1.0 Introduction
Computer architectures have now moved into the multi-core era, with projections pointing to commodity-class chips with hundreds of thousands of cores. We have also moved into the petaflop era (which will last less than a decade) to be followed by exascale systems, which are 1000 times faster: 1 million trillion flops. But before this level of performance can be realized, software and hardware engineering challenges will need to be addressed (see Torrellas, 2009). Though software environments that coordinate access to, and use of, distributed high performance computing resources have only entered a figurative toddlerhood, considerable progress has been made during the past decade (Armstrong, 2000). In this paper, I sketch out some likely effects of these developments on geocomputation.

2.0 Geography and Exascale Computing: Challenges and Opportunities
Exascale computing and cyberinfrastructure (CI) will revolutionize the way that we conceptualize and address geographical problems. A particular focus is placed on socio-economic analyses in this paper.

2.1 New Views of Optimality. Many geographical problems are computationally complex. As a consequence, heuristic optimization methods have been used for decades to address problems that could not be handled by exact methods. Thus, while large \( NP \) complete problems cannot be considered using exact methods, new scalable heuristics enable the exploration of the solution spaces of problems. The computation of legions of candidate solutions in near-real-time will enable decision-makers to design optimal (and near-optimal) solutions to complex policy problems in a way that is similar to the interactive use of CAD systems. Once candidate heuristic solutions are designed, they can be placed into a broader context through the use of exascale computing methods, involving either exhaustive enumeration or parallel versions of non-heuristic methods.

2.2 Moving from Pre-defined Aggregates to Individual-level Analysis. In the coming decade CI will support a move from existing pre-defined units (e.g., counties) to customized units of analysis (or quasi-continuous surfaces) that are created, bottom up, from massive amounts of individual-level information using geocoding methods (Rushton et al., 2006). While existing units, such as census tracts, will persist, their use as the sole lens through which socio-economic processes are viewed, will be diminished. This will lead to the creation of custom-designed optimal regions which can be defined using a variety of criteria, though typically requiring trade-offs.

2.3 Rethinking Site and Situation. There are two basic concepts in elementary economic geography: site and situation. Site refers to the characteristics of a place; situation places a site in relation to others. Site is usually conflated with a location, such as a surveyed parcel of land or settlement with an associated set of defining characteristics \((x, y, z, t, C)\), where \(x, y, z\) is a locational triplet, \(t\) is a time stamp and \(C\) is a vector of attributes. This is a relatively static conceptualization, even with the provision of a temporal snapshot.
In the current era, however, we can consider a site to be a more fluid entity, or a spatial process with transformations: birth-death, expand-contract, migrate, split-combine.

2.4 Real-time Bayesian Analysis. With massive input data streams, continuous (at a user-specified interval) updates to analyses will become a reality. Bayesian methods hold some promise in this regard, but they are typically very computationally intensive (Yan et al., 2007). The development of such approaches will also enable early-warning systems for extreme or unusual events (Apolloni et al., 2009). However, in such cases, robust methods will be needed for false positive detection.

2.6 Distributed Group Decision Support. Difficult policy problems require the application of different types of knowledge possessed typically by several individuals, who may not be co-located (Armstrong and Densham, 2008). CI supports the creation of virtual groups that can work to develop solutions to problems using geospatial tools. A compute-wiki approach can be used to create, modify, discuss and evolve solution processes by group members who may use it in different times and places. Synchronous use of a group decision support environment will require low-latency in both solution computation and during synchronous interactions across networks. This is problematic given the computational complexity of some solution processes and the rapidly expanding sizes of geospatial databases.

3.0 Geographical Aspects of Cyberinfrastructure Barriers-to-Use
In addition to generic problems that must be overcome by computer scientists in the coming years (e.g., Tallent and Mellor-Crummey, 2009), there are geographically-specific issues that must be addressed to achieve high levels of performance and efficient use of resources.

3.1 Decomposition and Locality. If we assume that high performance will come from large numbers of concurrently executing threads that are not suffering from data starvation and communication latency, decomposition and locality emerge as critically important issues. A mapping between memory hierarchies (Schneider et al., 2009) and geographic relationships also can be used to minimize latency. Wang and Armstrong (2003; 2008) describe a first approximation to locality-based decomposition, but much work remains. We particularly want to minimize latency, especially in distributed environments. Though latency can be reduced by exploiting locality in geospatial information, locality information needs to be added or computed to records. This presents a space-time tradeoff, but storing locality information (metadata) with geospatial records warrants investigation. This new type of metadata would be useful in determining tasks, task sizes, and their relative computational burdens, and would thus support load balancing as well as reduce network traffic and latency. Fortunately, many geographic units are already organized hierarchically, so that information might be used to construct computational tasks and support efficient approaches to exploiting locality (e.g., U.S. Census Blocks and Tracts).

3.2 Education. We aren’t training (many) GIS-types who know anything meaningful about parallelism and CI, and we aren’t training CS-types who know much about geographical analysis. Though multi-core parallelism is now normative, programmers trained to exploit the potential of such hardware are in short supply. This is recognized by industry and funding agencies, and as a consequence there are ongoing projects focused on addressing this issue (see Wilde et al., 2009; Torrellas et al., 2009; and initiatives at Illinois and Berkeley supported by Microsoft and Intel, among others). Researchers in the intersection between CS and GIScience are needed to guide future developments in this rapidly changing domain of knowledge.
References


