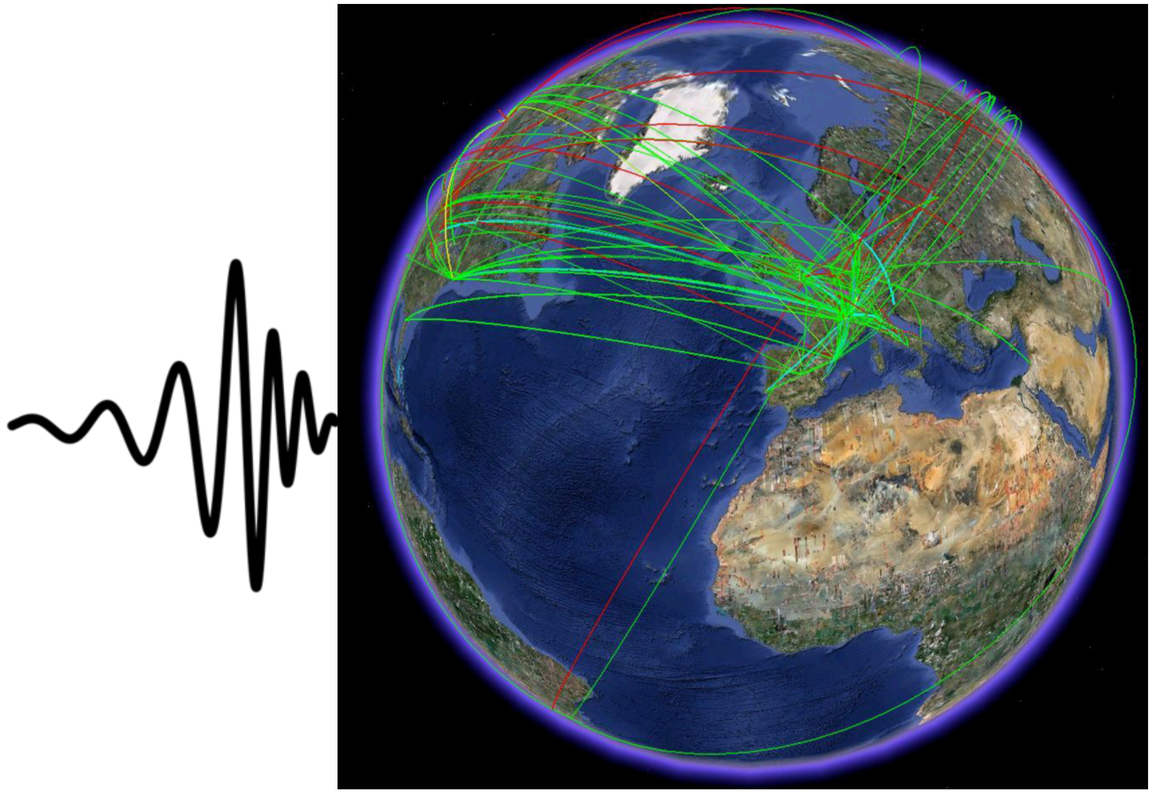
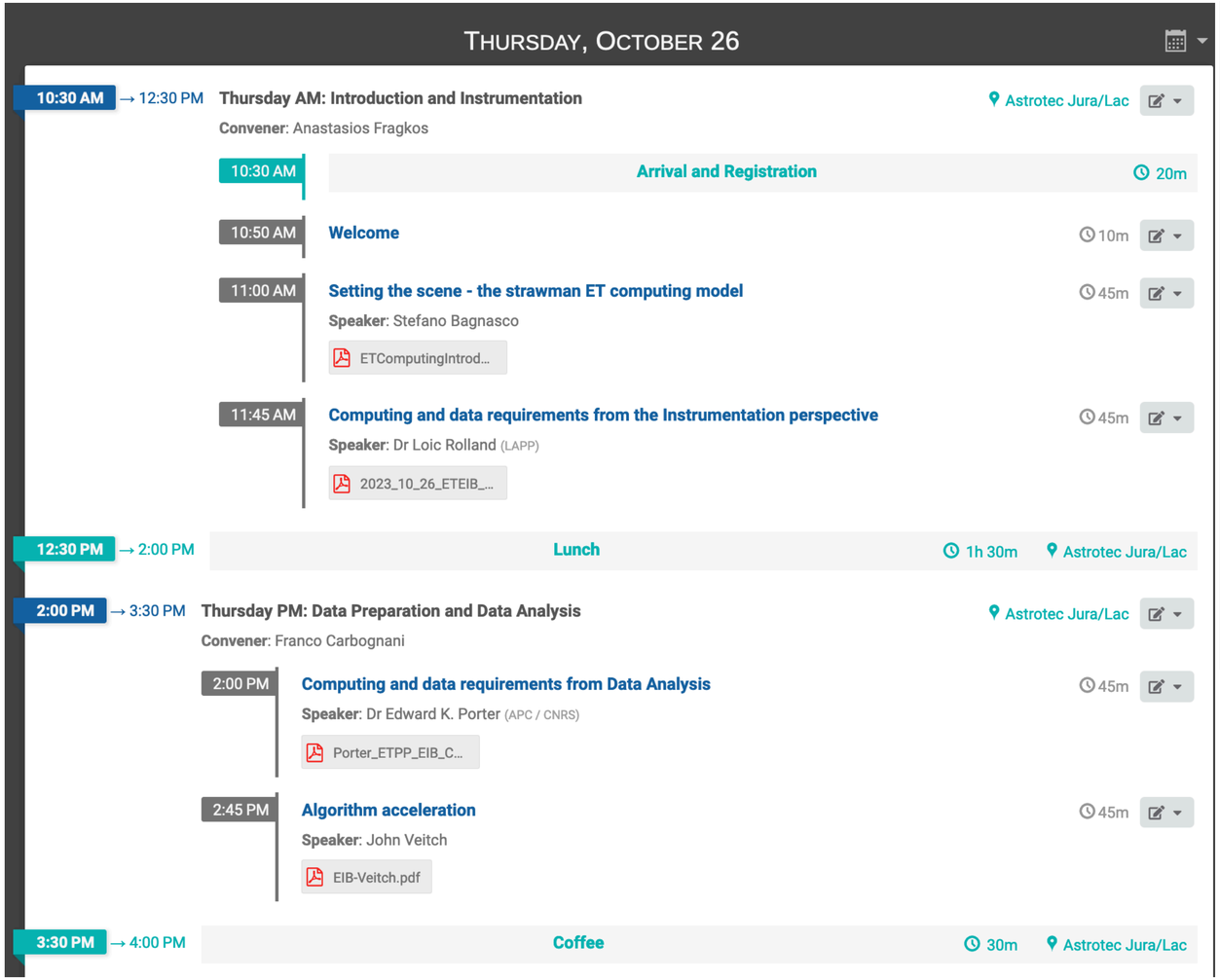
ET Computing and Data Requirements Workshop - Oct 26/27 2023



Workshop overview

- The workshop at Geneva Observatory (Versoix) had several aims:
  - **1st: Produce a draft of the ET computing and data requirements document**
    - This is the first deliverable (D8.1) for ET-PP WP8, deadline end of Feb 2024
    - The first day of the workshop was dedicated to getting feedback from ISB and OSB, attempting to define a reasonable baseline to make estimates, defining our assumptions (and dispelling myths!)
    - ***Huge thanks to the speakers, particularly Loïc Rolland, Ed Porter and John Veitch***
  - Secondary objectives included identifying existing solutions, common challenges with opportunities for collaboration, and unique challenges for ET
    - Provided context for ET computing, and helped identify where the real challenges are (and where ET will need to make significant investment)

Day 1 schedule



THURSDAY, OCTOBER 26

10:30 AM → 12:30 PM **Thursday AM: Introduction and Instrumentation** — Astrotec Jura/Lac

Convener: Anastasios Fragkos

- 10:30 AM — **Arrival and Registration** — 20m
- 10:50 AM — **Welcome** — 10m
- 11:00 AM — **Setting the scene - the strawman ET computing model** — 45m
  - **Speaker**: Stefano Bagnasco
  - ETComputingIntrod...
- 11:45 AM — **Computing and data requirements from the Instrumentation perspective** — 45m
  - **Speaker**: Dr Loic Rolland (LAPP)
  - 2023_10_26_ETEIB_...

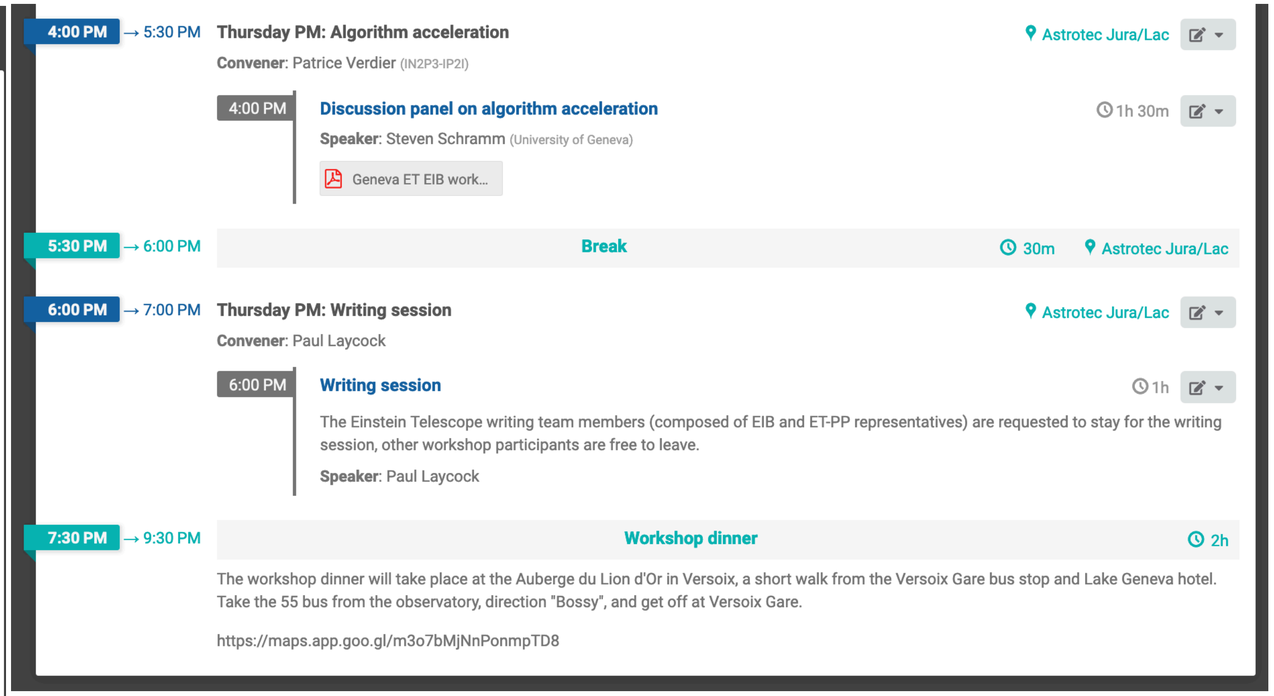
12:30 PM → 2:00 PM — **Lunch** — 1h 30m — Astrotec Jura/Lac

2:00 PM → 3:30 PM **Thursday PM: Data Preparation and Data Analysis** — Astrotec Jura/Lac

Convener: Franco Carbognani

- 2:00 PM — **Computing and data requirements from Data Analysis** — 45m
  - **Speaker**: Dr Edward K. Porter (APC / CNRS)
  - Porter_ETPP_EIB_C...
- 2:45 PM — **Algorithm acceleration** — 45m
  - **Speaker**: John Veitch
  - EIB-Veitch.pdf

3:30 PM → 4:00 PM — **Coffee** — 30m — Astrotec Jura/Lac



4:00 PM → 5:30 PM **Thursday PM: Algorithm acceleration** — Astrotec Jura/Lac

Convener: Patrice Verdier (IN2P3-IP2I)

- 4:00 PM — **Discussion panel on algorithm acceleration** — 1h 30m
  - **Speaker**: Steven Schramm (University of Geneva)
  - Geneva ET EIB work...

5:30 PM → 6:00 PM — **Break** — 30m — Astrotec Jura/Lac

6:00 PM → 7:00 PM **Thursday PM: Writing session** — Astrotec Jura/Lac

Convener: Paul Laycock

- 6:00 PM — **Writing session** — 1h

The Einstein Telescope writing team members (composed of EIB and ET-PP representatives) are requested to stay for the writing session, other workshop participants are free to leave.

Speaker: Paul Laycock

7:30 PM → 9:30 PM — **Workshop dinner** — 2h

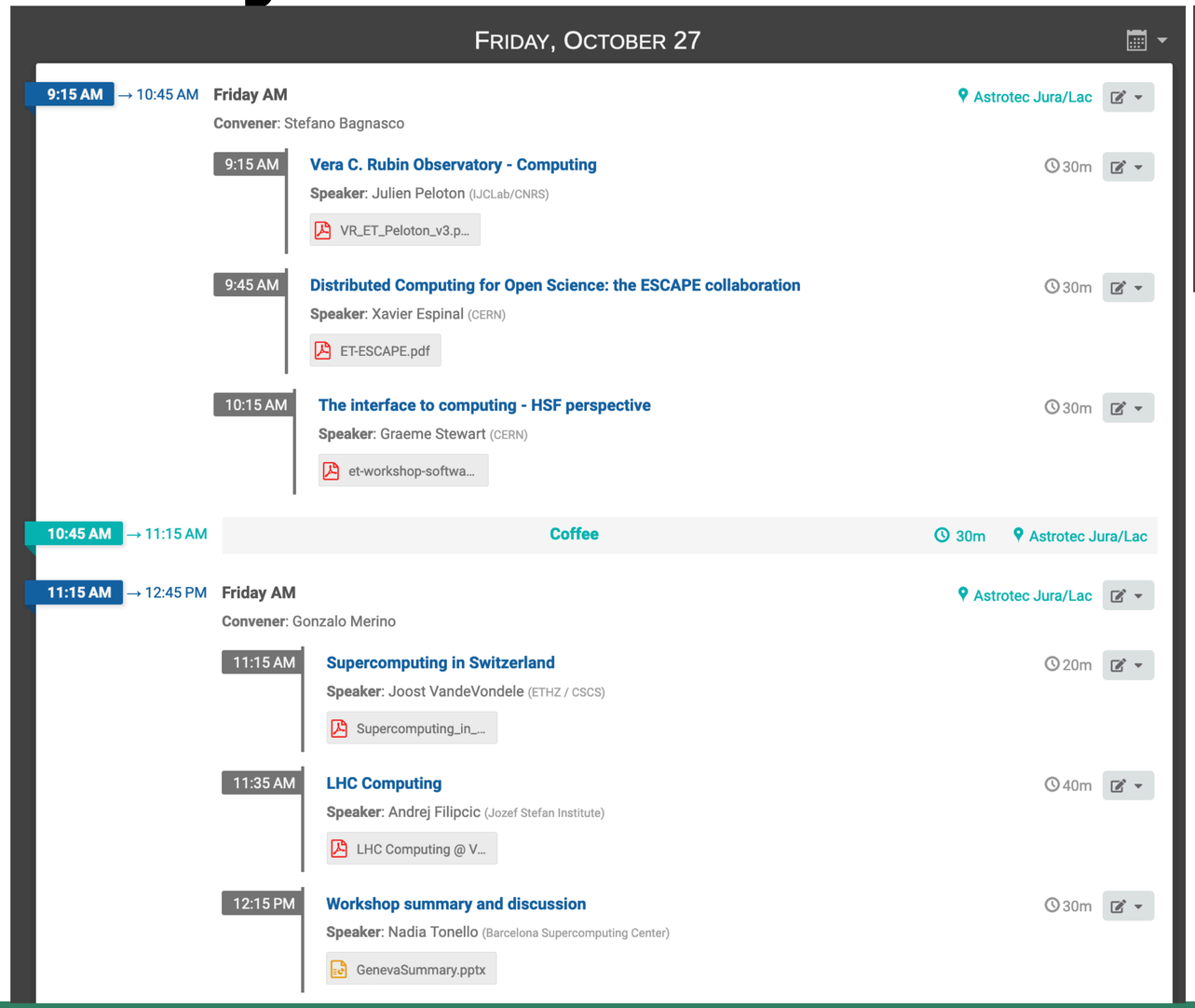
The workshop dinner will take place at the Auberge du Lion d'Or in Versoix, a short walk from the Versoix Gare bus stop and Lake Geneva hotel. Take the 55 bus from the observatory, direction "Bossy", and get off at Versoix Gare.

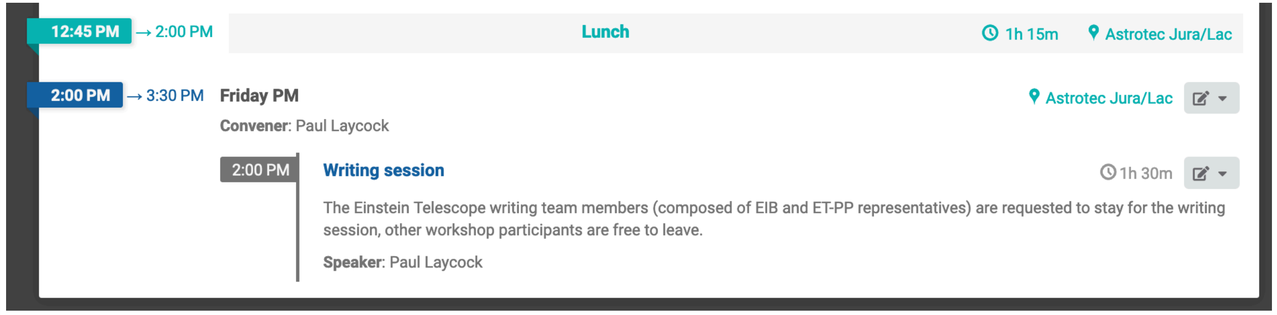
https://maps.app.goo.gl/m3o7bMjNnPonmpTD8

Dedicated to talks from ISB and OSB experts

A dedicated session to discuss how algorithms might be accelerated with new technologies and techniques

Day 2 schedule





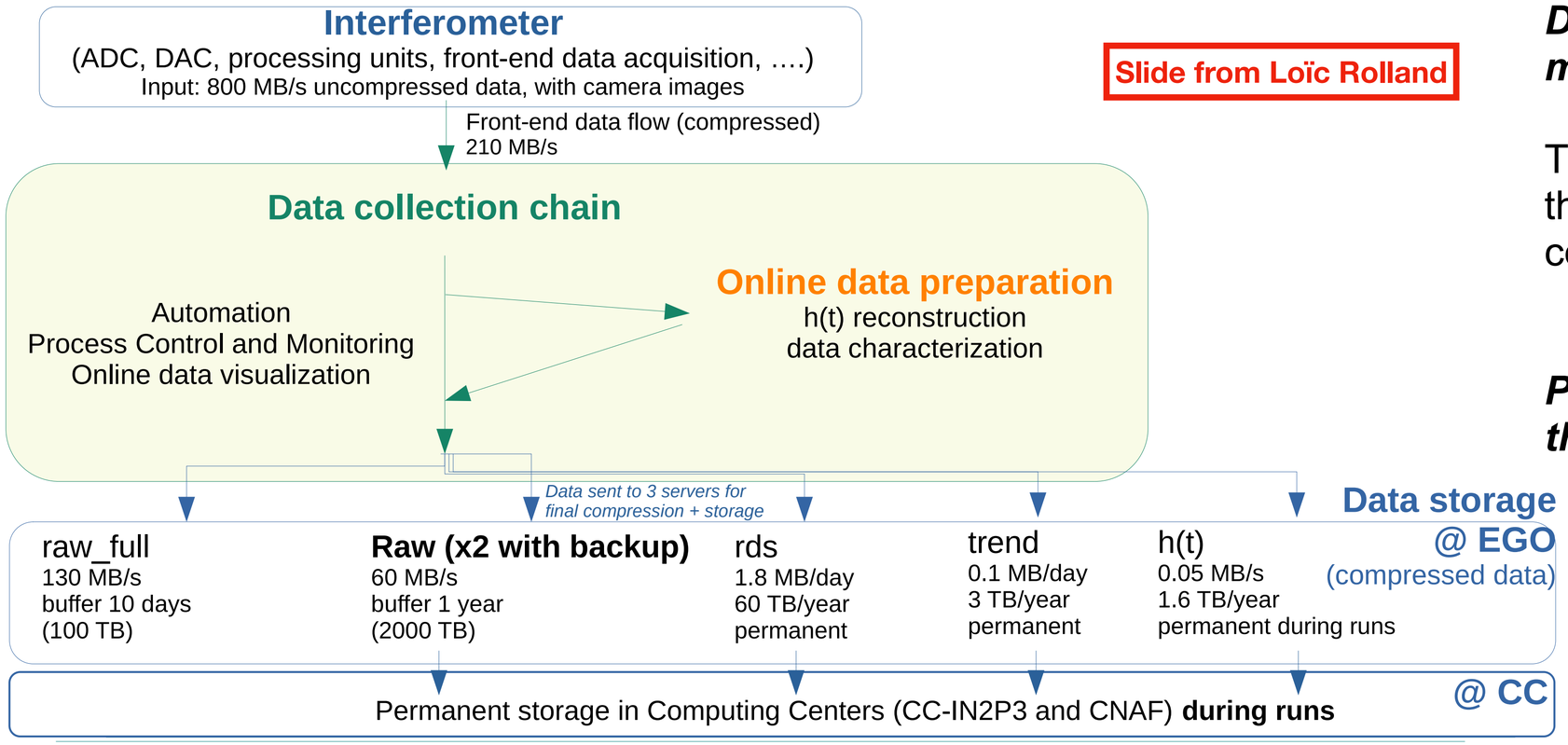
Invited talks from experts external to ET, ranging from the Vera Rubin Observatory, ESCAPE, HSF, LHC, and CSCS

Very useful to understand if there were emerging consensus views on particular topics

Collaboration is key, lots of scope for collaboration with the communities represented by the speakers

Online computing (DAQ, Data preparation)

Virgo data collection: main data flows



Slide from Loïc Rolland

Data volumes are significant but manageable

They do not scale with sensitivity or #events, they scale (modestly) with interferometer complexity and readout frequency

Processing power is modest (tens of cores), though large memory machines help

Online computing challenges

Some concerns expressed during the workshop:

Many home-grown tools exist for Virgo

This created a maintenance burden, common tools would be preferable

There are 31 separate simulation codes

i.e. there is no coherent full detector simulation, aka a “Digital Twin”

Effort is lacking and not CPU (each simulation “runs on a laptop”)

Designing something to handle that complexity likely requires software engineering expertise

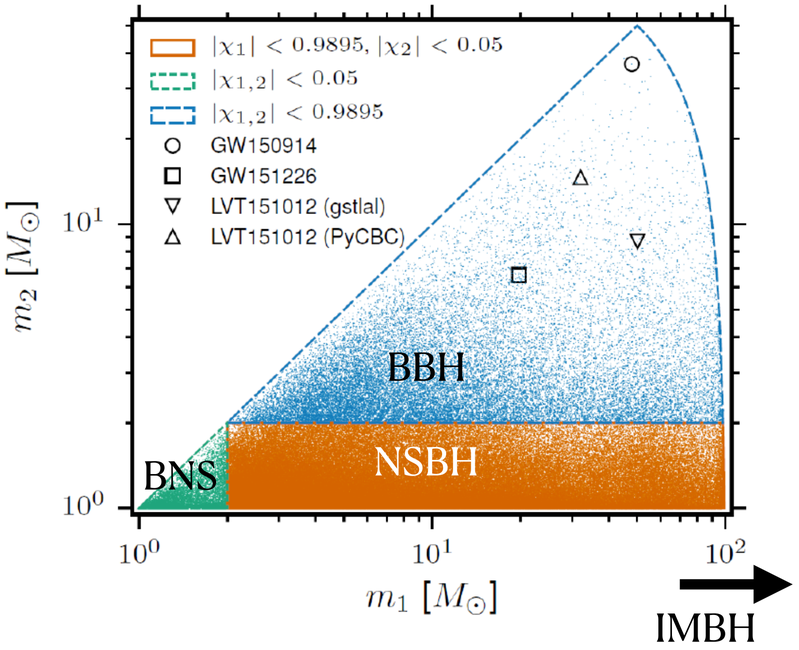
A common detector simulation would be a powerful tool, particularly for debugging operations issues

Offline computing (Analysis)

CBC Modelled Analysis: Template Banks

Slide from Ed Porter

- Today's template banks cover from:
  - ~0.5 solar masses (subsolar mass search)
  - a few hundred solar masses (intemediate mass BHs)
- An ET template bank would increase at the high mass range only
- Result might be x2-x5 times the cost today (more on that later)



Template banks for matched filtering, how do they scale to ET?

The main need for more templates in an ET template bank is at high masses

High mass is less densely populated for the same (optimised) coverage

An estimate of 2-5 increase, and this seems to be confirmed with a very first generation - thanks to Tito !

NOT the final word on this topic, but it does give a sense of how they scale

Offline computing (PE Analysis)

Novel methods

- Explosion in Machine Learning methods in last 5 years
  - CNNs, RNNs, CVAEs, GANs, Normalising flows, diffusion models, …
- Many off-the-shelf techniques work for images or text, but GW applications usually require customisation
- Enabled by and enables GPU computing as a general tool
  - Tensorflow, PyTorch, JAX main toolkits used in GWs so far
  - Python-driven with CUDA/C/Fortran backend
  - Can offers speedups of 1000x for certain problems
  - Other problems can be re-cast into GPU-friendly forms

Slide from John Veitch

Main computation bottleneck for ***parameter estimation*** is ***waveform generation*** and/or ***likelihood calculation*** runtime

This is significant

Speeding these components up is an active area of research

AI/ML looks very promising
GPUs in general look promising

Consensus view - Large general purpose data centres expect more and more GPU usage

Workshop highlights - the ET scale

The canonical statement is that ET will be something like 10% of an LHC experiment
Message - significant but feasible with existing technology *iff we can leverage the best in class solutions*

IGWN provides an excellent starting point for a computing model, the ET strategy will be to improve on this

However !

That computing power usage is very significant, meaning that the efficiency of ET software will be very important

Paradigm change: ET software is not your personal software

Best practice is to make the software open source, allowing among other things the people who wrote it to show off their work to future employers

ET software will be heavily vetted for performance, benefitting from the collective expertise of the collaboration, and using modern software practices. Automated CI/CD pipelines will constantly profile code and massively reduce if not eliminate buggy, inefficient code

Workshop highlights - expertise

The rapidly evolving computing landscape will be the biggest challenge for ET

Apart from onsite, ET will largely use shared computing resources, where ET will typically be a small %

Heterogeneity of computing architectures will be standard (not just CPU, GPU/TPU, hybrid clusters)

Meanwhile efficiency will become more important, from sustainability to demonstrating scientific value for money

All of the above motivates a paradigm change, ET will work *iff we can leverage the best in class solutions*:

ET will need experts in computing infrastructure to adapt to (r)evolutions (across many computing centres)

ET will need experts in software to design frameworks allowing optimisation for different compute architectures

This expertise is highly sought after in industry, experts will demand viable careers if we want to retain them

Deliverable status

Table of Contents

We have a good structure for the document, finalised during the workshop, with good progress on text

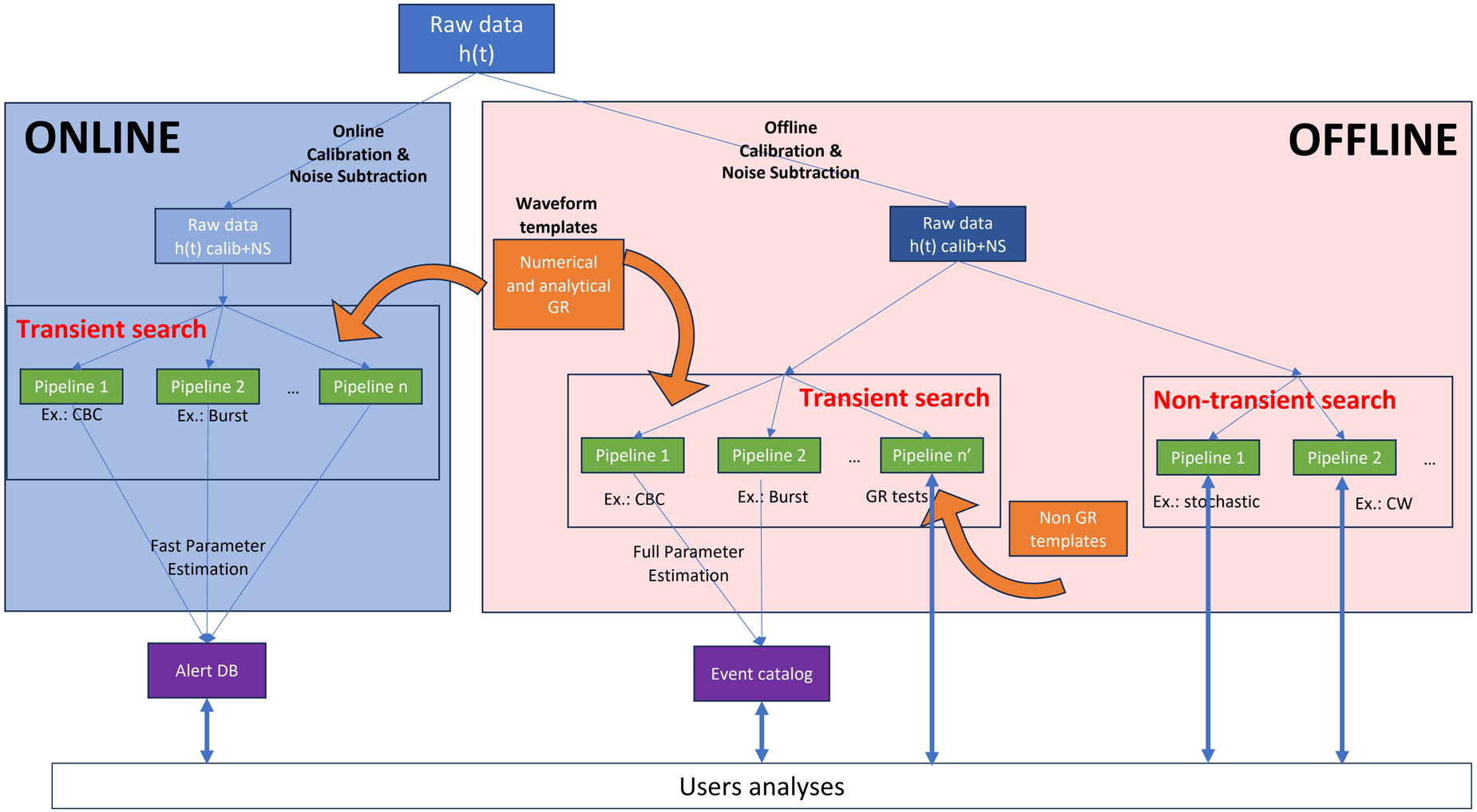
Expect the final document to be similar in length

After introducing the scope, defining the baseline and assumptions we lay out the computing requirements separately for

1. ***Online (onsite)***
2. ***Low latency***
3. ***Offline***

We will have a dedicated section on computing and software expertise, training and retention

Document highlights



A new figure to show the workflows and relationship between online and offline (*Patrice Verdier, discussed with Ed Porter and Viola Sordini*)

Complex, and expect that migration of algorithms between offline and online would be beneficial

This re-emphasises the need for good software engineering

Deliverable summary and timeline

We made a good start at the workshop, most important was to ***define the scope of the ET-PP deliverable D8.1***

This is not the final word on computing requirements for ET

This is a reasonable baseline that allows us to start thinking about an appropriate computing model

The baseline is based on reasonable assumptions, erring on the side of caution

It supports the "10% of an LHC experiment" statement, so we plan accordingly - use existing solutions

Similar-size experiments (e.g. Belle II and DUNE) follow the LHC computing model (WLCG)
ET's starting point is to build on IGWN, which is already WLCG-like, working with e.g. ESCAPE

The deadline for the document is end of February 2024

Aim to release a first draft before Christmas, around the ET-PP F2F meeting in mid-December

Conclusions - workshop outcomes

We made a good start at the workshop, most important was to ***define the scope of the ET-PP deliverable D8.1***

Aim to release a first draft before Christmas, final deadline is end of February 2024

Excellent input from ISB and OSB was and will continue to be absolutely critical !

The conversations have only just started and need to continue

Requirements should come from ISB and OSB, specifications for solutions need to come from Computing

e.g. OSB ask Computing *“what information do you need from the MDC?” (and Computing might need to help OSB gather it)*

Computing can’t define solutions top-down (top-down doesn’t get traction, bottom-up doesn’t scale, ask me about ATLAS !)

In other words, defining a computing model is a process that depends on excellent communication in both directions

Training and retaining computing and software experts will be crucial for the success of ET