Returning CP-observables to the frames they belong

Rahool Barman
Oklahoma State University

With
Jona Ackerschott, Dorival Gonçalves, Theo Heimel, Tilman Plehn

Based on
arXiv: 2308.00027

HEP Seminar
OSU

October 5, 2023
• Conventional LHC analysis involves comparing measured data with MC events simulated under NP hypothesis.

• Reconstructed LHC events present a convoluted version of the true underlying physics.

• Forward simulation chain can be highly resource intensive.

Invert simulation chain → apply on measured data → reconstruct parton-level

→ compare new physics hypotheses at the parton-level.
• Bin-by-bin unfolding

• Correct the information in each bin using correction factor $C_i$ computed from MC data.

$C_i = \frac{N_{\text{truth},i}}{N_{\text{reco},i}}$

Unfolded distribution: $x_{p,i} = x_{d,i} \times C_i$

✓ No assumptions on the shape of the distributions.

✓ Bin correlations not taken into account.

✓ Highly sensitive to MC statistics.
• Matrix inversion
  • Build response matrix $R \rightarrow$ each cell $\{i, j\}$ represents the fraction of events which have a true value in bin $i$ but get reconstructed in bin $j$.

Response matrix: $R_{ij} = \frac{N_{\text{truth},i}}{N_{\text{reco},j}}$

Unfolded events in bin $i$: $x_{p,i} = \sum_j x_{d,j} \times R$

✓ No assumptions on the shape of the distributions.
✓ Noise amplification.
✓ Limited by statistics and dimensionality.
Unfolding

• Iterative unfolding
  • Build response matrix $R_{ij}$.
  • Given a true distribution, use $R_{ij}$ to predict the reconstructed distribution.
  • Compare it with observed data to compute correction factors.
  • Correction factors are applied to the initial $R_{ij}$.
  • Iterate the process until the difference reaches below a threshold.

✓ Bin-dependent unfolding.
✓ Correlations among observables not considered.
Unfolding

Possible with machine learning based generative models.

- Generative Adversarial Networks (GAN)
- Normalizing Flows (NF)
- Variational Auto Encoders (VAE)

[Andreassen, Komiske, Metodiev, Nachman, Thaler (2020)]
[Bellagente, Butter, Kasiezka, Plehn, Winterhalder (2020)]
[Bellagente, Butter, Kasiezka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]
[Komiske, McCormack, Nachman (2021)]
In GANs, the generator and discriminator network competes against each other.

[Bellagente, Butter, Kasieczka, Plehn, Winterhalder(2019)]
[Butter, Plehn, Winterhalder(2019)]
In GANs, the generator and discriminator network competes against each other.

[Butter, Plehn, Winterhalder (2019)]
[Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2019)]
In GANs, the generator and discriminator network competes against each other.
In GANs, the generator and discriminator network competes against each other.

\[ D(x_p) \rightarrow 1, \quad D(x_G) \rightarrow 0 \]
In GANs, the generator and discriminator network competes against each other.

\[ L_D = \langle -\log D(x) \rangle_{x \sim P_p} + \langle -\log(1 - D(x)) \rangle_{x \sim P_G} \]

\[ D(x_p) \to 1, \quad D(x_G) \to 0 \]
In GANs, the generator and discriminator network competes against each other.

\[ D(x_p) \rightarrow 1, \quad D(x_G) \rightarrow 0 \]

\[ L_D = \langle -\log D(x) \rangle_{x \sim P_p} + \langle -\log(1 - D(x)) \rangle_{x \sim P_G} \]

\[ L_G = \langle -\log D(x) \rangle_{x \sim P_G} \]

- Discriminator works to distinguish generated data \( \{x_G\} \) from truth data \( \{x_p\} \). \([D(x_p) \rightarrow 1, D(x_G) \rightarrow 0]\)

- Generator works to fool the discriminator such that \( D(x_G) \rightarrow 1 \).
Naive GAN unfolding

\[ pp \rightarrow ZW \rightarrow (Z \rightarrow \ell^+ \ell^-)(W \rightarrow jj) \]

Training data

\[ p_{T,j} > 25 \text{ GeV}, |\eta_j| < 2.5 \]

2 \( \ell \) + 2 exclusive \( j \)
@ detector level

Parton-level

Detector-level

Figures taken from Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2019)
In GANs, the generator and discriminator network competes against each other.

[Image adopted from Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2019)]

[Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2019)]
[Butter, Plehn, Winterhalder (2019)]
Limitations

- Cannot exploit the pairing information between parton and detector level $\rightarrow$ training does not explore event-by-event matching.

- Fails if training and test data to not statistically similar.

$pp \rightarrow ZW \rightarrow (Z \rightarrow \ell^{+}\ell^{-})(W \rightarrow jj)$

- Training data
  - Parton-level $\rightarrow$ Detector-level
  - $p_{T,j} > 25 \text{ GeV}, |\eta_{j}| < 2.5$

- Test data
  - $2 \ell + 2 j$ exclusive @ detector level
  - $30 \text{ GeV} < p_{T,j_1} < 60 \text{ GeV}, 30 \text{ GeV} < p_{T,j_2} < 50 \text{ GeV}$
  - ($\sim 38 \%$ of events)

Figure taken from Bellagente, Butter, Kasieczka, Plehn, Winterhalder(2019)
Limitations

- Cannot exploit the pairing information between parton and detector level → training does not explore event-by-event matching.

- Fails if training and test data are not statistically similar.

FCGAN

[Image adopted from Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2019)]
FCGAN

\[ L_{D}^{FC} = \langle -\log D(x, y) \rangle_{x \sim P_{p}, y \sim P_{d}} + \langle -\log(1 - D(x, y)) \rangle_{x \sim P_{G}, y \sim P_{d}} \]

\[ L_{G} = \langle -\log D(x, y) \rangle_{x \sim P_{G}, y \sim P_{d}} \]

- Event-by-event matching \( \rightarrow \) exploit the pairing information between parton and detector level.

- Trained network can be applied to statistically different regions of phase space.

Test data

2 \( \ell^{+} + 2 j \) exclusive @ detector level

\[ 30 \text{ GeV} < p_{T,j_{1}} < 60 \text{ GeV}, \ 30 \text{ GeV} < p_{T,j_{2}} < 50 \text{ GeV}, \]

( \( \sim 14\% \) of events)

- Unfolding fails with harsher cuts.
- Challenges with invariant mass peak generation since MMD is not conditional.
- Dimensionality limitations.

[Image adopted from Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2019)]
Variational Autoencoders

Encoder ($E$)

Input $x$

Latent Space $z(\sigma, \mu)$

Decoder ($D$)

Output $y$

$$L = \left\| y - D(E(x)) \right\|^2 + \eta KL(p(z|x) \| q(Z))$$

Reconstruction Loss

KL divergence term

[Ianazi, Sato, Ambrozewicz, Blin, Melnitchouk, Battaglieri, Liu, Li (2021)]
Variational Auto Encoders

→ Regress detector response function starting from a generator-level jet

[Touranakou, Chernyavskaya, Duarte, Gunopulos, Kansal, Orzari, Pierini, Tomei, Vlimant (2022)]
[Otten, Caron, Swart, Beekveld, Hendriks, Leeuwen, Podareanu, Austri, Verheyen (2019)]
Variational Auto Encoders

→ Regress detector response function starting from a generator-level jet

[Touranakou, Chernyavskaya, Duarte, Gunopulos, Kansal, Orzari, Pierini, Tomei, Vlimant (2022)]

Process: $pp \rightarrow WW \rightarrow (W \rightarrow jj)(W \rightarrow jj)$

Jet constituents: $p_T > 250$ MeV, $|\eta| < 3.2$

Jets (anti-$k_T$ with $\Delta R = 0.5$): $p_T > 200$ GeV, $|\eta| < 2.5$

Input-target jet matched by minimizing $\Delta R$
Variational Auto Encoders

Process: \( pp \rightarrow WW \rightarrow (W \rightarrow jj)(W \rightarrow jj) \)

Loss \( L \propto \beta D_{KL}^i + (1 - \beta)(L_R^i + \alpha_m (m_{jet}^i - \tilde{m}_{jet}^i)^2 + \alpha_{p_T} (p_T^i - \tilde{p}_T^i)^2) \)

Figures taken from Touranakou, Chernyavskaya, Duarte, Gunopulos, Kansal, Orzari, Pierini, Tomei, Vlimant (2022)

Good agreement between reco and predicted distributions, but jet substructure quantities not well reproduced.
Variational Auto Encoders

→ Map detector data to the parton level phase space

Detector data \( x_d \) → Encoder \( E \) → Latent Space \( z(\sigma, \mu) \) → Decoder \( D \) → Parton Truth \( x'_p \)

- The Encoder maps the input detector data \( d \) to a more tractable latent space \( z = E(d) \) while preserving the essential features.

- The decoder maps \( z \) to the parton level \( p' = D(z) = E(l(d)) \).

[Otten, Caron, Swart, Beekveld, Hendriks, Leeuwen, Podareanu, Austri, Verheyen (2019)]
**FCGAN vs VAE**

- **Ease of conditioning external information** → VAEs are more challenging.

- **Likelihood estimation** → VAEs perform approximate estimation, FCGANs do not.

- **Training dynamics** → Sharp data (FCGAN) vs blurrier data (VAE), at the cost of training stability.

- **Latent space mapping** → VAEs typically map to a gaussian latent space.
  - Could also prove useful to learn the inherent relationship and correlation among input data.
  - The gaussian latent space, may not always be the most appropriate choice to map input data.

- **Exact likelihood estimation** not possible in either and invertibility can be ambiguous.
Normalizing flows

- Exact likelihood estimation
- **Invertibility**:
  - NF is capable of bi-directional mapping w/o information loss.
  - VAEs not strictly invertible due to stochasticity of the latent space.
  - FCGANs focus on generation, and invertibility is not strictly defined.
- **Flexibility**:
  - NF models a complex distribution to a simple distribution using a series of invertible transformations → models intricate distributions without making strict assumptions.
  - VAEs assume a Gaussian latent space → may not always capture the complexity of the distributions.
  - GANs focus on generating data that matches the target distribution → no explicit latent mapping and less statistical robustness.
Normalizing flows

• Series of bijective layers that transform complex ($Y$) to simple probability distributions ($Z$).

• Learns both directions of the mapping in parallel $\rightarrow$ bijectivity encoded in the same network.

• Building blocks $\rightarrow$ Invertible coupling layers. [Dinh, Krueger, Bengio (2016), Dinh, Sohl-Dickstein, Bengio (2016)]

[Image adapted from Nguyen, Ardizzone, Kothe (2019) and talk by A. Butter at Pheno-2022]
Normalizing flows

In the coupling layers, the coupling functions $s_2$ and $t_2$ take $x_2$ as input, and scale/translate $x_1$.

Forward pass:

$$y_1 = x_1 \odot e^{S_2(x_2)} + t_2(x_2)$$
$$y_2 = x_2 \odot e^{S_1(y_1)} + t_1(y_1)$$

Fully invertible coupling layer $\rightarrow [x_1, x_2]$ can be reconstructed given $[y_1, y_2]$.

Inverse transformations:

$$x_1 = (y_1 - t_2(x_2)) \odot e^{-S_2(x_2)}$$
$$x_2 = (y_2 - t_1(y_1)) \odot e^{-S_2(y_1)}$$
Normalizing flows

Forward pass:
\[ y_1 = x_1 \odot e^{s_2(x_2)} + t_2(x_2) \]
\[ y_2 = x_2 \odot e^{s_1(y_1)} + t_1(y_1) \]

For a coupling block transformation \( f(x) \sim y \)

tractable Jacobian \( J_f(x) : \frac{\partial f(x)}{\partial x} = \begin{bmatrix} e^{s_2(x_2)} & \text{finite} \\ 0 & e^{s_1(y_1)} \end{bmatrix} \)

→ rule of change of variables
\[ p_Y(x_d) = p_Z(x_p) \times |\det(J_f(x_p))|^{-1} \]

→ ensures bijective transformations and exact likelihood estimation

Inverse transformations:
\[ x_1 = (y_1 - t_2(x_2)) \odot e^{-s_2(x_2)} \]
\[ x_2 = (y_2 - t_1(y_1)) \odot e^{-s_2(y_1)} \]
Normalizing flows

- Coupling layers stacked together → **Invertible Neural Network (INN)**.
  - Transforms the input distribution to a general distribution through a series of invertible steps.

- Tractable Jacobian for each coupling layer → **Input and output densities can be related**.

- Typically, DNNs suffer an inherent information loss in the forward direction, making the inverse mapping ambiguous → **Not an issue with INNs**.

✓ Sampling and density estimation

[Image adapted from Nguyen, Ardizzone, Kothe (2019) and talk by A. Butter at Pheno-2022]
Naive INN unfolding

Process: $pp \rightarrow ZW \rightarrow (Z \rightarrow \ell^+\ell^-)(W \rightarrow jj)$

Forward simulation: $g \rightarrow (X_{\text{parton}}) \leftarrow$ Unfolding: $g \rightarrow (X_{\text{detector}})$

Training data

Exactly 2 jets: $p_T > 25$ GeV, $|\eta| < 2.5$
SFOS lepton pair near $m_Z$

Parton-level Event-wise matching Detector-level
Naive INN unfolding

Forward simulation: $\tilde{g} \rightarrow (x_{\text{detector}})$

Unfolding: $g \leftarrow (x_{\text{parton}})$
Naive INN unfolding

Forward simulation: $g \rightarrow \bar{g}$

Unfolding: $g \leftarrow$
Naive INN unfolding

\[(x_{\text{parton}}) \rightarrow \text{Forward simulation: } \bar{g} \rightarrow g \rightarrow (x_{\text{detector}})\]

\[
\begin{align*}
\text{Gen Parton} & \quad \bar{g}(x_d) \quad \text{INN} \quad g(x_p) \\
\text{True Parton} & \quad x'_p \quad \text{Gen detector} \quad x'_d \\
\text{Detector} & \quad x_d 
\end{align*}
\]
Naive INN unfolding

Forward simulation: $\bar{g} \rightarrow (x_{\text{parton}}) \rightarrow (x_{\text{detector}})$

Unfolding: $g \rightarrow (x_{\text{detector}}) \rightarrow (x_{\text{parton}})$

Process: $pp \rightarrow ZW \rightarrow (Z \rightarrow \ell^+\ell^-)(W \rightarrow jj)$

[Figure adopted from Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]

Loss: $L_{\text{MSE}}(x_p) + L_{\text{MSE}}(x_d) + L_{\text{MMD}}$
Naive INN unfolding

Process: $pp \rightarrow ZW \rightarrow (Z \rightarrow \ell^+\ell^-)(W \rightarrow jj)$

- INN generated $p_{T,j_1}$ and invariant masses (MMD) closely match with parton truth.
- Differences between generated and parton truth deviate in the soft $p_{T,j_2}$ region and tails.
- Typically inefficient in the inversion of features not included in event parametrization.
- Dimensionality limitations.

[Figure taken from Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]
Noise-extended INN

\[
\left( X_{\text{parton}} \right)_{r_p} \xrightarrow{\text{Forward simulation: } \bar{g} \rightarrow} \left( X_{\text{detector}} \right)_{r_d} \xleftarrow{\text{Unfolding: } g}
\]

- Allows mapping between unequal degrees of freedom at the parton and detector level.
- Random number vector extended on each side to account for unobservable degrees of freedom.
- MMD terms included for each observable and gaussian input → improves unfolding in the low and high \( p_T \) regions.

Process: \( pp \rightarrow ZW \rightarrow (Z \rightarrow \ell^+\ell^-)(W \rightarrow jj) \)

[Figure taken from Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]
Noise-extended INN: Limitations and Challenges

• Inclusive detector level information requires using large number of random variables.

• Requires careful calibration between MMD and MSE loss terms.

• Calibration of weights associated to different loss terms.

• Combination of several loss terms pose training challenges.

→ Upgrade to conditional INN
Conditional INN

- Generate probability distributions at the parton-level, given detector-level events $x_{\text{detector}}$
  - Conditional on $x_{\text{detector}}$
  - $x_{\text{parton}}$ mapped with $r$

Target phase space for unfolding can be chosen flexibly to include:
- ✔ QCD jet radiation
- ✔ Particle decays
Unfolding semileptonic $t\bar{t}h$ events

$pp \rightarrow t\bar{t}h \rightarrow (t \rightarrow \ell \nu b)(\bar{t} \rightarrow jj\bar{b})(h \rightarrow \gamma\gamma)$

<table>
<thead>
<tr>
<th>Parton-level:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1\ell + 2b + 2\gamma + \nu + 2j$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Detector-level:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1\ell + 2b + 2\gamma + \text{MET} + \leq 6 \text{ jets inclusive}$</td>
</tr>
</tbody>
</table>

Acceptance cuts

$|\eta_b| < 4, \quad |\eta_j| < 5, \quad |\eta_\ell| < 4, \quad |\eta_\gamma| < 4$

$\quad p_{T,b} > 25 \text{ GeV}, \quad p_{T,j} > 25 \text{ GeV}, \quad p_{T,\ell} > 15 \text{ GeV}, \quad p_{T,\gamma} > 15 \text{ GeV}$

Challenges:

★ Can the unfolding model correctly reconstruct the two hard jets at the parton level from a variable number of jets at the detector level?

★ How well can the dedicated BSM observables be reconstructed?

★ How model-dependent is the training?
Event parametrization

• Event information at the parton level can be parametrised through the 4-momentum of the final state particles \(\rightarrow\) may include redundant d.o.f.

• Reconstruction of sharp kinematic features like mass peaks can be challenging:
  ✓ Can be improved by adding targeted maximum mean discrepancy loss:
    ✓ Affects only the target distributions
    ✓ Avoids large model dependence
    ✗ Complications in training and performance limitations.

Alternative approach:
→ directly learn invariant mass features and important observable with appropriate phase-space parametrization.

→ may provide direct access to the most important BSM observables.
Conditional INN

- We use the Bayesian version of cINN \[\text{[Butter, Heimel, Hummerich, Krebs, Plehn, Rousselot, Vent (2021)]}\]
  - Stable network predictions
  - Allows the estimation of training-related uncertainties.

- Degrees of freedom:
  - Parton-level: \((t \to \ell \nu b) (\bar{t} \to jj\bar{b}) h\)
    
    22 d.o.f.

    A natural parametrization involving top mass:
    \[
    \{ m_t, \ell_T, \ell_t, \phi_t, m_W, \eta_W, \phi_W, \eta_{\ell_W}, \phi_{\ell_W} \}
    \]

- Alternatively, redefine the parton level parametrization including the important CP observables

Detector-level: 46 d.o.f.

\(1\ell + 2b + 2\gamma + \text{MET} + \leq 6 \) jets inclusive

\[
\bar{p}_{tt}, m_{t\ell}, |\vec{p}_{t\ell}|, \theta_{t\ell}^{CS}, \phi_{t\ell}^{CS}, m_{t\ell}, \\
\text{sign}(\Delta \phi_{t\ell}^{W}), m_W, |\bar{p}^{W}_{t\ell}|, \theta_{t\ell}^{W}, \phi_{t\ell}^{W}, |\bar{p}^{W}_{t\ell}| \\
\text{sign}(\Delta \phi_{d\ell}^{W}), m_{Wd}, |\bar{p}^{W}_{d\ell}|, \theta_{d\ell}^{W}, \Delta \phi_{d\ell}, |\bar{p}^{W}_{d\ell}|
\]
• New sources of CPV interactions can explain the matter-antimatter asymmetry in the universe.

• One such exciting scenario is CP violation in the Higgs sector $\sim$ mixing of CP-even and odd states $\rightarrow$ CP-mixed hypothesis is still allowed at the LHC.

• CPV in $hVV$ interactions is extensively tested at the LHC.

$\rightarrow$ desirable choice: $ht\bar{t}$
Direct probes at the LHC

\[ \mathcal{L} = -\frac{m_t}{v} \kappa_t h \bar{t}t (\cos \alpha + i \gamma_5 \sin \alpha) t \]

SM: \((\kappa_t, \alpha) = (1,0)\)

- \(pp \rightarrow h (\text{+ jets})\): indirect constraints.

- \(pp \rightarrow t\bar{t}h\) stands out as the viable direct probe:
  - Small rate at the LHC and complex topology.
  - Silver Lining: Observation at 5.2\(\sigma\) by ATLAS [2004.04545] and 6.6\(\sigma\) by CMS [2003.10866]

- Current limits: \(|\alpha| < 43^0\) (ATLAS) and \(|\alpha| < 55^0\) (CMS), at 95% CL.

Improved statistics @ HL-LHC paves the pathway for precision studies.

Importance matrix at the **non-linear level**

Sensitive only to non-linear new physics effects.

[RKB, Goncalves, Kling (2021)]
• Short lifetime for $t$ ($10^{-25}$ s) $\rightarrow$ Spin correlations can be traced back from their decay products.

• CP-odd observables constructed from antisymmetric tensor products

$$\epsilon(p_\tau, p_\tau, p_i, p_j) \sim \epsilon_{\mu\nu\rho\sigma} p_\mu^i p_\nu^i p_\rho^i p_\sigma^j :$$

$$\Delta \Phi_{ij} = \text{sgn} \left( \vec{p}_i \cdot (\vec{p}_i \times \vec{p}_j) \right) \arccos \left( \frac{\vec{p}_i \times \vec{p}_j}{|\vec{p}_i \times \vec{p}_j|} \right)$$

[Mileo, Kiers, Szynkman, Crane, Gegner (2016); Goncalves, Kong, Kim (2018); RKB, Goncalves, Kling (2021)]

Spin correlations scale with the spin analysing power $\beta_i$.

[Spin correlations scale with the spin analysing power] [Mileo, Kiers, Szynkman, Crane, Gegner (2016); Goncalves, Kong, Kim (2018); RKB, Goncalves, Kling (2021)]

• Kinematic reconstruction efficiency is limited at the detector level

Use Machine learning techniques to maximize the extraction of NP information from CP observables.
Likelihood inference

- Event likelihood ratio $r(x \mid \theta, \theta_{SM}) = p(x \mid \theta) / p(x \mid \theta_{SM})$ is intractable at the detector level.

\[ p(x \mid \theta) = \frac{1}{\sigma(\theta)} \frac{d^d \sigma(x \mid \theta)}{dx^d} \]

- Joint likelihood ratio $r(x, z \mid \theta, \theta_{SM})$ can be computed:

\[ r(x, z \mid \theta_1, \theta_0) \equiv \frac{p(x, z \mid \theta_1)}{p(x, z \mid \theta_0)} = \frac{d\sigma(z_p \mid \theta_0)\sigma(\theta_1)}{d\sigma(z_p \mid \theta_1)\sigma(\theta_0)} \]

- Uses $r(x, z \mid \theta, \theta_{SM})$ dependent loss functions:

\[ L[r(x \mid \theta_1, \theta_0)] \sim \frac{1}{N} \sum_{x_i, z_i} \left| r(x_i, z_i \mid \theta_1, \theta_0) - r(x_i \mid \theta_1, \theta_0) \right|^2 \]

The trained network is an estimator for $r(x \mid \theta, \theta_{SM})$

taken from [Brehmer, Cranner, Louppe, Pavez (2018)]
Dominant sensitivity from CP-even observables.

$|\alpha| < 23^0$ at 95% CL

Boost sensitivity through unfolding techniques.
Back to results from unfolding with cINN...

Challenges:

★ Can the unfolding model correctly reconstruct the two hard jets at the parton level from a variable number of jets at the detector level?

★ How well can the dedicated BSM observables be reconstructed?

★ How model-dependent is the training?
★ Unfolded distributions in good agreement with parton level truth despite added combinatorial ambiguity at the detector level.
Back to results from unfolding with cINN...

Challenges:

★ Can the unfolding model correctly reconstruct the two hard jets at the parton level from a variable number of jets at the detector level?

★ How well can the dedicated BSM observables be reconstructed?

★ How model-dependent is the training?
Reconstruction of dedicated observables

Parton level truth and unfolded SM for $\theta_{CS}$, $\Delta \phi_{t\bar{t}b}$ and $b_4$.

★ Unfolded distributions in close agreement with truth:
  ✓ Close agreement even for observables not included in event parametrization.
  ✓ Full phase space reconstruction.

★ Potential differences from the truth are covered by the uncertainty estimates of the Bayesian network.
Back to results from unfolding with cINN...

Challenges:

★ Can the unfolding model correctly reconstruct the two hard jets at the parton level from a variable number of jets at the detector level?

★ How well can the dedicated BSM observables be reconstructed?

★ How model-dependent is the training?
Model dependence

Unfolding SM events using networks trained on events with different amounts of CP-violation.

We train 3 networks on $\alpha = +\pi/4, -\pi/4$ and SM, respectively

Unfold SM dataset

★ Networks trained on $\alpha = \pi/4$ and $-\pi/4$ show only a slight bias towards broader $\theta_{CS}$ and flatter $b_4$ distributions.
★ $\sim 10 - 20\%$ bias $\rightarrow$ much smaller than the changes at parton truth from varying $\alpha$. 
Unfolding events with CP-violation using a network trained on SM events.

Unfold $\alpha = +\pi/4, -\pi/4$ and SM dataset

Again, the effect of bias is much smaller than the effect of $\alpha$ on the data.
Outlook

• Generative unfolding makes it possible to invert high-dimensional distributions and full phase-space reconstruction.

• The trained cINN behaves as an efficient kinematic reconstruction algorithm capable of tackling complex reconstruction challenges.

• The trained unfolding network was able to
  • extract various CP observables at the parton level with appropriate phase space parametrization.
  • resolve jet combinatorial ambiguity.
  • absolve any large model-dependence.

• While this study is clearly not the last word on this analysis technique, it presents a promising outlook for an experimental study, with a proper treatment of statistical limitations, continuum backgrounds, calibration, and iterative improvements of the unfolding network.
Thank you