Invertible Networks for Unfolding

Tilman Plehn

Universität Heidelberg

LHC-EW WG 2/2021

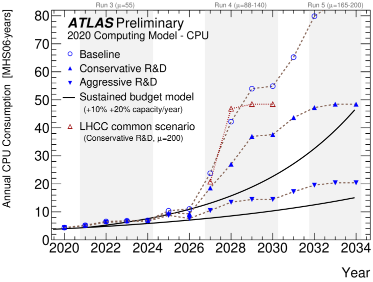
Machine Learning for HL-LHC

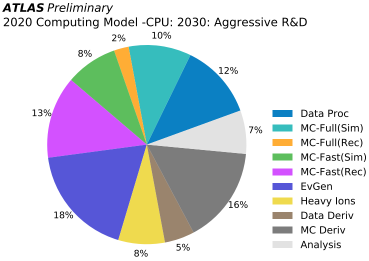
Searching for models ⟶ fundamental understanding of data

– precision theory

– precision simulations

– precision measurements





Machine Learning for HL-LHC

Searching for models $\longrightarrow$ fundamental understanding of data

- precision theory
- precision simulations
- precision measurements

Precision event generation

- simulated event numbers $\sim$ expected events [factor 25 for HL-LHC]
- general move to NLO/NNLO [1%-2% error]
- higher relevant multiplicities [jet recoil, extra jets, WBF, etc.]
- new low-rate high-multiplicity backgrounds
- cutting-edge predictions not through generators [N^3LO in Pythia?]
- interpretation beyond specific models [jets+MET]

Machine Learning for HL-LHC

Searching for models $\longrightarrow$ fundamental understanding of data

– precision theory
– precision simulations
– precision measurements

Precision event generation

– simulated event numbers $\sim$ expected events [factor 25 for HL-LHC]
– general move to NLO/NNLO [1%-2% error]
– higher relevant multiplicities [jet recoil, extra jets, WBF, etc.]
– new low-rate high-multiplicity backgrounds
– cutting-edge predictions not through generators [N^3LO in Pythia?]
– interpretation beyond specific models [jets+MET]

Three ways to use ML

– improve current tools: iSherpa, ML-MadGraph, etc
– new ideas, like fast ML-generator-networks
– conceptual ideas in theory simulations

Unfolding benefits

Power of kinematics

– distributions crucial for global EFT analyses
– searches with many applications: jets+MET
– searches not always used for intended purpose: $pp \to VH$
– proper unfolding means phase space, not STXS

SciPost Physics Submission

The Gauge-Higgs Legacy of the LHC Run II

Anke Biekötter[1], Tyler Corbett[2], and Tilman Plehn[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Niels Bohr International Academy and Discovery Centre, Niels Bohr Institute, University of Copenhagen, Denmark
biekoetter@thphys.uni-heidelberg.de

April 12, 2019

Abstract

We present a global analysis of the Higgs and electroweak sector based on LHC Run II and electroweak precision observables. We show which measurements provide the leading constraints on Higgs-related operators, and how the achieved LHC precision makes it necessary to combine rate measurements with electroweak precision observables. The SFitter framework allows us to include kinematic distributions beyond pre-defined ATLAS and CMS observables, independently study correlations, and avoid Gaussian assumptions for theory uncertainties. These Run II results are a step towards a precision physics program at the LHC, interpreted in terms of effective operators.

production	decay	ATLAS	CMS
	$h \to WW$	[120, 121]	[122–124]
	$h \to ZZ$	[121, 125]	[123, 124, 126, 127]
	$h \to \gamma\gamma$	[128]	[129]
	$h \to \tau\tau$	[121]	[123, 124, 130]
	$h \to Z\gamma$	[131]	[132]
WBF	$h \to$ inv		[133]
WBF	$h \to \tau\tau$		[130]
Vh	$h \to b\bar{b}$	[134]	[135]
Vh	$h \to \tau\tau$		[136]
Vh	$h \to$ inv	[137]	[138]
Vh	$h \to b\bar{b}$ (m_{Vh})	[139]	
$t\bar{t}h$	$h \to \gamma\gamma$	[118]	[129]
$t\bar{t}h$	$h \to ZZ \to 4\ell$	[118]	[126, 127]
$t\bar{t}h$	$h \to WW, ZZ, \tau\tau$	[121]	[123, 124]
$t\bar{t}h$	$h \to b\bar{b}$	[140]	[141]

Unfolding benefits

Power of kinematics

- distributions crucial for global EFT analyses
- searches with many applications: jets+MET
- searches not always used for intended purpose: $pp \to VH$
- proper unfolding means phase space, not STXS

SciPost Physics Submission

The Gauge-Higgs Legacy of the LHC Run II

Anke Biekötter[1], Tyler Corbett[2], and Tilman Plehn[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Niels Bohr International Academy and Discovery Centre, Niels Bohr Institute, University of Copenhagen, Denm
biekoetter@thphys.uni-heid

April 12, 2019

Abstract

We present a global analysis of the Higgs and el
Run II and electroweak precision observables.
provide the leading constraints on Higgs-related
LHC precision makes it necessary to combine rate
precision observables. The SFitter framework all
tributions beyond pre-defined ATLAS and CMS c
correlations, and avoid Gaussian assumptions f
Run II results are a step towards a precision phy
preted in terms of effective operators.

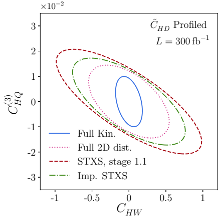
production	decay	ATLAS	CMS
	$h \to WW$	[120, 121]	[122–124]
	$h \to ZZ$	[121, 125]	[123, 124, 126, 127]
	$h \to \gamma\gamma$	[128]	[129]
	$h \to \tau\tau$	[121]	[123, 124, 130]
	[illegible]	[illegible]	[132]
			[133]
			[130]
			[135]
			[136]
			[138]

Benchmarking simplified template cross sections in WH production

Johann Brehmer,[a] Sally Dawson,[b] Samuel Homiller,[b,c] Felix Kling,[d,e] and Tilman Plehn[f]

[a] Center for Cosmology and Particle Physics, Center for Data Science, New York University, USA
[b] Department of Physics, Brookhaven National Laboratory, Upton, NY, 11973, USA
[c] C. N. Yang Institute for Theoretical Physics, Stony Brook University, NY, 11790, USA
[d] Department of Physics and Astronomy, University of California, Irvine, USA
[e] SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[f] Institut für Theoretische Physik, Universität Heidelberg, Germany

Abstract: Simplified template cross sections define a framework for the measurement and dissemination of kinematic information in Higgs measurements. We benchmark the currently proposed setup in an analysis of dimension-6 effective field theory operators for WH production. Calculating the Fisher information allows us to quantify the sensitivity of this framework to new physics and study its dependence on phase space. New machine-learning techniques let us compare the simplified template cross section framework to the full, high-dimensional kinematic information. We show that the way in which we truncate the effective theory has a sizable impact on the definition of the optimal simplified template cross sections.



Unfolding benefits

Power of kinematics

- distributions crucial for global EFT analyses
- searches with many applications: jets+MET
- searches not always used for intended purpose: $pp \to VH$
- proper unfolding means phase space, not STXS

SciPost Physics — Submission

The Gauge-Higgs Legacy of the LHC Run II

Anke Biekötter[1], Tyler Corbett[2], and Tilman Plehn[1]

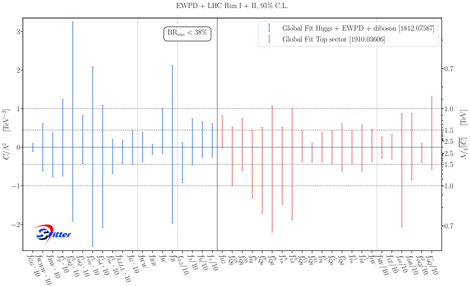
1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Niels Bohr International Academy and Discovery Centre, Niels Bohr Institute, University of Copenhagen, Denm
biekoetter@thphys.uni-heid

April 12, 2019

production	decay	ATLAS	CMS
	$h \to WW$	[120, 121]	[122–124]
	$h \to ZZ$	[121, 125]	[123, 124, 126, 127]
	$h \to \gamma\gamma$	[128]	[129]
	$h \to \tau\tau$	[121]	[123, 124, 130]
			[132]
			[133]
			[130]

Benchmarking simplified template cross sections in WH production

experiment		$\sqrt{S}$ (TeV)	$\mathcal{L}$ (fb^{-1})	channel	observable & K-factor		#bins	R	M	D	A
$pp \to t\bar{t}$											
CMS	[52]	8	19.7	$e\mu$	$\sigma_{t\bar{t}}$	[53]		✓	✓	.	.
ATLAS	[54]	8	20.02	lj	$\sigma_{t\bar{t}}$	[53]		✓	✓	.	.
CMS	[55]	13	2.3	lj	$\sigma_{t\bar{t}}$	[53]		✓	✓	.	.
CMS	[56]	13	3.2	ll	$\sigma_{t\bar{t}}$	[53]		✓	✓	.	.
ATLAS	[57]	13	3.2	$e\mu$	$\sigma_{t\bar{t}}$	[53]		✓	✓	.	.
ATLAS	[58]	8	20.3	lj	$\sigma^{-1}(d\sigma/dm_{t\bar{t}})$	[59–61]	7	.	✓	✓	.
CMS	[62]	8	19.7	lj	$\sigma^{-1}(d\sigma/dp_{T,t})$	[59–61]	7	.	.	✓	.
				ll	$\sigma^{-1}(d\sigma/dp_{T,1})$		5	.	.	✓	.
CMS	[63]	8	19.7	$e\mu$	$\sigma^{-1}(d^2\sigma/dm_{t\bar{t}}dy_{t\bar{t}})$	[64]	16	.	.	.	.
CMS	[65]	8	19.7	lj high p_T	$d\sigma/dp_{T,t}$		5	.	.	.	.
CMS	[66]	13	2.3	lj	$\sigma^{-1}(d\sigma/dm_{t\bar{t}})$		8	.	✓	✓	.
CMS	[67]	13	35.8	lj	$\sigma^{-1}(d\sigma/dp_T(t_h))$	[59–61]	12	.	.	✓	.
CMS	[68]	13	2.1	ll	$\sigma^{-1}(d\sigma/dp_{T,t})$	[59–61]	6	.	.	✓	.
CMS	[69]	13	35.9	ll	$\sigma^{-1}(d\sigma/d\Delta y_{t\bar{t}})$	[59–61]	8	.	.	.	✓
ATLAS	[70]	13	36.1	aj high p_T	$\sigma^{-1}(d\sigma/dm_{t\bar{t}})$		8	.	.	.	.
CMS	[71]	8	19.7	lj	A_C	[72]		.	.	.	✓
CMS	[73]	8	19.7	ll	A_C	[72]		.	.	.	✓
ATLAS	[74]	8	20.3	lj	A_C	[72]		.	.	.	✓
ATLAS	[75]	8	20.3	ll	A_C	[72]		.	.	.	✓
ATLAS	[76]	13	139	lj	A_C	[72]		.	.	.	✓



GAN algorithm

Generating events

- training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
- discriminator constructing $D(x)$ by minimizing [classifier $D(x) = 1, 0$ true/generator]

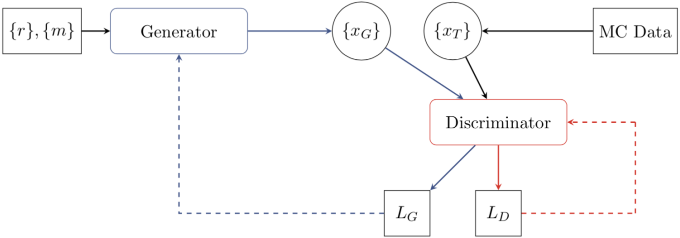
$$L_D = \Big\langle -\log D(x) \Big\rangle_{x_T} + \Big\langle -\log(1 - D(x)) \Big\rangle_{x_G}$$

- generator constructing $r \to x_G$ by minimizing [D needed]

$$L_G = \Big\langle -\log D(x) \Big\rangle_{x_G}$$

- equilibrium $D = 0.5 \;\Rightarrow\; L_D = 2L_G = -2\log 0.5$

$\Rightarrow$ statistically independent copy of training events



GAN algorithm

Generating events

- training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
- discriminator constructing $D(x)$ by minimizing [classifier $D(x) = 1, 0$ true/generator]
- generator constructing $r \to x_G$ by minimizing [D needed]

⇒ statistically independent copy of training events

Generative network studies

- Jets [de Oliveira (2017), Carrazza-Dreyer (2019)]
- Detector simulations [Paganini (2017), Musella (2018), Erdmann (2018), Ghosh (2018), Buhmann (2020)]
- Events [Otten (2019), Hashemi, DiSipio, Butter (2019), Martinez (2019), Alanazi (2020), Chen (2020), Kansal (2020)]
- Unfolding [Datta (2018), Omnifold (2019), Bellagente (2019), Bellagente (2020), Howard (2020)]
- Templates for QCD factorization [Lin (2019)]
- EFT models [Erbin (2018)]
- Event subtraction [Butter (2019)]
- Sherpa [Bothmann (2020), Gao (2020)]
- Basics [GANplification (2020), DCTR (2020)]
- Unweighting [Verheyen (2020), Backes (2020)]
- Superresolution [DiBello (2020), Baldi (2020)]

How to GAN away detector effects

Goal: invert standard simulation [Bellagente, Butter, Kasiczka, TP, Winterhalder]

– detector simulation typical Monte Carlo, random-number-driven

– inversion possible, in principle [MEM, but entangled convolutions]

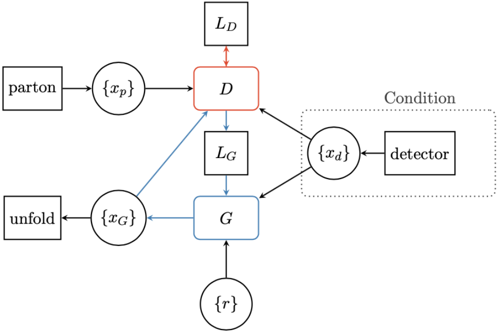
– GAN task

$$\text{partons} \xrightarrow{\text{DELPHES}} \text{detector} \xrightarrow{\text{GAN}} \text{partons}$$

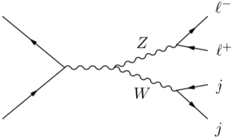
⇒ Full phase space unfolded

Conditional GAN

– random numbers to parton level
hadron level as condition
matched event pairs



Detector unfolding

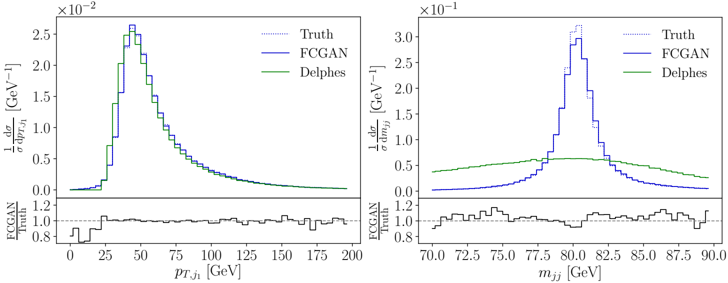


Reference process $pp \to ZW \to (\ell\ell)\ (jj)$

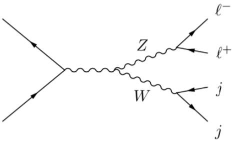
– broad jj mass peak
narrow $\ell\ell$ mass peak
modified $2 \to 2$ kinematics
fun phase space boundaries

– GAN same as event generation [with MMD]

Model (in)dependence



Detector unfolding



Reference process $pp \to ZW \to (\ell\ell)\ (jj)$

– broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
  fun phase space boundaries

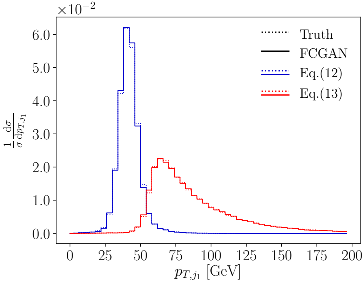
– GAN same as event generation [with MMD]

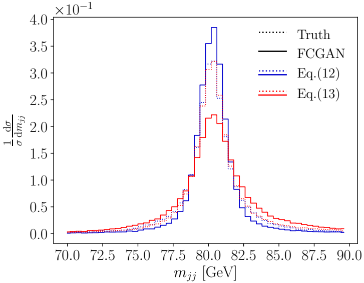
Model (in)dependence

– detector-level cuts [14%, 39% events, no interpolation, MMD not conditional]

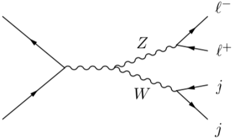
$$p_{T,j_1} = 30 \ldots 50\ \text{GeV} \quad p_{T,j_2} = 30 \ldots 40\ \text{GeV} \quad p_{T,\ell^-} = 20 \ldots 50\ \text{GeV} \qquad (12)$$

$$p_{T,j_1} > 60\ \text{GeV} \qquad (13)$$





Detector unfolding



Reference process $pp \to ZW \to (\ell\ell)\ (jj)$

– broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
  fun phase space boundaries

– GAN same as event generation [with MMD]

Model (in)dependence

– detector-level cuts [14%, 39% events, no interpolation, MMD not conditional]

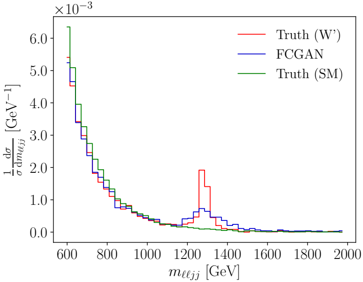
$$p_{T,j_1} = 30 \dots 50 \text{ GeV} \quad p_{T,j_2} = 30 \dots 40 \text{ GeV} \quad p_{T,\ell^-} = 20 \dots 50 \text{ GeV} \qquad (12)$$

$$p_{T,j_1} > 60 \text{ GeV} \qquad (13)$$

– model dependence [Thank you to Ben]

– train: SM events
  test: 10% events with W' in s-channel

⇒ Working fine, but ill-defined



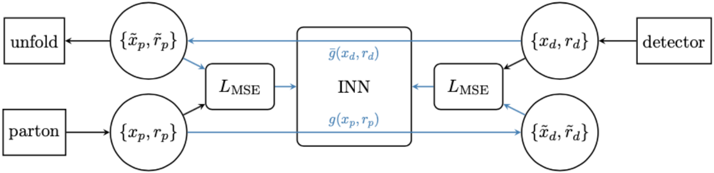
Invertible networks

Standard invertible networks [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

- network as bijective transformation — normalizing flow
  Jacobian tractable [specifically: coupling layer]
  evaluation in both directions — INN [Ardizzone, Rother, Köthe]
- mapping parton and detector phase spaces
  padding with random numbers [eINN, dimensionality, sampling for poor]

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

- training on event pairs (MSE) or samples (MMD)



Invertible networks

Standard invertible networks [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

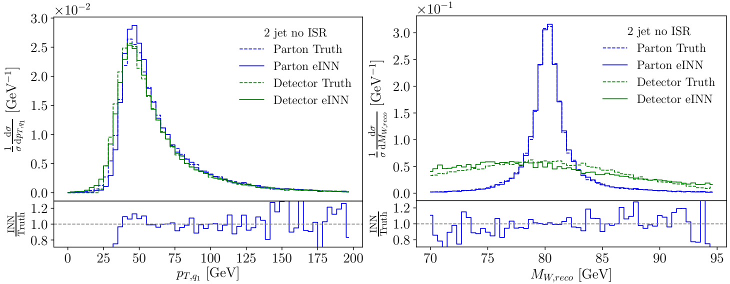
– network as bijective transformation — normalizing flow
Jacobian tractable [specifically: coupling layer]
evaluation in both directions — INN [Ardizzone, Rother, Köthe]

– mapping parton and detector phase spaces
padding with random numbers [eINN, dimensionality, sampling for poor]

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

– training on event pairs (MSE) or samples (MMD)

– same task as FCGAN, similar performance



Invertible networks

Standard invertible networks [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

- network as bijective transformation — normalizing flow
  Jacobian tractable [specifically: coupling layer]
  evaluation in both directions — INN [Ardizzone, Rother, Köthe]
- mapping parton and detector phase spaces
  padding with random numbers [eINN, dimensionality, sampling for poor]

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

- training on event pairs (MSE) or samples (MMD)
- same task as FCGAN, similar performance

⇒ Working okay, still ill-defined

Proper inverting with cINN

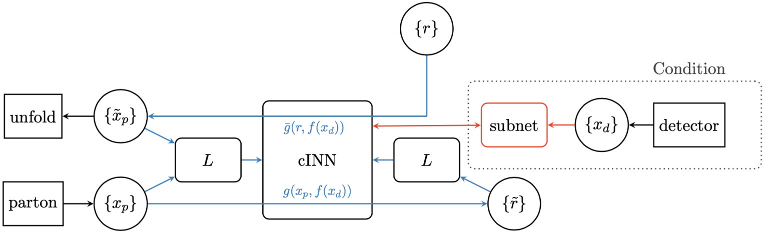
Statistical inversion [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

– task: construct parton-level pdf for (single) detector-level event

1- conditional INN: parton-level events from $\{r\}$

2- maximum likelihood loss

$$
\begin{aligned}
L &= - \left\langle \log p(\theta | x_p, x_d) \right\rangle_{x_p, x_d} \\
&\approx - \left\langle \log p(g(x_p, x_d)) + \log \left| \frac{\partial g(x_p, x_d)}{\partial x_p} \right| \right\rangle_{x_p, x_d} - \log p(\theta) \\
&= - \left\langle - \frac{||g(x_p, x_d)||_2^2}{2} + \log \left| \frac{\partial g(x_p, x_d)}{\partial x_p} \right| \right\rangle_{x_p, x_d} - \log p(\theta)
\end{aligned}
$$



Proper inverting with cINN

Statistical inversion [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

– task: construct parton-level pdf for (single) detector-level event

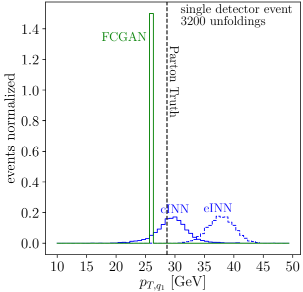
1- conditional INN: parton-level events from $\{r\}$

2- maximum likelihood loss

Again $pp \to ZW \to (\ell\ell)\ (jj)$

– performance like FCGAN

– distribution: single pair (x_p, x_d), unfolded many times [FCGAN is out]



Proper inverting with cINN

Statistical inversion [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

– task: construct parton-level pdf for (single) detector-level event

1- conditional INN: parton-level events from $\{r\}$

2- maximum likelihood loss

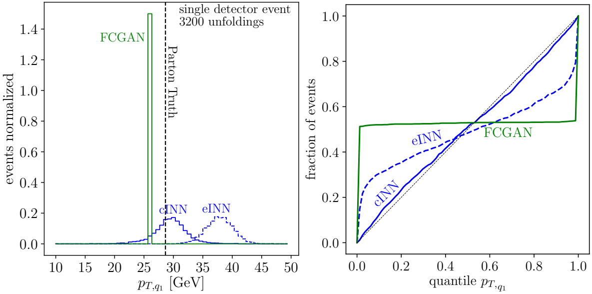
Again $pp \to ZW \to (\ell\ell)\ (jj)$

– performance like FCGAN

– distribution: single pair (x_p, x_d), unfolded many times [FCGAN is out]

– calibration: 1500 pairs (x_p, x_d), each unfolded 60 times, check for truth

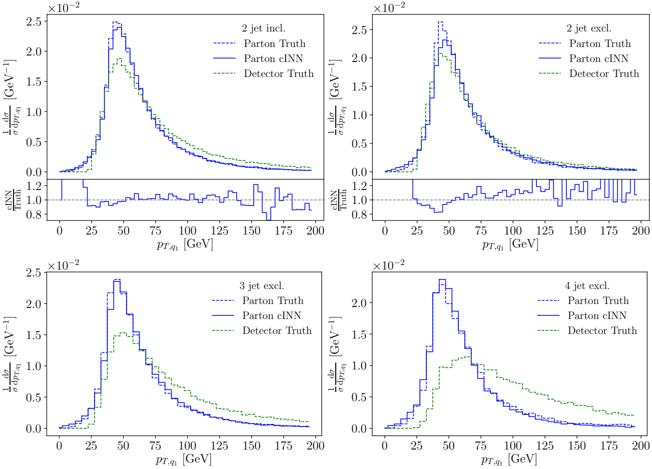
$\Rightarrow$ cINN well-defined!



Inverting to hard process

What theorists want: unfolding ISR

- detector-level process $pp \to ZW$+jets [variable number of objects]
- ME vs PS jets decided by network
- training jet-inclusively or jet-exclusively
  parton-level hard process extracted as $2 \to 2$



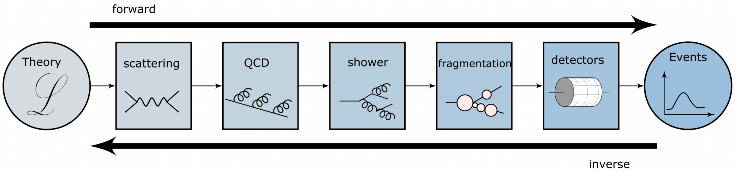
Inverting to hard process

What theorists want: unfolding ISR

– detector-level process $pp \to ZW$+jets [variable number of objects]
– ME vs PS jets decided by network
– training jet-inclusively or jet-exclusively
parton-level hard process extracted as $2 \to 2$

Towards systematic inversion

– detector unfolding on way
– QCD parton from jet algorithm standard
– jet radiation possible
$\Rightarrow$ Hard matrix element a proper goal?



Things are moving

Machine learning for LHC theory

– big data for fundamental physics
– GANs the cool kid
– INNs the theory hope
– Full inversion in reach