Returning CP-observables to the frames they belong

Rahool Barman
Oklahoma State University

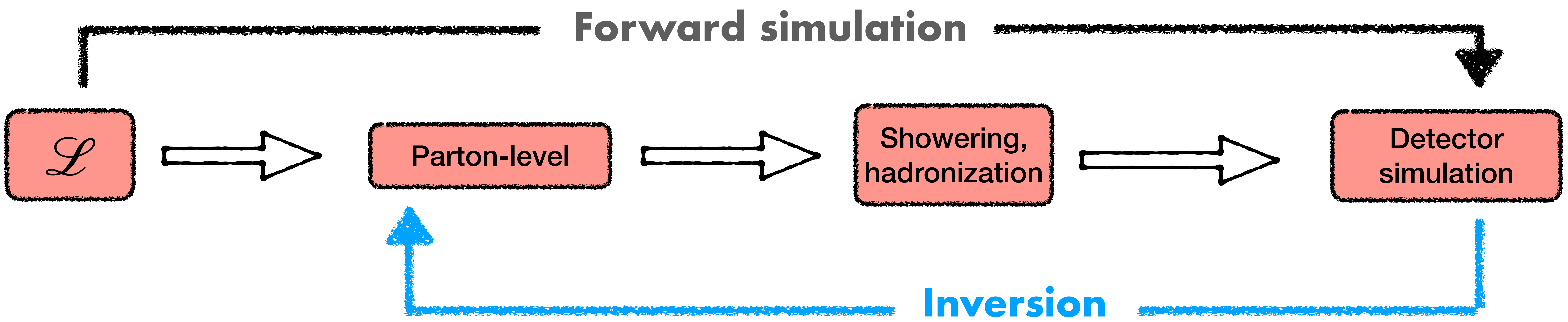
With
Jona Ackerschott, Dorival Gonçalves, Theo Heimel, Tilman Plehn

Based on
arXiv: 2308.00027

Particle Physics on the Plains
University of Kansas

October 14, 2023

Unfolding

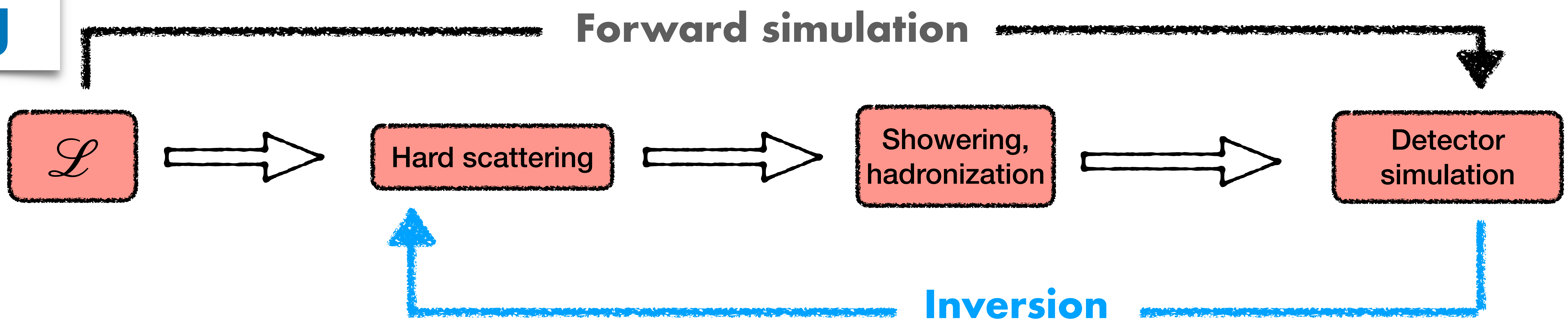


- Conventional LHC analysis involves comparing measured data with MC events simulated under NP hypothesis.
  - Reconstructed LHC events present a convoluted version of the true underlying physics.
  - Forward simulation chain can be highly resource intensive.

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

→ compare new physics hypotheses at the parton-level.

Unfolding



✦ Bin-independent

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

Possible with machine learning based generative models.

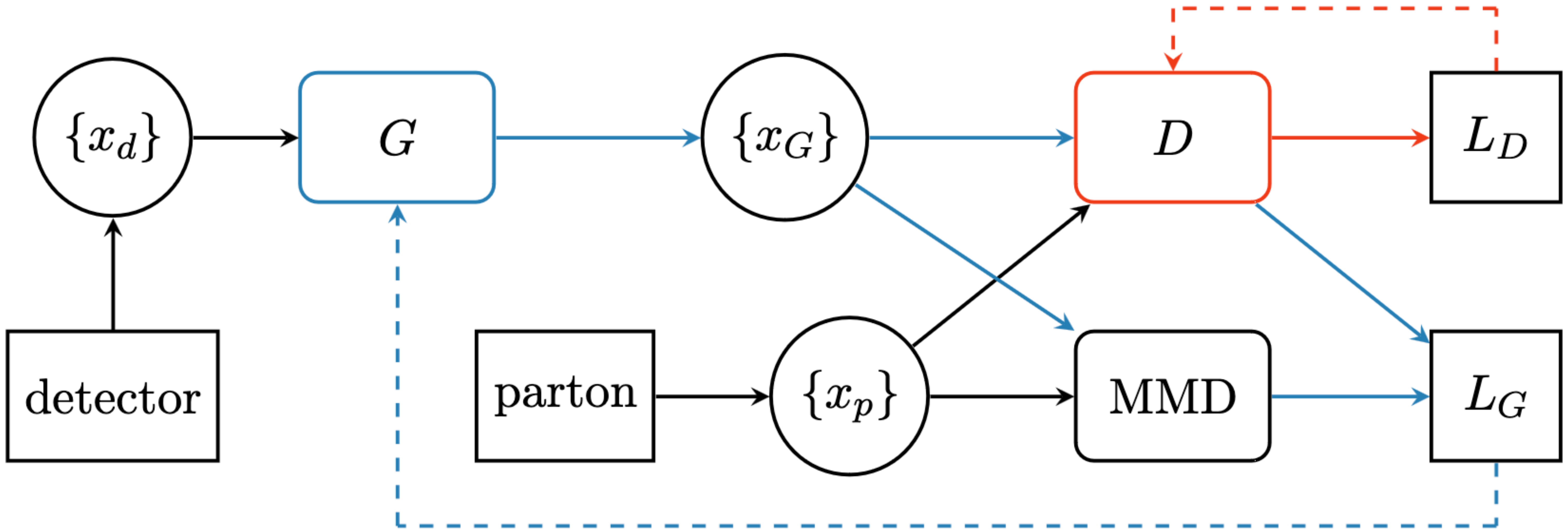
- Generative Adversarial Networks (GAN)
- Normalizing Flows (NF)
- Variational Auto Encoders (VAE)

[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)]

Generative Adversarial Network (GAN)

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

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



[Image 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 $D(x_G) \to 1$.

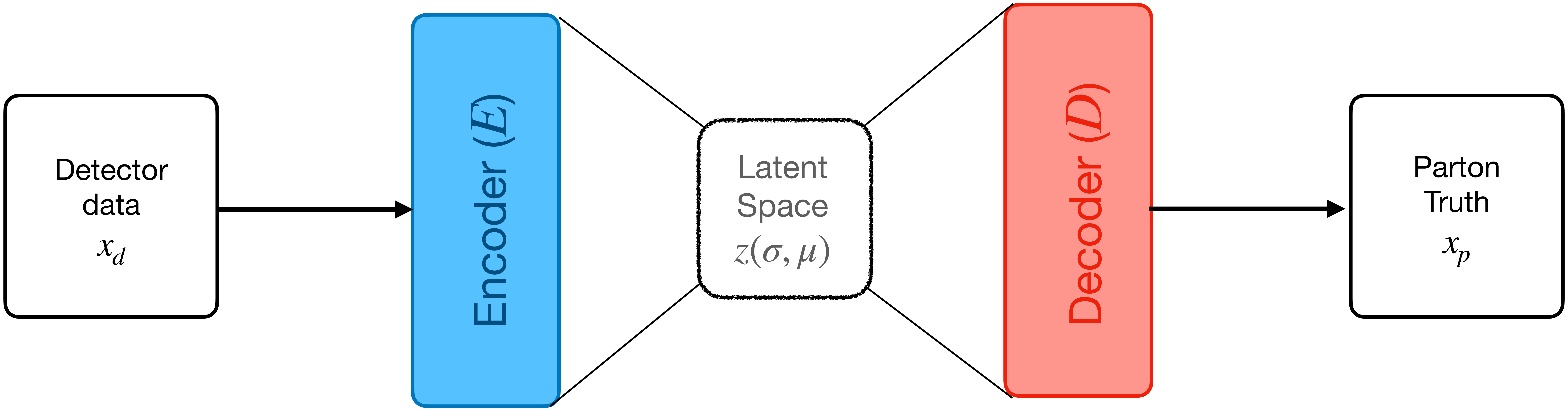
$$L_{\text{Discriminator}} = \langle -logD(x) \rangle_{x \sim P_p} + \langle -log(1 - D(x)) \rangle_{x \sim P_G}$$

$$L_{\text{Generator}} = \langle -logD(x) \rangle_{x \sim P_G}$$

Variational Auto Encoders

☑ Unfolding

→ Map detector data to the parton level phase space

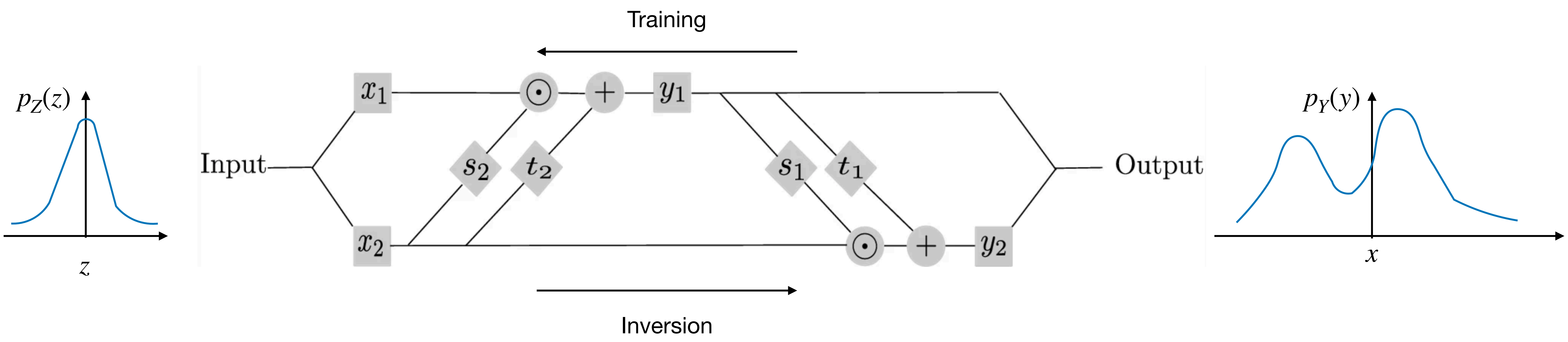


- The Encoder maps the input detector data d to a more tractable latent space $z = E(d)$ while preserving the essential features.
- The decoder maps z to the parton level $p' = D(z) = E(l(d))$.

[Otten, Caron, Swart, Beekveld, Hendriks, Leeuwen, Podareanu, Austri, Verheyen (2019)]

Normalizing flows

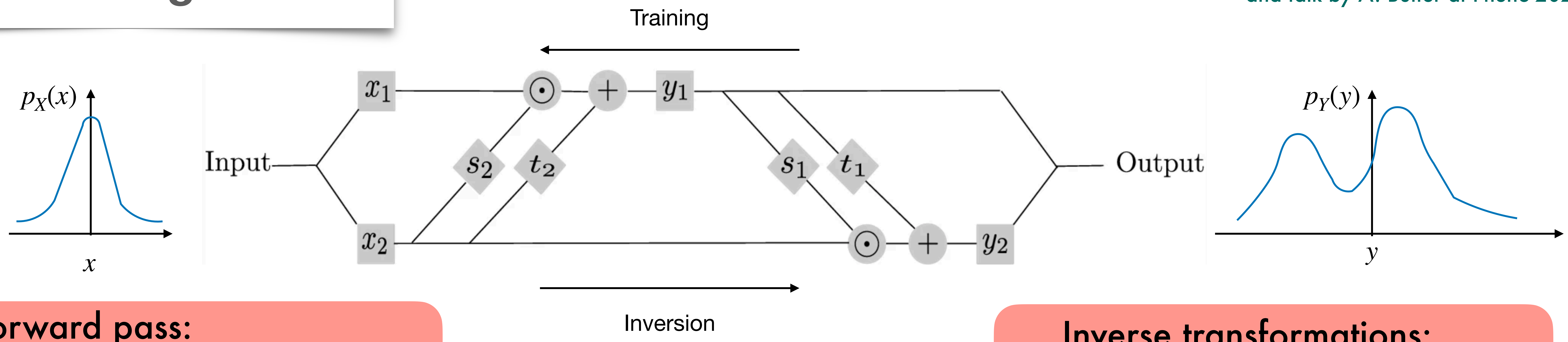
- Series of bijective layers that transform complex (Y) to simple probability distributions (Z).
- Learns both directions of the mapping in parallel → bijectivity encoded in the same network.
- Building blocks → Invertible coupling layers. [Dinh, Krueger, Bengio (2016), Dinh, Sohl-Dickstein, Bengio (2016)]



[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]



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

$$p_Y(x_d) = p_Z(x_p) \times |det(J_f(x_p))|^{-1}$$

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

☑ Sampling and density estimation

- Coupling layers stacked together → Invertible Neural Network (INN).

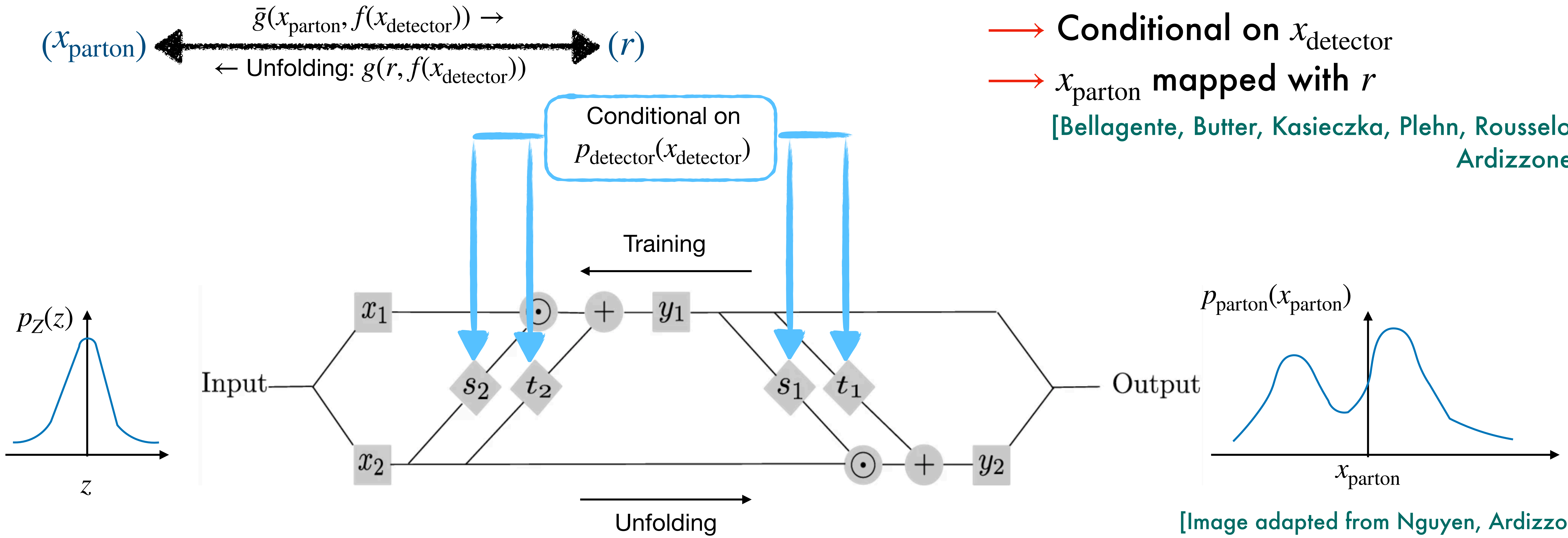
Conditional INN

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

⟶ Conditional on x_{detector}

⟶ 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]

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 → 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)]

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

(x_{parton}) $\bar{g}(x_{\text{parton}}, f(x_{\text{detector}})) \rightarrow$ (r)

$\leftarrow$ Unfolding: $g(r, f(x_{\text{detector}}))$

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

22 d.o.f.

Detector-level:

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

46 d.o.f.

A natural parametrization involving top mass:

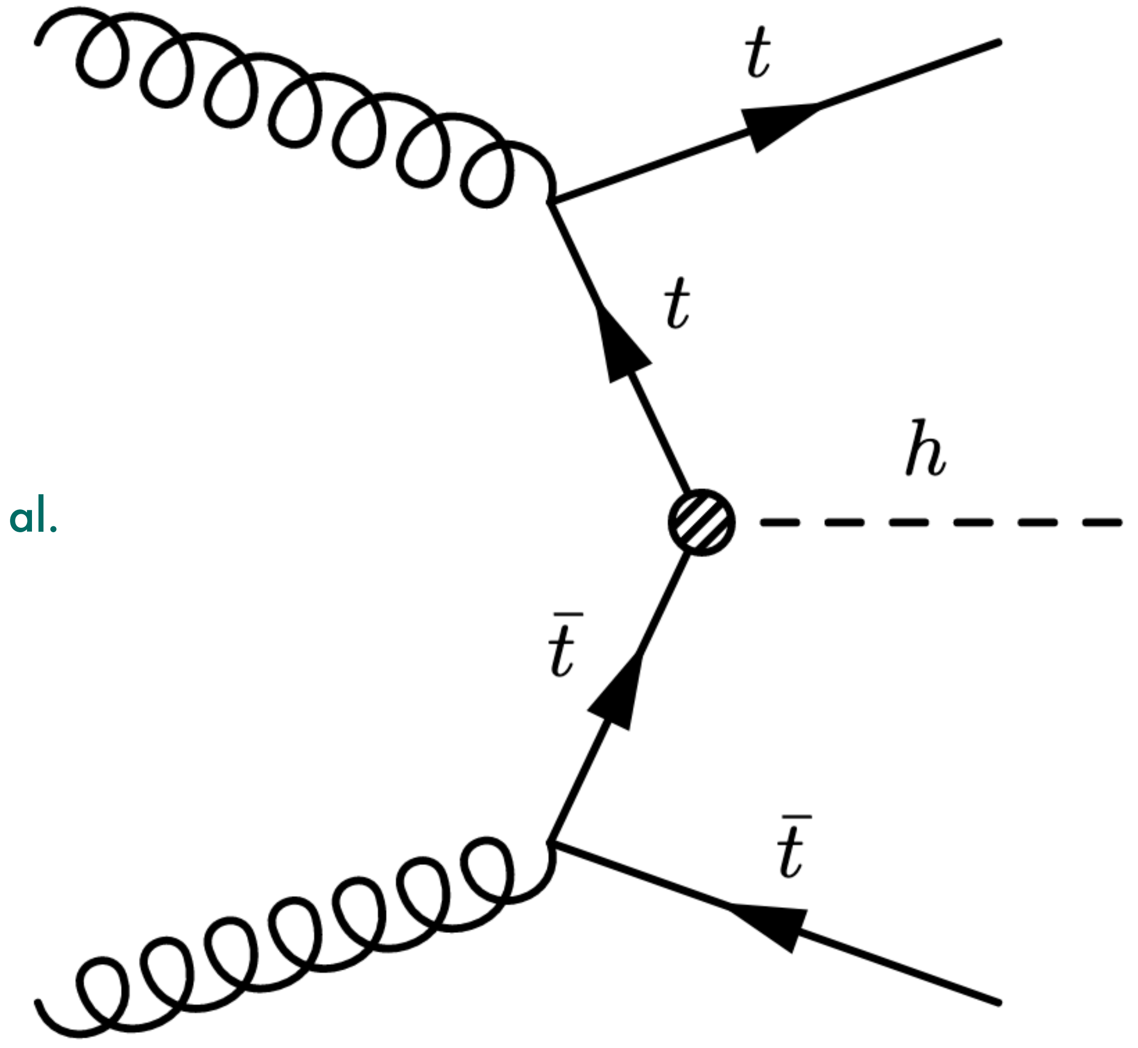
$$\left\{ m_t, p_{T,t}, \eta_t, \phi_t, m_W, \eta^t_W, \phi^t_W, \eta^W_{\ell,u}, \phi^W_{\ell,u} \right\}$$

- Redefine the parton level parametrization including the important CP observables

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

- We use the Bayesian version of cINN
  - Stable network predictions
  - Allows the estimation of training-related uncertainties.

CP measurement in Higgs-top interactions



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

- 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: $ht\bar{t}$

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

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

[Duca, Kilgore, Oleari, Schmidt, Zeppenfeld (2001), Klamke, Zeppenfeld (2007), Dolan, Harris, Jankowiak, Spannowsky (2014)]

Current limit (ATLAS: 2004.04545):
$|\alpha| < 43^0$ at 95 % CL

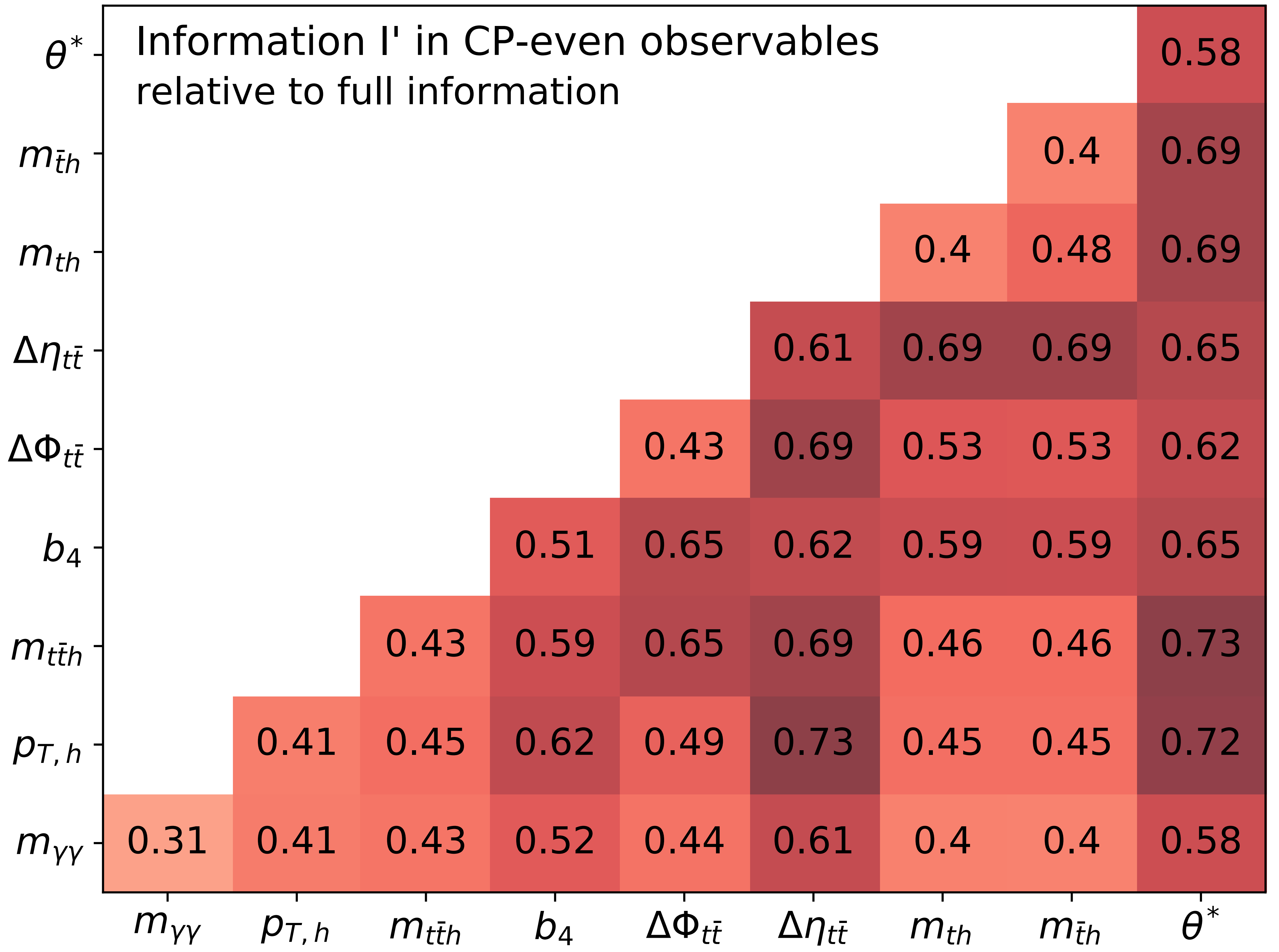
- $pp \rightarrow t\bar{t}h$: opportunity to directly probe α and κ_t

[Buckley, Goncalves (2016), Azevedo, Onofre, Filthaut, Goncalo (2017)]

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)]

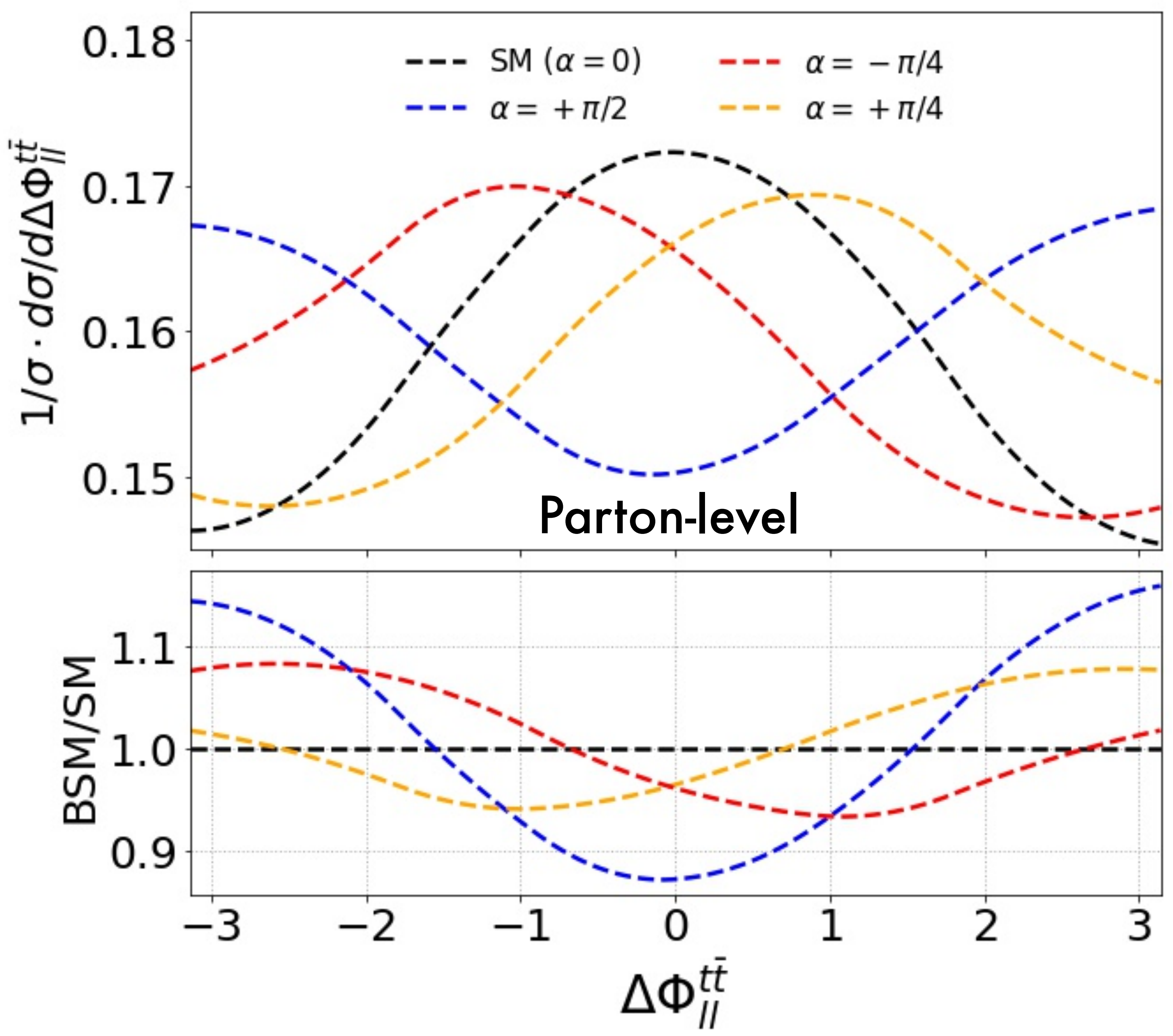
Sensitive only to non-linear new physics effects.

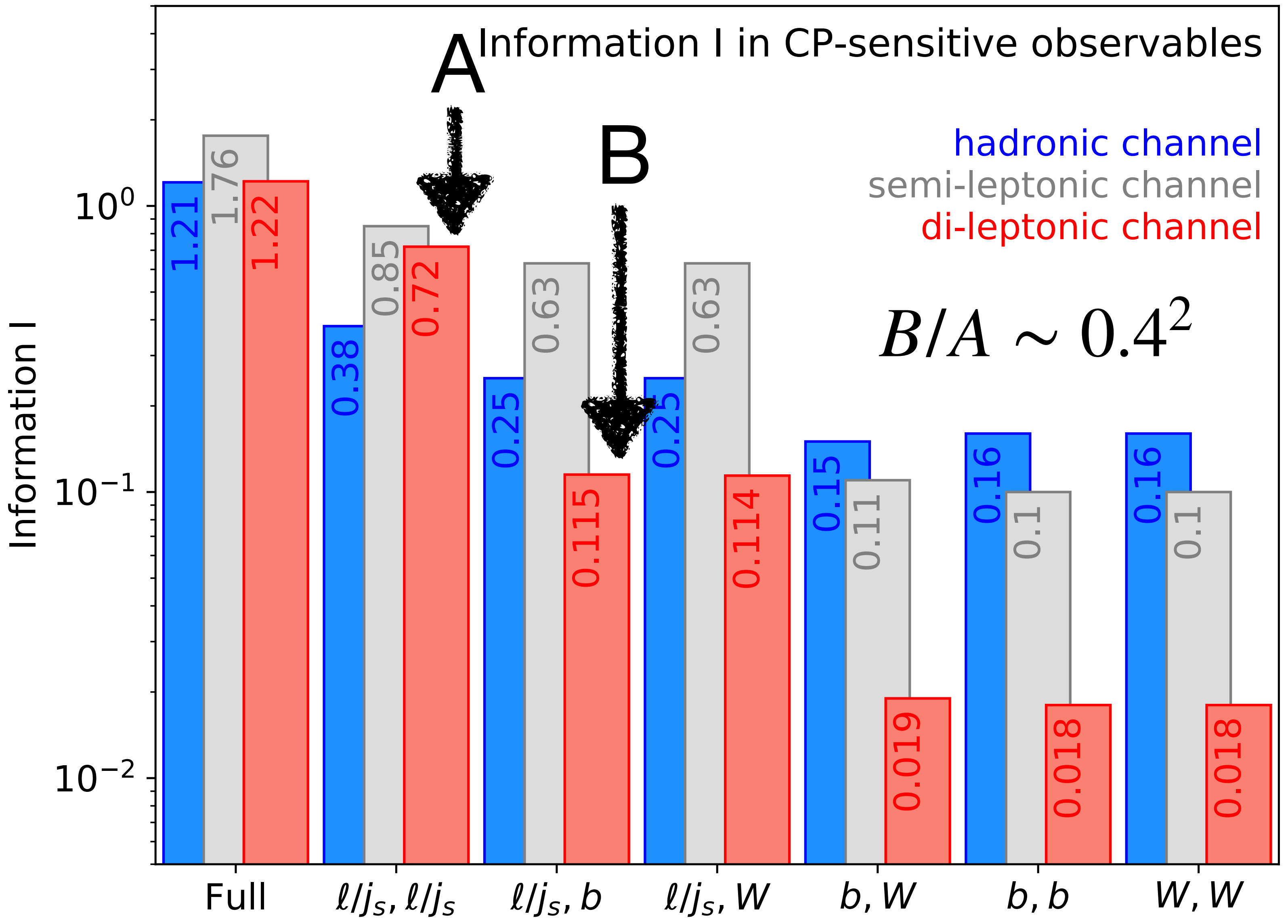
CP-odd observables

- 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
  $\epsilon(p_t, p_{\bar{t}}, p_i, p_j) \sim \epsilon_{\mu\nu\rho\sigma} p_t^\mu p_{\bar{t}}^\nu p_i^\rho p_j^\sigma$:

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





[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)]

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\xi_i} = \frac{1}{2}\left(1 + \beta_i P_t \cos\xi_i\right)$$

$$\text{Fisher Info} = \mathbb{E}\left[\frac{\partial \log p(x\,|\,\kappa_t, \alpha)}{d\alpha}\frac{\partial \log p(x\,|\,\kappa_t, \alpha)}{d\alpha}\right]$$

- Kinematic reconstruction efficiency is limited at the detector level

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

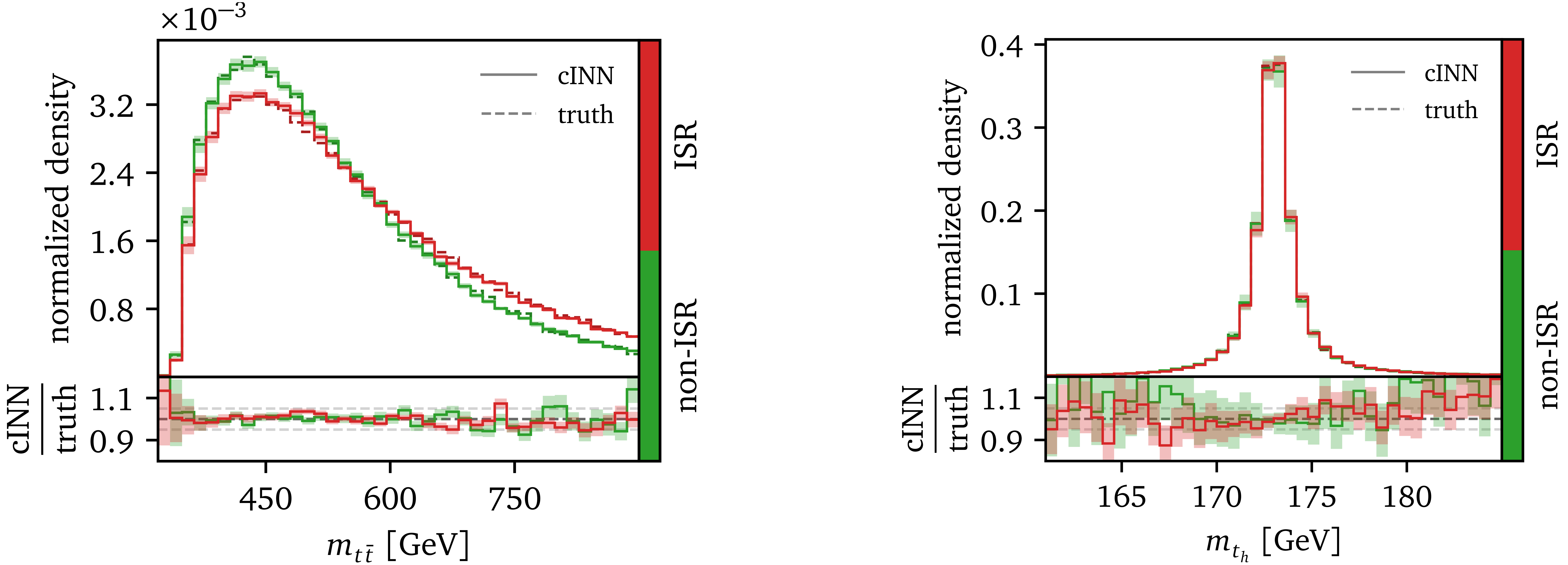
Back to results from unfolding with cINN...

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

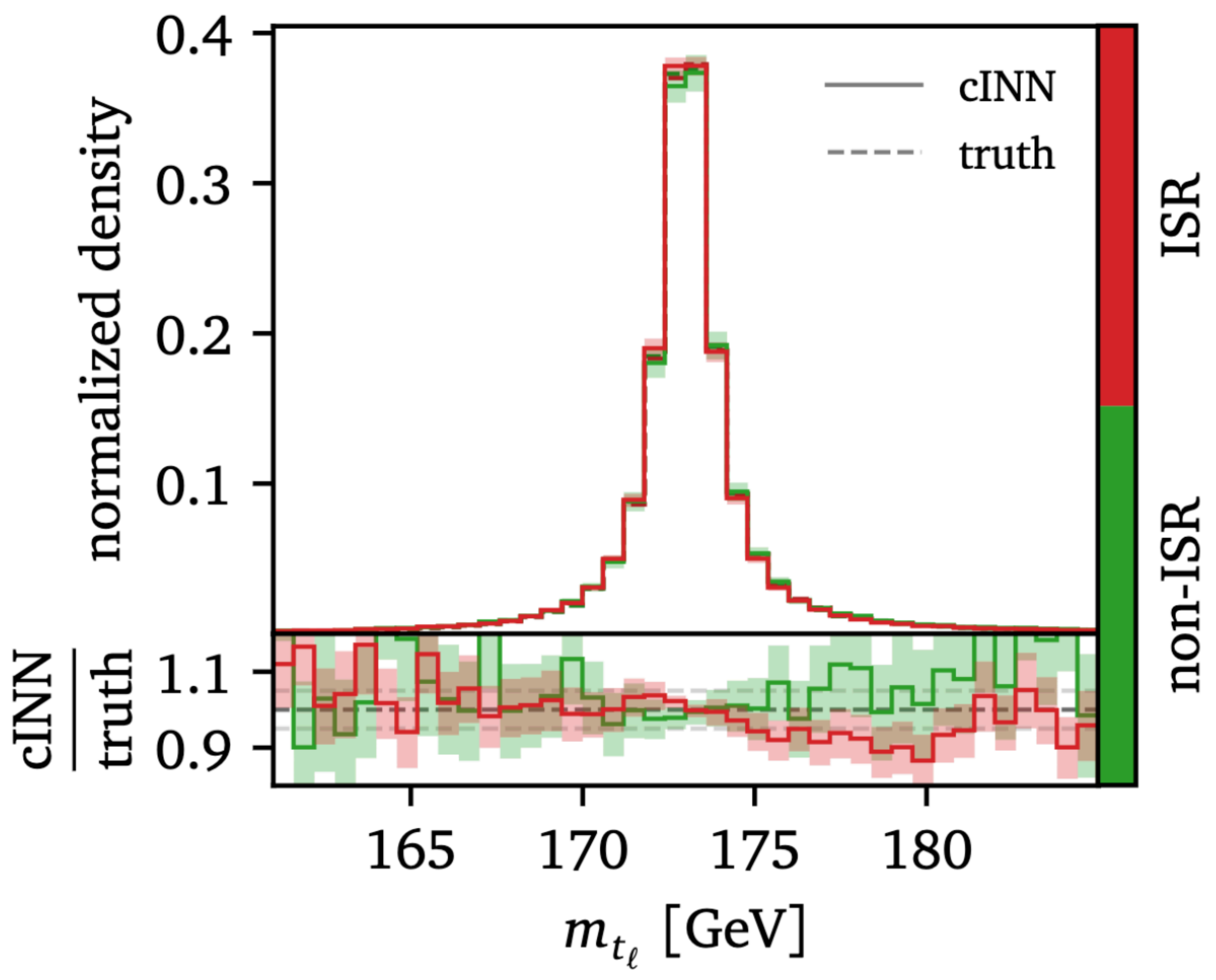
Parton level truth and unfolded top invariant masses $m_{t\bar{t}}$ and m_{t_h}



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

Leptonic top reconstruction

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



★ Unfolded distributions in good agreement with parton level truth despite missing d.o.f. (longitudinal neutrino momemtum) at the detector level.

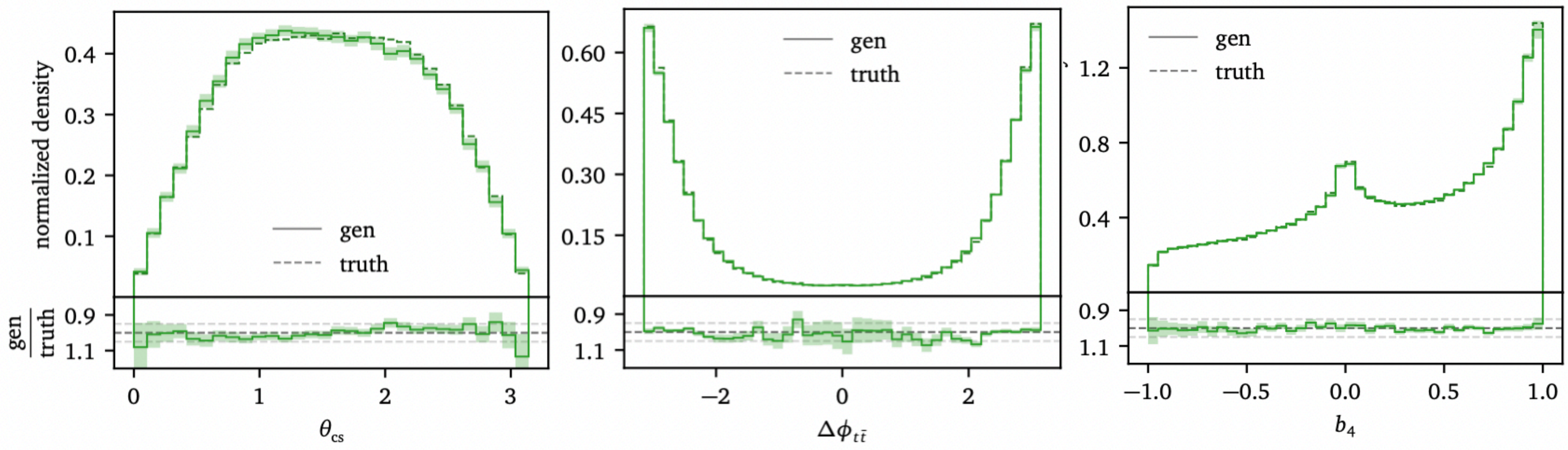
Back to results from unfolding with cINN…

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?

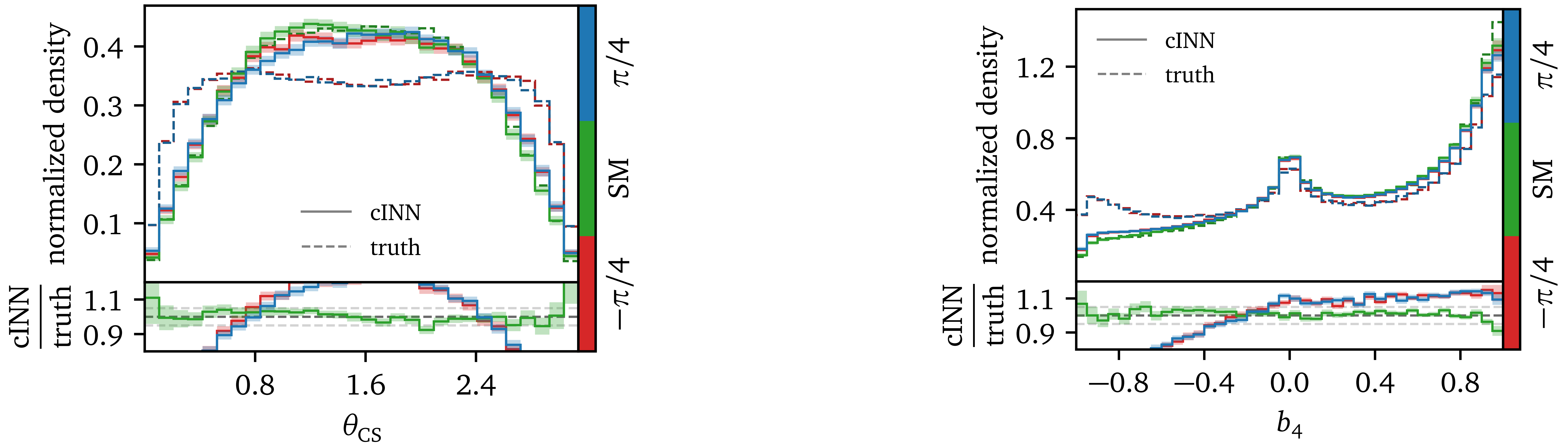
★ How well can the dedicated BSM observables be reconstructed?

★ 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.

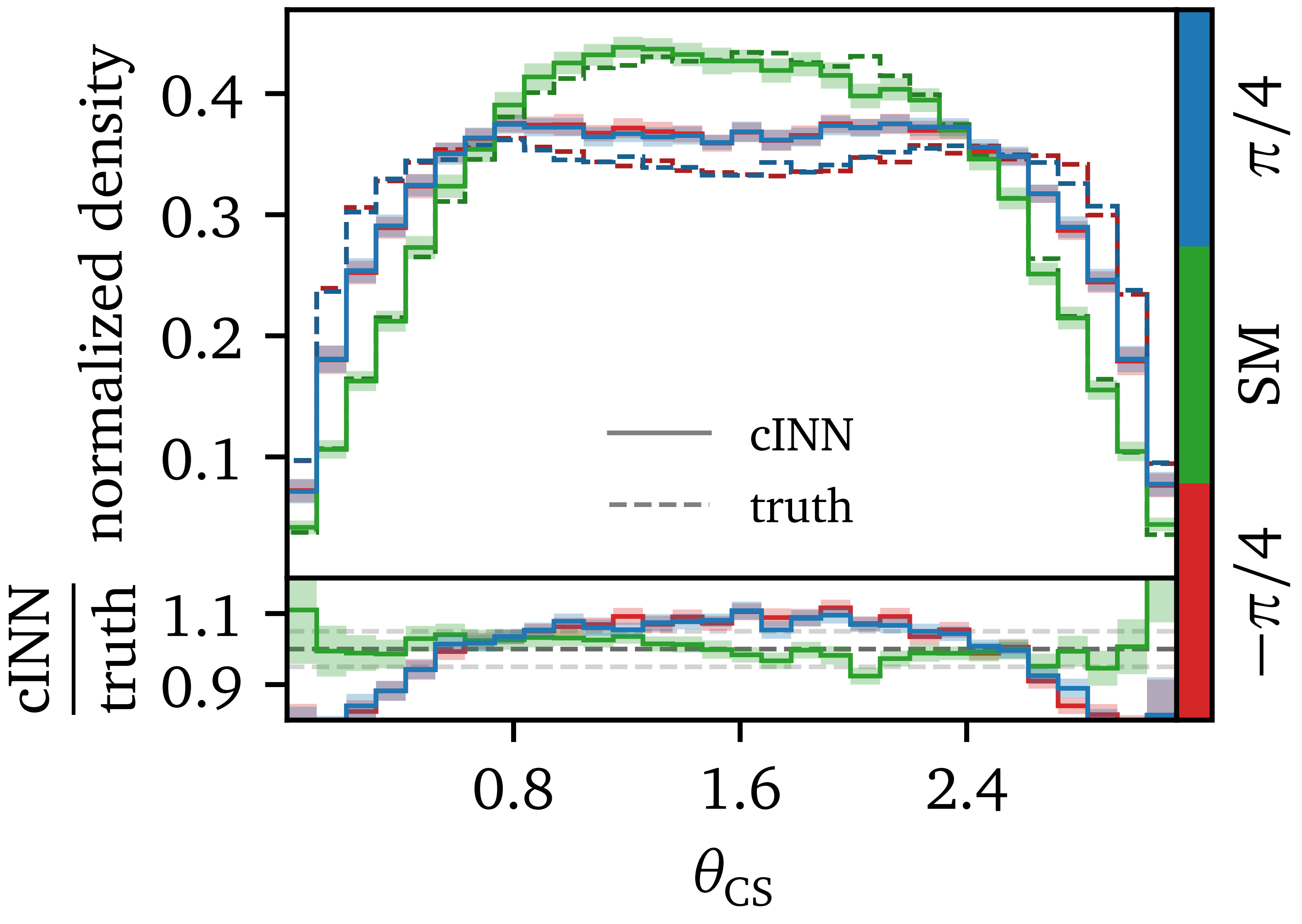
★ $\sim 10 - 20\,\%$ bias → much smaller than the changes at parton truth from varying α.

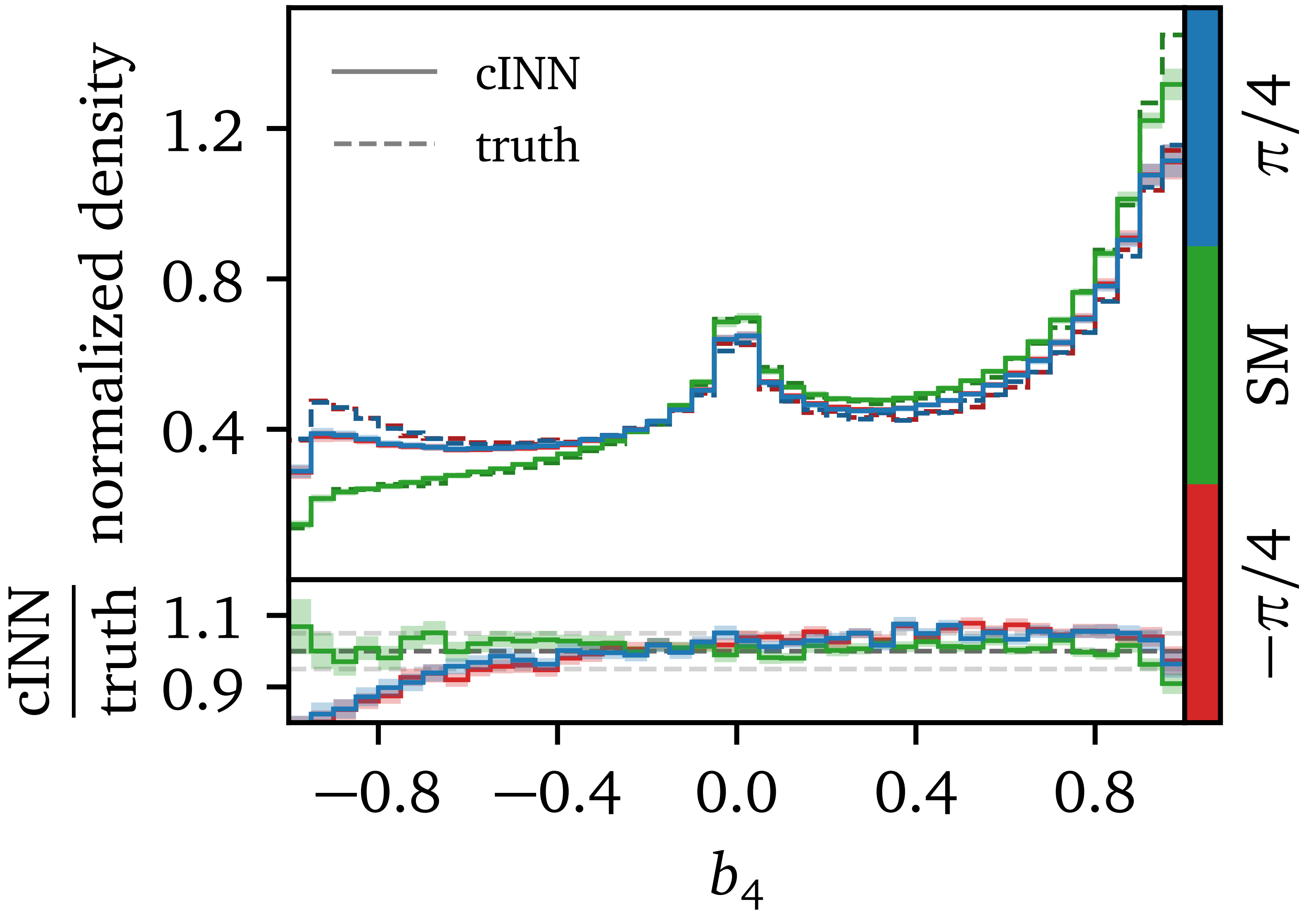
Model dependence

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.

    - Extract various CP observables.
    - Resolve jet combinatorial ambiguity.
    - Avoid large model-dependence.

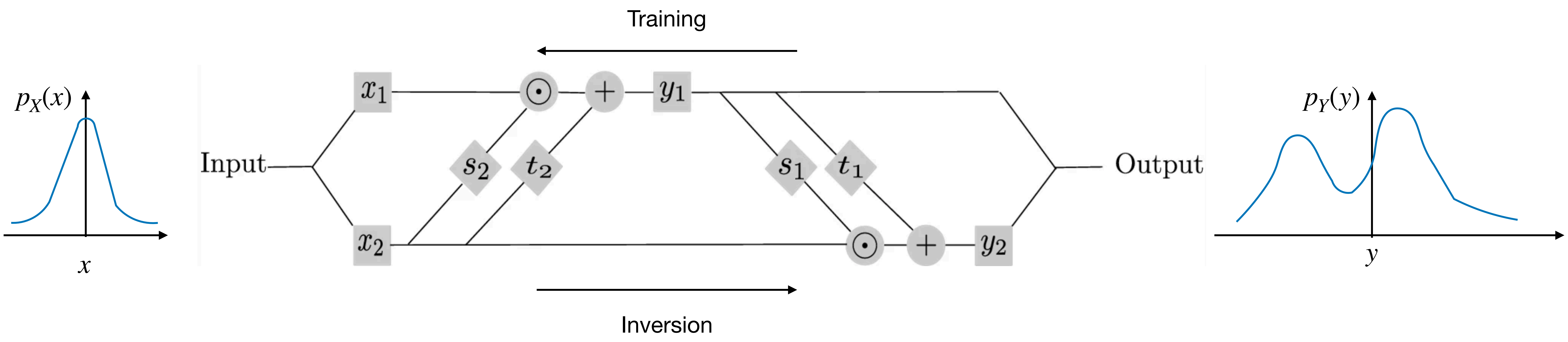
- Presents a promising outlook for an experimental study, with a proper treatment of statistical limitations, continuum backgrounds, calibration, and iterative improvements of the unfolding network.

Thank you

Backup slides

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

☑ 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.
- GANs 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.