Discussion on Algorithm Acceleration

Steven Schramm
October 26, 2023

The panel

- Ed Porter: Einstein Telescope
- Graeme Stewart: HEP Software Foundation
- John Veitch: Einstein Telescope
- Joost VandeVondele: Swiss National Supercomputing Centre (CSCS)

Framing our discussion

- Collectively, we have experience in how to handle the identification and characterisation of signals
  - Different search pipelines: CBC with matched filters, burst signals, stochastic signals, continuous wave signals, etc
  - Parameter estimation with Markov chain Monte Carlo and variants thereof
- It is thus easier to discuss how to evolve such approaches into the future
- In contrast, we do not have as much experience in early / pre-merger alerts
  - This is a field where we do not have any baseline to start from

- **For today, let us exclude pre-merger alerts and associated pipelines from our discussion**
  - This can be followed-up in subsequent discussions, when we have more understanding of what will be done

- **For the panel: does everyone agree with this strategy and framework for the discussion?**

Continuous waves

- In the preceding talks, continuous waves were mentioned as the computing resource driver
  - Challenging due to the need for large and ill-defined time samples (not localised in time)
  - Out of scope for low-latency discovery, as it's not a transient source

- Do we expect any of the methods employed to evolve or otherwise scale towards ET? This evolution could be through the use of GPUs, FPGAs, AI, and/or in some other way.

Matched filtering and template banks

- In the preceding talks, it was pointed out that template banks are not a limiting factor
  - Created once at the start of each run, and take 10-50% of the CPU
  - For clarification: do we expect 2-5x increase in computing costs, or 2000x increase?

- Are template banks expected to retain similar computing requirements, despite the need for higher-order post-Newtonian/post-Minkowski corrections to adapt to ET sensitivity?

Hardware improvements

- Computing hardware is evolving, even for CPUs: x86 looks like it may be replaced by ARM “soon”
  - Right now, GW community primarily relies upon CPUs
- GPUs are also increasingly in use in large data centres (especially HPCs), and FPGAs may in the future

- How do we expect the deployed computing hardware (CPU, GPU, otherwise) to evolve in the next decade, especially at large computing centres?
- How much of this change will be “for free”, and how much will require large adaptations to software?

Software and algorithm improvements

- Together with new hardware, algorithms and software continue to be improved
  - New languages and software environments may help to increase code efficiency
  - The rise of AI/ML is making significant inroads in many areas, including GW

- What improvements can we expect “for free” from the software domain in the next decade, including AI/ML developments and any other relevant areas?

Computing person-power

- Do we need some type of computing specialists in ET to deliver the science program
  - Software infrastructure experts? GPU and/or FPGA experts? Etc

- How can we ensure that we will have the person-effort required to enable the computing model
  - How can we support the necessary careers in computing, as a part of the computing requirements

The scale of the GW computing challenge

- From the previous talks, it seems that we expect resource demands to scale by O(10) or less, within the bounds of our current discussion
- Changes at this scale are not too concerning when discussing future GW computing challenges, especially given that we have also seen similar factors of improvements with GPUs and AI/ML

- Is it correct to say that computing needs are not expected to increase by more than an order of magnitude going from current LVK to the Einstein Telescope?

Other feedback (time permitting)

- Is there any other critical feedback that is important to keep in mind when considering the ET computing requirements and model, which we have not touched on here?