

Transforming Particle Theory

Tilman Plehn

Universität Heidelberg

Nikhef, April 2026



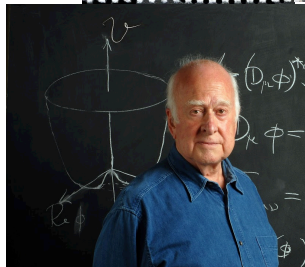
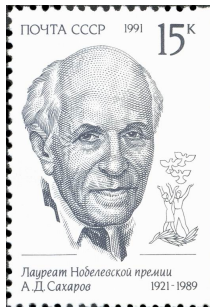
Modern LHC physics

Unchanged motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

Precision-LHC physics

- fundamental questions
- answers hidden in complexity
- huge data set
- first-principle simulations



Modern LHC physics

Unchanged motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

Precision-LHC physics

- fundamental questions
- answers hidden in complexity
- huge data set
- first-principle simulations

Starting with Lagrangian

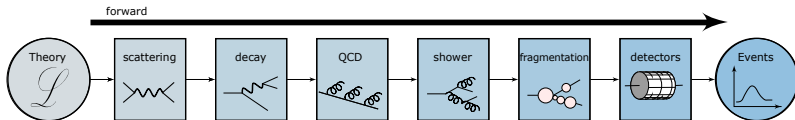
- calculate scattering
- simulate QCD [do not 'model QCD data']
- plus detector simulations

→ Events in hypothetical worlds

Analyses

- compare simulations and data
- complement with anomaly searches

→ Infer underlying Lagrangian



Representation learning

Just an upscaled fit

- approximate $f_\theta(x) \approx f(x)$
- x phase space
- f_θ numerical function

→ θ data representation

Phase space probabilities

- regression $x \rightarrow A_\theta(x)$
- classification $x \rightarrow p_\theta(x)$ [likelihood ratio]
- generation $r \sim \mathcal{N} \rightarrow x \sim p_\theta(x)$
- conditional generation $r \sim \mathcal{N} \rightarrow x \sim p_\theta(x|y)$

What to encode in θ ?

1. known symmetries
2. uncertainties
3. explainable structures

→ Physics representation learning [cusp.AI $\otimes$ NNPDF $\otimes$ Jesse Thaler]



Symmetry-aware networks

Benefit from known symmetries [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

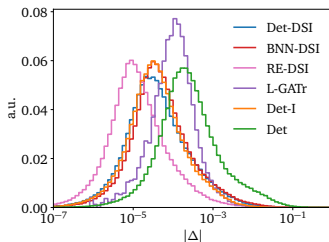
- leading-order loop amplitude $gg \rightarrow \gamma\gamma g(g)$
- curve of dimensionality unavoidable

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

- accuracy over phase space

$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)}$$

- impact of symmetric input



Symmetry-aware networks

Benefit from known symmetries [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

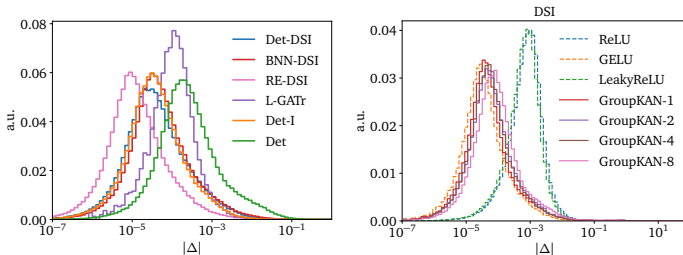
- leading-order loop amplitude $gg \rightarrow \gamma\gamma g(g)$
- curse of dimensionality unavoidable

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

- accuracy over phase space

$$\Delta(x) = \frac{A_{\text{NN}}(x) - A_{\text{true}}(x)}{A_{\text{true}}(x)}$$

- impact of symmetric input and activation function



...training, regularization, loss alchemy...

→ Symmetry-aware representations a theory task



Lorentz equivariance

Equivariant transformers [Brehmer, Bres0, de Haan, TP, Qu, Spinner, Thaler; Favaro, TP, Spinner + Hamprecht group]

- **permutation invariance** $\rightarrow$ graph or transformer
- **Lorentz equivariance** $\rightarrow \Lambda(f_\theta(x)) = f_\theta(\Lambda(x))$
- learned symmetry breaking
- idea vs implementation?

- 1- L-GATr geometric algebra representation
- 2- LLoCa local reference frame per particle



Lorentz equivariance

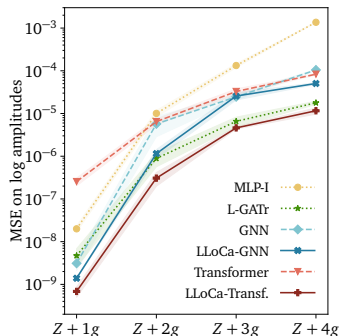
Equivariant transformers [Brehmer, Bresó, de Haan, TP, Qu, Spinner, Thaler; Favaro, TP, Spinner + Hamprecht group]

- **permutation invariance** → graph or transformer
- **Lorentz equivariance** → $\Lambda(f_\theta(x)) = f_\theta(\Lambda(x))$
- learned symmetry breaking
- idea vs implementation?

- 1- L-GATr geometric algebra representation
- 2- LLoCa local reference frame per particle

Higher-multiplicity amplitudes

- transition amplitudes $q\bar{q} \rightarrow Z + 1 \dots 4 g$
 - equivariant encoding
 - architectures: MLP, graph, transformer
- **Transformer for higher multiplicities**



Equivariant classification

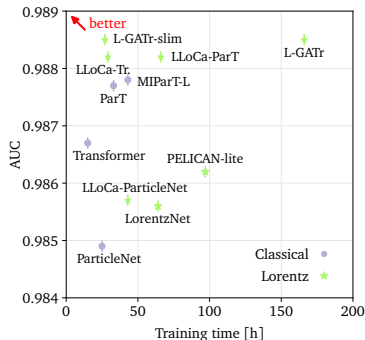
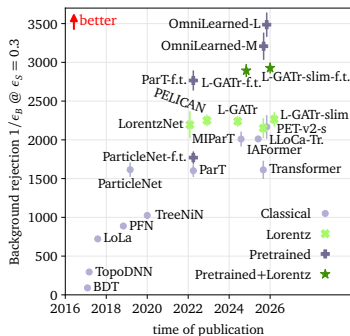
(Fat-) jet tagging [top, Higgs, W/Z]

- basis of every LHC analysis
- performance critical
- 30x better than best BDT [same for flavor tagging; ask ATLAS & CMS]

→ Transforming analyses

- jet axis breaking Lorentz-symmetry
- encoded vs learned symmetries [Breso, Greif, Mikuni, Wamorkar,...]

→ Resilience, uncertainties, implementation next



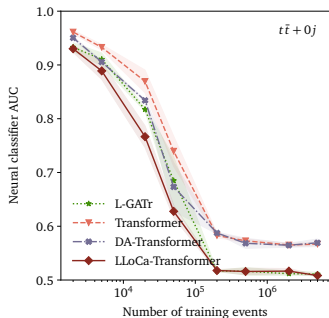
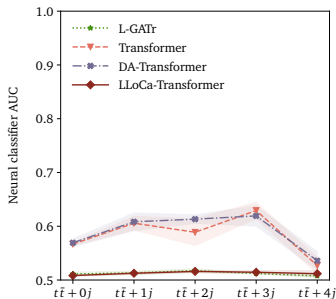
Equivariant generation

ML-event generation

- theory input to LHC analyses
- starting point for many applications
- jet-level events

$$pp \rightarrow t_{\text{had}} \bar{t}_{\text{had}} + 0 \dots 4j$$

- conditional flow matching with transformer-velocity
 - performance: [training-generation classifier](#)
- [Accurate phase space a solved problem...](#)



Learned uncertainties

Learned probability

- amplitude over phase phase

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

- internal representation θ of training data T [think Gaussian with mean and width]

$$p(A) = \int d\theta p(A|\theta) p(\theta|T)$$

→ θ -distribution defining Bayesian NN [cf NNPDF replicas]

Variational approximation

- definition of training

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

- similarity through minimal KL-divergence [Bayes' theorem to remove unknown posterior]

$$\begin{aligned} D_{\text{KL}}[q(\theta), p(\theta|T)] &= \int d\theta q(\theta) \log \frac{q(\theta)}{p(\theta|T)} \\ &= \int d\theta q(\theta) \log \frac{q(\theta)p(T)}{p(T|\theta)p(\theta)} \\ &\approx D_{\text{KL}}[q(\theta), p(\theta)] - \int d\theta q(\theta) \log p(T|\theta) \equiv \mathcal{L} \end{aligned}$$

→ Loss: regularization + likelihood



Learned uncertainties

Learned probability

- amplitude over phase phase

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

- internal representation θ of training data T [think Gaussian with mean and width]

$$p(A) = \int d\theta p(A|\theta) p(\theta|T)$$

→ θ -distribution defining Bayesian NN [cf NNPDF replicas]

Training-data-related uncertainties

- statistical — vanishing for perfect training: $q(\theta) \rightarrow \delta(\theta - \theta_0)$ [reducible, epistemic]

$$\sigma_{\text{stat}}^2 = \int d\theta q(\theta) [\bar{A}(\theta) - \langle A \rangle]^2$$

- systematic — vanishing for perfect data: $p(A|\theta) \rightarrow \delta(A - A_0)$ [irreducible, aleatoric]

$$\sigma_{\text{syst}}^2 = \int d\theta q(\theta) [\bar{A}^2(\theta) - \bar{A}(\theta)^2]$$

→ Complete description of training-related uncertainties



Evidential regression

Uncertainties without sampling [Bahl, Elmer, TP, Winterhalder]

- variational inference again

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

assuming $p(A|\lambda) = \mathcal{N}(A|\bar{A}, \sigma^2)$ or $\lambda \equiv (\bar{A}, \sigma^2)$

- choose conjugate prior [inverse Gamma function]

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

- explicit posterior [Student-t]

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

- learn $\{\gamma, \nu, \alpha, \beta\}(x)$

→ Explicit mean and uncertainties

$$A_{\text{NN}} = \gamma$$

$$\sigma_{\text{syst}}^2 = \frac{\beta}{\alpha - 1}$$

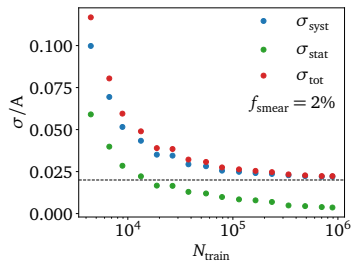
$$\sigma_{\text{stat}}^2 = \frac{\beta}{\nu(\alpha - 1)}$$



Calibrated uncertainties

Back to $gg \rightarrow \gamma\gamma g(g)$ [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

- systematics from artificial noise [statistics saturated]

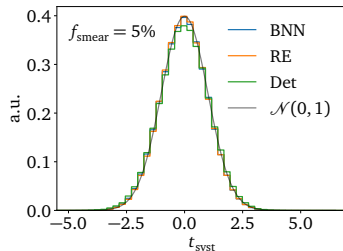
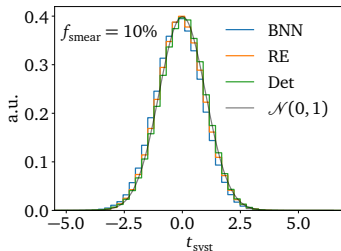


Calibrated uncertainties

Back to $gg \rightarrow \gamma\gamma g(g)$ [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

- systematics from artificial noise [statistics saturated]
- pull over phase space [like a closure test]

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



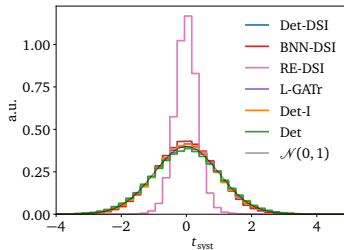
Calibrated uncertainties

Back to $gg \rightarrow \gamma\gamma g(g)$ [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

- systematics from artificial noise [statistics saturated]
- pull over phase space [like a closure test]

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

→ Independent of accuracy [ensembles?]



Calibrated uncertainties

Back to $gg \rightarrow \gamma\gamma g(g)$ [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

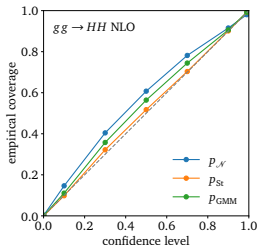
- systematics from artificial noise [statistics saturated]
- pull over phase space [like a closure test]

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

→ Independent of accuracy [ensembles?]

Trusted amplitudes [Bahl, Braun, Heinrich, TP, Revelli]

- HH 2-loop — calibrated uncertainties with Student- t likelihood



Calibrated uncertainties

Back to $gg \rightarrow \gamma\gamma g(g)$ [Bahl, Elmer, Favaro, Haußmann, TP, Winterhalder]

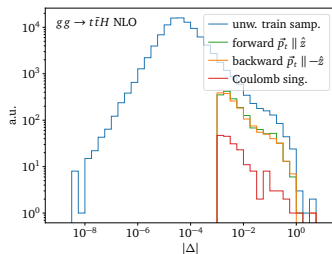
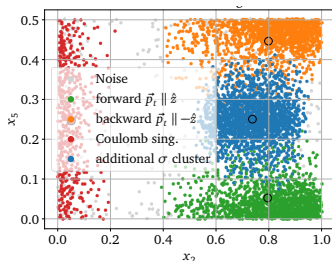
- systematics from artificial noise [statistics saturated]
- pull over phase space [like a closure test]

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

→ Independent of accuracy [ensembles?]

Trusted amplitudes [Bahl, Braun, Heinrich, TP, Revelli]

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



Generalization uncertainties

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

- ‘How many events can I sample from a network trained on n events?’

fewer, not perfect training

more, generative network smoothens $p(x)$

- learned phase space densities

$$p_{\text{gen}}(x) \approx p_{\text{train}}(x) \sim p_{\text{true}}(x) \quad \Rightarrow \quad p_{\text{gen}}(x) \stackrel{?}{\sim} p_{\text{true}}(x)$$

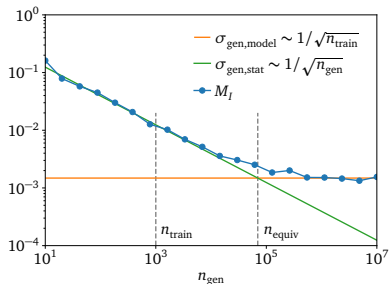
- two uncertainties on sampled events

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

- sweet spot

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

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



Generalization uncertainties

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

- ‘How many events can I sample from a network trained on n events?’

fewer, not perfect training

more, generative network smoothens $p(x)$

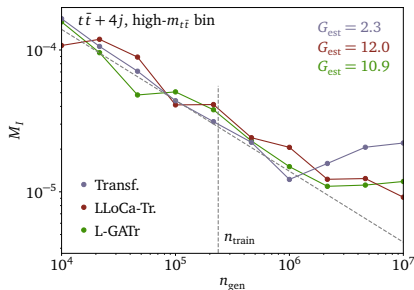
- learned phase space densities

$$p_{\text{gen}}(x) \approx p_{\text{train}}(x) \sim p_{\text{true}}(x) \quad \Rightarrow \quad p_{\text{gen}}(x) \stackrel{?}{\sim} p_{\text{true}}(x)$$

- test on generated data with measure M

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

→ L-GATr and LLoCa do amplify...



Applied by MadGraph, Sherpa & friends

Accelerating event generators [DeCrescenzo, Marino Villadamigo, Elmer, Heimel, TP, Winterhalder, Zaro]

- virtual amplitude surrogates [w/ uncertainties]
- real amplitude surrogates [w/ uncertainties]
- neural importance sampling [replace VEGAS by normalizing flow]
- learned channel weights
- NLO: conditional on FKS sectors

→ Standard method, ML-enhanced, GPU-ready

Sampling mode	Surrogates		$e^+e^- \rightarrow u\bar{u}g$		$e^+e^- \rightarrow u\bar{u}gg$	
	$R_\theta^{V/B}$	$\Sigma_{ij,\theta}$	σ_{NLO} [pb]	$\text{Var}(w_{\text{NLO}})/\sigma_{\text{NLO}}^2$	σ_{NLO} [pb]	$\text{Var}(w_{\text{NLO}})/\sigma_{\text{NLO}}^2$
VEGAS	$\times$	$\times$	0.10750(34)	100(14)	0.08769(27)	2400(130)
MADNIS	$\times$	$\times$	0.10760(19)	30.4(26)	0.08729(15)	720(90)
MADNIS	$\checkmark$	$\times$	0.10759(18)	28.8(29)	0.08711(18)	870(160)
MADNIS	$\checkmark$	$\checkmark$	0.10765(18)	27.4(11)	0.08738(15)	730(50)



Applied by MadGraph, Sherpa & friends

Accelerating event generators [DeCrescenzo, Marino Villadamigo, Elmer, Heimel, TP, Winterhalder, Zaro]

- virtual amplitude surrogates [w/ uncertainties]
- real amplitude surrogates [w/ uncertainties]
- neural importance sampling [replace VEGAS by normalizing flow]
- learned channel weights
- NLO: conditional on FKS sectors

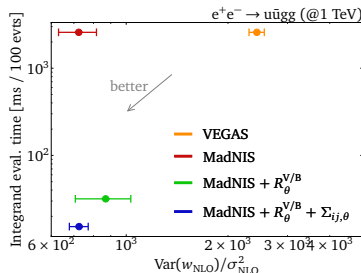
→ Standard method, ML-enhanced, GPU-ready

Sampling mode	Surrogates		$e^+e^- \rightarrow u\bar{u}g$		$e^+e^- \rightarrow u\bar{u}gg$	
	$R_\theta^{V/B}$	$\Sigma_{ij,\theta}$	σ_{NLO} [pb]	$\text{Var}(w_{\text{NLO}})/\sigma_{\text{NLO}}^2$	σ_{NLO} [pb]	$\text{Var}(w_{\text{NLO}})/\sigma_{\text{NLO}}^2$
VEGAS	✗	✗	0.10750(34)	100(14)	0.08769(27)	2400(130)
MADNIS	✗	✗	0.10760(19)	30.4(26)	0.08729(15)	720(90)
MADNIS	✓	✗	0.10759(18)	28.8(29)	0.08711(18)	870(160)
MADNIS	✓	✓	0.10765(18)	27.4(11)	0.08738(15)	730(50)

First lessons

- surrogates all the way
- learn ratios V/B
- still no solution for cancellations
- multiplicity scaling still hard

→ How close to 'NLO as fast as LO?'



Latent representation

Principle components quark–gluon tagging [Vent, Winterhalder, TP]

- PC₁: constituent number and diversity

$$n_{\text{pf}} = \sum_i 1 \quad \text{and} \quad S_{\text{PID}} = - \sum_{\text{type } j} f_j \log f_j$$

- PC₂: radial energy profile

$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \quad C_{0.2} = \frac{\sum_{i < j} p_{T,i} p_{T,j} (\Delta R_{ij})^{0.2}}{(\sum_i p_{T,i})^2} \quad r_\lambda = \frac{\lambda_{0.5}^1}{\lambda_1^2} \quad \lambda_k^\beta = \sum_i z_i^\beta \Delta R^k$$

- PC₃: fragmentation and energy dispersion

$$S_{\text{frag}} = - \sum_i z_i \log z_i$$

- PC_{4,5}: charge information etc



Latent representation

Principle components quark–gluon tagging [Vent, Winterhalder, TP]

- PC₁: constituent number and diversity

$$n_{\text{pf}} = \sum_i 1 \quad \text{and} \quad S_{\text{PID}} = - \sum_{\text{type } j} f_j \log f_j$$

- PC₂: radial energy profile

$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \quad C_{0.2} = \frac{\sum_{i < j} p_{T,i} p_{T,j} (\Delta R_{ij})^{0.2}}{(\sum_i p_{T,i})^2} \quad r_\lambda = \frac{\lambda_{0.5}^1}{\lambda_1^2} \quad \lambda_k^\beta = \sum_i z_i^\beta \Delta R^k$$

- PC₃: fragmentation and energy dispersion

$$S_{\text{frag}} = - \sum_i z_i \log z_i$$

- PC_{4,5}: charge information etc

Symbolic Regression

- classifier output

$$p_{\text{quark}} = \tanh^3 \left[0.55 \cdot C_{0.2} + 2 \left(-0.02 \cdot r_\lambda \cdot (C_{0.2} \cdot p_T D \cdot S_{\text{PID}} \cdot S_{\text{frag}} - 0.25) + 1 \right)^3 \right]$$

→ Formulas as effective regularizers? [Bahl, Fuchs, Menem, TP]

observables	model	AUC	Rej _{30%}
(n _{pf} , p _T D, C _{0.2} , r _λ , S _{PID} , S _{frag} , E _Q)	MLP	0.872	66.87
	PySR	0.871	66.58



Latent information geometry

Information geometry for learned representation [Kuntz, TP, Schäfer, Schosser, Vent]

- Fisher metric, Amari-Chentsov tensor, and Fisher-Rao distance

$$F_{ij}(\theta) = \int d\mu(z) p_{\theta}(z) \partial_i \log p_{\theta}(z) \partial_j \log p_{\theta}(z)$$

$$G_{ijk}(\theta) = \int d\mu(z) p_{\theta}(z) \partial_i \log p_{\theta}(z) \partial_j \log p_{\theta}(z) \partial_k \log p_{\theta}(z)$$

$$\rho_F = \int_0^1 dt \sqrt{F_{ij}(\gamma(t)) \dot{\gamma}^i(t) \dot{\gamma}^j(t)}$$

tracing curvature, nonmetricity, distance



Latent information geometry

Information geometry for learned representation [Kuntz, TP, Schäfer, Schosser, Vent]

- Fisher metric, Amari-Chentsov tensor, and Fisher-Rao distance

$$F_{ij}(\theta) = \int d\mu(z) p_{\theta}(z) \partial_i \log p_{\theta}(z) \partial_j \log p_{\theta}(z)$$

$$C_{ijk}(\theta) = \int d\mu(z) p_{\theta}(z) \partial_i \log p_{\theta}(z) \partial_j \log p_{\theta}(z) \partial_k \log p_{\theta}(z)$$

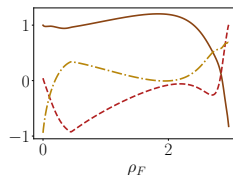
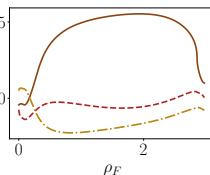
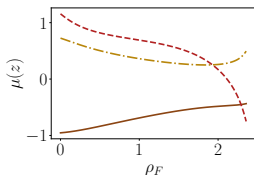
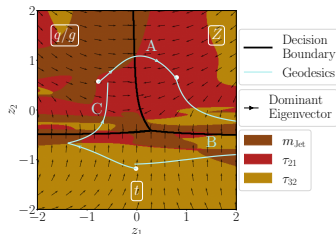
$$\rho_F = \int_0^1 dt \sqrt{F_{ij}(\gamma(t)) \dot{\gamma}^i(t) \dot{\gamma}^j(t)}$$

tracing curvature, nonmetricity, distance

Latent questions

- 1- Reconstruction and classification aligned?
- 2- Is a Z-jet in the middle between q/g and t ?

...



MadAgents

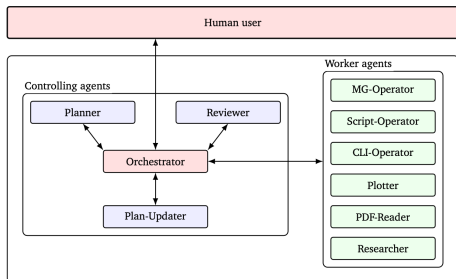
Interactive LLMs to accelerate workflows [TP, Schiller, Schmal]

1. install MadGraph-Pythia-Delphes-ROOT
2. support and train inexperienced users
3. support experienced users
4. generate events autonomously

→ Not physics, but great tool!

Standard structure

- benefit from public code & documentation
 - natural language interaction
command line interaction
 - agents through prompting
- ChatGPT w/ API keys or Claude



MadAgents at work

#3 Better $t\bar{t}$ simulations

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 at work

#3 Better $t\bar{t}$ simulations

- suggested 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 at work

#3 Better $t\bar{t}$ simulations

- suggested improvements
- ask for 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 at work

#3 Better $t\bar{t}$ simulations

- suggested improvements
- ask for events
- comparison of simulated 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 $\rightarrow$ 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. [...]

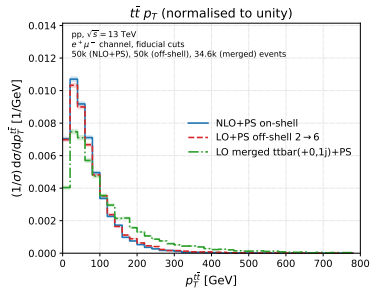


MadAgents at work

#3 Better $t\bar{t}$ simulations

- suggested improvements
- ask for events
- comparison of simulated events

→ Explained set of plots



MadAgents at work

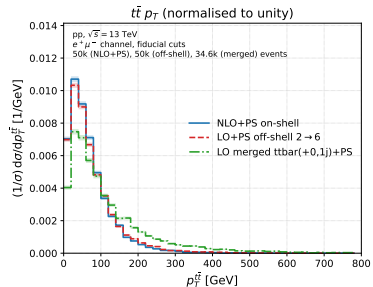
#3 Better $t\bar{t}$ simulations

- suggested improvements
- ask for events
- comparison of simulated events

→ Explained set of plots

#4 Autonomous event generation

- reproduce a dataset from a pdf file



User

Please implement and execute the workflow described in the paper. As a final deliverable, I want the generated events, a histogram of the resonance reconstruction, and a bash script that can execute the end-to-end workflow (including creation of the resonance histogram). All deliverables must be saved, you may create suitable subdirectories.

Constraint: Your generated events and reconstructions must match the paper as closely as possible.

Persistence: Do not present partial or simplified results. Present the final results only if you believe you have recreated the workflow to the best of your ability and that the outputs match the paper as closely as reasonably possible.

Assumptions & Decisions: If any part of the paper is unclear or ambiguous, do not ask me. Instead, make a reasonable decision. In your final reply, list every unclear or ambiguous point you encountered and, for each one, state the decision you made and briefly explain your reasoning.



MadAgents at work

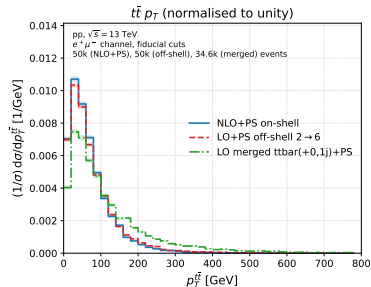
#3 Better $t\bar{t}$ simulations

- suggested improvements
- ask for events
- comparison of simulated events

→ Explained set of plots

#4 Autonomous event generation

- reproduce a dataset from a pdf file



Plan

- Step 1:** Locate PDF and set up project directories.
- Step 2:** Extract physics and workflow specifications from the PDF and references.
- Step 3:** Inspect and prepare the software environment.
- Step 4:** Implement MadGraph and event generation workflow.
- Step 5:** Implement reconstruction, resonance analysis, and histogram production.
- Step 6:** Create end-to-end bash workflow script.
- Step 7:** Validate full workflow and prepare final deliverables and ambiguity summary.



MadAgents at work

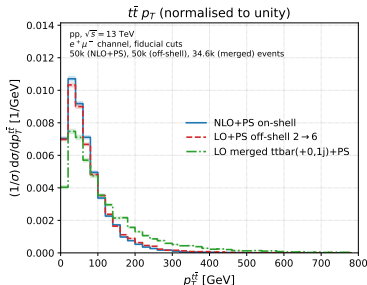
#3 Better $t\bar{t}$ simulations

- suggested improvements
- ask for events
- comparison of simulated events

→ Explained set of plots

#4 Autonomous event generation

- reproduce a dataset from a pdf file
- provide an analysis note and know your limits



MadAgents (shortened)

Simulation and analysis setup

The signal under study is a simplified scalar leptoquark S_1 with quantum numbers $S_1 \sim (\bar{3}, 1, 1/3)$, pair-produced in proton-proton collisions, followed by the decay chain

$$pp \rightarrow S_1 S_1^\dagger \rightarrow (\ell^+ j)(\ell^- j) \Rightarrow 2\ell + 2j \text{ final state}, \quad (5)$$

closely following the workflow outlined in the HEPTAPOD [51] short note and the MC4BSM-style tutorials.

Event generation.

Events are generated with MadGraph5_aMC@NLO v3.7.0 [5] interfaced to Pythia 8.3 [1] for parton showering and hadronisation. We use an existing public leptoquark UFO model, LQn1o_5FNS_v5_UFO, originally constructed in the FeynRules framework [54], and identify the HEPTAPOD scalar S_1 with the particle LQ1d

[...]

All non-trivial choices that are not fixed by the HEPTAPOD note (collider energy, PDF set, specific UFO model, Yukawa pattern, use of hadron-level rather than detector-level objects, exact cuts and binning) are documented in a separate “ambiguities and decisions” note that accompanies this analysis.



MadAgents at work

#3 Better $t\bar{t}$ simulations

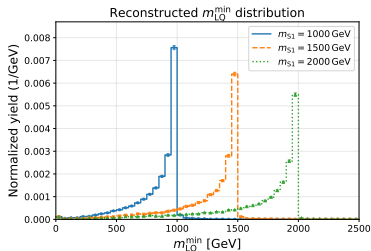
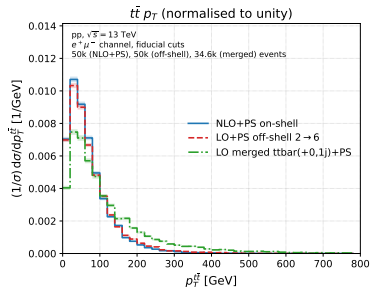
- suggested improvements
- ask for events
- comparison of simulated events

→ Explained set of plots

#4 Autonomous event generation

- reproduce a dataset from a pdf file
- provide an analysis note and know your limits
- make the money plot

→ Just the starting point for research...



ML for particle theory

Scientific AI

- just another tool for a numerical field?
an avalanche which wiping out our scientific culture?
transformative new tool box for LHC theory?

- representation learning:

symmetries
uncertainties
explainability

- applications:

event generators
higher orders
hadronization
simulation-based inference
unfolding

...

→ It's delivery time!

221.01421v4 [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



Bitter lesson?

Scaling with compute [Bahl, Bresó, Butter, Iturriza Ramirez]

- Richard Sutton (2019): 'The biggest lesson is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law. Most AI research has been conducted as if the computation available were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time massively more computation inevitably becomes available.'



Bitter lesson?

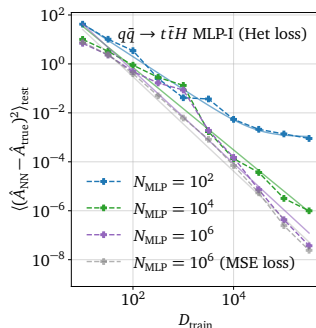
Scaling with compute [Bahl, Bres0, Butter, Iturriza Ramirez]

- accuracy given intrinsic dimension $[x \in \{N, D, C\}]$

$$\text{MSE}(x) \sim \frac{1}{x^{\alpha_x}} + K_x \quad \text{with} \quad \alpha_x > \frac{4}{d}$$

- $\alpha_x \sim 1 \dots 3$
scaling with d ? [motivation for transformer]
effects inducing K_x
improved by loss, activation, representation?

→ Moore's law relevant for HL-LHC?



Bitter lesson?

Scaling with compute [Bahl, Bres0, Butter, Ituriza Ramirez]

- accuracy given intrinsic dimension $[x \in \{N, D, C\}]$

$$\text{MSE}(x) \sim \frac{1}{x^{\alpha_x}} + K_x \quad \text{with} \quad \alpha_x > \frac{4}{d}$$

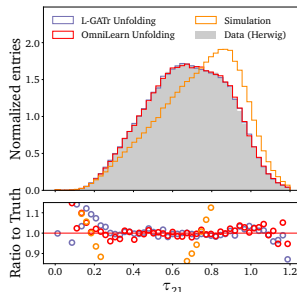
- $\alpha_x \sim 1 \dots 3$
scaling with d ? [motivation for transformer]
effects inducing K_x
improved by loss, activation, representation?

→ Moore's law relevant for HL-LHC?

Explicit vs implicit physics knowledge [Breso, Greif, Mikuni, Wamorkar,...]

- L-GATr vs OmniLearn
- classification with similar classes:
Z+jets unfolding
H1 unfolding
anomaly detection
- known: competition driving performance
unclear: what works better?
challenge: make networks more efficient

→ If the question is 'or' the answer is 'and'



Extrapolating ISR

Universal QCD jet radiation [Z + 1...8 jets]

- from n to $n + 1$ jets

$$R_{(n+1)/n} = \frac{\sigma_{n+1}}{\sigma_n} \quad \text{and} \quad P(n) = \frac{\sigma_n}{\sigma_{\text{tot}}} \quad \text{with} \quad \sigma_{\text{tot}} = \sum_{n=0}^{\infty} \sigma_n.$$

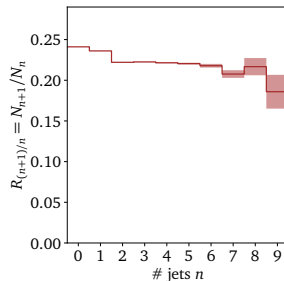
- large scale drop: Poisson scaling

$$R_{(n+1)/n} = \frac{\bar{n}}{n+1} \quad \Leftrightarrow \quad P(n) = \frac{\bar{n}^n e^{-\bar{n}}}{n!}.$$

- democratic scales: staircase scaling

$$R_{(n+1)/n} = e^{-b} = 1 - \tilde{\Delta}_g(Q^2) \equiv P(n+1|n)$$

→ Universal pattern learnable?



Extrapolating ISR

Universal QCD jet radiation [Z + 1...8 jets]

- democratic scales: staircase scaling

$$R_{(n+1)/n} = e^{-b} = 1 - \tilde{\Delta}_g(Q^2) \equiv P(n+1|n)$$

→ Universal pattern learnable?

Autoregressive transformer [Butter, Charton, Villadamigo, Ore, TP, Spinner]

- maximize factorized probability loss

$$p(x_i | x_{1:i-1}) = p_{\text{kin}}(x_i | x_{1:i-1}) p_{\text{split}}(x_{1:i-1})$$

$$p(x_{1:n}) = \left[\prod_{i=1}^n p_{\text{kin}}(x_i | x_{1:i-1}) \right] \left[\prod_{i=1}^n p_{\text{split}}(x_{1:i-1}) \right] [1 - p_{\text{split}}(x_{1:n})]$$

- train up to 6 jets, generate 7 and 8 jets

→ full extrapolation with right latent representation

