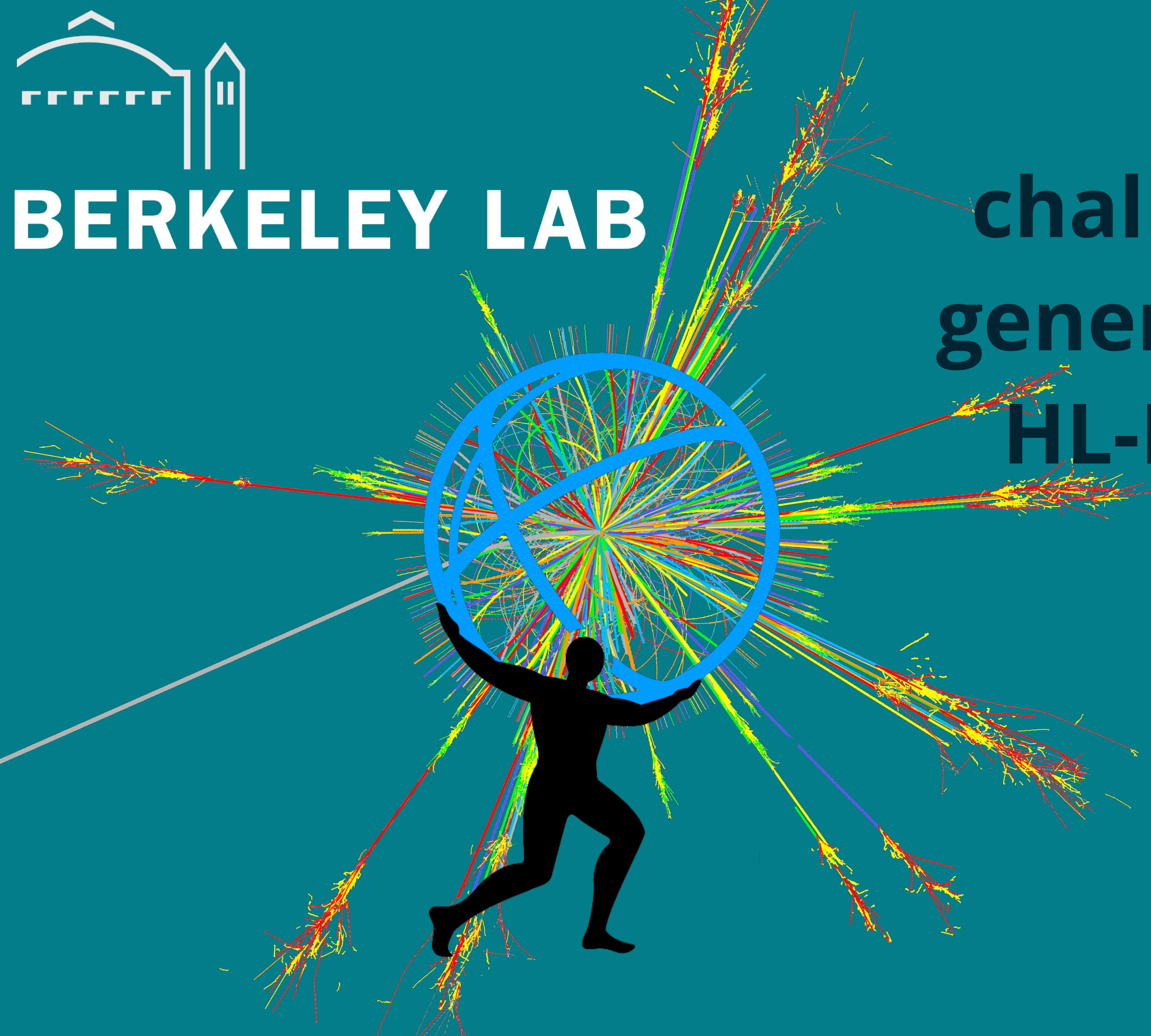# Computational challenges for event generation in view of HL-LHC and beyond
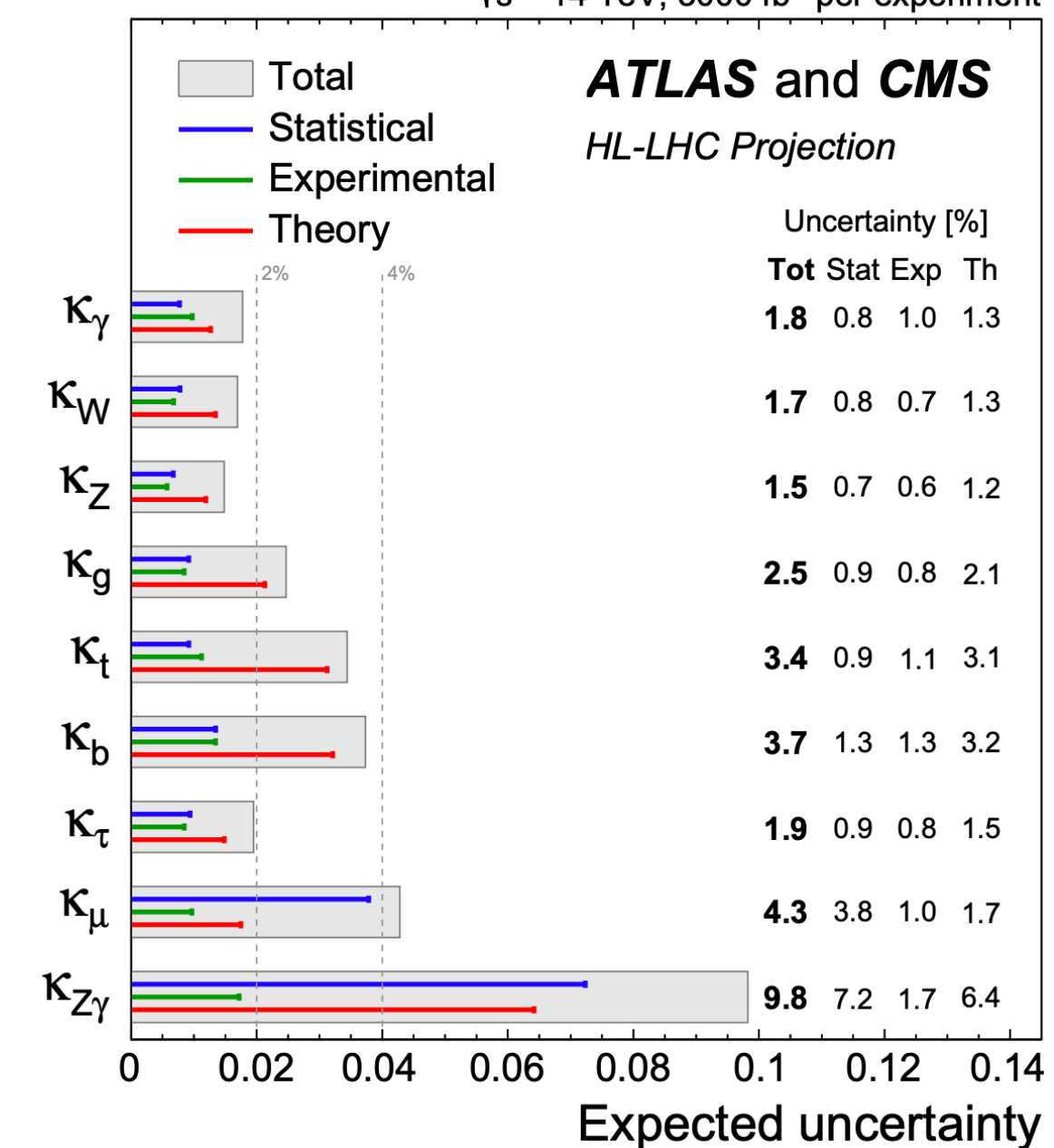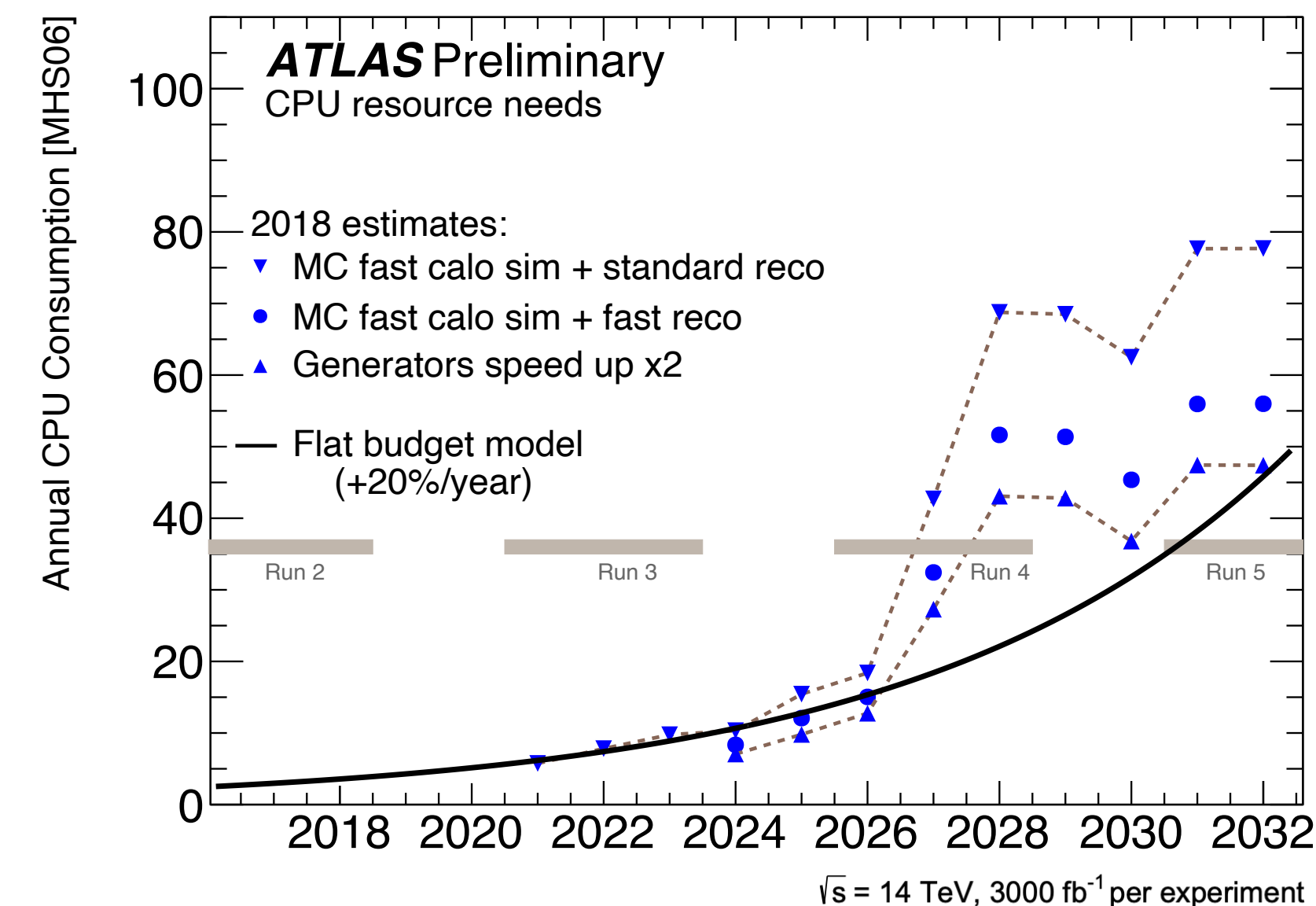
Snowmass CPM

7/10/2020

Josh McFayden (LBNL)

ATLAS EXPERIMENT

Event generation importance

▸ Event generation has already taken up to ~15% of the LHC experiments' computing share.

▸ **Assuming no development in software this fraction will increase for HL-LHC**:

   ▸ Simulation and reconstruction are becoming faster.

   ▸ Precision and therefore complexity of calculations is increasing: (***NLO-merging***, ***EWK corrections***, ***NNLO***).

▸ There are already issues with available resources:

   ▸ Some analyses have **significant uncertainties** arising through limited MC statistics and/or precision of physics modelling.

   ▸ **Projections for HL-LHC** physics output can be quite severely limited by theory uncertainties (even when halved!).

ATLAS Preliminary
CPU resource needs

2018 estimates:
- ▼ MC fast calo sim + standard reco
- ● MC fast calo sim + fast reco
- ▲ Generators speed up x2
- — Flat budget model (+20%/year)

Annual CPU Consumption [MHS06]

Run 2   Run 3   Run 4   Run 5

$\sqrt{s}$ = 14 TeV, 3000 fb^{-1} per experiment

ATLAS and **CMS**
HL-LHC Projection

- Total
- Statistical
- Experimental
- Theory

	Uncertainty [%]			
	Tot	Stat	Exp	Th
κ_γ	1.8	0.8	1.0	1.3
κ_W	1.7	0.8	0.7	1.3
κ_Z	1.5	0.7	0.6	1.2
κ_g	2.5	0.9	0.8	2.1
κ_t	3.4	0.9	1.1	3.1
κ_b	3.7	1.3	1.3	3.2
κ_τ	1.9	0.9	0.8	1.5
κ_μ	4.3	3.8	1.0	1.7
$\kappa_{Z\gamma}$	9.8	7.2	1.7	6.4

Expected uncertainty

Josh McFayden ǀ Snowmass ǀ 7/10/2020

Development | Short-term view

▶ Following topics being discussed in HEP Software Foundation Generators WG:

   ▶ *"Challenges in MC event generator software for HL-LHC"* [2004.13687]

▶ Performance benchmarking

   ▶ Comparison between experiments.

   ▶ Comparison between generators.

      ▶ Discovered factor ~2 saving in compute time for scale choice

      ▶ Generator authors have also been profiling code bases.

▶ Efficiency savings

   ▶ Weight-based uncertainties now ubiquitous - big saving!

   ▶ Negative weights (and other dilution of effective MC statistics) still cause significant issues.
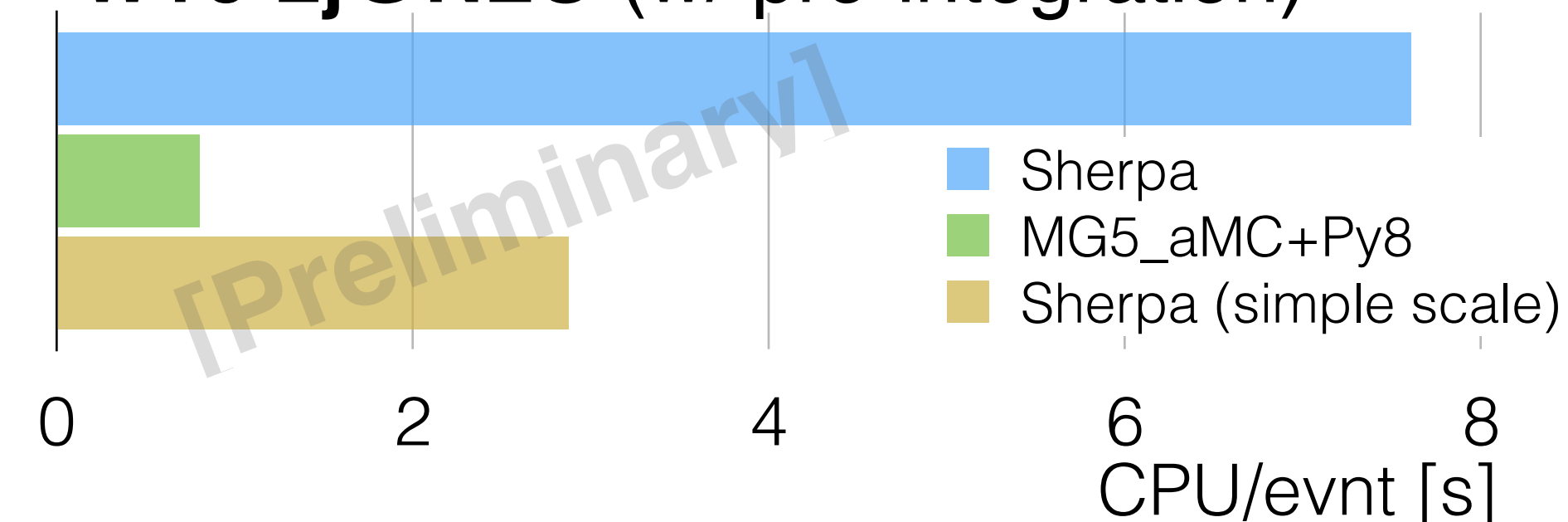
      ▶ Several recent developments:

         ▶ MC@NLO-Δ [2002.12716]

         ▶ NN-based resampling [2007.11586]

   ▶ Positive resampling [2005.09375]

W+0-2j@NLO (w/ pre-integration)

Sherpa
MG5_aMC+Py8
Sherpa (simple scale)

[Preliminary]

0   2   4   6   8
CPU/evnt [s]

	MC@NLO	MC@NLO-Δ
	441	Δ-441
$pp \to e^+e^-$	3.2% (1.1)	2.0% (1.1)
$pp \to e^+\nu_e$	3.4% (1.2)	2.3% (1.1)
$pp \to H$	3.4% (1.2)	0.5% (1.0)
$pp \to H b\bar{b}$	38.0% (17)	31.3% (7.2)
$pp \to W^+ j$	15.7% (2.1)	7.4% (1.4)
$pp \to W^+ t\bar{t}$	15.1% (2.1)	11.5% (1.7)
$pp \to t\bar{t}$	19.6% (2.7)	7.7% (1.4)

Fraction of -ve weights (Statistical dilution factor)

3

Development | Longer-term view

▸ Need to cope with heterogeneous future HPC architectures

  ▸ Existing studies to scale generator codes to HPCs (ANL, SLAC).

    ▸ Not yet well integrated into experiments…

  ▸ Migration to GPUs and vectorised code using dataparallel paradigms.

    ▸ Ongoing work in MG5_aMC (MG5_aMC authors, CERN, ANL).

▸ Optimisation of most time-consuming stages of event generation (mainly using Machine Learning techniques)

  ▸ E.g. phase space sampling, integration, PDF evaluation.

  ▸ Recent efforts: Normalising Flows/VegasFlow/PDFFlow.

▸ Promote collaboration, training, funding and career opportunities in generator area.

  ▸ Has not always been easy to attract effort in this area.

  ▸ Highlighted quite strongly in European Strategy Update.

Performance Share of Accelerators

4
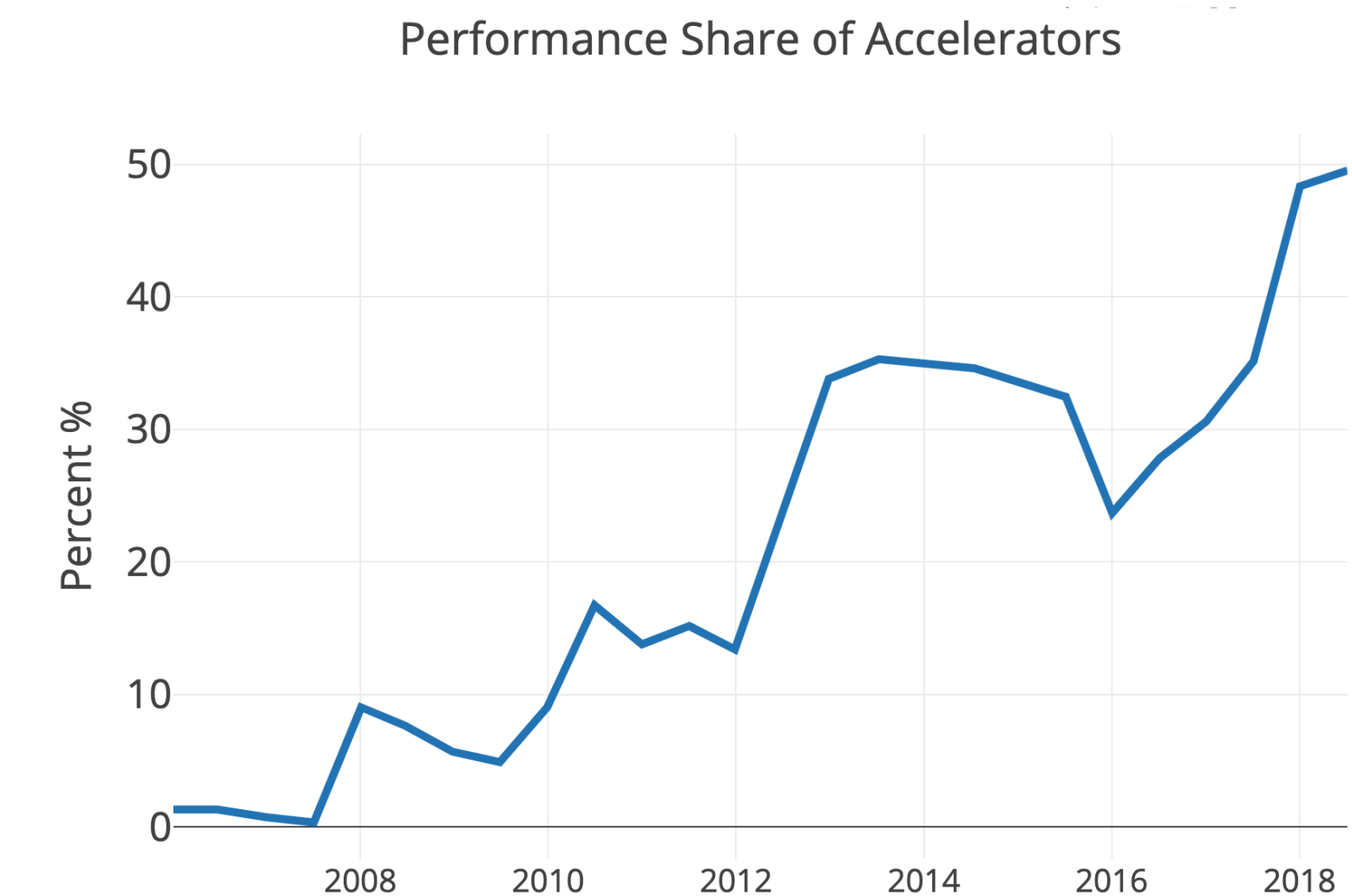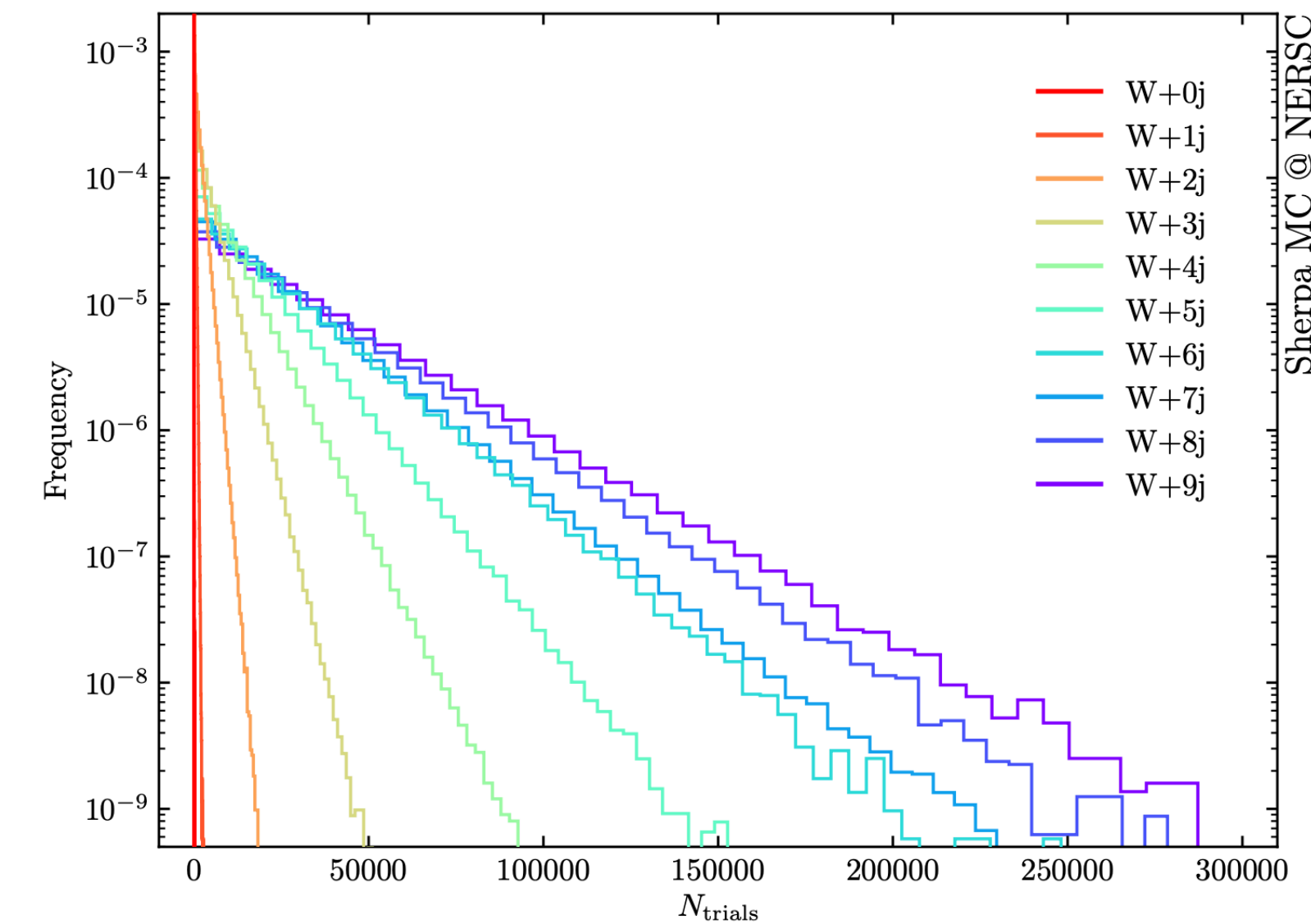
Back-ups
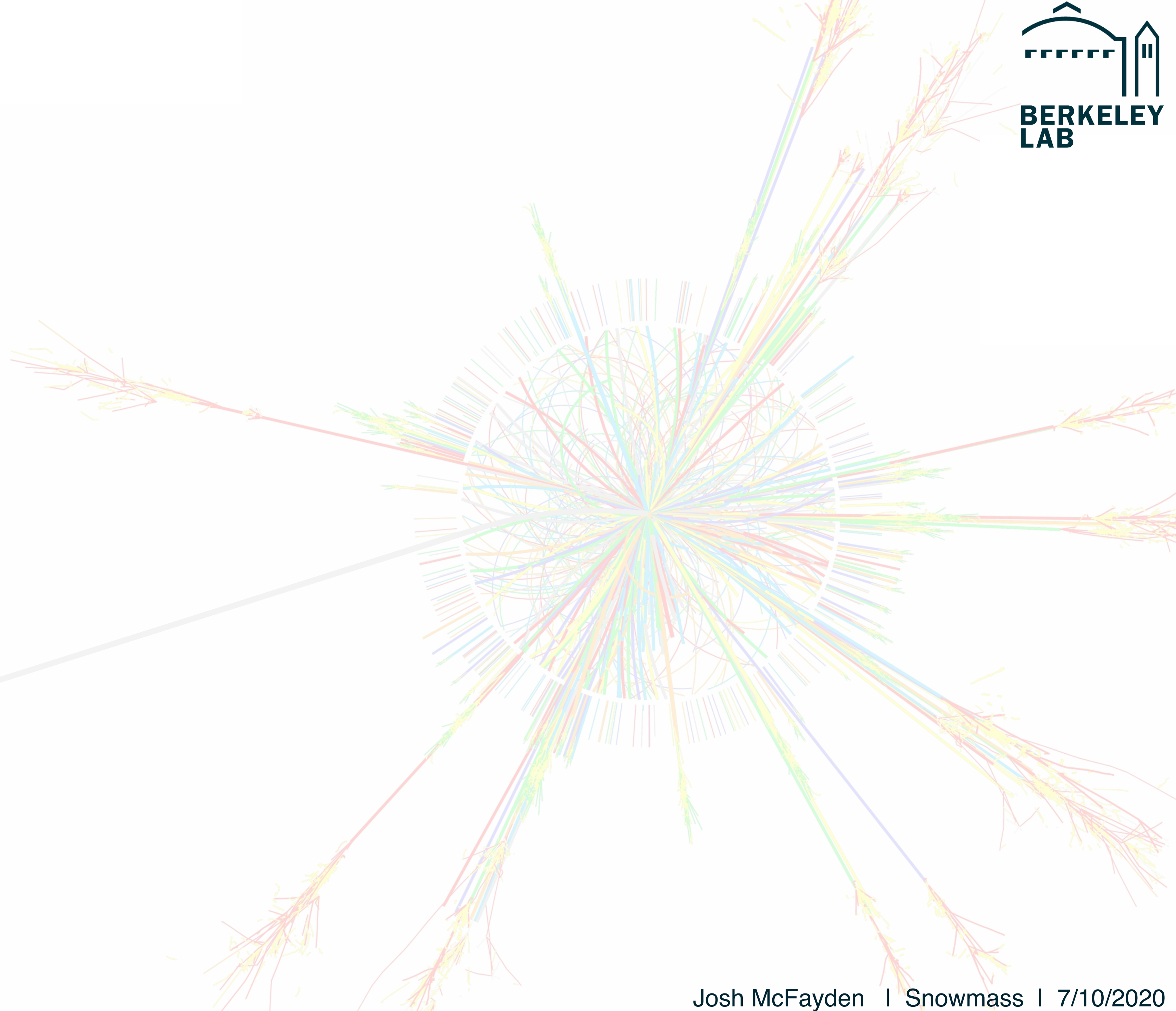
European Strategy Update

▸ From the strategy document itself:

  ▸ "Both exploratory research and theoretical research with direct impact on experiments should be supported, including recognition for the activity of providing and developing computational tools."

  ▸ "Large-scale data-intensive software and computing infrastructures are an essential ingredient to particle physics research programmes. The community faces major challenges in this area, notably with a view to the HL-LHC. As a result, the software and computing models used in particle physics research must evolve to meet the future needs of the field."

▸ From the deliberation document:

  ▸ "Theory plays an essential role in assessing the strategic importance for future investments in accelerators and experimental infrastructure"

  ▸ "Calculation-intensive areas such as precision phenomenology at colliders, lattice field theory or the development of Monte-Carlo event generators and other software tools require long time scales to yield results"

  ▸ "Computing and software are profound R&D topics in their own right and are essential to sustain and enhance particle physics research capabilities"

  ▸ "More experts need to be trained to address the essential needs, especially with the increased data volume and complexity in the upcoming HL-LHC era, and will also help in experiments in adjacent fields."

HEP Software Foundation Gen WG

‣ HSF Event Generators Working Group:

    ‣ https://hepsoftwarefoundation.org/workinggroups/generators.html

    ‣ ~Monthly meetings

‣ Mandate to provide a common forum for discussion and technical work on the physics event generators used by HEP experiments. It promotes the collaboration of experimental and theoretical physicists from different experiments and generator teams and of software and computing engineers, with the aim of having them work together on improving the current codes and production workflows and on making new theoretical advances easier to implement in a computationally efficient way.

‣ Recent publication submitted to *Computing and Software for Big Science*

    ‣ *"Challenges in MC event generator software for HL-LHC"* [2004.13687]

    ‣ Extended version of document prepared for LHCC.

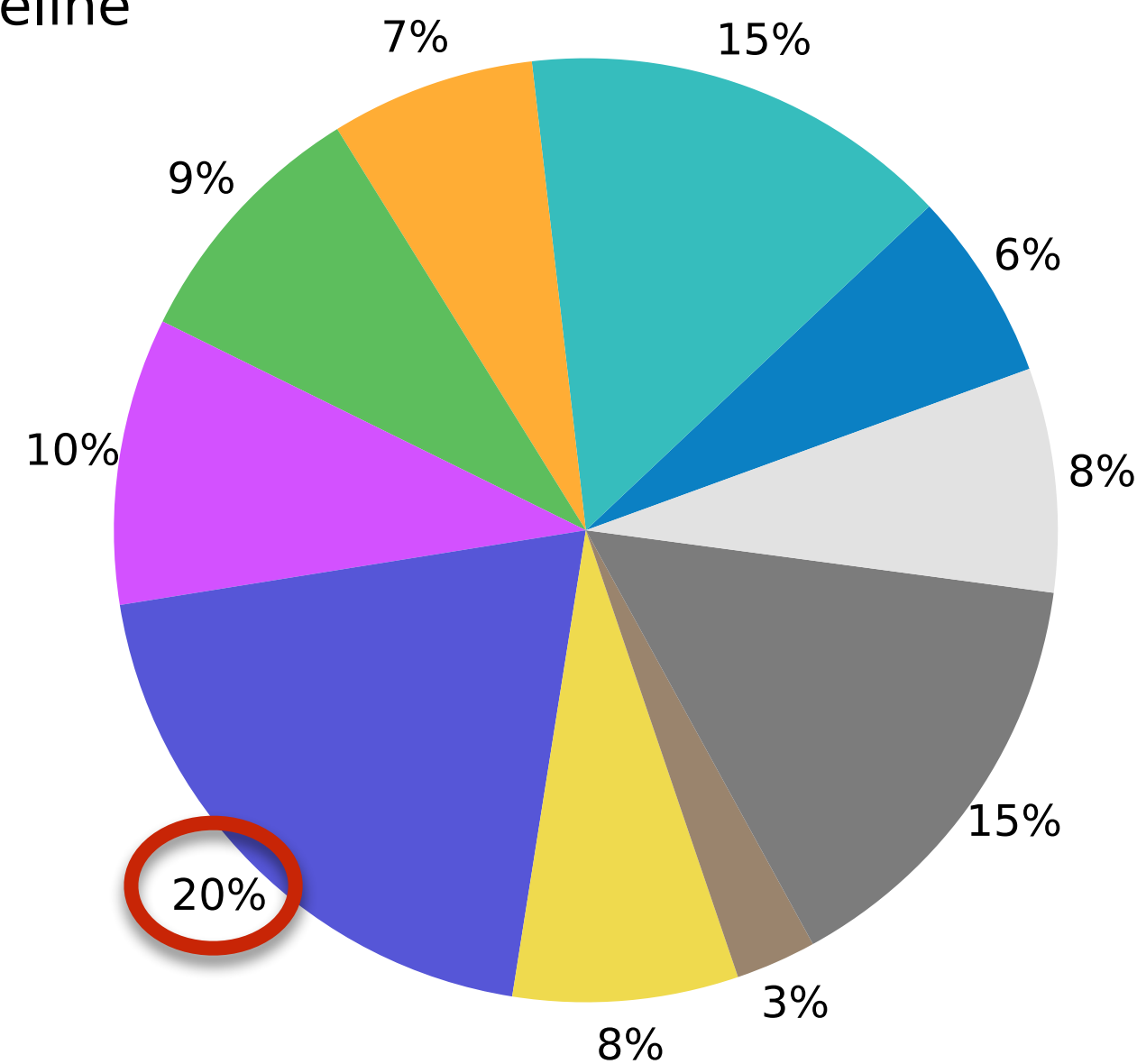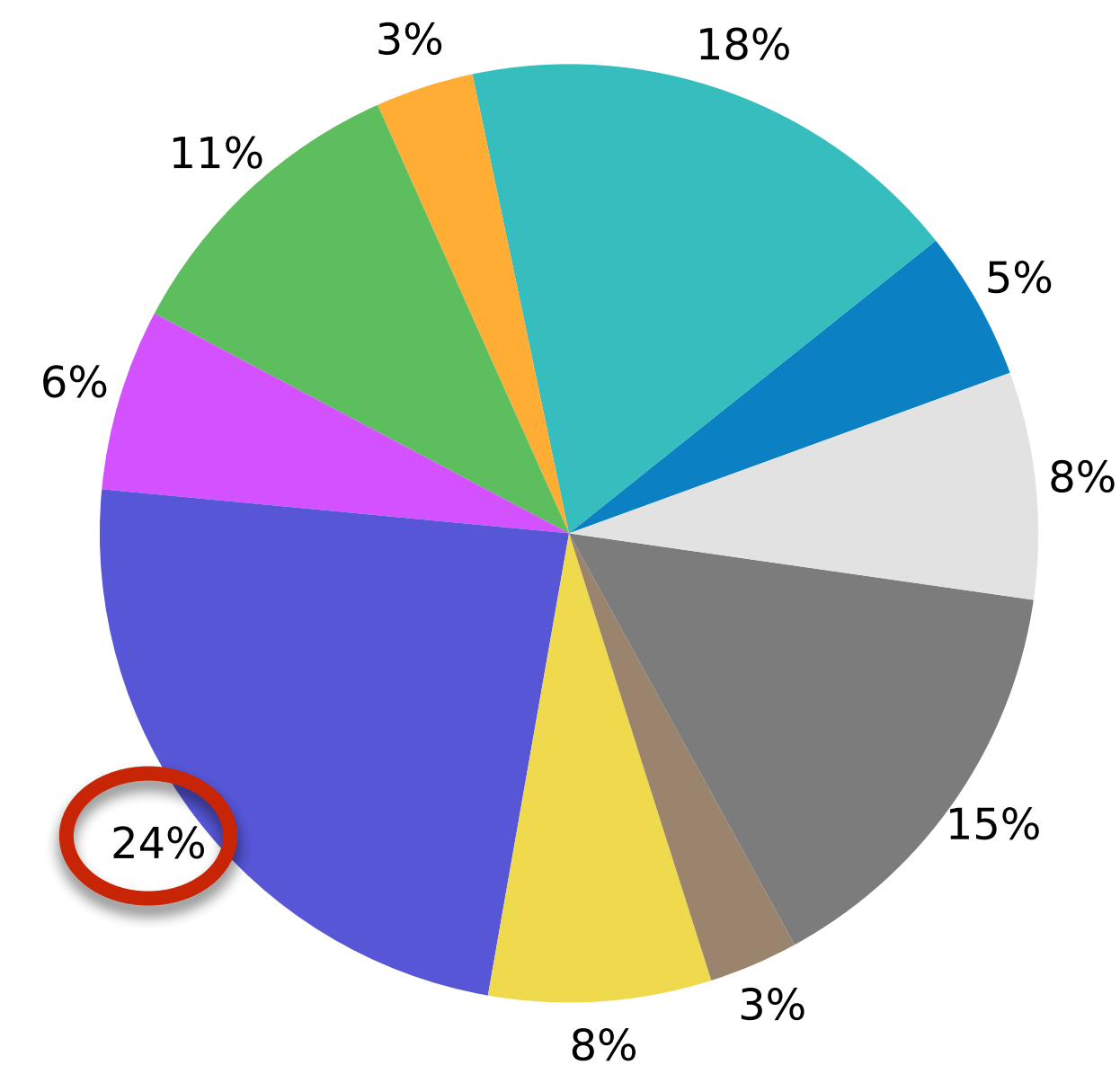# ATLAS Projections

ATLAS Preliminary
2020 Computing Model

Baseline

Aggressive R&D

Conservative R&D

ATLAS Preliminary
2020 Computing Model - CPU

○ Baseline
▲ Conservative R&D
▼ Aggressive R&D
— Sustained budget model
  (+10% +20% capacity/year)
△ LHCC common scenario
  (Conservative R&D, μ=200)

Annual CPU Consumption [MHS06·years]

Year

Run 3 (μ=55)   Run 4 (μ=88-140)   Run 5 (μ=165-200)

Data Proc
MC-Full(Sim)
MC-Full(Rec)
MC-Fast(Sim)
MC-Fast(Rec)
EvGen
Heavy Ions
Data Deriv
MC Deriv
Analysis