

GPU computing at SLAC: 2013-11-22 2pm

From: Deborah Joanne Bard [<mailto:djbard@slac.stanford.edu>]

Sent: Friday, November 22, 2013 7:18 PM

To: Todd Martinez

Cc: Marshall, Stuart L.; Adesanya, Adeyemi; Abel, Tom; Kaehler, Ralf; Brian Moritz; Thomas Peter Devereaux; Eriksson, Thomas; Voss, Johannes; O'Grady, Christopher P.; Perazzo, Amedeo; Gaponenko, Igor; Radmer, Randall James; Ho, Ling Cherd; Hoeche, Stefan; Dubois, Richard; Boehnlein, Amber; Luitz, Steffen; Gruber, Shirley; Yan, Jun; van den Bedem, Henry

Subject: Re: GPU computing at SLAC: meeting 22nd Nov 2pm

Hi All,

the minutes of today's meeting are below. Please let me know if you have any comments or suggestions!

cheers,

Debbie

Minutes

attending: Tom Abel, Yemi Adesanya, Debbie Bard, Stefan Hoeche, Igor Gaponenko, Stuart Marshall.

Summary: There is demand for a small central GPU resource, where users can practice and develop code and run small-scale analyses. It is likely that there is not demand for a larger facility. It is also likely that the affordable option of an interactive/batch server using desktop GPUs would be sufficient to meet the needs of users, but there are drawbacks to such a scheme. We will prepare a proposal outlining the use cases for such a resource, with input from groups across slac.

Here's an outline of our discussions:

- How are people using GPUs at SLAC right now?

Kipac: the KIPAC GPU cluster is used mostly by individuals using GPUs for analysis work on small/medium scales. Visualisation group has their own GPUs.

LCLS: Three GPUs are currently used by individual users with a specific code base.

- Do existing GPU users have an interest in using central GPU resources?

What kind of demand is there (among existing and potential GPU users) for this kind of resource?

Kipac: There is some interest in further small-scale use of GPUs for processing/fitting data.

LCLS: Users in LCLS do not have full access to SLAC resources, so they are unlikely to use central resources. Only pre=processing, filtering etc is done at SLAC, otherwise the user typically takes data back to the home

institution for analysis. However, a configurable workflow is being developed which may include GPU modules.

Theory: Potential to use GPUs for event generation. This is a large code that would have to be re-written for the GPU, as yet it's unclear whether this is worth it in terms of possible speed-up, and in terms of end user resources. A central GPU resource would help scope out the possible gains.

- What kind of resources are useful to people now? What would be useful to the wider SLAC community?

Yemi: There's already demand from groups (Kipac, theory, SIMES) for resources to try out GPU implementation without investing in the hardware themselves.

Igor: there are plenty of libraries/tools for GPUs, perhaps we could bring some in and set them up.

- What problems do we need to be aware of (e.g. over-subscribed interactive nodes)?

Yemi: New LSF protocol manages access to GPU slots in exclusive process mode, so a clash of resources is avoided. However, this won't work for interactive nodes and is not an efficient use of CPU/GPU.

Tom: It may be better to invest in consumer-grade GPUs that don't have the capacity/precision, but are sufficient for user needs. 1/10th price!

Igor: Have to be careful about the precision, and whether memory will be sufficient. Need a good, fast connection to where the data is stored!

Stuart: Can't fit consumer GPUs into a regular server farm. Need to consider options there.

Igor: remember NVidia/cuda is not the only game in town! OpenCL/ATI GPUs may be a good option.