

Machine Learning for LHC Predictions

Tilman Plehn

Universität Heidelberg

QCD@LHC, October 2026



Better theory predictions

Theory [Lots of talks here]

- pdf sets [talks: Stefano Forte & Eva Groenendijk]
- perturbative QFT
- resummed QFT [talk: Rikab Gambhir]
- hadronization

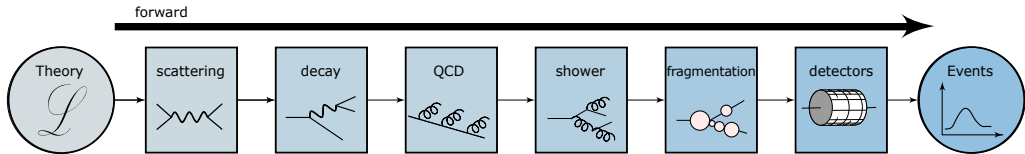
→ 'Modeling' in terms of first principles

Tools moving to AI

- NNPDF, Colibri
- MadGraph, Sherpa
- MC@NLO, Powheg
- Pythia, Herwig → MLhad, HadML
- NNLO/N3LO...

Enhancements

- more precise
- faster
- user-friendly



Neural importance sampling

Sampling

Amplitudes

NLO

Hadronization

Feynman diag

MadAgents

Generative Jacobians

- VEGAS $\rightarrow$ normalizing flow [i-flow: Gao, Höche, Isaacson, Krause, Schulz]

$$\int dx f(x) = \int dx g_\theta(x) \frac{f(x)}{g_\theta(x)} \equiv \left\langle \frac{f(x)}{g_\theta(x)} \right\rangle_{x \sim g_\theta}$$

- w/ neural multi-channel weights [Heimel, Hütsch, Maltoni, Mattelaer, TP, Winterhalder]

$$\sum_{i=1}^{n_c} \int dx \alpha_{i,\varphi}(x) f(x) = \sum_{i=1}^{n_c} \left\langle \frac{\alpha_{i,\varphi}(x) f(x)}{g_{i,\theta}(x)} \right\rangle_{x \sim g_{i,\theta}}$$



Neural importance sampling

Generative Jacobians

- VEGAS $\rightarrow$ normalizing flow [i-flow: Gao, Höche, Isaacson, Krause, Schulz]

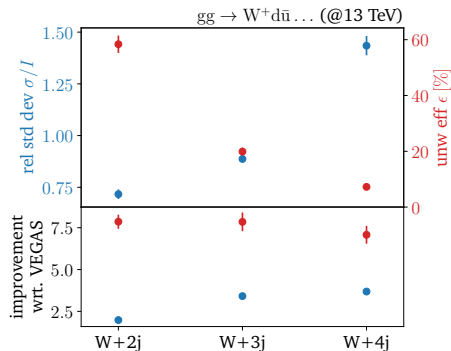
$$\int dx f(x) = \int dx g_\theta(x) \frac{f(x)}{g_\theta(x)} \equiv \left\langle \frac{f(x)}{g_\theta(x)} \right\rangle_{x \sim g_\theta}$$

- w/ neural multi-channel weights [Heimel, Hütsch, Maltoni, Mattelaer, TP, Winterhalder]

$$\sum_{i=1}^{n_c} \int dx \alpha_{i,\varphi}(x) f(x) = \sum_{i=1}^{n_c} \left\langle \frac{\alpha_{i,\varphi}(x) f(x)}{g_{i,\theta}(x)} \right\rangle_{x \sim g_{i,\theta}}$$

MadNIS [Sherpa/Pepper similar]

- VEGAS pre-training of spline flows
 - online + buffered training
 - fewer channels
- $\rightarrow$ [MadGraph7](#)



Neural control variates

Sampling

Amplitudes

NLO

Hadronization

Feynman diags

MadAgents

Negative weights [Heimel, TP, Revelli, Vent, Winterhalder]

- learned subtraction

$$\begin{aligned} \int dx f(x) &= \int dx \left[f(x) - C_\phi p_\phi(x) \right] + C_\phi \int dx p_\phi(x) \\ &= \left\langle \frac{f(x) - C_\phi p_\phi(x)}{g_\theta(x)} \right\rangle_{x \sim g_\theta} - C_\phi \langle 1 \rangle_{x \sim p_\phi} \end{aligned}$$

- LO: reduce weight range
move evaluations to cheap flow
 - benchmark $pp \rightarrow t\bar{t}gg$
 - NLO: avoid negative weights
finite soft/collinear subtraction
 - benchmark $e^+e^- \rightarrow t\bar{t}g, q\bar{q}g$
- Proof of concept for now...



Amplitudes

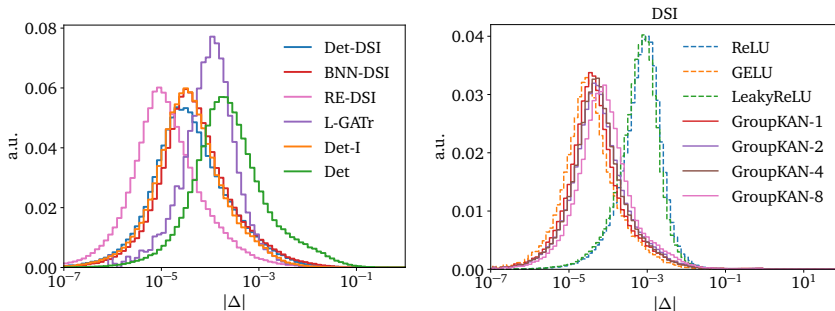
Accurate surrogates [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

- training data: loop amplitudes $gg \rightarrow \gamma\gamma g(g)$
- curse of dimensionality unavoidable [Bahl, Bresó, Butter, Iturriza]

$$A_{\text{NN}}(\{p_i\}) \quad \text{vs} \quad A_{\text{NN}}(\{(p_i p_j)\})$$

- Accuracy from symmetric input...
...activation, ensemble, training, regularization, factorization

[Janßen, Maitre, Schumann, Siegert, Truong; Villadamigo, Frederix, TP, Vitos, Winterhalder]



Amplitudes

Accurate surrogates [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

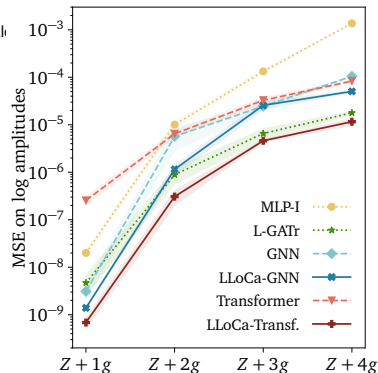
- training data: loop amplitudes $gg \rightarrow \gamma\gamma g(g)$
- curse of dimensionality unavoidable [Bahl, Bresó, Butter, Iturriaza]

$$A_{\text{NN}}(\{p_i\}) \quad \text{vs} \quad A_{\text{NN}}(\{(p_i p_j)\})$$

- Accuracy from symmetric input...
...activation, ensemble, training, regularization, factorization
[Janßen, Maitre, Schumann, Siegert, Truong; Villadamigo, Frederix, TP, Vitos, Winterhalder]

Lorentz equivariance [Brehmer, Bresó, de Haan, TP, Qu, Spinner, Thaler,...]

- equivariant transformers: L-GATr & LLoCa
 - learnable symmetry breaking
- 1 amplitudes $q\bar{q} \rightarrow Z + 1\dots 4 g$



Amplitudes

Accurate surrogates [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

- training data: loop amplitudes $gg \rightarrow \gamma\gamma g(g)$
- curse of dimensionality unavoidable [Bahl, Bresò, Butter, Iturriza]

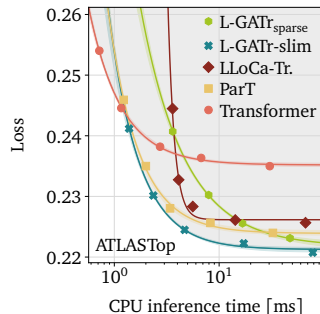
$$A_{\text{NN}}(\{p_i\}) \quad \text{vs} \quad A_{\text{NN}}(\{(p_i p_j)\})$$

- Accuracy from symmetric input...
...activation, ensemble, training, regularization, factorization

[Janßen, Maitre, Schumann, Siegert, Truong; Villadamigo, Frederix, TP, Vitos, Winterhalder]

Lorentz equivariance [Brehmer, Bresò, de Haan, TP, Qu, Spinner, Thaler,...]

- equivariant transformers: L-GATr & LLoCa
 - learnable symmetry breaking
- 1 amplitudes $q\bar{q} \rightarrow Z + 1\dots 4 g$
 - 2 jet tagging
 - 3 event generation
- Knowledge driving performance [also OmniLearned]



Calibrated learned uncertainties

Statistics vs systematics [Bahl, Elmer, TP, Winterhalder]

- variational inference [Gaussian likelihood]

$$\langle A \rangle = \int dA p(A) A$$

$$\sigma^2 = \int dA p(A) (A - \langle A \rangle)^2$$

$$p(A) = \int d\lambda p(A|\lambda) p(\lambda|T) \approx \int d\lambda p(A|\lambda) q(\lambda)$$

- conjugate prior [$p(A|\lambda) = \mathcal{N}(A) \rightarrow$ inverse Gamma fct]

$$q_{\gamma, \nu, \alpha, \beta}(\lambda) = \frac{\beta^\alpha \sqrt{\nu}}{\Gamma(\alpha) \sqrt{2\pi\sigma^2}} \left(\frac{1}{\sigma^2} \right)^{\alpha+1} \exp\left(-\frac{2\beta + \nu(\gamma - \bar{A})^2}{2\sigma^2} \right)$$

- integrating to Student- t

$$p(A) = \text{St} \left(A \middle| \gamma, \frac{\beta(1+\nu)}{\nu\alpha}, 2\alpha \right)$$



Calibrated learned uncertainties

Statistics vs systematics [Bahl, Elmer, TP, Winterhalder]

- variational inference [Gaussian likelihood]

$$\langle A \rangle = \int dA p(A) A$$

$$\sigma^2 = \int dA p(A) (A - \langle A \rangle)^2$$

$$p(A) = \int d\lambda p(A|\lambda) p(\lambda|T) \approx \int d\lambda p(A|\lambda) q(\lambda)$$

- conjugate prior $[p(A|\lambda) = \mathcal{N}(A) \rightarrow \text{inverse Gamma fct}]$

$$q_{\gamma, v, \alpha, \beta}(\lambda) = \frac{\beta^\alpha \sqrt{v}}{\Gamma(\alpha) \sqrt{2\pi\sigma^2}} \left(\frac{1}{\sigma^2} \right)^{\alpha+1} \exp\left(-\frac{2\beta + v(\gamma - \bar{A})^2}{2\sigma^2} \right)$$

→ Learned uncertainties [heteroscedastic, BNN, ensembles, conformal, evidential]

$$A_{\text{NN}} = \gamma \quad \sigma_{\text{stat}}^2 = \int d\lambda [\bar{A}(\lambda) - A_{\text{NN}}]^2 q(\lambda) = \frac{\beta}{v(\alpha - 1)}$$

$$\sigma_{\text{syst}}^2 = \int d\lambda \sigma^2(\lambda) q(\lambda) = \frac{\beta}{\alpha - 1}$$



Calibrated learned uncertainties

Statistics vs systematics [Bahl, Elmer, TP, Winterhalder]

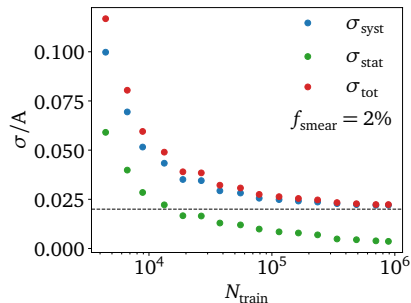
- variational inference [Gaussian likelihood]

$$p(A) = \int d\lambda p(A|\lambda) p(\lambda|T) \approx \int d\lambda p(A|\lambda) q(\lambda)$$

→ **Learned uncertainties** [heteroscedastic, BNN, ensembles, conformal, evidential]

Back to $gg \rightarrow \gamma\gamma g(g)$

- systematics from artificial noise



Calibrated learned uncertainties

Statistics vs systematics [Bahl, Elmer, TP, Winterhalder]

- variational inference [Gaussian likelihood]

$$p(A) = \int d\lambda p(A|\lambda) p(\lambda|T) \approx \int d\lambda p(A|\lambda) q(\lambda)$$

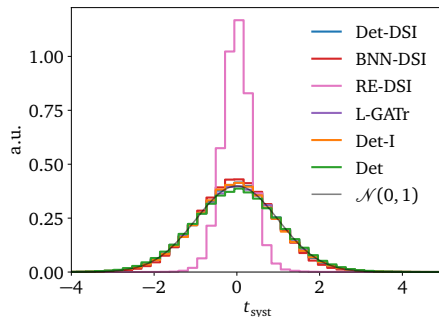
→ **Learned uncertainties** [heteroscedastic, BNN, ensembles, conformal, evidential]

Back to $gg \rightarrow \gamma\gamma g(g)$

- systematics from artificial noise
- pull over phase space

$$t_{\text{syst}}(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{\sigma_{\text{syst}}(x)}$$

→ **Calibration universal** [ensemble-bias]



Generalization uncertainties

Generative amplification [Bahl, Diefenbacher, Elmer, TP, Spinner]

- ‘How many events can I sample from a network trained on n_{train} events?’ [Louis Lyons]
- smoothly learned $p_{\text{gen}}(x)$

$$p_{\text{gen}}(x) \approx D_{\text{true}}^{n_{\text{train}}}(x) \sim p_{\text{true}}(x)$$

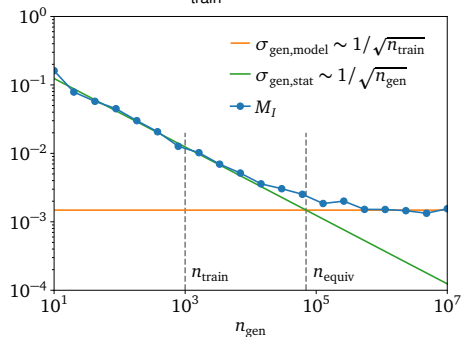
- Gaussian uncertainties on generated sample

$$\sigma_{\text{gen,stat}}^2 + \sigma_{\text{gen,model}}^2 = \begin{cases} \sigma_{\text{gen,stat}}^2 \sim \frac{a}{n_{\text{gen}}} & n_{\text{gen}} \ll n_{\text{train}} \\ \sigma_{\text{gen,model}}^2 \sim \frac{b}{n_{\text{train}}} & n_{\text{gen}} \gg n_{\text{train}} \end{cases}$$

- sweet spot

$$\sigma_{\text{gen,stat}}^2 \Big|_{n_{\text{gen}}=n_{\text{equiv}}} \stackrel{!}{=} \sigma_{\text{gen,model}}^2$$

$$\Leftrightarrow \frac{n_{\text{equiv}}}{n_{\text{train}}} = \frac{a}{b}$$



Generalization uncertainties

Generative amplification [Bahl, Diefenbacher, Elmer, TP, Spinner]

- ‘How many events can I sample from a network trained on n_{train} events?’ [Louis Lyons]
- smoothly learned $p_{\text{gen}}(x)$

$$p_{\text{gen}}(x) \approx D_{\text{true}}^{n_{\text{train}}}(x) \sim p_{\text{true}}(x)$$

- test with measure M

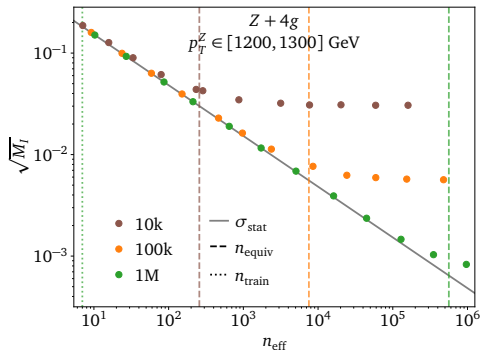
$$M[D_{\text{true}}^n, p_{\text{true}}] \Big|_{n=n_{\text{equiv}}} \stackrel{!}{=} M[D_{\text{gen}}^n, p_{\text{true}}] \Big|_{n \rightarrow \infty}$$

Surrogate Monte Carlo [Bahl, TP, Revelli]

- amplitude $q\bar{q} \rightarrow Z + \text{gluons}$
- M_I integral over tail region

$$\frac{n_{\text{equiv}}}{n_{\text{train}}} = \frac{560.000}{7} = 80.000$$

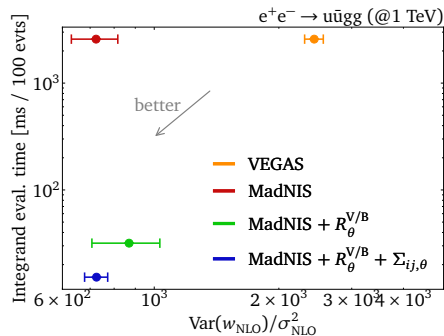
→ Amplification works, better for surrogates



Putting everything together [DeCrescenzo, Marino Villadamigo, Elmer, Heimel, TP, Winterhalder, Zaro]

- virtual amplitude surrogates [Born ratios w/ uncertainties]
- real amplitude surrogates [w/ uncertainties]
- neural importance sampling & channel weights
- conditional on FKS sectors

→ How close to 'NLO as fast as LO?' [NNLO: Janßen, Poncelet, Schumann]



Putting everything together [DeCrescenzo, Marino Villadamigo, Elmer, Heimel, TP, Winterhalder, Zaro]

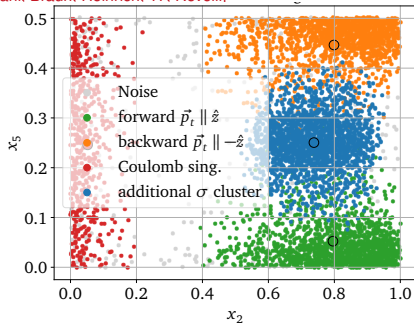
- virtual amplitude surrogates [Born ratios w/ uncertainties]
- real amplitude surrogates [w/ uncertainties]
- neural importance sampling & channel weights
- conditional on FKS sectors

→ **How close to 'NLO as fast as LO?'** [NNLO: Janßen, Poncelet, Schumann]

Accelerated loop amplitudes [Breso, Heinrich, Magerya, Olsson; Bahl, Braun, Heinrich, TP, Revelli]

- HH 2-loop
calibrated with Student- t likelihood
- $t\bar{t}H$ 1-loop
clustered large uncertainties for adaptive training

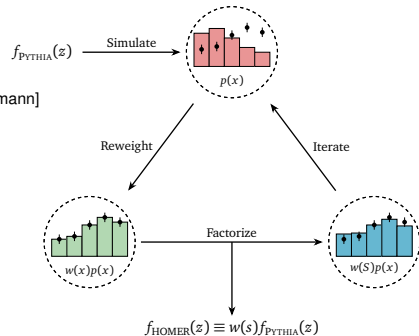
→ **Solving actual problems**



Hadronization: MLhad & HadML

Iterative HOMER [Kim, Ore, Palacios Schweitzer, TP, Szewc + MLhad]

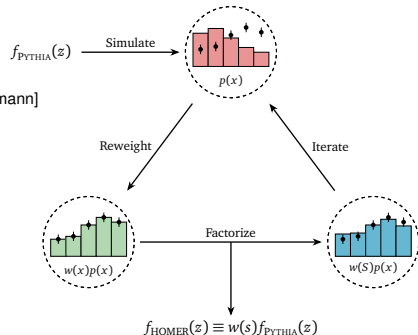
- better hadronization within string model [talk: Steffen Schumann]
- learn weights between theory and data
transfer weights to theory description
- learned uncertainty



Hadronization: MLhad & HadML

Iterative HOMER [Kim, Ore, Palacios Schweitzer, TP, Szewc + MLhad]

- better hadronization within string model [talk: Steffen Schumann]
- learn weights between theory and data transfer weights to theory description
- learned uncertainty



Beyond toy model [preliminary, using AUSSIE]

- variable number of particles
 - information loss
 - expressive beyond string fragmentation?
- Critical input to ML-tagger training [Jennifer Rohloff's talk]



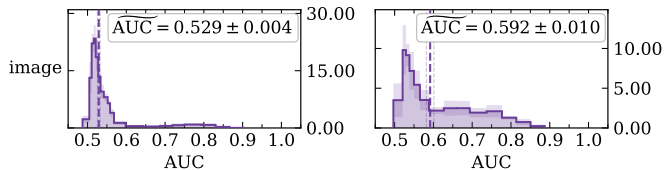
Feynman diagrams

Conditioned autoregressive transformer

- $q\bar{q} \rightarrow e^+e^-$ conditioned on $m_Z \rightarrow$ clearly possible
- event generation from Feynman diagrams?

Zero-shot generation [Bahl, TP, Schiller, Sivagnanalingam]

- train on 1352 processes [100k events each]
generate for 139 new processes
test with NN classifier



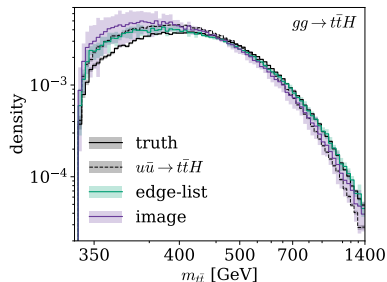
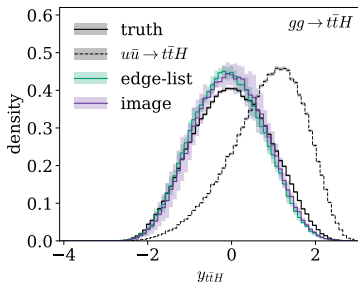
Feynman diagrams

Conditioned autoregressive transformer

- $q\bar{q} \rightarrow e^+ e^-$ conditioned on $m_Z \rightarrow$ clearly possible
- event generation from Feynman diagrams?

Zero-shot generation [Bahl, TP, Schiller, Sivagnanalingam]

- train on 1352 processes [100k events each]
 - generate for 139 new processes
 - test with NN classifier
 - $t\bar{t}H$ kinematics
- Any use case?



Neural networks everywhere

Examples for non-mentioned ML:

- high-dimensional & unbinned unfolding [talk: Hanfei Cui]
 - SBI [talks: Robert Schöfbeck & Chris Pollard]
 - amplitude relations [Cai, Charton, Cranmer, Dixon,...]
 - IBP simplifications [Shih]
 - ISR transformers [Butter, Charton, Villadamigo, Ore, TP, Spinner]
 - anomalies in jets [Das, Ostertag, TP, Vogel; talk: Radha Mastandrea]
- [Add your project...](#)

U.S. INTERNATIONAL CANADA ESPAÑOL 中文

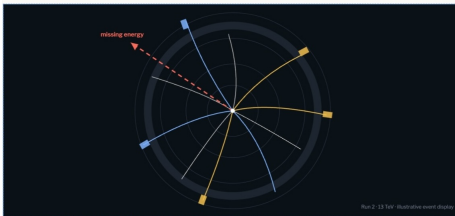
The New York Times

PHYSICS

Anthropic Says Its A.I. Found Dark Matter Hiding in CERN's Public Data

A model sifting a decade of openly released collider records flagged a faint, persistent imbalance that its creators say points to a long-sought particle. Physicists are intrigued, and deeply skeptical.

By Eleanor Vazquez-Hart | Sept. 24, 2026, 6:02 a.m. ET



A reconstruction of the kind of collision the model singled out: sprays of particles fly outward, but the momentum does not balance, testing that something invisible escaped the detector.

ILLUSTRATION



MadAgents

Agents running code, specifically MadGraph [TP, Schiller, Schmal]

- installation
- student training
- expert support
- autonomous generation
- new: SMEFT re-casting



MadAgents

Agents running code, specifically MadGraph [TP, Schiller, Schmal]

- installation
- student training
- expert support
- autonomous generation
- new: SMEFT re-casting

Expert user support

- improvements

User

I am a PhD student working on precision simulations for top-quark processes at colliders.

I currently have a baseline leading-order, on-shell simulation for a given process and observable.

Task:

Critically assess the limitations of such a setup and propose systematic and conceptually distinct ways to improve its theoretical precision.

For each proposed improvement:

- Explain which physical effects are being captured that are missing at baseline level.
- Clarify when and why this improvement becomes important for realistic analyses.
- Indicate, at a high level, how such an improvement is typically implemented in modern Monte-Carlo simulations.

Reason from first principles and standard practice in collider phenomenology. The goal is to identify the dominant axes along which precision is improved, not to enumerate tools.



MadAgents

Agents running code, specifically MadGraph [TP, Schiller, Schmal]

- installation
- student training
- expert support
- autonomous generation
- new: SMEFT re-casting

Expert user support

- improvements

MadAgents (shortened)

[...]

Dominant and conceptually distinct precision axes

Starting from a leading-order, on-shell baseline for top-quark processes, the main orthogonal directions along which theoretical predictions can be systematically improved are:

1. **Perturbative QCD accuracy in production and decay:** fixed-order calculations at NLO or NNLO, matching to parton showers (NLO+PS), and higher-order resummation at NNLO accuracy.
2. **Description of additional hard radiation:** inclusion of multi-jet matrix elements and their consistent merging with parton showers at LO or NLO accuracy.
3. **Finite-width and off-shell effects:** treatment beyond the narrow-width approximation, including the full resonance structure, non-resonant contributions, and interference effects.
4. **Electroweak and QED corrections:** NLO electroweak effects and photon radiation, which become increasingly relevant for high-scale observables and leptonic final states.
5. **All-order logarithmic structure:** higher-logarithmic resummation and improved parton-shower descriptions for observables sensitive to large scale hierarchies.
6. **Non-perturbative, PDF, and scheme uncertainties:** effects from hadronisation, the underlying event, colour reconnection, parton distribution functions and α_s , as well as choices of mass and renormalisation schemes.

In realistic precision top-quark analyses, several of these axes are typically combined. The appropriate level of sophistication is guided by which physical effects dominate the theoretical uncertainties for the observables and phase-space regions under consideration.



MadAgents

Agents running code, specifically MadGraph [TP, Schiller, Schmal]

- installation
- student training
- expert support
- autonomous generation
- new: SMEFT re-casting

Expert user support

- improvements
- generate events

User (shortened)

Based on the precision-improvement axes identified above, the task is to construct and compare concrete simulation setups.

Simulation task: For each of the following strategies:

1. Higher-order QCD corrections (NLO+PS),
2. Finite-width and off-shell effects beyond the narrow-width approximation,
3. Hard additional radiation via matrix-element multi-jet merging,

do the following:

- Define a representative simulation setup implementing the respective strategy.
- Generate 50000 events for each setup using reasonable, standard choices.
- Clearly state all assumptions and simplifications, such as on-shell versus off-shell decays, jet multiplicities, or matching and merging schemes.

You may choose suitable tools, schemes, and defaults where needed, but avoid unnecessary over-optimisation.
[...]



MadAgents

Agents running code, specifically MadGraph [TP, Schiller, Schmal]

- installation
- student training
- expert support
- autonomous generation
- new: SMEFT re-casting

Expert user support

- improvements
- generate events
- compare events

MadAgents (shortened)

Concrete setups and files

Common to all three:

- Collider: pp at $\sqrt{s} = 13$ TeV.
- $m_t = 172.5$ GeV, $\Gamma_t \approx 1.32$ GeV.
- PDFs: NNPDF3.1 NLO (LHAPDF ID 303400).
- Final state used in the comparison: **only** $t\bar{t} \rightarrow be^+ \nu_e \bar{b}\mu^- \bar{\nu}_\mu$ (no $e^-\mu^+$).
- Fiducial selection (analysis level): exactly 1 e and 1 μ (OS), $p_T > 25$ GeV, $|\eta| < 2.5$, isolated; at least 2 jets with $p_T > 25$ GeV, $|\eta| < 2.5$ and at least 1 truth b-tag.

Samples and locations:

1. NLO+PS on-shell (higher-order QCD axis)

- Process: $pp \rightarrow t\bar{t}$ [QCD] at **NLO QCD**, on-shell tops.
- Decays: LO MadSpin, $t \rightarrow be^+ \nu_e$, $\bar{t} \rightarrow \bar{b}\mu^- \bar{\nu}_\mu$.
- Matching: aMC@NLO to Pythia8, hadronisation+UE on. [...]

2. LO+PS off-shell 2→6 (off-shell axis)

- Process: $pp \rightarrow e^+ \nu_e \mu^- \bar{\nu}_\mu b\bar{b}$ at **LO QCD**, with complex-mass scheme and all doubly-, singly- and non-resonant diagrams (full $W^+W^-b\bar{b}$ structure).
- No separate decay step: leptons and b's appear directly from the ME, with finite widths.
- Shower: Pythia8, hadronisation+UE on. [...]

3. LO MLM-merged $t\bar{t}$ (+0,1j)+PS (hard-radiation axis)

- Processes: $pp \rightarrow t\bar{t}$, $pp \rightarrow t\bar{t}j$ at **LO QCD** (0 and 1 extra parton; 2-jet MEs were dropped due to a technical issue).
- Merging: MLM k_T with Pythia8, $x_{\text{cut}} \approx 30$ GeV, dynamic scale $\sim H_T/4$.
- Decays: LO MadSpin to the same $e^+\mu^-$ channel as in NLO+PS. [...]



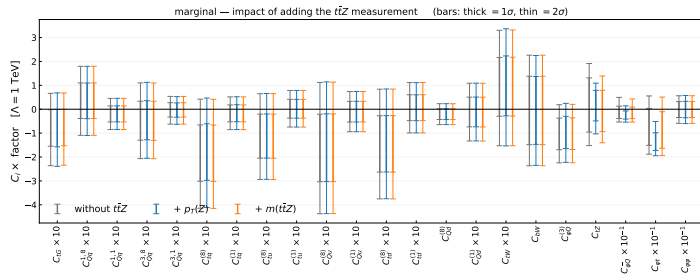
MadAgents

Agents running code, specifically MadGraph [TP, Schiller, Schmal]

- installation
- student training
- expert support
- autonomous generation
- new: SMEFT re-casting

‘Dramatically’ fast SMEFT update [Diefenbacher,TP, Schiller, Schmal; talk Matt LeBlanc]

- reproduce SFitter analysis
 - search for news $\rightarrow t\bar{t}Z$
 - implement $m_{t\bar{t}Z}$ and $p_{T,Z}$
 - re-simulate SMEFT
- Just LLMs running codes...



ML-event generation

Scientific AI

- just another tool for a numerical field?
an avalanche wiping out Math and then us?
a great opportunity!
- QCD applications:
parton densities
event generators
higher orders
hadronization
simulation-based inference
unfolding
global analyses ...
→ **It's delivery time!**

2211.01421v4 [hep-ph] 24 Apr 2025

Modern Machine Learning for LHC Physicists

Tilman Plehn^{a,b*}, Anja Butter^{a,c}, Barry Dillon^{a,d},
Theo Heimel^{a,e}, Claudius Krause^{a,f}, and Ramon Winterhalder^{a,g}

^a Institute for Theoretical Physics, Heidelberg University, Germany

^b Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany

^c LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France

^d Intelligent Systems Research Centre, Ulster University, Northern Ireland

^e CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium

^f HEPHY, Austrian Academy of Sciences, Vienna, Austria

^g TIFLab, Università degli Studi di Milano & INFN Sezione di Milano, Italy

April 25, 2025

Abstract

Depending on the point of view, modern machine learning is either providing an unprecedented boost to the numerical methods of particle physics, or it is transforming the way we do science with vast amounts of complex data. In any case, it is crucial for young researchers to stay on top of this development and apply cutting-edge methods and tools to all LHC physics tasks. These lecture notes lead students with basic knowledge of particle physics and significant enthusiasm for machine learning to relevant applications. They start with an LHC-specific motivation and a non-standard introduction to neural networks and then cover classification, unsupervised classification, generative networks, data representations, and inverse problems. Three themes defining much of the discussion are statistically defined loss functions, uncertainties, and accuracy. To understand the applications, the notes include some aspects of theoretical LHC physics. All examples are chosen from particle physics publications of the last few years, and many of them come with corresponding [tutorials](#).¹

