Still-in-progress
Mandate for the Software and Computing Team

David Lange

Princeton University

Charge for task force formed to propose organization and mandate of the FCC "Software and Computing" work package within the FCC PED pillar.

To prepare an updated proposal for the organization and the mandate of the ”Software and Computing” work package for the FCC technical and financial feasibility study

1. Mandate and goals of the group for the 2021-2025 period, possibly with final objectives, and a rough idea of milestones and deliverables Structure of the group, optimized to build the community around high- level tasks
2. Ideas or concepts to help more easily engage with possible newcomers and participating institutes
3. How to organize relations with the other Work Packages of the PED pillar (Detector Concepts, Physics Performance, Physics Programme, MDI, EPOL) and with the ECFA PED workshops
4. List of possible co-conveners

Who are we?

- Patrizia Azzi (INFN Padova)
- Tommaso Boccali (INFN Pisa)
- Gloria Corti (CERN)
- Giulio Eulisse (CERN)
- Frank Gaede (DESY)
- Gerardo Ganis (CERN)
- Paolo Giacomelli (INFN Bologna)
- Clément Helsens (CERN)
- Patrick Janot (CERN)
- David Lange (Princeton, Chair)
- Edward Moyse (Univ. Massachusetts)
- Emmanuel Perez (CERN)
- Frank Simon (MPI Physik)

Our process and status

- We did a survey the current state of the software and computing ecosystem (thank you Gerri and Clement for efficiently handling all the requests for talks…)
  - It became clear to us that a solid basis is in place. Many of the remaining gaps were already in the process of being closed as we talked about them.

- We identified topic areas from which to formulate recommendations in terms of goals, needs and connections

- Current state: We have a second draft report that has helped identify remaining issues to be discussed. We hope to be done after one more round of discussion (being an optimist..)

Two-fold feasibility study for the S&C group

1. To establish the needed scale of computing facilities needed to enable the physics program of the FCC-ee

2. To develop and support the software ecosystem and computing infrastructure needed to allow detector and physics experts to complete the studies needed for their aspects of the feasibility study program

Establishing the S&C needs of FCC-ee

- Scale
  - The Z-pole running requirements drive the scale of computing need for FCC-ee
  - The precision goals of FCC-ee suggest multiple ways event reconstruction and Monte Carlo simulation will be needed to address systematic errors. The current “rules of thumb” for scaling based on event rates from LHC needs to be adjusted (eg, multipliers like MC/data ratio need to be reconsidered)

- Cost
  - Exponential cost per performance of computing hardware is expected to continue. But even small uncertainties in exponents diverge over decades
  - One way to formulate is relative to the computing LHC uses today (this also applies to the feasibility of running a computational facility (online or distributed)

Current results suggest this is not insurmountable

Using LHC-scale computing is nearly sufficient (eg, within 10x) for all the simulation needed for the Z-pole run of a FCC-ee detector

	Generation	Simulation	Reconstruction	DELPHES
Computing unit	$3.5–5.2 \cdot 10^{10}$	$2.6–3.9 \cdot 10^{6}$	$5.2–7.8 \cdot 10^{6}$	$2.4–3.6 \cdot 10^{10}$
ATLAS equivalent	$3.5–5.2 \cdot 10^{13}$	$2.6–3.9 \cdot 10^{9}$	$5.2–7.8 \cdot 10^{9}$	$2.4–3.6 \cdot 10^{13}$

Events simulated per day using the equivalent of the ATLAS computing facilities

RAW storage similar to the full HL-LHC

Run	$\sqrt{s}$ (GeV)	Statistics	RAW data
Z	91.2	$3 \cdot 10^{12}$ Z decays (visible)	3–6 EB
WW	160	10^{8} W^+W^- events	0.1–0.2 PB
ZH	240	10^{6} ZH events	1–2 TB
$t\bar{t}$	350, 365	10^{6} $t\bar{t}$ events	1–2 TB

Analysis level data similar to LHC Run 2

Run	$\sqrt{s}$ (GeV)	Statistics	AOD data
Z	91.2	$3 \cdot 10^{12}$ Z decays (visible)	15–30 PB
WW	160	10^{8} W^+W^- events	0.5–1 TB
ZH	240	10^{6} ZH events	5–10 GB
$t\bar{t}$	350, 365	10^{6} $t\bar{t}$ events	5–10 GB

Ganis, Helsens:
https://arxiv.org/abs/2111.10094

Draft Recommendations

- Build upon existing analyses of computational needs to establish feasibility considering both online and offline aspects including trigger-less systems.
  - The committee's belief, based on current information, is that this is relatively straightforward if the scale of LHC facility funding is assumed. However, revised estimates of data sizes or of compute needed to process data should be considered as they become available.

- Establish timeline for gathering and updating inputs and finalizing the assessment of resources needed for the FCC-ee program.
  - There are a number of important parameters to determine using inputs from detector and analysis experts. Work with detector and physics group experts to establish timescales for needed inputs and sufficient context to capture any large sources of uncertainty.

Supporting the broader FCC-ee feasibility assessment

- The software and computing work package has a central role in facilitating many aspects of the FCC-ee feasibility study.

- Detector designers, algorithm developers, and analysts depend on the success of the software and computing infrastructure.

- Priorities differ from those of current LHC software and computing work. Instead of stability and robustness, the feasibility study needs a low barrier of entry and continuous support of multiple detector configurations and their evolution.
  - Correctness, ease-of-use, commonality, traceability, interoperability and agility are all important.
  - Nevertheless, infrastructure, lessons learned, and best practices from the LHC community are important to consider

The Current direction of the FCC software ecosystem is solid

- We found the current approach based on edm4hep, key4hep, and dd4hep to be a solid basis for the feasibility study (and beyond).
  - Leveraging commonalities across future-collider community broadens engagement

- Continuous need to emphasize the bridge from core development to integration and support of FCC detector geometrie and algorithms together with FCC detector experts and analysts.
  - Encourage documentation (from the core outward..), good examples, clear recipes for getting the “latest and greatest” application stack
  - Be welcoming of software contributions / encourage developers to take the last step of integrating their work with the collaboration software
  - A comment beyond our mandate: The enthusiasm to make this a reality needs to be bidirectional..

Recommendations (abbreviated, more words in backup..)

- Identify any missing algorithmic components of the current FCC software ecosystem.
- Work with detector and physics stakeholder groups to establish and evolve the needed scale and schedule for Monte Carlo production required to successfully complete the FCC-ee feasibility study.
- Support and facilitate the software release and distribution for FCC-ee.
- Identify computational and storage resources necessary, and a rough time profile of needs, for Monte Carlo production and analysis needs.
- Ensure that FCC analysis procedures and resources are well documented accessible to FCC analysts

Onboarding and Engagement (abbreviated, more words in backup..)

- Continue emphasis on training
- Ensure documentation and examples are functional and up to date.
- Foster an active user forum for discussion and questions.
- Use surveys to better understand successes and ways to improve
- Participate in HSF and advertise FCC / ECFA events there
- Keep up with industry / open-source practices.
- Look for collaborative opportunities via new projects (and funding)

Connections to FCC PED groups and beyond

Suggested topical areas where strong connections are needed (those not already captured in recommendations)

- Generators: A cross PED generators working group to ensure that needs are communicated and prioritized (especially FCC specific aspects)
- Geometry: Ensure that detector and subdetector components are integrated and documented. Anticipate and prioritize developments for new detector concepts require integration.
- Software Integration: Lead regular software integration fora/meetings to gather needs with detector and physics experts. Work with PED leadership to establish and evolve release schedules and goals.

Connections to FCC PED groups and beyond

- MC Production: Plan (scale of and schedule for) and carry out MC campaigns. Establish regular forum to organize sample requests; prioritization; expectations for availability; etc
- Computing resources: Work with resource providers (eg, CERN, WLCG sites, HPC centers) to enable FCC-ee Monte Carlo production and analysis needs.
- Analysis Tools and Infrastructure: A forum for analysts to discuss their experience with or extensions to the FCC analysis infrastructure.
- Documentation and training: Establish and maintain entry point to FCC-ee software and computing documentation and tutorial materials. Ensure that relevant community training events are advertised to FCC analysts.

Conclusion

- We are not finished as a task force, but we expect to be soon.

- We see that FCCSW is a solid basis for the feasibility study (and beyond), and that practices are evolving to be more general and more sustainable.
  - In many cases, recommendations are essentially to sustain current practice

- No Org chart?
  - We are formulating task-based roles that we feel are important to meet recommendations and to explicitly build/enhance the connections with other groups.

- Please reach out with questions or concerns – especially to help us identify aspects that we have failed to consider..

Backups

Recommendations

- Identify any missing algorithmic components of the current FCC software ecosystem. The committee identified solutions in place or in progress for all high-level components that it discussed. However, specific items (e.g., specific generators, MDI codes, etc) are expected to require continued work.
- Work with detector and physics stakeholder groups to establish and evolve the needed scale and schedule for Monte Carlo production required to successfully complete the FCC-ee feasibility study. This should include estimates of when Monte Carlo samples using a fast simulation approach are insufficient for the analysis to be be performed, requiring a full detector simulation are required.
- Support and facilitate the software release and distribution for FCC-ee. Building on current practices, establish a planning process to ensure that all stakeholder needs are incorporated, and that schedule constraints are clearly communicated and updated.
- Identify computational and storage resources necessary, and a rough time profile of needs, for Monte Carlo production and analysis needs.
- Ensure that FCC analysis procedures and resources are well documented accessible to FCC analysts recognizing that many will be engaged part time and/or only for a limited period of time

Onboarding and Engagement

- Focus on training. Continue efforts on starter kit. Ensure that the FCC-ee community is aware of software and computing skill building opportunities from HSF and elsewhere.
- Ensure documentation and examples are functional and up to date. Growing the continuous integration system is one way to avoid code ”rot” in examples. Establish links to the physics communities to ensure that examples stay current and topical.
- Foster an active User forum for discussion and questions. Encourage developers and analysts to engage with this forum and help to answer questions as they become more experienced.
- Use surveys as major milestones for algorithm development or analysis results are reached to better understand project components that are most essential, face difficulties, were not sufficiently robust, or lack sufficient documentation and/or tests.
- Participate in HSF (one FCC-ee expert now co-leads the HSF software frameworks working group). Ensure that software and computing related events are on the HSF community calendar and advertised as appropriate.
- Keep up with industry. This avoids duplicating effort, and enables attracting / retaining
- Consider what opportunities are available when partnering with FCC-ee collaborators (and others) to build new projects (and funding). It is often easier to bridge gaps to other communities (eg, generators, MDI experts, etc) via joint projects

Connections to FCC PED groups and beyond

The recommendations have identified a number of important connections already. Here I discuss a few suggestions areas where a dedicated forum is suggested (which can mean a meeting, but can also be more informal):

- Generators: A cross PED generators working group would be an effective mechanism to ensure that needs communicated effectively and prioritized properly (especially for FCC specific aspects)
- Geometry: Ensure that detector and subdetector components are integrated into FCC software releases and included in documentation and continuous integration tests. Anticipate and prioritize developments required in cases when a new detector of subdetector concept is not supported by existing infrastructure.
- Software Integration: Lead regular software integration fora/meetings to gather needs with detector and physics experts. Work with PED leadership to establish and evolve release schedules and goals.

Connections to FCC PED groups and beyond

- MC Production: Plan (scale of and schedule for) and carry out Monte Carlo production campaigns. Establish regular forum to establish and evolve sample requests; prioritization; expectations for availability; etc
- Computing resources: Work with resource providers, including CERN, WLCG grid sites involved in FCC-ee research, and HPC centers to enable FCC-ee Monte Carlo production and analysis needs.
- Analysis Tools and Infrastructure: A forum for analysts to discuss experience with or extensions to the FCC analysis infrastructure.
- Documentation and training: Establish and maintain entry point to FCC-ee software and computing documentation and tutorial materials. Ensure that relevant community training events are advertised to FCC analysts.