A top to bottom framework approach to vectorization

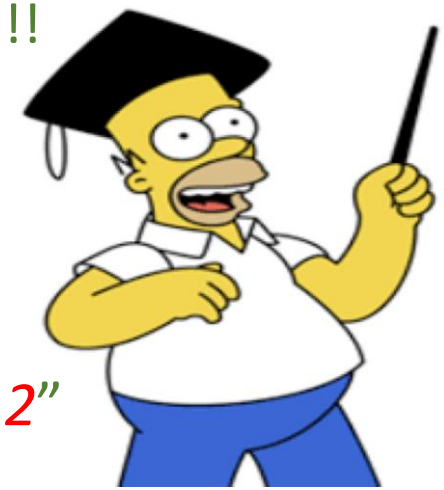
A. Gheata for the GeantV team

CHEP 2018

9-13 July 2018, Sofia, Bulgaria

Preamble

"You need to go faster for Run3 !!!



"Manufacturers made a longer model for you!
It's called *AVX-512*"

"You maaay need to fill it up…"

"Some coordination is also needed"

A top to bottom, bottom-up approach, CHEP'18

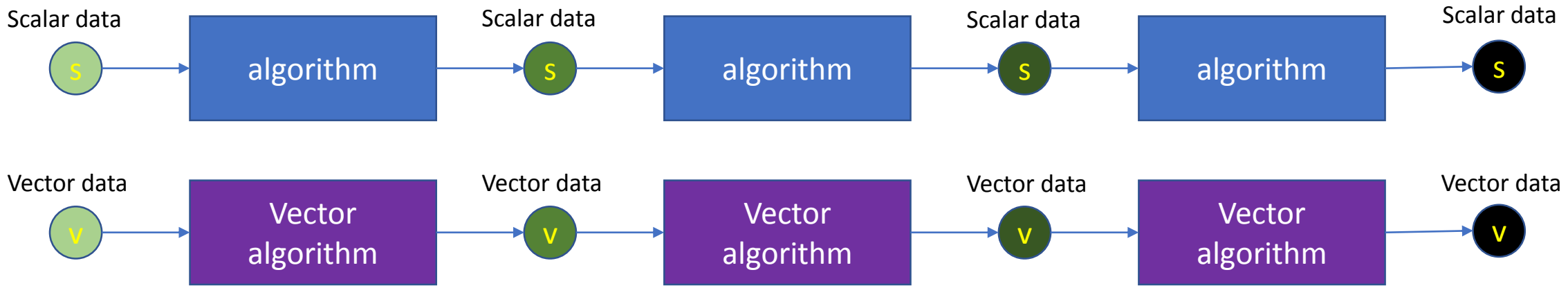
"But… I've reached the theoretical stroke rate limit!…"

"Tried it, doesn't go faster …"

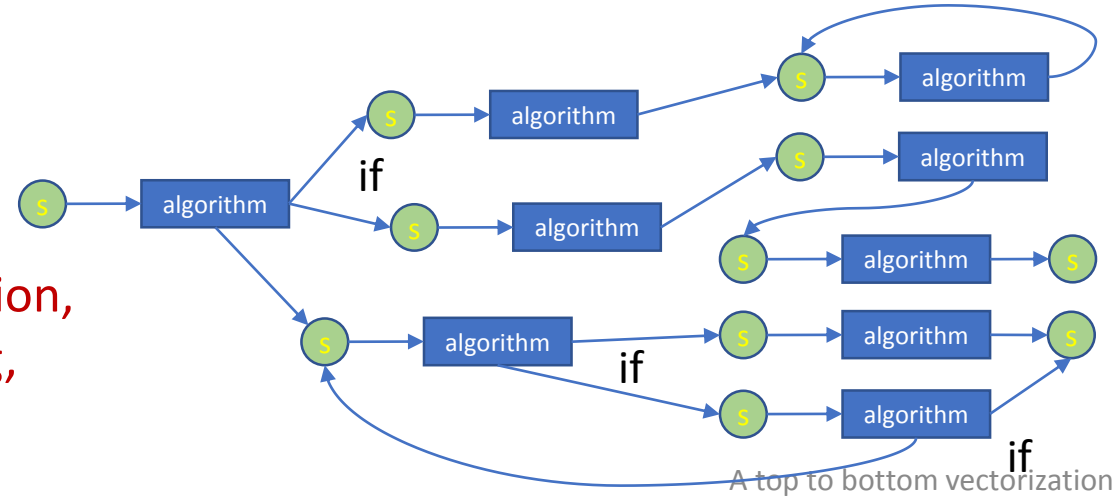
"I have few ideas how to do that…"

"Can I use my old whistle?"

Vector workflows: vectorizing the data loop



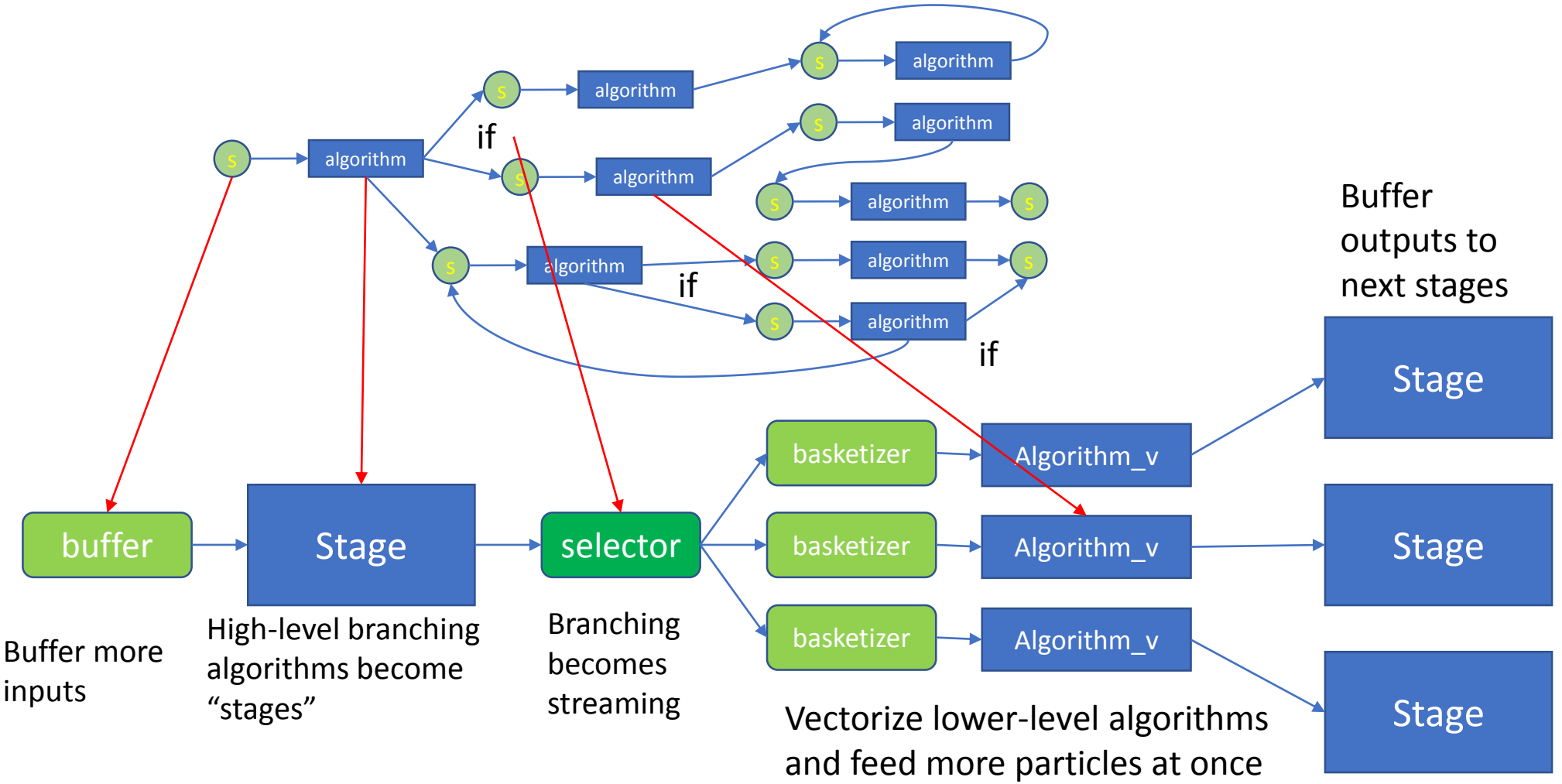
Online



Simulation, tracking, analysis

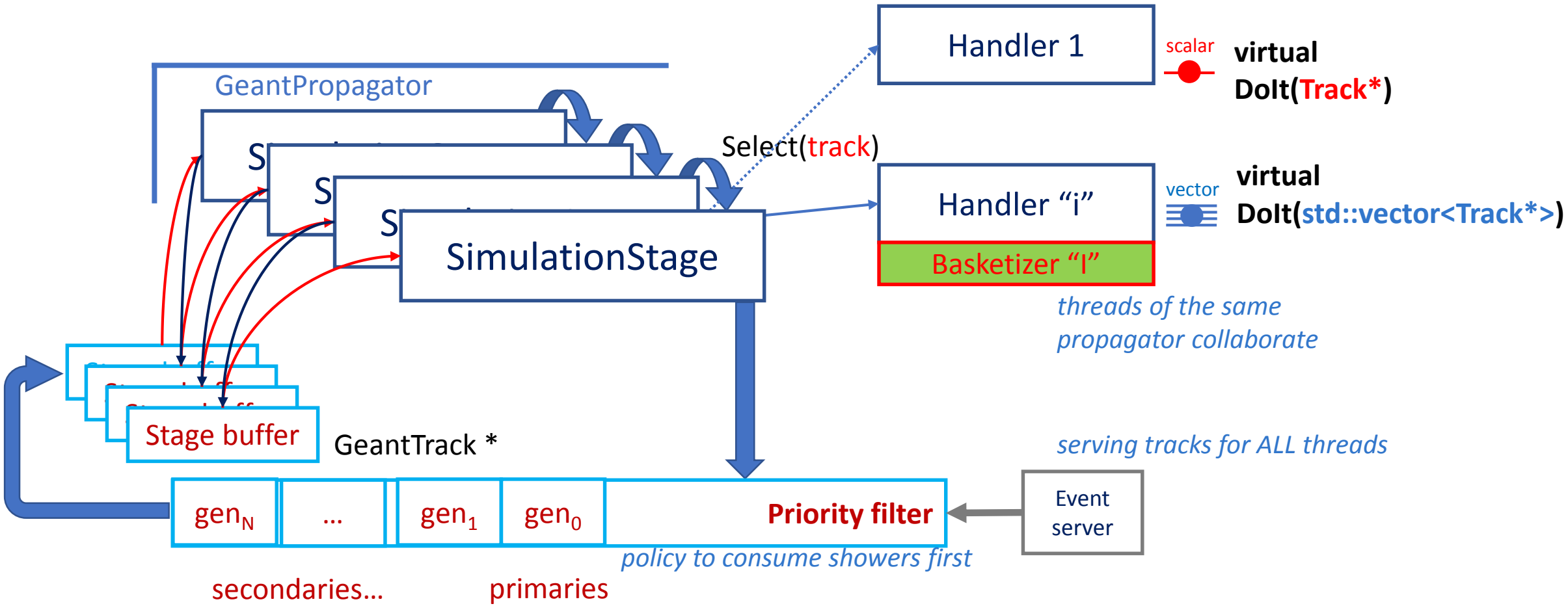
Complex workflows in vector approach

To which extent?
Potential benefits?
Programmability?
Maintainability?
-> GeantV R&D



GeantV version3 workflow

Both scalar/vector flow are supported

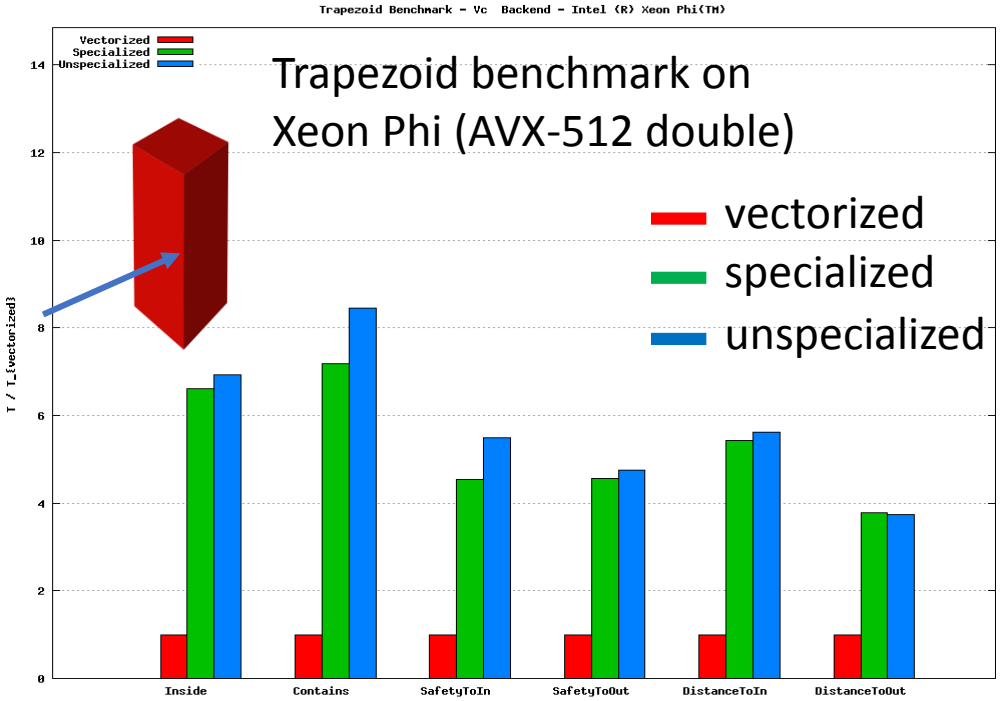


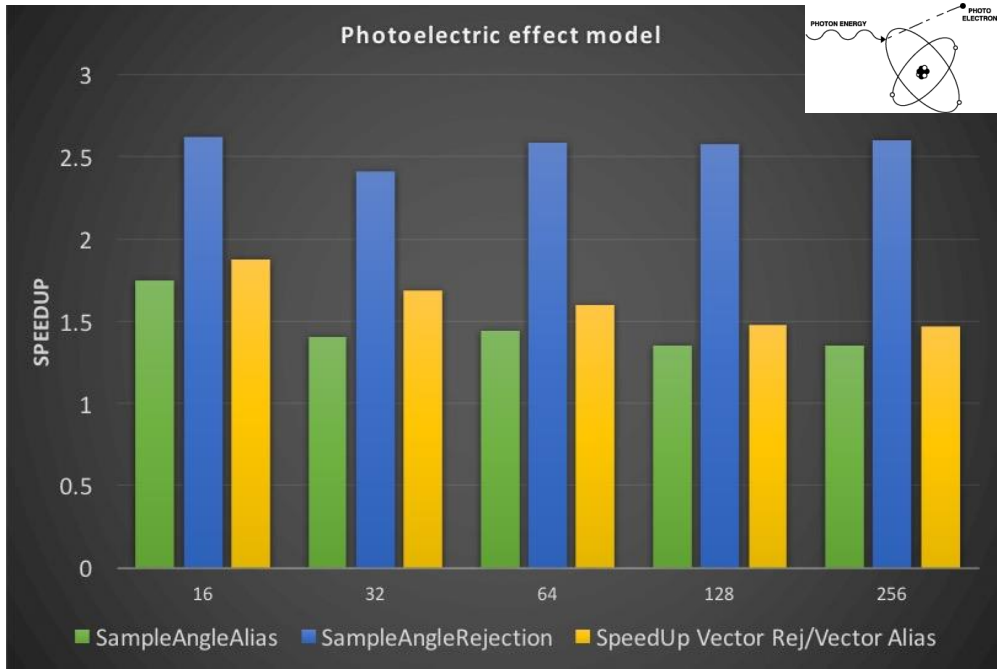
The bottom view: core components

https://github.com/root-project/veccore

https://github.com/root-project/vecmath

- Vectorization of low-level algorithms
  - Geometry (M.Gheata talk), physics (M. Bandieramonte talk), propagation in field
  - Vectorizing on multi-particle inputs, but also on internal loops when possible



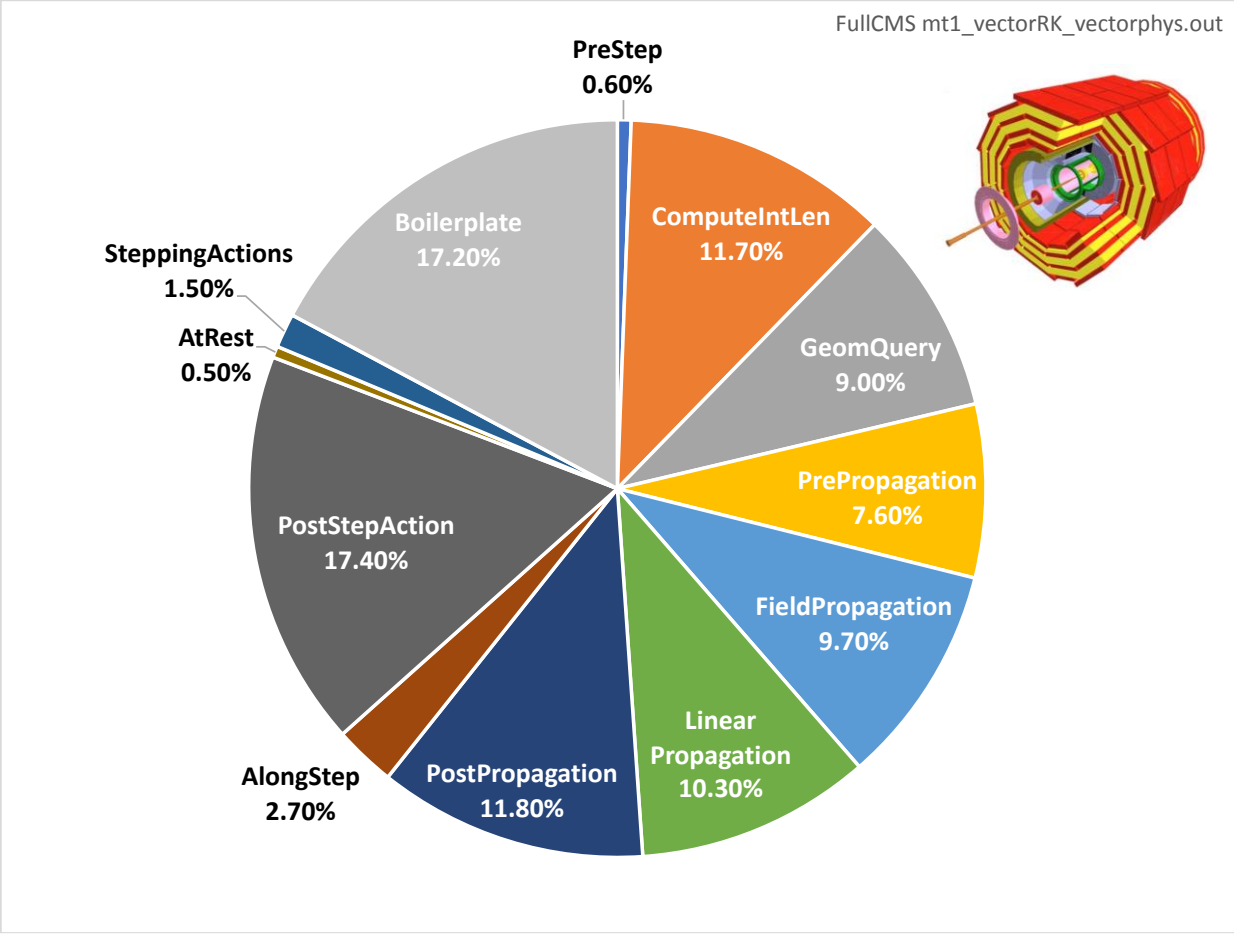


Arch specs, White background

The top view: simulation stages

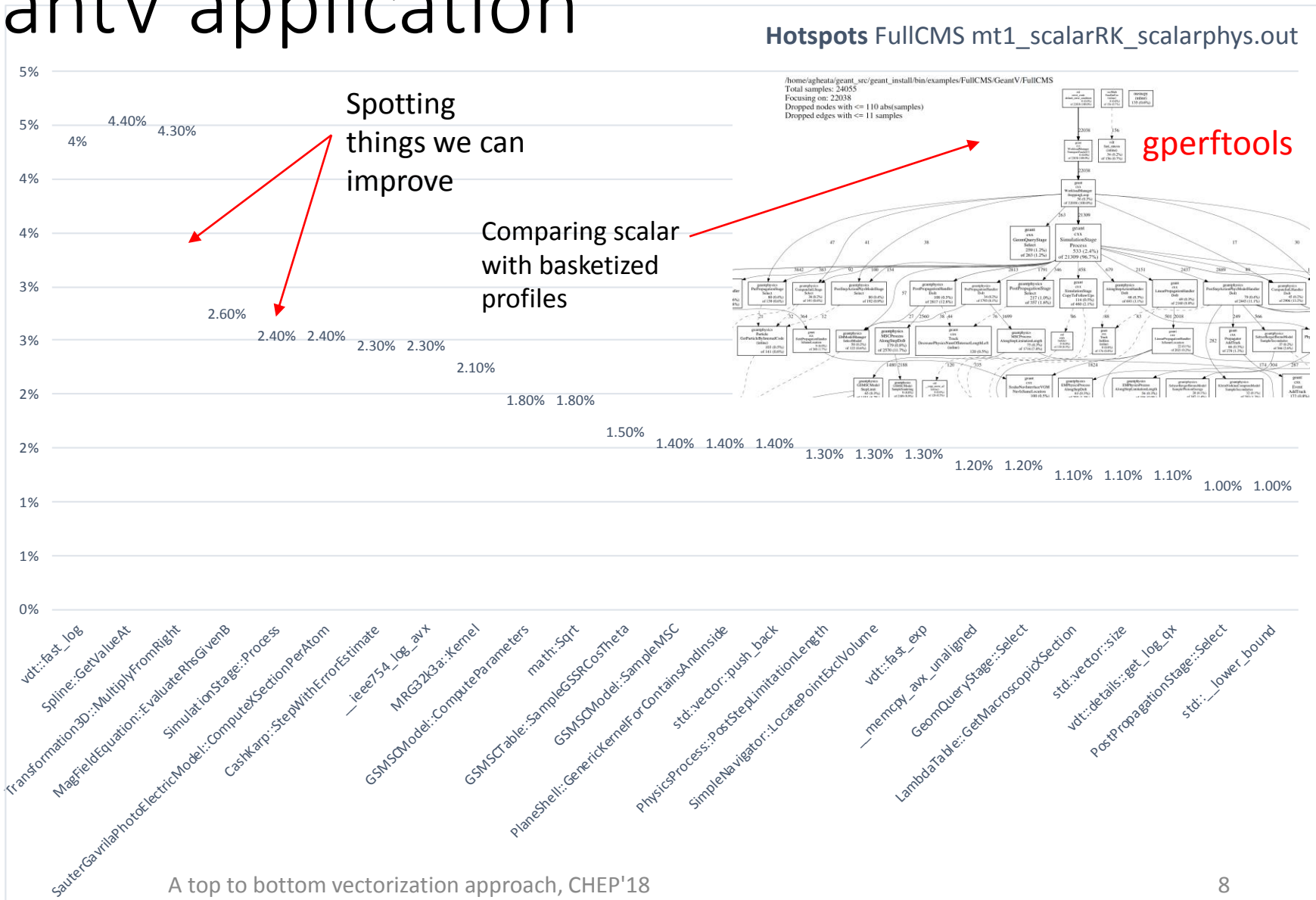
- The relative weight depends on the setup
  - Geometry complexity, physics list, user cuts
- Many stages containing already vectorized code
  - Field propagation giving expected speed-up in the full flow
  - Not yet the case for geometry and post step action physics stage
- Thorough benchmarking against equivalent Geant4 application ongoing
- Current phase: performance optimizations
  - Detailed profiling to understand overheads and bottlenecks

Top-level profile of basketized GeantV simulation of CMS



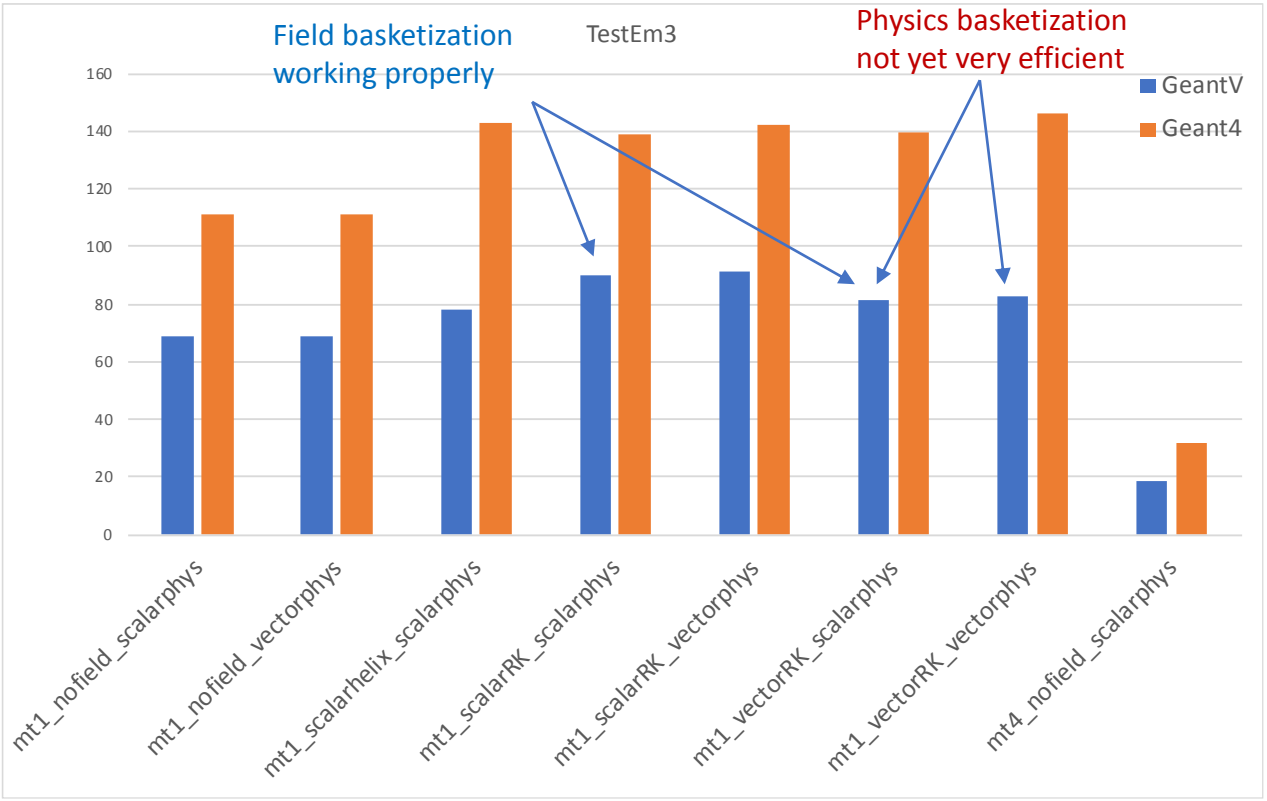
Profiling GeantV application

- Understanding hotspots
- Understanding basketization overheads
- Working on reducing scalar bottlenecks



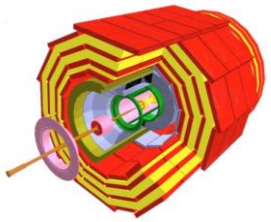
Benchmarking metrics

TestEm3 example, 100 GeV e- fixed gun, GeantV EM physics list



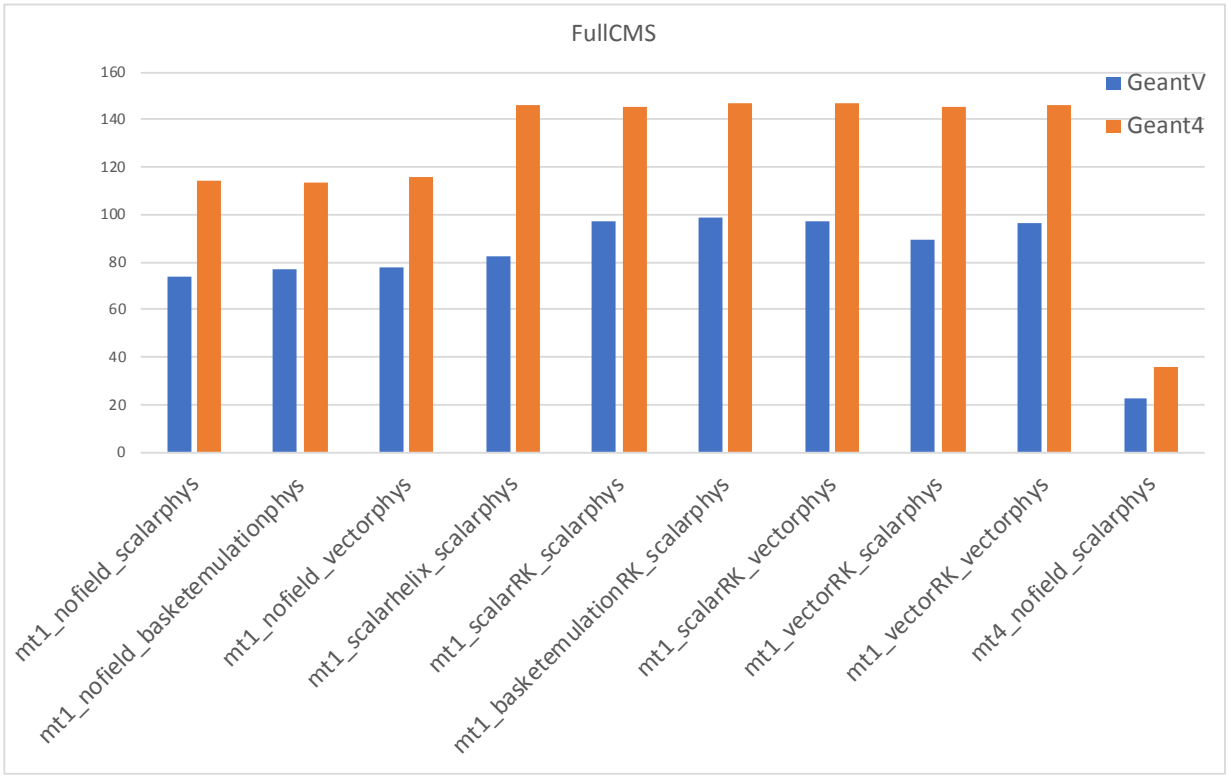
- Benchmarking several different configurations
  - Magnetic field ON/OFF
  - Field Runge-Kutta vs. helix
  - Physics, field, (geometry in future) basketizing ON/OFF
  - Basket emulation as scalar loops (understand basketization overheads)
  - MT simulation
- Corresponding profiles extracted
  - Plan to do the same for Geant4 with different field configurations

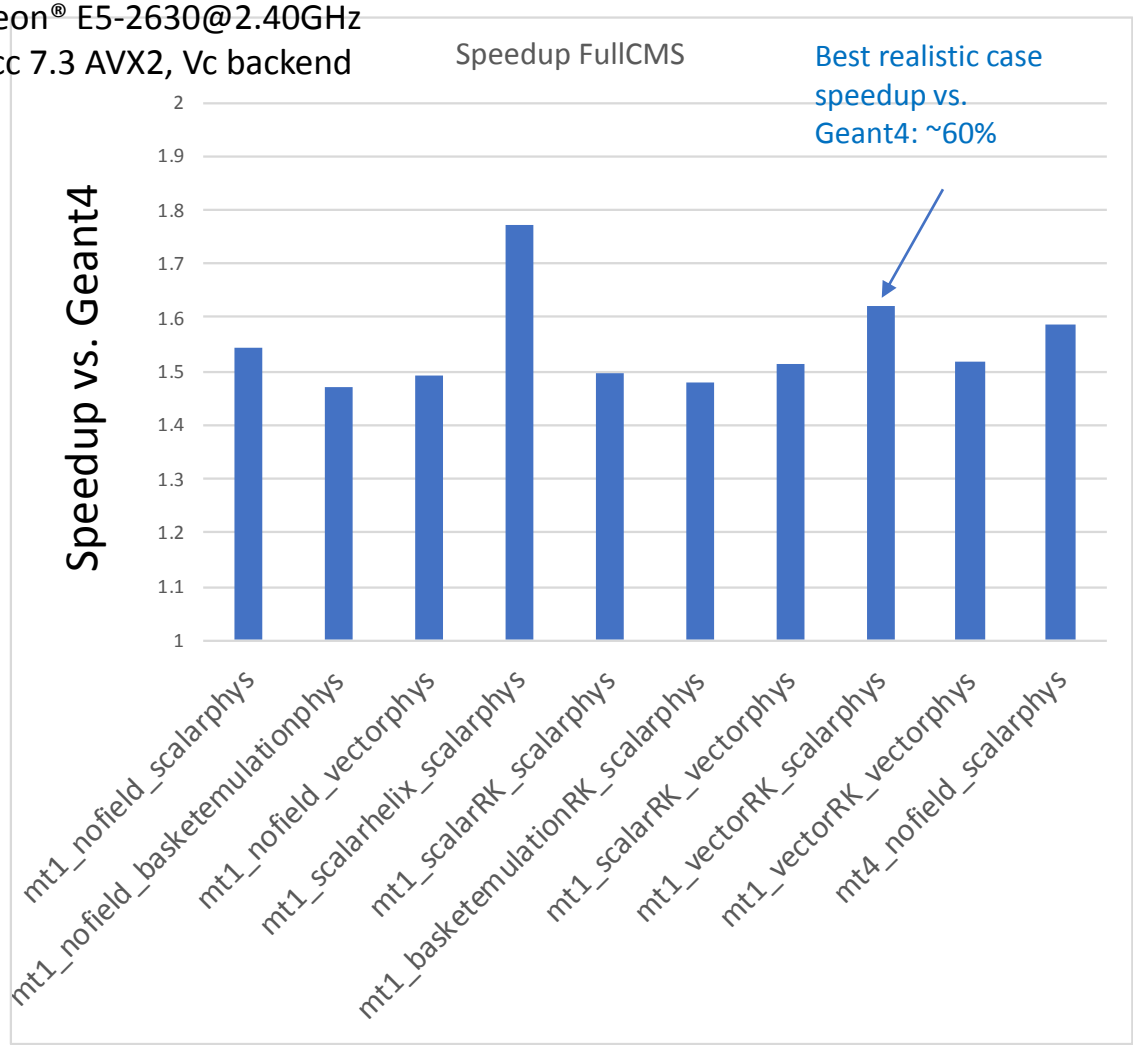
Current performance benchmarks - CMS



CMS geometry, 100 GeV e^- isotropic gun
GeantV EM physics list, same geometry & production cuts GV/G4

Xeon® E5-2630@2.40GHz
gcc 7.3 AVX2, Vc backend





Ongoing work

- Vectorization work and connection of vectorized code
  - Multiple scattering and photoelectric model (first implementation available)
- Reducing basketization overhead
  - Light version (no track exchange among threads)
- Geometry optimizations
  - Specialized volume navigators (reducing scalar bottlenecks + de-virtualizing + caching)
- Physics optimizations + new features
  - Selective switch-on alias tables for ”hot” regions instead of rejection sampling, for models where it brings profit
  - Use of floating point instead of double case by case
  - Work on pRNG to support reproducibility in MT + basketized mode
- These can bring to a factor of ~2x for the full CMS benchmark for the beta tag end 2018 compared to Geant4
  - Out of which 25-40% from basket vectorization

Outlook

- Top to bottom vectorization on input data opens new optimization opportunities
  - New algorithms become vectorizable
  - Pipeline workflows are perfect candidates, but more complex workflows can also benefit
  - Basketizing data comes with benefits (better caching, vectorization), but also overheads (data copying, larger memory footprint)
- Benefits of basketized approach become visible in GeantV
  - Not as large as initially expected due to the large complexity of simulation code
  - A detailed profiling analysis and optimizations ongoing
- Beta release by the end of 2018
  - Demonstrator for realistic simulation of EM showers in the context of LHC
  - Targeting a factor of 2 speed-up, out of which up to ~40% from vectorization

GeantV collaborators

- **BARC:** S. Behera, A. Bhattacharyya, H. Kumawat, R. Sehgal
- **CERN**: A. Ribon, A. Gheata, D. Konstantinov, F. Hariri, F. Carminati, G. Cosmo, G. Amadio, GR Khattak, I. Goulas, J. Apostolakis, M. Bandieramonte, M. Gheata, M.Novak, O. Shadura, P. Mendez Lorenzo, PR Karpinki, R. Brun, S. Wenzel, S. Vallecorsa, V. Ivantchenko, W. Pokorski, K. Nikolics
- **FNAL**: D. Elvira, G. Lima, K. Genser, P. Canal, SY Jung, A. Miranda-Aguillar, J. Martinez
- **Students 2017-2018**: R. Schmitz, V. Drogan, S. Sharan, E. Orlova, Ananya, D. Savin

Thank you !