



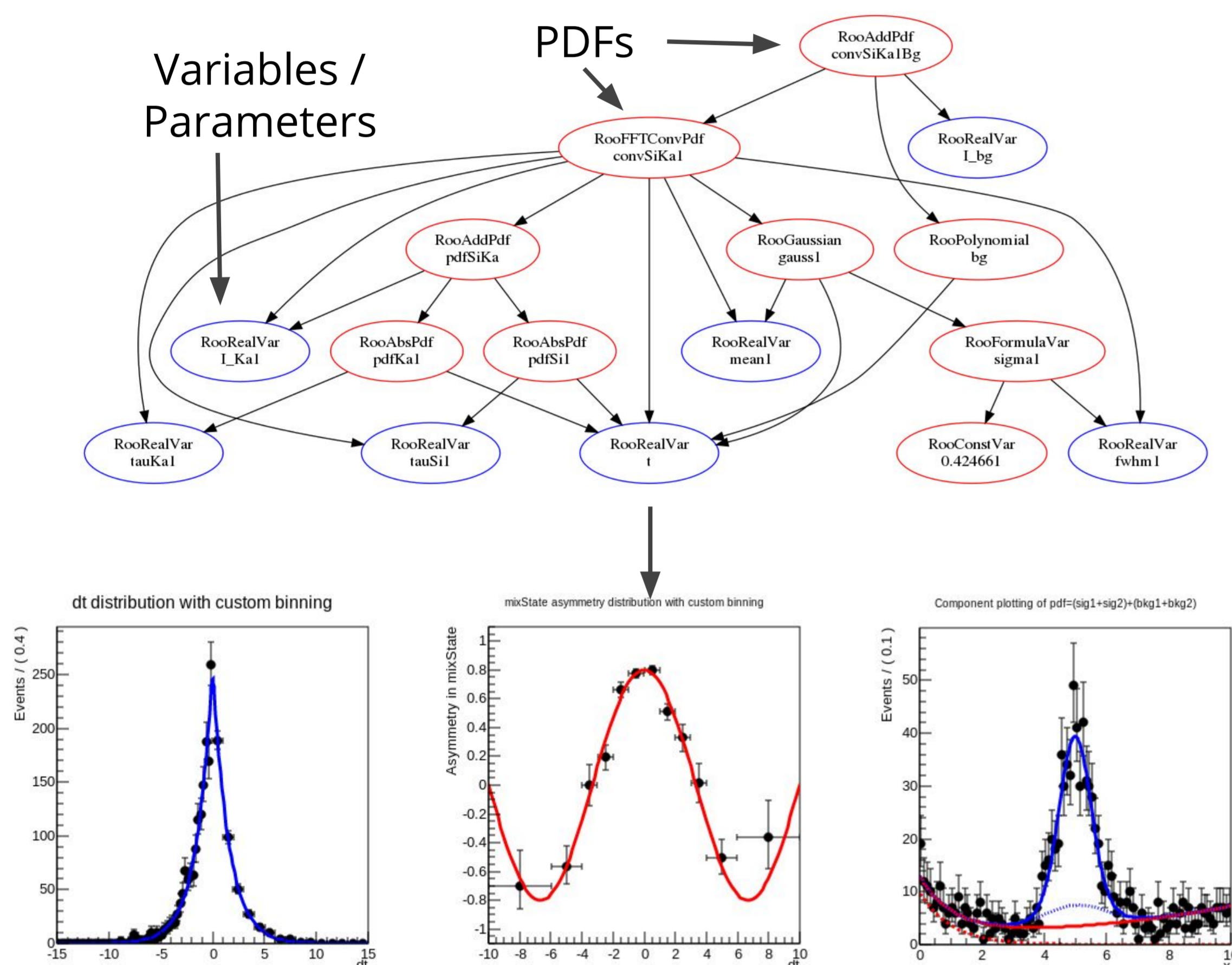
# Making RooFit Ready for Run 3 & Beyond

Stephan Hageboeck  
EP-SFT CERN

<https://root.cern>

## Introduction: RooFit

- ▶ C++ package in ROOT for creating statistical models [1]
- ▶ Tools to build composite PDFs (binned/unbinned)
- ▶ Fit to data / generate toys / integrate & normalise

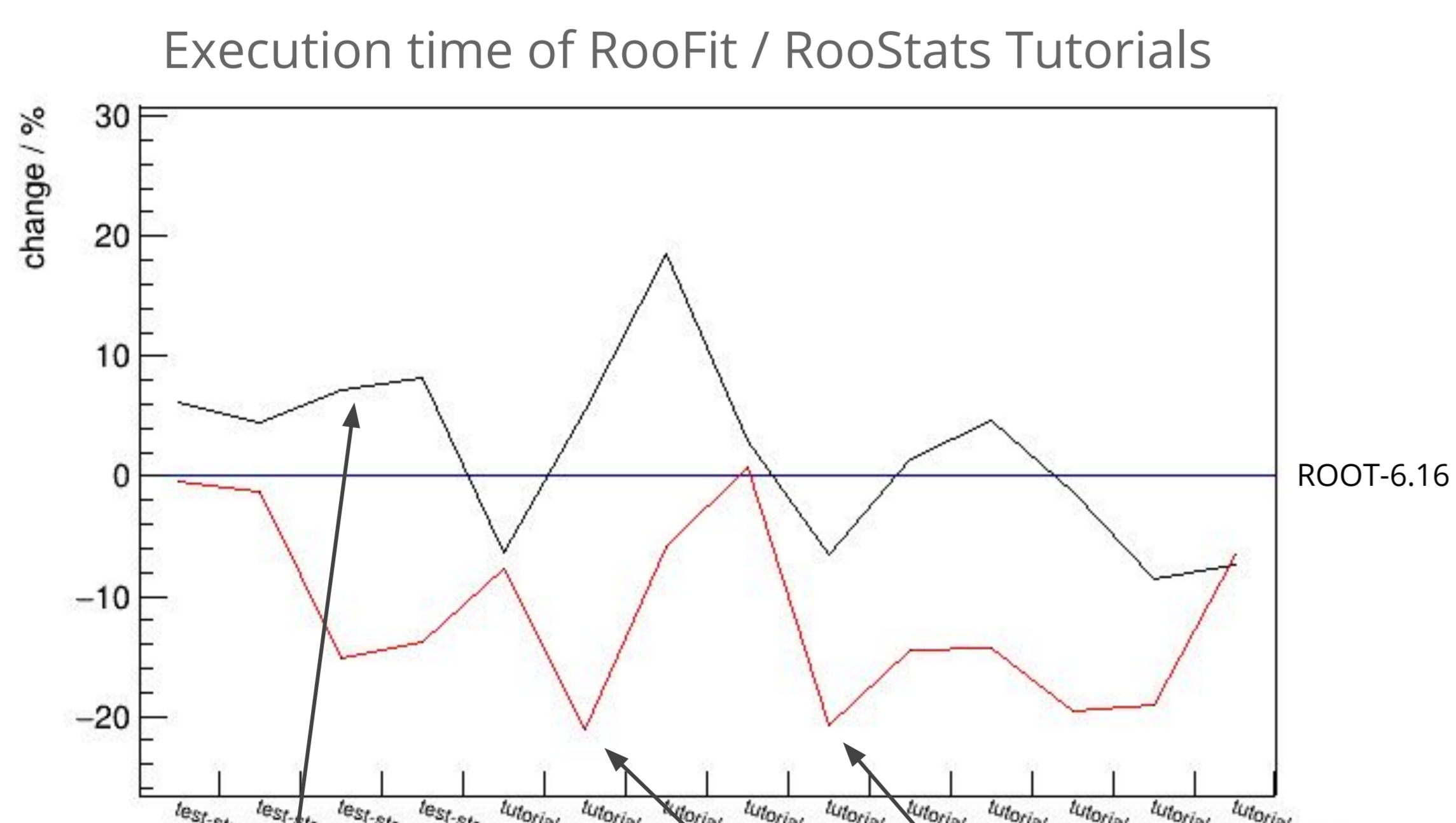


- ▶ Includes:
  - **RooStats**: Statistical tests with RooFit PDFs
  - **HistFactory**: Create models from histograms
- ▶ Used by all LHC experiments for statistical inference
- ▶ LHC's Run 3: >2x more data and more elaborate fits. **RooFit needs to handle this!**

## Faster Binned Fits

ROOT 6.18

- ▶ Iterating over RooFit's core collection "RooLinkedList" identified as bottleneck
    - RooFit core collections now based on std::vector
    - Provide STL-like interface (size, begin, end)
    - Existing code still works: legacy iterators re-implemented
  - ▶ **Results:**
    - Easier coding with range-based for loops
    - **Fits with STL iterators ~20% faster**
- ATLAS H→bb Run 1 measurement [2]:  
11:30 min → 9:20 min, identical result



Heavy use of legacy iterators:  
Sometimes gains, sometimes slowdowns

Critical legacy iterators replaced:  
**20% faster**

```
Iterator* paramIter = paramList.createIterator();
RooAbsArg* param;
while((param = (RooAbsArg*)paramIter->Next())) {
    _paramList.add(*param);
}
delete paramIter;
```

```
for (const auto param : paramList) {
    _paramList.add(*param);
}
```

## Faster HistFactory Models

ROOT 6.18

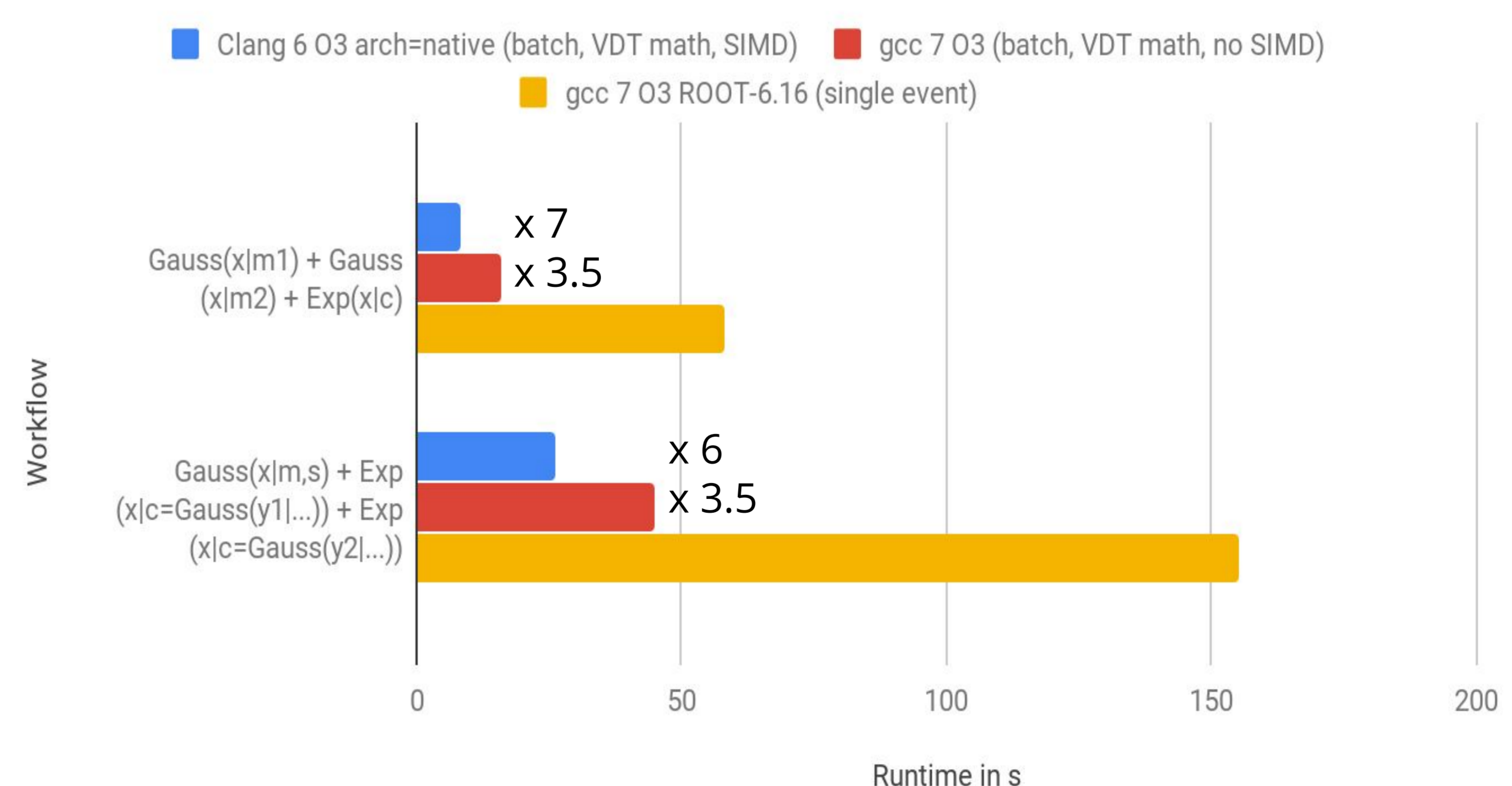
- ▶ Create statistical models from ROOT histograms [3]
    - Provide histograms + XML/C++/Python description
    - Implement combined measurement with
      - Multiple channels
      - Multiple signal/background samples
      - Various shape, cross-section and statistical uncertainties
    - Generating large models requires handling thousands of histograms
  - ▶ Move semantics for HistFactory classes implemented, memory management of histograms improved
  - ▶ **~ 10x speedup**
    - E.g. large\* ATLAS model:  
ROOT-6.16: 1800s  
ROOT-6.18: 150s
- \* Model similar to ATLAS H→bb [2] with 704 observables, 253 nuisance parameters, 28 channels, 2705 Gaussian constraints, 533 Poisson constraints, 10832 histograms

## Batched PDF Computations

Experimental

- ▶ Current RooFit evaluates probabilities for single events
- ▶ One function call for **each PDF\*** in expression tree **for every data event**
  - \* Subtree expressions can be cached if parameters remain constant
- ▶ **Batched Computations**
  - Read full batch of data using std::span
  - Call evaluation functions only **once** per fit step
  - Backward compatibility: "hybrid" mode
    - PDFs that support batch evaluation in batch mode
    - Other PDFs (e.g. user-supplied) in single-event mode
  - **Speedup of 2x - 3.5x**, depending on model size
- ▶ **SIMD computations**
  - If evaluating in batches: SIMD computations possible
  - Experimental vectorised loops for Gauss, Poisson, Exp using VDT math [3] alternatives for std::exp, std::log
  - **Speedup of up to 7x**

## Fitting 2 Million Events in Batch Mode



## References

- [1] W. Verkerke and D. Kirkby, 2003, *The RooFit toolkit for data modeling*, [arXiv:physics/0306116](https://arxiv.org/abs/physics/0306116)
- [2] ATLAS Collaboration, 2015, *Search for the H→bb decay of the Standard Model Higgs boson in associated (WZ)H production with the ATLAS detector*, [JHEP01\(2015\)069](https://arxiv.org/abs/1506.069)
- [3] K. Cranmer et al., 2012, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, [CERN-OPEN-2012-016](https://arxiv.org/abs/1201.016)
- [4] D. Piparo, V. Innocente and T. Hauth, 2014, *Speeding up HEP experiment software with a library of fast and auto-vectorisable mathematical functions*, [J. Phys.: Conf. Ser. 513 052027](https://arxiv.org/abs/1305.2027)

[stephan.hageboeck@cern.ch](mailto:stephan.hageboeck@cern.ch)