

Archaeology in the Age of Supercomputing

Devin A. White

Abstract As a discipline, archaeology is poised to fully embrace both the power and the peril of big data analysis. Our datasets are growing ever larger, especially those generated via remote sensing and geospatial processing activities, as is the computational complexity of algorithms designed to exploit them. Analyses are quickly outpacing what can be done using a single processing core on a desktop computer, leveraging off-the-shelf commercial and open source software. Our research needs are becoming increasingly sophisticated, to the point where relying wholly on outside experts in computer science and related fields is untenable. While the above statements could be viewed primarily as challenges, it is better to think of them as opportunities for archaeology to grow technologically and retain more ownership of our hardest problems. High performance computing, i.e., supercomputing, is already having an impact on the field, but we are moving into an era that promises to put the power of the world's largest and fastest computers at archaeologists' fingertips. What does the state of the art look like? How could we use the coming power? What lines of inquiry and analysis could we pursue once long-standing technical limitations have been removed? This chapter will focus on the present and the future of archaeological high performance computing, using several ongoing projects across a broad swath of the discipline as examples of where we are now and signposts for where we are heading.

Introduction

As a discipline, archaeology is poised to fully embrace both the power and the peril of big data analysis. Our datasets are growing ever larger, especially those generated via remote sensing and geospatial processing activities, as is the computational

D.A. White (✉)

Geographic Data Sciences, Geographic Information Science and Technology Group,
Oak Ridge National Laboratory, Oak Ridge, TN, USA
e-mail: whiteda1@ornl.gov

D.A. White

Department of Anthropology, University of Tennessee, Knoxville, TN, USA

complexity of algorithms designed to exploit them. Analyses are quickly outpacing what can be done using a single processing core on a desktop computer, leveraging off-the-shelf commercial and open source software. Our research needs are becoming increasingly sophisticated, to the point where relying wholly on outside experts in computer science and related fields is untenable. While the above statements could be viewed primarily as challenges, it is better to think of them as opportunities for archaeology to grow technologically and retain more ownership of our hardest problems. High performance computing, i.e., supercomputing, is already having an impact on the field, but we are moving into an era that promises to put the power of the world's largest and fastest computers at archaeologists' fingertips. What does the state of the art look like? How could we use the power already available, much less what is coming next? What lines of inquiry and analysis could we pursue once long-standing technical limitations have been removed? How will fieldwork be transformed?

This chapter, after an introduction to the world of high performance computing, will focus on the present and the future of archaeological supercomputing, using several ongoing projects across a broad swath of the discipline as examples of where we are now and signposts for where we are heading, concluding with some thoughts on the art of the possible, given current and emerging technological trends. It is hoped that the reader will come away feeling less intimidated by the idea of using supercomputing to solve archaeological problems, and knowing that they can and should take full advantage of the computing power available today as well as help drive how the systems of tomorrow are designed.

The Current State of the Art for High Performance Computing

“Supercomputing” is the widely-recognized name for what has now become a much more varied computational landscape, driven largely by the commoditization, democratization, and more recently, miniaturization of the necessary hardware and software. The three trends have created, and will continue to create, opportunities for archaeologists to generate, process, analyze, visualize, and contextualize data more quickly, accurately, and creatively. To do so most effectively, and to better imagine what could come next, it is important to appreciate the current state of the art and to understand that there are resources available to scientists at multiple levels, some of which are relatively straightforward to access and leverage. This section will serve as a relatively non-technical primer on high performance computing (the more proper term), focusing on the definition and function of a broad set of technologies archaeologists are already using, would like to use, or may not even know exist. To that end, it lays the foundation and establishes a frame of reference for the illustrative use cases highlighted below and sets the stage for the subsequent discussion on the art of the possible.

The most fundamental concept associated with high performance computing (HPC) is that of *parallelism*. In its simplest form, parallelism is the process by which a large problem is broken up into smaller, more manageable pieces and distributed to a large number of computers (hereafter referred to as nodes) that can each work on their assigned piece independently. The two most common problems encountered are (1) single case, where one larger set of calculations is run to produce a single result, and (2) ensemble case, where multiple smaller sets of calculations are run to produce multiple results. The latter is generally used when exploratory research is required to understand the full behavior of an algorithm or model, so many variations on inputs are supplied and the output is analyzed statistically to derive meaningful trends. When all of the pieces have been worked on, the individual results are brought together to form the final output. This is the ideal case, where little to no communication between nodes is required. There are far more complicated cases, which is more of the norm, where nodes must update each other on what they are doing and exchange data. There are several different ways, with respect to both hardware and software, that parallelism is “expressed” in HPC systems, and they are not always mutually exclusive. Each one, discussed below, was born of a different computational need and thus will differentially apply to a problem of interest—and may even be combined when it is advantageous to do so. A scientist working within the HPC domain is primarily trying to figure out how to break up a problem so that it can run most efficiently on the hardware and software they have available (square peg in round hole) or trying to seek out the right hardware and software to meet the requirements of their current solution (square peg in square hole). Both avenues create challenges and opportunities. It is best to begin with an overview of hardware, after which software will be discussed.

HPC Hardware

Modern computers, with rare exceptions, have a central processing unit (CPU) that is made up of multiple cores. Each core is capable of executing a computational task independently of the others, but can do so within the same workspace (system memory), enabling each of them to easily talk to one another and exchange data while they are working on a common problem. In this tightly-coupled arrangement, sometimes referred to as *strong parallelism*, each core can be thought of as a very efficient node. Sadly, most of today’s desktop software—especially the applications focused on image processing and geospatial data analysis—is not written to take advantage of multiple cores despite their widespread availability, but the situation is slowly improving and archaeologists are beginning to benefit from the more rapid data analysis available through tools they are already using. The size of strongly parallel solutions, which tend to work best for problems that require a significant amount of communication between nodes and/or for inputted data to remain whole throughout the process, are most often limited by the number of cores and the amount of available memory on the system. While these numbers are always

increasing (for example, this manuscript was written on a system with 32 cores and 512 GB of RAM, which is close to the high end at present), they cannot keep pace with the exponentially-growing size of big data problems. Larger problems are handled one of two ways: (1) build a larger tightly-coupled system or (2) find a way to use a more loosely-coupled approach.

Larger systems, known as Symmetric Multiprocessing (SMP) solutions, can be thought of as gigantic workstations. They generally have hundreds or thousands of cores that all have access to a large pool of shared memory. SMPs are purpose-built and not very common, but for certain classes of problems that simply cannot be solved in a loosely-coupled way, or where doing so would be prohibitively expensive, they are the best option. For scientists who have access to such systems, they are also a great way to test the *scalability* of an existing tightly-coupled approach, i.e., the efficiency of it when a much larger amount of resources are available and used. In some cases, simply throwing a bigger system at the problem does not help, which is when the overall approach has to be rethought.

A special case of strong parallelism is *massive parallelism*, which is more commonly referred to as hardware acceleration. Massively parallel arrangements are the domain of General Purpose Graphics Processing Units (GPGPUs) and more specialized devices called coprocessors. Both, but primarily GPGPUs, have revolutionized HPC over the past decade. Originally designed to rapidly process data for display on individual computers, in particular to support the video game industry, they have been repurposed, or even purpose-built, to instead execute mathematical operations of interest to scientists. The highest end GPGPU available today, the NVIDIA Tesla K80, has the equivalent of 4992 cores, which is on par with what one finds in SMP solutions, but the required hardware is the size of a small book (Fig. 1). The smaller size, and its design legacy, imposes some very significant restrictions on the type of problems it can solve. First, it is not a computer in its own right—it has to be connected to a traditional computer, but more than one can generally be connected to the same computer (an important distinction that will be revisited below). Second, it is designed to execute a massive number of very simple calculations that don't depend on one another, which is what is needed for data display. As communication requirements gradually increase, the utility of a GPGPU, or even a coprocessor, rapidly decreases. Developing software that can effectively run on this kind of hardware is often a laborious, expensive, frustrating, and counter-intuitive process, but when done right, the resulting speedups are extremely impressive. One area where archaeologists are already benefiting from GPGPUs is in viewshed analysis, exactly the kind of problem the hardware was designed to solve. You are in essence creating a three-dimensional environment and determining what can be seen from a particular vantage point, which is a crucial element of modern, immersive video games.

A loosely-coupled arrangement, sometimes referred to as *weak parallelism*, is traditionally associated with the notion of a *cluster*, which many readers are likely familiar with. A cluster is a collection of individual nodes that can act as a greater whole through software-based orchestration, but the connections between them are very tenuous and computationally expensive to use, and resources like memory are

Fig. 1 GPGPUs, and related hardware known as coprocessors, are small devices capable of quickly executing massive numbers of calculations, but they have to be connected to a regular computer to function. The NVIDIA Tesla K80 GPGPU, currently the fastest in the world, is pictured here. Image courtesy of NVIDIA Corporation



not shared. As such, they work best when communication is kept to a minimum. It might be surprising to the reader, but all of the leading supercomputers in the world, including Titan at Oak Ridge National Laboratory, employ this arrangement (Fig. 2). Titan is composed of 18,688 nodes (16 cores and 32 GB of RAM per node), each of which is capable of operating as an individual workstation. The modern science (and art) of HPC is centered on how to use a system like that to its fullest potential. Apart from ease of maintenance (nodes can be easily replaced), one of the main advantages of loose coupling is that the solution can more easily scale in a physical sense, i.e., nodes can be quickly added or removed as needed to meet the requirements of a specific problem. There are two common types of loose coupling arrangements, most often differentiated by whether or not the nodes exist in the same physical location. When that is the case, and every effort is made to accelerate communication between nodes through specialized hardware and software, the arrangement is *scientific*. When off-the-shelf hardware and software are used, or more importantly, the nodes are scattered across a wide geographic area and talk to one another over the internet, the arrangement is *commercial*, what we generally think of as “the cloud.” The former (think Titan) is built for absolute speed, the latter (think Google) is built for more general purpose computing tasks like sending email and streaming movies. An apt comparison would be that of a sports car to a minivan: They each excel at certain tasks, and can each do what the other is good at, but not necessarily very well. They also have very different price tags, due to how they are designed and built. Examples of archaeologists using large clusters are few and far between, but that is beginning to change as they begin to explore large-scale modeling and simulation frameworks (scientific) and/or need to process massive quantities of data using off-the-shelf software (commercial).

An increasingly common approach to problem-solving in the HPC domain is to combine the different types of parallelism, leveraging each one for its strengths, an arrangement known as *hybrid parallelism*. For example, a single case problem can



Fig. 2 ORNL's Titan supercomputer, which employs weak parallelism to connect a large number of individual computers for complex problem-solving. This is the most common arrangement for today's high-end scientific systems. Only a small part of Titan, currently the second-fastest computer in the world, is pictured here

be broken down into several large chunks, each of which is passed to a strongly parallel node on a weakly parallel system, which then further breaks down the chunk so each of its CPU cores is working on a different sub-chunk, and some of those cores might be passing part of their sub-chunk to one or more locally-connected massively parallel GPGPUs for rapid analysis (because at this point the calculations are simple). A scientist that is able to successfully harness the available parallelism at all of these levels is akin to a symphony conductor. It should be noted that it is not necessary, or even advantageous, to do this for every problem. Archaeologists reconstructing three-dimensional scenes from large volumes of imagery have probably taken advantage of hybrid parallelism and did not even know it. Most desktop commercial photogrammetry software (e.g., PhotoScan and Pix4D) can take advantage of multiple CPU cores and multiple GPGPUs, if they are present, but the advantage does not extend beyond a single machine.

There are two additional hardware trends worth noting here, as they will be particularly useful for archaeologists: miniaturization and specialization. Within the past few years, innovative hardware manufacturers have found ways to shrink the size of the components required to make a functional computer, resulting in affordable units that are the size of candy bars (and smaller). The two most well-known examples at present are the Raspberry Pi and the NVIDIA Jetson. Both are smaller than a GPGPU and are full-fledged computers with no moving parts, running regular operating system software, capable of completing a wide range of tasks. It did not take long for curious scientists to find ways to link several of them together to create what is essentially a desktop supercomputer that runs exactly the same software as larger machines and, with some creativity, can harness hybrid

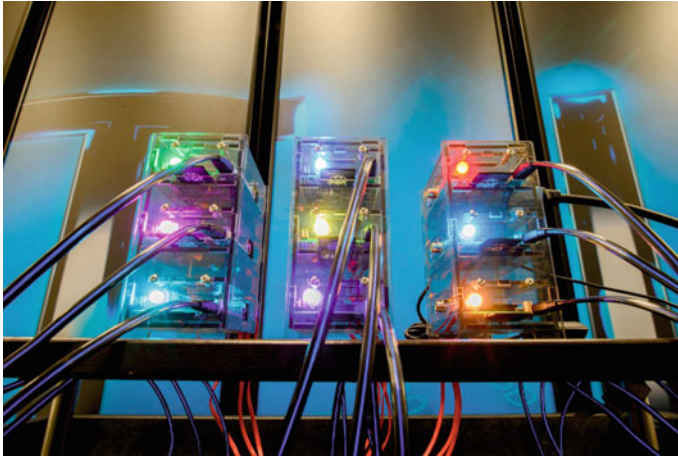


Fig. 3 ORNL's Tiny Titan supercomputer, built from nine Raspberry Pi computers, pictured in front of Titan for comparison. It can run many of the same applications as its larger cousin, albeit more slowly and at a smaller scale

parallelism to solve problems. One of these, dubbed Tiny Titan (Fig. 3), is used at ORNL as an instructional aid during visits by local schoolchildren, but it is capable of doing much more and cost very little to make. Imagine being able to take a small, lightweight, durable, cheap supercomputer with you to the field. At the other end of the spectrum are increasingly specialized machines that are designed to solve one type of problem extremely well, which in some ways brings HPC full circle to its early days, when custom-built computers were the norm. The most notable example at present is the “graph discovery” appliance built by Cray, which at every level, from basic design and construction to the software it runs, is focused on quickly analyzing extremely large networks, which has been a very difficult problem for traditional HPC systems to solve due to the amount of shared memory and communication required. Given that archaeologists are beginning to more seriously focus on social network analysis, and are rapidly running into limits with respect to the amount of data they can process, a specialized system like this could be very helpful.

HPC Software

Hardware serves as the physical foundation layer for HPC, but there is an equally important virtual foundation layer that is important to briefly review here: The software that governs how the various components of the system are exposed to the scientist, communicate with one another, and exchange data. As with the hardware

discussion above, the review will begin with what can run on a single computer (node) and build upwards and outwards from there.

For strongly parallel systems, where CPU cores and shared memory are the main units of currency, there are several software frameworks available that make it relatively easy for a scientist to develop a parallel application. The most popular of them is OpenMP (<http://openmp.org/wp/>), but an alternative is Intel's Threading Building Blocks (<https://www.threadingbuildingblocks.org/>), also called TBB. They are both written in C/C++, work across multiple operating systems, and enjoy widespread support within the scientific HPC community. C/C++ is not often considered to be an approachable language, though, so it is important to note that similar functionality is available in many of the cross-platform languages that are popular with archaeologists, specifically Python, R, MATLAB, IDL, Java, and JavaScript. In many cases, all one has to do is add a small amount of new code to an existing application to get an instantaneous boost in performance. It may not be the largest possible boost, but it generally helps and allows you to better understand what your application is capable of. Some languages, like MATLAB and IDL, will automatically detect and use all available cores for certain tasks.

For massively parallel systems, where GPGPUs and coprocessors are the main units of currency, there are two main software frameworks available, both of which are becoming easier for scientists to use, but generally speaking require significantly more experience to leverage to their fullest potential. The most popular one by a large margin is CUDA (http://www.nvidia.com/object/cuda_home_new.html), which only works with GPGPUs manufactured by NVIDIA. The alternative, which works with all types of hardware accelerators, is OpenCL (<https://www.khronos.org/opencv/>). Both are written in C/C++ (CUDA also offers a Fortran version) and are relatively difficult to master. They each, like OpenMP and Threading Building Blocks, enjoy widespread use within the scientific HPC community—especially since many of the world's largest and fastest computers have at least one GPGPU or coprocessor attached to each node, as do miniaturized systems like Raspberry Pi and NVIDIA Jetson. They are also widely used in the computer vision and computational photogrammetry communities due to their unique ability to quickly process imagery. As a result, many desktop image processing packages used by archaeologists are already taking advantage of CUDA and/or OpenCL. It should be noted that many of the archaeologist-friendly languages mentioned above do provide methods for communicating with CUDA and OpenCL, but they are primarily limited to offering “faster” versions of specific, popular computational tasks (e.g., multiplying matrices, filtering images) if acceptable hardware is available on the computer. In other words, one gives up all control, but some results may come back faster than if no hardware acceleration was available.

For weakly parallel systems, where nodes are the main units of currency, there are also two main software frameworks available. One is specific to scientific systems and the other to commercial systems. Message Passing Interface, or MPI, is the standard way nodes communicate with one another on a scientific system, and there are several flavors available, including MPICH, MVAPICH, and Open MPI. They are all written in C/C++ and have moderate learning curves with respect to

picking up the basics. MapReduce, specifically the free and open source implementation of it called Hadoop (<https://hadoop.apache.org/>), is hugely popular within the commercial HPC space. The paradigm was originally developed by Google, but quickly caught on with non-scientists because it is written in Java, runs on almost anything, and has a relatively shallow learning curve. The core concept behind the framework is that a single problem can be broken down into smaller pieces and distributed to several nodes (mapping), where the pieces are analyzed to generate a small number of meaningful values (reducing), which are then aggregated to create output. A very large community has built up around it, even on the academic side, but interestingly enough, Google has already moved on to something else. MPI and MapReduce each have strengths and weaknesses, and there are passionate supporters on both sides, but at a higher level, while they share the basic concept of passing information between nodes, MapReduce performs only a subset of what MPI is capable of—but it is a very useful subset for a wide variety of relatively simple problems that require quickly processing a large amount of data and for which it is possible for each node to work independently. It is possible to be very creative in using Hadoop to solve far more complex problems, but there is a lot of overhead, and some risk, involved in doing so. Unlike MPI, which is tied to a specific system, it is also capable of using networked nodes that are scattered across a wide geographic region (i.e., the cloud) and most commercial cloud vendors already support it. It is at a distinct disadvantage compared to MPI when it comes to hybrid parallelism, though. For example, a single node with multiple CPU cores is generally treated as multiple single-core nodes in MapReduce, greatly limiting the potential to break down a large problem even further or leverage the strengths of all available hardware. Again, archaeologist-friendly languages like Python and R provide ways for scientists to quickly and easily tackle big data problems on weakly parallel systems, especially in the area of statistical analysis, by leveraging MPI or Hadoop behind the scenes if either framework is present, but one loses a great deal of control over how that happens. Lower-level access, where one can more fully control how nodes communicate with one another, is also available, but it does require the programmer to do a lot more work to manage the entire process.

As a footnote to the above discussions on hardware and software, the line between weakly parallel scientific and commercial HPC systems has started to blur in the past few years. Scientific HPC researchers are finding ways to run Hadoop efficiently on their systems so that they do not have to rewrite useful software and commercial HPC providers are teaming up with those researchers to find ways to run MPI-based scientific applications in the cloud. There is a price to pay in terms of performance when going either of those directions, but the end results are that (1) software that is easier to write is becoming easier to run on large supercomputers that already exist, lowering the barrier to entry for archaeologists who might already have access to those kinds of systems through their home institutions, and (2) it is becoming possible to temporarily and inexpensively build a traditional scientific supercomputer whose nodes could exist all over the world, when the need arises.

At this point in the discussion, the reader's head is most likely spinning, given the wide array of available hardware and software options that vary substantially

with respect to availability and usability. Table 1 provides a summary of the software paradigms mentioned above, organized by hardware paradigm, and also indicates which ones can be accessed through several archaeologist-friendly languages (and at what level). The take home message for the reader should be that there are several levels at which HPC can be applied to archaeological problems, and those levels can be initially explored easily through languages like Python and R. When more sophisticated control is required, or a problem outgrows locally-available computing resources, more complicated frameworks do exist that can help, but finding the right one(s) to use and exploiting them to their fullest extent may require teaming up with a computer scientist.

Available Resources Beyond the Desktop

There is a wide array of HPC hardware and software resources available to archaeologists, some of which might be free or very low cost to use—if one knows who to talk to and what questions to ask.

On the scientific side, the prime example is the National Science Foundation's Extreme Science and Engineering Discovery Environment (<https://www.xsede.org/>), also known as XSEDE. Scientists can apply for computing grants, ranging from small seed projects that test out ideas to large projects that might consume a significant amount of resources, and the program has several machines distributed around the country that specialize in different aspects of HPC. Before a large project grant is awarded, the scientists submitting it must document their application's performance on several systems of increasing size and complexity and thoroughly justify why one of the NSF's largest machines is required. No direct funding is awarded at any level. Instead, grantees are given time allocations that roughly equate to the number of hours a single CPU core can be used. So, for example, if a grant is awarded for 100,000 h on a system with 100,000 CPU cores (Titan has almost 300,000), a scientist could theoretically use up their entire allotment in an hour if they ran an application such that it requested and used all available resources. Generally speaking, multiple projects are running on large systems at the same time, so it is very rare to have access to all of it, but it is something that should be kept in mind.

Other countries with robust scientific research programs have similar grant initiatives, as do most universities. In almost all cases, both during the application process and after a grant is awarded, the scientists are paired with one or more technical support liaisons whose job may include translating algorithms so that they can run on the requested system—especially when the scientists come from a non-traditional computing discipline like archaeology. It is important to note that as of the time this chapter is being written, social science is still considered a novel focus area for scientific HPC, so there is great interest from that community and many opportunities to find support for little or no cost.

Table 1 A summary of HPC software paradigms, organized by hardware paradigm, with indicators for the level of support available through a selection of archaeologist-friendly languages

	Domain	Popularity	Language	Sophistication	Learning curve	Python	R	Java	JavaScript
<i>Strong parallelism</i>									
OpenMP	Scientific	High	C/C++	High	Moderate	High	Medium	High	Medium
TBB	Scientific	Low	C/C++	High	Steep	High	Medium	High	Medium
<i>Massive parallelism</i>									
CUDA	Scientific	High	C/C++	High	Steep	Low	Low	Low	Low
OpenCL	Scientific	Low	C/C++	High	Steep	Low	Low	Low	Low
<i>Weak parallelism</i>									
MPI	Scientific	High	C/C++	High	Moderate	Medium	Medium	Low	Low
Hadoop	Commercial	High	Java	Low	Shallow	Medium	Medium	High	High

The situation on the commercial side is more heterogeneous, given that the suppliers of available computing resources are focused on making a profit. The three largest are Amazon Elastic Compute Cloud (<http://aws.amazon.com/ec2/>), Microsoft Azure (<http://azure.microsoft.com/en-us/>), and Google Compute Engine (<https://cloud.google.com/compute/>). All of them, as noted above, support Hadoop, but support for MPI is minimal to nonexistent. For scientists working with satellite imagery and derived products, there are two additional options worth considering: Google Earth Engine (<https://earthengine.google.org/>), which is focused on analyzing coarse-grained data produced by sensors like Landsat and MODIS, and DigitalGlobe's Geospatial Big Data platform (<https://www.digitalglobe.com/>), which at present is focused on analyzing the fine-grained data produced by their own sensors. In both cases, the user can access data stored in the cloud and analyze it in place, which can be very fast and convenient. The simplest usage option in this for-profit environment is to pay for the compute time you need, but that can quickly deplete project funds, even when the expense is built into a traditional research grant—especially if you underestimate your project's resource requirements. Most companies have nonprofit arms that award time grants that are similar to those available on the scientific side of HPC, but the biggest differences between the two are that the awards tend to be a lot smaller and, more importantly, you are largely on your own to figure out how to use those resources to solve your problem.

Archaeological Supercomputing: Illustrative Use Cases

The goal of this section is to highlight several active research areas within archaeology that have benefitted from, or could definitely benefit from, the use of HPC hardware and software as described above. Specific projects will be discussed for each one to give the reader an idea of what has been attempted, where successes were achieved, and where challenges still remain. In other words, this is a tour of the state of art in archaeological supercomputing. To that end, the emphasis will be on what the researchers did, not what they specifically discovered. As a discipline, we are only beginning to take advantage of available resources, altering our thinking about the scale and scope of the questions we can ask and the problems we can solve. Once the tour is complete, this chapter will conclude with a discussion of the art of the possible, i.e., thoughts on where we can go from here.

Four areas will be highlighted here: landscape recording and reconstruction, terrain analysis, social network analysis, and complex adaptive systems. The examples of each discussed below do not, and cannot, cover the entire breadth and depth of archaeological supercomputing. They do, however, represent a reasonable cross-section of computationally intense problem solving and touch upon many topics of current interest to the discipline. One thread that ties all of them together is their geospatial focus. While that is likely not surprising to the reader, it is important to point out.

Landscape Recording and Reconstruction

The processing of capturing accurate, detailed three-dimensional information at the feature, site, and regional levels has undergone a revolution over the past decade. While traditional survey methods are still widely used, active scanning technologies like LIDAR and passive methods like photogrammetry-based Structure from Motion (SfM) are starting to play integral roles in most field projects (Opitz and Cowley 2013). Each is capable of producing massive numbers of precise point measurements, on the order of billions and trillions, that can be used to create models of the environment that can be analyzed and visualized in a wide variety of ways, a process that often involves fusing imagery, often collected simultaneously, to make the results more realistic-looking and to provide more quantitative depth. While the analytical and visual techniques employed are generally not novel, working with that much data is definitely a new frontier for archaeology, but has been explored at great lengths within the established HPC community—especially on the scientific side. Desktop software designed to exploit point clouds, as they are called, quickly breaks down at that scale. To make the situation even more challenging, archaeologists are working at multiple levels, from documenting excavations at individual sites to scanning huge swaths of jungle.

The two most noteworthy large-scale landscape recording LIDAR projects in recent years, ones that have pushed the boundaries of the discipline, are centered on the site of Caracol in Belize and Angkor Wat in Cambodia. Over two collection campaigns spanning several years, Arlen and Diane Chase have collected LIDAR data for more than 1200 km² of the triple-canopied Belizean rainforest, a region that they have been painstakingly, and slowly, exploring via traditional pedestrian survey for three decades (Chase et al. 2012, 2014). The resulting point cloud, consisting of trillions of points, required a great deal of trial-and-error processing via specialized software in order to remove the trees and underlying vegetation that was obscuring features of archaeological interest, ranging from agricultural terraces to roads and household groups. The result, even though far from perfect, was an incredibly detailed bare earth digital elevation model that could be run through traditional GIS software to generate standard products like shaded relief maps and hydrological flow models. They have also, through a hybrid parallelized application provided by the author of this chapter, created a Sky-View Factor map for the entire region, which is a substantial improvement over shaded relief due to its omni-directionality (Zakšek et al. 2011), as can be seen in Fig. 4. Not satisfied with stopping there, the Chases are now embarking on the creation of an automated framework for classifying features of interest in the terrain data to make map creation, and subsequent interpretation, possible for the entire region. Doing so manually would require many years and a substantial amount of funding, but the machine-learning-based automated approach carries a cost as well: required computing power. It simply cannot be done within a desktop environment. Damian Evans has faced similar survey challenges in Cambodia and was able to reveal a much larger, better organized, and more varied landscape surrounding Angkor Wat

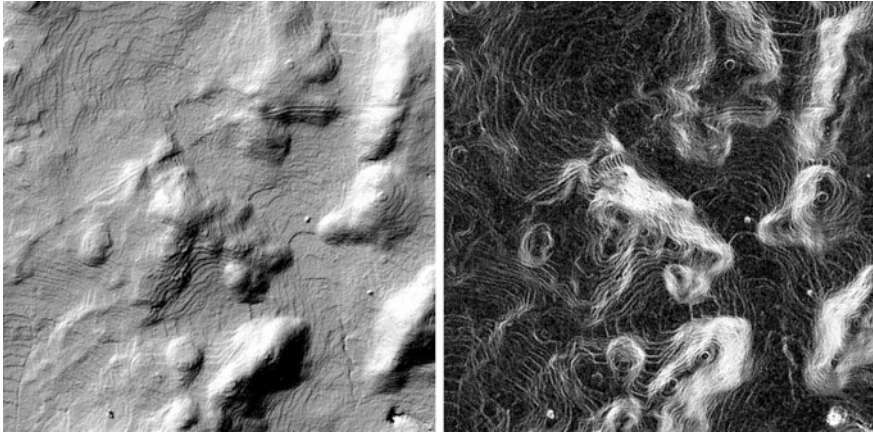


Fig. 4 High spatial resolution LIDAR data from Caracol, Belize, visualized using a traditional shaded relief method (*left*) and an inverted version of a more computationally intensive method known as Sky-View Factor (*right*). Note how many more potential features of interest are visible in the latter image

than anyone anticipated (Evans et al. 2013). The data collected for that project included next-generation full waveform LIDAR, but the computational requirements for fully exploiting it far outstrip the capacities of all but the largest computers in the world, so only a traditional point cloud was provided by the vendor. Full waveform, already shown to be valuable to archaeologists who are attempting to produce more accurate bare earth models in challenging terrain (Doneus et al. 2008; Lasaponara et al. 2011), will become more common in the next decade. What both projects clearly demonstrate is that this is just the beginning. LIDAR technology, deployed via manned aircraft and drones, is quickly dropping in price and it is only a matter of time before it is a mainstay for any field project.

Landscapes come in many sizes, though, and the challenges being faced by those working at the regional level are faced by those working at the individual site level, too. LIDAR data can be collected using a sensor mounted on a tripod, and can produce point counts of a similar volume, but what is far more practical for most projects on a budget is to use SfM (Structure from Motion) to record individual site features and, when possible, every aspect of an excavation (Green et al. 2014; Opitz and Cowley 2013; Remondino 2011). SfM uses specialized photogrammetry software and a large number of images of an object, taken from different perspectives, to reconstruct that object's three-dimensional characteristics and even build immersive environments. It requires a great deal of computing power if one wants the results for a small area quickly, or to do a large area at all. For an individual site, running the software on a desktop computer, which is common for archaeologists working with true color or thermal imagery collected by a low-flying drone to create digital elevation models and orthophotos, is usually sufficient (Casana et al. 2014). However, what happens when your area of interest is much

larger? The Center for Advanced Spatial Technologies at the University of Arkansas has experimented with running PhotoScan on their institutional scientific cluster in an ad hoc hybrid parallel framework, with mixed results, but the approach shows a lot of promise. As with the LIDAR examples discussed above, the amount of three-dimensional data being collected, and associated imagery, is only going to increase. In light of how inexpensive it is to produce mass quantities of photos, and how affordable handheld active scanners like Google's Project Tango (<https://www.google.com/atap/project-tango/>) are quickly coming to market, we will all be drowning in useful, multidimensional data. We have mastered many of the analytical and visual techniques we want to use through existing software, but there is a huge gap between where that software and expertise ends and where we want to (and need to) be as a discipline.

A special case of recording and reconstruction worth mentioning here is the CORONA Atlas of the Middle East (<http://corona.cast.uark.edu/>), a multi-year project that was originally designed to recreate the landscape of the Fertile Crescent during the time when now-declassified spy satellites collected imagery over the entire region (Casana and Cothren 2013). In many cases, sites visible in those images no longer exist, destroyed through processes as diverse as agricultural expansion, urban sprawl, and armed conflict. Working with the imagery is extremely difficult, but in partnership with a photogrammetrist, archaeologists were able to properly geolocate a large number of historical stereo pairs, which can now be publicly accessed and used to produce orthophotos and terrain models using the techniques mentioned above. Creating those usable images required a hybrid parallel computing framework that leveraged GPGPUs, where possible. There is a great deal more the authors want to do with the Atlas, including an expansion into new regions and creating an automated process for detecting unrecorded sites, but recording damage at known sites has become a more pressing focus (Casana 2014; Casana and Panahipour 2014), at least in the short term.

Terrain Analysis

Terrain analysis, specifically the extraction of meaningful information from analyzing digital elevation models, has been a staple of archaeological GIS for at least two decades. The most common products generated are slope and shaded relief (mentioned above), both of which are relatively fast to calculate and are relatively easy to move into HPC environments because the mathematical operations required are *embarrassingly parallel*, as in they can map well into strong, weak, or even hybrid parallel frameworks because each one is completely independent and not very complex. This is also the case for Sky-View Factor, but that approach does require more time to compute than the others. There are three types of analysis that are far from *embarrassingly parallel*, though, and they are growing in significance for archaeologists who are interested in how things (people, water, ideas, social connections, goods, etc.) flow across landscapes. They are focused on watershed,

peoplesheds, and viewsheds. It should be noted that all three are of interest to the researchers recording and reconstructing landscapes in the ways discussed in the previous section, in particular those struggling with how to effectively process large datasets without sacrificing fidelity.

The first is hydrological flow modeling, which has been used by archaeologists for many years, but requires a significant amount of computing power to execute and has been limited to the desktop environment, so regions of interest have remained somewhat small. A given terrain model is analyzed to find all high points from which will always flow downhill. Water is placed in those locations and the flow to all possible low points is modeled, after which a virtual stream network is extracted and its components are classified by rate of flow, creating a “realistic” representation of local watersheds along the way. While this can be done in traditional GIS software, that software cannot handle the size of landscapes that are now of interest to archaeologists, where “size” refers to how much data one has, not its physical geographic extents. Work has been done in the HPC domain to solve this problem, though, so archaeologists are encouraged to use tools like TauDEM (<http://hydrology.usu.edu/taudem/taudem5/index.html>), which leverages MPI and can be accessed through an ArcGIS extension, if needed.

The second is least cost analysis. What if one is interested in modeling how people, not water, flow across landscapes? Traditional GIS software allows for the generation of a small number of least cost paths across relatively small landscapes (it suffers the same technical limitations as hydrological analysis), but that is rarely the scale at which archaeologists are thinking about connections between, and travel to, locations of interest within a region. What if you are not sure where travelers are coming from or going to, but instead want a more general sense of how a landscape might channel movement, akin to water flow? The only feasible way to answer either of those questions, especially when one is working with a very large landscape, is to use some form of HPC. The From Everywhere To Everywhere (FETE) project, initially focused on the state of Oaxaca in Mexico, is doing just that (White and Barber 2012). The software written for FETE is capable of using strong parallelism on a desktop or hybrid parallelism on a cluster to quickly generate tens to hundreds of millions of theoretical travel routes across a region, which are then aggregated into a map that indicates the rate of people-flow, creating something akin to a “peopleshed.” If locations of interest are known, they can be used, but it starts with the assumption of no a priori knowledge and instead samples terrain to build up an understanding of how it directs movement. The more samples requested, the more complex the overall set of calculations. Figure 5, where all of Mesoamerica has been analyzed at relatively high spatial resolution to highlight potential pedestrian trade routes across the entire region, demonstrates some of what can be done with the approach. It is somewhat reminiscent of a circulatory system, which makes sense.

The third is viewshed analysis, which like the previous two, can and is often done to a limited extent with GIS software packages. The traditional approach is to pick a small number of points of interest, specify a visibility extent, and generate a map of what can be seen from which points. Also like the previous two, calculating

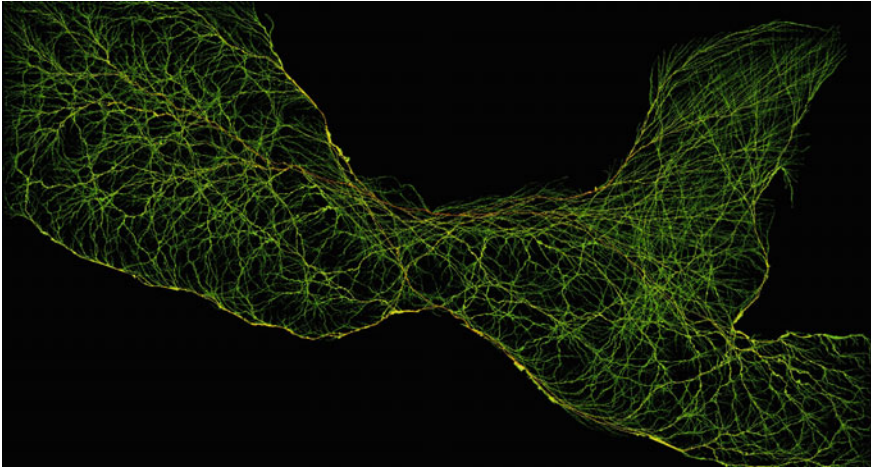


Fig. 5 Theoretical, terrain-based pedestrian trade routes throughout all of Mesoamerica, generated at relatively high spatial resolution using the FETE HPC application. Hundreds of millions of least cost routes were required to create the map. *Green* routes are high traffic, *yellow* are higher traffic, and *red* are highest traffic

a viewshed is computationally expensive, but the end result is extremely useful. As early as 2003, archaeologists began to hit a substantial computational wall with respect to viewshed analysis, expressing an interest in producing two types of products: *aggregate viewsheds*, where results are built up in a fashion similar to FETE through directed or systematic sampling, and *total viewsheds*, where results are built up for every single cell in the supplied elevation model (Llobera 2003). Figure 6 shows an example of an aggregate viewshed, created for a large region in the North American Southwest using a massively parallelized algorithm running on a GPGPU. In either approach, the main goal is to highlight prominent features on a landscape through sheer computational brute force, which requires HPC. To date, that brute force has been expressed most elegantly through the visual prominence research of Bernardini, who was interested in finding out which communities on a landscape could see the same prominent features and might then be considered part of the same “sight communities” (Bernardini et al. 2013; Bernardini and Peeples 2015), which could possibly share other things in common as well—despite great distances between them. Creating the baseline visual prominence map, derived from skylines extracted from viewsheds, involved deploying traditional GIS software in a weak parallel arrangement in a commercial-style cloud, a process that required a month to complete due to the inefficient, single-core nature of the software. Work is ongoing to translate the algorithms so that they can run within a hybrid parallel framework and leverage GPGPUs for the viewshed calculations, leveraging concepts developed by the broader Geographic Information Science community (Zhao et al. 2013). When complete, the anticipated speedup, and the overall extent of a region that can be analyzed, will be significant.

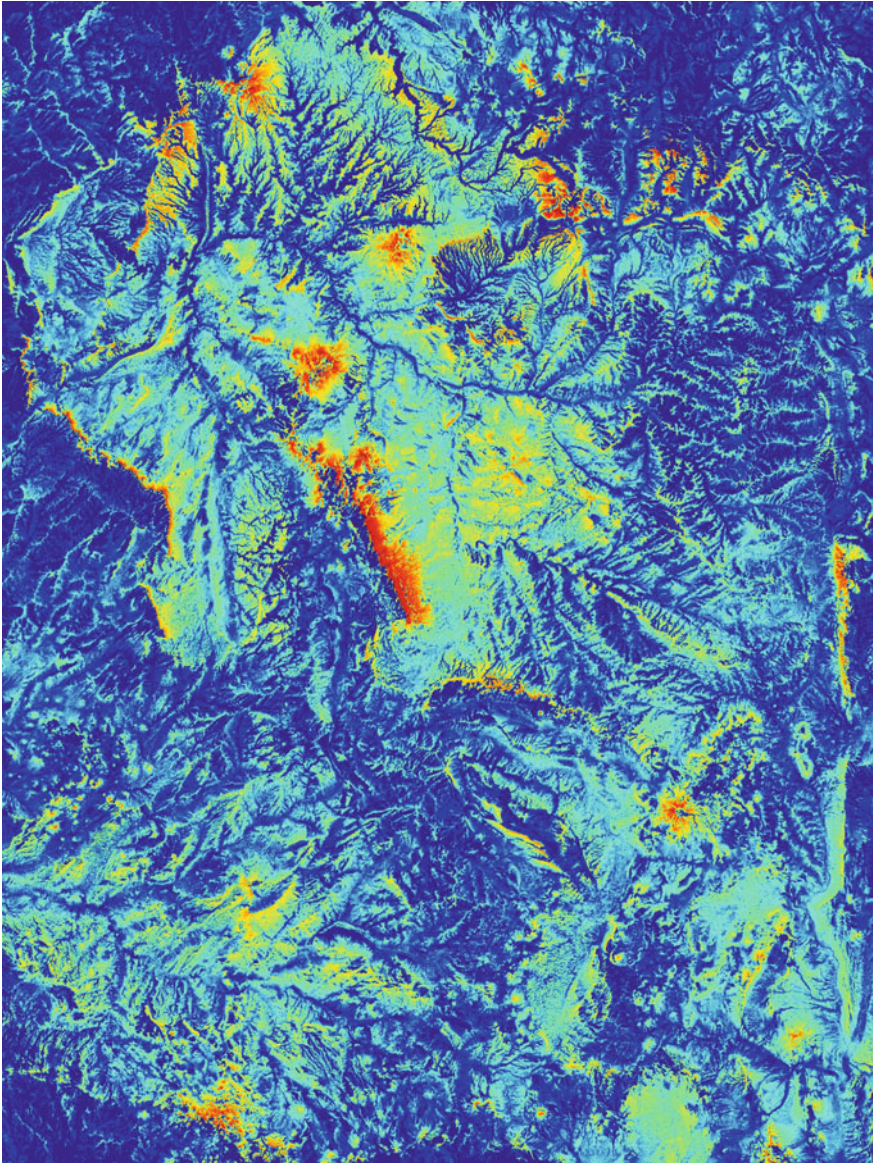


Fig. 6 An aggregate viewshed model for the entire Chaco regional system, which spans 158,000 km² in the North American Southwest. *Blue* areas are the least visible, *red* areas the most visible. Hundreds of thousands of viewshed analyses were executed using a GPGPU at regular spatial intervals, and then consolidated, to produce the output

Social Network Analysis

Social network analysis (SNA) has only recently been embraced by archaeologists, but the broader quantitative social science community has been exploring its utility for almost fifteen years (Borgatti et al. 2009) and it is an integral underpinning of Silicon-Valley-based social media outlets like Facebook and Twitter. Archaeologists have been able to benefit from those earlier explorations, as well as more recent research within the commercial sector, translating some of the core methods so that they have meaning within our discipline. At its heart, SNA is graph theory, which for the purposes of this discussion is a branch of mathematics and computer science concerned with how entities are connected to one another and how information flows between them. Given that archaeologists are interested in the flow of ideas, goods, and people between discrete locations, for example, using SNA would appear to be a natural fit for the discipline. The most substantial and ambitious efforts to date have come out of the Southwest Social Networks Project, which is examining community interactions (via a large standardized ceramics and architecture database) across space and time in the North American Southwest (Mills et al. 2013, 2015). Where their research, and ultimately all SNA projects, run into issues is when their graphs get so large that they cannot be processed on a single workstation or, if they can be processed, doing so requires a great deal of time. That is becoming increasingly common. As mentioned above, graph computing is a challenging problem, even for HPC. Many advances have been made, including the development of specialized computers and open source software packages like GraphLab and GoldenOrb, the latter of which is based on another Google standard named Pregel, but none of these are particularly easy for archaeologists to use at present. Where the situation becomes even more interesting is when SNA is combined with one or more of the terrain analysis techniques discussed above (Bernardini and Peebles 2015). As the reader has seen, each one is challenging on its own, but the results are highly complementary because they each speak to a different aspect of flow across a landscape. The use of SNA, in particular the combining of it with other more well-established quantitative methods, will continue to grow in the coming years and eventually become a common approach within the discipline. Making it a practical one will take time, though.

Complex Adaptive Systems

Modeling and simulating complex adaptive systems, like SNA, has quickly moved from a niche research space to one that is being more fully embraced by archaeologists as established projects have published their findings and clearly demonstrated its utility. The software required has also become much easier to use, with two packages being the most prominent at the time of writing: NetLogo (<https://ccl.northwestern.edu/netlogo/>) and Repast Symphony (<http://repast.sourceforge.net/>).

Both make it relatively easy to create what are known as agent-based models (ABMs), where complex (and nonlinear) interactions between people can be simulated over varying amounts of time and space. The most well-known ABM effort in archaeology to date is the Village Ecodynamics Project (VEP), a multi-year effort focused on understanding the formation and eventual abandonment of villages in the Mesa Verde Region of the North American Southwest (Kohler and van der Leeuw 2007; Kohler and Varien 2012). ABMs are by far the most computationally demanding entities discussed in this chapter. Communication between all of the elements of a model is required, so as the model scales to larger numbers of people and/or larger or more fine-grained space-time contexts, desktop software solutions quickly break down. NetLogo, by its own admission, is designed to be an educational tool, not a production-level solution, so models must be kept small. Repast Symphony can operate at the production level, but even it runs into technical limitations as a model grows in size. VEP has run into several issues related to computing capacity, which has constrained their ability to look at more regions at great levels of spatial and temporal detail. This is a common problem for archaeologists working with ABMs, who are generally interested in exploring the interactions of many people in very detailed ways. One solution is available at present, which is to employ an HPC-enabled version of Repast, unsurprisingly called Repast HPC (http://repast.sourceforge.net/repast_hpc.html). One archaeologist has already attempted to use Repast HPC to explore an entire Hohokam irrigation system in southern Arizona and the results to date are promising (Murphy 2012). As with many of the technologies discussed above, there is a steep learning curve associated with the framework, but the hope is that it will become more accessible in the coming years because the disciplinary need is definitely present.

Concluding Thoughts: The Art of the Possible

Roughly speaking, there has been a trillion-fold increase in computing power since 1956 (for more details, see <http://pages.experts-exchange.com/processing-power-compared/>). To put that increase in perspective, the newly-released Apple Watch is on par with a Cray-2 from 1985, which means we can now walk around with a thirty-year-old supercomputer on our wrist, one that can connect to a much deeper pool of computing power via the internet. Smartphones are even more powerful (fifteen to twenty times greater). Where will computing power be in thirty years? Will we be able to walk around with the equivalent of Titan on our wrists? What can archaeologists possibly do with that much power, given that we take so little advantage of what is already available to us on the scientific and commercial sides? Granted, it is not necessarily easy to use what is available today, but it is hoped that the survey provided above will open new doors for archaeologists and help them connect with the right technical resources who can help them, with the ultimate goal being empowered archaeologists, fluent in the languages of computer science,

helping themselves. We should, in the end, completely own our problems and their solutions.

A more important question to ask is this: If unlimited computing power was available, how would archaeologists interact with that technology, what questions could they ask, and what problems could they solve? As archaeologists broaden their regions of interest and/or examine smaller areas at increasingly finer levels of detail, with respect to both space and time, the amount, variety, and complexity of the data they must work with grows exponentially, which means that HPC will inevitably become a deeply integrated and transformative element of the discipline, much in the same way as radiocarbon dating and LIDAR/SfM. What would that look like?

A natural place to start would be the “grand challenge” topics proposed by Kintigh et al. in PNAS (2014), topics that until very recently could not be realistically explored at a global scale due to data sparsity, lack of sufficient technical expertise, and lack of available computing power:

- Emergence, communities, and complexity
- Resilience, persistence, transformation, and collapse
- Movement, mobility, and migration
- Cognition, behavior, and identity
- Human-environment interactions

Beneath each of these general topics are multiple questions and the reader is encouraged to consult the article for more detailed information (Kintigh et al. 2014). Archaeologists have attempted to address these topics in relatively small, focused ways over the past several decades, but momentum has built up recently to address them in a cross-cultural way that incorporates as much space and time as the extant archaeological record will allow. The recent work by Kohler on the spatially variable Neolithic Demographic Transition in the North American Southwest is an excellent example of this trend (Kohler and Reese 2014). That is not just a big data problem, it is a massive data problem, one that is fraught with peril due to the fragmented and inconsistent nature of global archaeological datasets. Two initiatives are taking the first steps towards creating a consolidated archaeological database for the world: the Digital Archaeological Record (<http://core.tdar.org/>), also known as tDAR, which is focused on archiving all types of critical archaeological data, from site reports to datasets, and the Digital Index of North American Archaeology (<http://ux.opencontext.org/blog/archaeology-site-data/>), also known as DINAA, which is focused on creating interoperability models between archaeological site databases, thus enabling analysis at larger scales and finer spatiotemporal resolutions. Between the two, assuming enough compute power is present, archaeologists can now ask regional and even continental-level questions in a way that was not possible previously. Whether tDAR and DINAA specifically persist is not really the point here: consolidated archaeological databases are the future and will enable researchers to finally, after centuries of trying, ask the really big questions in a quantitatively defensible way (Kintigh 2006). It is important to note,

however, that those topics are just a subset of a much richer tapestry of “big questions” that HPC and the data analytics it supports, in the hands of archaeologists, can address.

Putting aside grand challenges, there are practical areas where increasingly-available and increasingly-accessible HPC can and should transform the discipline. They are summarized here as a series of desires, something all archaeologists should want: everything global, everything detailed, everything mobile, everything fast, and everything smart.

What is meant by everything global is the desire to have the world’s archaeological knowledge base available at one’s fingertips for analysis, visualization, and contextualization, no matter the location, with the ability to contribute back to it in real time. Repositories like tDAR and DINAA are an excellent start, but the broader knowledge base should include the entire breadth and depth of the discipline. Given how compute power and storage capacities continue to rapidly increase, this is an achievable goal.

What is meant by everything detailed is the desire to have fine-grained spatiotemporal (four-dimensional) models of archaeological features, sites, regions, cultures, continents, and even the entire world. The standard can and should be the digital capture of archaeological information in multiple dimensions and the ability to immerse oneself in it at multiple scales, from virtually exploring the intricacies of an individual artifact to experience a reconstruction of an ancient city, complete with people. More nuanced reconstructions based on existing data is one path, but another is to more fully embrace the wide array of options now available for recording data in the field, not the least of which are smartphones, drones, Microsoft Kinect, and laser scanners. A logical extension of this desire would be the ability to test out hypotheses in virtual environments, seeing how events might play out under varying circumstances over large expanses of space and/or time. Many of the foundational technologies, including compute power, are already available to reach this goal, but the discipline (like so many others) currently lacks the technical expertise and resources to take full advantage of it. As costs continue to drop, and technological barriers continue to fall, the situation will greatly improve.

What is meant by everything mobile and everything fast is the desire to have HPC resources available at one’s fingertips, regardless of location, as quickly as possible (at the speed of research). That means being able to collect and analyze mass quantities of data in the field, perhaps in a disconnected fashion (no internet access), including accurate real-time recording and analysis of excavations by multiple sensors (terrestrial and airborne) and being able to use augmented reality displays while one works. This mobility and speed should extend to the lab environment as well.

Lastly, what is meant by everything smart is the desire to have the equivalent of IBM’s Watson for archaeology. Artificial intelligence research is currently undergoing a renaissance and human-trained systems like Watson are now able, in fields like medicine, to digest vast storehouses of information, find non-obvious connections between elements, and make suggestions to researchers, in close to real time. By connecting a system with this kind of potential to others that address the

previous four desires, archaeologists will be able to explore the past in ways that we cannot even possibly imagine.

At the center of this new type of exploration are archaeologists who are not just passive recipients of technologies and methods developed by others. We can, and should, harness the potential of HPC for ourselves. This chapter, only the latest step in that direction, has introduced the reader to the current landscape of HPC, how to take advantage of it, and some of what archaeologists have already attempted. It is very exciting to think of what could happen next.

Acknowledgments This submission was written by the author acting in his own independent capacity and not on behalf of UT-Battelle, LLC, or its affiliates or successors.

References

- Bernardini, Wesley, Alicia Barnash, Mark Kumler and Martin Wong 2013 Quantifying visual prominence in social landscapes. *Journal of Archaeological Science* 40:3946–3954.
- Bernardini, Wesley and Matthew A. Peeples 2015 Sight Communities: The Social Significance of Shared Visual Landmarks. *American Antiquity* 80(2):215–235.
- Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass and Giuseppe Labianca 2009 Network Analysis in the Social Sciences. *Science* 323(5916):892–895.
- Casana, Jesse 2014 Regional-Scale Archaeological Remote Sensing in the Age of Big Data: Automated Site Discovery vs. Brute Force Methods. *Advances in Archaeological Practice* 2(3):222–233.
- Casana, Jesse and Jackson Cothren 2013 The CORONA Atlas Project: Orthorectification of CORONA Imagery and Regional-Scale Archaeological Exploration in the Near East. In *Mapping Archaeological Landscapes from Space*, edited by D. C. Comer and M. J. Harrower. Springer, New York.
- Casana, Jesse, John Kantner, Adam Wiewel and Jackson Cothren 2014 Archaeological aerial thermography: a case study at the Chaco-era Blue J community, New Mexico. *Journal of Archaeological Science* 45:207–219.
- Casana, Jesse and Mitra Panahipour 2014 Satellite-Based Monitoring of Looting and Damage to Archaeological Sites in Syria. *Journal of Eastern Mediterranean Archaeology and Heritage Studies* 2(2):128–151.
- Chase, Arlen F., Diane Z. Chase, Jaime Awe, John F. Weishampel, Gyles Iannone, Holley Moyes, Jason Yaeger and M. Kathryn Brown 2014 The Use of LiDAR in Understanding the Ancient Maya Landscape. *Advances in Archaeological Practice* 2(3):208–221.
- Chase, Arlen F., Diane Z. Chase, Christopher T. Fisher, Stephen J. Leisz and John F. Weishampel 2012 Geospatial revolution and remote sensing LiDAR in Mesoamerican archaeology. *Proceedings of the National Academy of Sciences of the United States of America* 109(32):12916–12921.
- Doneus, Michael, Christian Briese, Martin Fera and Martin Janner 2008 Archaeological prospection of forested areas using full-waveform airborne laser scanning. *Journal of Archaeological Science* 35:882–893.
- Evans, Damian H., Roland J. Fletcher, Christophe Pottier, Jean-Baptiste Chevance, Dominique Soutif, Boun Suy Tan, Sokrithy Im, Darith Ea, Tina Tin, Samnang Kim, Christopher Cromarty, Stéphane De Greef, Kasper Hanus, Pierre Bâty, Robert Kuszinger, Ichita Shimoda and Glenn Boomazian 2013 Uncovering archaeological landscapes at Angkor using lidar. *Proceedings of the National Academy of Sciences of the United States of America* 110(31):12595–12600.

- Green, Susie, Andrew Bevan and Michael Shapland 2014 A comparative assessment of structure from motion methods for archaeological research. *Journal of Archaeological Science* 46: 173–181.
- Kintigh, Keith 2006 The Promise and Challenge of Archaeological Data Integration. *American Antiquity* 71(3):567–578.
- Kintigh, Keith W., Jeffrey H. Altschul, Mary C. Beaudry, Robert D. Drennan, Ann P. Kinzig, Timothy A. Kohler, W. Fredrick Limp, Herbert D. G. Maschner, William K. Michener, Timothy R. Pauketat, Peter Peregrine, Jeremy A. Sabloff, Tony J. Wilkinson, Henry T. Wright and Melinda A. Zeder 2014 Grand challenges for archaeology. *Proceedings of the National Academy of Sciences of the United States of America* 111(3):879–880.
- Kohler, Timothy A. and Kelsey M. Reese 2014 Long and spatially variable Neolithic Demographic Transition in the North American Southwest. *Proceedings of the National Academy of Sciences of the United States of America* 111(28):10101–10106.
- Kohler, Timothy A. and Sander E. van der Leeuw (editors) 2007 *The Model-Based Archaeology of Socionatural Systems*. SAR Press, Santa Fe, NM.
- Kohler, Timothy A. and Mark D. Varien (editors) 2012 *Emergence and Collapse of Early Villages: Models of Central Mesa Verde Archaeology*. University of California Press, Berkeley, CA.
- Lasaponara, R., R. Coluzzi and N. Masini 2011 Flights into the past: full-waveform airborne laser scanning data for archaeological investigation. *Journal of Archaeological Science* 38:2061–2070.
- Llobera, M. 2003 Extending GIS-based visual analysis: the concept of visualsapes. *International Journal of Geographical Information Science* 17(1):25–48.
- Mills, Barbara J., Jeffery J. Clark, Matthew A. Peeples, W. Randall Haas, Jr., John M. Roberts, Jr., J. Brett Hill, Deborah L. Huntley, Lewis Borck, Ronald L. Breiger, Aaron Clauset and M. Steven Shackley 2013 Transformation of social networks in the late pre-Hispanic US Southwest. *Proceedings of the National Academy of Sciences of the United States of America* 110(15):5785–5790.
- Mills, Barbara J., Matthew A. Peeples, W. Randall Haas, Jr., Lewis Borck, Jeffery J. Clark and John M. Roberts, Jr. 2015 Multiscalar Perspectives on Social Networks in the Late Prehispanic Southwest. *American Antiquity* 80(1):3–24.
- Murphy, John 2012 A High-Performance Computing Simulation of an Irrigation Management System: The Hohokam Water Management Simulation II. In *Archaeology in the Digital Era Volume II: e-Papers from the 40th Conference on Computer Applications and Quantitative Methods in Archaeology*, edited by G. Earl, T. Sly, A. Chrysanthi, P. Murrieta-Flores, C. Papadopoulos, I. Romanowska and D. Wheatley, pp. 225–230. Amsterdam University Press, Amsterdam.
- Opitz, Rachel S. and David C. Cowley (editors) 2013 *Interpreting Archaeological Topography: 3D Data, Visualisation and Observation*. Oxbow Books, Oxford, United Kingdom.
- Remondino, Fabio 2011 Heritage Recording and 3D Modeling with Photogrammetry and 3D Scanning. *Remote Sensing* 3:1104–1138.
- White, Devin A. and Sarah B. Barber 2012 Geospatial modeling of pedestrian transportation networks: a case study from precolumbian Oaxaca, Mexico. *Journal of Archaeological Science* 39:2684–2696.
- Zakšek, Klemen, Kristof Oštir and Žiga Kokalj 2011 Sky-View Factor as a Relief Visualization Technique. *Remote Sensing* 3:398–415.
- Zhao, Yanli, Anand Padmanabhan and Shaowen Wang 2013 A parallel computing approach to viewshed analysis of large terrain data using graphics processing units. *International Journal of Geographical Information Science* 27(2):363–384.