Profiling basf2 for fun and profit

Hadrien Grasland

LAL – Orsay

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 654168.

Some context

- France is joining Belle 2 this year, via LAL & IPHC
- I am working on cross-experiment software projects:
  - Thread-safe condition handling in Gaudi
  - ACTS tracking toolkit
- I think ACTS could be useful to Belle 2
  - Genfit lacks maintainers, known to have perf issues
  - ACTS wants to be fast and usable by many experiments
- In 2018, I'll have some time to make it happen

General plan

- Analyze performance of Belle 2 tracking
- Look for hot spots, see how I can help
- Where sensible, study viability of ACTS integration
- Improve ACTS whenever it isn’t ready

Analysing performance

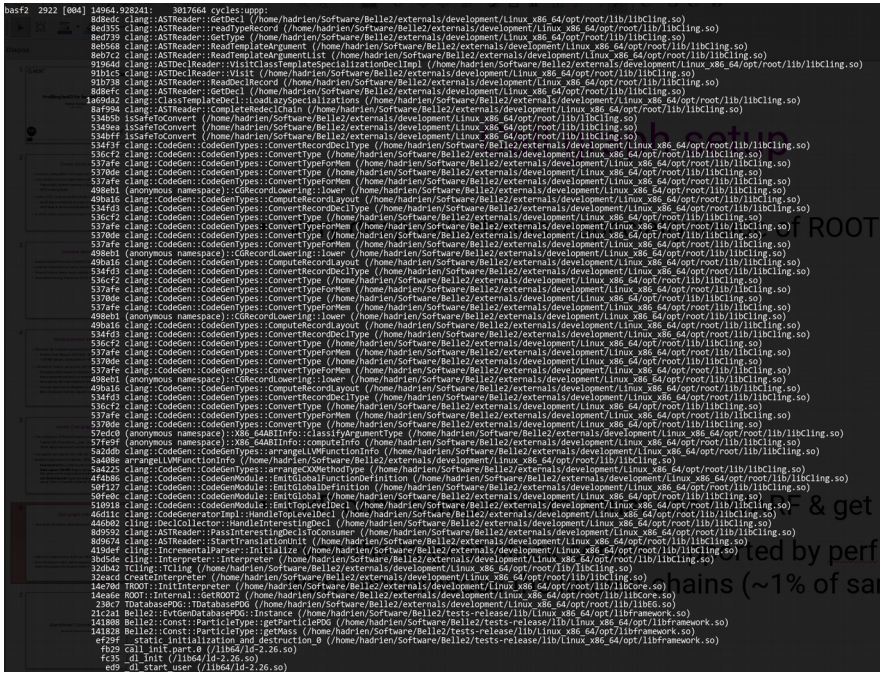
- Nils gave me a simple fitting job to experiment with
  - Perform track fitting on 100 Y(4S)→BB events
  - ~200 MB dataset, running time of ~40s
- I started to study it using Linux perf (aka perf_events)
  - **Sampling** profiler based on hardware performance counters
  - **Native** code execution, very low measurement overhead
  - More **precise** than callgrind on CPU events, **system-wide**
  - **Free** and open-source, integrated in the **Linux** kernel
  - Main drawback: Features depends on CPU + kernel version

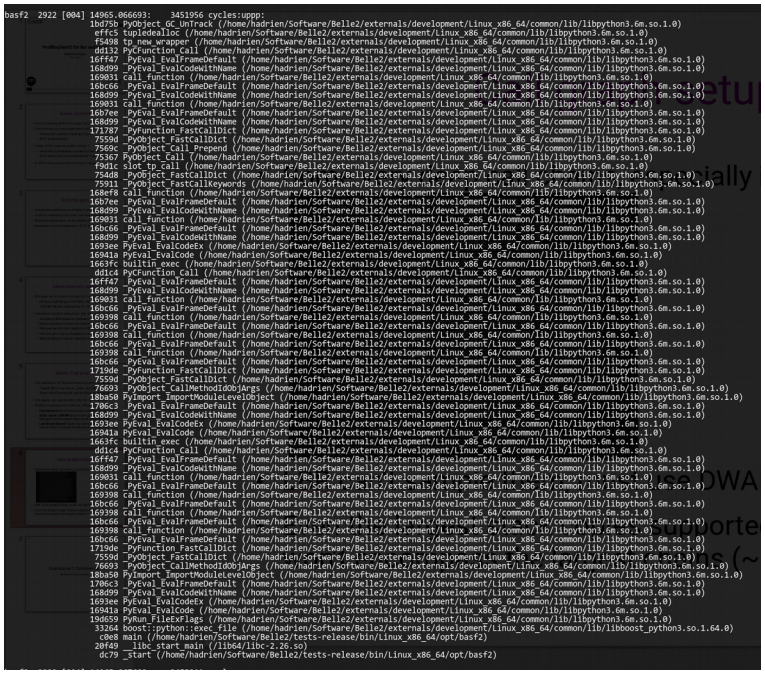
A word about call graphs

- The usefulness of flat profiles is limited
  - I spend 20% of my time in __acos… but I have no idea why
  - Small utility functions are spread around, no big picture
- Call graphs help by explaining who calls what function
- Non-trivial to measure, perf supports several methods:
  - **Frame pointer:** Nice in theory, but compilers broke it…
  - **Stack copies + DWARF:** Easy to set up, but high overhead that grows w/ stack chunk size + some artifacts
  - **Last Branch Record:** Fast & reliable, but need recent kernel (4.1+) and CPU (>=Haswell); short call chains (O(10) frames)

What call graph setup?

- Some very deep stack traces, especially in ROOT & Python





- Too much for LBR, must use DWARF (and get debug info)
  - Even the largest stack copies supported by perf (~64KB) won't span the largest traces (< 2% of stack samples)

Measurement setup

- 1 kHz sampling of 64 KB stack copies means perf must process >64 MB of data per second!
  - Put raw output on tmpfs to reduce IO pressure[1]
  - For larger profiles, RT priority is also an option
- In the end, I used the following profiling setup:

```
perf record
    -a                                  // Measure all system activity
    --call-graph=dwarf,65528            // Use DWARF w/ ~64KB samples
    -F 1000                             // Use 1 kHz sampling rate
    -o /dev/shm/perf.data               // Write output to tmpfs

    basf2 02_fit.py -i simulated.root  // Command under study
```

[1] If you use this trick yourself, keep in mind that 1/it is easy to run out of RAM and 2/all unsaved data will be lost on reboot.

Top-level results

- Need to tell perf report about our crazy stack traces:

```
perf report --max-stack=65535 -i /tmp/perf.data
```

- …and in the end we get this at the top of the profile:
  - 86% of correct stack samples from basf2[2]
    - 98% of that in Belle2::EventProcessor::process
      - 53% in Belle2::EventProcessor::processInitialize
      - 47% in Belle2::EventProcessor::processCore
- That is a suprising lot of initialization for a 40s job…

[2] As for what corrupted the remaining 14% samples, I'm currently studying it with the help of perf experts.

Initialization profile

- Under Belle2::EventProcessor::processInitialize, we find:
  - 95% in Belle2::GeometryModule::initialize
    - 56% G4GeometryManager::CloseGeometry
    - 27% Belle2::GeoMagneticField::create
      - 52% decompression of gzipped text stream
      - 43% string → double conversions (strtod)
    - 9% Belle2::EKLM::GeoEKLMCreator::create
      - 97% in Belle2::EKLM::AlignmentChecker::checkAlignment, itself mostly calling Belle2::EKLM::Polygon2D::* funcs
    - Remaining 8% scattered in <3% funcs, not worth investigating today

Initialization conclusions

- Initialization is not the most pressing concern, but we might get easy performance in smaller jobs by…
  - Understanding why Geant4 geometry initialization is so slow
  - Using a more efficient B field map format than gzipped CSV
- Would **not** bother with EKLM yet, as if we go back to absolute numbers it's only 3% of the total CPU time…

Core processing profile

- Most important part: will scale up with input size!
  - 82% in genfit::DAF::processTrackWithRep
    - **65% in genfit::RKTrackRep::Extrap**
    - 8% in genfit::KalmanFitterRefTrack::fitTrack
    - 6% in genfit::DAF::calcWeights
    - 5% in Belle2::CDCRecoHit::constructMeasurementsOnPlane
    - Remaing 16% spread across many <3% functions…
  - 15% in Belle2::RootInputModule::event
    - 99% is spent in ROOT's TTree::GetEntry
    - Note that the TTreeReader API is preferred these days…
- 20% of absolute time spent doing Runge-Kutta, why?

RKTrackRep::Extrap profile

- 48% in G4Navigator::*
  - Doing various kinds of geometry lookups
- 20% in genfit::RKTrackRep::RKPropagate
  - 64% in genfit::FieldManager::getFieldVal
  - Rest seems to be computations
- 11% in genfit::MaterialEffects::effects
- 7% in genfit::RKTrackRep::calcForwardJacobianAndNoise
- Remaining 14% scattered in <2% functions

Core processing conclusions

- TrackFitter has **very deep** call chains, many small funcs
  - Spent lots of time flattening data for you in these slides :)
  - We probably lose nontrivial CPU just in function calls/rets…
  - May want to investigate inlining, link-time opts, PGO
- Top regions of interest for optimization:
  - Why so much time in Geant4 geometry again?
  - Suspicious time in BField lookups too
  - After that, can study ROOT IO & actual computations…

General conclusions

- Overall, I would say the first performance priorities are…
  - Geant4-based geometry (does it do the right thing, and is it doing it as efficiently as possible?)
  - Magnetic field map (better on-disk format & runtime access)
  - Why the compiler inlines so little of our hot code
- As far as possible collaborations go…
  - I'm working in the same project as the VecGeom team, which promises faster and G4-compatible geometry code
  - ACTS magnetic field is currently receiving lots of attention

Questions? Comments?