Returning CP-observables to the frames they belong

Rahool Barman
Oklahoma State University

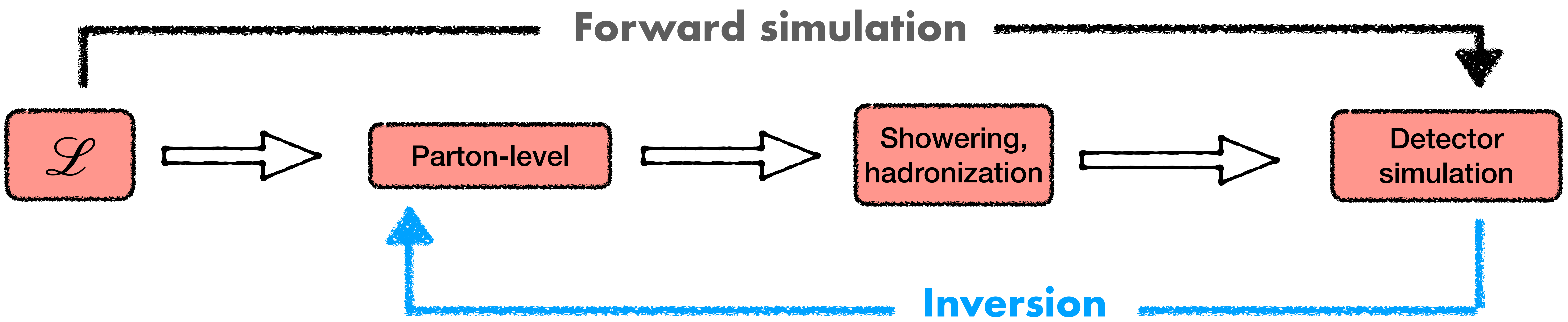
With
Jona Ackerschott, Dorival Gonçalves, Theo Heimel, Tilman Plehn

Based on
arXiv: 2308.00027

HEP Seminar
OSU

October 5, 2023

Unfolding



- 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

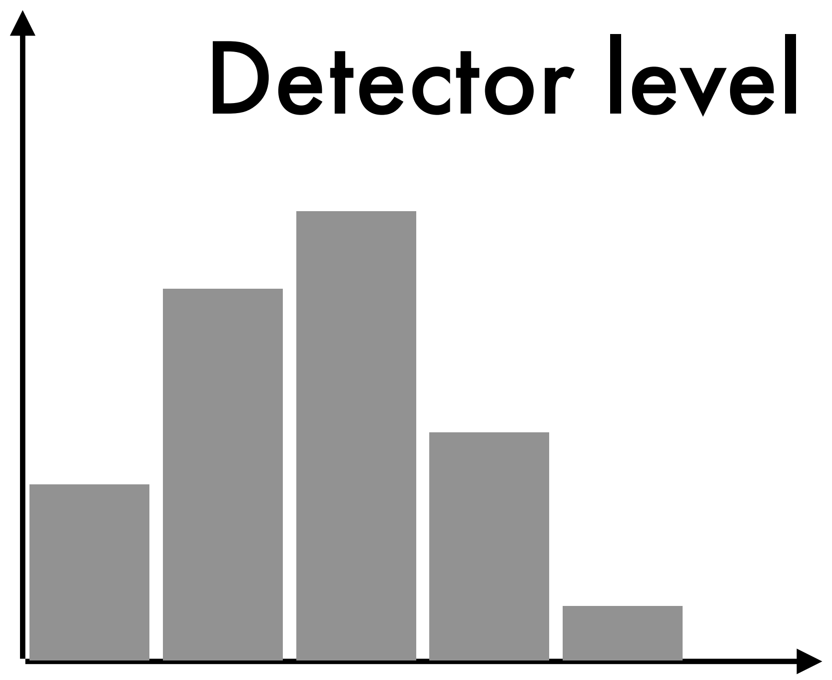
→ compare new physics hypotheses at the parton-level.

Unfolding

- Bin-by-bin unfolding
  - Correct the information in each bin using correction factor $\mathscr{C}_i$ computed from MC data.

$$\mathscr{C}_i = N_{\text{truth},i}/N_{\text{reco},i}$$

Unfolded distribution: $x_{p,i} = x_{d,i} \times \mathscr{C}_i$



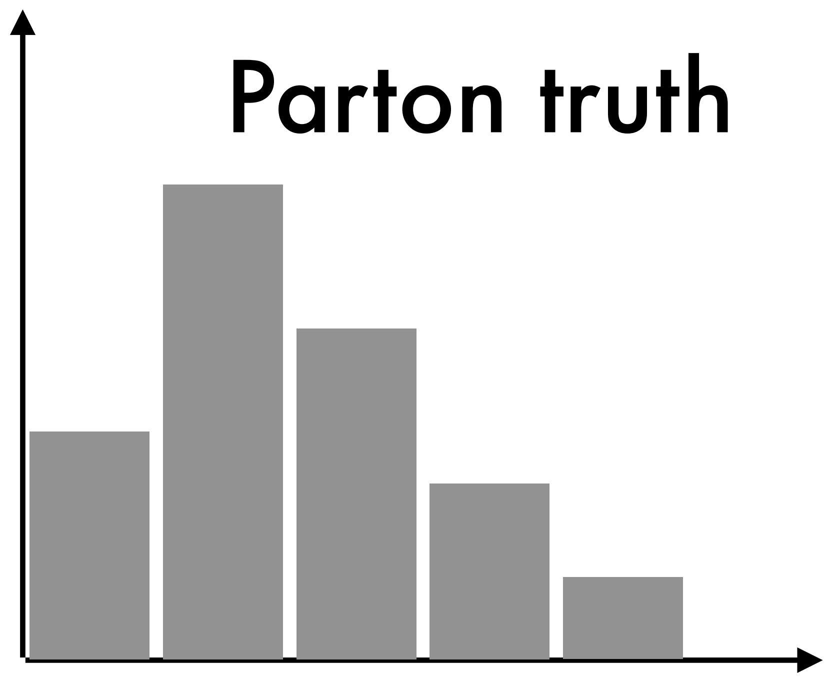
✓ No assumptions on the shape of the distributions.

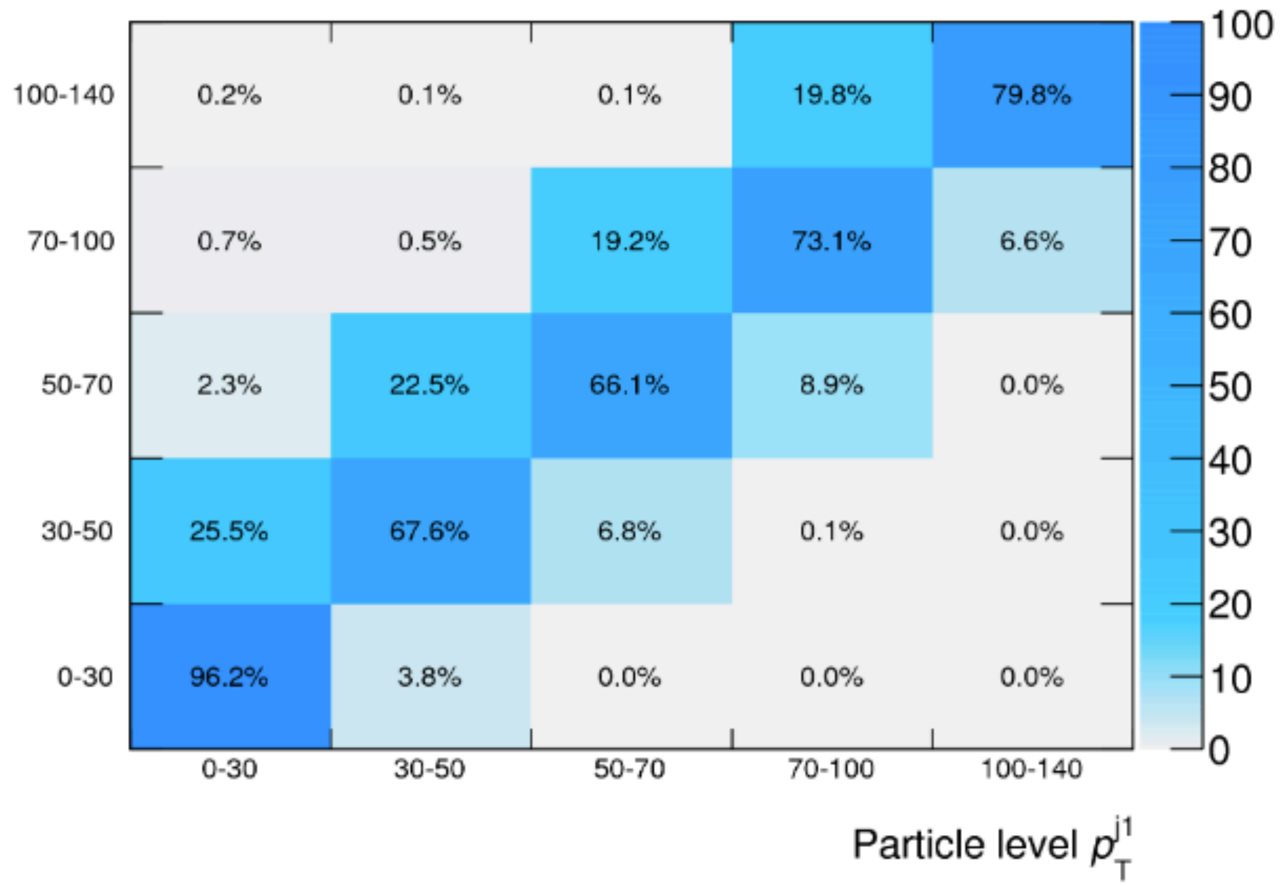
✓ Bin correlations not taken into account.

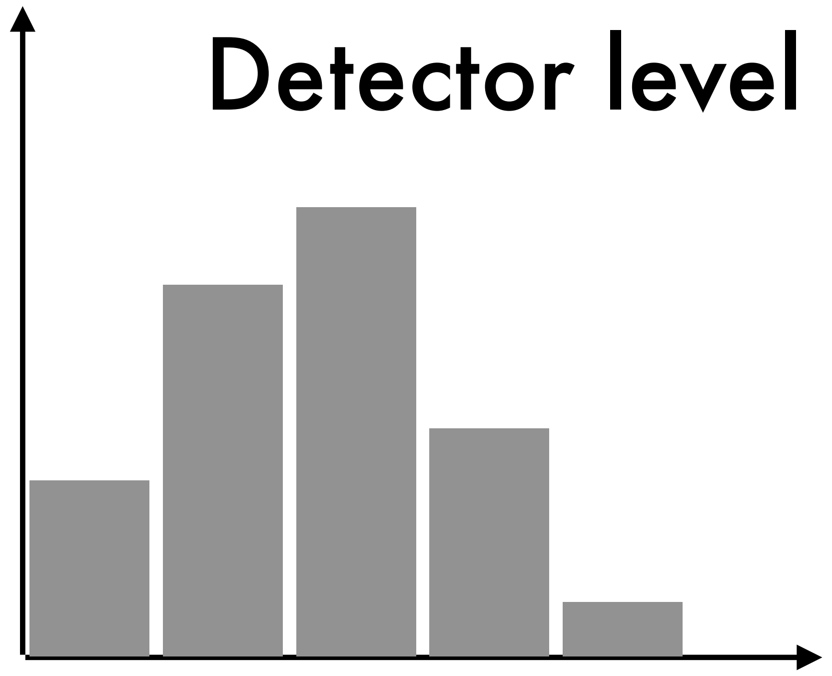
✓ Highly sensitive to MC statistics.

Unfolding

- 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: $R_{ij} = N_{\text{truth},i} / N_{\text{reco},j}$ Unfolded events in bin i :$x_{p,i} = \Sigma_j \, x_{d,j} \times R$

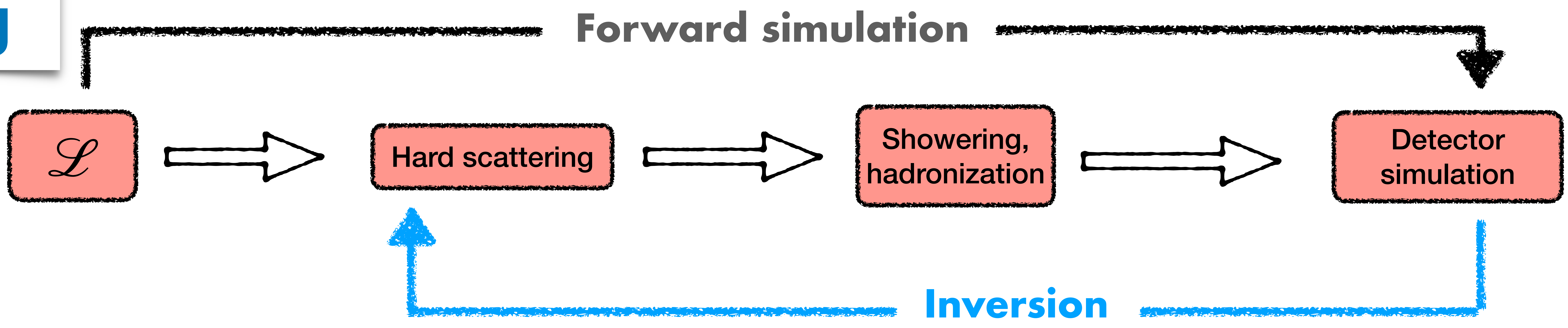
✓ No assumptions on the shape of the distributions.
✓ Noise amplification.
✓ Limited by statistics and dimensionality.

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



✦ Bin-independent

✦ Able to invert multi-dimensional d.o.f.

Possible with machine learning based generative models.

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

[Bellagente, Butter, Kasieczka, Plehn, Winterhalder (2020)]
[Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]
[Andreassen, Komiske, Metodiev, Nachman, Thaler (2020)]
[Komiske, McCormack, Nachman (2021)]

GANs

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

In GANs, the generator and discriminator network competes against each other.

G

D

GANs

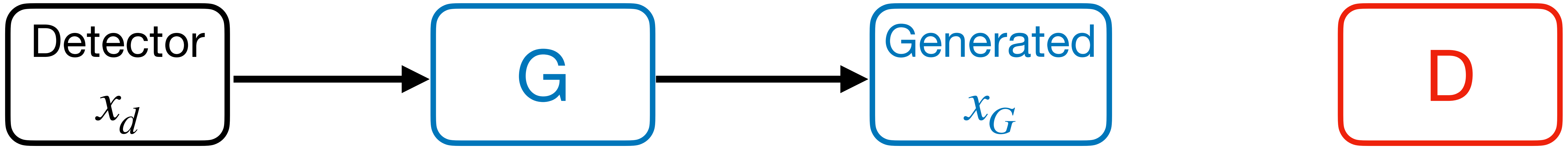
[Bellagente, Butter, Kasieczka, Plehn, Winterhalder(2019)]
[Butter, Plehn, Winterhalder(2019)]

In GANs, the generator and discriminator network competes against each other.

GANs

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

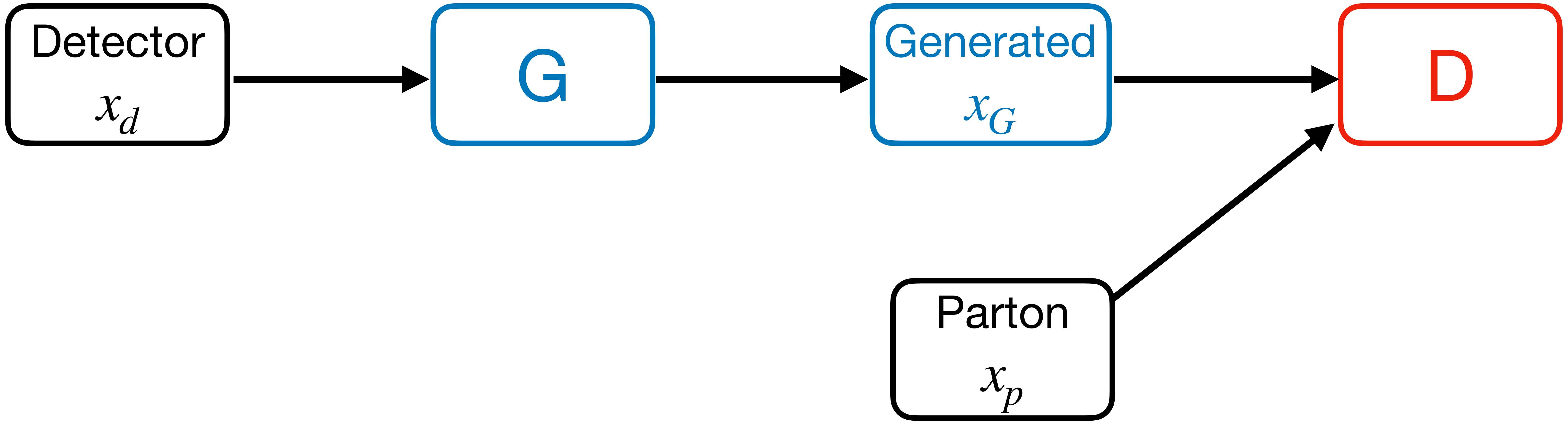
In GANs, the generator and discriminator network competes against each other.



GANs

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

In GANs, the generator and discriminator network competes against each other.

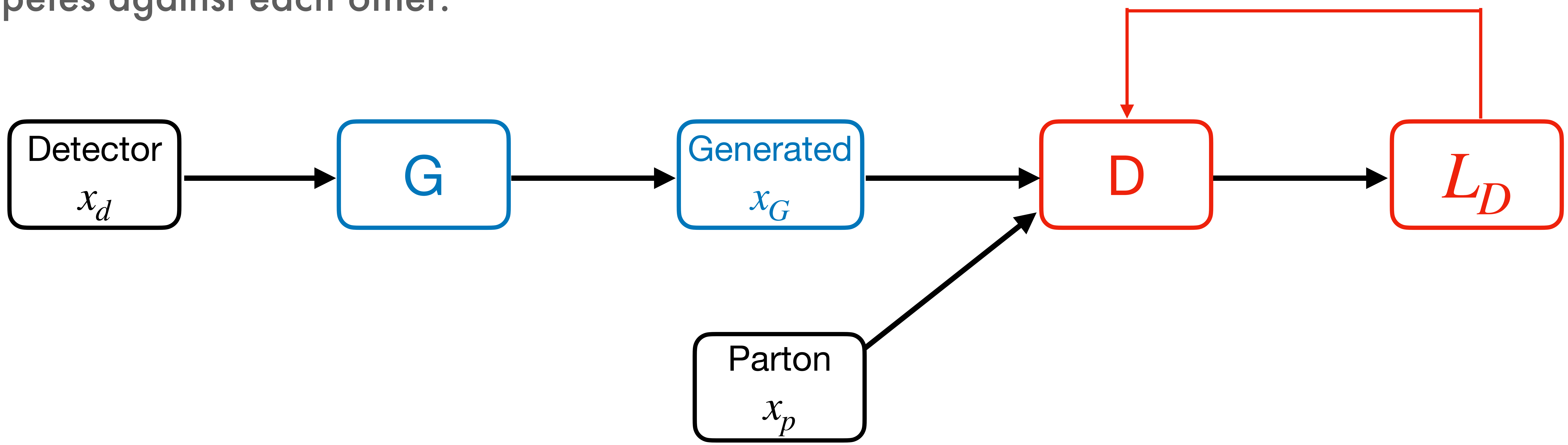


$$D(x_p) \to 1, \quad D(x_G) \to 0$$

GANs

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

In GANs, the generator and discriminator network competes against each other.



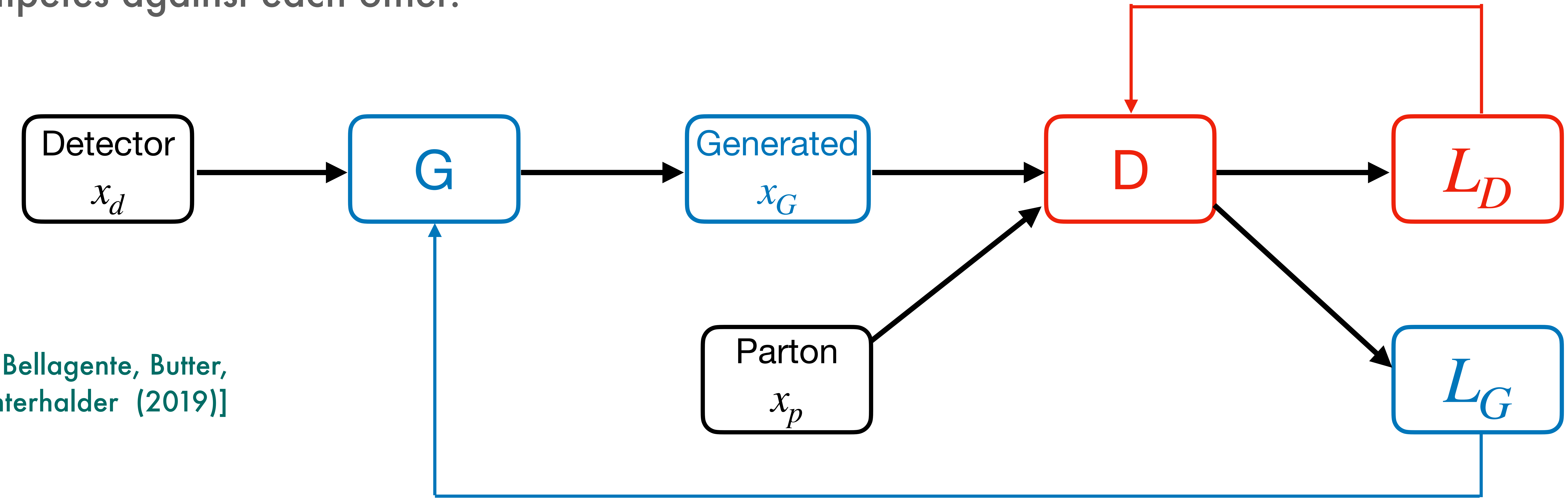
$$L_\mathrm{D} = \langle -logD(x)\rangle_{x\sim P_p} + \langle -log(1 - D(x))\rangle_{x\sim P_G}$$

$$D(x_p) \to 1, \quad D(x_G) \to 0$$

GANs

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

In GANs, the generator and discriminator network competes against each other.



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

- Discriminator works to distinguish generated data $\{x_G\}$ from truth data $\{x_p\}$. $[D(x_P) \to 1, D(x_G) \to 0]$
- Generator works to fool the discriminator such that $D(x_G) \to 1$.

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

$$L_G = \langle -logD(x) \rangle_{x \sim P_G}$$

$$D(x_p) \to 1, \quad D(x_G) \to 0$$

Naive GAN unfolding

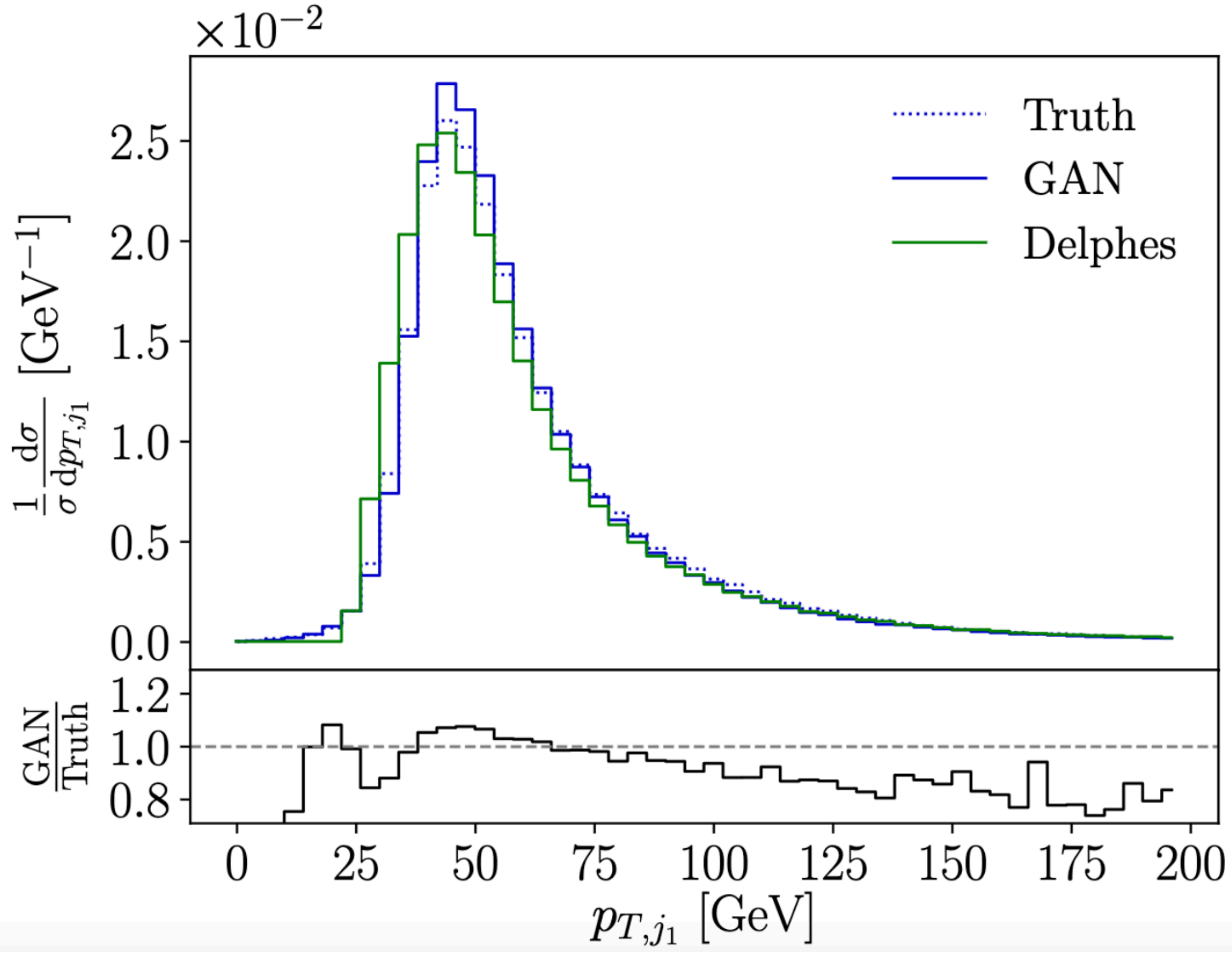
$$pp \to ZW \to (Z \to \ell^+\ell^-)(W \to jj)$$

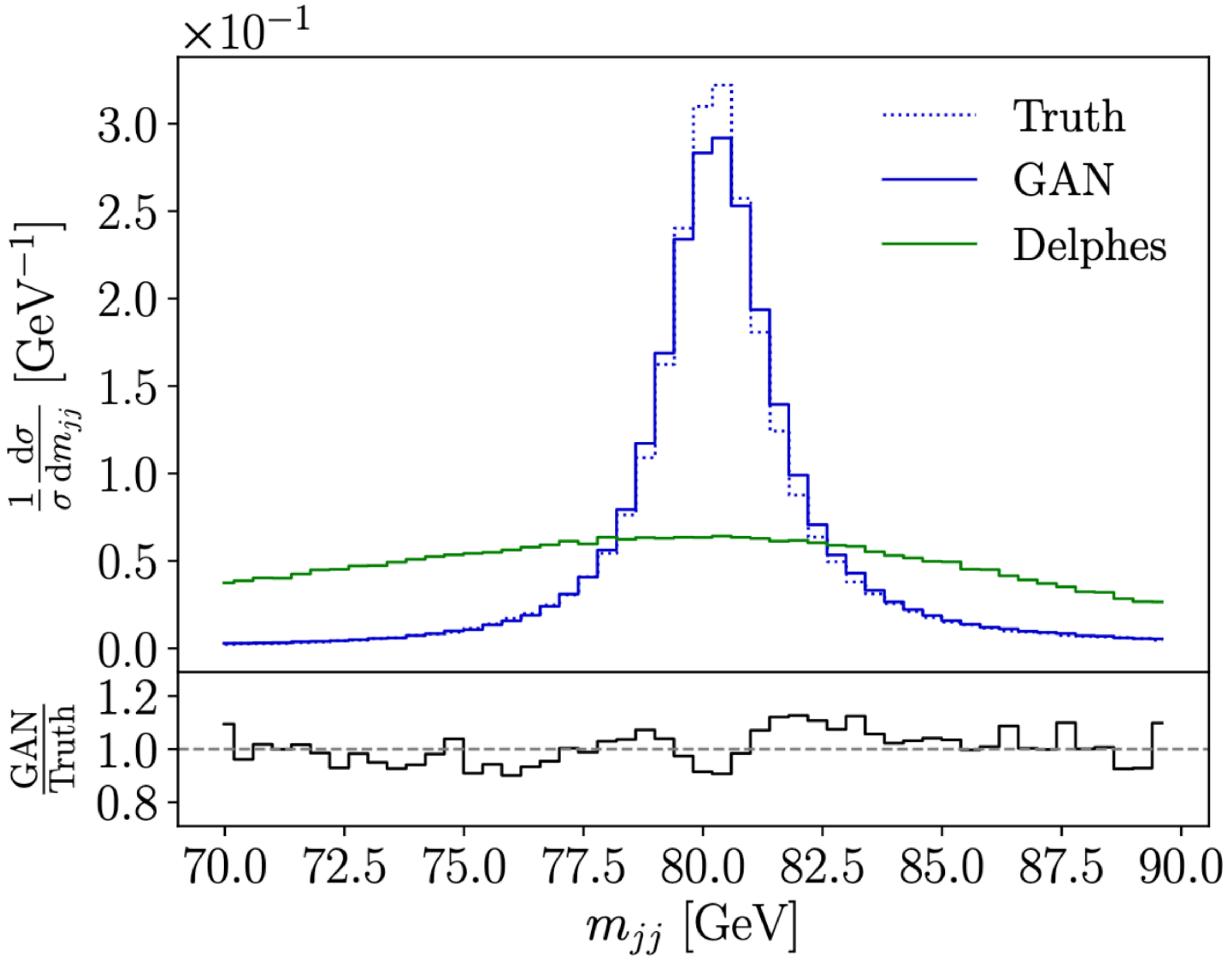
Training data

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

2 ℓ + 2 exclusive j
@ detector level

Parton-level ⟷ Detector-level



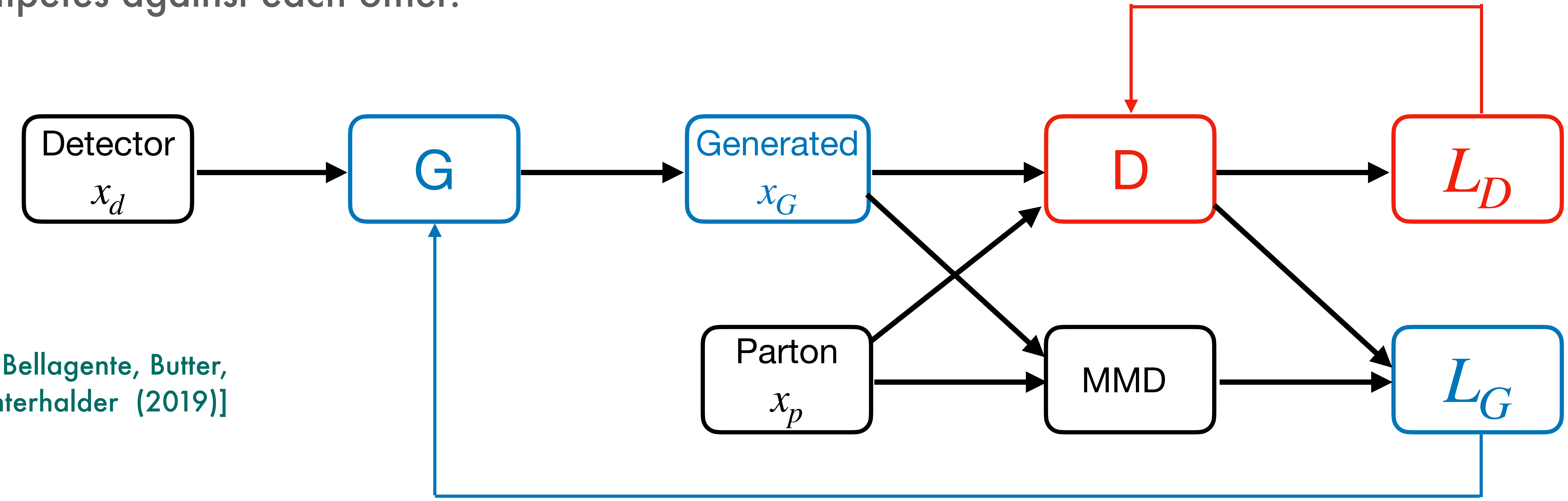


Figures taken from Bellagente, Butter, Kasieczka, Plehn, Winterhalder(2019)

GANs

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

In GANs, the generator and discriminator network competes against each other.



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

Limitations

$$pp \to ZW \to (Z \to \ell^+\ell^-)(W \to jj)$$

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

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

Training data — Parton-level ⟷ Detector-level ($2\,\ell + 2\,j$ exclusive @ detector level)

$p_{T,j} > 25$ 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)

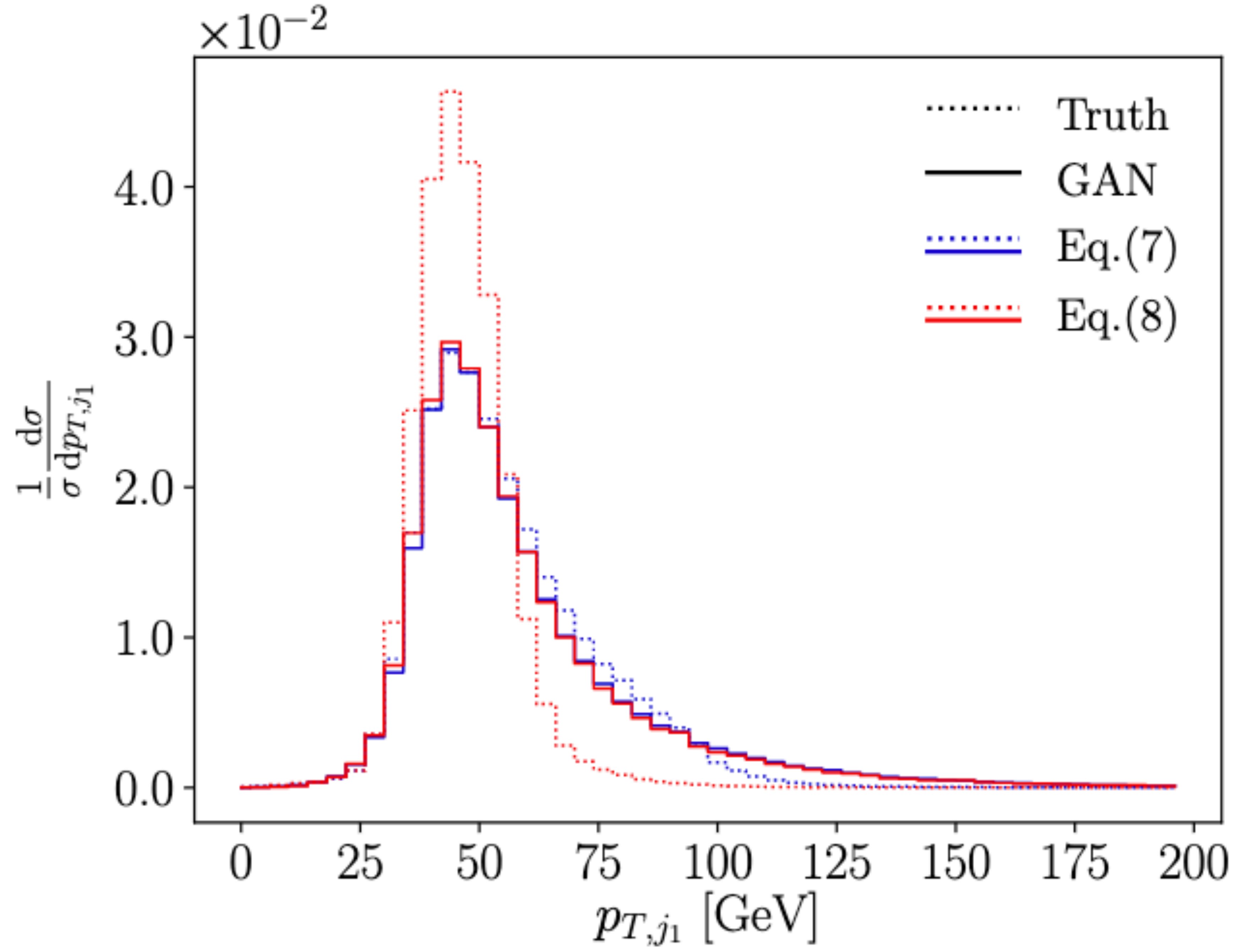
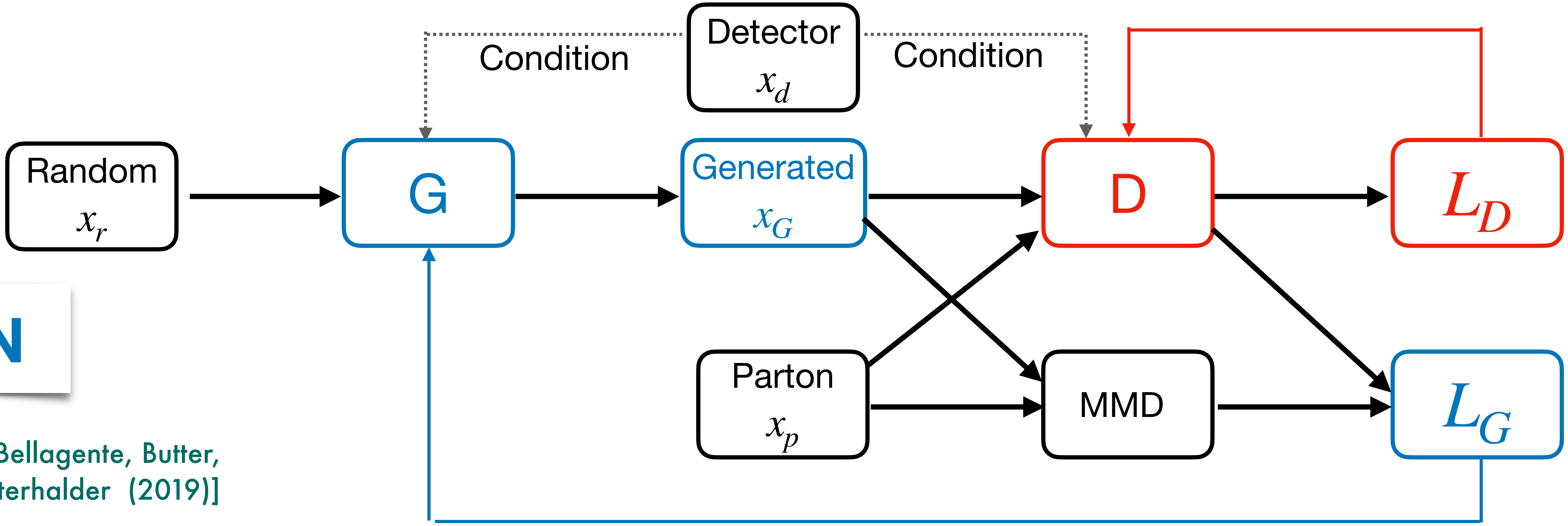


Figure taken from Bellagente, Butter, Kasieczka, Plehn, Winterhalder(2019)

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 to not statistically similar.



FCGAN

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

FCGAN

$$L_D^{FC} = \langle -logD(x,y)\rangle_{x\sim P_p, y\sim P_d} + \langle -log(1 - D(x,y))\rangle_{x\sim P_G, y\sim P_d}$$

$$L_G = \langle -logD(x,y)\rangle_{x\sim P_G, y\sim P_d}$$

- Event-by-event matching → exploit the pairing information between parton and detector level.

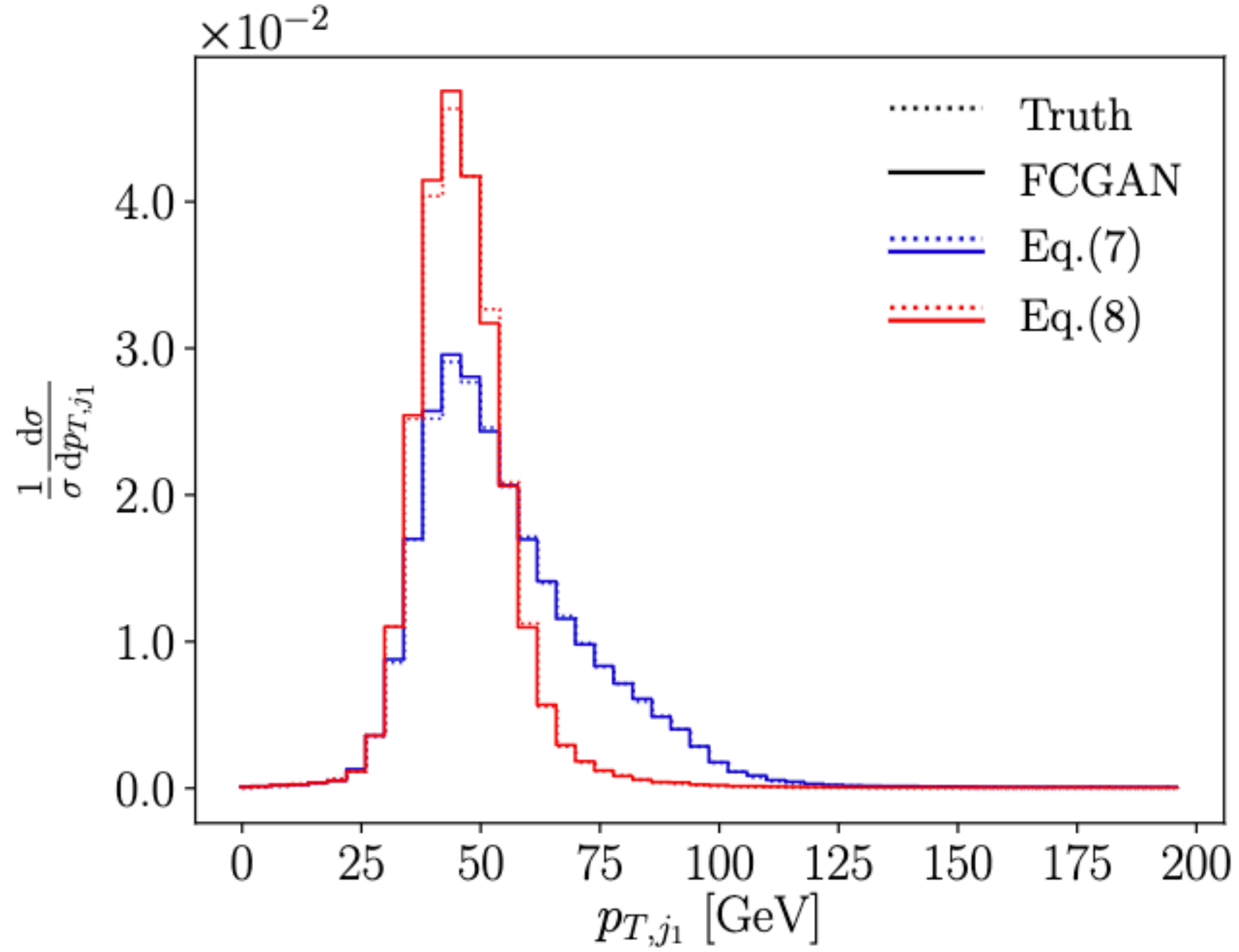
- Trained network can be applied to statistically different regions of phase space.

Test data

$2\,\ell + 2\,j$ exclusive @ detector level

$$30\text{ GeV} < p_{T,j_1} < 60\text{ GeV},\ 30\text{ GeV} < p_{T,j_2} < 50\text{ GeV},$$

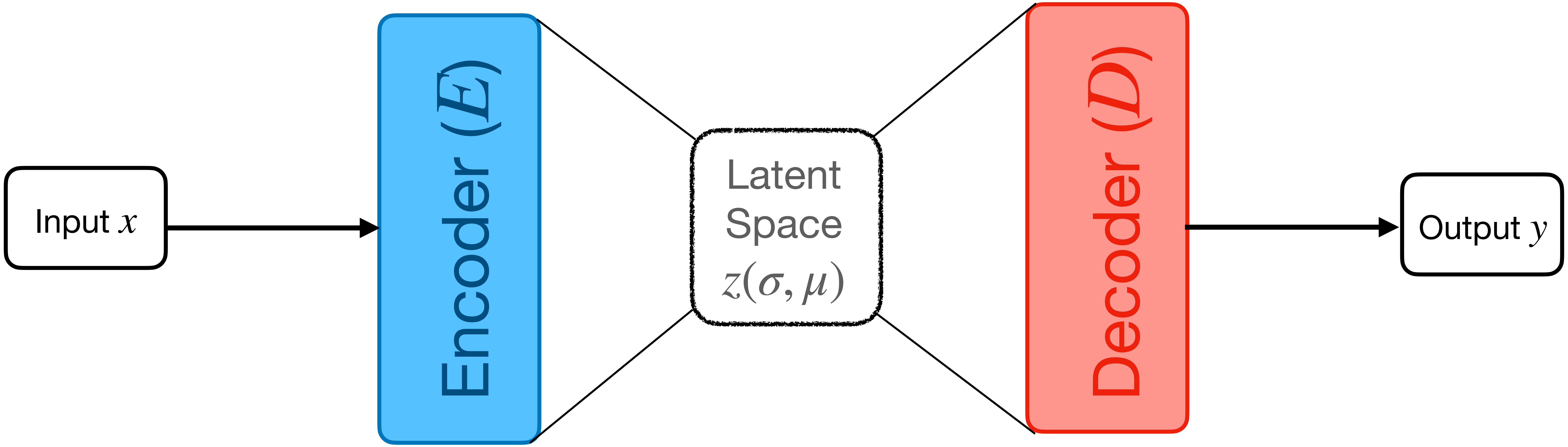
($\sim 14\,\%$ of events)



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

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

Variational Autoencoders



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

Reconstruction Loss

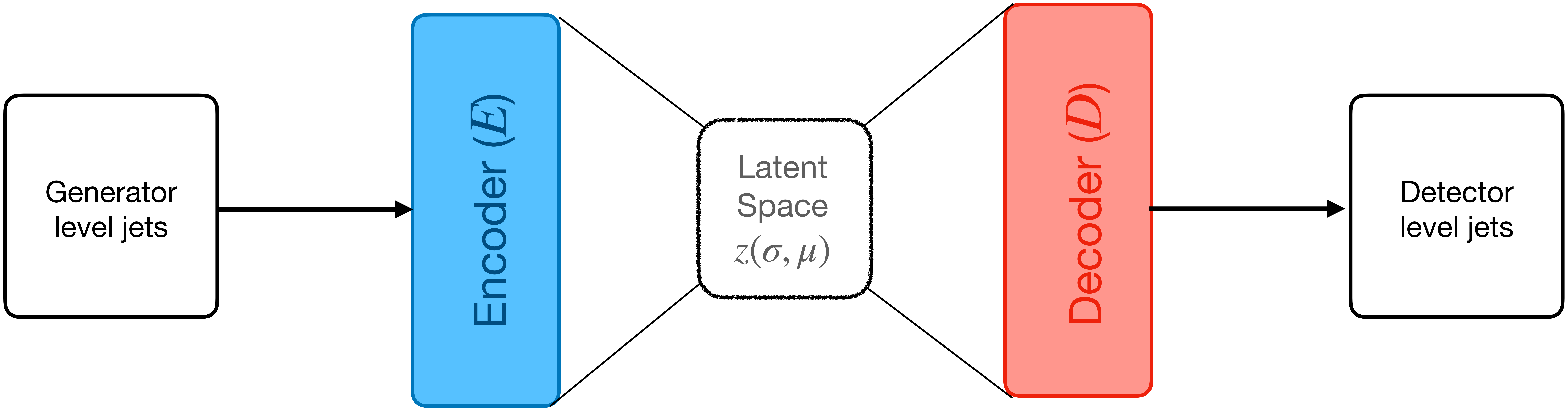
KL divergence term

[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: $pp \to WW \to (W \to jj)(W \to jj)$

Training data

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

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

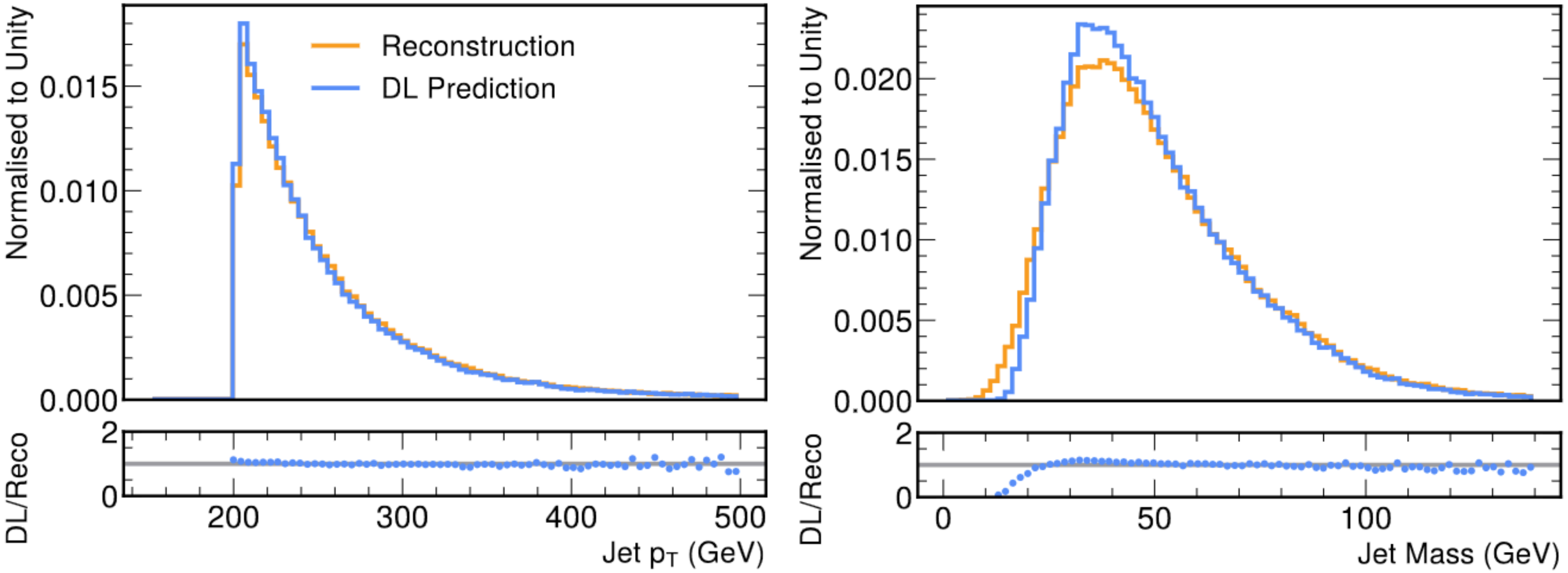
Generator-level ⟷ Detector-level

Input-target jet matched by minimizing ΔR

Variational Auto Encoders

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

Loss $L \propto \beta D^i_{KL} + (1-\beta)(L^i_R + \alpha_m(m^i_{jet} - \tilde{m}^i_{jet})^2 + \alpha_{p_T}(p^i_T - \tilde{p}^i_T)^2)$



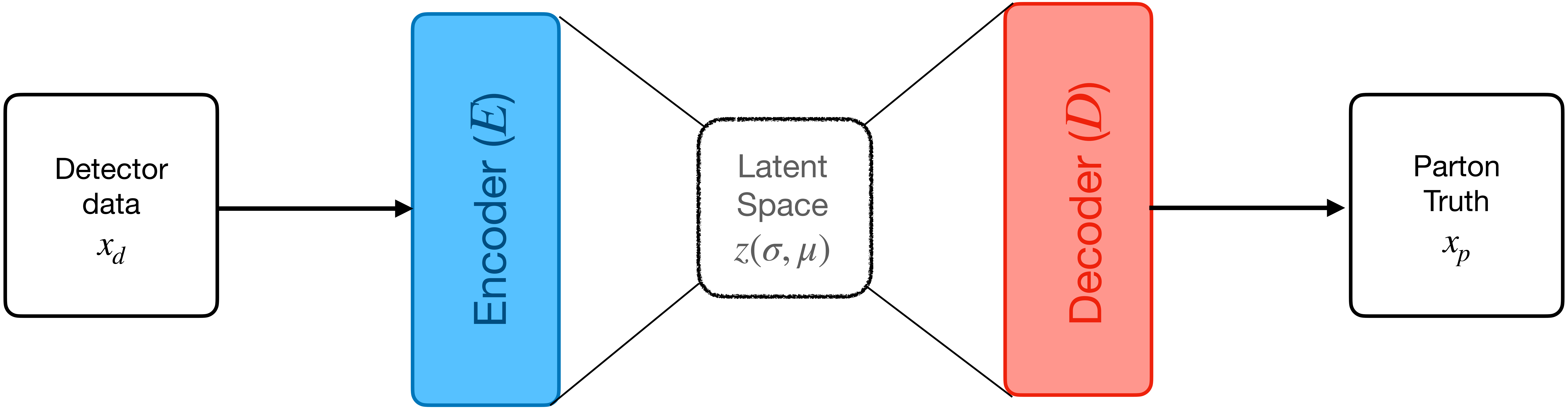
Figures taken from Touranakou, Chernyavskaya, Duarte, Gunopulos, Kansal, Orzari, Pierini, Tomei, Vlimant (2022)

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

Variational Auto Encoders

☑ Unfolding

→ Map detector data to the parton level phase space



- 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:

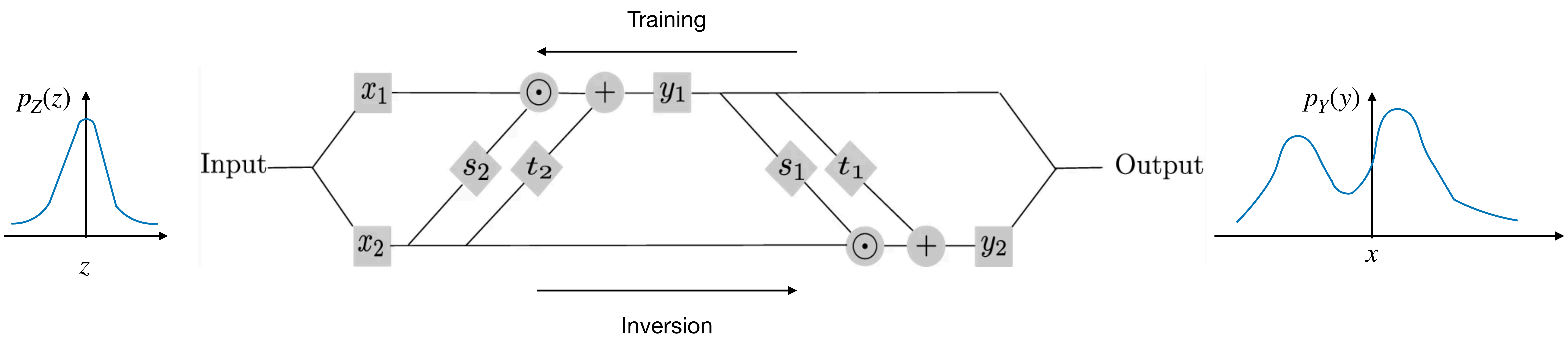
- NF models a complex distribution to a simple distribution using a series of invertible transformations → models intricate distributions without making strict assumptions.

- 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

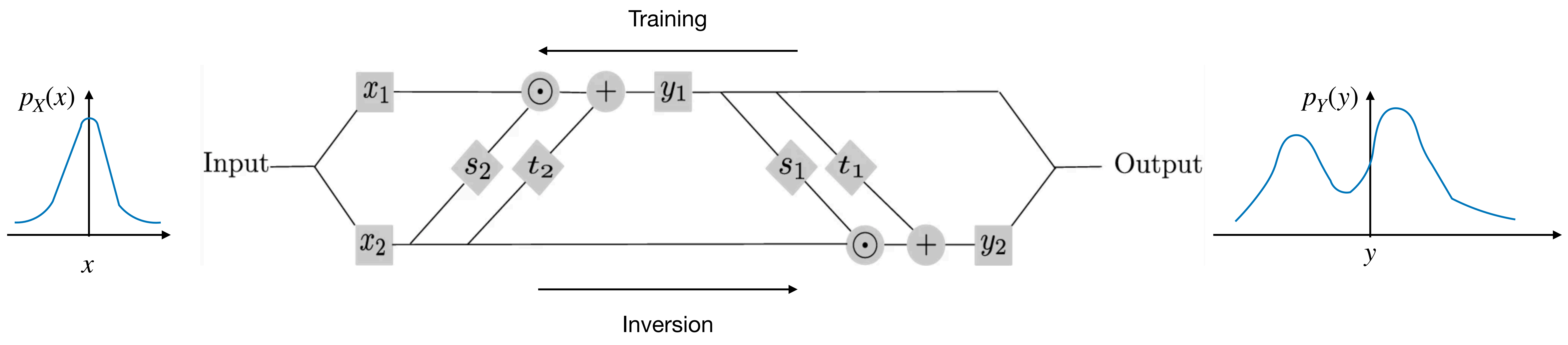
- Series of bijective layers that transform complex (Y) to simple probability distributions (Z).
- Learns both directions of the mapping in parallel → bijectivity encoded in the same network.
- Building blocks → Invertible coupling layers. [Dinh, Krueger, Bengio (2016), Dinh, Sohl-Dickstein, Bengio (2016)]



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

Normalizing flows

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



- In the coupling layers, the coupling functions s_2 and t_2 take x_2 as input, and scale/translate x_1.
- Fully invertible coupling layer → $[x_1, x_2]$ can be reconstructed given $[y_1, y_2]$

Forward pass:

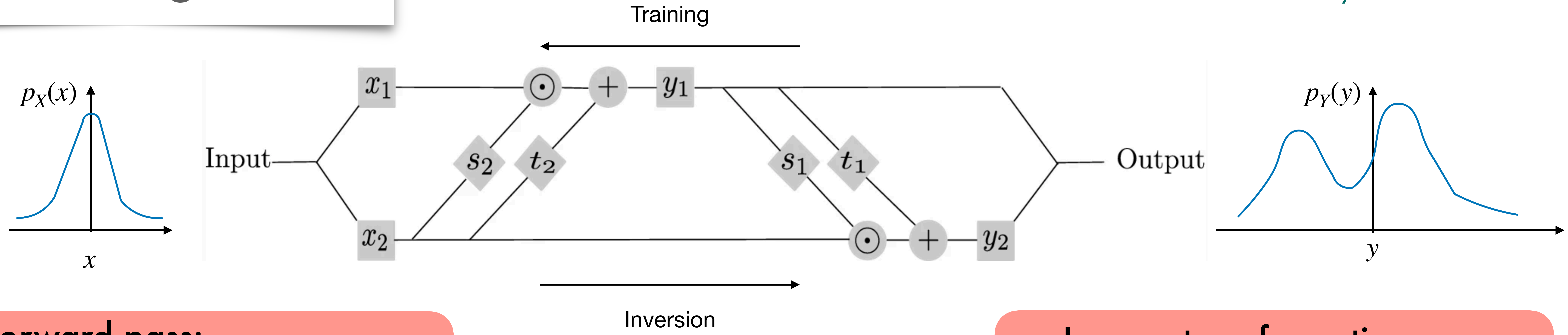
$$y_1 = x_1 \odot e^{S_2(x_2)} + t_2(x_2)$$
$$y_2 = x_2 \odot e^{S_1(y_1)} + t_1(y_1)$$

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

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



Forward pass:

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

Inverse transformations:

$$x_1 = (y_1 - t_2(x_2)) \odot e^{-s_2(x_2)}$$
$$x_2 = (y_2 - t_1(y_1)) \odot e^{-s_2(y_1)}$$

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

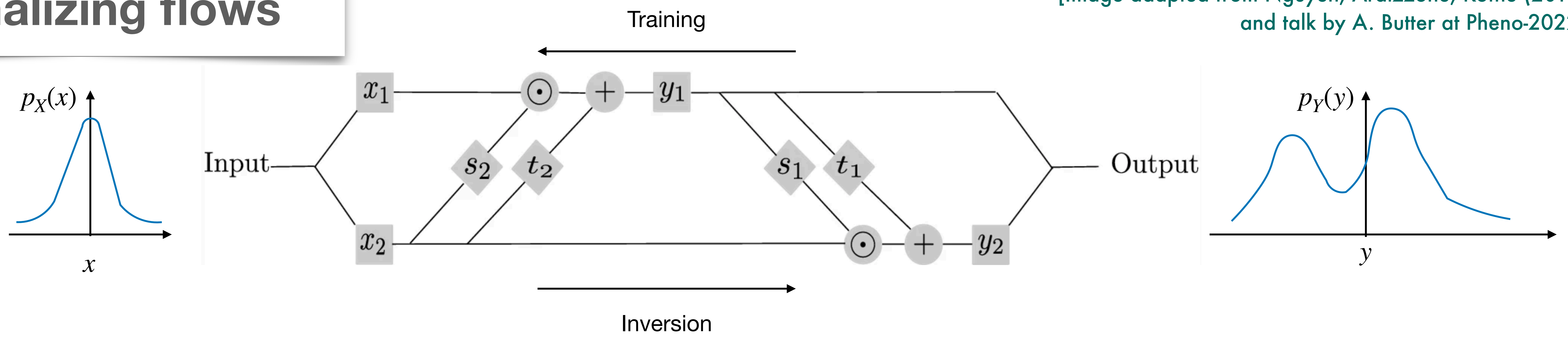
tractable Jacobian $J_f(x)$: $\frac{\partial f(x)}{\partial x} = \begin{bmatrix} e^{S_2(x_2)} & \text{finite} \\ 0 & e^{S_1(y_1)} \end{bmatrix}$

→ 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

Normalizing flows

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



- 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

Naive INN unfolding

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

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

(x_{parton}) $\longleftrightarrow$ (x_{detector})

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

$\leftarrow$ Unfolding: g

Training data

Exactly 2 jets: $p_T > 25\ \text{GeV}, |\eta| < 2.5$

SFOS lepton pair near m_Z

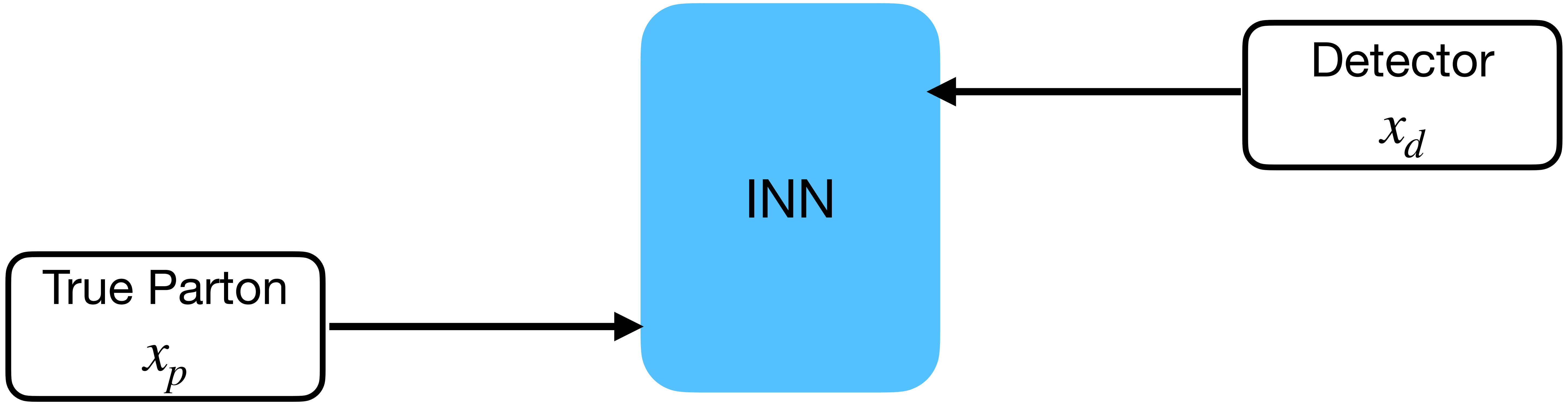
Parton-level $\longleftrightarrow$ Detector-level

Event-wise matching

Naive INN unfolding

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

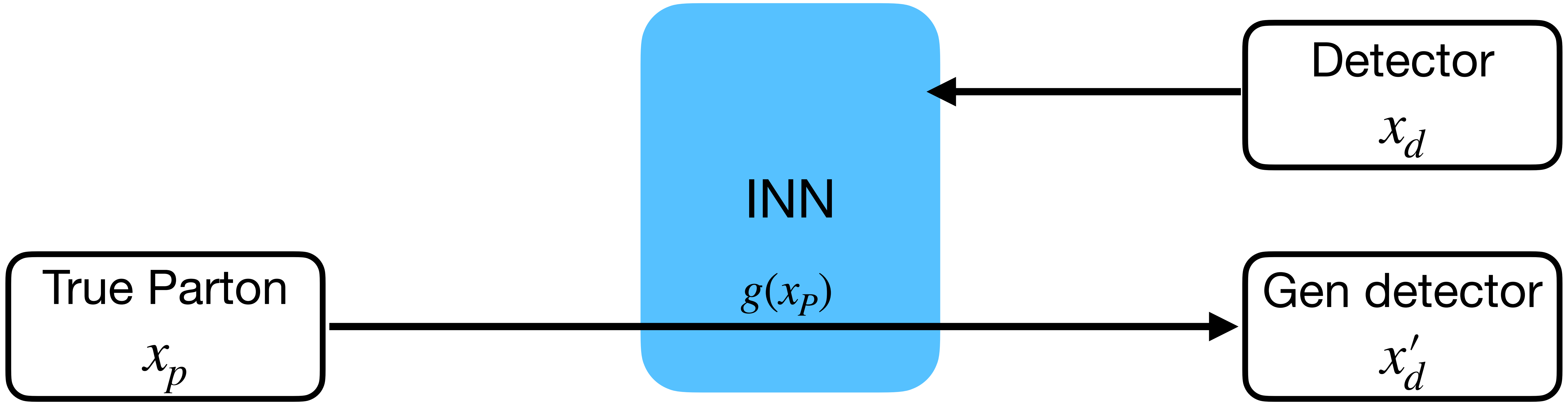
← Unfolding: g



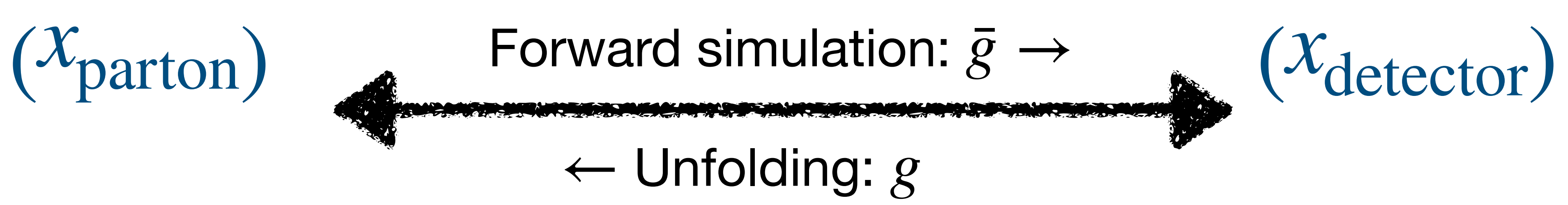
Naive INN unfolding

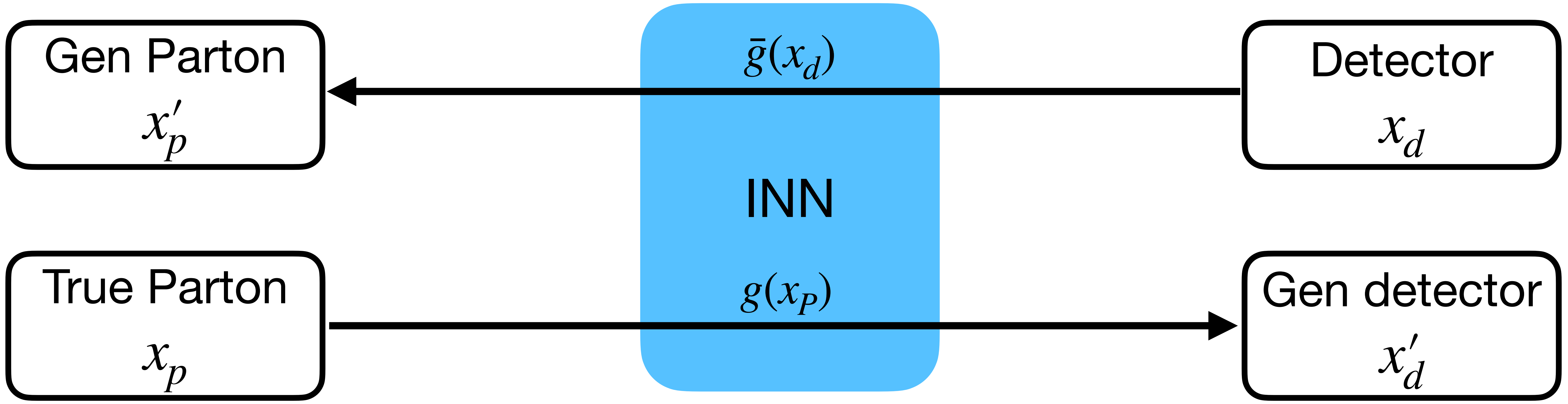
(x_{parton}) Forward simulation: $\bar{g} \rightarrow$ (x_{detector})

$\leftarrow$ Unfolding: g



Naive INN unfolding

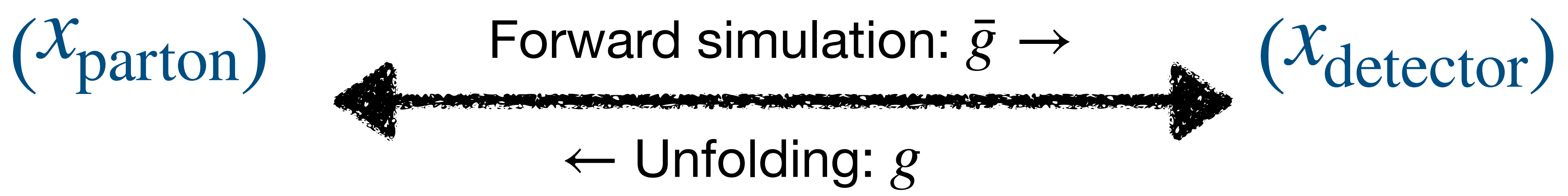


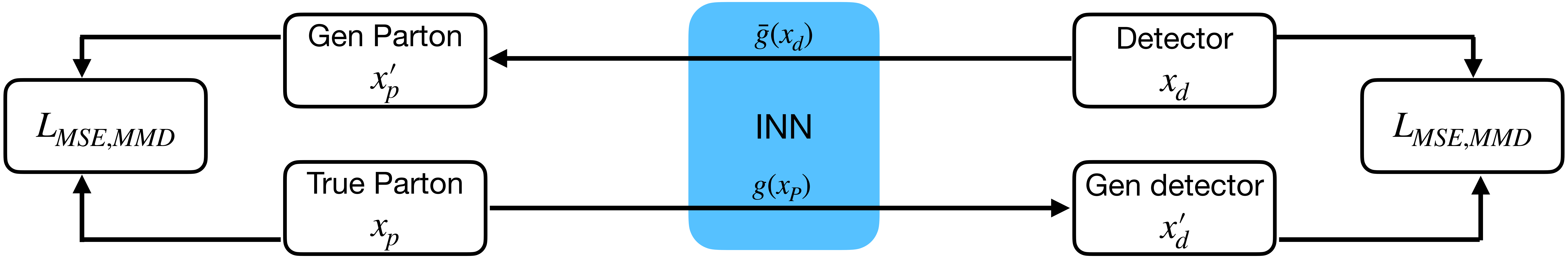


Naive INN unfolding

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

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





Loss: $L_{MSE}(x_p) + L_{MSE}(x_d) + L_{MMD}$

Naive INN unfolding

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

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

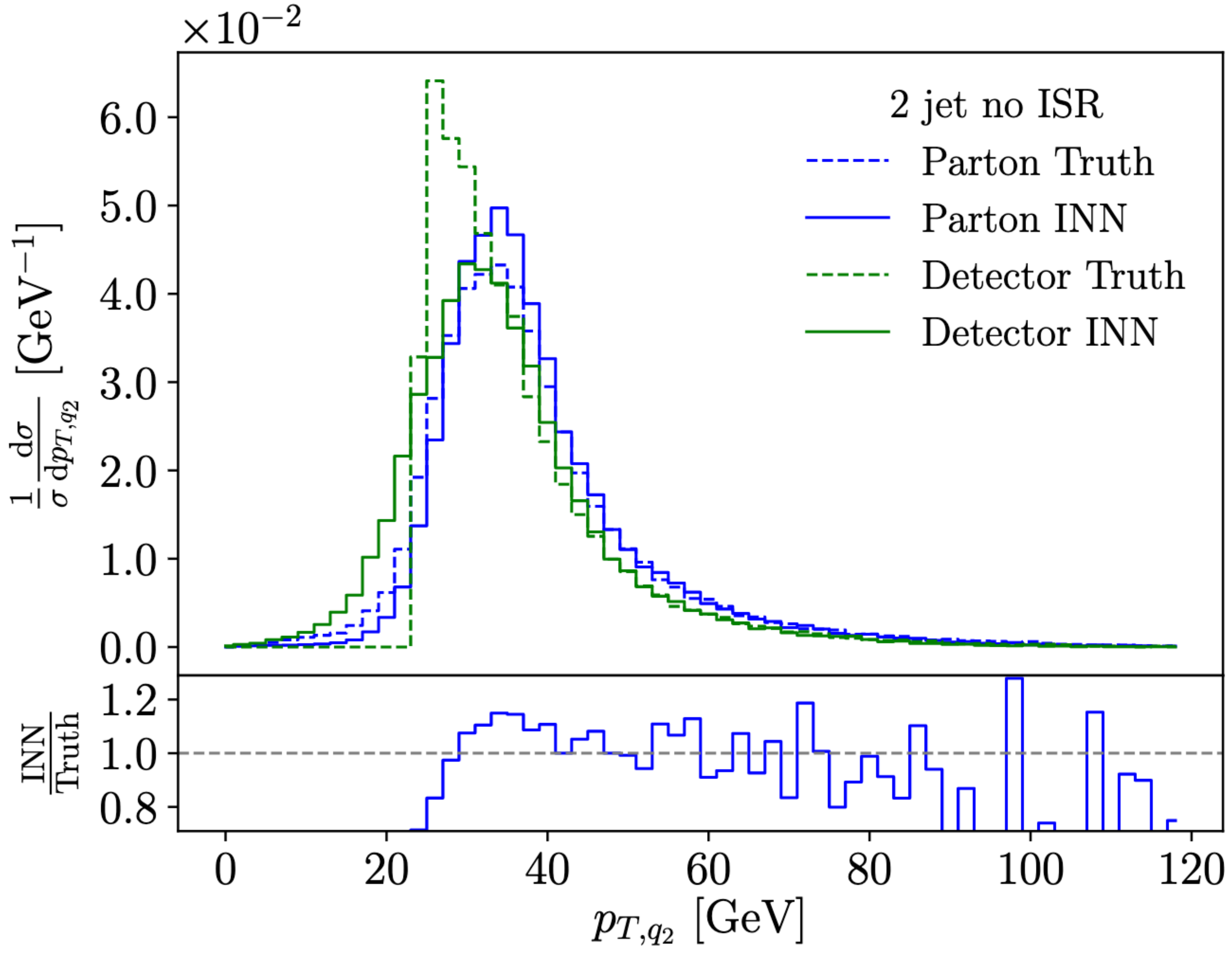
(x_{parton}) $\xrightarrow{\text{Forward simulation: } \bar{g}}$ / $\xleftarrow{\text{Unfolding: } g}$ (x_{detector})

- 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.

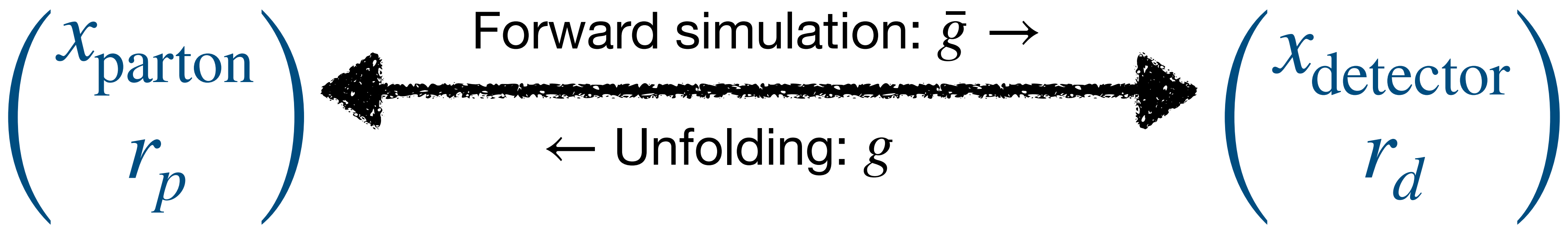
- Dimensionality limitations.



Noise-extended INN

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

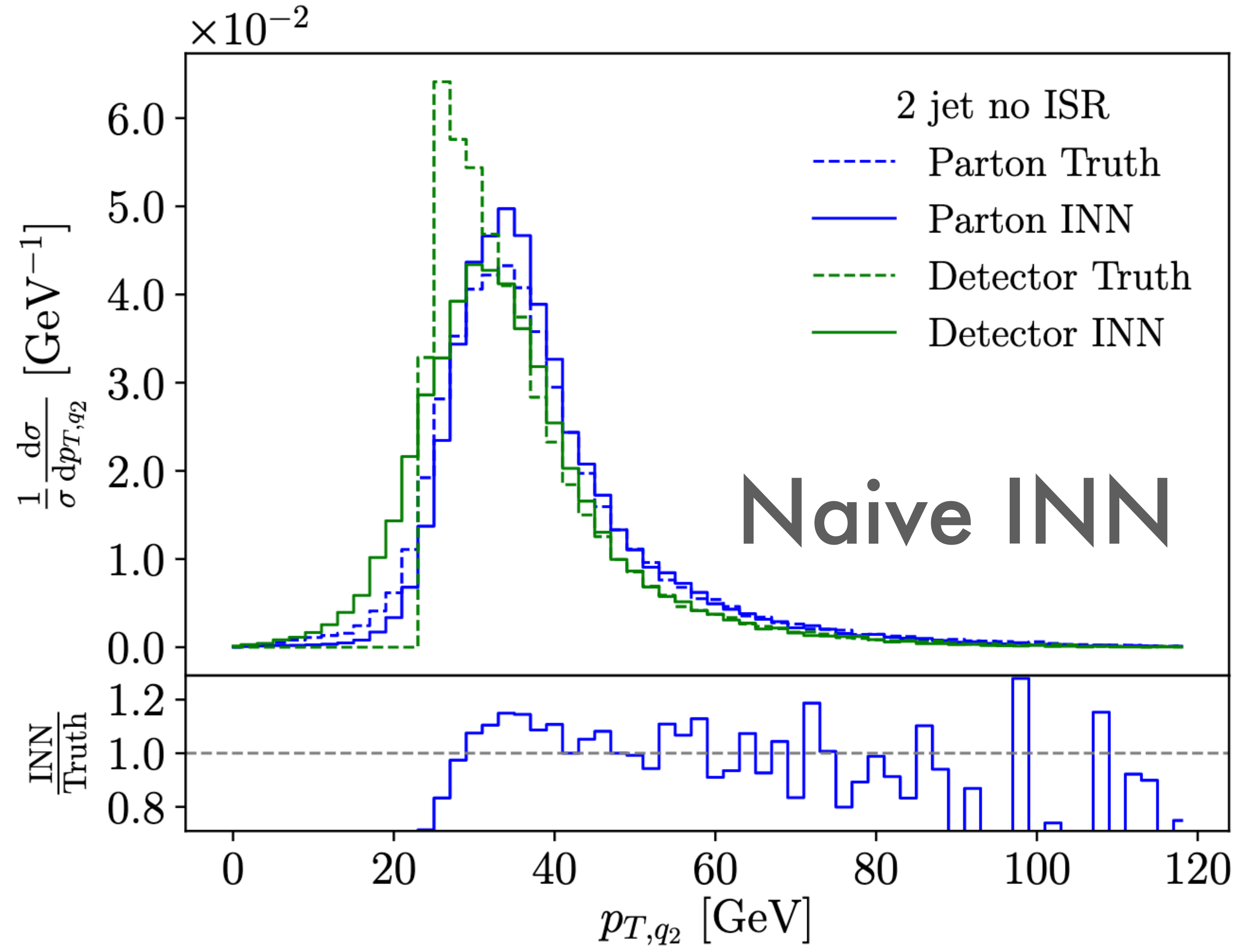
[Figure taken from Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]

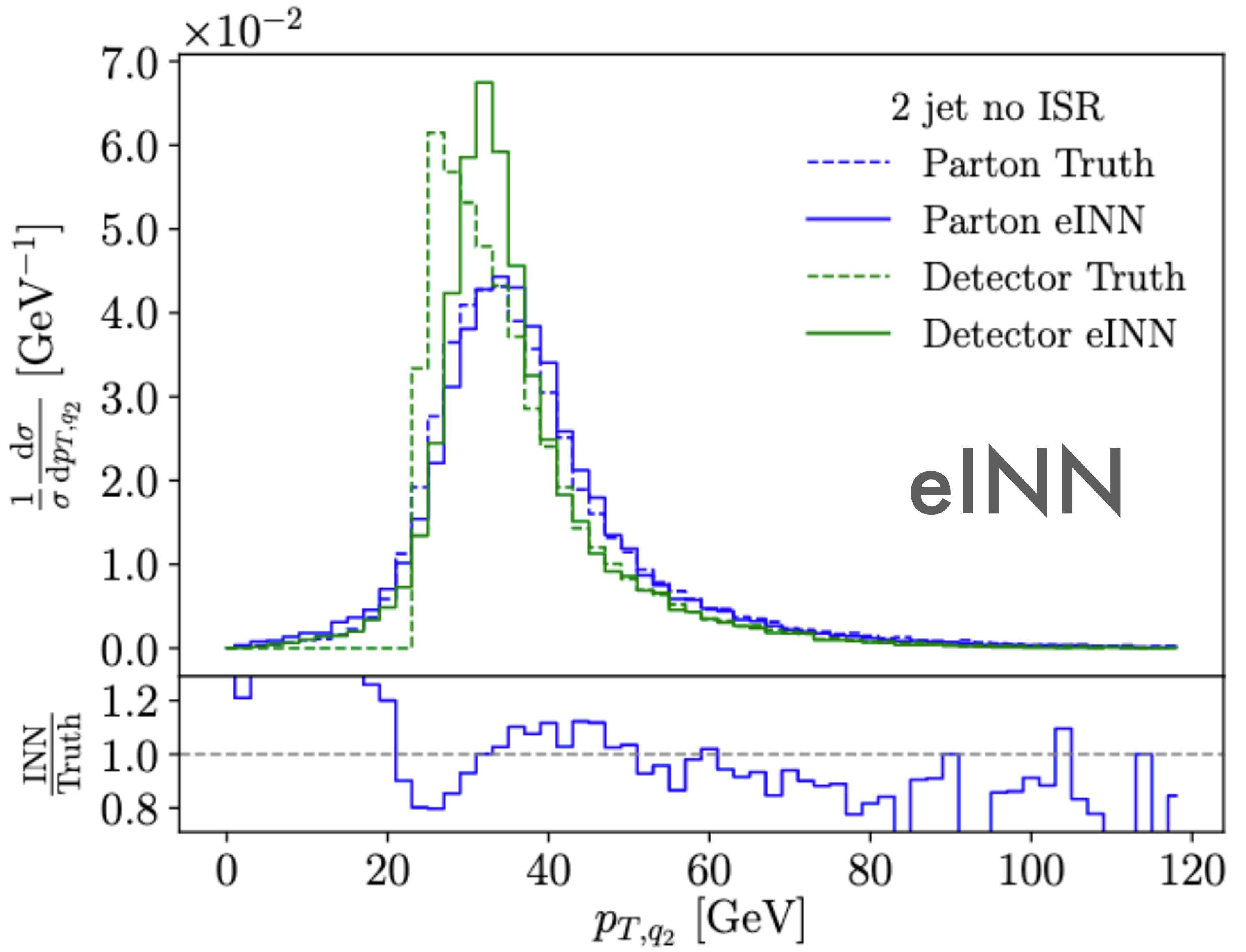


- Allows mapping between unequal degrees of freedom at the parton and detector level.

- 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.





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.

$\rightarrow$ Upgrade to conditional INN

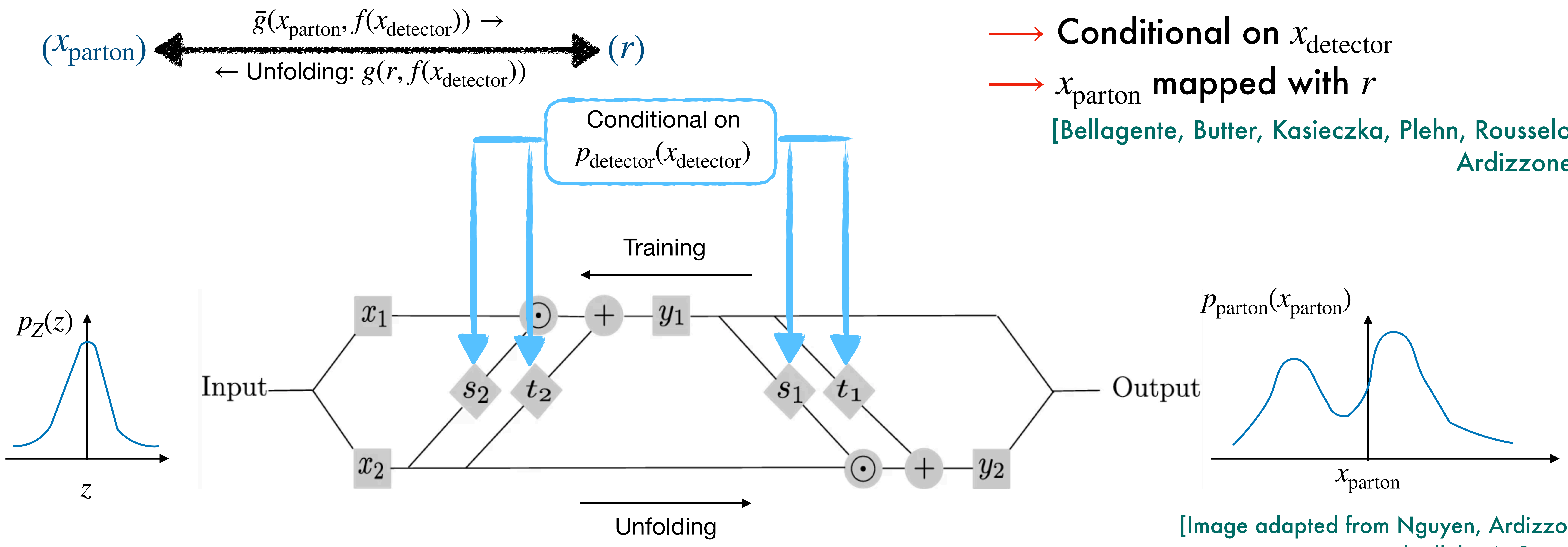
Conditional INN

- Generate probability distributions at the parton-level, given detector-level events x_{detector}

$\longrightarrow$ Conditional on x_{detector}

$\longrightarrow$ x_{parton} mapped with r

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



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

- Sharp features are included via MMD loss terms.

Target phase space for unfolding can be chosen flexibly to include:

☑ QCD jet radiation

☑ Particle decays

Unfolding semileptonic $t\bar{t}h$ events

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

➡ Parton-level:

$$1\ell + 2b + 2\gamma + \nu + 2j$$

➡ Detector-level:

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

Acceptance cuts

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

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

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.

[Butter, Plehn, Winterhalder (2019)]
[Bellagente, Butter, Kasieczka, Plehn, Rousselot, Winterhalder, Ardizzone, Kothe (2020)]

<u>Alternative approach:</u>

$\rightarrow$ directly learn invariant mass features and important observable with appropriate phase-space parametrization.

$\rightarrow$ may provide direct access to the most important BSM observables.

Conditional INN

$$(x_{\text{parton}}) \xleftrightarrow[\leftarrow \text{Unfolding: } g(r, f(x_{\text{detector}}))]{\bar{g}(x_{\text{parton}}, f(x_{\text{detector}})) \rightarrow} (r)$$

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

- Degrees of freedom:

Parton-level: $(t \to \ell \nu b)(\bar{t} \to j j \bar{b}) h$

22 d.o.f.

Detector-level: 46 d.o.f.

$1\ell + 2b + 2\gamma + MET + \leq 6$ jets inclusive

A natural parametrization involving top mass:

$$\left\{ m_t, p_{T,t}, \eta_t, \phi_t, m_W, \eta_W^t, \phi_W^t, \eta_{\ell,u}^W, \phi_{\ell,u}^W \right\}$$

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

$$\vec{p}_{t\bar{t}}, m_{t_\ell}, |\vec{p}_{t_\ell}^{\,\text{CS}}|, \theta_{t_\ell}^{\text{CS}}, \phi_{t_\ell}^{\text{CS}}, m_{t_h},$$
$$\text{sign}(\Delta\phi_{\ell\nu}^{t\bar{t}})\, m_{W_\ell}, |\vec{p}_\ell^{\,t\bar{t}}|, \theta_\ell^{t\bar{t}}, \phi_\ell^{t\bar{t}}, |\vec{p}_\nu^{\,t\bar{t}}|$$
$$\text{sign}(\Delta\phi_{du}^{t\bar{t}})\, m_{W_h}, |\vec{p}_d^{\,t\bar{t}}|, \theta_d^{t\bar{t}}, \Delta\phi_{\ell d}^{t\bar{t}}, |\vec{p}_u^{\,t\bar{t}}|$$

CP measurement in Higgs-top interactions

- New sources of CPV interactions can explain the matter-antimatter asymmetry in the universe.

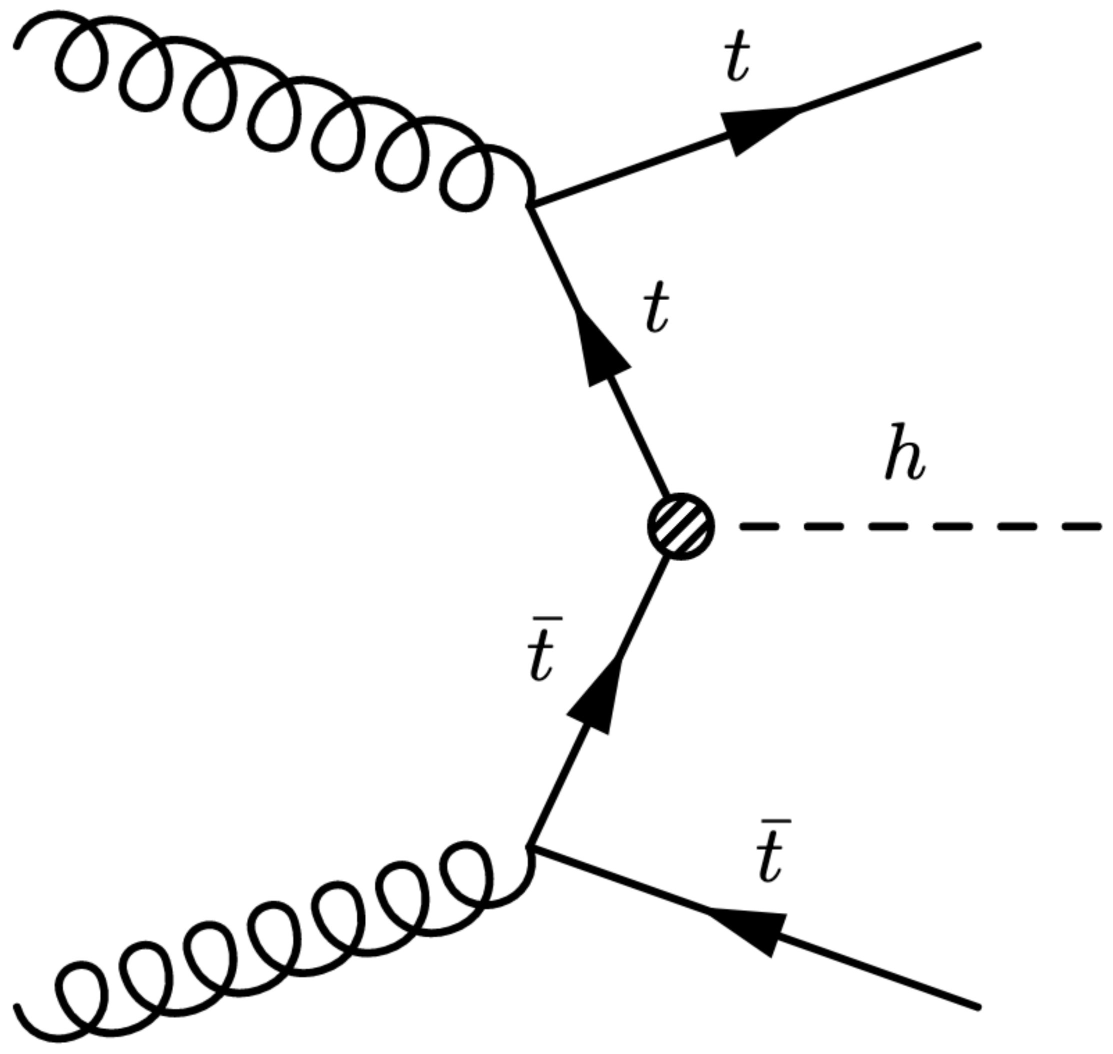
- One such exciting scenario is CP violation in the Higgs sector $\sim$ mixing of CP-even and odd states $\rightarrow$ *CP-mixed hypothesis is still allowed at the LHC.*

- CPV in hVV interactions is extensively tested at the LHC.

  [See for instance: G. Aad et al. (1506.05669), G. Aad et al. (1602.04516), A. M. Sirunyan et al. (1707.00541), A. M. Sirunyan et al. (1903.06973), A. M. Sirunyan et al.(1901.00174), G. Aad et al. (2002.05315), Bernreuther, Gonzalez, Wiebusch (2010), Englert, Goncalves, Mawatari, Plehn (2012), Djouadi, Godbole, Mellado, Mohan (2013), Anderson, Bolognesi, Caola, Gao et al. (2013)]

- CPV in $hf\bar{f}$ couplings manifest at tree-level:

  $\rightarrow$ desirable choice: $ht\bar{t}$



Direct probes at the LHC

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

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

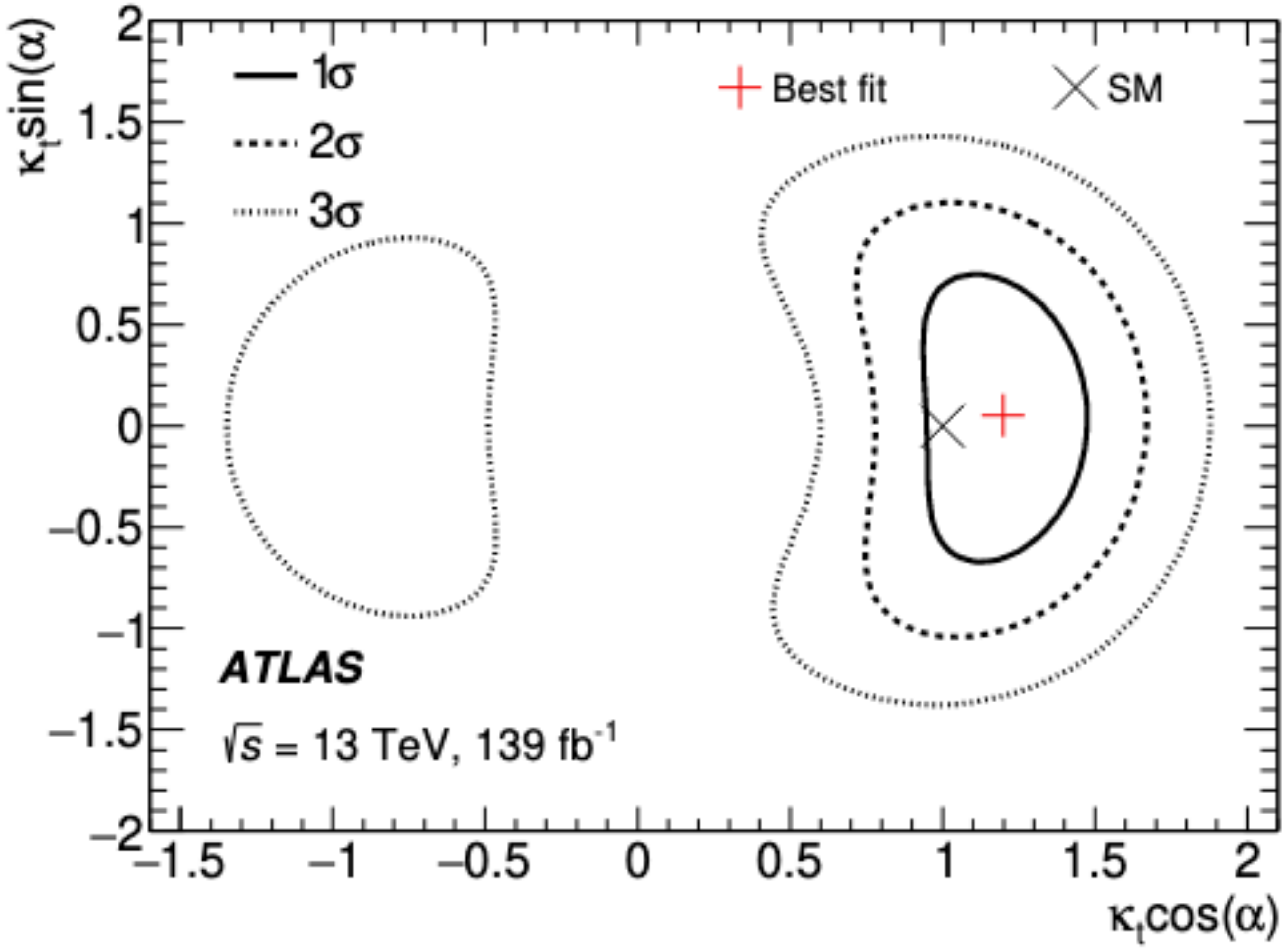
- $pp \to h$ (+ jets): **indirect constraints.** [Duca, Kilgore, Oleari, Schmidt, Zeppenfeld (2001), Klamke, Zeppenfeld (2007), Grojean et al. (2013), Dolan, Harris, Jankowiak, Spannowsky (2014)]

- $pp \to t\bar{t}h$ **stands out as the viable direct probe:**

  → *Small rate at the LHC and complex topology.*

  → Silver Lining: Observation at 5.2σ by ATLAS [2004.04545] and 6.6σ by CMS [2003.10866]

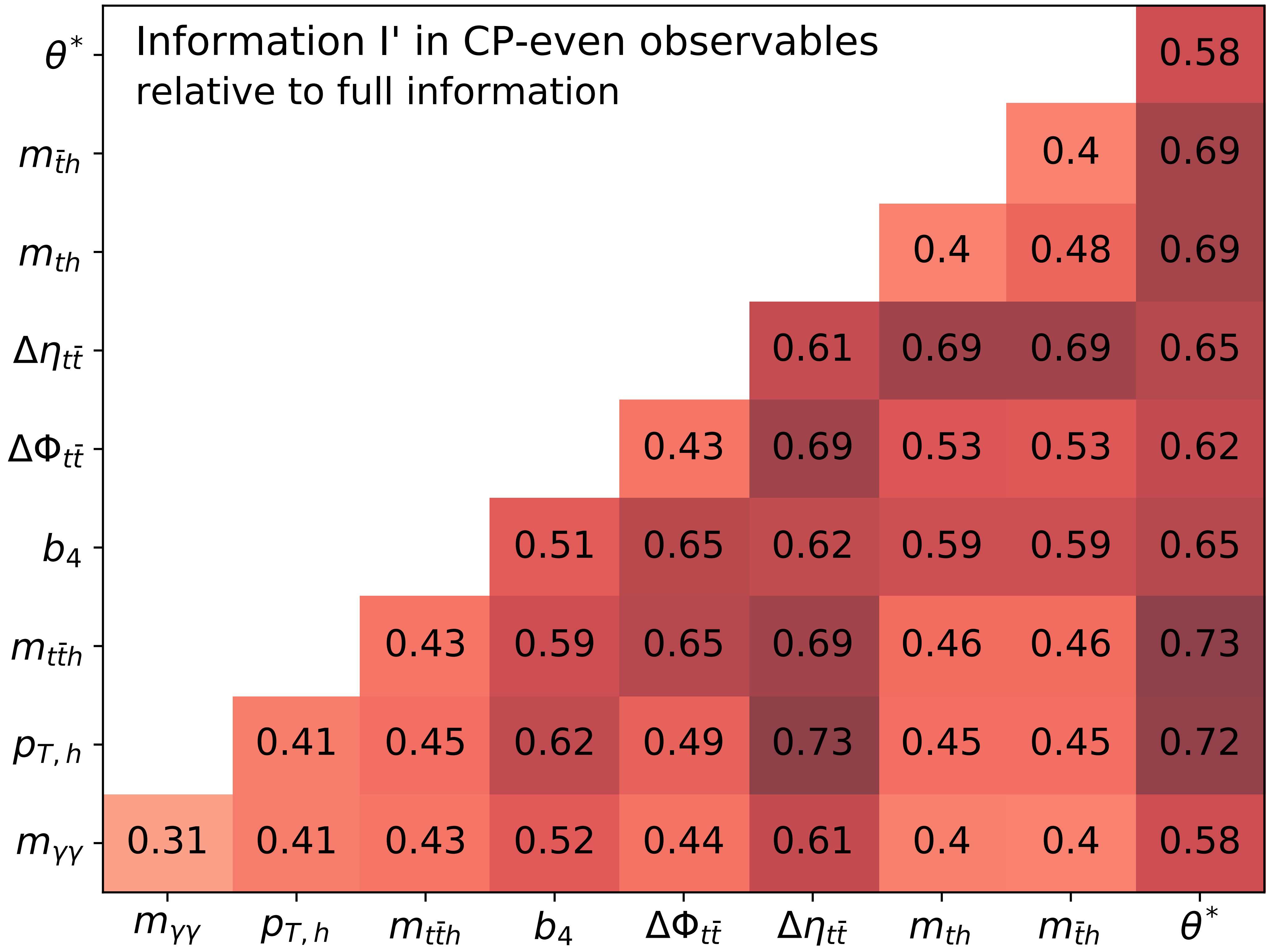
- Current limits: $|\alpha| < 43^0$ (ATLAS) and $|\alpha| < 55^0$ (CMS), at $95\ \%$ CL.



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

$t\bar{t}(h \to \gamma\gamma)$ @ HL-LHC

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



[RKB, Goncalves, Kling (2021)]

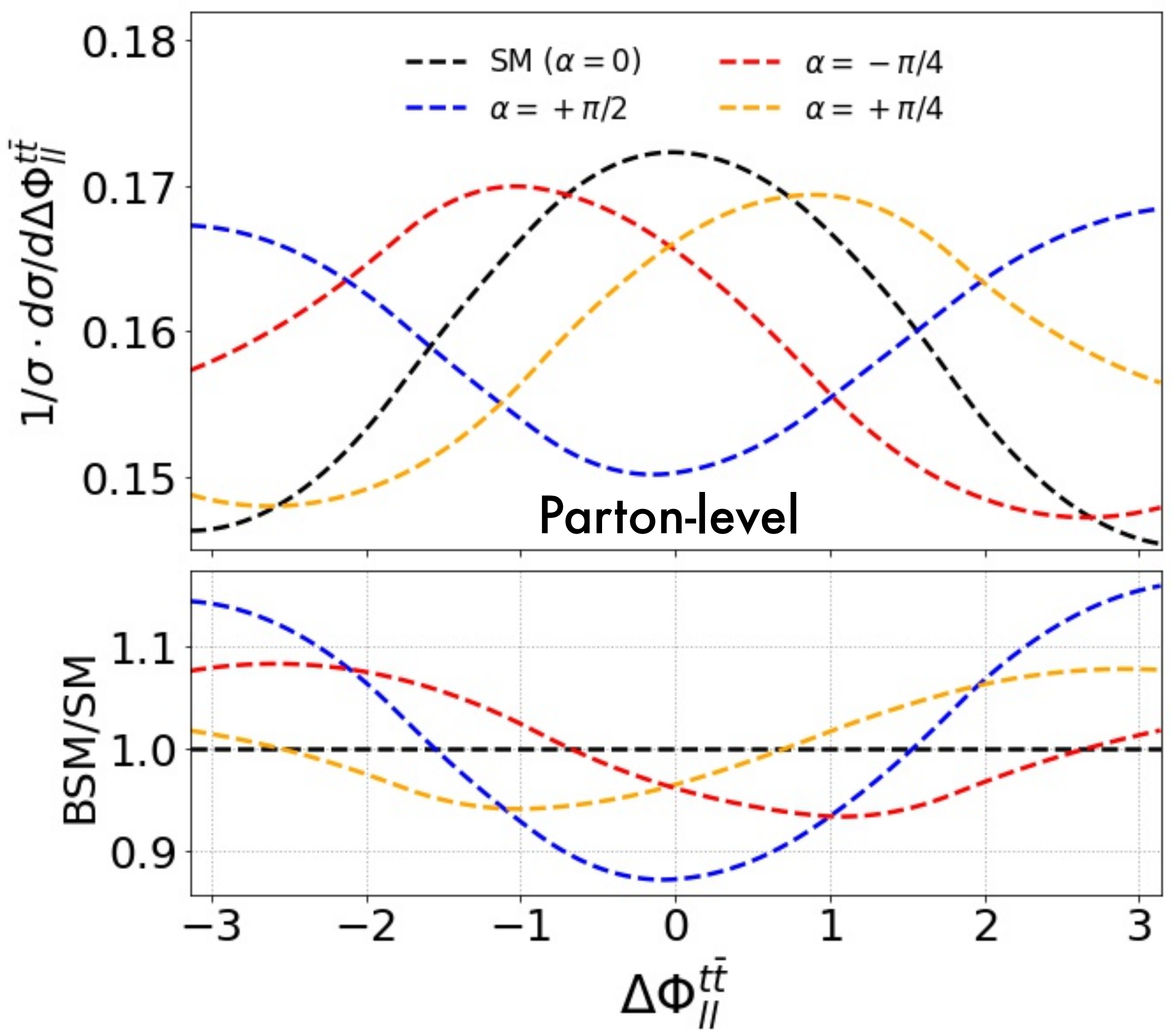
Sensitive only to non-linear new physics effects.

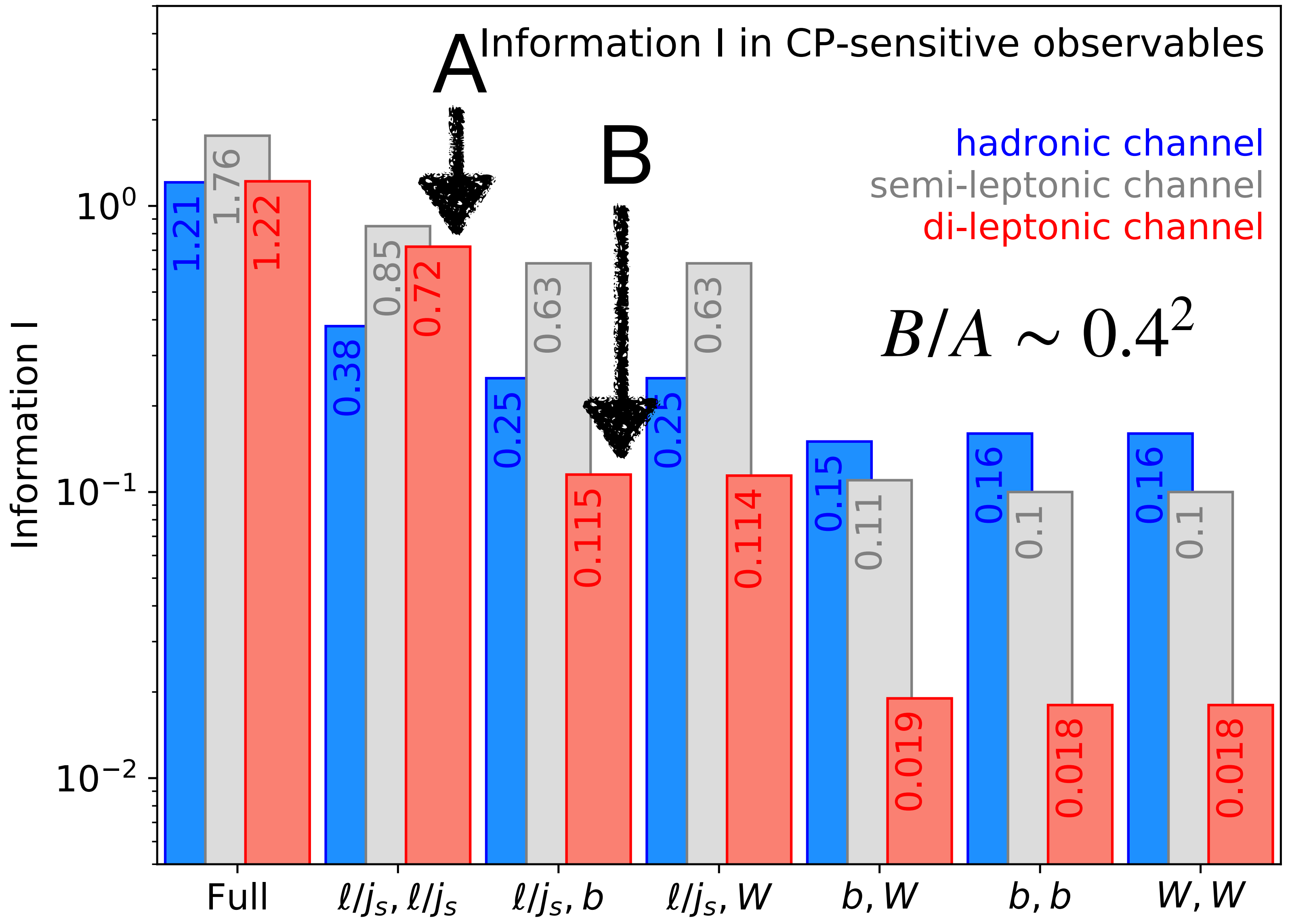
CP-odd observables

- Short lifetime for t $(10^{-25}\ s)$ → Spin correlations can be traced back from their decay products.

- CP-odd observables constructed from antisymmetric tensor products $\epsilon(p_t, p_{\bar{t}}, p_i, p_j) \sim \epsilon_{\mu\nu\rho\sigma} p_t^\mu p_{\bar{t}}^\nu p_i^\rho p_j^\sigma$:

$$\Delta\phi_{ij}^{t\bar{t}} = \mathbf{sgn}\left[\vec{p}_t \cdot (\vec{p}_i \times \vec{p}_j)\right] \arccos\left[\frac{\vec{p}_t \times \vec{p}_i}{|\vec{p}_t \times \vec{p}_i|} \cdot \frac{\vec{p}_t \times \vec{p}_j}{|\vec{p}_t \times \vec{p}_j|}\right]$$





[RKB, Goncalves, Kling (2021)]

← Spin correlations scale with the spin analysing power β_i.

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

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\xi_i} = \frac{1}{2}\left(1 + \beta_i P_t \cos\xi_i\right) \qquad \text{Fisher Info} = \mathbb{E}\left[\frac{\partial \log p(x\,|\,\kappa_t, \alpha)}{d\alpha}\frac{\partial \log p(x\,|\,\kappa_t, \alpha)}{d\alpha}\right]$$

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

$$p(x\,|\,\theta) = \frac{1}{\sigma(\theta)} \frac{d^d\sigma(x\,|\,\theta)}{dx^d}$$

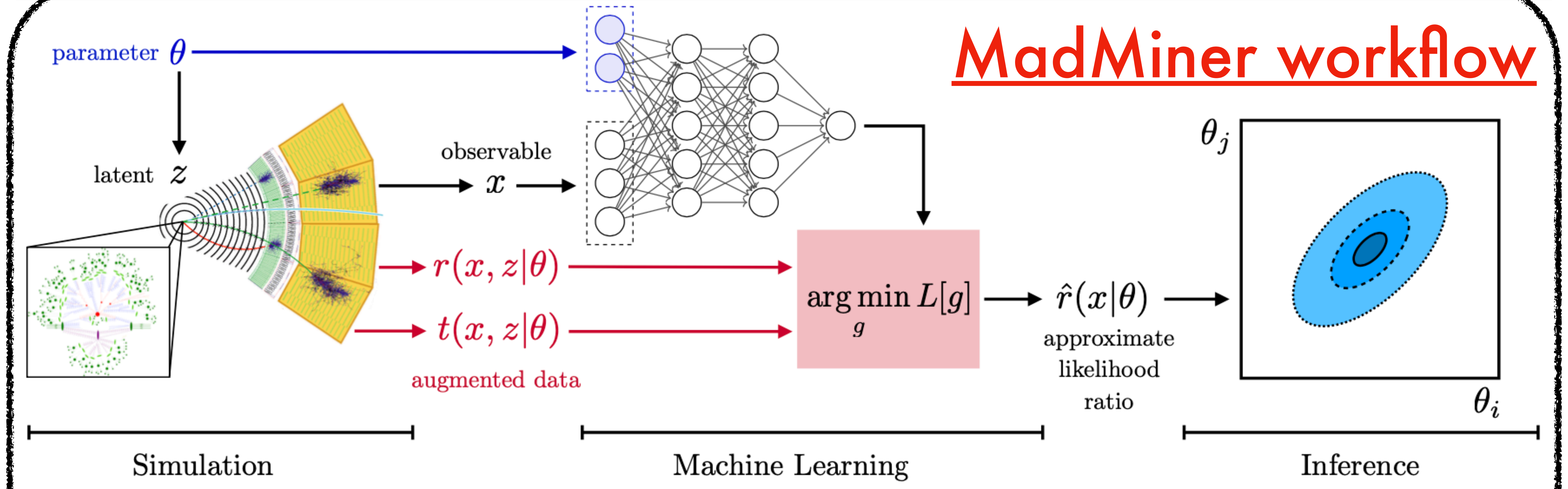
$$p(x\,|\,\theta) = \int dz_d \int dz_s \int dz_p \; \overbrace{p(x\,|\,z_d)\; p(z_d\,|\,z_s)\; p(z_s\,|\,z_p)\; p(z_p\,|\,\theta)}^{p(x|\theta)}$$

Latent parameters z : Detector response, Shower effects, Parton Level

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

$$r(x, z\,|\,\theta_1, \theta_0) \equiv \frac{p(x, z\,|\,\theta_1)}{p(x, z\,|\,\theta_0)} = \frac{d\sigma(z_p\,|\,\theta_0)\sigma(\theta_1)}{d\sigma(z_p\,|\,\theta_1)\sigma(\theta_0)}$$

MadMiner workflow



taken from [Brehmer, Cranner, Louppe, Pavez (2018)]

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

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

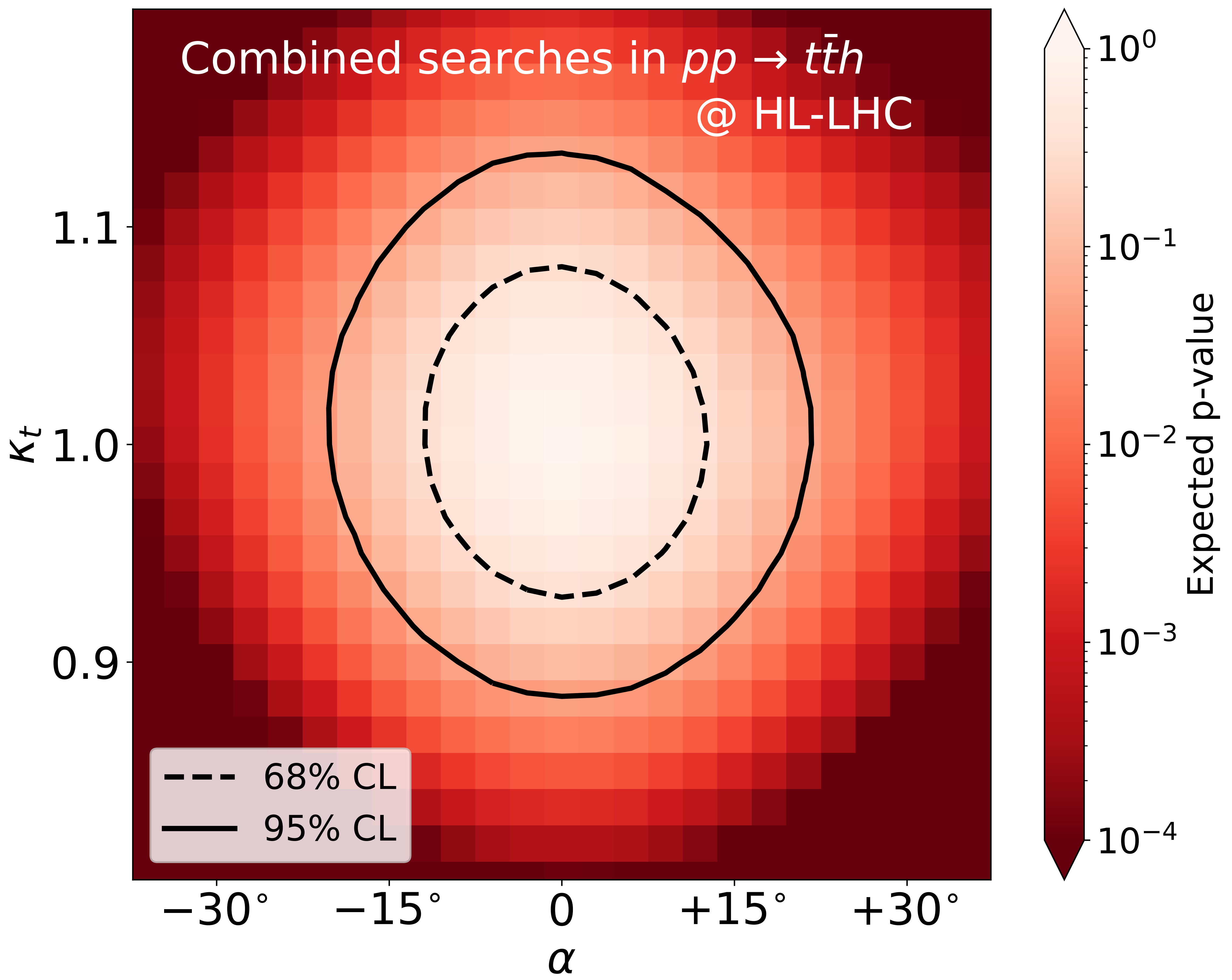
- The trained network is an estimator for $r(x\,|\,\theta, \theta_{SM})$

$t\bar{t}(h \to \gamma\gamma)$ @ HL-LHC

[RKB, Goncalves, Kling (2021)]

→ **Dominant sensitivity from CP-even observables.**

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



➡ Boost sensitivity through unfolding techniques.

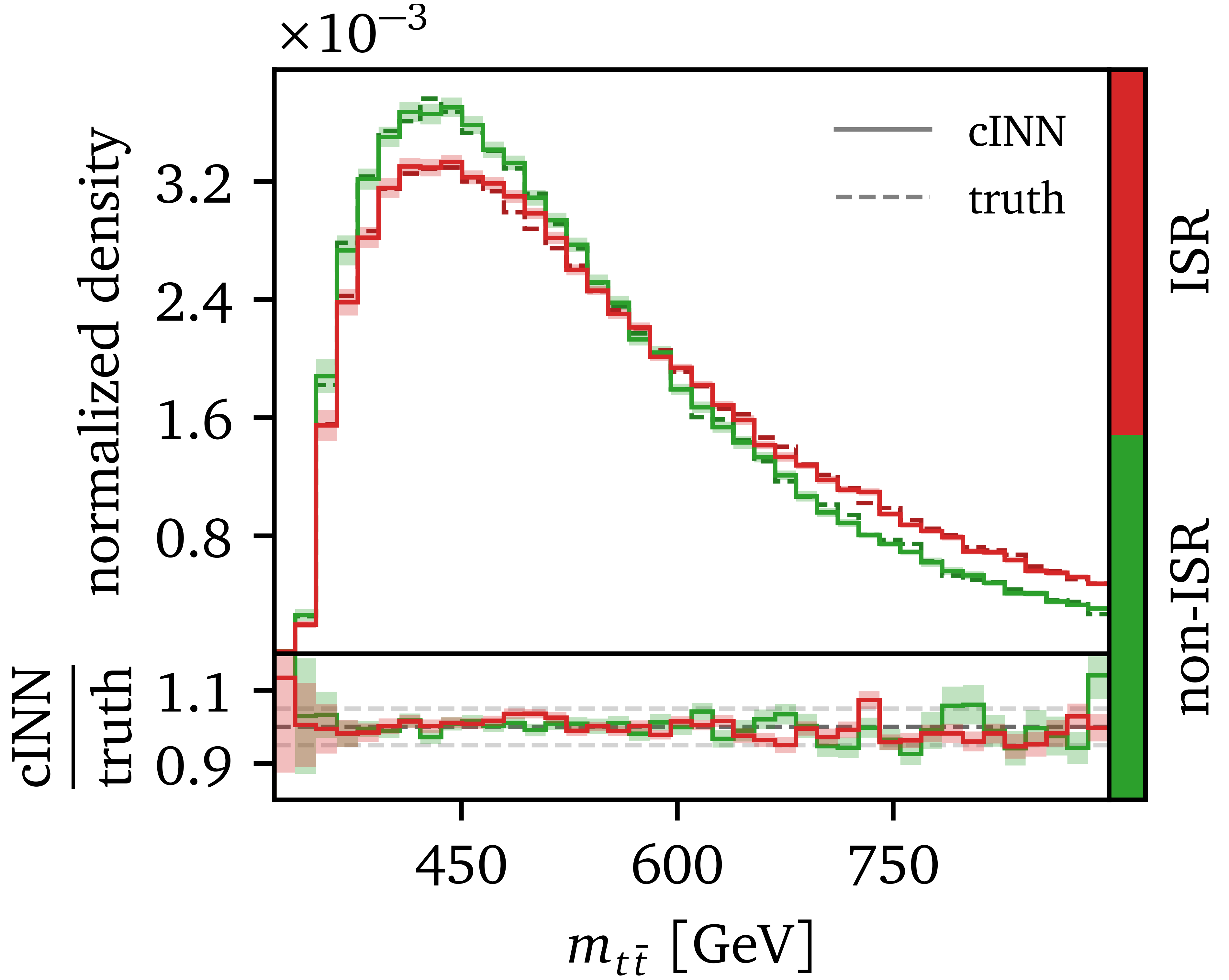
Back to results from unfolding with cINN…

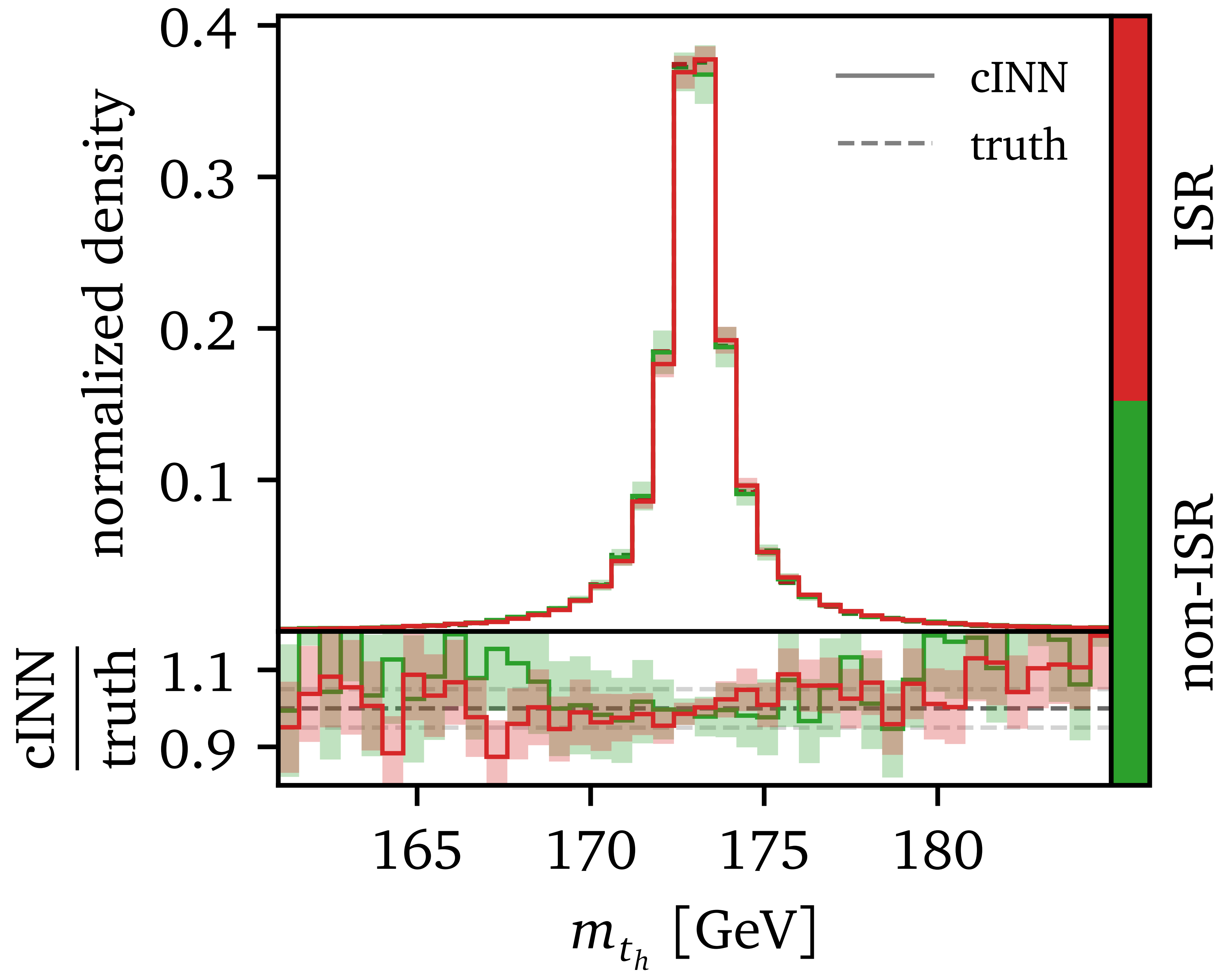
Challenges:

- ★ 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?

Jet combinatorics

Parton level truth and unfolded top invariant masses m_{t_ℓ} and m_{t_h}





★ Unfolded distributions in good agreement with parton level truth despite added combinatorial ambiguity at the detector level.

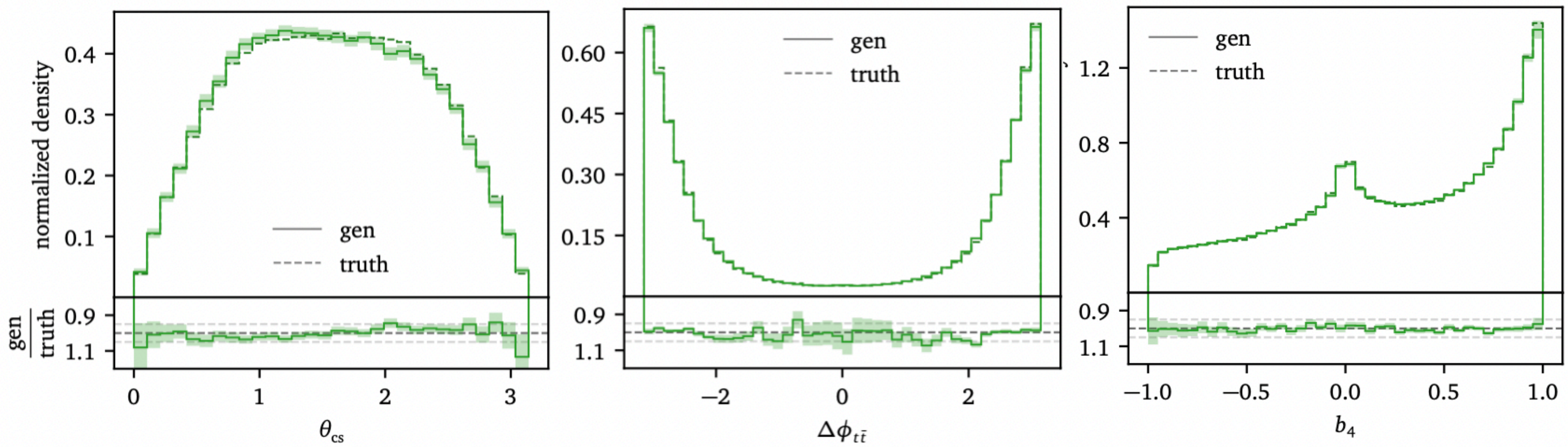
Back to results from unfolding with cINN…

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 θ_{CS}, $\Delta\phi_{t_\ell t_h}$ 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:

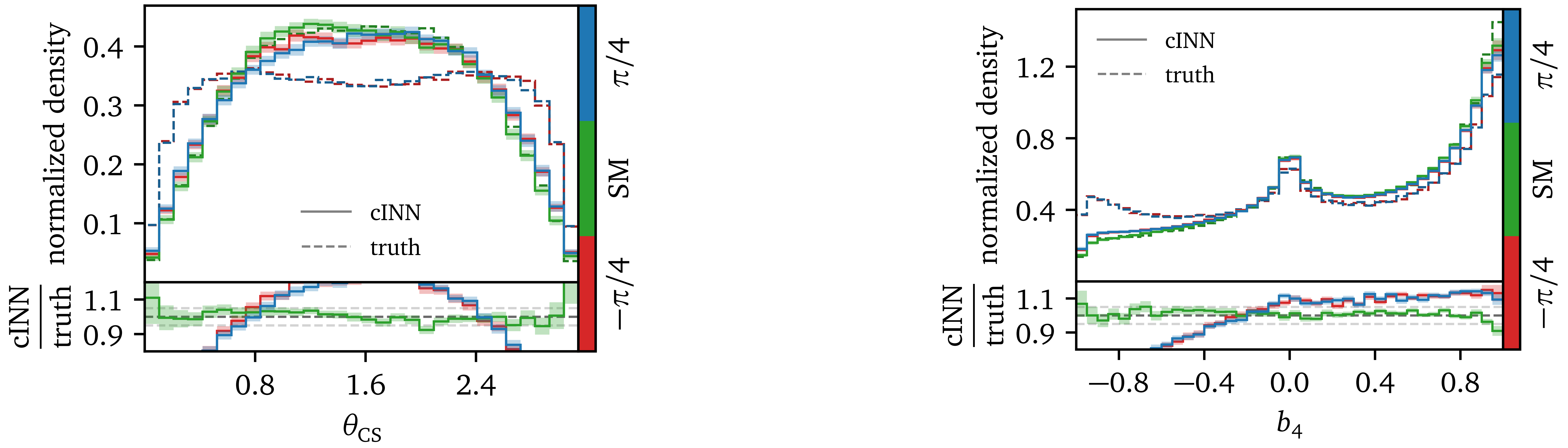
- ★ Can the unfolding model correctly reconstruct the two hard jets at the parton level from a variable number of jets at the detector level?
- ★ How well can the dedicated BSM observables be reconstructed?
- ★ How model-dependent is the training?

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



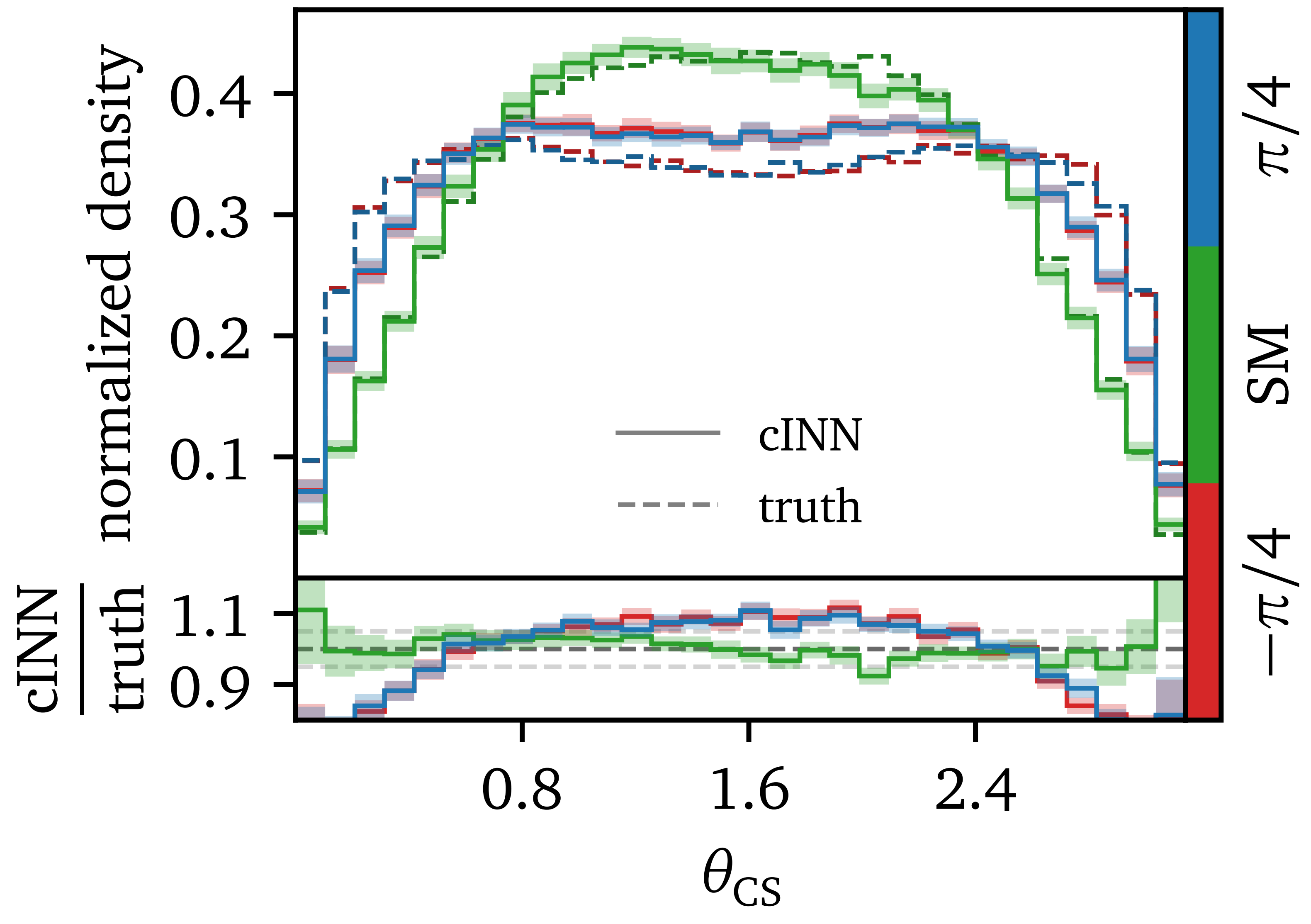
- ★ Networks trained on $\alpha = \pi/4$ and $-\pi/4$ show only a slight bias towards broader θ_{CS} and flatter b_4 distributions.
- ★ $\sim 10 - 20\,\%$ bias → much smaller than the changes at parton truth from varying α.

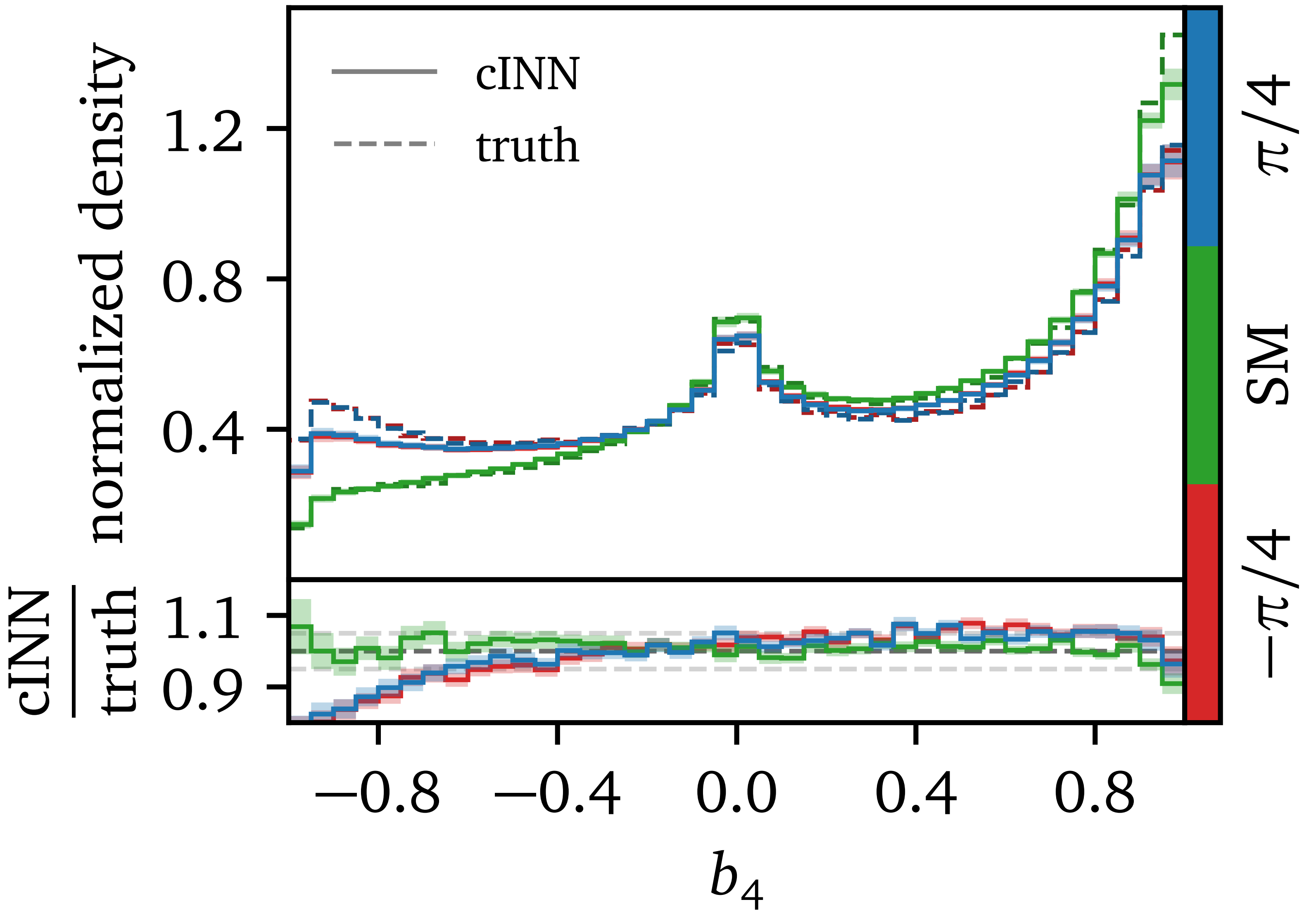
Model dependence

Unfolding events with CP-violation using a network trained on SM events.

Train network on SM dataset

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





★Again, the effect of bias is much smaller than the effect of α 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