# High Throughput Computing as an Enabler for Interdisciplinary Collaboration

**Sara Kraemer**
Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 W. Johnson Street, Madison, WI, 53704
sbkraeme@wisc.edu
+608-265-5624

**Christopher Thorn**
Wisconsin Center for Education Research
University of Wisconsin-Madison
1025 W. Johnson Street, Madison, WI, 53704
cathorn@wisc.edu
+608-263-2709

**ABSTRACT**
In this paper, we outline a conceptual discussion related to potential key sociotechnical characteristics of high throughput computing (HTC). In particular, we are interested in how various characteristics of HTC resources, as well as other contextual factors, enable or hinder interdisciplinary virtual teams' productivity across a range of scientific domains and disciplines. We focus on the Condor Project at the University of Wisconsin-Madison as an example of a widely-diffused model of high throughput computing and compare and contrast their philosophy to groupware design tenets. Research needs center on defining user requirements under various organizational models of HTC as well as criteria for group-level decision support for interdisciplinary teams using HTC.

**Author Keywords**
High throughput computing, sociotechnical systems, virtual teams, interdisciplinary, collaboration

**ACM Classification Keywords**
H.5.3 Group and Organization Interfaces: Computer-supported collaborative work

## INTRODUCTION
There is increasing evidence that HTC—and grid technologies in general—are ubiquitous in and mission-critical to interdisciplinary research that requires large amounts of computing power as well as access to expert advice [1]. We define HTC as an environment that can deliver large amounts of processing capacity over long periods of time. In addition to computational power delivered, there is a second, critical measure of system quality: HTC systems are designed to be extremely fault-tolerant and require minimum human intervention [2]. By design, these technologies enable and support distributed teams. These and other characteristics drive the interactions and forms of collaboration that emerge when users from various scientific domains use HTC resources to work on computational problems. Virtual teams in high throughput computing may vary across: time and geography, domains of science, team size, background or culture, type of task, type of research problems (e.g., applied, basic), computational needs, fluidity of membership in the HTC community, and degree of interdisciplinarity either within their scientific domain and/or across research projects.

Some HTC systems, such as the Condor Project software at UW-Madison, have unique characteristics that foster interdisciplinary virtual team collaboration, such as high degrees of resource flexibility, end-user control, open-ended planning, and distributed resource management [3]. HTC generally, and Condor specifically, are drivers of leading-edge science in local research teams and in large-scale, internationally distributed production environments.

## HIGH THROUGHPUT COMPUTING AND SCIENTIFIC RESEARCH COLLABORATION
The core characteristic of HTC is its ability to provide large amounts of computing for sustained periods of time. However, HTC has a number of additional characteristics that provide greater access to a wide range of disciplines and toolsets. For example, because Condor runs on many computing platforms and can execute any software that does not require user interaction, a wide range of tools are readily available—from commercial research software to scripting engines and compilers. In addition, the abundance of available scientific tools allows individual scientists or teams to engage with the Condor HTC environment using tools familiar to them. Enabling of existing tools in an HTC setting provides critical social and technological gateways for new adopters of HTC. Access to the HTC environment also exposes new adopters to tools and methods used by others to address similar computational problems. In this way, scientists' skills and knowledge are affected by the

capabilities and characteristics of HTC technologies and tools.

Because of its "opt-in" nature, reports of Condor use under represent actual levels of use. Many users are unable to register their existence with the central Condor administrative site because of security concerns or technical limitations of their local network infrastructure. The Condor Project maintains a map page that reports total Condor pools and hosts and is updated daily [4]. On May 9$^{th}$, 2008, there were 32,644 known Condor hosts and 47 Condor pools in Wisconsin. This represents 32.7% of the known systems in the U.S. and 23.9% of the known systems in the world. Condor pools have been created in 38 U.S. states – including the second largest group in the District of Columbia. Condor is also deployed on 136,730 hosts in 39 countries with the second largest deployment being in Switzerland.

The Condor embodies a philosophy of *flexibility*; this philosophy that has served the allowed the design to flourish in a highly unpredictable distributed operating environment [3]. International distributed systems are heterogeneous in numerous ways: they are composed of many types and brands of hardware; they run various operating systems and applications; they are connected by unreliable networks; they change configuration constantly as old components become obsolete and new components are become online, and they have many owners with private policies and requirements that control their participation in the community. Condor has adopted a five-component flexibility philosophy to address these barriers and enable virtual team collaboration:

1. *Let communities grow naturally.* Given tools of sufficient power, people will organize the computing structures they need. However, human relationships are complex, and people invest their time and resources to varying degrees and relationships and requirements change over time. Therefore, Condor design permits but does not require cooperation.
2. *Leave the owner in control, whatever the cost*. To attract the maximum number of participants in a community, the barriers to participation must be low. Users will not donate their property to the "common good" unless they maintain some control over how it is used. Therefore, owners of computing resources are given the tools to set policies and retract resources for private use.
3. *Plan without being picky*. Plan for slack resources as well as resources that are slow, misconfigured, disconnected, or broken. The designers of Condor spend time and resources contemplating the consequences of failure than the potential benefits of success.
4. *Lend and borrow*. The Condor project has developed a large body of expertise in distributed resource management and aims to give the research community the benefits of their expertise while accepting and integrating knowledge and software from other sources. They have also instituted a mechanism for collective problem-sharing and solving among its users.
5. *Understand previous research*: The Condor Project continually updates its organizational knowledge with previous research to apply well-known fundamentals as well as cutting-edge techniques to emergent problems. The inclusion of current user innovations keeps the work focused on the edge of discovery rather than wasting effort remapping known territory.

As outlined in the philosophy of flexibility, the Condor approach is much more than a complex set of computational resources. The Condor team maintains a close intellectual partnership of computer and domain scientists working together on the challenges of HTC in the context of break-though science. Condor has advanced HTC technology via improvements in their software coupled with innovations in the computational approaches to the domain scientists. These interactions have made Condor privy to numerous sets of interdisciplinary virtual team as well as numerous types of sociological and technological problems.

### Specific types of scientific collaboration

Experiences at the Condor Project suggest that multidisciplinary and interdisciplinary collaboration around HTC resources reveal important sociotechnical implications for collaborative research as well as technology design [5]. Interesting interactions arise when the barriers to computational resource access are removed. For example, some virtual teams using HTC embody what is known as "tool-based" specializations, that is, they view their interaction with HTC as a function of their research areas while other teams exemplify a "perspective-based" specialization where the HTC ceases to be part of their research area and problem situations define their research space [6, 7]. Those operating in perspective-based specialization may be more likely to share local research methods and solutions with other team members or teams and view HTC as an enabling technology, rather than simply as resource sharing.

As HTC technologies mature, they move out of a tool-based environment and become a part of the research infrastructure. The move from "everyday" research means those users' perceptions of risk and trust will be enhanced, thereby enabling new models of financial support and social engagement. When the tools themselves move out of basic research into applied research and production work, new risks and benefits appear. For example, computer scientists may perceive a risk of being identified as merely an infrastructure provider On the other hand, for those on the contemplating the adoption of HTC, the emergence of HTC as a commodity service greatly reduces their perception of risk of adoption as a potential new user. As HTC systems provide ubiquitous access to robust computing environments, new potential members will see HTC as a relevant and important aspect of their work.

Perspective-based collaboration is particularly common in experimental research involving complex instrumentation, such as telescopes, particle accelerators, or CT scanners [8]. High-energy particle physics is a well-known domain that has embodied perspective-based collaboration characterized by collectivism, erasure of the individual epistemic subject, non-bureaucratic mechanisms of work, lack of over-bearing formal structures, and an absence of rigid rules [9]. Similar types of collaborations and configurations have been anecdotally found by the Condor team, based on their experiences working closely with domain scientists. These examples are representative of possible sociotechnical components of a virtual interdisciplinary team model or typology of virtual team scientific collaborations using HTC resources.

## PRINICIPLES OF CSCW DESIGN AND CONDOR

In Grudin's [10] paper on groupware design challenges for developers, eight themes were outlined as key obstacles to creating truly usable groupware technologies. While the main focus of this paper was on communication technologies, the principles of good groupware design can also be compared and contrasted to the Condor project design themes:

1. *Disparity between who does the work and who gets the benefit.* This tenet refers to the perception of group benefit technology use and Condor has worked to build a perception that there are benefits to participating in their high throughput computing system via their philosophy of flexibility. Condor allows users to determine their level of participation in the system and have intentionally left relationships and requirements loosely defined so that research communities can grow naturally.
2. *Critical mass and prisoner's dilemma problems*. Most groupware will only be useful if a high percentage of group members use it. To incentivize group members, Condor has set low barriers to participation and allows users to set use policies, even if they only use it for single use.
3. *Social, political, and motivational factors.* Condor's philosophy of allowing communities to grow organically and in their own scientific contexts acknowledges the various contextual factors that affect HTC use. Condor engages deeply with users; who engage very deeply with users on scientific problems, are critical sources of knowledge for user requirements and problem specification. Condor team members are "complex problem unpackers"; they assist with understanding the problem's context and removing barriers to large, complex problem spaces and communities.
4. *Exception handling in workgroups* is related to #3 and recognizes that specific contextual circumstances create *ad hoc* problem solving. Because Condor imposes little organizational controls on their users, a greater level of flexibility is created. They anticipate the consequences of system failure and have created a fault-tolerant system that also allows for fluid policies and use.
5. *Designing for infrequently used features.* Groupware features will fare better if integrated with features that support individual activity. Leaving the owner in control not only supports this tenet but incentivizes users to donate their resources. For example, Condor group members can participate in "Condor Week", an annual event where the Condor team and its users convene to share information and ideas. This is also a time for developers to engage directly with users about unique forms of use or emerging needs.
6. *Underestimating the difficulty of evaluating groupware*. Certainly, it is difficult to evaluate the effectiveness of Condor use in a distributed environment and generalizing from a use case or small set of use cases can also be risky. Condor is developing artifacts and systems to engage users and disseminate some of their results (see next section for our proposed research with Condor).
7. *Breakdown of intuitive decision-making.* These problems occur when decision-makers are drawn to an application that benefits one group of users or a particular use. Condor must present their resource and services so that all user sets, including the decision-makers, can clearly understand the benefits of using the Condor system. This includes maintaining a deep and working knowledge of previous research in the various facets of computer science.
8. *Managing acceptance*. Condor's acceptance and use will be directly related to how well they can understand the group member's work environments and design Condor system and services to meet real needs. We hope to work in this area with Condor to further develop knowledge about users' work environments in our proposed research of interdisciplinary virtual team performance.

## DESIGNING FOR COLLABORATION

The sociotechnical approach embodied in this paper can be used to further develop the effective design of cyberinfrastructure to support interdisciplinary collaborative research. An effective approach to uncovering the HTC user needs would certainly investigate contextual factors under various models of HTC organization. For example, the Rosen Center for Advanced Computing at Purdue University offers a stark contrast to the Condor Project's organizational model. The Rosen Center is a support center for advanced computing based in a central IT unit while the Condor Project organizes several large HTC initiatives on campus. In addition, the Condor team both supports the use of Condor generally as well as engages very deeply with local clients' and their projects, supporting them in the defining and redefining of their research problems, assisting with interdisciplinary collaboration, and providing technical assistance. The Condor Project and Rosen Center users include biostatistics and medical informatics, genomics and genetics, engineering physics, chemical engineering, chemistry, computer sciences, medical physics, particle physics, and astrophysics.

### Proposed research

We are currently proposing research to identify and describe interdisciplinary team performance via team modeling techniques developed in human factors engineering (see CSCW and Human Factors workshop paper for in-depth discussion). The main outcome of this research will be a model of interdisciplinary virtual team performance spanning a number of dimensions, such as: space, time, cultural backgrounds, social norms, workflows, and computational environments. This model will focus on interdisciplinary virtual teams working in distributed computing environments across computational research and scientific domains (the Condor Project and Rosen Center will serve as sampling pools). Through iterative sets of interviews and focus groups conducted with HTC computer scientists and the interdisciplinary virtual teams they serve, we aim to: (1) describe the sociotechnical factors that contribute to and hinder virtual team performance; (2) specify the process and task performance outcomes of virtual teams; and (3) model effective virtual team performance (with a specific emphasis on interdisciplinarity and collaboration).

To date, there has not been an in-depth analysis of virtual team collaboration in scientific communities, or of multi- and interdisciplinary virtual teams, specifically. We will address these gaps by exploring the sociotechnical systems aspects of virtual teams conducting scientific research with HTC. An analysis of the social impacts of the technical configurations of HTC software such as Condor will lead to deeper understanding of how HTC is used as an effective enabler of new scientific problem sets, solutions, and collaboration configurations, as well as how the HTC technology can be designed and deployed to meet emerging scientific problems and configurations.

## ACKNOWLEDGMENTS

We thank Miron Livny and Todd Tannenbaum of the Condor Project at the University of Wisconsin-Madison for the comments and input into the development of these ideas and for their reviews on earlier drafts of this prose.

## REFERENCES

1. Berman, F. and T. Hey, *The Scientific Imperative*, in *The Grid 2: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Editors. 2004, Morgan Kaufmann Publishers: San Francisco, CA. p. 13-23.

2. Livny, M. and R. Raman, *High throughput resource management*, in *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, Editors. 1998, Morgan Kaufmann Publishers Inc.: San Francisco, CA. p. 311-337.

3. Thain, D., T. Tannenbaum, and M. Livny, *Distributed computing in practice: The Condor experience.* Concurrency and Computation: Practice and Experience, 2005. **17**(2-4): p. 323-356.

4. Condor Project, *Condor World Map*, http://www.cs.wisc.edu/condor/map/. 2008, Department of Computer Sciences, University of Wisconsin-Madison.

5. Thain, D., T. Tannenbaum, and M. Livny, *How to measure a large open-source distributed system.* Concurrency and Computation: Practice and Experience, 2006. **18**(5): p. 1989-2019.

6. Schmidt, K., *Cooperative work: A conceptual framework*, in *Distributed Decision-Making: Cognitive Models for Cooperative Work*, J. Rasmussen, B. Brehmer, and J. Leplat, Editors. 1991, Wiley: Chicester, West Sussex, England.

7. Whitely, R., *Cognitive and social institutionalization of scientific specialties and research areas*, in *Social Processes of Scientific Development*, R. Whitely, Editor. 1974, Routledge and Kegan Paul: London p. 69-95.

8. Katz, J.S. and B.R. Martin, *What is research collaboration?* Research Policy, 1997. **26**: p. 1-18.

9. Chompalov, I., J. Genuth, and W. Shrum, *The organization of scientific collaborations.* Research Policy, 2002. **31**: p. 749-767.

10. Grudin, J., *Groupware and social dynamics: eight challenges for developers.* Communications of the ACM, 1994. **37**(1): p. 92-105