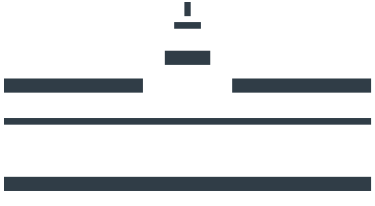
Learning Off-Shell Effects in Top-Pair Production with Direct Diffusion Neural Networks

Tomáš Ježo

Institute of Theoretical Physics
University of Münster

based on: [**TJ**, Nason '15], [**TJ**, Lindert, Nason, Oleari, Pozzorini '16], [Ferrario Ravasio, **TJ**, Nason, Oleari '18], [Ferrario Ravasio, **TJ**, Nason, Oleari '19], [Herwig, **TJ**, Nachman '19], [Ferrario Ravasio, **TJ** '21], [**TJ**, Lindert, Pozzorini '23], [Butter, **TJ**, Klasen, Kuschick, Palacios Schweitzer, Plehn '23] [Kuschick '24]



Top LHC France 2025
30/04/25

Learning Off-Shell Effects in Top-Pair Production with Direct Diffusion Neural Networks

- Motivations:
  - Why top quark? Because it's a versatile probe of the SM and a window to NP.
    - a.) Coloured object that b.) decays electroweakly and c.) couples strongly to the Higgs boson
  - Why top quark at LHC? Because "several hundred million tops produced" ...
    - ...implies theory will soon lag behind the experiment.
    - ...means it is major background in many other LHC analyses.

Precise simulations of top quark production and decay at LHC imperative!

Correspondingly we have: NLO QCD, NNLO QCD, NLO EW, NNLO QCD+NLO EW, analytic resummations, NLO QCD+PS and NNLO QCD+PS

Learning Off-Shell Effects in Top-Pair Production with Direct Diffusion Neural Networks

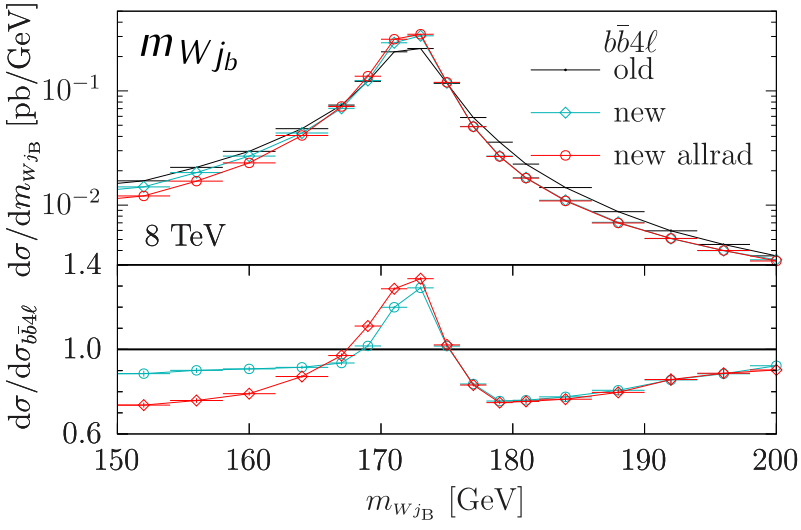
- Motivations:
  - Why top quark? Because it's a versatile probe of the SM and a window to NP.
    - a.) Coloured object that b.) decays electroweakly and c.) couples strongly to the Higgs boson
  - Why top quark at LHC? Because "several hundred million tops produced" ...
    - ...implies theory will soon lag behind the experiment.
    - ...means it is major background in many other LHC analyses.
  - But do we also need off-shell effects?
    - They modify shapes of spectra used for measurements of top properties,
    - and allow the inclusion of quantum interferences between different production modes and radiation from production and decay

There is: NLO QCD, NLO EW and NLO QCD+PS in the dileptonic channel

Do we need off-shell effects?

Do we need off-shell effects?

- Off-shell effects distort the top mass shape and other distributions



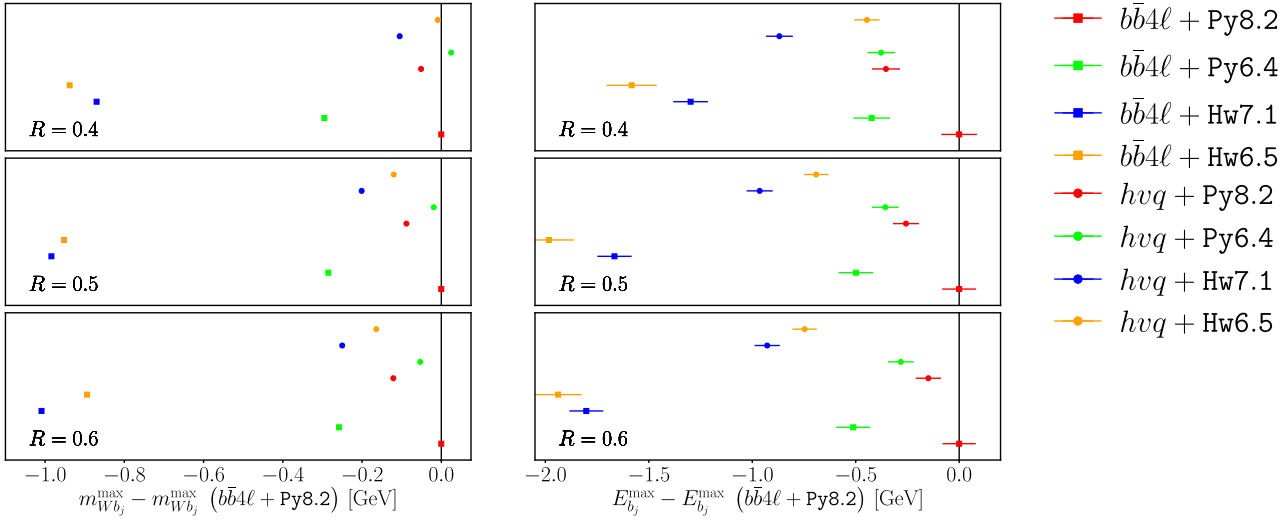
[**TJ**, Nason '15],
[**TJ**, Lindert, Nason, Oleari, Pozzorini '16]

  - Potentially affecting m_t and y_t measurements

- Proper treatment of interference required
  - To describe the data
  - And if you have it, you can try measuring Γ_t in tails

Do we need off-shell effects?

- Off-shell effects distort the top mass shape and other distributions
  - Potentially affecting m_t and y_t measurements

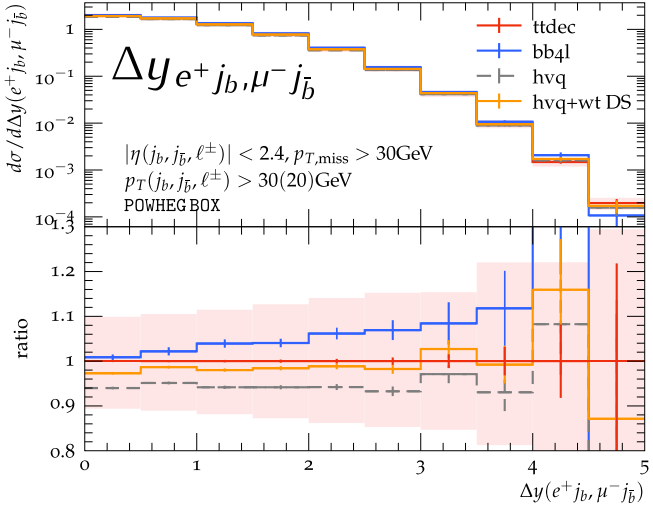


[Ferrario Ravasio, **TJ**, Nason, Oleari '18, '19], [ATL-PHYS-PUB-2021-042]

- Proper treatment of interference required
  - To describe the data
  - And if you have it, you can try measuring Γ_t in tails

Do we need off-shell effects?

- Off-shell effects distort the top mass shape and other distributions
  - Potentially affecting m_t and y_t measurements

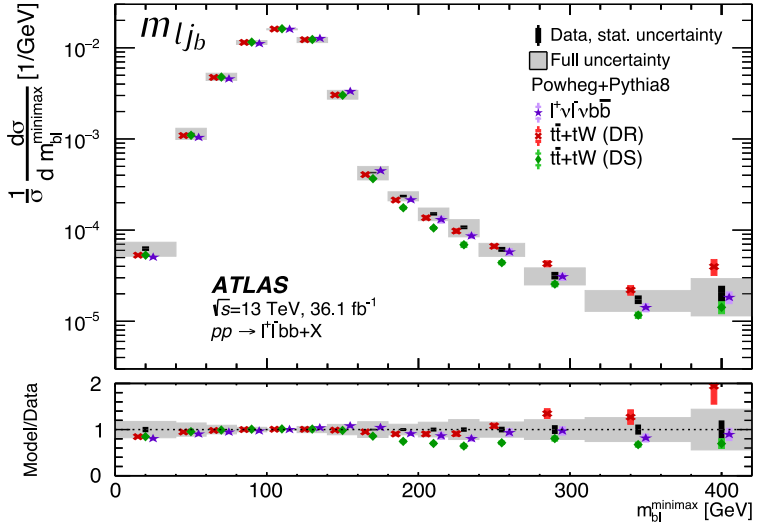


[Ferrario Ravasio, **TJ** '21]

- Proper treatment of interference required
  - To describe the data
  - And if you have it, you can try measuring Γ_t in tails

Do we need off-shell effects?

- Off-shell effects distort the spectra used for measurements of top properties
  - Potentially affecting m_t and y_t measurements
- Proper treatment of interference required
  - To describe the data

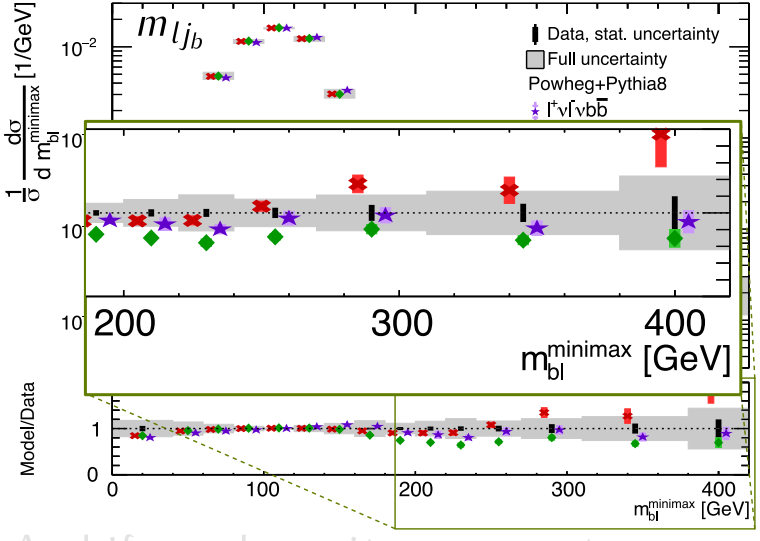


[PRL 121, 152002]

  - And if you have it, you can try measuring Γ_t in tails

Do we need off-shell effects?

- Off-shell effects distort the spectra used for measurements of top properties
  - Potentially affecting m_t and y_t measurements
- Proper treatment of interference required
  - To describe the data

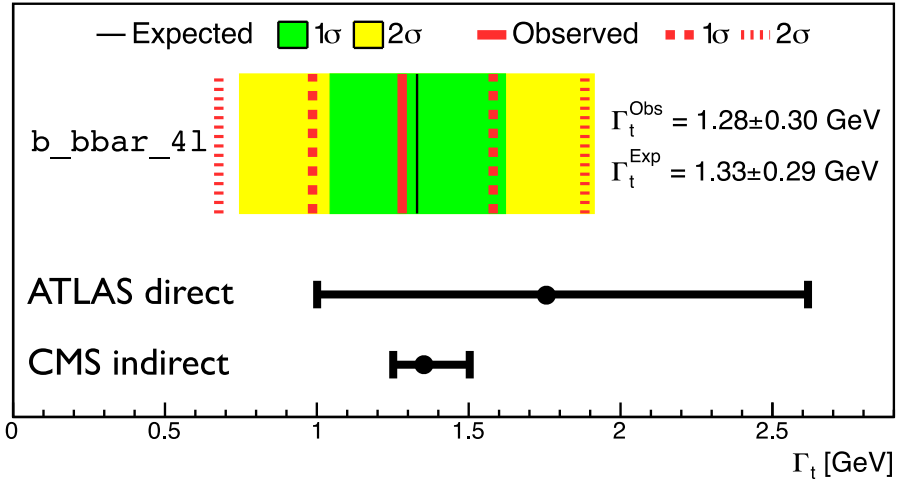


[PRL 121, 152002]

  - And if you have it, you can try measuring Γ_t in tails

Do we need off-shell effects?

- Off-shell effects distort the spectra used for measurements of top properties
  - Potentially affecting m_t and y_t measurements
- Proper treatment of interference required
  - To describe the data
  - And if you have it, you can try measuring Γ_t in tails



[Herwig, **TJ**, Nachman '19]

bb4l: $t\bar{t}$ off-shell

bb4l: $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ @ NLO+PS

[TJ et al. '15, '16, '23]

- We published a MC event generator `POWHEG BOX RES/bb4l`
  - Implementing process $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ up to $\mathcal{O}(\alpha_S^2 \alpha^4 \times \alpha_S)$, l, ℓ different
  - ME in 4FNS ($m_b > 0$) but 5FNS PDFs also possible (CGN '98 matching)
  - Matching to PS using the resonance-aware version of the POWHEG method
- Two important developments
  - POWHEG style matching for processes with resonances possible
  - Modelling of emission in the decay with exact matrix element

Traditional NLOPS

ME

Multiple-radiation-improved NLOPS (`allrad`)

ME

bb4l: $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ @ NLO+PS

[TJ et al. '15, '16, '23]

- We published a MC event generator `POWHEG BOX RES/bb4l`
  - Implementing process $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ up to $\mathcal{O}(\alpha_S^2 \alpha^4 \times \alpha_S)$, l, ℓ different
  - ME in 4FNS ($m_b > 0$) but 5FNS PDFs also possible (CGN '98 matching)
  - Matching to PS using the resonance-aware version of the POWHEG method
- Two important developments
  - POWHEG style matching for processes with resonances possible
  - Modelling of emission in the decay with exact matrix element

Traditional NLOPS

ME

Multiple-radiation-improved NLOPS (`allrad`)

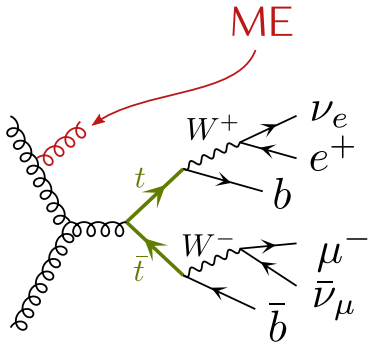
ME

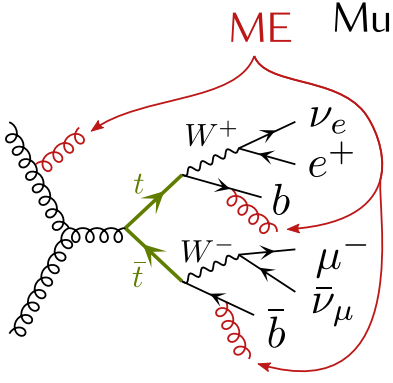
bb4l: $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ @ NLO+PS

[TJ et al. '15, '16, '23]

- We published a MC event generator `POWHEG BOX RES/bb4l`
  - Implementing process $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ up to $\mathcal{O}(\alpha_S^2 \alpha^4 \times \alpha_S)$, l, ℓ different
  - ME in 4FNS ($m_b > 0$) but 5FNS PDFs also possible (CGN '98 matching)
  - Matching to PS using the resonance-aware version of the POWHEG method
- Two important developments
  - POWHEG style matching for processes with resonances possible
  - Modelling of emission in the decay with exact matrix element

Traditional NLOPS





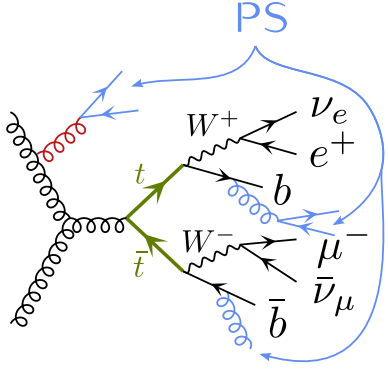
Multiple-radiation-improved NLOPS (`allrad`)

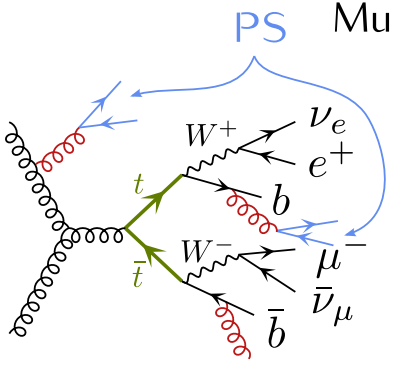
bb4l: $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ @ NLO+PS

[TJ et al. '15, '16, '23]

- We published a MC event generator `POWHEG BOX RES/bb4l`
  - Implementing process $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ up to $\mathcal{O}(\alpha_S^2 \alpha^4 \times \alpha_S)$, l, ℓ different
  - ME in 4FNS ($m_b > 0$) but 5FNS PDFs also possible (CGN '98 matching)
  - Matching to PS using the resonance-aware version of the POWHEG method
- Two important developments
  - POWHEG style matching for processes with resonances possible
  - Modelling of emission in the decay with exact matrix element

Traditional NLOPS





Multiple-radiation-improved NLOPS (`allrad`)

bb4l: $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ @ NLO+PS

[TJ et al. '15, '16, '23]

- We published a MC event generator `POWHEG BOX RES/bb4l`
  - Implementing process $pp \to l^+ \nu_l \ell^- \bar{\nu}_\ell b\bar{b}$ up to $\mathcal{O}(\alpha_S^2 \alpha^4 \times \alpha_S)$, l, ℓ different
  - ME in 4FNS ($m_b > 0$) but 5FNS PDFs also possible (CGN '98 matching)
  - Matching to PS using the resonance-aware version of the POWHEG method
- We have done it ... ... but it is very computationaly intensive!
  - Genuine calculation with $2 \to 6(7)$ external legs at LO(NLO)
  - Warm-up stage: 32 cd (core days) to reach ~0.5% on inclusive σ
  - Even generation stage: 3 evt/s (~1700 cd for 50M events)
  - Note: no uncertainties (scale, pdf, matching, ...), unweighting up to virtuals
- Limitations
  - Too computationally costly to "play with"
  - Publishing the "full" output impractical
  - Validation of samples too time consuming

ML off-shell effects

Learning off-shell effects with NNs

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Can machine learning help?
  - Train a NN that generates events directly
  - Make integration and unweighting methods more efficient
  - Learn matrix elements
  - Transform samples using reweight
- Our idea: transform samples beyond reweighting
  - Take advangate of existing or inexpensive samples (`POWHEG/hvq`)
  - Train a NN that transforms it into the costly sample
- Our goal: publish "full output" (= inexpensive sample run card + NN)
  - To play with, to benchmark, to validate samples against, etc.
  - Part of a hybrid calculation which aims at higher formal accuracy

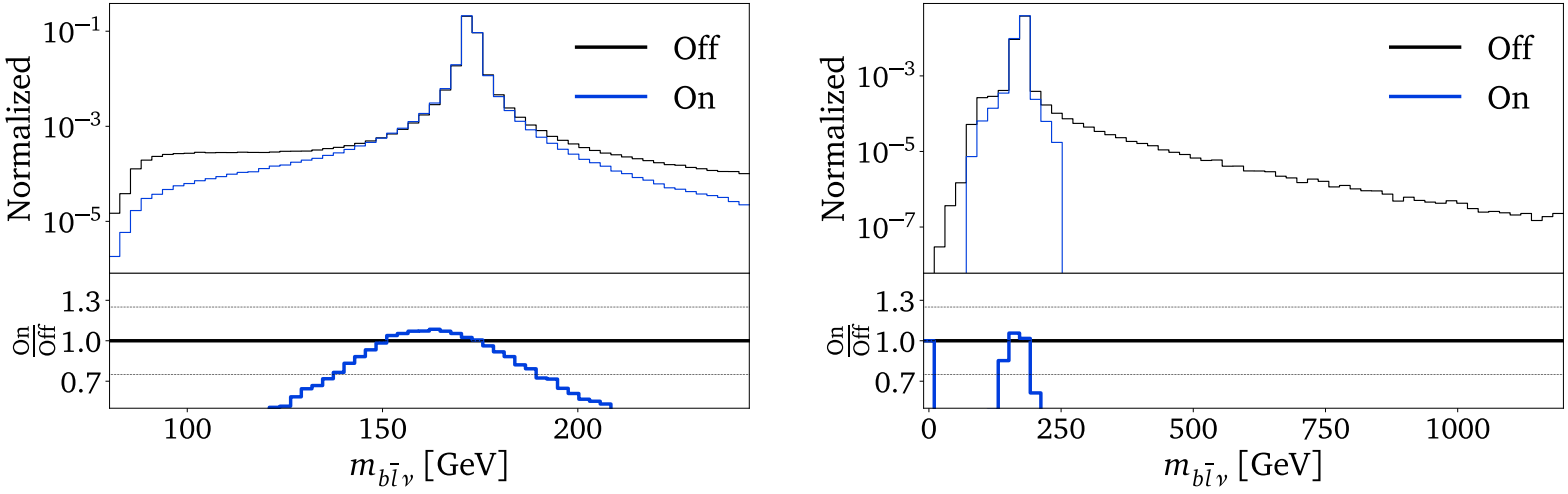
Learning off-shell effects with NNs

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Conceptually:
  - We are after a recipe to transform one sample into another, event by event
  - Goal sample: events with full off-shell effects (`POWHEG/bb4l`)
  - Starting sample: events with approximate off-shell effects (`POWHEG/hvq`)
- Technically:
  - "Unshowered" parton level events (`LHE` format) on input and output
  - Going beyond reweighting is crucial, due to incomplete phase-space coverage
  - Use direct diffusion based neural network supplemented by a classifier
    - Requires samples of same dimensionality
- Important points to address:
  - Could parton shower be included?
  - Detector simulation is still the most costly component

The phase-space population issue

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

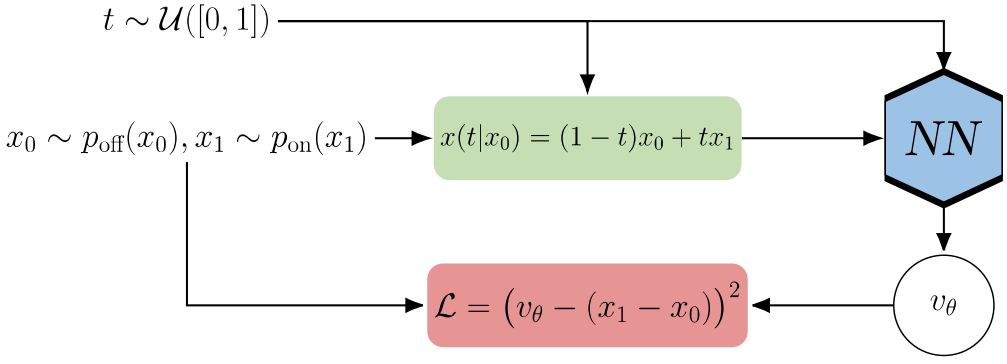


- Reconstructed top mass with approximate vs. full off-shell effects:
  - —— On `POWHEG/hvq` generator with approximate top decay, no radiative corrections in decay, only finite-width effects (+ approximate spin correlations)
  - —— Off `POWHEG/bb4l` generator with full off-shell top decay, including corrections in decay, finite-width, non-resonant and interference effects

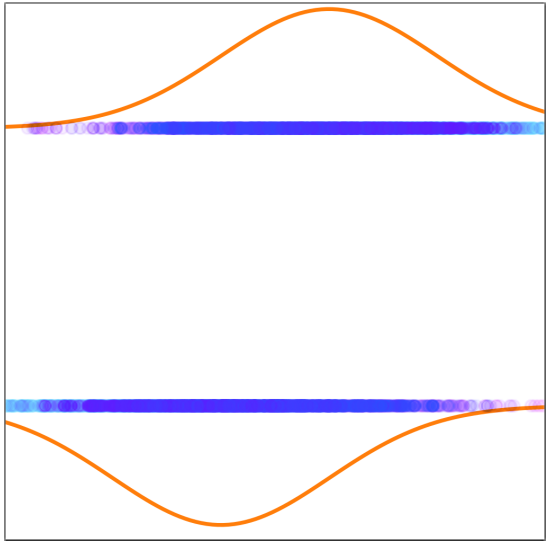
The direct diffusion neural network

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Off-Shell event $x_{off}(t=0)=x_0$, on-shell events $x_{on}(t=1)=x_1$ respectively



[arXiv:2209.15571, arXiv:2210.02747, arXiv:2209.03003, arXiv:2305.10475v2]

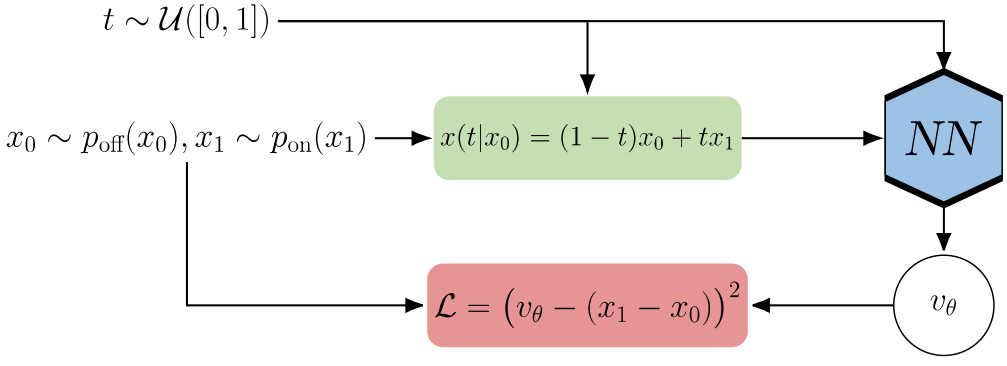


† slide credit M. Kuschick

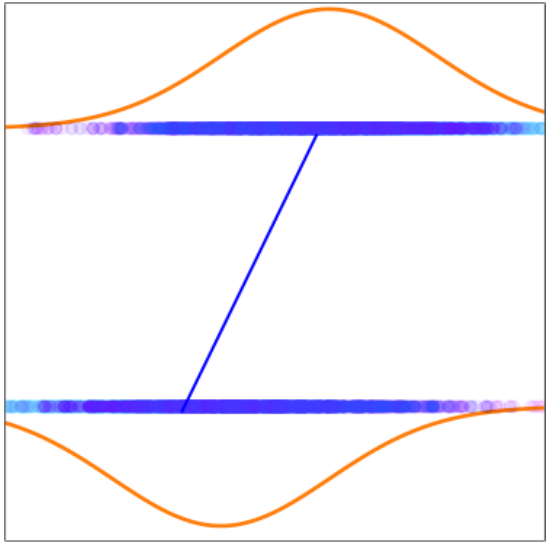
The direct diffusion neural network

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Off-Shell event $x_{off}(t=0)=x_0$, on-shell events $x_{on}(t=1)=x_1$ respectively



[arXiv:2209.15571, arXiv:2210.02747, arXiv:2209.03003, arXiv:2305.10475v2]

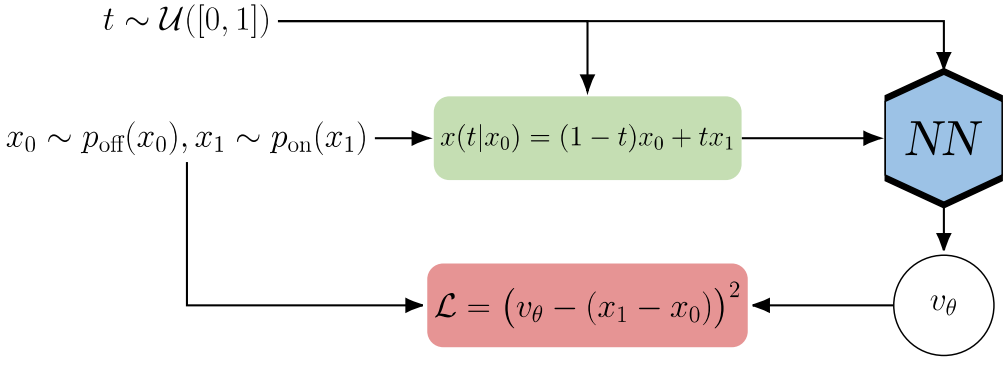


† slide credit M. Kuschick

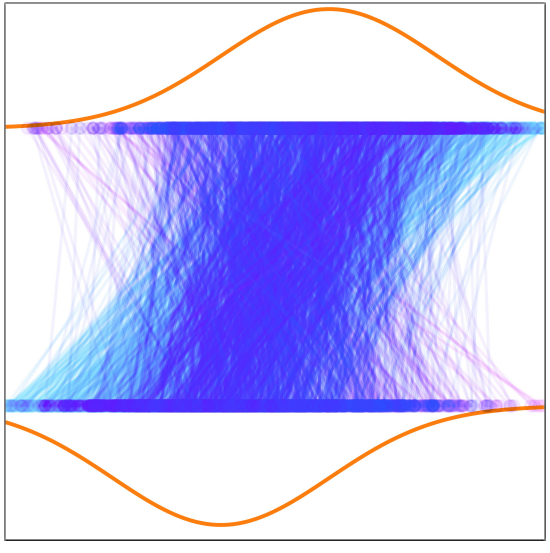
The direct diffusion neural network

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Off-Shell event $x_{off}(t=0)=x_0$, on-shell events $x_{on}(t=1)=x_1$ respectively



[arXiv:2209.15571, arXiv:2210.02747, arXiv:2209.03003, arXiv:2305.10475v2]

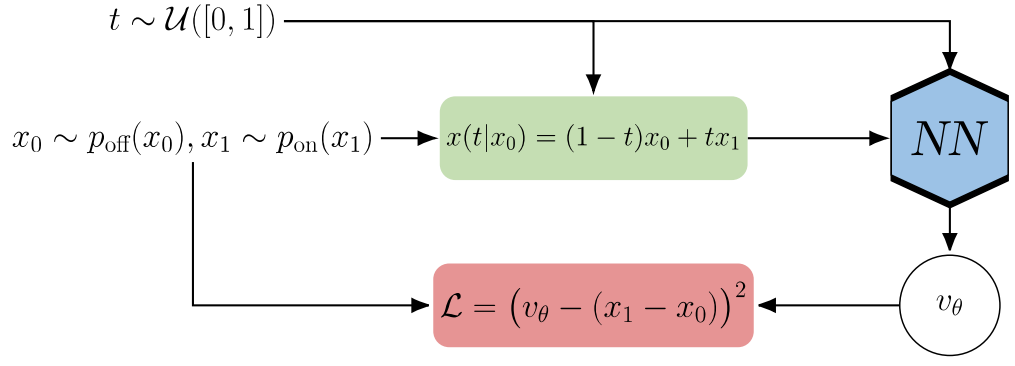


† slide credit M. Kuschick

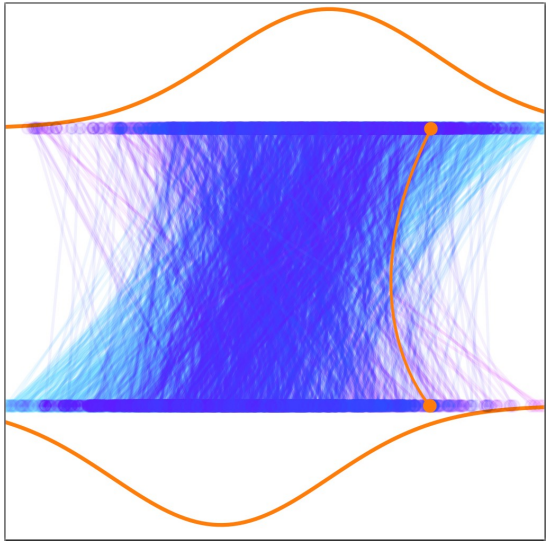
The direct diffusion neural network

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Off-Shell event $x_{off}(t=0)=x_0$, on-shell events $x_{on}(t=1)=x_1$ respectively



[arXiv:2209.15571, arXiv:2210.02747, arXiv:2209.03003, arXiv:2305.10475v2]

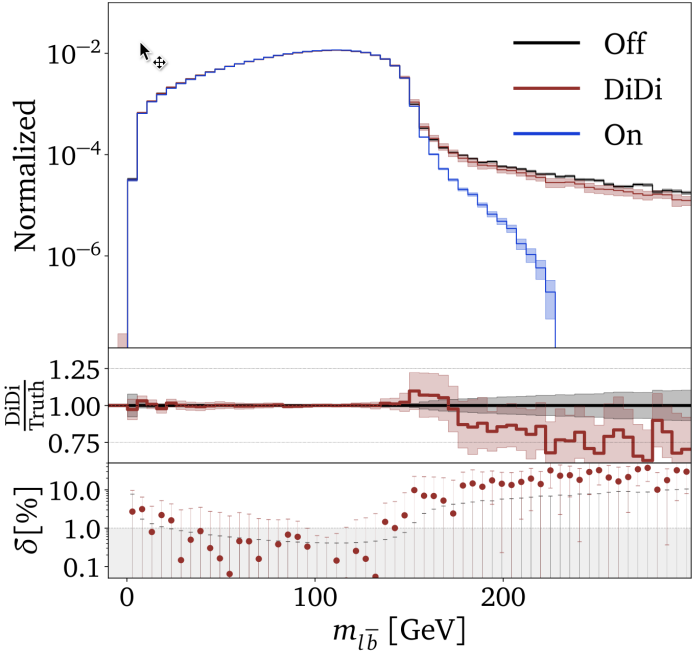


† slide credit M. Kuschick

The direct diffusion neural network

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Does it work?
  - On input sample with approximate top decay modelling
  - Off target sample with full off-shell top decay modelling
  - DiDi NN prediction
- It succeeds in filling regions of phase-space absent in the input sample
- Precision could be better



Direct diffusion + reweighting

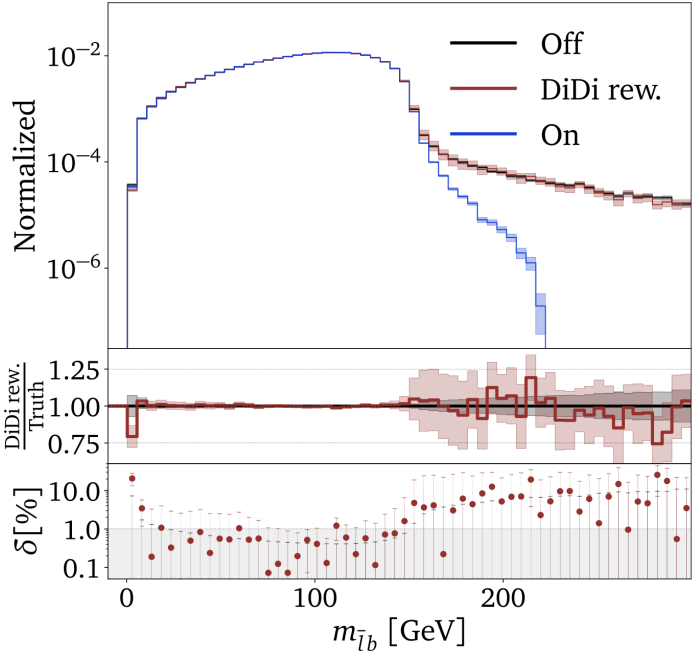
[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Furthermore, we can reweight the DiDi events with

$$w(x) = \frac{p_{\text{Off,data}}(x)}{p_{\text{Off,DiDi}}(x)} \text{ with}$$

$$C(x) = \frac{p_{\text{Off,data}}(x)}{p_{\text{Off,data}}(x) + p_{\text{Off,DiDi}}(x)}$$

- To appreciably improve the precision



Results

Results

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]

- Setup
  - $e^+\nu_e\mu^-\bar{\nu}_\mu b\bar{b}$ at LO QCD $\mathcal{O}(\alpha_S^2\alpha^4)$ at 13TeV LHC
  - Input sample: `POWHEG/hvq`; Target sample: `POWHEG/bb4l`
  - $m_t = 172.5$ GeV, $\Gamma_t = 1.453$ GeV, $m_b = 4.75$ GeV, etc.

DiDi

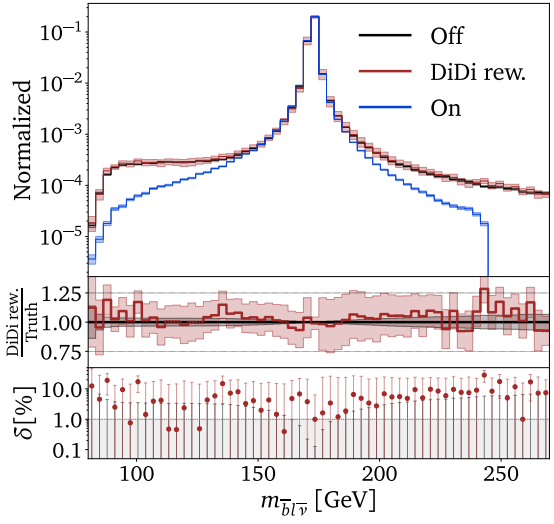
Hyperparameter	
Embedding dimension	64
Layers	8
Intermediate dimensions	768
LR scheduling	OneCycle
Starter LR	10^{-4}
Max LR	10^{-3}
Epochs	1000
Batch size	16384
c	10^{-3}
# Training events	3 M

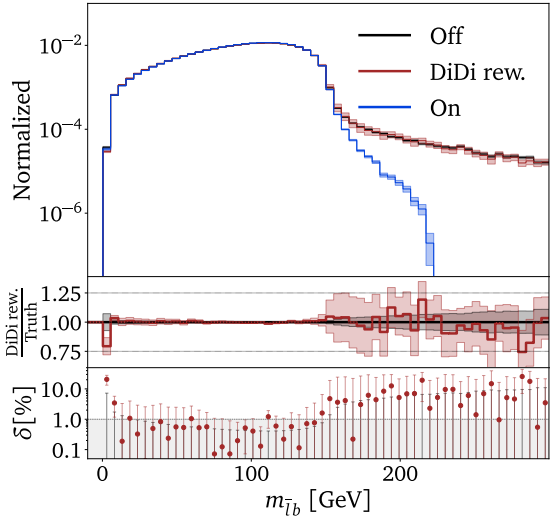
Reweighting classifier

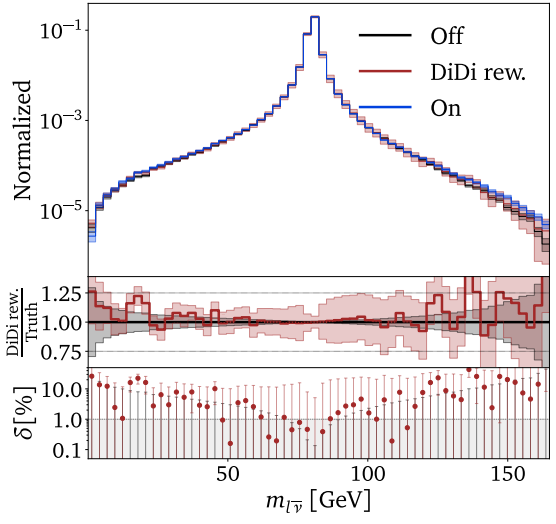
Hyperparameter	
Layers	5
Intermediate dimensions	512
Dropout	0.1
Normalization	BatchNorm1d
LR scheduling	ReduceOnPlateau
Starter LR	1^{-3}
Patience	10
Epochs	100
Batch size	1024
# Training events	2.5 M

Results

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]



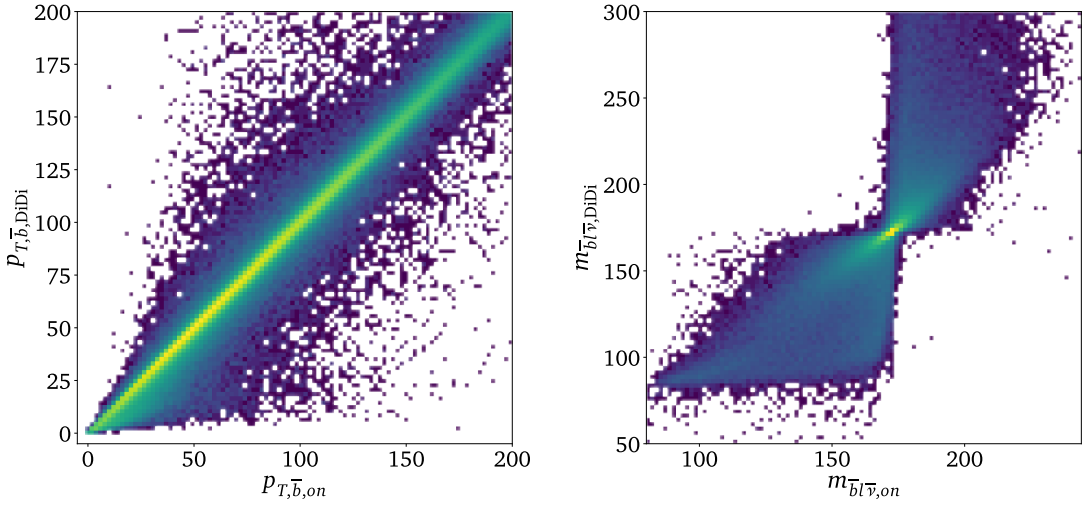




- Prediction (DiDi Rew.) matches the target sample (Off) very well
  - Devitations down to ~1% in the bulk and ~10% in the tails
- Uncertainties smaller than the size of the off-shell effect in the off-shell regions
  - DiDi is a Bayesian NN, but not the reweighting classifier

Results

[Butter, TJ, Klasen, Kuschick, Palacios Schweitzer, Plehn '23]



- Migration plots visualize the mapping from the input data set to the prediction
  - Offer insights as to what the NN actually does under the hood
- We find that no events moving from below to peak above the peak and vice versa

Towards a realistic example

[Kuschick '24]

- Radiative corrections
  - Extra radiation increases the dimensionality, does DiDi scale? Yes, see next slide!
  - At NLO `bb4l` and `hvq` do not match in dimensionality. DiDi applied only to radiation in production.
- Other production modes
  - Do we gain anything if tW production is added to the input sample? No benefits found thus far (no plots shown in this presentation).
- Parton Shower
  - No intention to include, but preparing for parton shower requires extra steps (shower starting scale, colour flows, etc.)
  - Could it be included in principle? Yes, in principle.

Summary and outlook

- New ML method for transforming event samples
- "Learns" off-shell effects in top quark production
  - Performs kinematic shifts and reweights
  - Combines a direct diffusion NN and a classifier
  - Proof of concept limited to LO, but realistic application is WIP [Kuschick '24]
- Other ideas
  - Can it also learn dependence on input parameters, e.g. top mass and width?
  - Can it encode scale dependence and be independent of PDFs for uncertainty estimates?
  - Can it be extended to include parton shower and detector simulation steps?