Machine Learning for Phase Space Sampling

An exploration for LHC simulated event generation with SHERPA

Enrico Bothmann

Institut für Theroetische Physik, Universität Göttingen

ACAT 2022

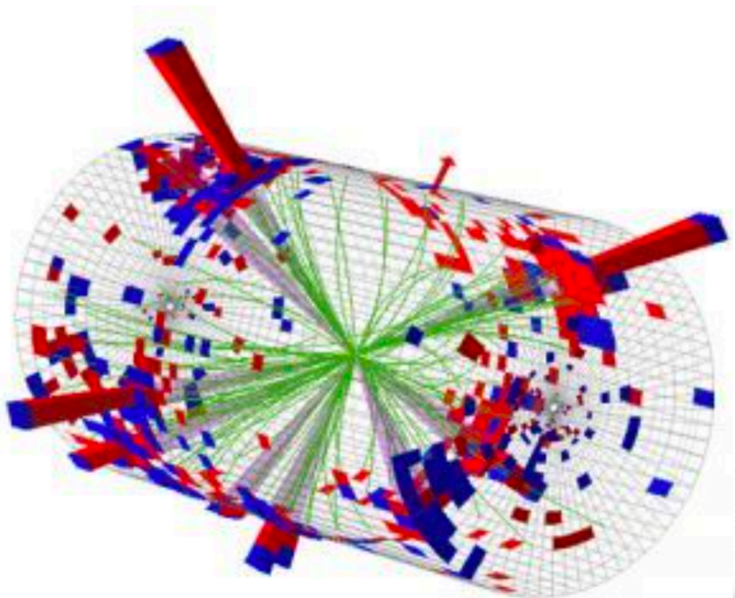
Several slides adapted from Steffen Schumann

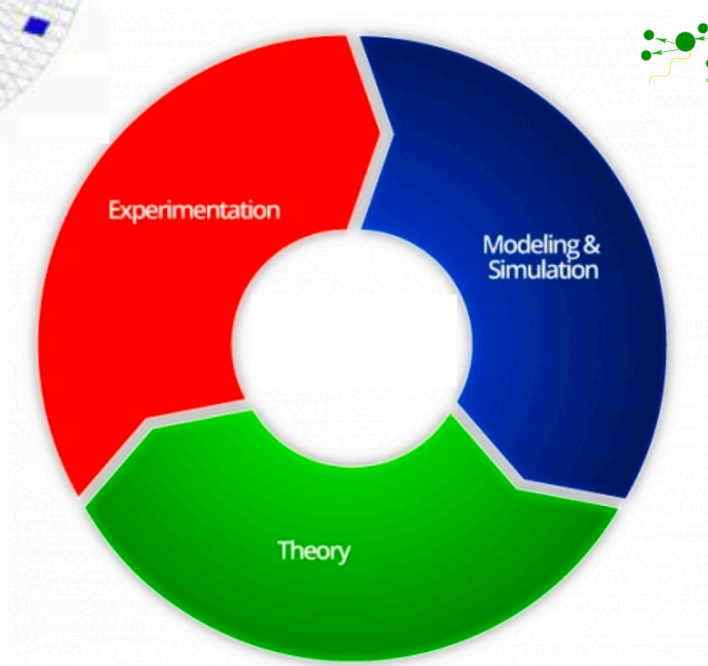
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

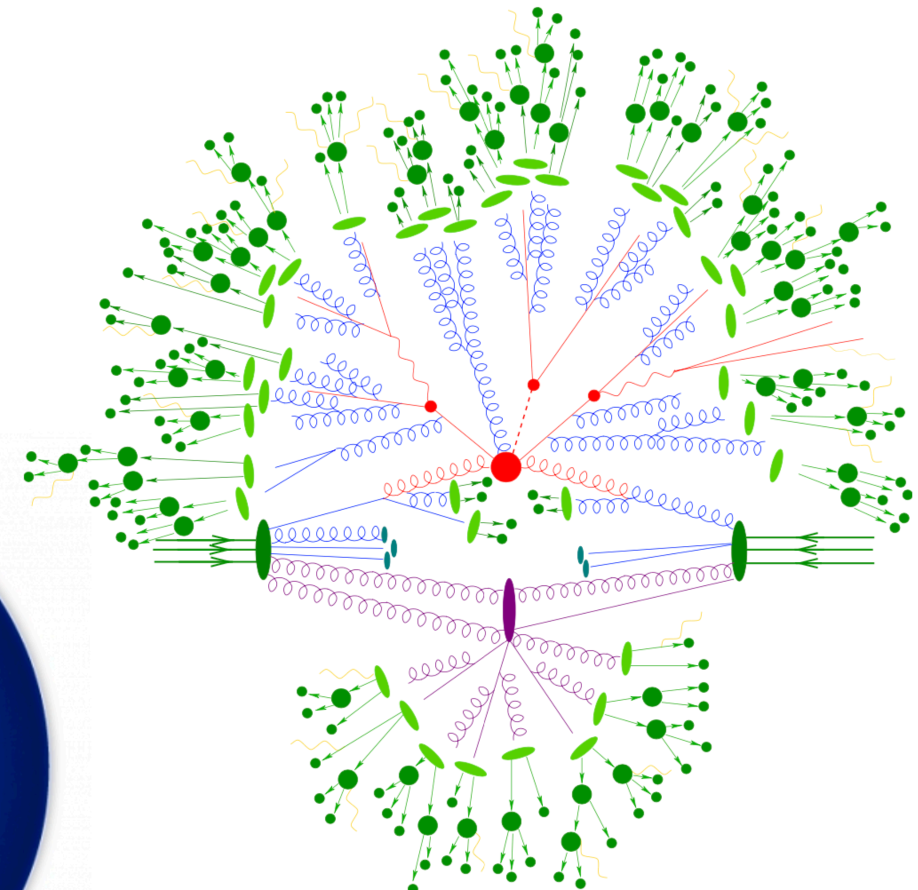
MCnet

Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

The High Energy Physics Triad







$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\Psi}\not{D}\Psi + h.c.$$

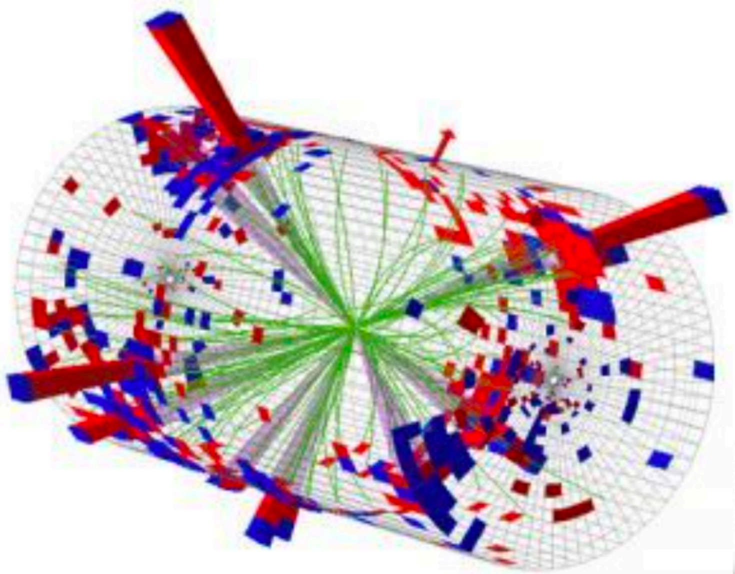
Slide adapted from Stefan Höche

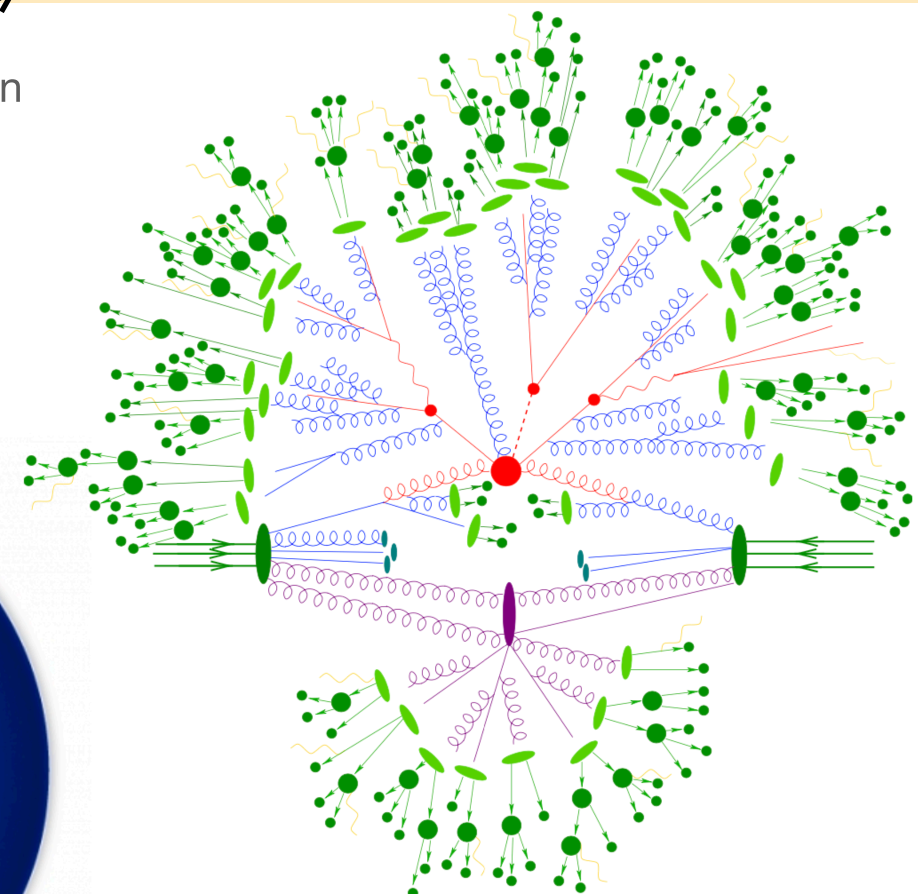
The High Energy Physics Triad

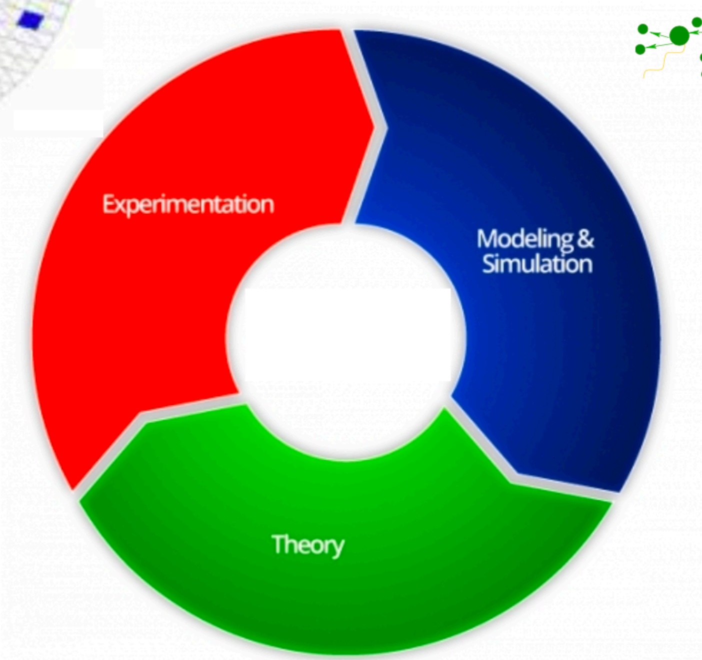
Random collision ▶ Detector response ▶ Trigger ▶ Event Reconstruction ▶ Event Sample

Random numbers ▶ Monte Carlo event generator ▶ Detector simulation ▶ Event Reconstruction ▶ Simulated Event Sample

statistical comparison



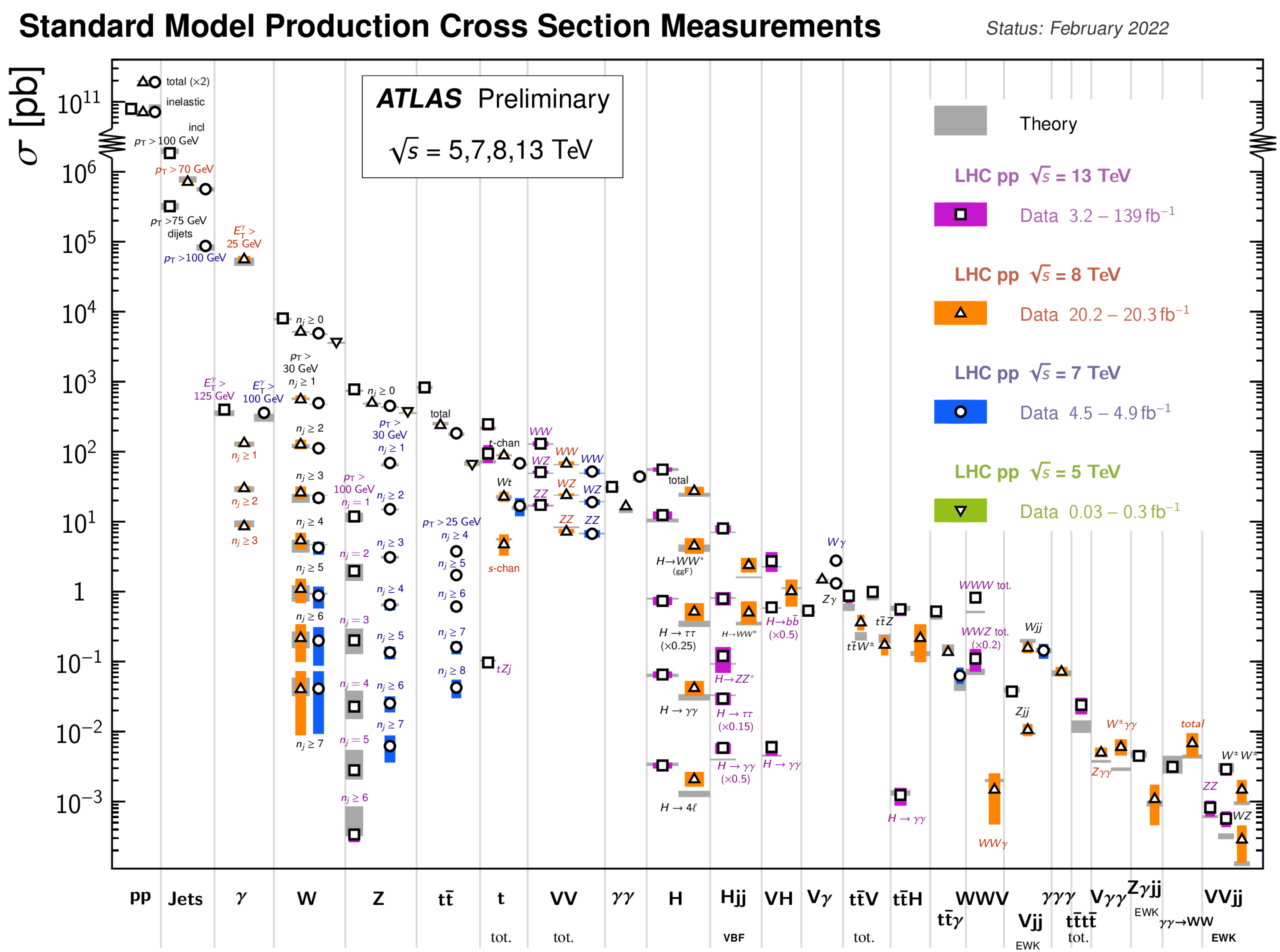




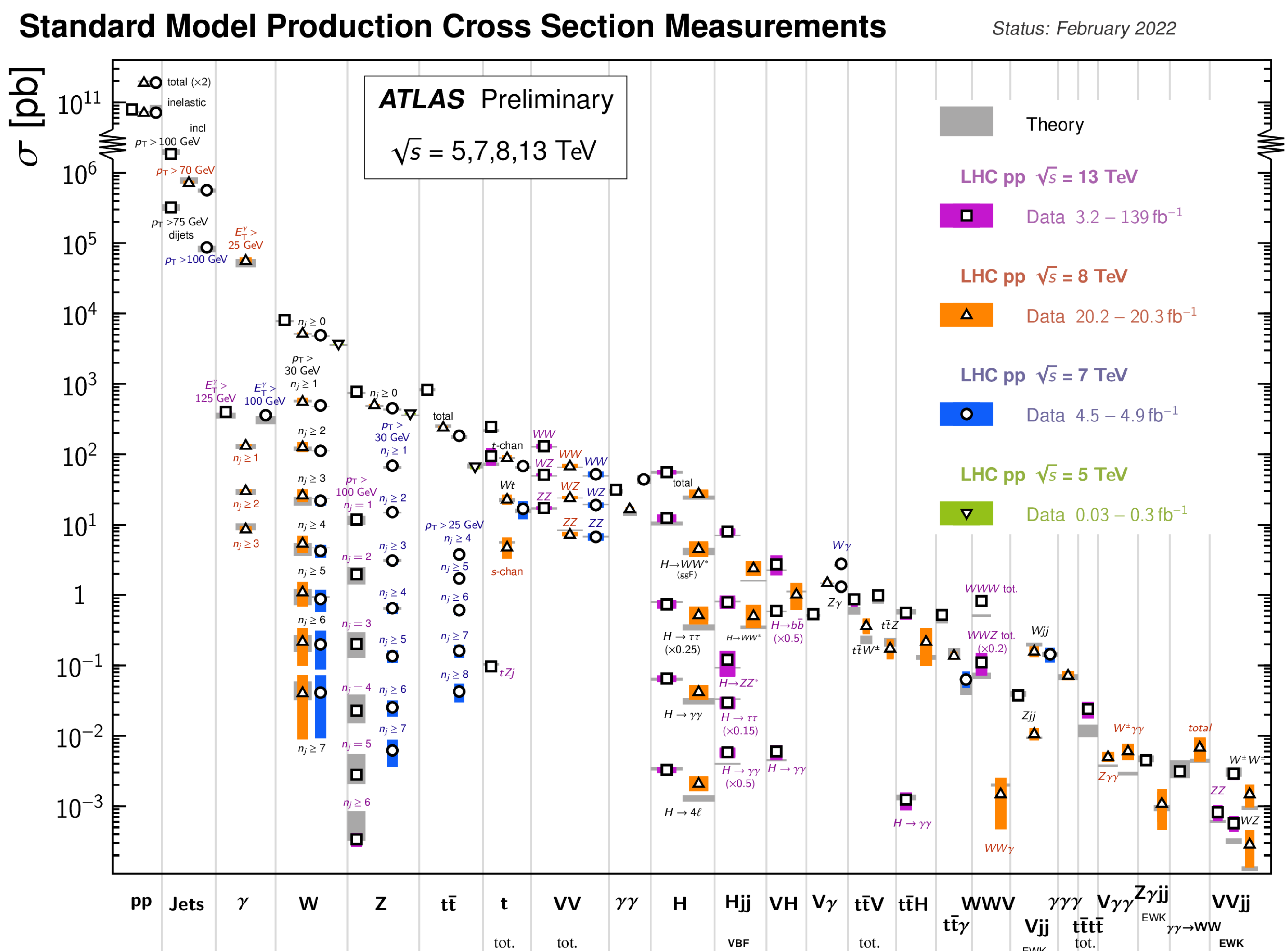
$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi} \not{D} \psi + h.c.$$

Slide adapted from Stefan Höche

LHC Process Zoo



LHC Process Zoo

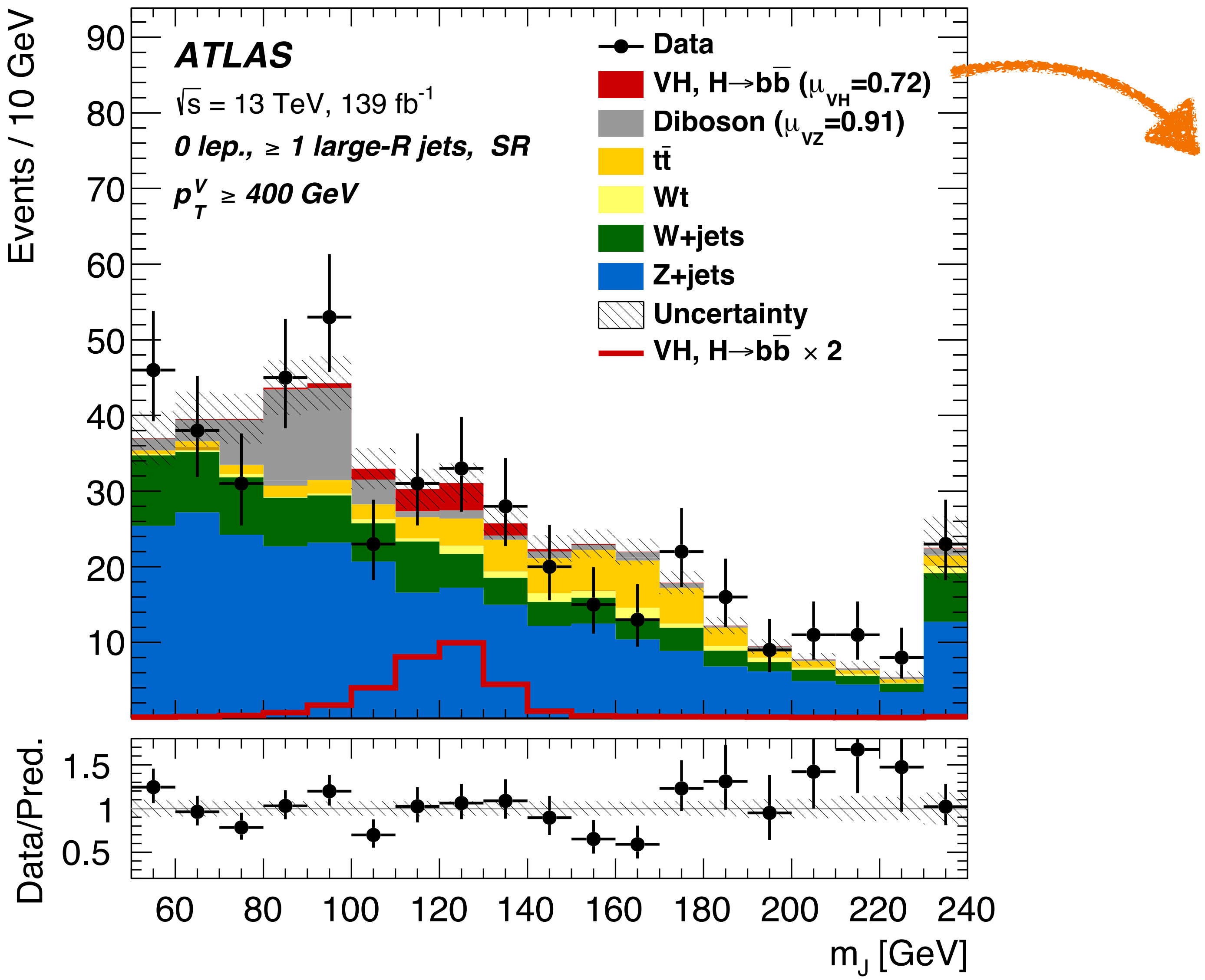


signal processes for SM measurements & BSM searches
often associated with several jets

V+jets: background to most other processes & large rates even with several extra jets

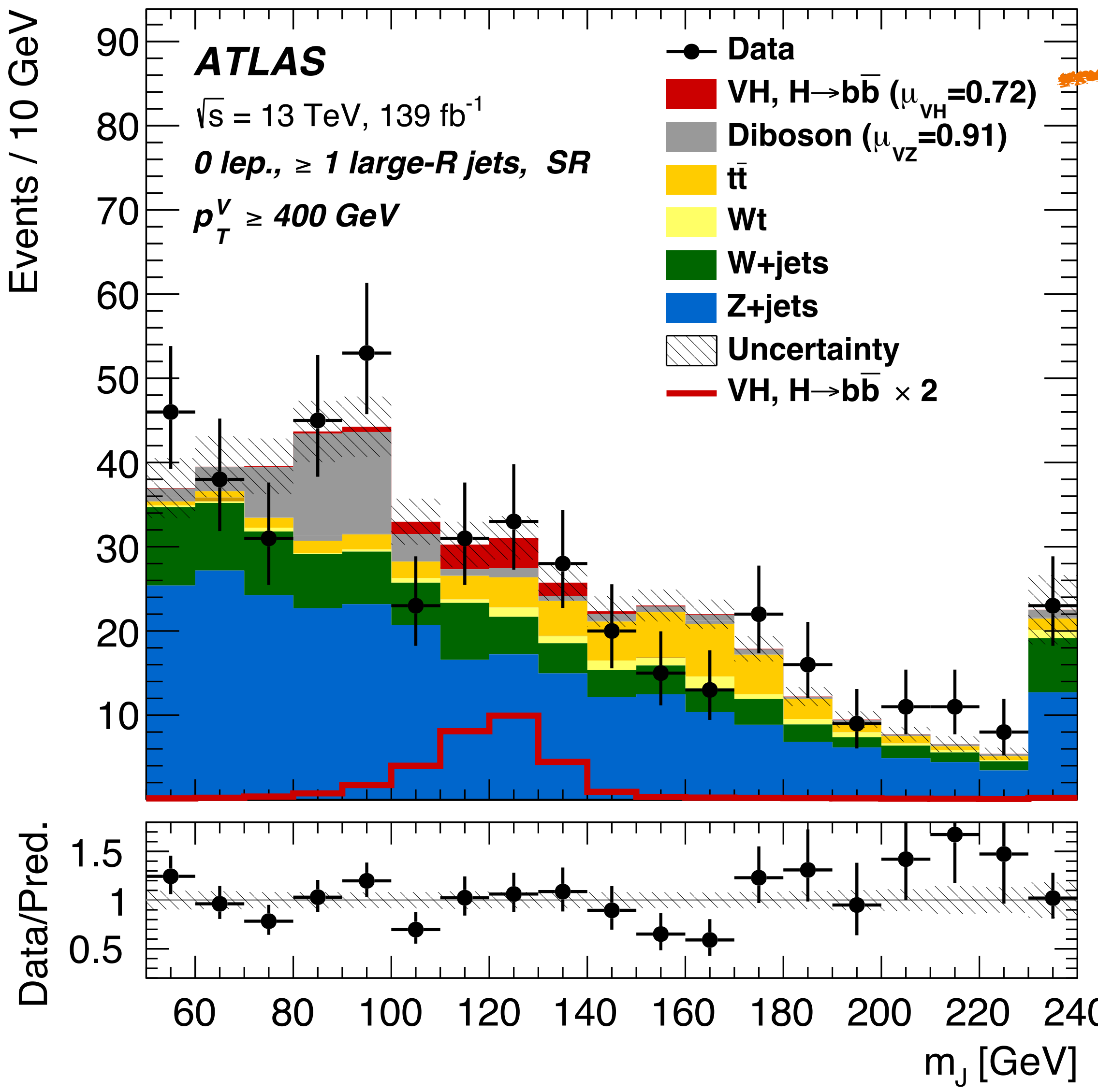
Typical Use Case & Need for Precision

[G. Aad et al. [ATLAS], Phys. Lett. B **816** (2021), 136204]



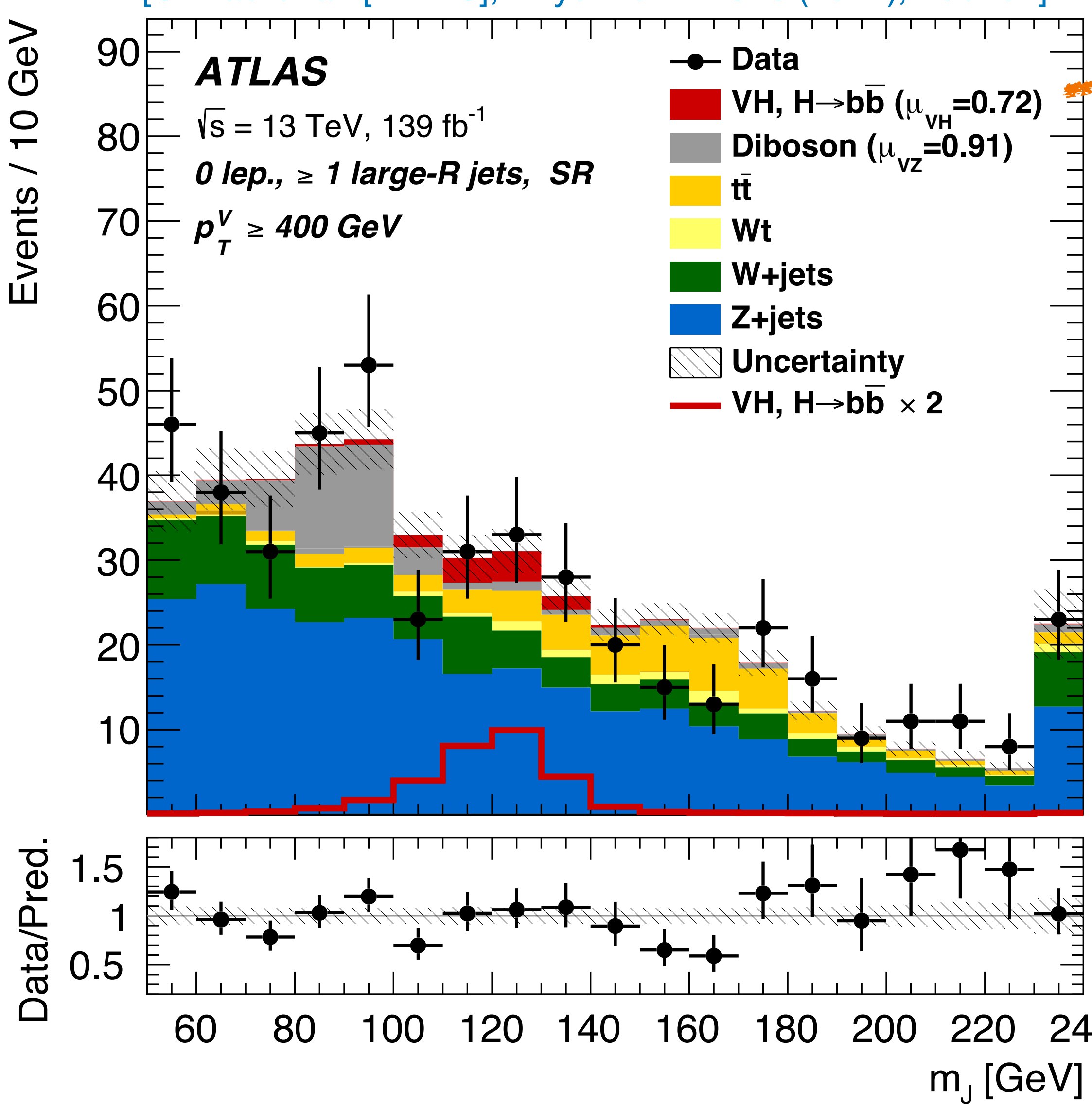
Typical Use Case & Need for Precision

[G. Aad et al. [ATLAS], Phys. Lett. B **816** (2021), 136204]



“measured signal yield to that predicted by the Standard Model, is $0.72^{+0.39}_{-0.36}$ ”

Typical Use Case & Need for Precision



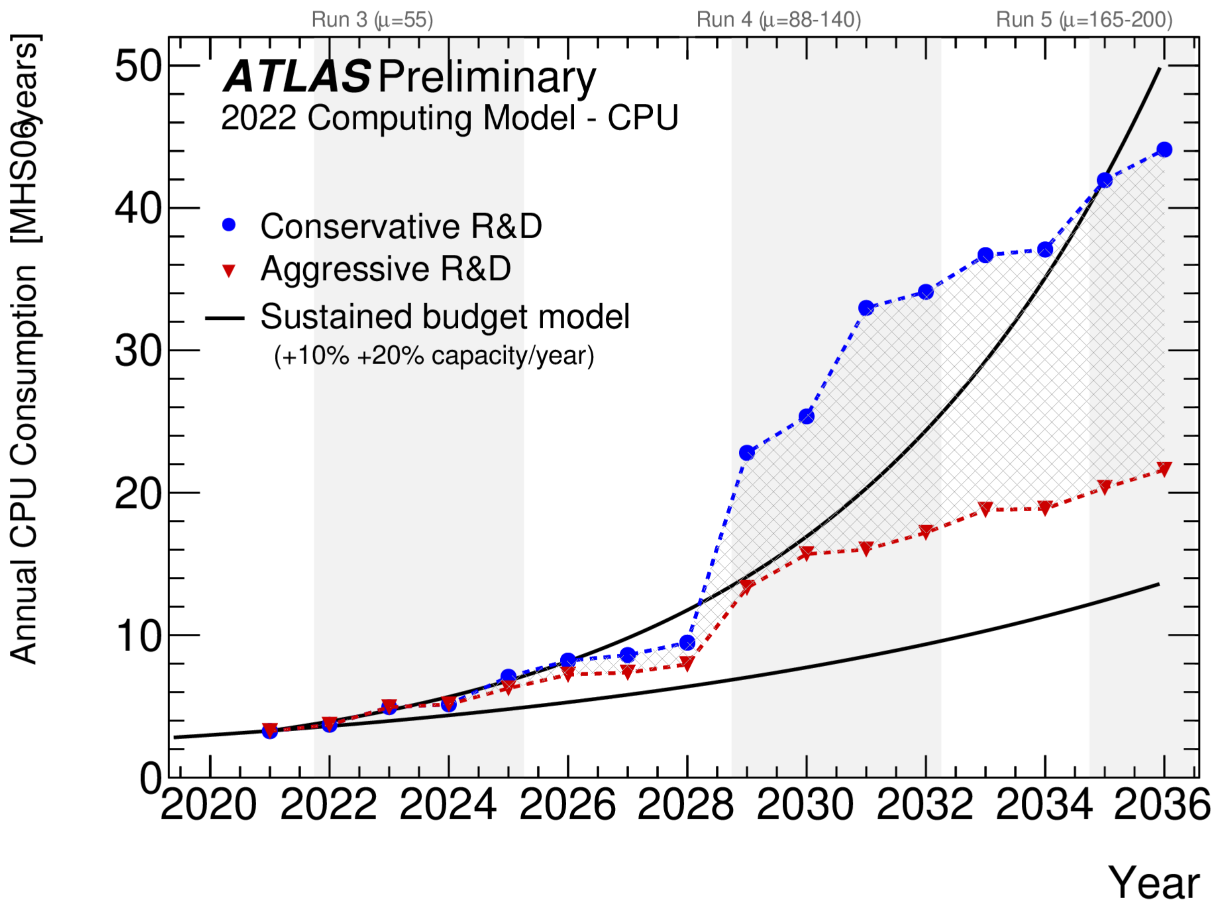
"measured signal yield to that predicted by the Standard Model, is $0.72^{+0.39}_{-0.36}$ "

- results like this need (automated) first-principle forward simulation for all relevant processes
  - accuracy in perturbative expansion: (N)NLO QCD, (N)LO EW, NLL etc.
  - multi-jet predictions
  - quantifiable uncertainties for model parameters, perturbation theory, etc.

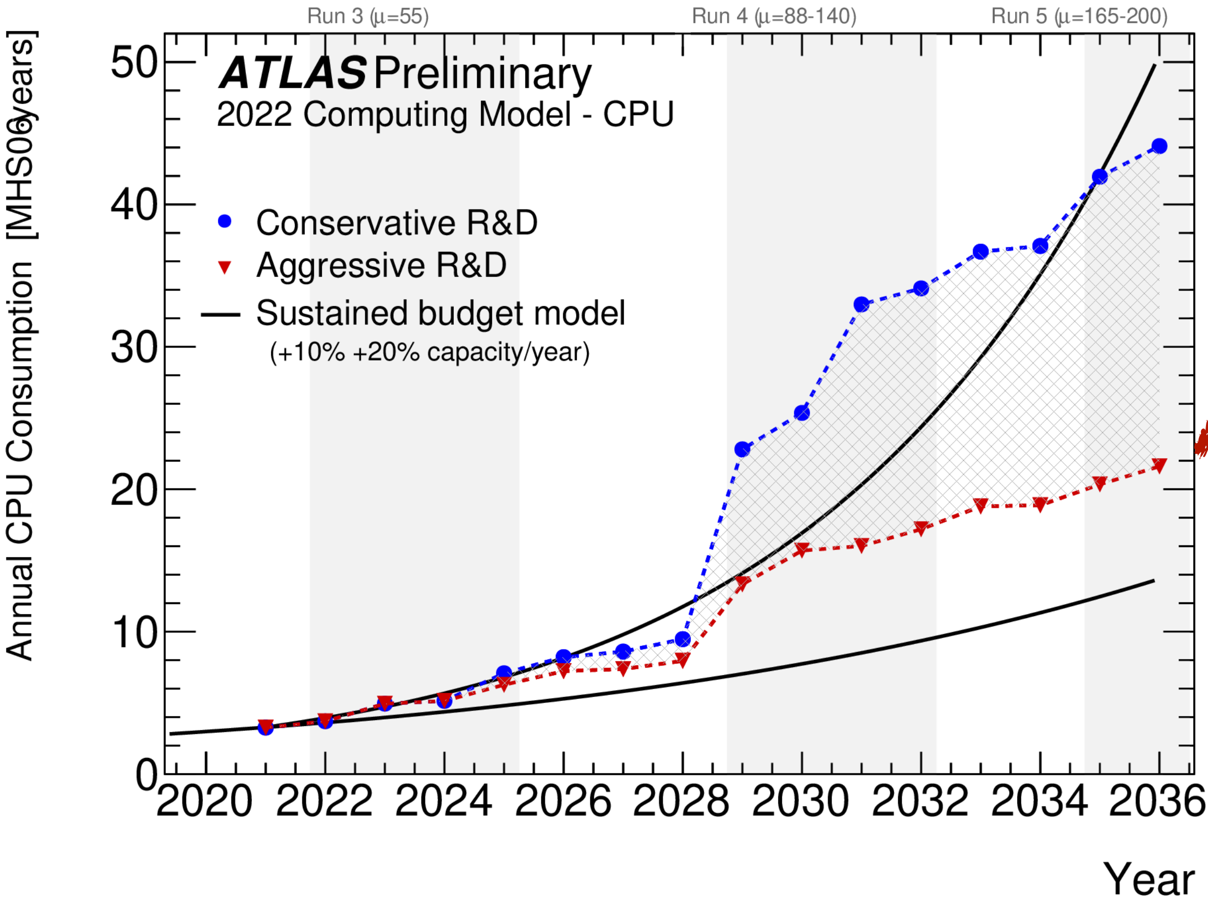
➡ MCEG provide this

➡ but precision simulations are expensive

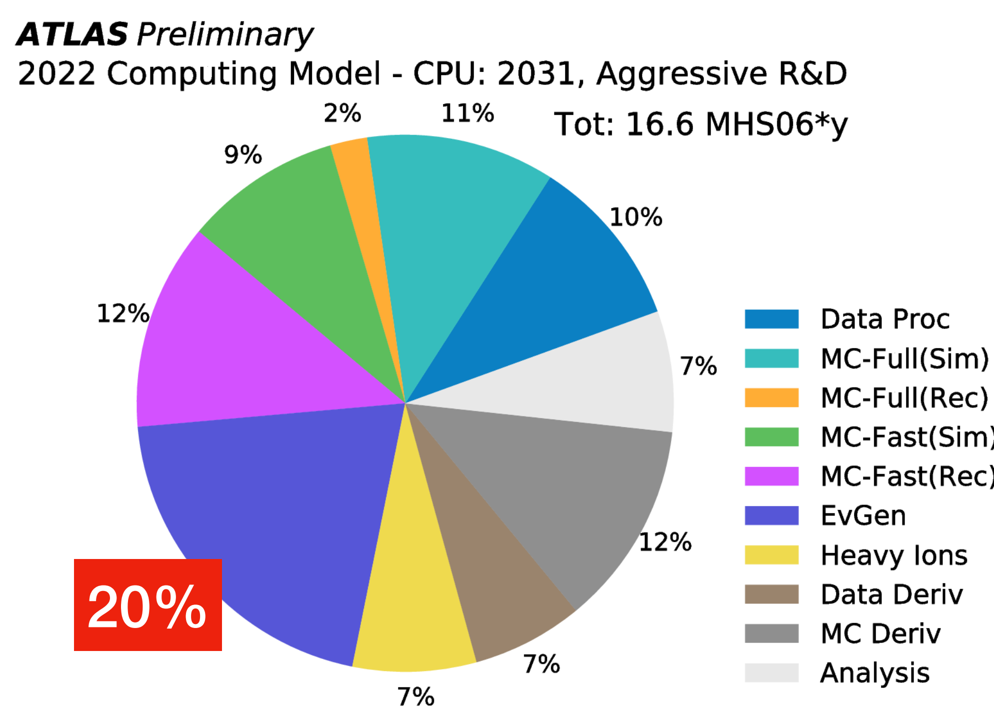
Event generation computing challenge



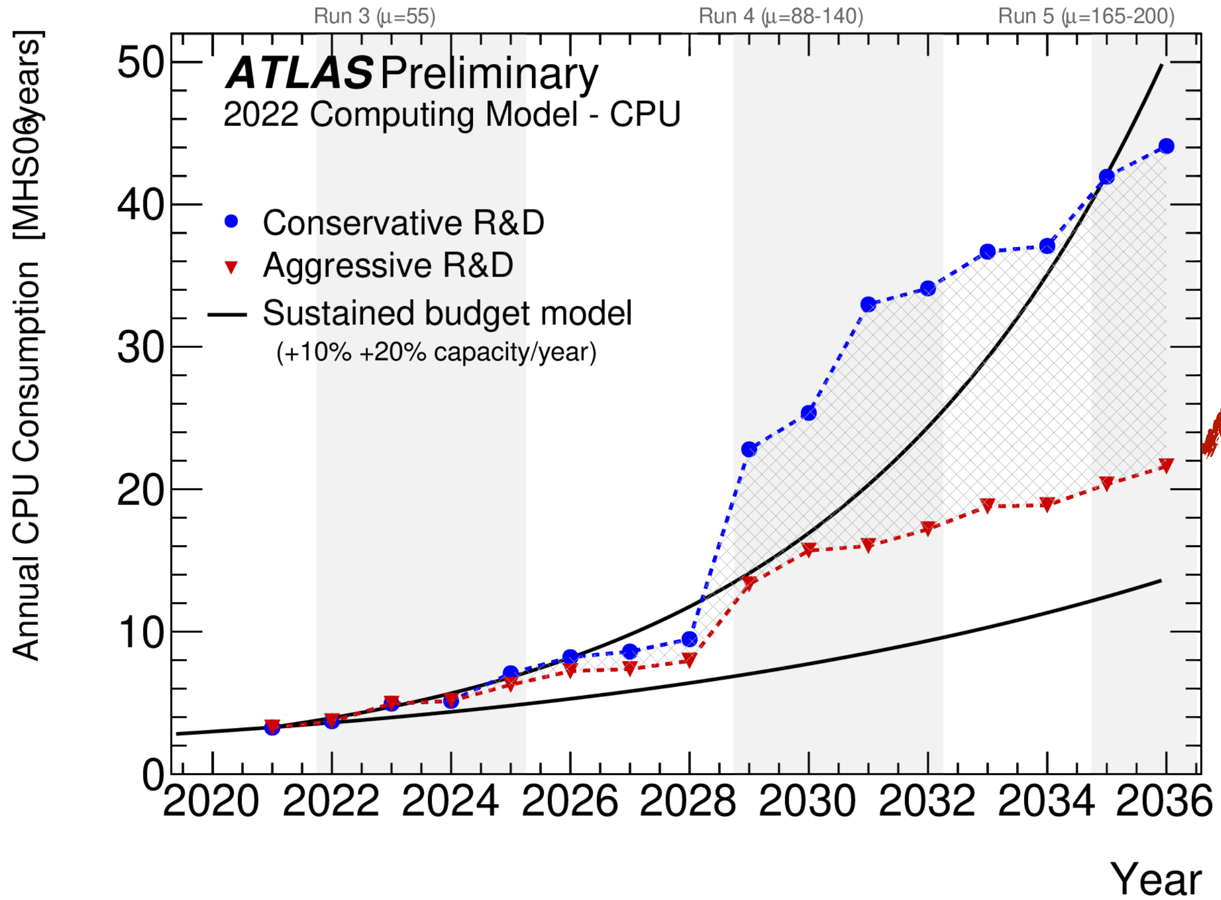
Event generation computing challenge



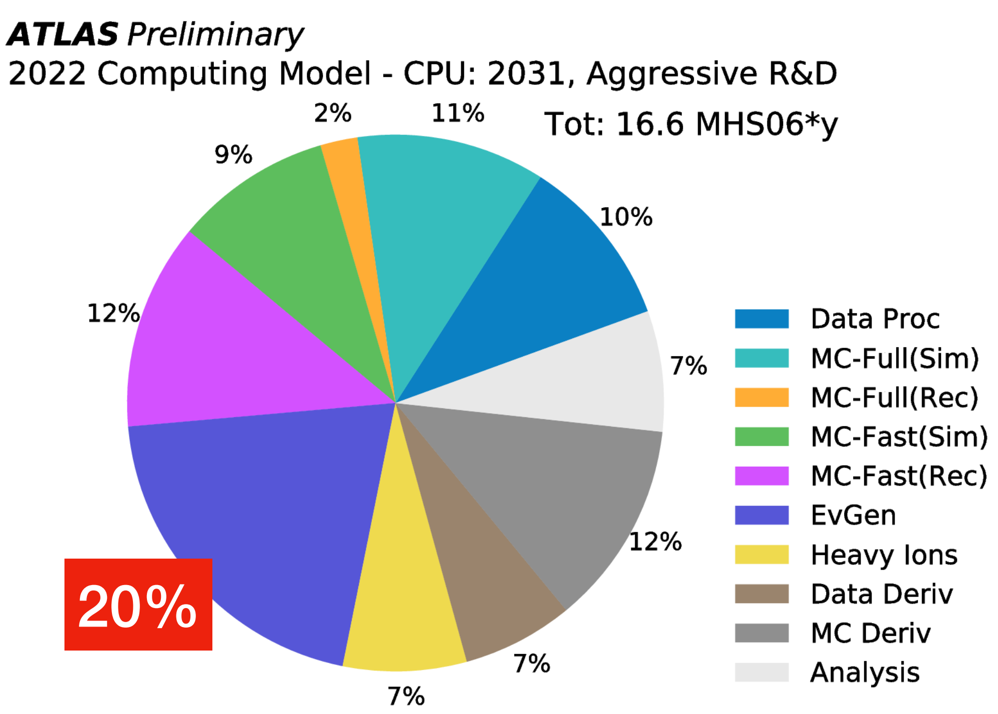
composed of …



Event generation computing challenge



composed of …

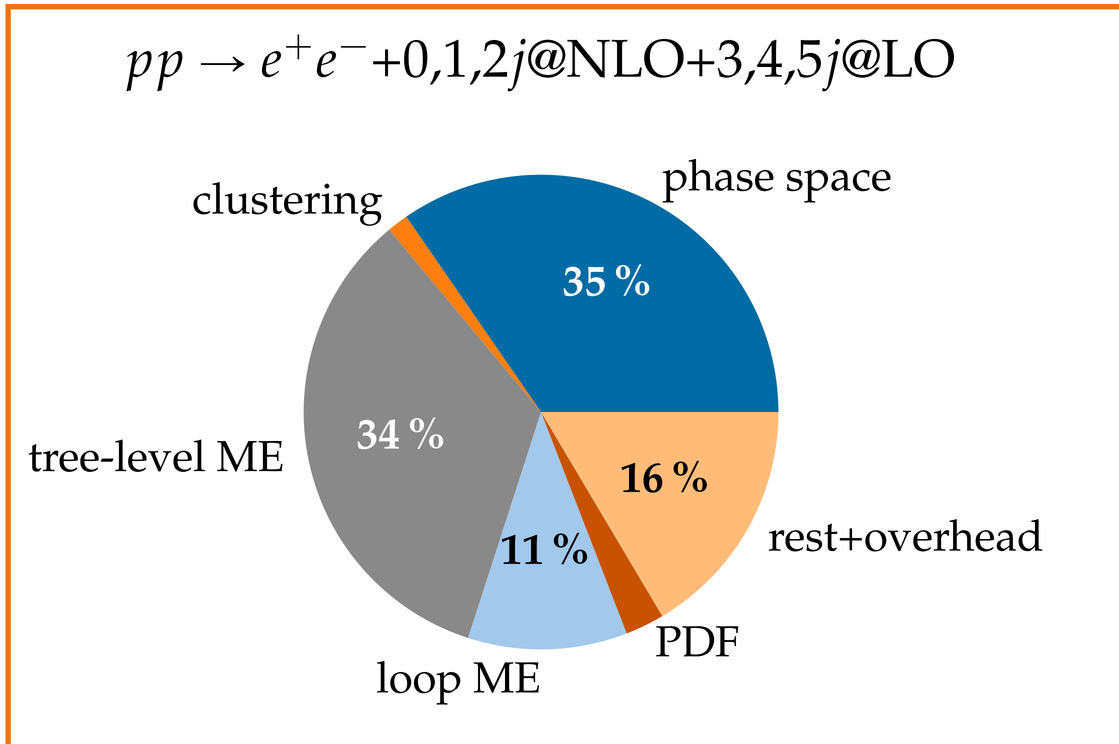


- Run-III & HL-LHC CPU requirements spark renewed interest in efficiency
  [ATLAS HL-LHC Computing CDR, Valassi et al Challenges in MCEG software for HL-LHC arXiv:2004.13687]
- Event Generators is significant consumer of CPU hours
- Beyond addressing this important issue …
  - … faster simulation also enables physics: better accuracy, higher jet multiplicity, etc.

Event Generation Bottlenecks

Our job is to sample this integral:

$$\sigma_{pp\to X_n} = \sum_{ab} \int \mathrm{d}x_a \mathrm{d}x_b \, \mathrm{d}\Phi_n \; f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \, |\mathcal{M}_{ab\to X_n}|^2 \, \Theta_n(p_1, \ldots, p_n)$$



relative CPU time usage for a typical ATLAS V+jets setup with SHERPA

(after series of optimisations which give a ~40x speed-up, thus ticking off one of the major HSF generator WG milestones)

[EB et al. 2209.00843]
[Talk by C. Gutschow]

➡ relevant remaining bottleneck: phase-space & matrix elements $|\mathcal{M}|^2$ (ME)

- better phase-space sampling directly reduces number of ME evaluations

When expensive integrands meet poor sampling efficiencies

Our job is to sample this integral:

$$\sigma_{pp\to X_n} = \sum_{ab} \int \mathrm{d}x_a \mathrm{d}x_b \, \mathrm{d}\Phi_n \, f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \, |\mathcal{M}_{ab\to X_n}|^2 \, \Theta_n(p_1, \ldots, p_n)$$

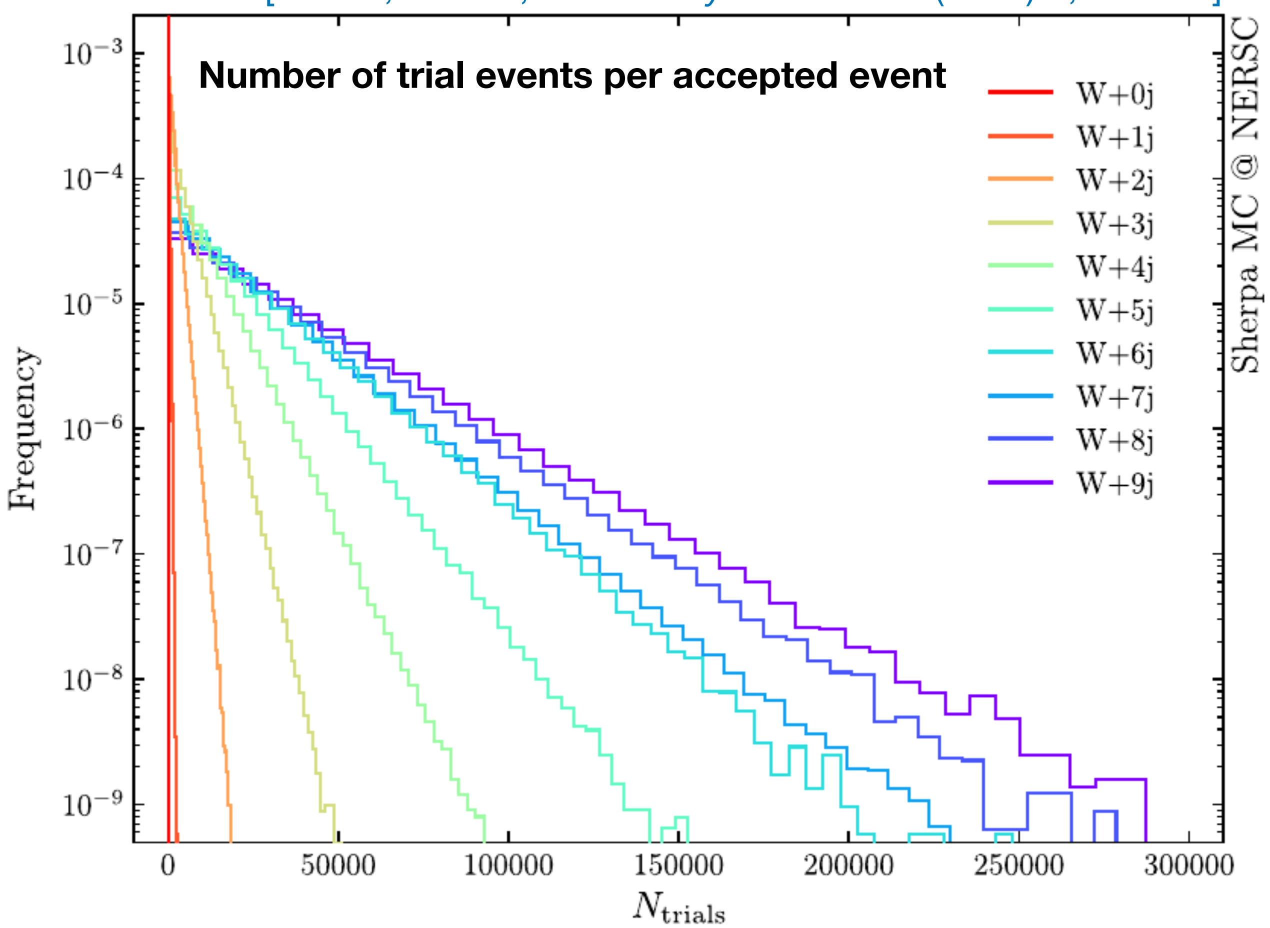
First figure of merit:

$$\epsilon = \frac{1}{\langle N_{\mathrm{trials}} \rangle}$$

(rejection sampling efficiency)

- Low sampling efficiency ϵ, why?
  - multi-modal, wildly fluctuating target distribution $f_a f_b |\mathcal{M}|^2$
  - subject to non-trivial acceptance cuts Θ_n
  - high dimensionality $\dim[\Phi_n] = 3n - 4$
- Good news: Fairly generic sampling/ integration problem

[Höche, Prestel, Schulz *Phys.Rev.D* 100 (2019) 1, 014024]



Can we use ML/NN as a remedy without compromising on precision requirements?

Multi-channel importance sampling

- Consider generic integral over target function $f(x), x \in V \subseteq \mathbb{R}^d$
- Choose variable mapping $y : V \to U \subseteq \mathbb{R}^d$

$$I = \int_V \mathrm{d}^d x\, f(x) = \int_U \mathrm{d}^d y\, \frac{f(x)}{g(x)}\bigg|_{x \equiv x(y)} \quad \text{with} \quad \left|\frac{\partial y(x)}{\partial x}\right| = g(x)$$

↪ reduce variance of MC estimate
through suitable $g(x)$, such that $w = f/g \approx \text{const}\,.$

Second figure of merit:

$$\mathrm{Var} \approx \frac{\langle w^2 \rangle - \langle w \rangle^2}{N}$$

(Monte-Carlo variance)

↪ for multi-modal target use multi-channel $g(x) = \sum_i \beta_i g_i(x)$ with $\sum_i \beta_i = 1$

$$I = \int_V \mathrm{d}^d x\, f(x) = \sum_i \int_V \mathrm{d}^d x\, \beta_i\, g_i(x) \frac{f(x)}{g(x)} = \sum_i \int_{U_i} \mathrm{d}^d y_i\, \beta_i\, \frac{f(x)}{g(x)}\bigg|_{x \equiv x(y_i)}$$

↪ ME generators use physics knowledge about $f_a f_b |\mathcal{M}|^2$ to construct channels, mapping out prominent features/singularities, but not all features are known

We can embed ML into this for additional optimisation of $y(x)$.
But we need to (i) guarantee phase-space coverage and (ii) cheap evaluation.

Neural Importance Sampling (Normalising Flows)

[Bothmann et al., SciPost Phys. 8 (2020) no.4, 069], [Gao et al., Phys. Rev. D 101 (2020) no.7, 076002]

- Further generic optimisation of random number mapping entering phase-space channels $g(x)$

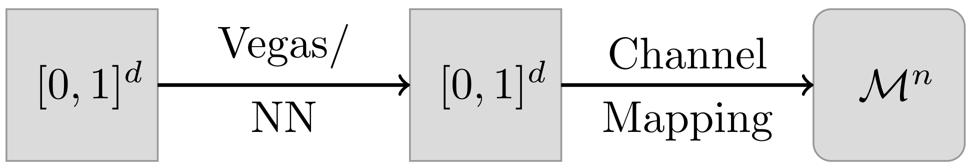
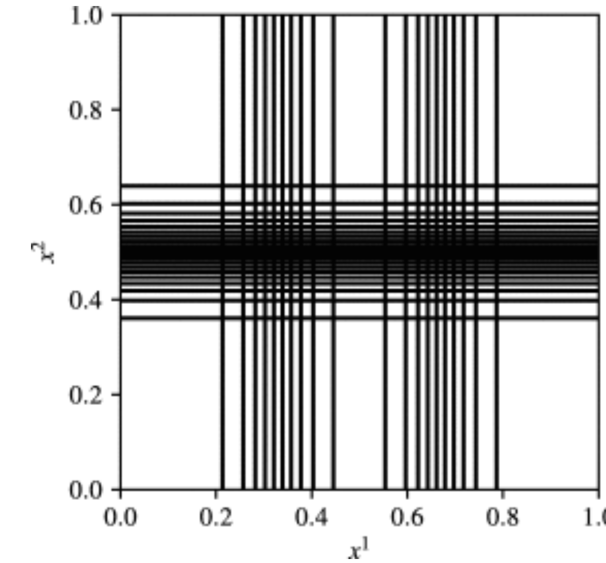


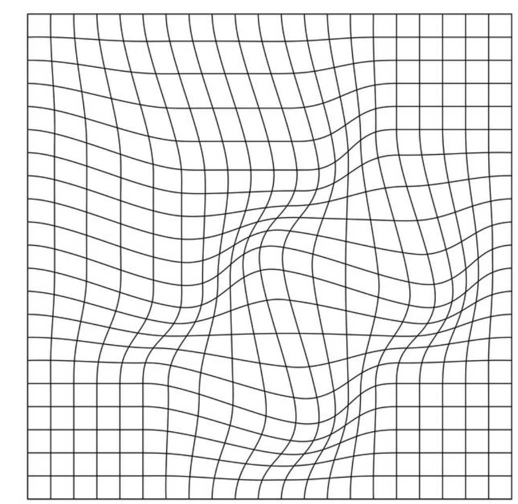
Illustration: Linear VEGAS grid



- Chain of bijective maps, called coupling layers ℓ

$$\left.\begin{array}{lll} x^A & \rightarrow & y^A := x^A \\ x^B & \rightarrow & y^B := C\left(x^B; m(x^A)\right) \end{array}\right\} \; J = \left| \begin{pmatrix} \mathrm{diag}(1) & 0 \\ \frac{\partial C}{\partial m}\frac{\partial m}{\partial x_A} & \frac{\partial C}{\partial x_B} \end{pmatrix} \right| = \left| \frac{\partial C}{\partial x_B} \right|$$

Illustration: Nonlinear NIS grid

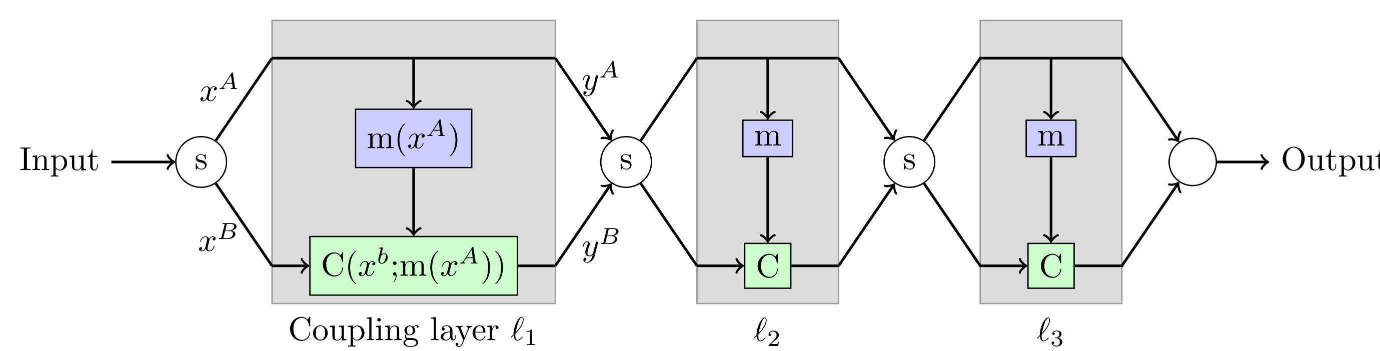


- C invertible+separable yields cheap Jacobian ~ $\mathcal{O}(d)$

- m arbitrary function → use DNN

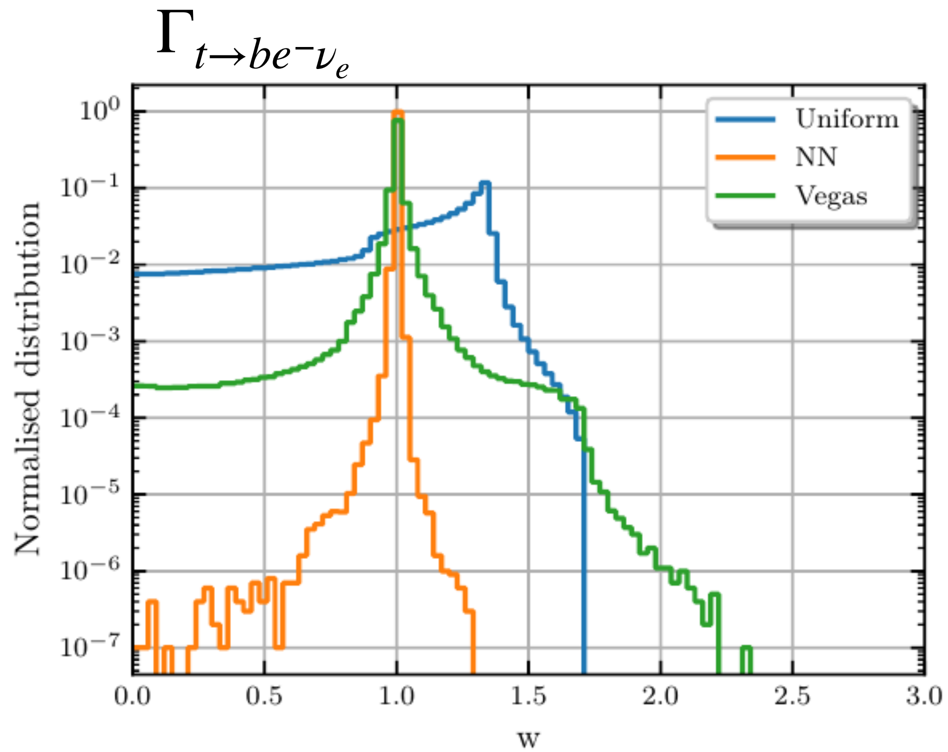
- C piecewise quadratic → „Neural Importance Sampling"
  first applied to ray tracing in 3D scenes
  [Müller et al. arXiv:1808.03856]

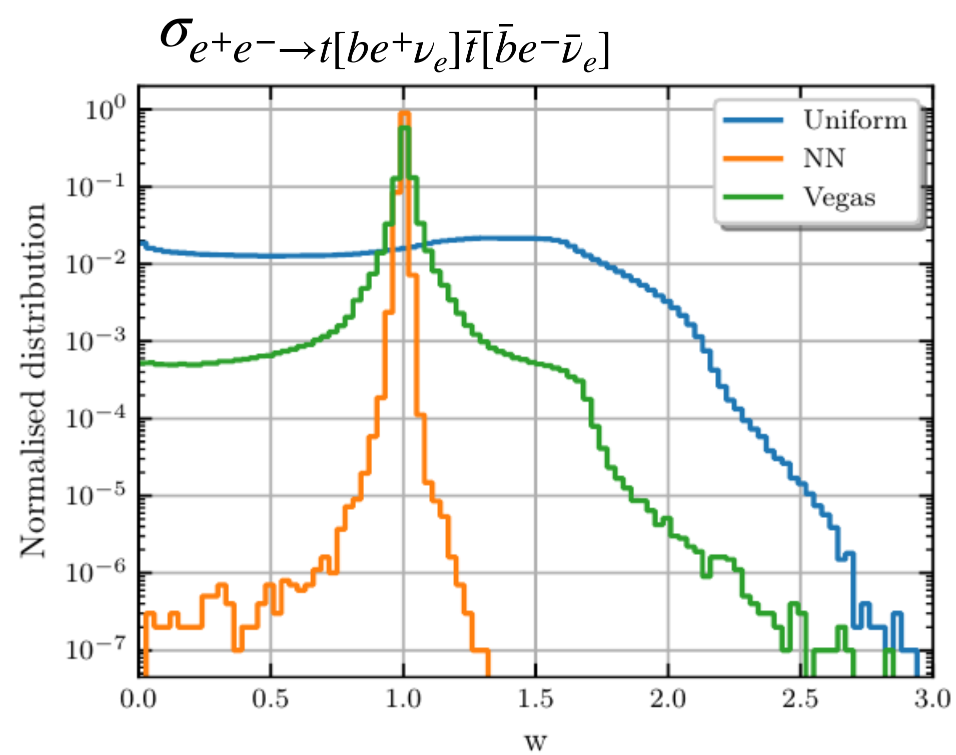
- very expressive+cheap non-linear variable transformations (non-factorisable!)



Neural Importance Sampling – Results

remember: aim for $w = f/g \approx 1$, i.e. peaked distribution of w





	top decays		top-pair production		$gg \to 3g$		$gg \to 4g$	
Sample	ϵ_{uw}	E_N [GeV]	ϵ_{uw}	E_N [fb]	ϵ_{uw}	E_N [fb]	ϵ_{uw}	E_N [fb]
Uniform	59 %	0.1679(2)	35 %	1.5254(8)	3.0 %	24806(55)	2.7 %	9869(20)
Vegas	50 %	0.16782(4)	40 %	1.5251(1)	27.7 %	24813(23)	31.8 %	9868(10)
NN	84 %	0.167865(5)	78 %	1.52531(2)	64.3 %	24847(21)	48.9 %	9859(10)

- Smaller impact for more complicated (multi-channel) processes, similar in [Gao et al., Phys. Rev. D 101 (2020) no.7, 076002]
- GPU evaluation of MEs desirable for efficient training cf. talks by M. Knobbe, R. Wang and A. Valassi
- Alternative to ML-*assisted* phase space sampling: directly learn target distribution using autoregressive flows, GANs, VAEs [Stienen and Verheyen SciPost Phys. **10**, 038 (2021)], [Butter, Plehn and Winterhalder, SciPost Phys. **7** (6), 075 (2019)], [Sipio et al. JHEP **08**, 110 (2019)], [Otten et al. Nature Commun. **12** (1), 2985 (2021)], [Choi and Lim, J. Korean Phys. Soc. **78** (6), 482 (2021)]
    - if no surjectivity guarantee → might miss tails of distributions and get small bias in overall integration result

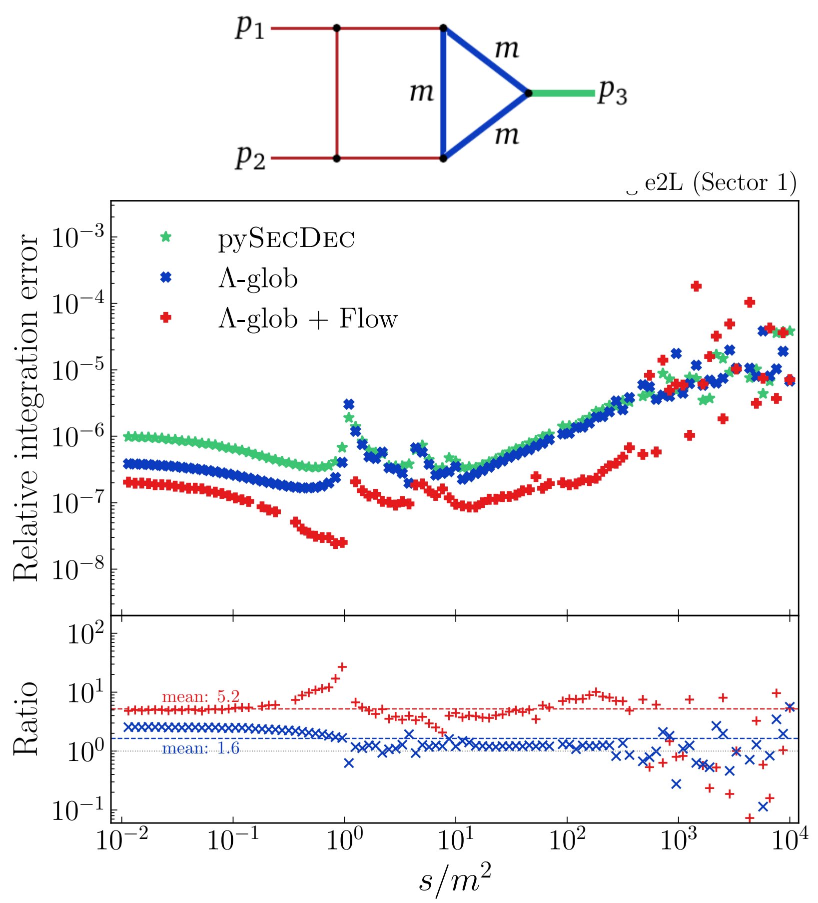
Neural Importance Sampling – Spin-Offs

Aim: Improve numerical evaluation of integral over Feynman parameters in multi-loop diagrams.

Method: Use Neural Importance Sampling to improve sampling

2-loop, one massive

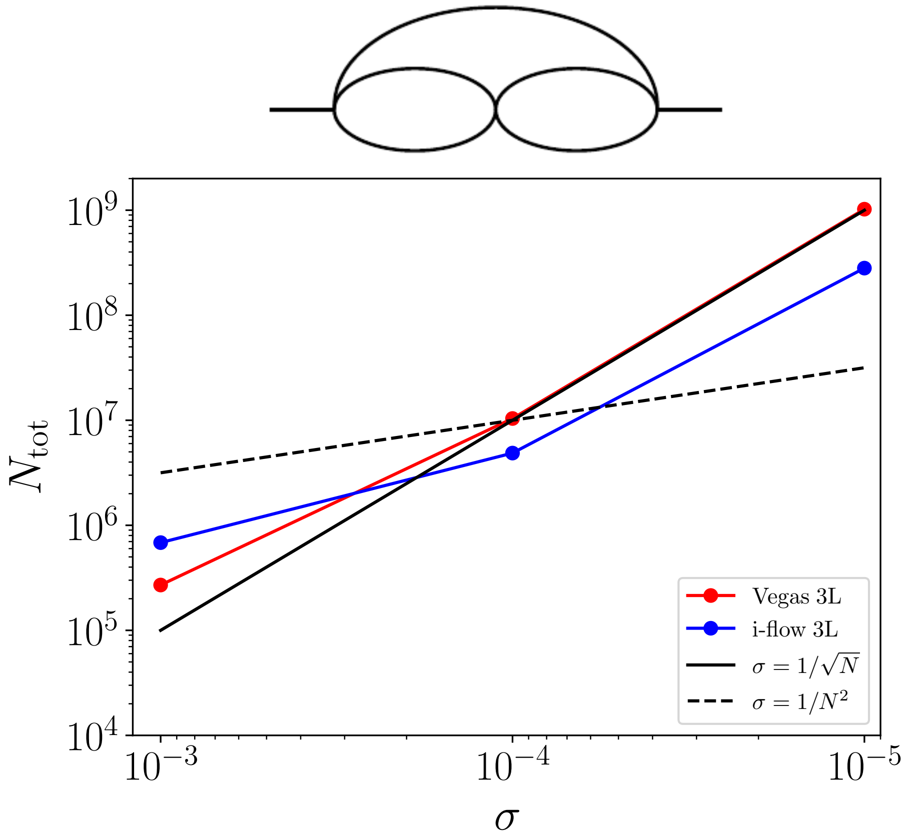
[Winterhalder et al., "Targeting multi-loop integrals with neural networks", arXiv:2112.09145 [hep-ph]]



- 6 dimensions
- Achieve reduction of numerical uncertainties by 2–10x

1,2,3-loop binary dynamics in GR

[Jinno, Kälin, Liu and Rubira: "Machine Learning Post-Minkowskian Integrals", arXiv:2209.01091 [hep-th]]



- 9 dimensions
- training phase included here
- Achieve precision target with less evaluations compared to VEGAS 2–3x

NN integration of multi-loop integrals
[Talk by D. Maître]

Related: application of Normalising Flows in Lattice Field Theory
[Talk by A. Singha]

More on integrating loops
[Plenary Talk by S. Badger, Talks by E. de Doncker, A. Butter]

Exploring Phase Space with Nested Sampling

[Yallup, Janßen, Schumann and Handley, Eur. Phys. J. C 82 (2022), 8]

- Transfer Bayesian inference algorithm to our sampling problem
  - applications in cosmology, statistical thermodynamics, material science
  - wide range of existing tools, e.g. `PolyChord` [Handley, Hobson and Lasenby]
- Consider uniform prior, posteriori matching target distribution
- aim: optimise rejection sampling efficiency ϵ, example $gg \to ng$:

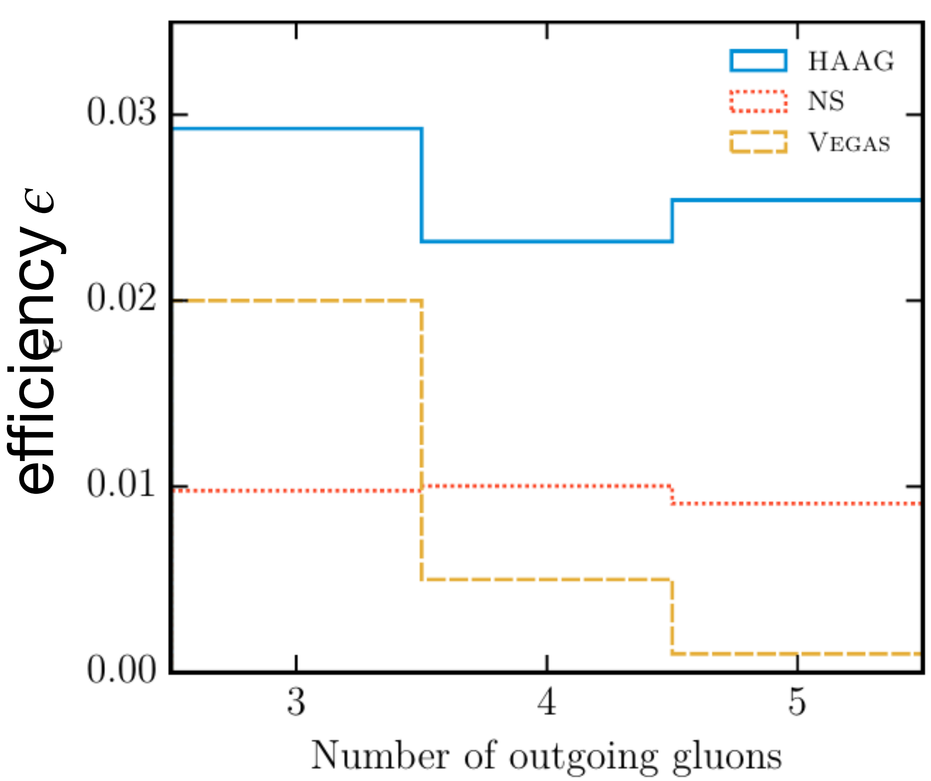
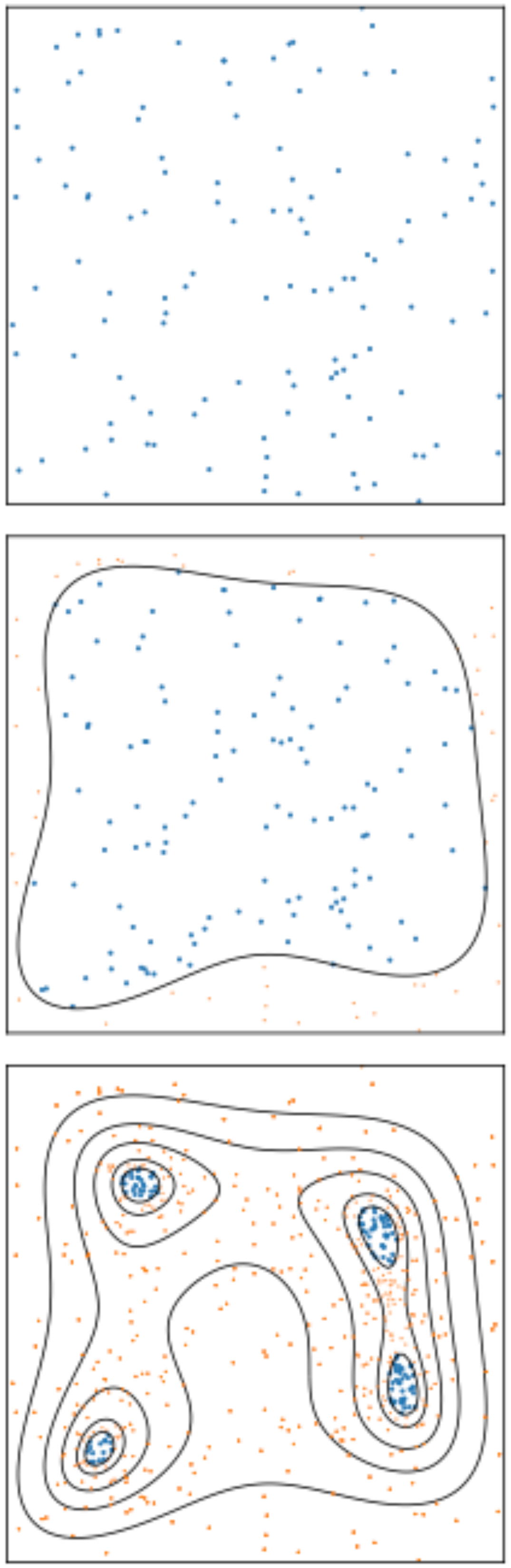
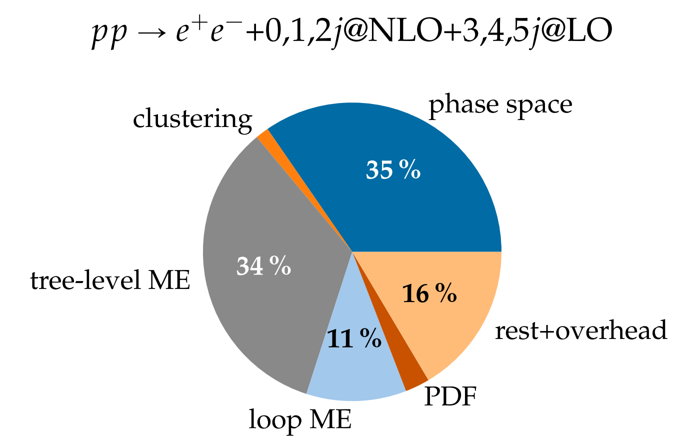


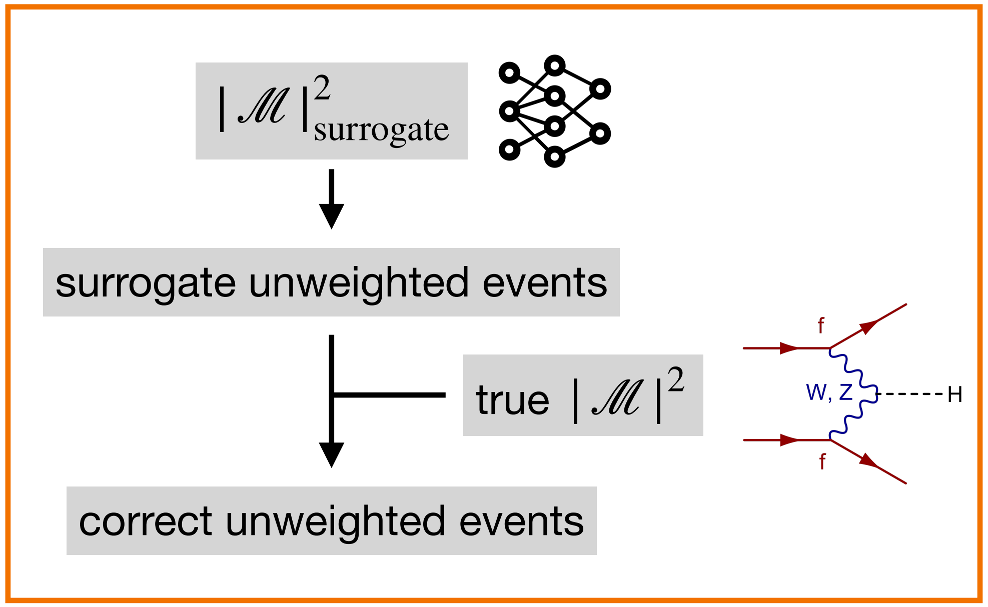
Illustration of Algorithm:

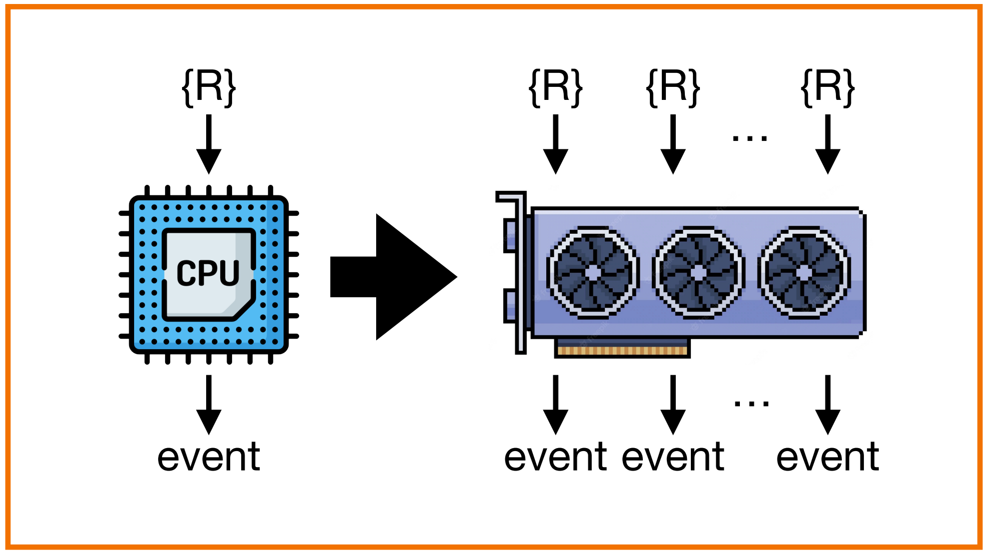


Active research on alternative approaches

- live with low phase-space efficiency, but make $|\mathcal{M}|^2$ evaluation much faster
  - surrogate unweighting
    [Danziger et al. 2109.11964], [Talk by T. Janßen]
    - use integrand emulator for trial events
    - correct accepted events later to exact result by 2nd rejection step
    - emulation must be close to the real thing, but failure to do so only reduces gain factor
    - proof-of-concept: simple DNN gives effective gain factors between 2 and 10
    - provides use case for more sophisticated ME emulators
      [Maître and Truong, JHEP 11 (2021), 066], [Aylett-Bullock, Badger and Moodie, JHEP 08 (2021), 066], [Badger et al., arXiv:2206.14831 [hep-ph]], [Janßen, Maître, Schumann, Siegert and Truong, tbp soon]
  - Accelerated $|\mathcal{M}|^2$: GPU, Vector Engines, …
    - madgraph4gpu
      [Talk by A. Valassi]
    - BlockGen
      [Talks by M. Knobbe, T. Childers]







Conclusions

- The Problem
  - LHC physics programme & computing demands efficient event generation, for signal and background processes
  - Main event generator bottleneck identified:
    - Expensive (N)NLO ME and many-jets LO ME evaluations
    - Combined with inefficient phase-space sampling
- Solutions (active & explorative research!)
  - Development and implementation of novel sampling algorithms
    Neural Importance Sampling, Nested Sampling, …
  - Beyond toy examples, traditional approaches not so easy to beat
  - Many more ideas (see previous slide)
    surrogate NN models, faster (GPU-accelerated) ME, …
- Interdisplinary relevance
  for range of integration/sampling problems
  - Cross talk to many other fields
    ML, Lattice FT, cosmo, industry, …