TECHNISCHE UNIVERSITÄT DRESDEN
INSTITUT FÜR KERN- UND TEILCHENPHYSIK
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

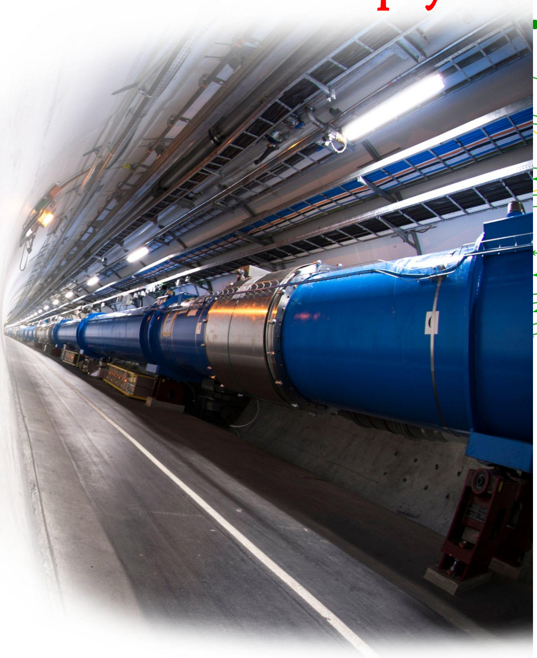
Matrix element surrogates for realistic LHC event generation setups

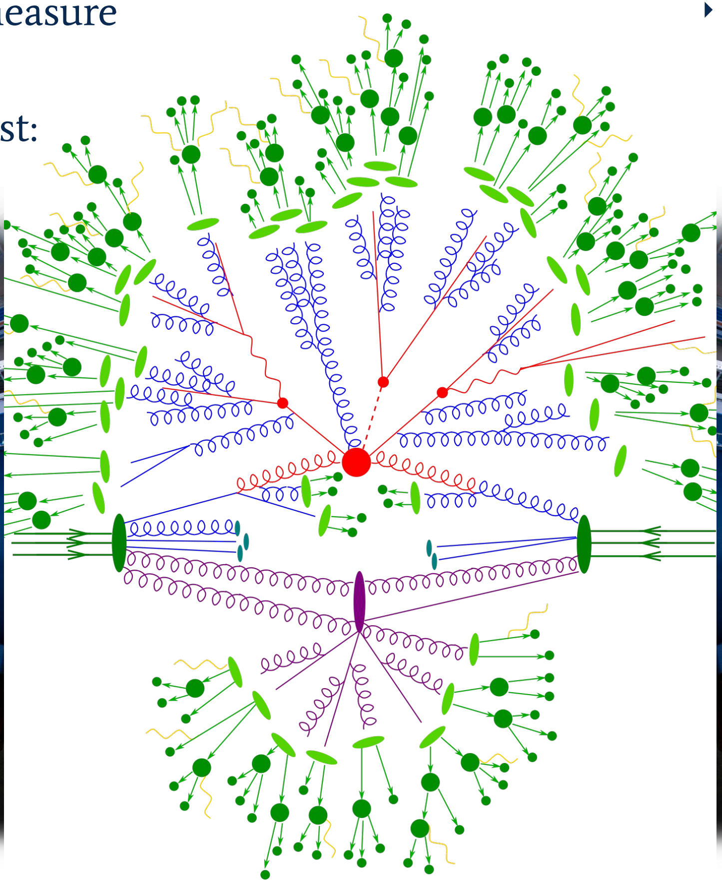
Tim Herrmann, Timo Janßen, Frank Siegert, Steffen Schumann

ErUM-Data KISS Annual Meeting
March 2025

Digital twins: LHC collision events

- In our detectors we measure **stable particles**
- Our theoretical interest: **fundamental physics**

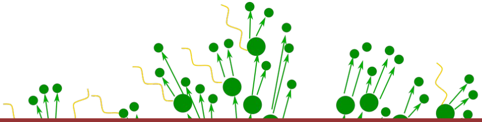




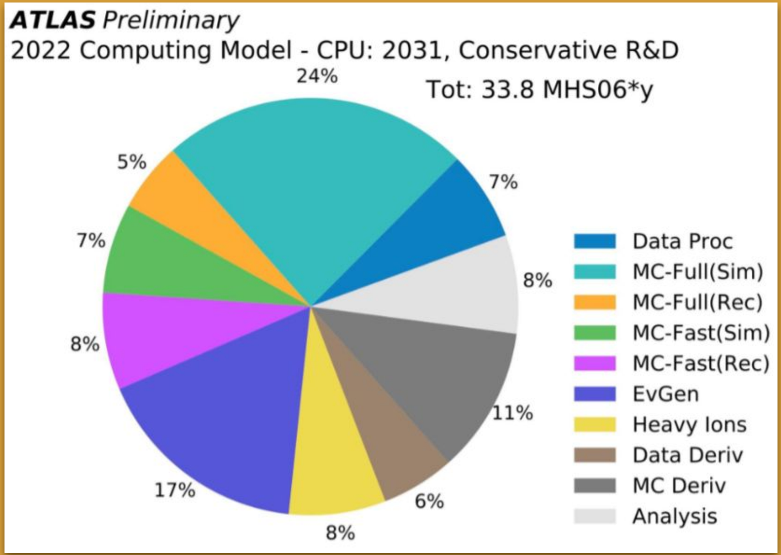
- **Connection** : Monte Carlo simulation of (QCD) dynamics

Is event generation expensive?

- In our detectors we measure **stable particles**
- Our theoretical interest: **fun**



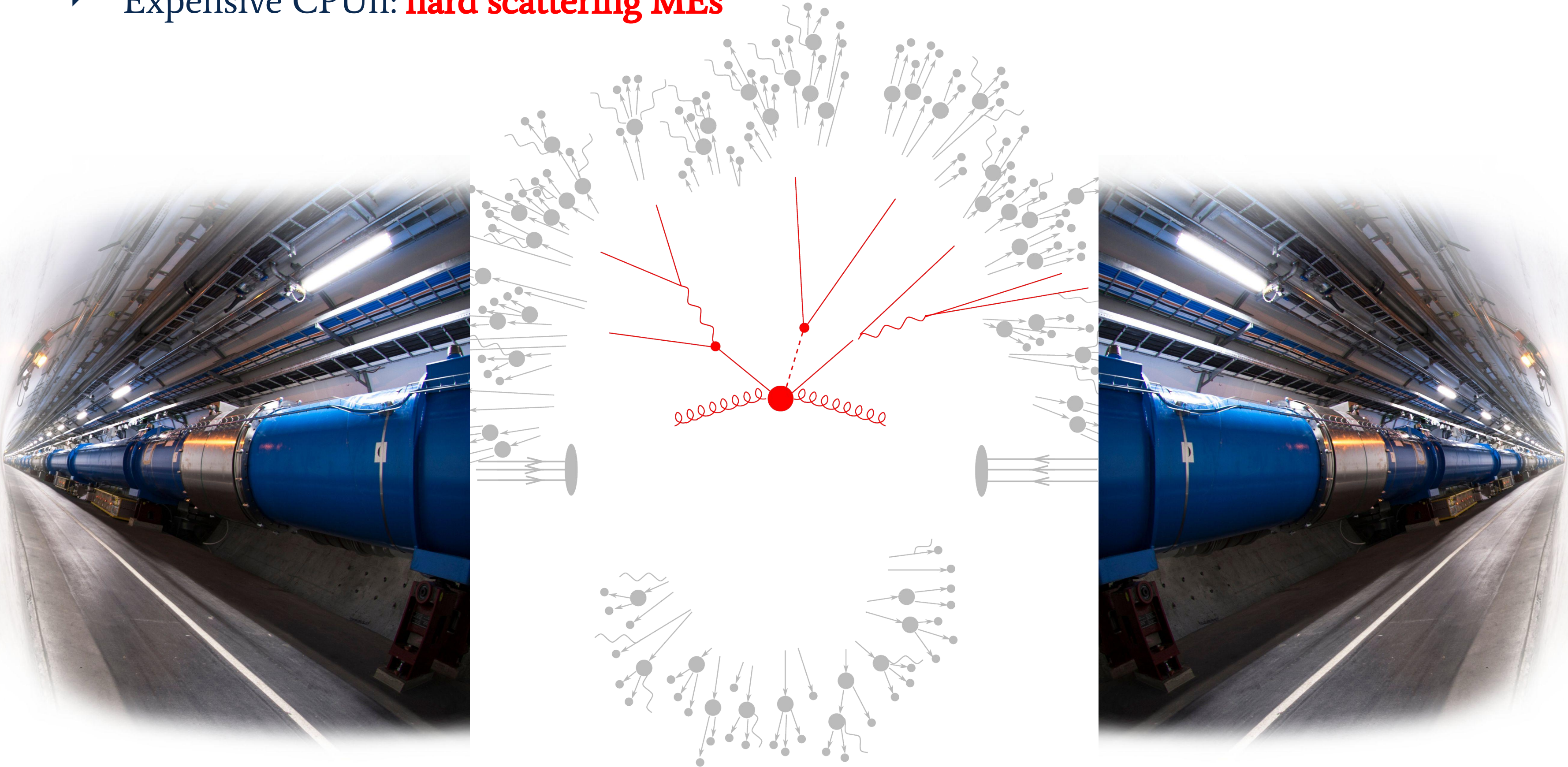
- **Connection** : Monte Carlo simulation of (QCD) dynamics



[ATLAS HL-LHC Computing Roadmap]

1 phys. core
≈ 20 HS23
≈ 20 HS06

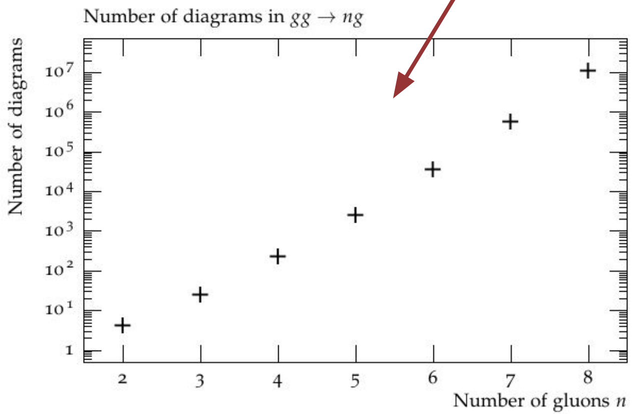
- Expensive CPUh: **hard scattering MEs**



- Expensive CPUh: **hard scattering MEs**

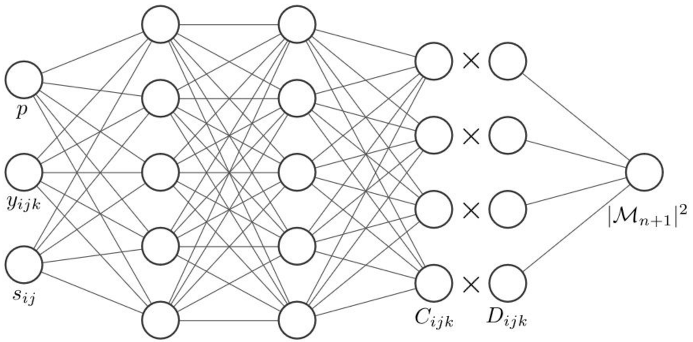
$$\hat{\sigma}_N = \int_{\text{cuts}} \mathrm{d}\hat{\sigma}_N = \int_{\text{cuts}} \left[\prod_{i=1}^{N} \frac{\mathrm{d}^3 q_i}{(2\pi)^3 2E_i} \right] \delta^4 \left(p_1 + p_2 - \sum_i^N q_i \right) |\mathcal{M}(p_1, p_2, q_1, \ldots, q_N)|^2$$

n	#diagrams
2	4
3	25
4	220
5	2,485
6	34,300
7	559,405
8	10,525,900

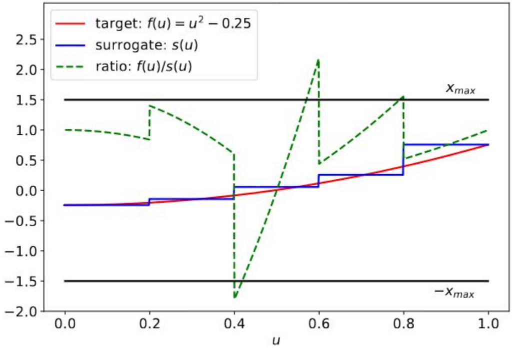


KISS A1 : Can we replace MEs with ML-based surrogates?

Machine Learning:
Design and train **suitable surrogate models** .



Monte Carlo:
Develop **unbiased algorithm** to use surrogates.



→ Monte Carlo unweighting in 2 stages: surrogate, real ME

Can we save substantial computational time on a real life example?

TECHNISCHE UNIVERSITÄT DRESDEN

Realistic LHC ATLAS setup for Z+njets

Run card similar to the **ATLAS Z+≤5jets** setup
⇒ use surrogates to speed it up

Includes **phase space enhancement** → increased statistics in relevant regions:

$$h(\Phi) = \left(\max\left(\frac{\sum_{i\in P} p_{T,i}}{20\,\mathrm{GeV}}, \frac{p_{T,e^+e^-}}{20\,\mathrm{GeV}}\right)\right)^2$$

Multiplicity N	Enhancement $\langle h\rangle_{\text{process group}}$	Cross section $\sigma_{\text{process group}}$ [pb]	$\langle h\rangle \cdot \sigma$ [pb]
0	1.0	1793.2(7)	1793.2(7)
1	5.7	650.0(9)	3689(5)
2	39.3	248.3(6)	9749(24)
3	131	96.4(4)	12 628(57)
4	315	34.8(2)	10 973(73)
5	647	11.2(1)	7225(82)
6	1191	3.5(1)	4112(129)

HL-LHC scenario ($L = 3000\,\mathrm{fb}^{-1}$): required events = $\langle h\rangle \cdot \sigma \cdot L \cdot 10$

```
ME_GENERATORS: Comix
EVENT_GENERATION_MODE: PartiallyUnweighted

MI_HANDLER: None
FRAGMENTATION: None
ME_QED: {ENABLED: false}

YFS:
    MODE: None

# collider setup
BEAMS: 2212
BEAM_ENERGIES: 6500

EW_SCHEME: alphamZ

#Comix color summed vs sampled
COLOR_SCHEME: 1 #1 summed #2 sampled (default)

PSI:
    MAXOPT: 2
    NOPT: 5
    ITMIN: 200000.0
    NPOWER: 0.6

PROCESSES:
- 93 93 -> 11 -11 93{6}:
    Order: {QCD: Any, EW: 2}
    2->3-9:
        Enhance_Function: VAR{max(pow(sqrt(H_T2)-PPerp(p[2])-PPerp(p[3])
          ↪ PPerp2(p[2]+p[3]))/400.0}
        Max_N_Quarks: 4
    Max_Epsilon: 1e-3
    CKKW: 20

SELECTORS:
- [Mass, 11, -11, 66, E_CMS]

ANALYSIS: Rivet
RIVET:
  --analyses:
    - MC_ZJETS
    - MC_ZINC
  --skip-weights: 1
  --ignore-beams: 1
   JETCONTS: 1
```

Reducing number of DNNs required

Many subprocesses with **different parton flavour combinations** within Z + <N> partons

- Use same DNN for identical matrix elements (only PDF weight differs) → “reduced”
- Consider only diagrams with at most 2 quark lines

  ⇒ only 256 reduced subprocesses for Z + ≤ 6 jets

	all quarks ($\leq$ 4 quarks)	
Number of partons N	subprocesses	reduced
0	5	2
1	15	6
2	95	30
3	145	50
4	485 (160)	199 (56)
5	635 (160)	277 (56)
6	1595 (160)	836 (56)
7	1945 (160)	1054 (56)

Reducing number of DNNs required

Many subprocesses with **different parton flavour combinations** within Z + <N> partons

- Use sa … ed”
- Cons

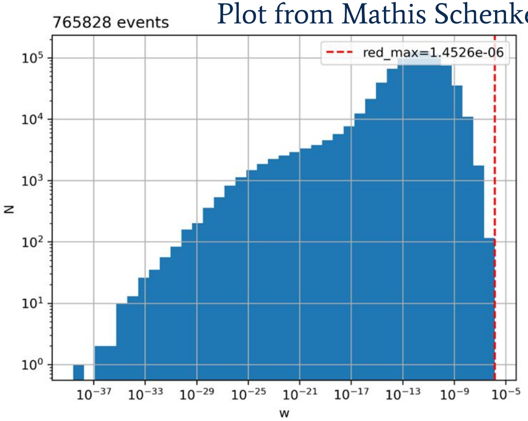
Still further improvement possible:

Process 1	Process 2	Correlation
Highest and lowest correlation for Z + 5-jets		
$\mathrm{d_1\bar{d}_1} \to e^+e^-\mathrm{ggg d_1\bar{d}_1}$	$\mathrm{u_1\bar{d}_2} \to e^+e^-\mathrm{ggg u_1\bar{d}_2}$	0.99
$\mathrm{gu_1} \to e^+e^-\mathrm{gg d_2 u_1\bar{d}_2}$	$\mathrm{u_1\bar{d}_2} \to e^+e^-\mathrm{ggg u_1\bar{d}_2}$	0.001
Highest and lowest correlation for Z + 6-jets		
$\mathrm{gg} \to e^+e^-\mathrm{gg d_1 d_2\bar{d}_1\bar{d}_2}$	$\mathrm{gg} \to e^+e^-\mathrm{gg u_1 d_2\bar{u}_1\bar{d}_2}$	0.98
$\mathrm{g\bar{u}_1} \to e^+e^-\mathrm{ggg d_2\bar{u}_1\bar{d}_2}$	$\mathrm{u_1\bar{u}_2} \to e^+e^-\mathrm{gggg u_1\bar{u}_2}$	0.001
Correlation between top 3 run time contributing subprocesses		
$\mathrm{gu_1} \to e^+e^-\mathrm{ggggg u_1}$	$\mathrm{gd_1} \to e^+e^-\mathrm{ggggg d_1}$	0.95
$\mathrm{gu_1} \to e^+e^-\mathrm{ggggg u_1}$	$\mathrm{g\bar{d}_1} \to e^+e^-\mathrm{ggggg\bar{d}_1}$	0.83
$\mathrm{gd_1} \to e^+e^-\mathrm{ggggg d_1}$	$\mathrm{g\bar{d}_1} \to e^+e^-\mathrm{ggggg\bar{d}_1}$	0.80

7	1945 (160)	1054 (56)

Surrogate unweighting

Plot from Mathis Schenker



- Event weights w (≙ differential cross section) spread 30 orders of magnitude
- Rejection sampling to make the weight distribution uniform → **unweighted events**
- Acceptance probability:

$$\epsilon_{\text{full}} \approx \frac{\langle w \rangle}{w_{\max}} \qquad (\approx 0.0005 \text{ for Z+6jets})$$

→ **99.95** % of exactly determined weights not used, including expensive ME weight

- Idea: Use surrogate s in first unweighting step

$$\epsilon_{\text{1st,surr}} \approx \frac{\langle s \rangle}{w_{\max}}$$

- Calculate exact weight afterwards and correct with overweight in second rejection step

$$x \equiv \frac{w}{s} \qquad \epsilon_{\text{2nd,surr}} \approx \frac{\langle x \rangle_{\text{weight}=s}}{x_{\max}}$$

- Large time savings (gains) expected, unbiased prediction

Effective gain factor

Effective gain factor

How much do we gain overall?

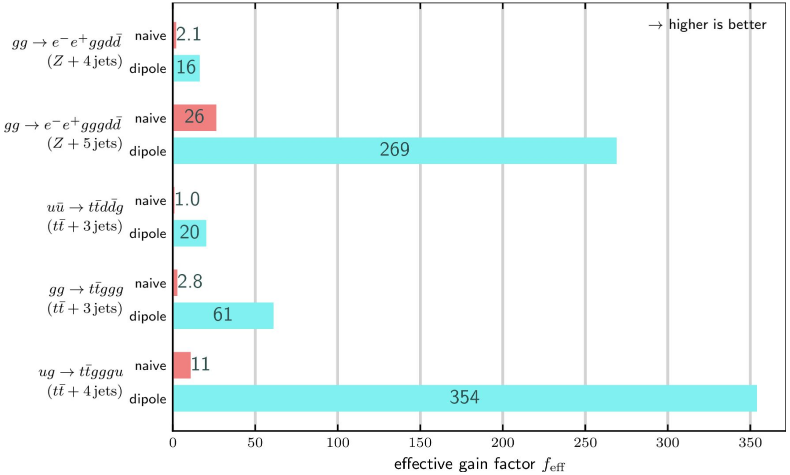
$$f_{\text{eff}} := \frac{T_{\text{standard}}}{T_{\text{surrogate}}}$$

$$f_{\text{eff}} = \frac{\frac{1}{\epsilon_{\text{full}}} \cdot \left(\langle t_{\text{ME}} \rangle + \langle t_{\text{PS}} \rangle\right)}{\frac{1}{\epsilon_{\text{1st,surr}}\epsilon_{\text{2nd,surr}}} \left(\langle t_{\text{surr}} \rangle + \langle t_{\text{PS}} \rangle\right) + \frac{1}{\epsilon_{\text{2nd,surr}}} \cdot \langle t_{\text{ME}} \rangle}$$

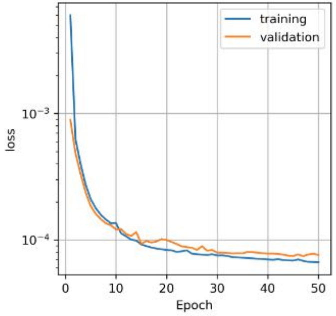
$$= \frac{1}{\frac{\langle t_{\text{surr}} \rangle + \langle t_{\text{PS}} \rangle}{\langle t_{\text{ME}} \rangle + \langle t_{\text{PS}} \rangle} \cdot \frac{\epsilon_{\text{full}}}{\epsilon_{\text{1st,surr}}\epsilon_{\text{2nd,surr}}} + \frac{\langle t_{\text{ME}} \rangle}{\langle t_{\text{ME}} \rangle + \langle t_{\text{PS}} \rangle} \cdot \frac{\epsilon_{\text{full}}}{\epsilon_{\text{2nd,surr}}}}$$

We need:

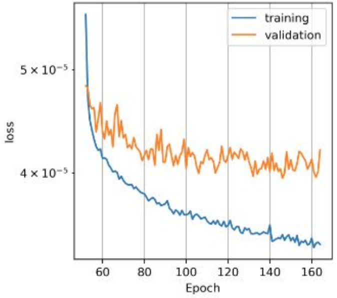
- low original unw. eff.: $\varepsilon_{\text{full}} \ll 1$
- fast surrogate: $\langle t_{\text{surr}} \rangle \ll \langle t_{\text{ME}} \rangle$
- accurate surrogate: $\varepsilon_{\text{2nd,surr}} \cong 1$



Gain factor as loss function



(a) MSE loss function



(b) Gain loss function

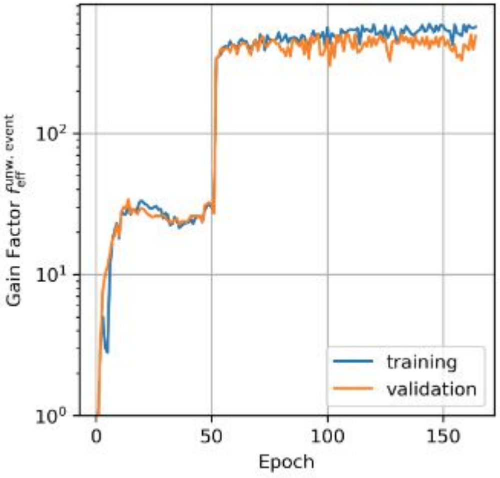
Train model directly on final figure of merit:
gain factor

Parameter	Value
Hidden layers	4
Nodes in hidden layers	128
Activation function	swish [48, 49]
Weight initialiser	He uniform [50]
Loss function	1) MSE on arcsinh 2) gain factor
Loss change epoch	50
Batch size	1024
Optimiser	ADAM [51]
Initial learning rate	10^{-3}
Callbacks	EarlyStopping

Hyperparameters optimised with OPTIMA

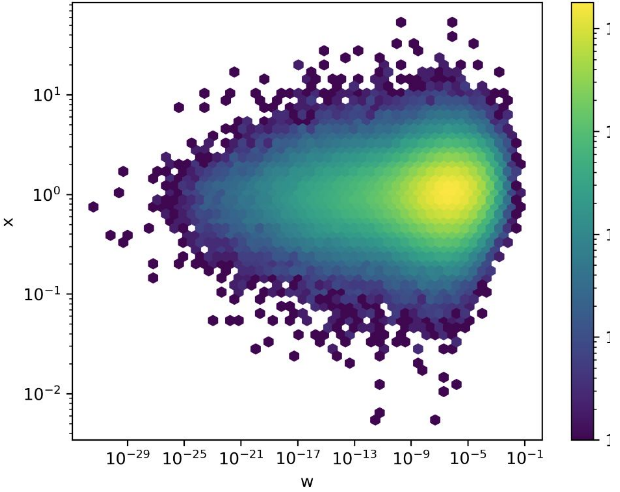
Gain factor as loss function

For gu→e⁺e⁻gggggu:

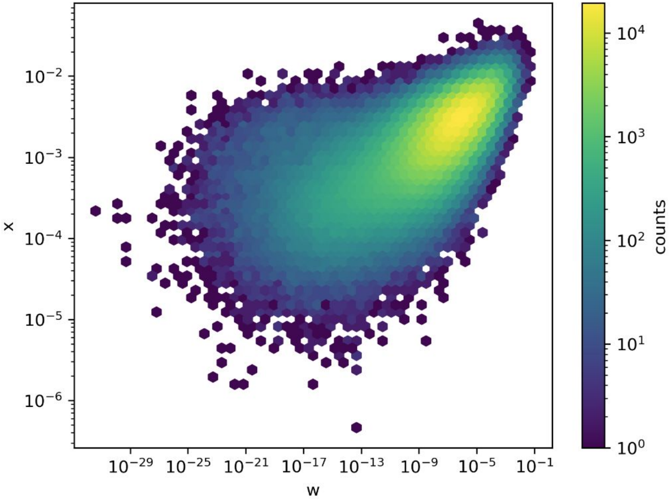


(c) Gain factor

Epoch 50: before loss change



Epoch 105: best model



Summed vs Sampled vs Surrogate

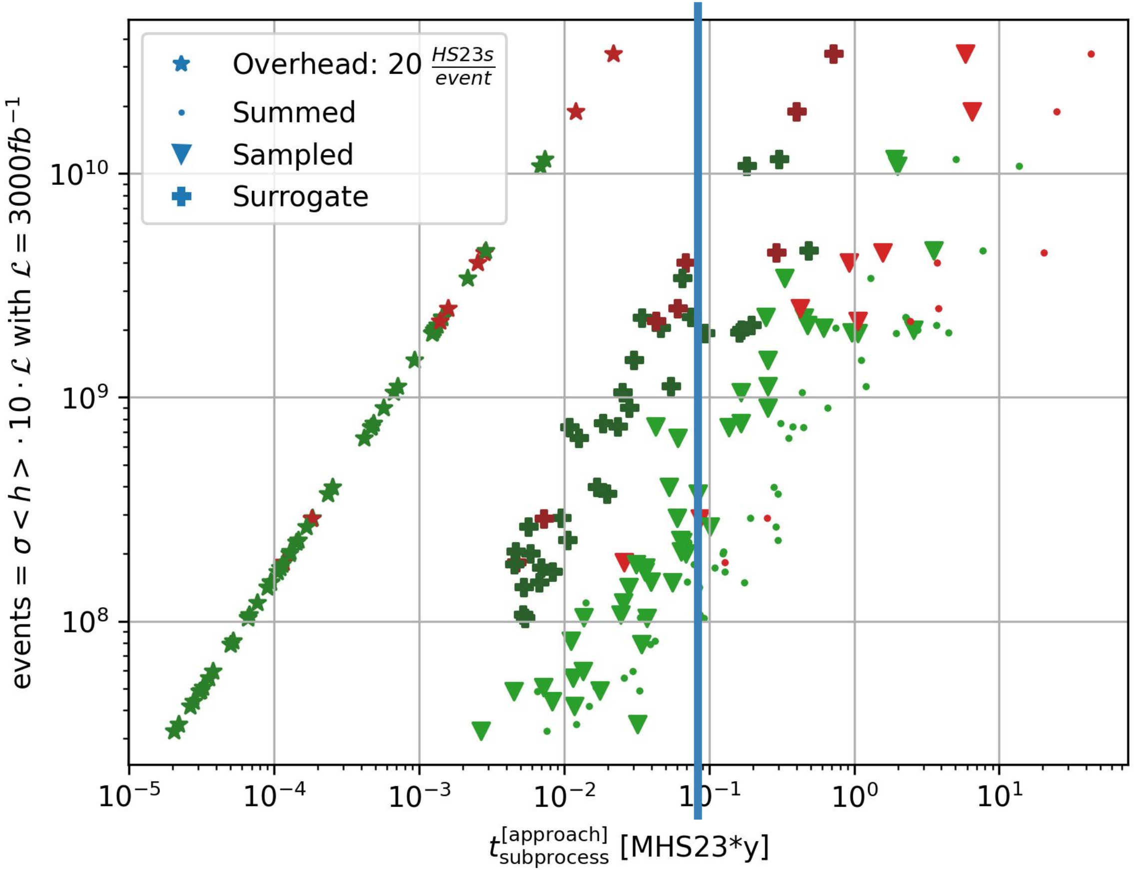
(table needs update)

Configuration	$\langle t_{\text{init}} \rangle$ [ms]	$\langle t_{\text{PS}} \rangle$ [ms]	$\langle t_{\text{ME}} \rangle$ [ms]	PS points	gen points	unw. effi	$t_{\text{unw}}^{\text{normal}}$ [s]	$t_{\text{unw}}^{\text{ME+PS surr}}$ [s]
Z7j color sampled	2.97	16.9	13.3	90 496	6	0.000 066 3	500	1013
Z6j color summed	1.17	4.98	288	33 757	23	0.000 681	432	32.5
Z6j color sampled	1.29	5.34	3.69	83 984	20	0.000 238	43.3	97.0
Z5j color summed	0.475	1.45	14.6	178 043	66	0.000 371	44.6	20.1
Z5j color sampled	0.541	1.44	1.22	426 802	68	0.000 159	20.1	47.7
Z4j color summed	0.318	0.454	1.55	137 146	89	0.000 649	3.58	5.52
Z4j color sampled	0.370	0.456	0.499	455 757	81	0.000 178	7.46	21.1
Z3j color summed	0.283	0.159	0.363	46 604	112	0.002 40	0.335	1.03
Z3j color sampled	0.335	0.169	0.240	125 756	108	0.000 859	0.866	3.12
Z2j color summed	0.227	0.0592	0.160	5700	94	0.0165	0.0271	0.119
Z2j color sampled	0.276	0.0679	0.144	12 101	94	0.007 77	0.0628	0.277
Z1j color summed	0.108	0.0251	0.0889	269	34	0.126	0.001 76	0.0117
Z1j color sampled	0.102	0.0301	0.0769	347	25	0.0720	0.002 90	0.0202
Z0j color summed	0.0742	0.0144	0.0459	57	13	0.228	0.000 590	0.005 83
Z0j color sampled	0.0656	0.0157	0.0429	56	19	0.339	0.000 366	0.003 89

⇒ use surrogate starting from summed: train 60/256 most time saving subprocesses

Comparison of computational time on test data

Summed reduced subprocess trained right of this line



Z+6jets: $gu_1 \rightarrow e^+e^- gggggu_1$

Z+6jets: $gd_1 \rightarrow e^+e^- gggggd_1$

Z+6jets: $u_1u_1 \rightarrow e^+e^- ggggu_1u_1$

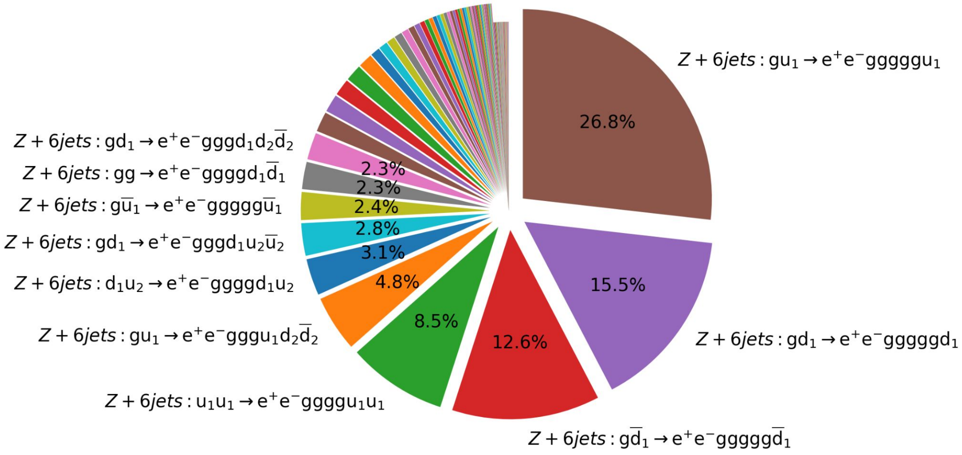
Z+6jets: $g\bar{d}_1 \rightarrow e^+e^- ggggg\bar{d}_1$

Z+6jets: $gu_1 \rightarrow e^+e^- gggu_1d_2\bar{d}_2$

Performance estimate of $pp \to e^+e^- + \leq 6j$ on test dataset

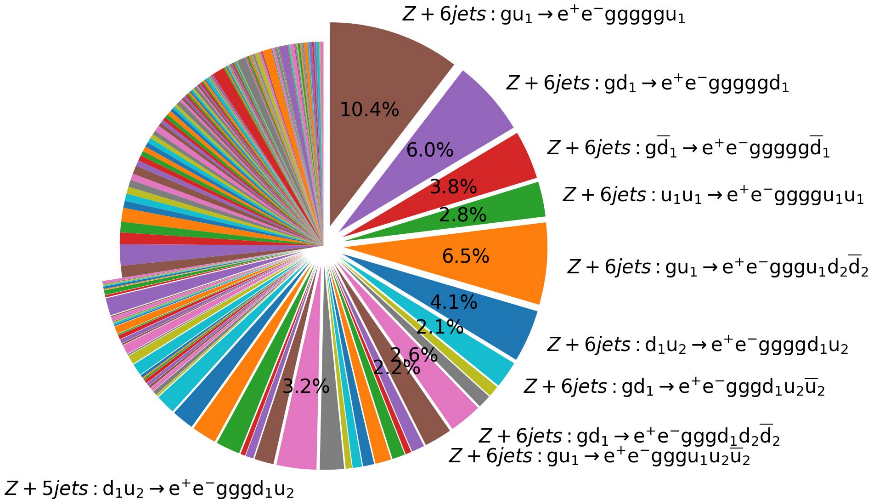
Without surrogates:

$\approxeq$ **3380** HS23s/event



With surrogates:

$\approxeq$ **160** HS23s/event



Colour sampled: $\approxeq$ **1084** HS23s/event

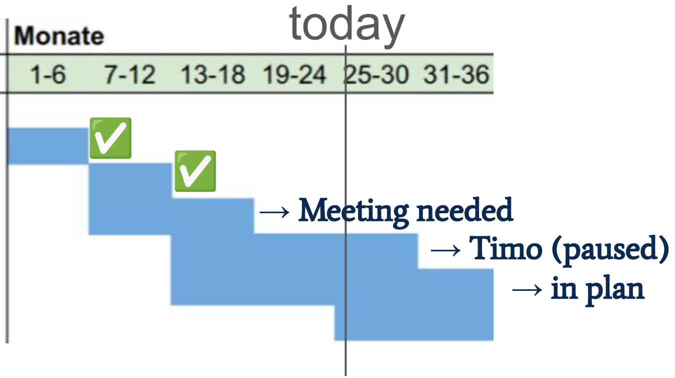
Comparison of computational time on test data

Process	Events [M]	Summed [MHS23*y]	Sampled [MHS23*y]	Surrogates [MHS23*y]	**Ratio**
Z + $\leq$0 jets	53 797	0.03	0.03	-	
Z + $\leq$1 jets	164 479	0.10	0.10	-	
Z + $\leq$2 jets	456 960	0.29	0.30	-	
Z + $\leq$3 jets	835 797	0.58	0.74	-	
Z + $\leq$4 jets	1 164 974	1.16	2.25	-	
Z + $\leq$5 jets	1 381 719	**8.93**	18.08	**3.10**	**2.8**
Z + $\leq$6 jets	1 505 067	161.32	**51.68**	**7.57**	**6.7**

Previous work and KISS plan (03/23 - 03/26)

- Proof of principle:
  - MA theses (Johannes Krause, TU Dresden 2015 & Katharina Danziger, TU Dresden 2020)
  - Danziger, Janßen, Schumann, Siegert [2109.11964]
- First generalisation and more advanced NN:
  - Janßen, Maitre, Schumann, Siegert, Truong [2301.13562]
- Milestones for KISS:

Arbeitspaket		Meilenstein	Beteiligte AGs
Teilprojekt A:			
A1: Erzeugung von Teilchenereignissen	M1	Schnittstelle Sherpa/Surrogate	TUD/UGOE
	M2	Benchmarking	TUD/UGOE
	M3	Dynamisches Training	UHD/TUD
	M4	NLO-Genauigkeit	UGOE/TUD
	M5	Optimierung zu Multi-Leg	TUD/UHD/UGOE
	M6	Invertierung	UHD/UGOE



- BA theses:
  - Mathis: Effective gain factor as loss (11/24 - 2/25, M5)
  - Leonard: Network structure: L-GATr (3/25 - 6/25, M5)

Appendix

Network details

Data taken from integration step:

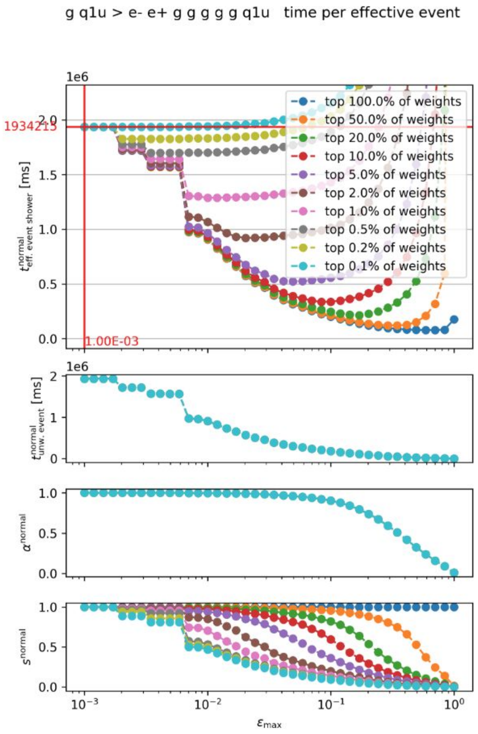
- Momenta and derived dipole Terms

$$|\mathcal{M}_{n+1}|^2 \simeq \sum_{\{ijk\}} C_{ijk} D_{ijk}$$

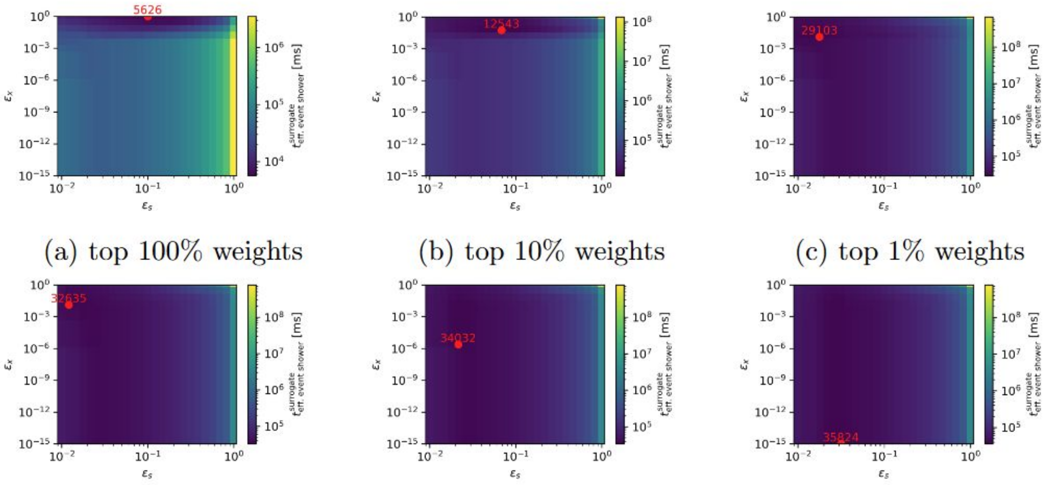
Data split into:

- 800k train
- 200k val
- 1 000k test

Optimal epsilon max



(a) $t_{\text{rest}} = 1\,\text{s}$



(a) top 100% weights (b) top 10% weights (c) top 1% weights

(d) top 0.1% weights (e) top 0.01% weights (f) top weight

Modularity: Sherpa Interface

Interface available for external surrogate providers

→ available in both Sherpa2 and Sherpa3

Output from Sherpa for training:

- **momenta** of initial and final state ME partons and **ME (+PhS) weight**

Training outside of Sherpa (your NN could go here):

- Train Onnx model

Input to Sherpa:

- Onnx model which calculates **ME weight surrogate** from **momenta**

Integration

~5M events: **momenta** and **ME weight**

Train

Onnx NN model

Event generation

1. Pick PS point
2. Keep event according to ME weight

momenta

NN

ME weight surrogate

The deliverables that are a part of the conservative and aggressive scenario are described in detail in the subsequent sections. While they all contribute to achieve the full physics potential of the ATLAS detector, the deliverables with the largest impact on the resource estimates in the aggressive scenario are: it is assumed that the fraction of full simulation used by the collaboration will be reduced from 33% to 10%; the event generation and reconstruction times will be reduced by 20% each; a smaller fraction of the AODs will be

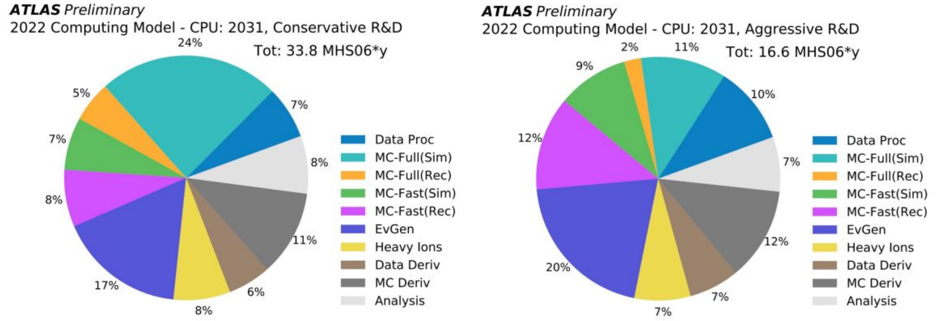


Figure 3: projection for Run 4 of the breakdown of compute (upper row), disk (middle row) and tape (lower row) usage, for the conservative (left) and aggressive (right) R&D scenarios. The expected totals in million HS06*years and exabytes are also displayed.

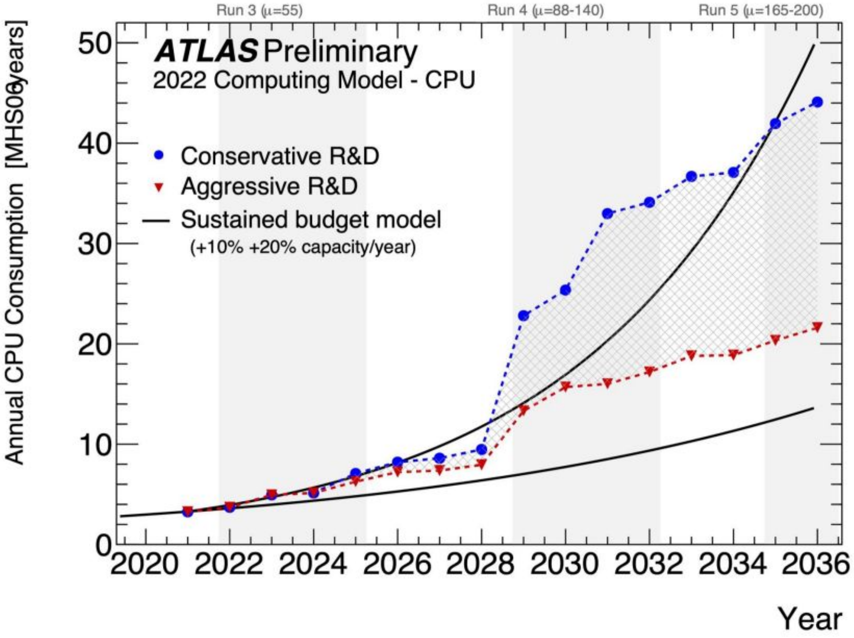


Figure 1: projected evolution of compute usage from 2020 until 2036, under the conservative (blue) and aggressive (red) R&D scenarios. The grey hatched shading between the red and blue lines illustrates the range of resources consumption if the aggressive scenario is only partially achieved. The black lines indicate the impact of sustained year-on-year budget increases, and improvements in new hardware, that together amount to a capacity increase of 10% (lower line) and 20% (upper line). The vertical shaded bands indicate periods during which ATLAS will be taking data.

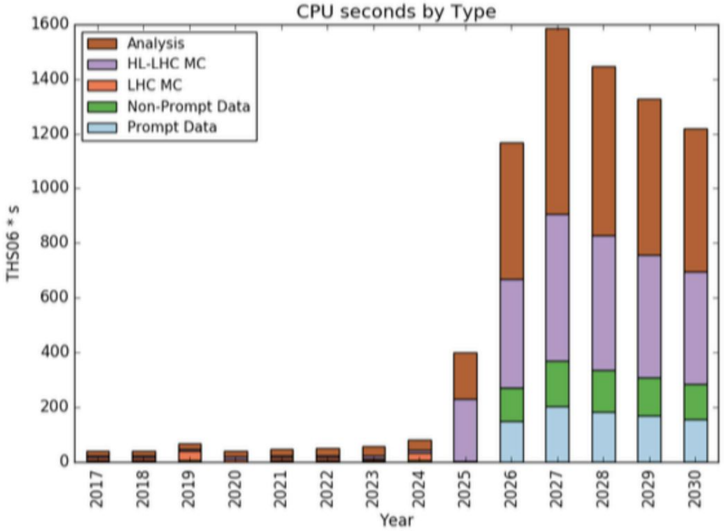


Figure 3: Current estimates for CMS CPU and disk needs through the initial years of the HL-LHC program.

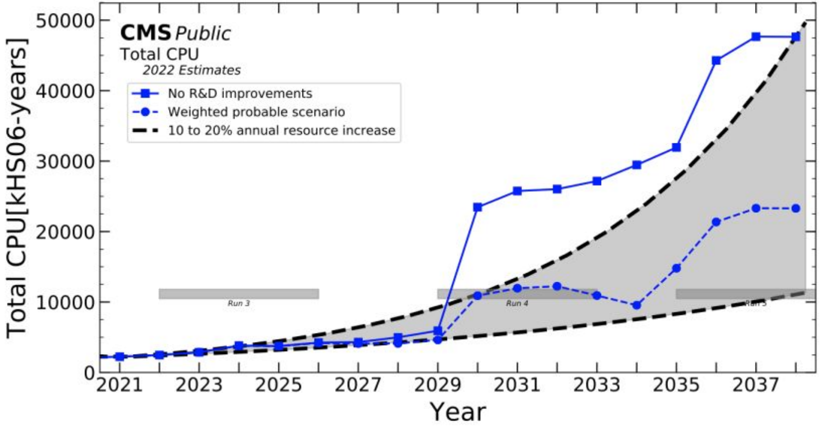


Figure 1. Estimated computing resource needs for CMS [10]. Shown are the modeled annual projections of total CPU and disk needs for CMS through Run 4. The estimated needs for each computing model scenario are shown by the blue lines. The gray band shows the projected resource availability for an example scenario that extrapolates the 2021 CMS pledged resources using an annual increase in available resources of between 10% and 20%. This assumes current WLCG cost projections [12] and a warranty + 3 years replacement cycle of hardware.

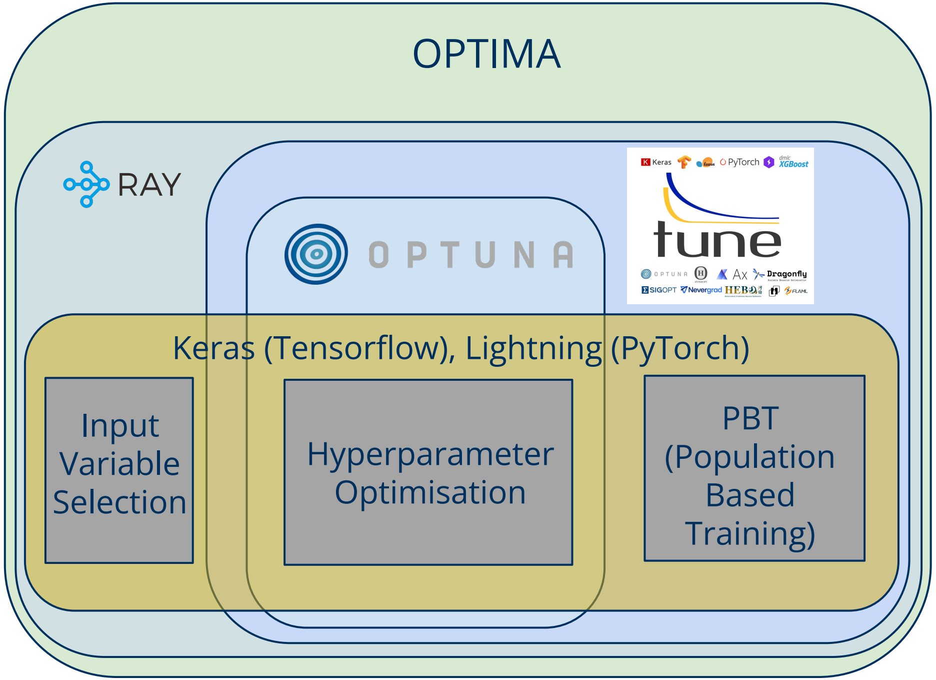
NN optimisation: OPTIMA

Original prototype: fixed hyperparameters
Problem: different scattering processes

→ very different optimal hyperparameters?
→ need flexible and automatic optimisation

Solution: OPTIMA

- Hyperparameter optimization using:
  - Bayesian optimisation and/or
  - Population based training (PBT)
- Erik Bachmann: master thesis and
  - on pypi: optima_ML

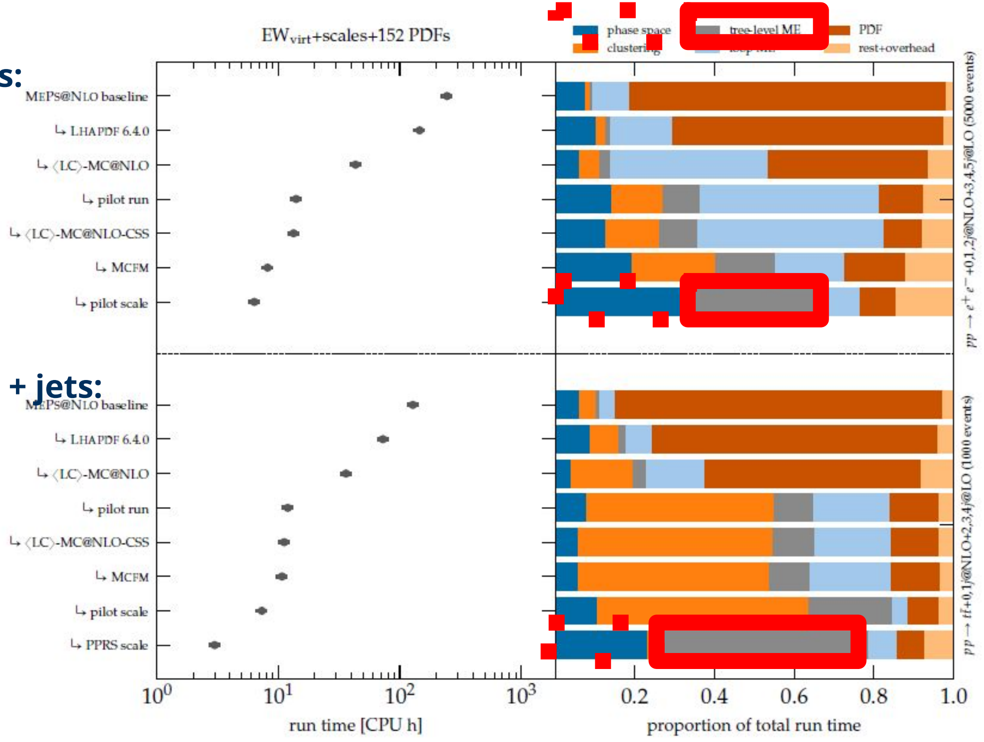


Need for Faster Event Generation

arXiv:2209.00843 **[hep-ph]**
(reproduced in PhD thesis of Timo Janßen)

Z+jets:

ttbar + jets:



ATLAS Software and Computing HL-LHC Roadmap - v 2.1 (cern.ch)

1 HS06 ≅ 10 core HEP-SPEC06 (HS06) (hepix.org)