

Scientific Sub-Council Meeting 11/24/08

Rough Notes 11/24/08, by Les Cottrell

Attendees: Gunther Haller, Steffen Luitz, Richard Dubois, Greg Dubois-Felsman, Stuart Marshall, Cho-Kuen Ng, Clemens Wermelskirchen, Sebastian Boutet, John Arthur, Tom Devereux

Not Present: Richard Mount, Sayed Rokni, Michael Soltis,

Persis attended for 20 mins at the start

Persis ~~opened up by stating that~~: Steffen is responsible for computing as acting head of SCCS. The sub-council is to assist Steffen. Requirements of the sub-council are driven by the lab requirements.

- The first challenge is to find: what are the computing requirements for the next few years (mid-term)
- The second challenge is what is the business model for SCCS to give user groups what they need. SCCS is the most complicated budget of the Lab. What is the right financial model in this new lab? Only one option for example, say SCCS provides a minimum common service level, then if one does something special then pay for it oneself. May copy the business model from another Lab.

This could be the most critical sub-council. Need to help line management make decisions. We have not made transition to where we need to be in computing. What is the Vision what are the requirements to meet the vision?

There was a question and answer session. The following are rough notes:

Compare whether or not everyone (or a majority) needs something versus only some need it. There is a grey area where some people do it themselves. Listen carefully to one another. Come up with common services. This leverages money for science.

Q: How are things partitioned? This sub-council will make recommendation and present to Sandy/Persis and then decide how to fund.

Persis: Some Labs do their computing model well, find out how they do it what is their business model?

Q: How does this differ from the public council of IT?

Persis: They are more concerned with business computing. The IT council will probably address common infrastructure such as desktops and networking. We need to make a list of what falls in which sub-council.

Q: Some groups have brought in standard stuff which is easy to manage. The non-standard stuff to take data/control is different. Standard stuff goes to IT sub-council.

Other is it to this committee. Is there some commonality? Can some of this be supported centrally, e.g. MCC/LCLS use central networking.

Persis left.

Might do easier stuff first, e.g. scientific computing, networking, rather than controls.

Computing for LCLS has not been fully estimated or costed. The budget was eliminated from the project to be funded later from operating. Now challenge is to figure out what is needed for LCLS and then find funding. This is part of this sub-council's business. Decide capacity required gather, put together and present for PULSE, LCLS. It is not in anyone's budget at this time.

Gunter showed a list of services.

Formatted: Heading 3 Char

We need to turn the list of services into a matrix of program areas and what services are [provided](#). Discussion on desktop support, Macs at Kavli not part of common SLAC infrastructure. Should it be part of standard infrastructure? Postdocs will not come to work at SLAC if it is too straight jacket (e.g. only 1 OS supported), will go elsewhere. Need an open dialogue. Security model somewhat drives the supportable infrastructure. Will be more constricted than an open university. But there is room for balance. Tom D. brought in unique kernel mods at the university, they were done by individuals, not centrally, they needed full root access. There was not much security overview.

Gunter will send round the services list to solicit input from each area so one can decide what is common, could be supported centrally, and to clarify clearly what each service means and the service level agreements.

Laundry list of Services (from Stuart):

- Floor space, power, cooling, physical installation
- Network infrastructure
- OS Install & updates
- Monitoring, detection
- Hardware repair
- Authentication, authorization,
- Group management
- Net file system, security infrastructure, backup/archive systems.
- Job scheduling & management
- Process, network statistics
- Email services (this is GP)
- Compilers
- Debugging facilities
- MPI software & batch interface
- Automation of sysadmin tasks
- Collaboration tasks.

Given a list like this how do we assign to sub-councils. How much of this [is](#) centralized and shared, what does one do by group? We will need a second level description giving a detailed description and service level so everybody understands it to be the same.

People agreed to come up with a list of services needed for their area and then it will be reviewed.

Meet again one more time before Xmas.

Would be interesting to learn from ANL and LBL (visit or invite someone from there). The other Lab needs to be similar size as SLAC and support central computing. Have to be able to say we looked at other Labs. Can we learn a lot from org charts? How much do names of groups have unique meanings. Need a leader to head up effort to explore this (how other labs do things).

Stuart will recruit people to assist in learning how other Labs operate. Gregory, Tom and Richard will provide information on what they learnt at a recent meeting.

LUSI Data intro (Sebastian)

Data rate gated by 120 images/sec (pulse rep rate), detector size = 1.2 Mpixel. A frame is 132MB/s (750px²) to 1.3GB/s (2300px²). One experiment has between 10⁵ and 10⁷ frames 110GB (750px² 10⁵ frames) to 110TB (2300px², 10⁷ frames). This is for one of the more intensive experiments (CXI). Initially LCLS will be similar to today, i.e. run for week, walk away with hard disk (couple hundred GBytes/week). [This is true but if the experiments are successful, we would move to a new regime within 1 week. The initial phase may not last more than a few days.](#)

For CXI future, need to decide if it is a good image on the fly (filtering). Then realtime pixel correction using calibration constants, followed by processing of corrected image (e.g. imaging of single molecules, combine 10⁵ – 10⁷ images into a 3D dataset.) Learn/pattern recognition/classify/sort images, e.g. 1000 bins (this is a very tricky bit). Average bins, alignment and reconstruction. ~5Hz real-time display of 3D images (1kx1kx1k).

Brute force method: Compare each image to one another. Need scalar product of every pair 120 images/s, 4Mpixels -> 5x10¹⁰ operations per second. Total data set: 10⁶ images -> 10¹⁸ operations, 15mins on world's fastest computer, many years on desktop. If transfer data to other sites to analyze then need high speed network and time to transfer data. Currently the algorithm is being developed by simulation. There are probably better algorithms such as adding compression.

The algorithm may not scale with size. May have very intensive IO requirements. Will one need a high class computing center. How will it get funded. [The expectation right now is that users will work out the problem. This issue of concern is that SLAC needs to figure out a way to work it out because leaving it to the users may cause an problem for SLAC when the data goes unused. Expectation is that SLAC will work out the kinks of the algorithm and processes.](#) If problem is not solved then data will sit idle. Could be a very slow start while going through the R&D. Note the facility will get a lot more use if one can extract the information/results quickly and early on. Not everyone agrees the problem is solvable in the near future. Can one get funding (e.g. from SCIDAC) to tackle the challenges.

CXI is one of the experiments with high data needs. The initial experiments will probably be simpler and lower data rates. By 2015 when all 6 beam-lines are running expect an even distribution of experiments.

LCLS does have a review proposal council. Expect to have users on the system within weeks of turn on.

For SSRL there was a dedicated set of users who developed simplified packages. Later things got standardized and then the field exploded.

Babar has demonstrated ability to send several TBytes/day to remote sites. [This has proven to be much more efficient than manually send physical media \(DVDs, disks, tapes etc.\).](#)

May need a review panel for LCLS computing to figure out how to understand/develop the data gathering/analysis etc. with results presented to funding agencies with the hope of getting funding to address the data gathering, analysis. Now is a good time to do this given that the construction phase is coming to a finish and data gathering analysis is about to take off. Handling of large dataset, analysis etc. is supposed to be a core competency of this Lab. We need to mobilize our resources while they are still here.

It was discussed that it is important to figure out what algorithms are needed and to perform simulations now over the next two years to have a chance of being ready when data is collected. Need to be success oriented and expect a well performing LCLS. Lab can't afford that users collect data and then take years to produce science. That would be bad and considering that XFEL has already started the computing effort even so XFEL is scheduled to turn on 5 years after LCLS, SLAC needs to get started. SLAC needs to be able to offer computing services to users so e.g. CXI experiment users can produce science and publications as soon as reasonable after data taking. That users take the large volume set of data off site and figure out at each university/lab on how to analyze it on any reasonable time-scale is not realistic plus not efficient. SLAC needs to step up and since this is a core competency of the lab there is a good match.

Action Items

Meet again one more time before Xmas.

Gunther will send round the services list to solicit input from each area so one can decide what is common, could be supported centrally, and to clarify clearly what each service means and the service level agreements. We will develop a matrix of service vs group from this.

Stuart will recruit people to assist in learning how other Labs operate. Gregory, Tom and Richard will provide information on what they learnt at a recent meeting.

| PS: Gunteher will also send out the process on how to proceed from where we are now to at least answer the two questions Persis asked in the beginning.