Invertible Networks
or
Partons to Detector and Back Again

ML4jets

Anja Butter

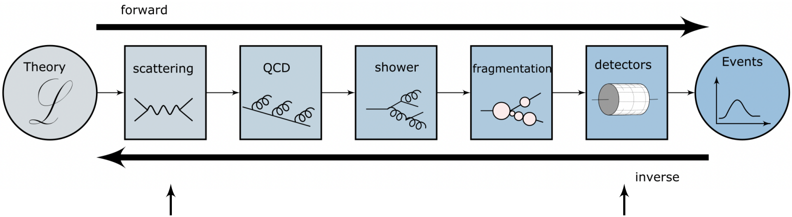
ITP, Universität Heidelberg

arXiv:2006.06685

with M. Bellagente, G. Kasieczka, T. Plehn, A. Rousselot,

R. Winterhalder, L. Ardizzone, U. Köthe

Can we invert the simulation chain?

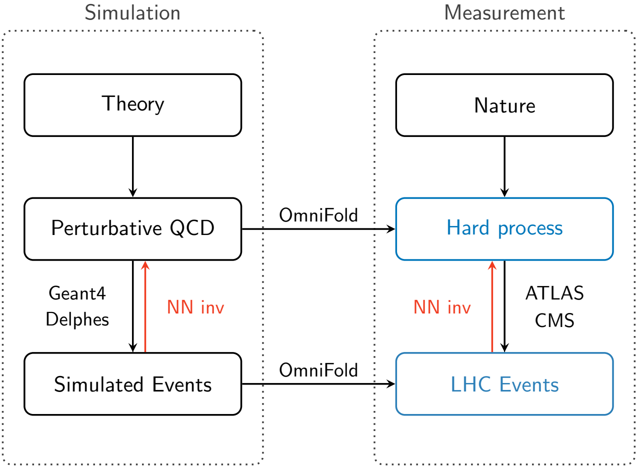


What we
want to know

What we
measure or simulate

wish list:
- ☐ multi-dimensional
- ☐ bin independent
- ☐ statistically well defined

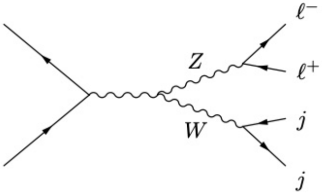
Inverting the simulation chain



- 1D GAN [1806.00433] Datta et al., FCGAN [1912.00477] Bellagente et al.
- cINN (this talk) [2006.06685] Bellagente et al., VAE (Jessica's talk) [2101.08944] Howard et al.
- Orthogonal approach: OmniFold [1911.09107] Andreassen et al.

Setup

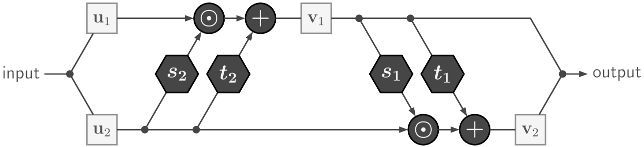
$$pp \to ZW^{\pm} \to (\ell^-\ell^+)\,(jj)$$



- event selection:
  - detector level: 2 jets & 2 leptons
  - $p_{T,j} > 25$ GeV $\&|\eta_j| < 2.5$ GeV.
- for now: # dof at parton = # dof at detector level
- spoiler: algorithm will be suitable for any level of unfolding and variable dof

Invertible networks

$$\left(x_{part}\right) \xleftrightarrow[\leftarrow \text{ unfolding}:\bar{g}]{\text{PYTHIA,DELPHES}:g\rightarrow} \left(x_{det}\right)$$



[1808.04730] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner,

E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, U. Köthe

+ Bijective mapping

+ Tractable Jacobian (see Lyntons' talk)

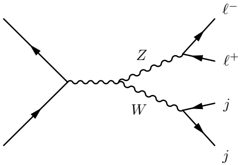
+ Fast evaluation in both directions

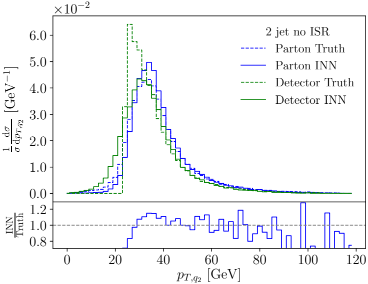
+ Arbitrary networks s and t

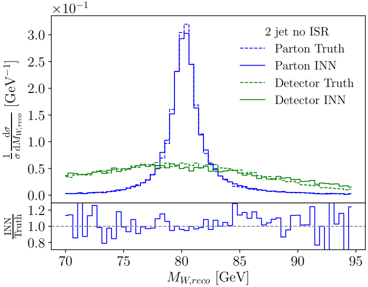
+ Affine coupling blocs yield smoother distributions

Inverting detector effects

- $pp \to ZW \to (ll)(jj)$
- Train: parton $\to$ detector
- Evaluate: parton $\leftarrow$ detector







multi-dimensional ✓ bin independent ✓ statistically well defined ?

Including the probabilistic aspect

$$\left(x_{part}\right) \xleftrightarrow[\leftarrow \text{inversion:}\bar{g}]{\text{PYTHIA,DELPHES:}g\rightarrow} \left(x_{det}\right)$$

with $\mathcal{L} = \mathcal{L}_{part} + \mathcal{L}_{det}$

Including the probabilistic aspect

$$\begin{pmatrix} x_{part} \\ r_{part} \end{pmatrix} \xleftrightarrow[\leftarrow \text{inversion:}\bar{g}]{\text{Pythia,Delphes:}g\rightarrow} \begin{pmatrix} x_{det} \\ r_{det} \end{pmatrix}$$

with $\mathcal{L} = \mathcal{L}_{part} + \mathcal{L}_{det} + \mathcal{L}_r$

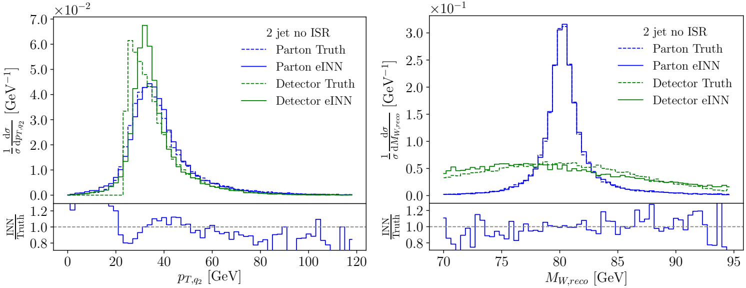
Exact choice of $\mathcal{L}_{part}, \mathcal{L}_{det}, \mathcal{L}_r$ determines (un)supervised training

Including the probabilistic aspect

$$\begin{pmatrix} x_{part} \\ r_{part} \end{pmatrix} \xleftrightarrow[\leftarrow \text{ inversion:} \bar{g}]{\textsc{Pythia,Delphes:}g \rightarrow} \begin{pmatrix} x_{det} \\ r_{det} \end{pmatrix}$$

with $\mathcal{L} = \mathcal{L}_{part} + \mathcal{L}_{det} + \mathcal{L}_r$

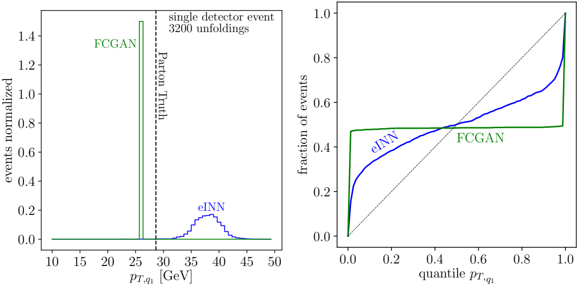
Exact choice of $\mathcal{L}_{part}, \mathcal{L}_{det}, \mathcal{L}_r$ determines (un)supervised training
$\mathcal{L}$: MSE to link x_{part} and x_{det}, MMD for correct shapes



Calibration of stochastic nature

$$\begin{pmatrix} x_p \\ r_p \end{pmatrix} \xleftrightarrow[\leftarrow \text{ inversion:}\bar{g}]{\text{PYTHIA,DELPHES:}g\rightarrow} \begin{pmatrix} x_d \\ r_d \end{pmatrix}$$

Sample r_d for fixed detector event
How often is Truth included in distribution quantile?



- Problem: arbitrary balance of many loss functions

A complex loss functions

$$\mathcal{L} = \mathcal{L}_{part} + \mathcal{L}_{det} + \mathcal{L}_r$$

$$\begin{aligned} \mathcal{L}_{part} &= \lambda_1 \, ||x_{part} - \bar{g}_{part}(x_{det}, r_{det})|| \\ &\quad + \lambda_2 \, \text{MMD}(x_{part}, \bar{g}_{part}(x_{det}, r_{det}) \end{aligned}$$

$$\begin{aligned} \mathcal{L}_{det} &= \lambda_3 \, ||x_{det} - g_{det}(x_{part}, r_{part})|| \\ &\quad + \lambda_4 \, \text{MMD}(x_{det}, g_{det}(x_{part}, r_{part}) \end{aligned}$$

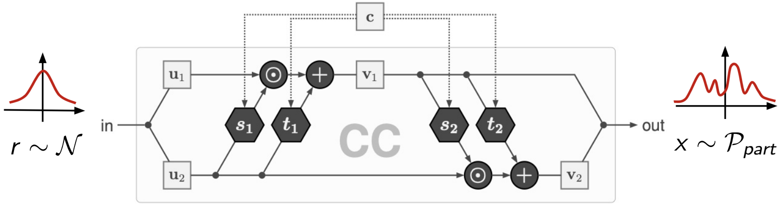
$$\begin{aligned} \mathcal{L}_r &= \lambda_5 \, \text{MMD}(\bar{g}_r(x_{det}, r_{det}), \mathcal{N}) \\ &\quad + \lambda_6 \, \text{MMD}(g_r(x_{part}, r_{part}), \mathcal{N}) \end{aligned}$$

$\rightarrow$ Result depends on careful tuning of $\lambda_1, \lambda_2, \lambda_3, ...$

Rephrasing the problem

Given detector level information [$\rightarrow$ condition c]
$\rightarrow$ What is the probability density at parton level?

$\rightarrow$ Conditional INN



Training the network

$$x_p \xleftrightarrow[\leftarrow \text{ unfolding: } \bar{g}(r, f(x_d))]{g(x_p, f(x_d)) \rightarrow} r$$

$\rightarrow$ Training: Maximize posterior over model parameters

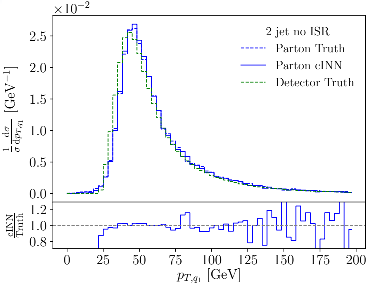
$$\begin{aligned}
L &= -\left\langle \log p(\theta | x_p, x_d) \right\rangle_{x_p \sim P_p, x_d \sim P_d} \\
&= -\left\langle \log p(x_p | \theta, x_d) \right\rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta) + \text{const.} \quad \leftarrow \text{Bayes} \\
&= -\left\langle \log p(g(x_p, x_d)) + \log \left| \frac{\partial g(x_p, x_d)}{\partial x_p} \right| \right\rangle - \log p(\theta) \leftarrow \text{change of var} \\
&= \left\langle 0.5 ||g(x_p, f(x_d))||_2^2 - \log \left| \frac{\partial g(x_p, x_d)}{\partial x_p} \right| \right\rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta)
\end{aligned}$$

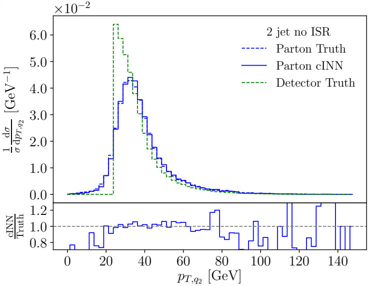
Gaussian latent variable $(g(x_p, x_d)) \rightarrow r$

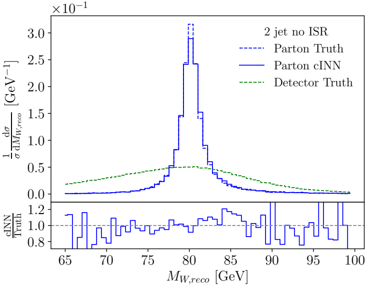
Jacobian

Regularization $||\theta||^2$

Cross check distributions



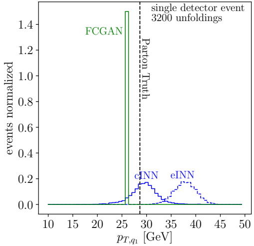


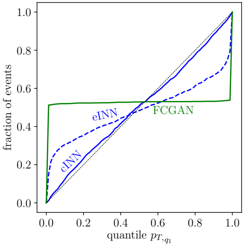


cINN result for calibration [2006.06685]

$$x_p \xleftrightarrow[\leftarrow \text{ unfoldiing: } \bar{g}(r, f(x_d))]{g(x_p, f(x_d)) \rightarrow} r$$

Minimizing $L = \left\langle 0.5 ||\bar{g}(x_p, f(x_d)))||_2^2 - \log|J| \right\rangle_{x_p \sim P_p, x_d \sim P_d} - \log p(\theta)$



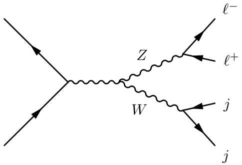


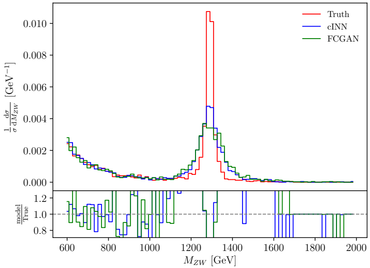
multi-dimensional ✓ bin independent ✓ statistically well defined ✓

Stability under new physics?

Insert W' resonance

- Train: SM events
- Test: 10% events with W' in s-channel





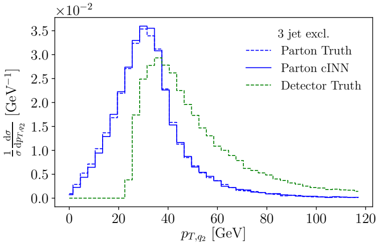
Inverting the full event I

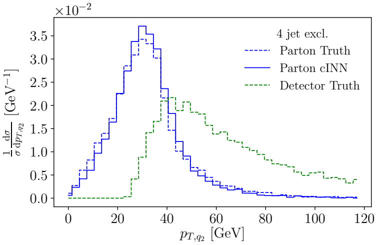
- $pp > WZ > q\bar{q}l^+l^- +$ ISR

$\rightarrow$ ISR leads to large fraction of 2/3/4 jet events

$\rightarrow$ # dof parton $\neq$ # dof detector

- Train and test on exclusive 2/3/4 jet channels

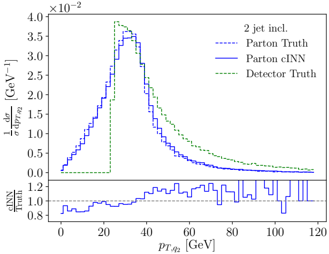




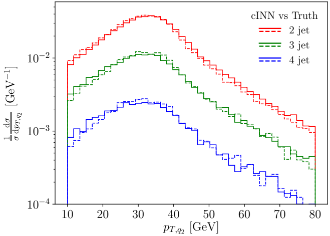
Inverting the full event II

$pp > WZ > q\bar{q}l^+l^- + \text{ISR}$

Train on inclusive dataset



Evaluate
exclusive 2/3/4 jet channels



Summary

- Unfold high-dim. detector level distributions with cGANs and INN
- Inversion is stable under insertion of new data
- cINN guaranties correct calibration
- Simultaneous unfolding of different exclusive channels