Thoughts on SciDAC 2.0:
Opportunities for Improvement & Features to Preserve

SciDAC Center for Component Technology for Terascale Simulation Software (CCTTSS)
`http://www.cca-forum.org/scidac/`
Rob Armstrong, Sandia National Laboratories, Lead PI
rob@sandia.gov

File Generated: November 24, 2004 at 6:10 p.m.

With the end of "SciDAC 1.0" approaching, members of the Center for Component Technology for Terascale Simulation Software (CCTTSS), which leads the development of the Common Component Architecture (CCA), have been discussing the future of both the project and the SciDAC program. Based on our experience within the program, we offer a distillation of our observations and some suggestions for the future, in the hope of helping the DOE Office of Science plan a follow-on "SciDAC 2.0" that will be even more successful than the current version.

We have tried to be brief, but would be happy to expand on any of the items here and to participate in further discussions about SciDAC 2.0.

1 Preserving and Strengthening Things that Worked

The original SciDAC did a lot of new things that we would like to see preserved in any follow-on program. We list them below in no particular order.

- **Emphasizing research on a scale that can only be done by teams.** Programs like SciDAC which emphasize large-scale science done by teams of researchers provide an important venue to allow the formation of the critical mass of intellect and effort to solve problems that are beyond the scope of smaller single- and few-investigator projects. This is as true in the ISICs as in the scientific domains – the infrastructure needed for large-scale computational science is itself a large-scale challenge. "Critical mass" funding levels facilitate the delivery of high quality results while providing the team with the flexibility to adapt to "targets of opportunity" and unforseen problems. They also promote the development and growth of communities around the scientific areas, which can have a lasting impact on large-scale computational science.

- **Charge SciDAC ISICs with the full range of research, development, code production, and deployment.** Development and research are too tightly bound in advanced computing to be separated; building the computational support required by modern applications has led to research questions, while research accomplishments may lead to new development challenges and may obviate earlier development problems. Likewise, deployment in actual scientific applications may lead to new, related, R&D activities. Deployment of new tools and technologies, such as those bring developed by the ISICs, requires a close collaboration between the applications and the ISICs.

- **Expect that key participants spend a signficant fraction of their time on their SciDAC project.**

Most participants were funded at a significant fraction of time (25% or more), so that the SciDAC project was a major daily activity for them. This is an important part of insuring the critical mass (by giving participants enough time to focus on the project instead of continually "context switching") and also facilitates collaboration among participants.

- **Maintain five year funding cycles instead of three year.** Grand scale collaborations take longer to stand up and larger problems take longer to solve. Five year cycles are much better suited to projects of the "SciDAC scale", and give project more room to respond and adapt to dynamic situations.

- **Encourage empowerment and inclusivity.** As much as practical, our project was run democratically and even unfunded participants felt empowered. CCTTSS has encouraged participation at its quarterly meetings, and gives voting power to any participant who has attended two of the last three meetings. This has led to a vital infusion of ideas and collaboration from industry, universities, and other DOE labs.

2 Raising the Bar: Suggestions for Improvements

Exploit the power of collaboration to produce new science not otherwise possible — this was the motivation of the original SciDAC, and we believe its successor should take this vision to the next level. While keeping the focus on great science, we advocate increased scrutiny of methodology in this second round of SciDAC. Competitive proposals should reach farther to make a compelling case for grand-scale collaborative efforts. Teams should have to reach out beyond their established circles to find and incorporate new solutions. Specifically, we recommend a renewed emphasis on interdisciplinary coordination (not just interactions), clearer delineation of roles (both computational and organizational), and a wholistic approach to leadership-class computing. For each of these points, we itemize our concerns, suggestions, and rationale below:

- **Increase the time horizons for computational science software.**

  - **Concern:**
    The SciDAC program has strongly emphasized doing new large-scale computational science on the application side, and cast the ISICs in role of supporting that . Some of the tools and technologies developed by the ISICs (and elsewhere in the Math/Computer Science community) can have an immediate impact on the applications, while others involve a longer-term investment for longer-term gain. We have observed that many current applications projects are so focused on the immediate need to produce major scientific results within the time frame of the SciDAC program, that they are reluctant to make investments in new techniques or technologies that would position them better in the long run.
  - **Suggestion:** In *addition* to producing scientific results, the SciDAC proogram should also encourage applications projects to develop/enhance their software and environment with a view to their use as a long-term base for scientific discovery, beyond the end of the SciDAC program itself. Examples of such activities might include:
    * Investing in a significant restructuring of code or algorithms in order to take advantage of new math or computer science techniques that will greatly expand the potential for science in that area, but may not be mature enough to provide that impact immediately.
    * Transforming software from a small team product into community codes, to which many can contribute – and leverage to tackle even more challenging scientific problems. This might involve rearchitecting software, working with others in the community to determine common interfaces, and other activities designed to improve the ability to collaborate around the *software* as well as the science.

  - **Rationale:** ???

- **Increase Incentives for Interdisciplinary Collaborations.**

  - **Concern:** The original SciDAC model heavily encouraged ISIC/Application interactions, but we would like to see more follow-through to actual interdisciplinary collaborations.
  - **Suggestion:** Add formal incentives for interdisciplinary collaborations to the relatively informal (albeit vocal) encouragement already present in SciDAC.
    1. Reward application groups who send interns to ISICs, and vice versa.
    2. Establish a sabbatical program for interested scientists to cross over from/to Application or ISIC teams (Faculty and Lab people).
    3. Award out-year mini-grants targeted to 1-2 year ISIC/Application joint initiatives, perhaps building on Scientific Application Pilots (SAPs).
    4. Encourage coding camps. CCA successfully used these week-long workshops to transition interested application groups into practicing ones. Often this served as a great networking opportunity between creators and customers of the technology.
    5. Demand computational science goals that can *only* be acheived by grand scale collaborations. This would likely take the form of multi-physics or multiscale interdisciplinary applications involving tens of different computational investigators from diverse universities and laboratories.
  - **Rationale:** SciDAC is changing the way scientists do business, and we firmly believe that this change is a positive one that delivers added value to the taxpayer dollar. The goal is to foster integrated computational science that makes use of all of the DOE funded assets together to produce scientific simulations that are only possible with many and diverse investigators. However, culture shifts take time; this one is far from complete. Communities that were hungry for change already have, and those who've resisted are likely to continue doing so without extra incentive.

- **Reconsider Inherent Value of Inter-ISIC Collaborations.**

  - **Concern:** The primary goal — which should remain intact — is for SciDAC ISICs is to impact Applications. However, the importance of building strong inter-ISIC collaborations should not be overlooked as an important first step of that process.
  - **Suggestion:** Recognize ISIC technologies are much more likely to impact each other than the Applications in the short-term because the cultural gaps are smaller. Successes in inter-ISIC exchanges may also serve as an early indicator for eventual Application impact.
  - **Rationale:** Widespread use of a technology across ISICs builds credibility with Applications. Conversely, Application teams are rightfully suspicious of technologies that have failed to garner some level of community backing.

- **More clearly dilineate roles and interactions between Base and SciDAC so that DOE gets maximal value from both.**

  - **Concern:** SciDAC interactions can produce new questions as well as new results. In attempts to aquire new funding to tackle these spin-off research issues, a few of our members have encountered unexpected difficulties. Most troubling was an emerging image problem of being too closely associated with SciDAC "tool development" and not "real science."
  - **Suggestion:** Ammend the current view of Base Program-funded research progressing to SciDAC funding for productization and wide dissemination to include a "feedback loop". It should be expected (and encouraged) that SciDAC tools be employed to foster spin-off research projects suitable for the Base Program portfolio.

  - **Rationale:** New tools — whether microscopes, telescopes, particle accellerators, or software — also present new avenues of scientific inquiry and exploration. Denying legitimate scientific inquiry using Base Program because the tools were developed under the auspices of SciDAC is a missed opportunity to capitalize on investments already made.

- **Fund all Aspects to Achieve Leadership-Class Computing.**

  - **Concern:** Top 500 is a hardware benchmark, not a measure of actual scientific productivity. Maintaining the US lead in computational science requires coordinated and integrated research on the entire vertical stack of technology: hardware, system software, middleware, algorithms, numerics, data management, visualization, physics, system integration, testing, etc. Even though some aspects are harder to measure — or too obscure to effectively promote to nontechnical people — they are all important.
  - **Suggestions:**
    1. Put more weight on winning international prizes that value the entire computational science effort instead of hardware alone: examples include Gordon Bell (particularly in special categories) and Sidney Fernbach awards.
    2. Have DOE/SC generate a quarterly email/webzine directed to all its PIs. This internal publication would promote new developments and opportunities coming from the DOE/SC research portfolio, including the ones that would otherwise not make mainstream press. The reason is not just to inform, but to repeatedly demonstrate that the organization values the entire portfolio, in spite of the mainstream's prediliction on focusing on big iron, a few applications and ignoring all the things in between.
  - **Rationale:** The chain is only as strong as weakest link.

- **Provide better support for communication and collaboration both within projects and between them.**

  - **Concern:** While the SciDAC program is very much about teamwork and collaboration, the projects have been left on their own to deploy collaboration tools for use within the project. Likewise, no support has been provided to facilitate "broadcast"-style communications between projects (such as mailing lists for project PIs/contacts on which ISICs might announce new software releases/capabilities).
  - **Suggestions:** While individual projects can (and our case *have*) deployed "collaboration servers" to facilitate their work together, this can be a significant burden, and requires project members with appropriate experience and willingness to setup and maintain. We believe that it would be more economical and productive for collaboration servers to be centrally operated, maintained, and supported to provide all SciDAC projects with at least a basic electronic collaboration capability. It might even be beneficial to employ and "evangelist", tasked with helping groups unfamiliar with the tools understand how to take advantage of them.

    At the program level, it would be useful to have a hierarchy of mailing lists to allow communication with the PIs of projects under a given program manager/office and larger aggregates. At a minimum, a web-accessible list of projects, PIs, and their email addresses should be available to any SciDAC PI.
  - **Rationale:** You can't have teamwork without communication.

3 CCA-Specific Features

These are some characteristics of our large, multi-institutional collaboration that we feel were important in our success so far. We do not advocate that this be applied to all SciDAC projects, but feel this worked well for us and encourage others to follow suit.

- We had collaborative aims already: The CCA is a grass-roots effort that first aquired funding as a group as a SciDAC ISIC under the name "Center for Component Technology for Terascale Simulaiton Science" or CCTTSS after interacting informally for several years.
- We have used the SciDAC funding we received as a core, around which we have developed a true community organization (the CCA Forum), open to all interested parties.
- The group meets quarterly. Thus when someone new becomes interested, there is not a large delay before integration in the community is possible. The CCA Forum meeting hosting duties rotate among volunteer member sites, this increases our visibility and reduces strain on hosts and travel abilities of participants.
- Technical interest drives expanded participation. Many regular participants in the CCA do not share SciDAC funding (and occasionally not even DOE funding.) Voting membership is open to all who attend two out of the last three meetings (telephone attendance included).
- We are following a successful engineering and business pattern applied to software: components. Collaborative work on the scale needed to handle multi-discipline simulations (10's to 100's of programmers and scientists) is done only via component and plugin architectures. Current examples include any graphical desktop environment, web servers, web browsers, and large scientific visualization packages.
- We have deployed our own "collaboration server" (cca-forum.org), which provides numerous mailing lists, source code repositories, a project web site, and other features which have proven invaluable to our highly interactive project. In the previous section, we suggested making such capabilities available to all projects from a centralized provider, but in the absence of that, we would strongly urge individual projects to establish their own.