A top to bottom framework approach to vectorization

A. Gheata for the GeantV team

CHEP 2018

9-13 July 2018, Sofia, Bulgaria

“A view on performance”

by Homer Simpson

“You need to go faster for Run 3 !!!”

“Manufacturers made a better model for you! It’s called AVX-512”

“SIMD is about many doing the same thing at once…”

“For efficiency, some synchronization is also needed”

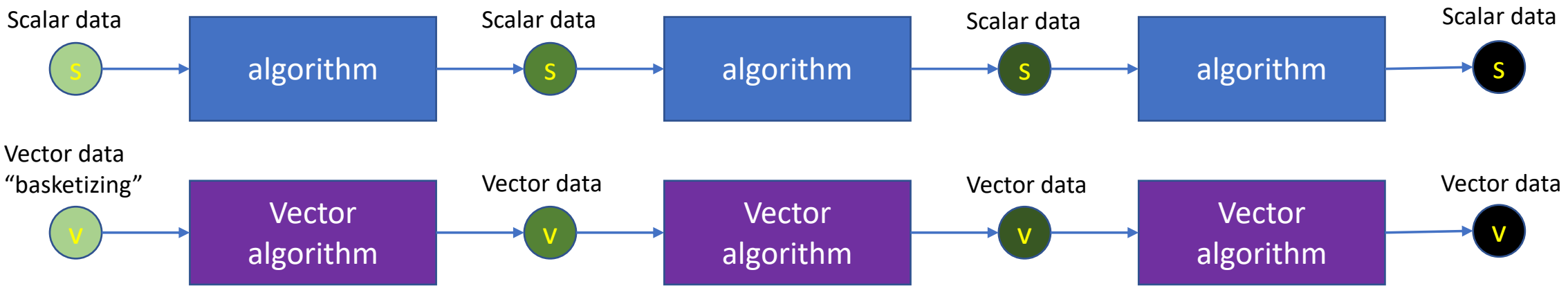
“But… I’ve reached the theoretical limit for the stroke rate!…”

“Tried it with my best paddle, doesn’t go faster …”

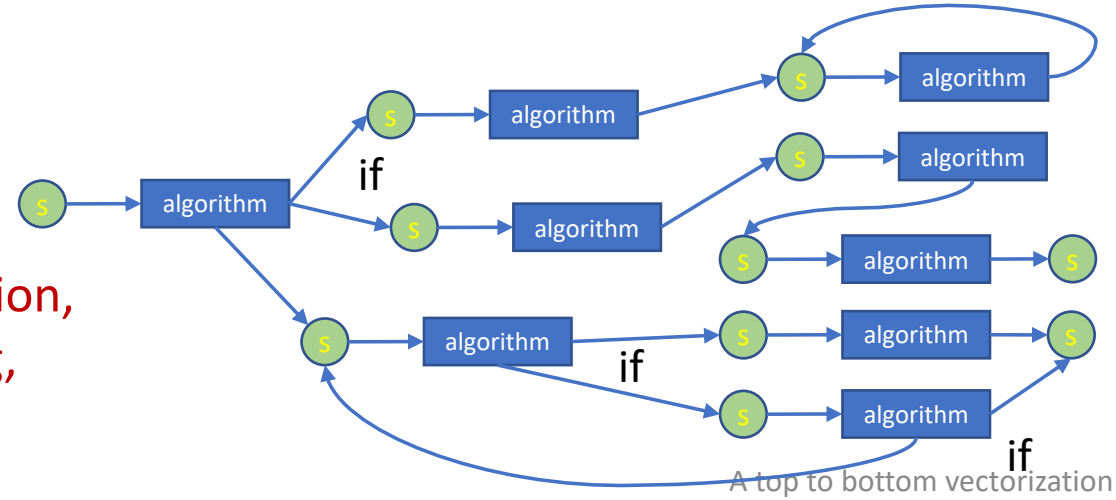
“I understand now…”

“Can I use my old whistle?”

Beyond loop vectorization



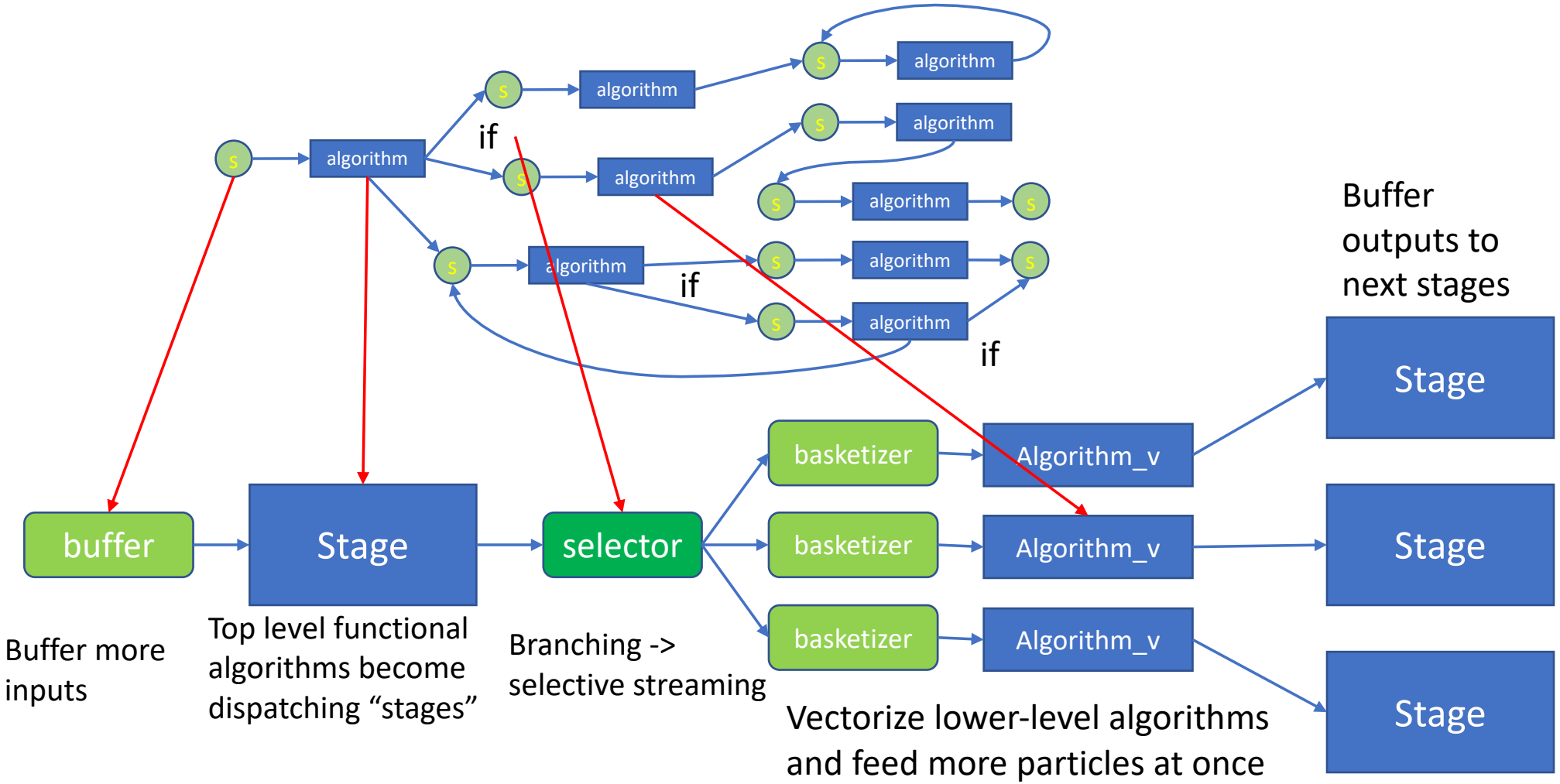
Online



Simulation, tracking, analysis

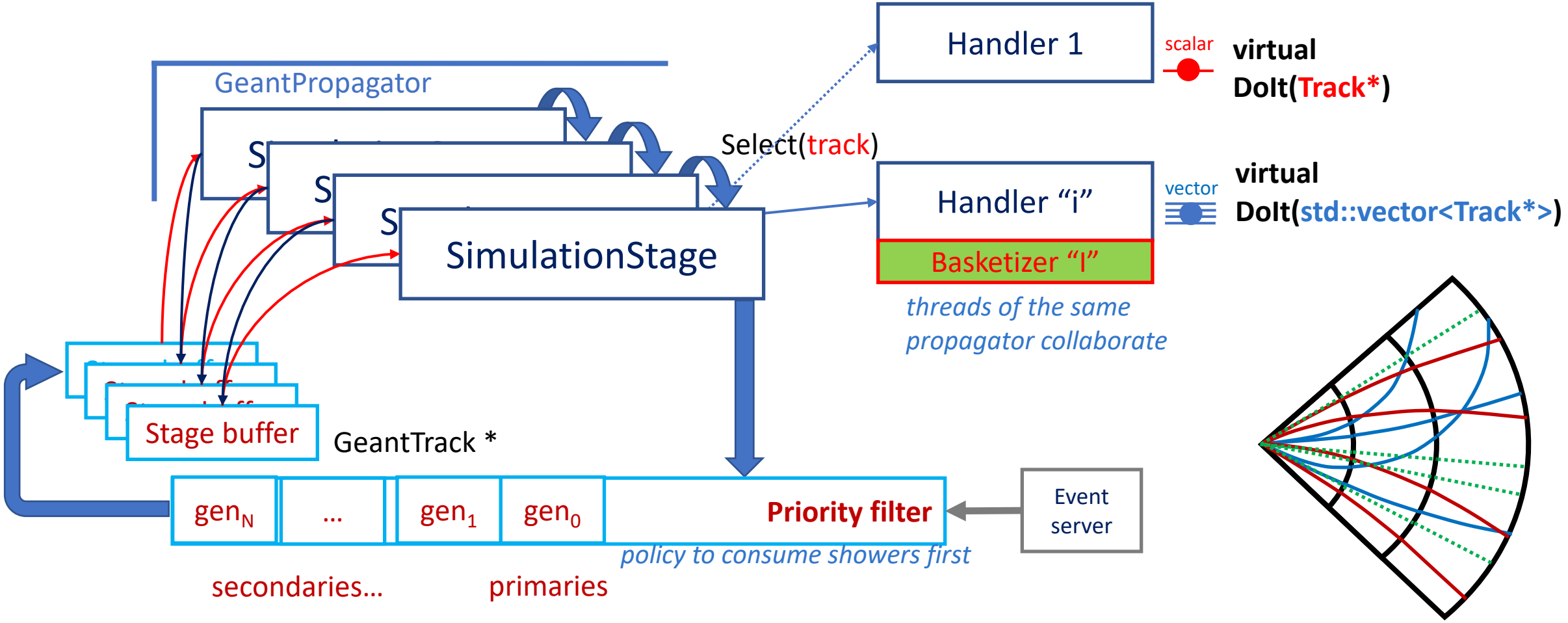
Complex workflows in vector approach

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



GeantV multi-particle processing

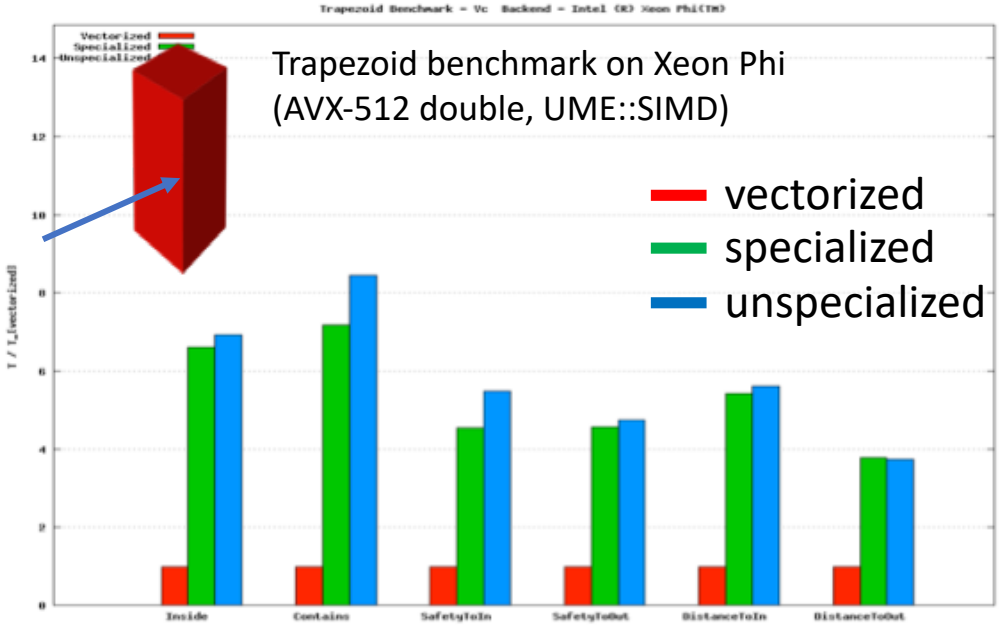
Both scalar/vector flow are supported

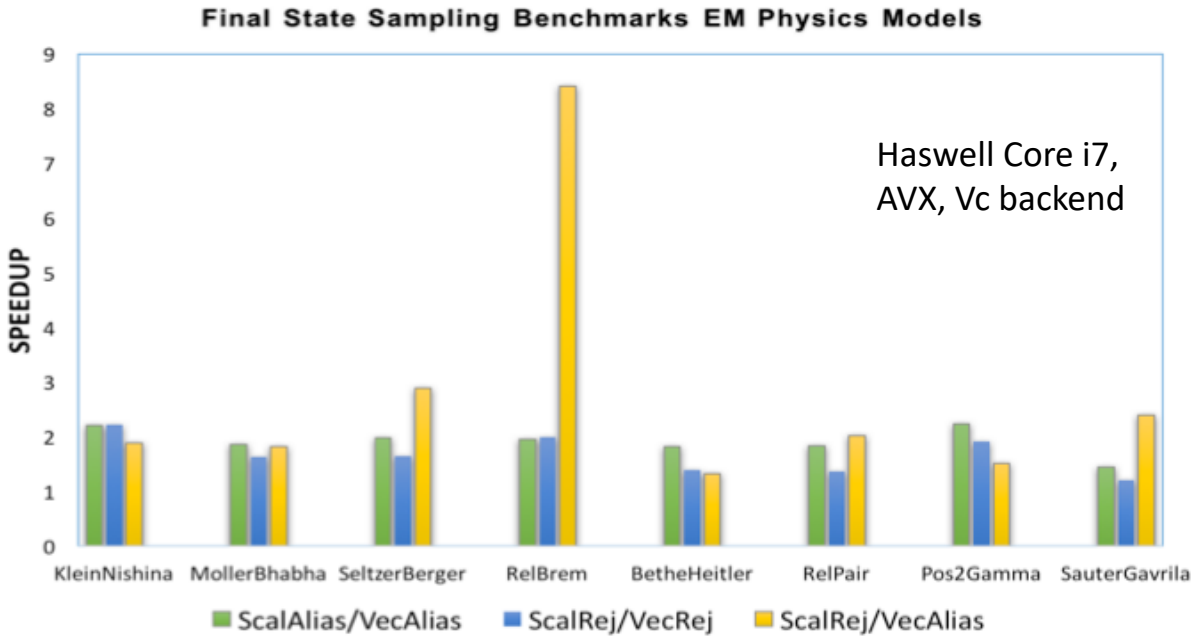


The bottom view: core components

Abstraction layer: https://github.com/root-project/veccore
Common math, RNG, …: https://github.com/root-project/vecmath

- Vectorization of low-level algorithms using VecCore
  - See ”Vectorization of ROOT Mathematical Libraries” talk by L. Moneta for VecCore details
  - Geometry (M.Gheata talk), physics (M. Bandieramonte talk), propagation in field
  - Vectorizing on multi-particle inputs, but also on internal loops when possible

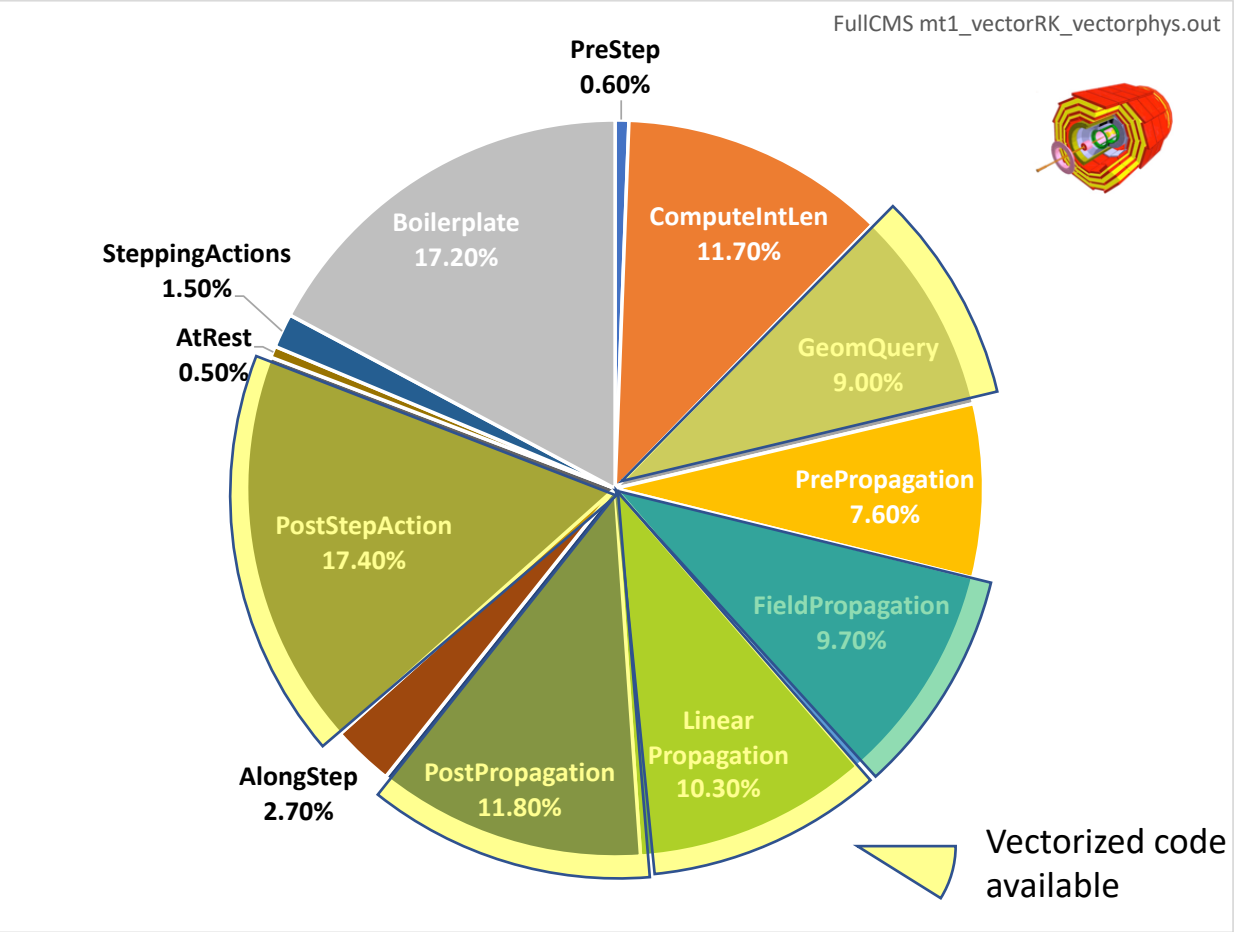




The top view: simulation stages

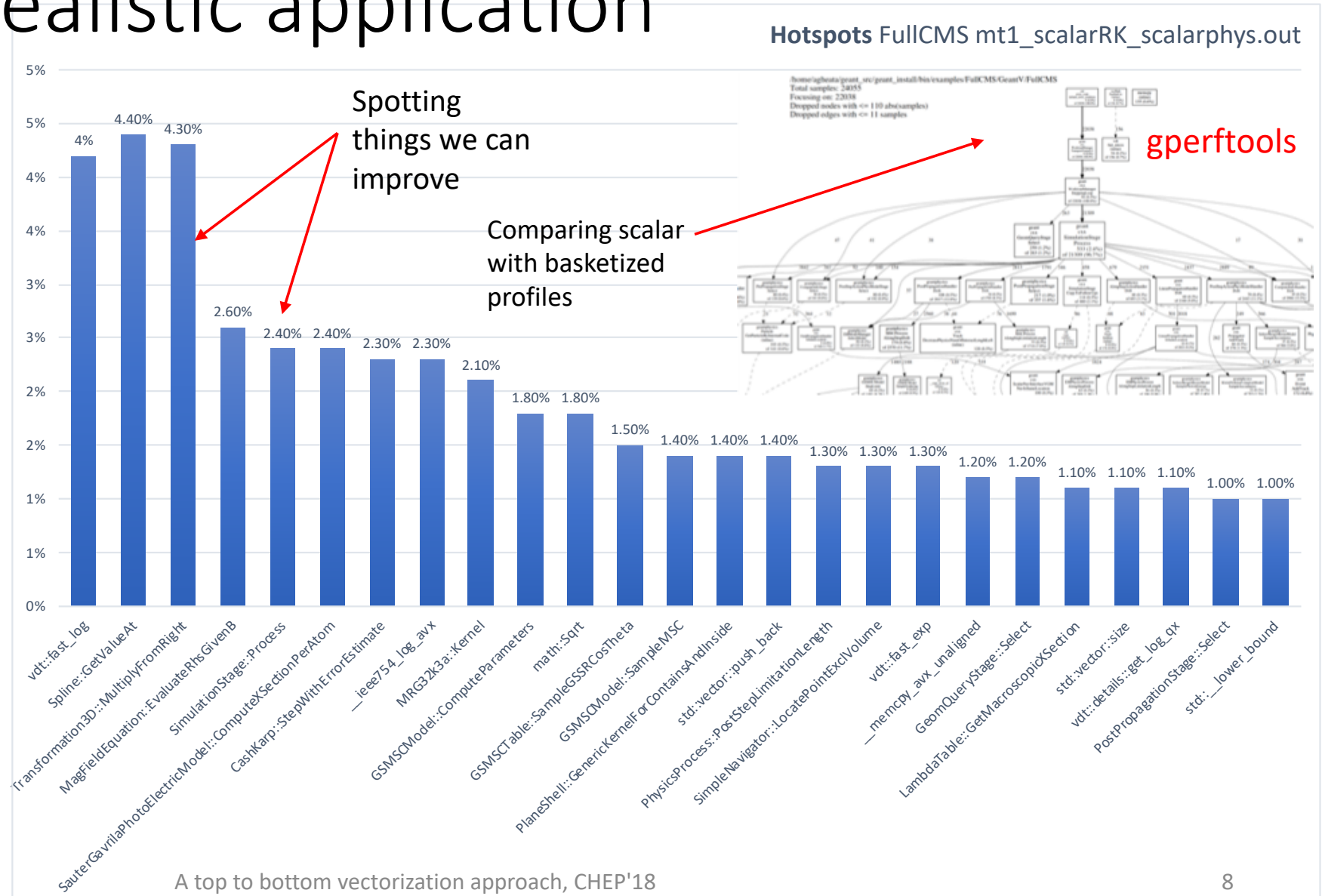
- After having validated a realistic application against Geant4, we need to understand the performance
  - Detailed profiling of the simulation stages
- Many stages containing already vectorized code
  - Field propagation giving expected speed-up in the full flow
  - Not yet the case for geometry and post step action physics stage
- Thorough benchmarking against equivalent Geant4 application ongoing
  - Baseline for future optimizations

Top-level profile of basketized GeantV simulation of CMS

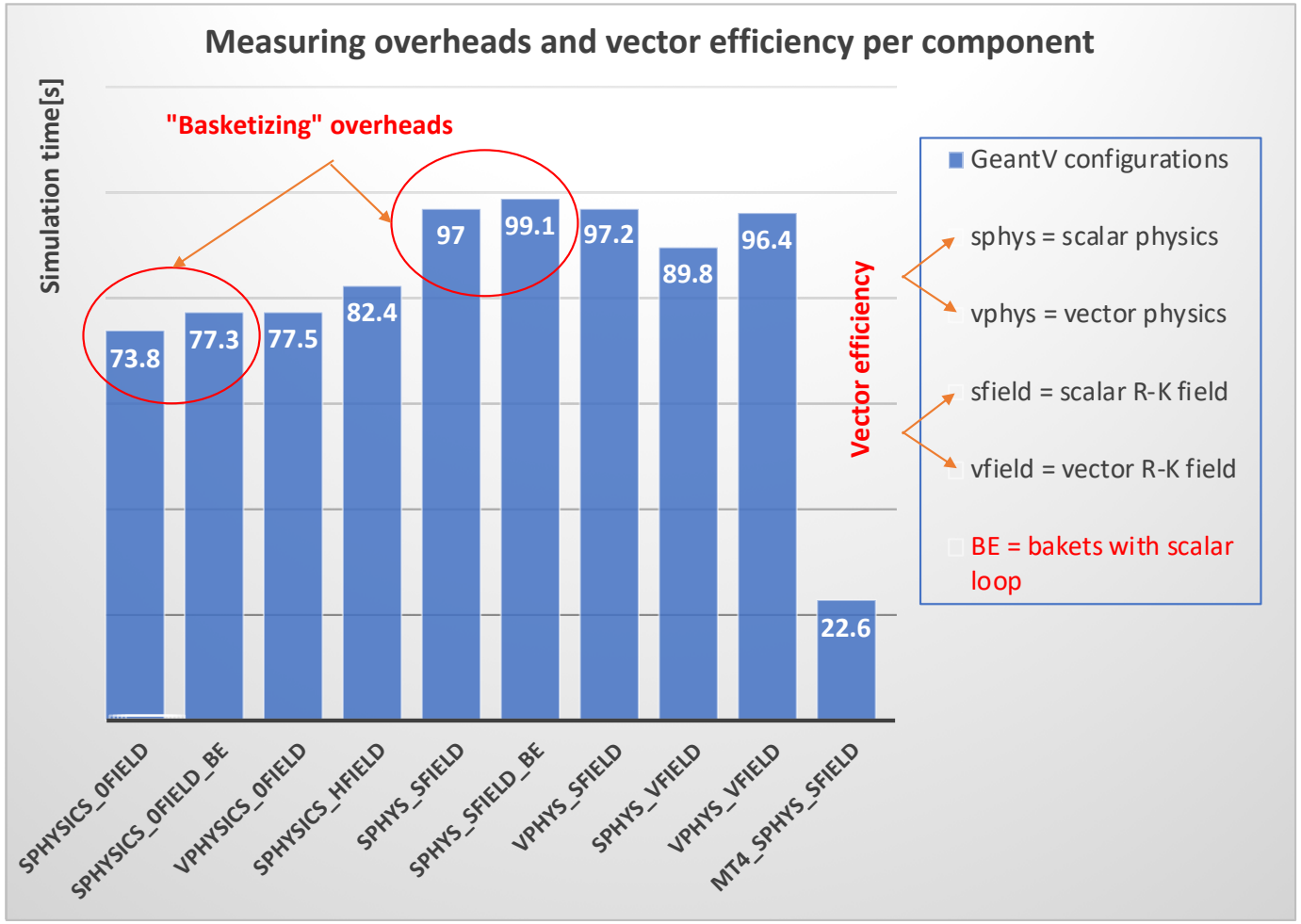


Profiling a realistic application

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

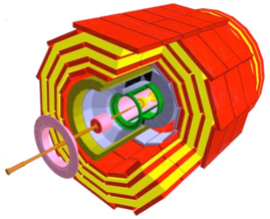


Benchmarking configurations



- Magnetic field ON/OFF, propagator type
  - Understand behavior of different component implementations
- Basketizing ON/OFF (Physics, field, geometry in future)
  - Understand vectorization efficiency
- Basketize but perform scalar loops
  - Understand framework bottlenecks for basketizing
- MT simulation
  - Understand scalability, memory footprint
- Corresponding profiles generated
- Putting in place systematic performance monitoring

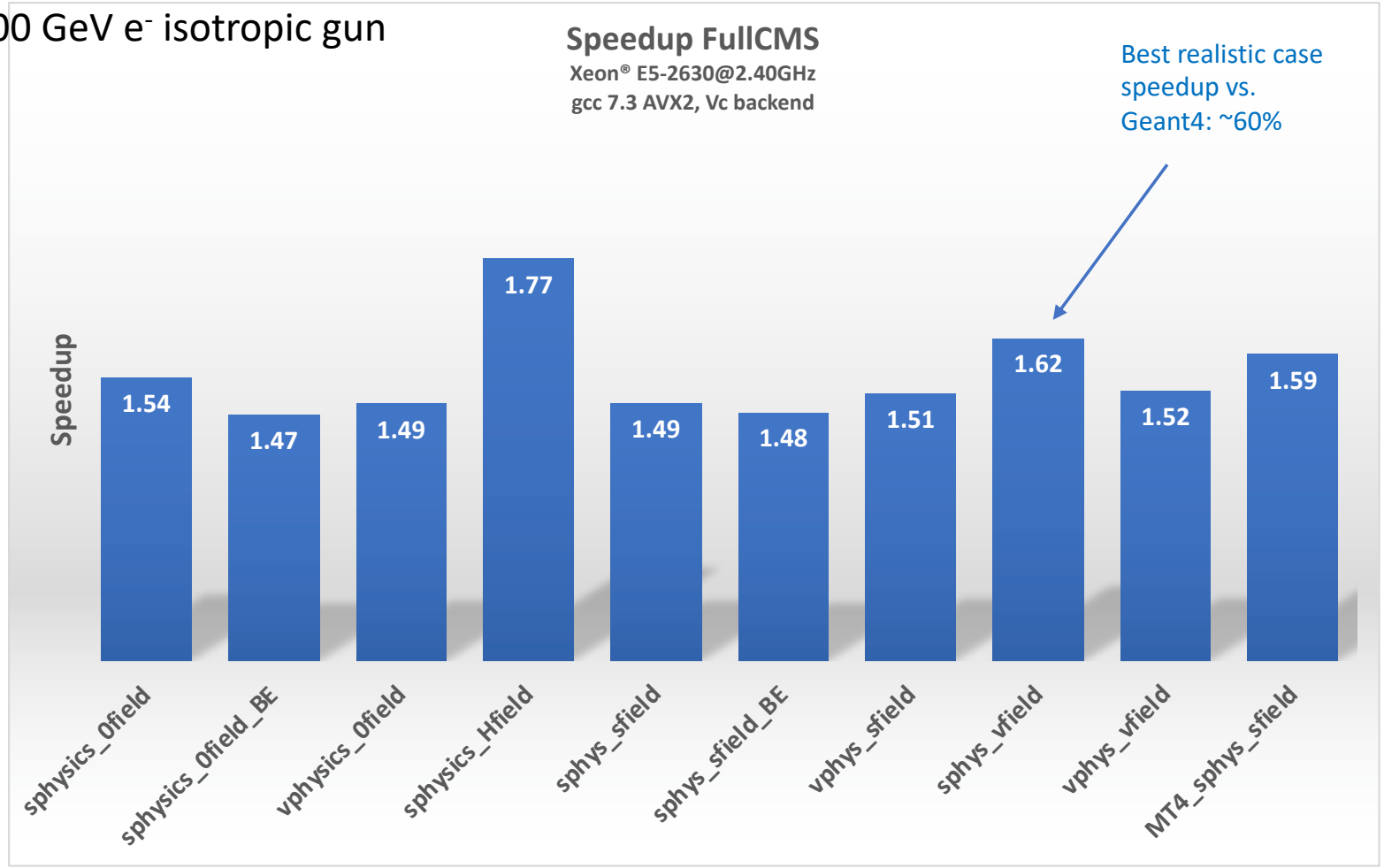
Current generic performance benchmark



CMS geometry, 100 GeV e^- isotropic gun

Geant4 10.4 Rev2

- Generic benchmark with standard EM processes for e^+, e^-, γ, field ON/OFF
- Full geometry imported from GDML, including production cuts
- Scoring:
  - mean energy deposit
  - mean charged and neutral step lengths
  - mean number of steps
  - mean number of secondary particles



Ongoing work

- After understanding the main bottlenecks, now working on:
  - More vectorization work and connection of existing vectorized code
    - Multiple scattering and photoelectric model
  - Reducing basketization overheads
    - Light version (no track exchange among threads)
  - Geometry optimizations
    - Specialized volume navigators (reducing scalar bottlenecks + de-virtualizing + caching)
  - Physics optimizations + new features
    - Tuning usage of sampling methods depending on model and energy/material
    - Use of floating point instead of double case by case
  - Work on pRNG to support reproducibility in MT + basketized mode
- Aiming for 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 in simulation but not only
  - 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 of GeantV by the end of 2018
  - Demonstrator for realistic simulation of EM showers in the context of LHC

GeantV collaborators

- **BARC (India):** S. Behera, A. Bhattacharyya, H. Kumawat, R. Sehgal
- **CERN**: G. Amadio, J. Apostolakis, M. Bandieramonte, R. Brun, F. Carminati, G. Cosmo, A. Gheata, M. Gheata, I. Goulas, V. Ivantchenko, F. Hariri, P.R. Karpinski, G.R. Khattak, D. Konstantinov, P. Mato, P. Mendez, K.Nikolics, M. Novak, W. Pokorski, A. Ribon, O. Shadura, S. Vallecorsa, S. Wenzel
- **CIC-IPN (Mexico)**: J. Martínez Castro, A. Miranda Aguillar
- **FNAL (USA)**: D. Elvira, G. Lima, K. Genser, P. Canal, SY Jung, K Pedro
- **Students 2017-2018**: R. Schmitz, V. Drogan, S. Sharan, E. Orlova, Ananya, D. Savin

Thank you !