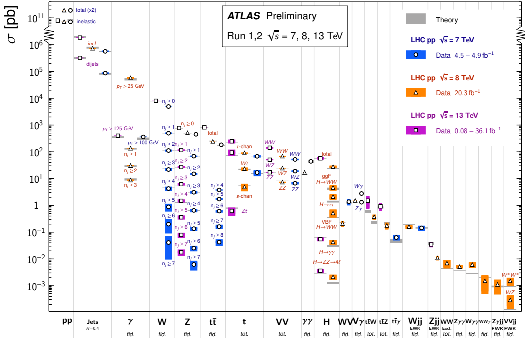
– many processes

– vastly different rates

– high precision

– predicted by theory

– but completely useless!



LHC goals

Fundamental questions

– particle nature of dark matter?

– origin of the Higgs mechanism? [hierarchy problem?]

– matter-antimatter asymmetry? [CP-symmetry]

– Standard Model all there is?

Rate measurements

– many processes

– vastly different rates

– high precision

– predicted by theory

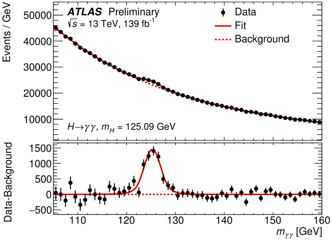
– but completely useless!

Rates not interesting

– energy distributions dropping

– new physics heavy

$\Rightarrow$ bumps, tails, kinematics instead

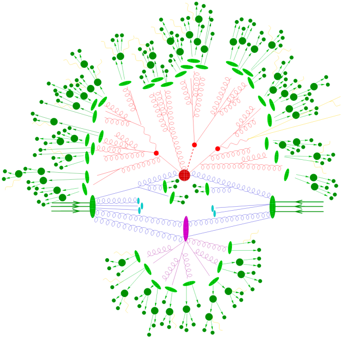


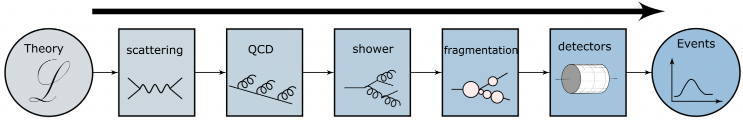
LHC simulations

Simulation-based inference [likelihood-free inference]

- start with Lagrangian
- calculate scattering in perturbative QFT
- simulate events [theory: Sherpa, Madgraph, Pythia]
- simulate detectors [experiment: ATLAS, CMS, Delphes]

$\Rightarrow$ LHC events in virtual worlds





LHC simulations

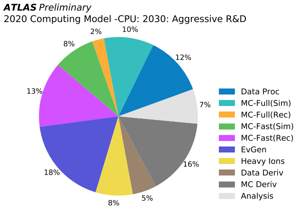
Simulation-based inference [likelihood-free inference]

- start with Lagrangian
- calculate scattering in perturbative QFT
- simulate events [theory: Sherpa, Madgraph, Pythia]
- simulate detectors [experiment: ATLAS, CMS, Delphes]

$\Rightarrow$ LHC events in virtual worlds

HL-LHC: preparing for 25-fold data set

- simulated event numbers $\sim$ expected events
- statistics requiring 1%-2% uncertainty [NNLO/N^{3}LO]
- flexible signal hypotheses [time-dependent]
- low-rate high-multiplicity backgrounds



LHC simulations

Simulation-based inference [likelihood-free inference]

- start with Lagrangian
- calculate scattering in perturbative QFT
- simulate events [theory: Sherpa, Madgraph, Pythia]
- simulate detectors [experiment: ATLAS, CMS, Delphes]

$\Rightarrow$ LHC events in virtual worlds

HL-LHC: preparing for 25-fold data set

- simulated event numbers $\sim$ expected events
- statistics requiring 1%-2% uncertainty [NNLO/N^{3}LO]
- flexible signal hypotheses [time-dependent]
- low-rate high-multiplicity backgrounds

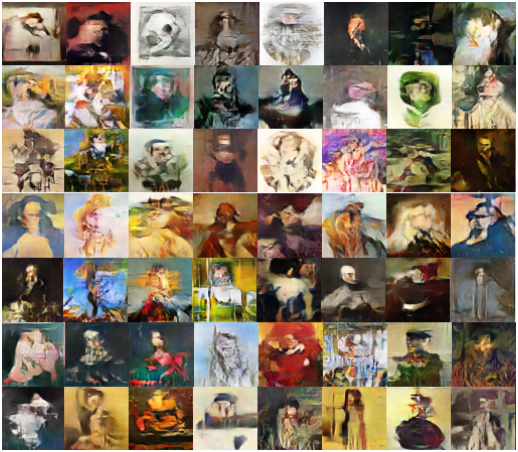
Three ways to use ML

- improve current tools: iSherpa, ML-MadGraph...
- new ideas: fast ML-generator-networks...
- conceptual ideas for simulations and analyses

Generative neural networks

GANGogh [Bonafilia, Jones, Danyluk (2017)]

- can networks create new pieces of art?
  map random numbers to image pixels
- train on 80,000 pictures [organized by style and genre]
- generate portraits



Generative neural networks

GANGogh [Bonafilia, Jones, Danyluk (2017)]

- can networks create new pieces of art?
  map random numbers to image pixels
- train on 80,000 pictures [organized by style and genre]
- generate portraits

Edmond de Belamy [Caselles-Dupre, Fautrel, Vernier (2018)]

- trained on 15,000 portraits
- sold for $432.500

$\Rightarrow$ ML all marketing and sales

Generative neural networks

GANGogh [Bonafilia, Jones, Danyluk (2017)]

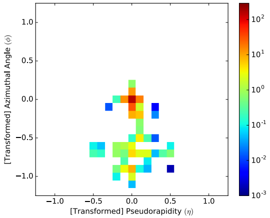
– can networks create new pieces of art?
map random numbers to image pixels
– train on 80,000 pictures [organized by style and genre]
– generate portraits

Edmond de Belamy [Caselles-Dupre, Fautrel, Vernier (2018)]

– trained on 15,000 portraits
– sold for $432.500
⇒ ML all marketing and sales

Jet portraits [de Oliveira, Paganini, Nachman (2017)]

– calorimeter or jet images
– reproduce valid jet images from training data
– organize them by QCD vs W-decay jets
⇒ Generative networks also useful



How to GAN

Adversarial training for LHC events

- training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
- discriminator classifier function $D(x)$ from minimizing [$D(x) = 1(\text{T}), 0(\text{G})$]

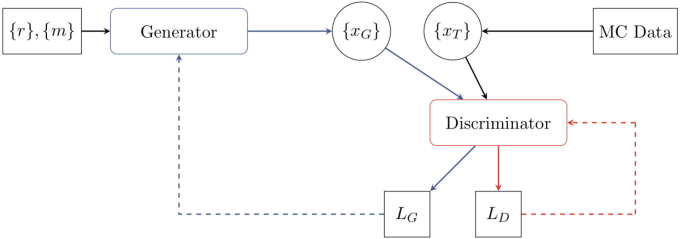
$$L_D = \big\langle -\log D(x) \big\rangle_{x_T} + \big\langle -\log(1 - D(x)) \big\rangle_{x_G}$$

- generator mapping $r \to x_G$ by minimizing [D needed]

$$L_G = \big\langle -\log D(x) \big\rangle_{x_G}$$

- equilibrium $D = 0.5 \;\Rightarrow\; L_D/2 = L_G = -\log 0.5$

$\Rightarrow$ statistically independent copy of training events



How to GAN

Adversarial training for LHC events

– training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
– How to GAN LHC Events [Butter, TP, Winterhalder]
  train a GAN on LHC events
– How to GAN Away Detector Effects [Bellagente, Butter, Kasieczka, TP, Winterhalder]
  use conditional GAN to unfold LHC events
– How to GAN Event Subtraction [Butter, TP, Winterhalder]
  train GAN on two samples and generate difference
– How to GAN Event Unweighting [Backes, Butter, TP, Winterhalder]
  train GAN on weighted events, generate 'unweighted' events
– How to GAN Higher Jet Resolution [Heidelberg-Irvine]
  train GAN to improve jet image resolution
– GANplifying Event samples [Butter, Diefenbacher, Kasieczka, Nachman, TP]
  show how GAN beats training statistics

How to GAN

Adversarial training for LHC events

– training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
– How to GAN LHC Events [Butter, TP, Winterhalder]
  train a GAN on LHC events
– How to GAN Away Detector Effects [Bellagente, Butter, Kasieczka, TP, Winterhalder]
  use conditional GAN to unfold LHC events
– How to GAN Event Subtraction [Butter, TP, Winterhalder]
  train GAN on two samples and generate difference
– How to GAN Event Unweighting [Backes, Butter, TP, Winterhalder]
  train GAN on weighted events, generate 'unweighted' events
– How to GAN Higher Jet Resolution [Heidelberg-Irvine]
  train GAN to improve jet image resolution
– GANplifying Event samples [Butter, Diefenbacher, Kasieczka, Nachman, TP]
  show how GAN beats training statistics

⇒ Two big LHC questions

  How precise are GANned distributions?
  What is their uncertainty?

GANplification

Gain beyond training data [Butter, Diefenbacher, Kasieczka, Nachman, TP]

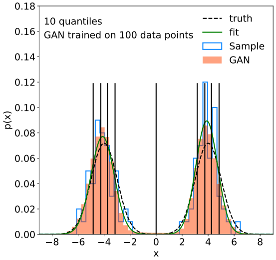
– true function known
compare sampling vs GAN vs fit

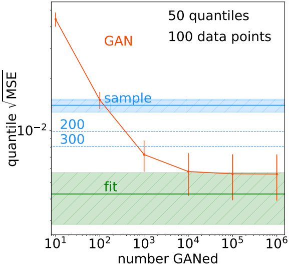
– quantiles with χ^2-values

– start with 100 sampled points
fit like 700 sampled points
GAN like 500 sampled points ...
... but requiring 10,000 GANned events

– interpolation and resolution the key [NNPDF]

$\Rightarrow$ Generative networks beyond training data





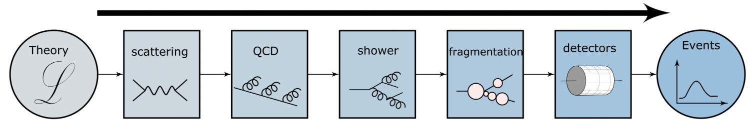
INN-generator

Challenging ML-event generators

– training from event samples
  no energy-momentum conservation
  no detector effects [sharper structures]

1- top-quark pairs $t\bar{t} \to 6$ jets
   resonance peaks for $t/\bar{t}$ and $W^\pm$

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets
   Z-peak, variable jet number, jet-jet topology



INN-generator

Challenging ML-event generators

– training from event samples
no energy-momentum conservation
no detector effects [sharper structures]

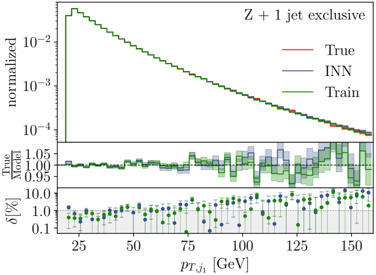
1- top-quark pairs $t\bar{t} \to 6$ jets
resonance peaks for $t/\bar{t}$ and $W^\pm$

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets
Z-peak, variable jet number, jet-jet topology

INN-generator [Butter, Heimel, TP, many students (soon)]

– map phase space $\leftrightarrow$ latent space
bijective mapping, Jacobian known
sample from Gaussian latent space

– training on 5.4M Z+jets events
goal: 1% precision relative to truth



INN-generator

Challenging ML-event generators

– training from event samples
  no energy-momentum conservation
  no detector effects [sharper structures]

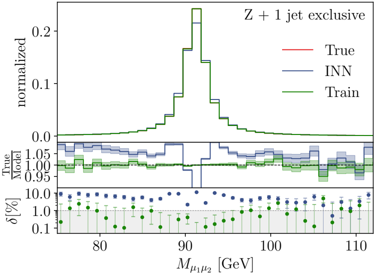
1- top-quark pairs $t\bar{t} \to 6$ jets
   resonance peaks for $t/\bar{t}$ and $W^\pm$

2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets
   Z-peak, variable jet number, jet-jet topology

INN-generator [Butter, Heimel, TP, many students (soon)]

– map phase space $\leftrightarrow$ latent space
  bijective mapping, Jacobian known
  sample from Gaussian latent space

– training on 5.4M Z+jets events
  goal: 1% precision relative to truth



INN-generator

Challenging ML-event generators

– training from event samples
no energy-momentum conservation
no detector effects [sharper structures]

1- top-quark pairs $t\bar{t} \to 6$ jets
resonance peaks for $t/\bar{t}$ and $W^\pm$

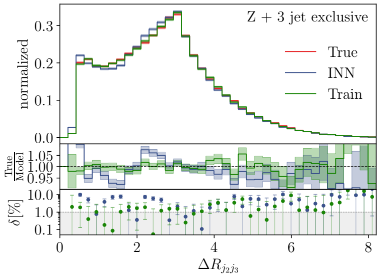
2- $Z_{\mu\mu} + \{1, 2, 3\}$ jets
Z-peak, variable jet number, jet-jet topology

INN-generator [Butter, Heimel, TP, many students (soon)]

– map phase space $\leftrightarrow$ latent space
bijective mapping, Jacobian known
sample from Gaussian latent space

– training on 5.4M Z+jets events
goal: 1% precision relative to truth

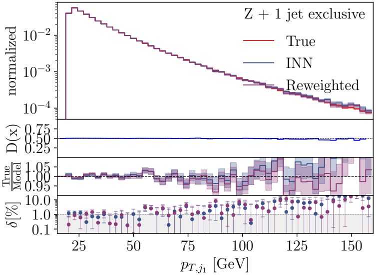
$\Rightarrow$ Precision promising, not yet perfect



Discriminator-generator network

GAN spirit — include discriminator

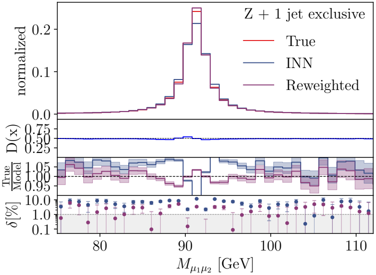
- discriminator: training vs generated events
  input $\{p_T, \eta, \phi, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$
  separate networks for jet multiplicities
- decent generator training $D \approx 0.5$
- add'l event weight $w_D = D/(1-D)$



Discriminator-generator network

GAN spirit — include discriminator

- discriminator: training vs generated events
  input $\{p_T, \eta, \phi, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$
  separate networks for jet multiplicities
- decent generator training $D \approx 0.5$
- add'l event weight $w_D = D/(1-D)$

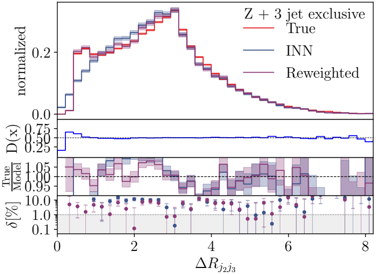


Discriminator-generator network

GAN spirit — include discriminator

- discriminator: training vs generated events
  input $\{p_T, \eta, \phi, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$
  separate networks for jet multiplicities
- decent generator training $D \approx 0.5$
- add'l event weight $w_D = D/(1-D)$

$\Rightarrow$ Precision possible



Discriminator-generator network

GAN spirit — include discriminator

– discriminator: training vs generated events
  input $\{p_T, \eta, \phi, M_{\mu\mu}, \Delta R\}$
  output $D = 0(\text{generator}), 1(\text{truth})$
  separate networks for jet multiplicities

– decent generator training $D \approx 0.5$

– add'l event weight $w_D = D/(1 - D)$

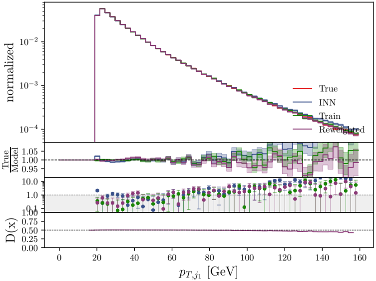
$\Rightarrow$ Precision possible

Joint training

– combine discriminator-generator training
  all information in generator
  unweighted events

– GAN-like training unstable
  Nash equilibrium hard for INN-generator

– coupling through weights
  move 'truth' to emphasize disagreement



Discriminator-generator network

GAN spirit — include discriminator

– discriminator: training vs generated events
  input $\{p_T, \eta, \phi, M_{\mu\mu}, \Delta R\}$
  output $D = 0$(generator), 1(truth)
  separate networks for jet multiplicities

– decent generator training $D \approx 0.5$

– add'l event weight $w_D = D/(1-D)$

$\Rightarrow$ Precision possible

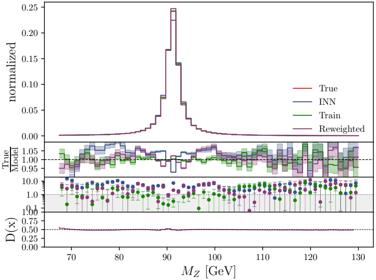
Joint training

– combine discriminator-generator training
  all information in generator
  unweighted events

– GAN-like training unstable
  Nash equilibrium hard for INN-generator

– coupling through weights
  move 'truth' to emphasize disagreement

$\Rightarrow$ DiscFlow getting there



One number not a prediction

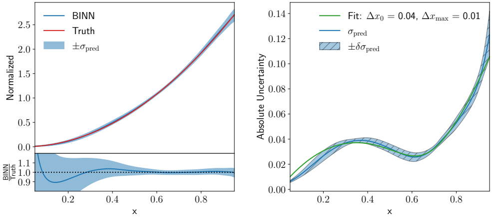
Bayesian INN [Bellagente, Haußmann, Luchmann, TP]

– learn network weight distributions
sample for network output with error bar
possible for regression, classification, density estimate

– generate events with error bars
learn density and uncertainty maps over phase space

– 2D toy models: wedge ramp, kicker ramp, Gaussian ring

$\Rightarrow$ Side remark: see how INN learns



One number not a prediction

Bayesian INN [Bellagente, Haußmann, Luchmann, TP]

– learn network weight distributions
  sample for network output with error bar
  possible for regression, classification, density estimate

– generate events with error bars
  learn density and uncertainty maps over phase space

– 2D toy models: wedge ramp, kicker ramp, Gaussian ring

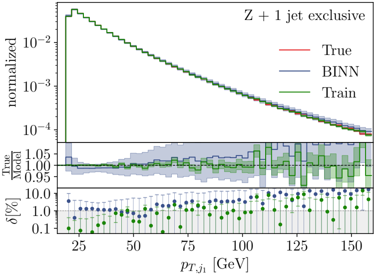
⇒ Side remark: see how INN learns

Back to Z+jets

– kinematic distributions with error bars

– limitation: training statistics

– error estimate conservative



One number not a prediction

Bayesian INN [Bellagente, Haußmann, Luchmann, TP]

– learn network weight distributions
  sample for network output with error bar
  possible for regression, classification, density estimate

– generate events with error bars
  learn density and uncertainty maps over phase space

– 2D toy models: wedge ramp, kicker ramp, Gaussian ring

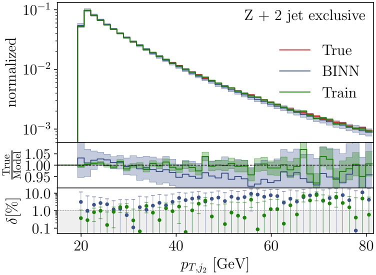
⇒ Side remark: see how INN learns

Back to Z+jets

– kinematic distributions with error bars

– limitation: training statistics

– error estimate conservative



One number not a prediction

Bayesian INN [Bellagente, Haußmann, Luchmann, TP]

– learn network weight distributions
sample for network output with error bar
possible for regression, classification, density estimate

– generate events with error bars
learn density and uncertainty maps over phase space

– 2D toy models: wedge ramp, kicker ramp, Gaussian ring

⇒ Side remark: see how INN learns

Back to Z+jets

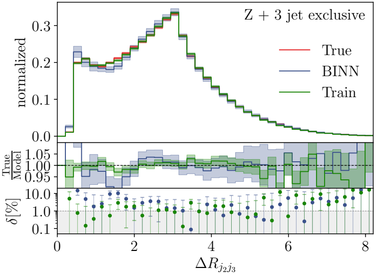
– kinematic distributions with error bars

– limitation: training statistics

– error estimate conservative

⇒ Precision and error bars, check!!

⇒ For systematic uncertainties, ask...

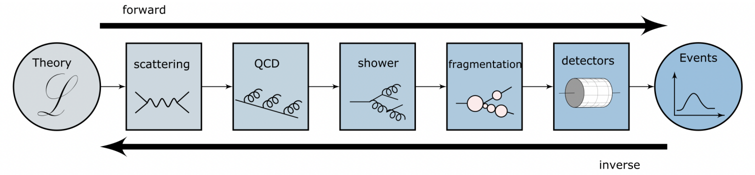


INNverting event simulations

Inverting LHC simulations

– unfolding QCD-shower to hard parton standard [jet algorithm]
unfolding detector common
unfolding top-quark decays useful
matrix element method complete unfolding

⇒ systematic approach through generative network



INNverting event simulations

Inverting LHC simulations

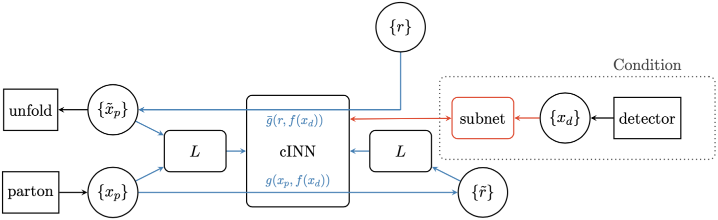
- unfolding QCD-shower to hard parton standard [jet algorithm]
  unfolding detector common
  unfolding top-quark decays useful
  matrix element method complete unfolding

⇒ systematic approach through generative network

Conditional INN

- standard INN/cINN setup: parton-level events from $\{r\}$
- maximum likelihood loss

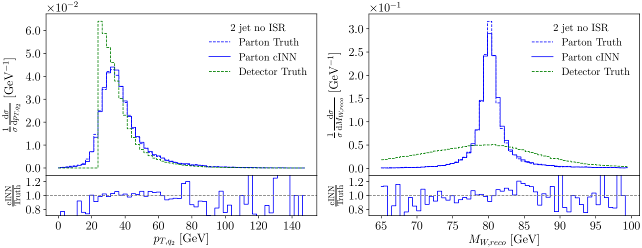
$$L = -\left\langle \log p(\theta|x_p, x_d) \right\rangle_{x_p, x_d}$$
$$= -\left\langle \log p(g(x_p, x_d)) + \log \left| \frac{\partial g(x_p, x_d)}{\partial x_p} \right| \right\rangle_{x_p, x_d} - \log p(\theta) + \text{const.}$$



INNverting event simulations

Hard process $q\bar{q} \to ZW \to (\ell\ell)\ (jj)$

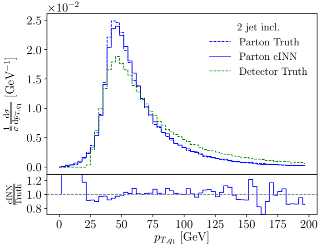
- stochastic inverse problem
  model assumption like in forward direction
- invert detector effects

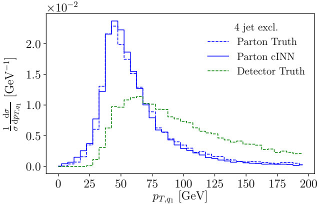


INNverting event simulations

Hard process $q\bar{q} \to ZW \to (\ell\ell)\ (jj)$

- stochastic inverse problem
  model assumption like in forward direction
- invert detector effects
- invert QCD jet radiation [matrix element vs parton shower]





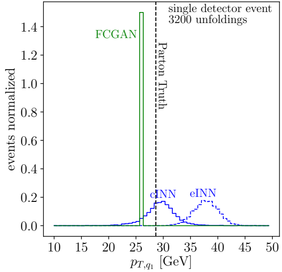
INNverting event simulations

Hard process $q\bar{q} \to ZW \to (\ell\ell)\,(jj)$

- stochastic inverse problem
  model assumption like in forward direction
- invert detector effects
- invert QCD jet radiation [matrix element vs parton shower]

Proper statistical inversion

- distribution
  single detector event
  3200 unfoldings to partonic phase space



INNverting event simulations

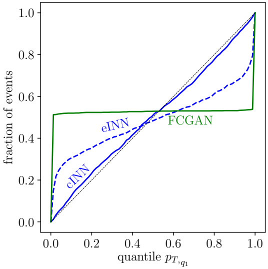
Hard process $q\bar{q} \to ZW \to (\ell\ell)\ (jj)$

– stochastic inverse problem
  model assumption like in forward direction
– invert detector effects
– invert QCD jet radiation [matrix element vs parton shower]

Proper statistical inversion

– distribution
  single detector event
  3200 unfoldings to partonic phase space
– calibration
  1500 detector-parton event pairs
  60 unfoldings per pair, in quantiles
  truth within given quantile for fraction of pairs

$\Rightarrow$ Probability distribution in parton phase space!



Machine learning for LHC theory

Machine learning for the LHC

– Classification/regression standard

  uncertainties?
  symmetries?
  experimental realities?

– GANs the cool kid

  generator producing best events
  discriminator checking generator
  limited in precision and uncertainty control

– INNs my theory hope

  flow networks for control and precision
  Bayesian for error bars
  condition for inversion

– Progress means young people playing with ideas