

Breakout 3: Long Term Futures: Notes from room 2 (David Halstead, Frank Wurthwein)

1) LHC: in light of Moore's law having slowed down, lengthening the time for hardware to become "half price" by factor ~3-6, biggest R&D item in s&c for next 5 years is cost-containment for computing 10-20 years from now.

Impedance mis-match between Large Facilities Propose/design/build/use timescale (decades) and CI award lifespan (3-5 years), but there are opportunities:

CI providers/facilitators: National SuperComputer centers (and XSEDE) have exceptional State-of-the-art visibility and historical perspective, but there will always be a challenge projecting 5+ years ahead on architectures/costs. They can help to avoid lock-in risk.

data science curriculum should include data from the big facilities
data scientists as a profession must be a viable career path
minimal data provenance to allow data reuse of data from large facilities

LSST: how will scientists analyze the PB catalogues to extract science in 20 years?
Will they be subsetting heavily, to extract small samples to take home for analysis?
Will we provide compute power centrally? What will be our business model?

2) 5 years from now we will have HEP scale experimental facilities in other sciences.
How will students, post-docs, faculty at campuses throughout the US be able to access the data from the facility? Will they extract data to bring home? Will there be infrastructure at campuses to receive that data, and support its processing? Will we be able to train students across the community to do science at that scale?

Campus organizations are too inward looking. Ill prepared for inter-campus collaborations among PIs from different institutions. NSF could, and maybe should play a role in instigating a change here. Preparing campuses for a future where intercampus collaboration is the norm across an ever widening range of science.

LIGO: 5 years from now the change of computing will be about x6.
Absolutely critical to buy CI as late as possible.
CI needs to scale fast late. Partner with SP/XSEDE on Dev/Test evaluation and CI specification?
Distributed networks of observatories that want to be analyzed together, and collected reliably.
Moore's law will be just fine for LIGO.

Use the next 5 years to build (inter)national relationships required for the kind of facilities that will be operational in 20 years.

NSF awards are normally 3 to 5 year awards. CI providers and organizations that manage large facilities struggle with this policy, because there's a risk NSF will not fund the ongoing operations and maintenance of the facility.

CI providers should apply NSF funds to form a community of partners to grow the infrastructure through shared investments. The NSF IRNC program provides excellent examples of NSF funds being matched by international partners. In some cases there was a 10 to 1 match for every dollar. That level of investment can be applied to develop long-term agreements with providers that can sustain the community in the event NSF funding does not persist. For example, the contracts the AmLight project has negotiated has provided additional network capacity to the community. Capacity has at the minimum doubled every 5 years, and we have been able to do even more when new technology, such as 100Gbps, became available.

3) What steps does the community need to take:

Within the next 20 years, will there be new communities that will be using the data of the NSF facilities? Communities we may not yet anticipate?

Or are we talking about the same communities that we have today? Those are presumably well defined.

Foster Industry partnerships, commercialization

Realistic lifecycle resource planning (design, construction, maintenance) for Software based on clear specifications.

Concern over programming language and scalable algorithmic problem description paradigm adoption and lifecycle (fortran, C, C++, Java, Python, MatLab etc), SIMD, vector, asynchronous execution, MPI, Open MP, Cuda.

Tool reusability: move from CS research project to reuse

Constant personnel challenge: what skill-sets? DO you train within, or hire "off-the-shelf". How to retain/ or ease replacement (design/process documentation and compartmentalization)

Need mutually beneficial standards.

4) What steps does the NSF have to take?

If we want to engage more the general public then we need to be funded to do so.

Outreach etc. tends to be often an unfunded mandate.

Consistency across agencies would be nice, e.g. for Data Management plans.

Within 20 years you will have significant technical transitions.

Example: non-volatile memory to increase write speeds.

Hierarchical memory/storage infrastructures.

Vectorization becoming crucial for even single threaded applications.