ArcheoSciences

Revue d'archéométrie

33 (suppl.) | 2009

Mémoire du sol, espace des hommes

“My Other Computer’s a BEAR”: A potential strategy for processing Large Volume Datasets

James Adock

Electronic version

URL: <https://journals.openedition.org/archeosciences/1641>

DOI: 10.4000/archeosciences.1641

ISBN: 978-2-7535-1599-4

ISSN: 2104-3728

Publisher

Presses universitaires de Rennes

Printed version

Date of publication: 30 October 2009

Number of pages: 247-249

ISBN: 978-2-7535-0943-6

ISSN: 1960-1360

Electronic reference

James Adock, “My Other Computer’s a BEAR”: A potential strategy for processing Large Volume Datasets”, *ArcheoSciences* [Online], 33 (suppl.) | 2009, Online since 30 October 2011, connection on 21 September 2021. URL: <http://journals.openedition.org/archeosciences/1641> ; DOI: <https://doi.org/10.4000/archeosciences.1641>

“My Other Computer’s a BEAR”: A potential strategy for processing Large Volume Datasets

James ADCOCK^{*,**}

Key words: GPR, Modelling, Landscapes, Super-computers, Remote access.

The use of increasingly complex visualisations, a desire for landscape-scale site evaluations, and the fusion of multiple data sources to draw more informed conclusions about a given site, is leading to exponential rises in computer processing demands, gradually pulling away from the capabilities of the average desktop computer. This paper looks at available resources for dealing with large volume datasets but by using an approach which is perhaps more widely accessible, rather than simply open to academic institutions or well-funded research groups. With the current financial downturn, any available technologies that shorten processing times and help projects stay within increasingly tight budgets (without reducing the quality of output) must be taken advantage of. Whilst the inspiration for this work came from working with ground penetrating radar (GPR) data, it should be obvious that the concepts and solutions discussed are equally applicable to the output from any large-scale geophysical survey.

Over the last 10 years, the increasing speed of desktop computers coupled with improvements in processing software algorithms has seen a vast improvement in the timescales involved in producing, for example, amplitude maps (time-slices) of GPR data. However, more recently

there have been parallel shifts in both processing and survey practices that are pushing processing times back up again. Firstly there is a move towards improved visualisations – for example the use of isosurfaces (Leckebusch, 2003) to form volumetric displays, (Fig. 1) requiring dense datasets. Secondly, there is now an increasing requirement to look at archaeological features within a wider context, rather than in isolation, leading to larger surveys and analysis at a landscape scale – even the reconstruction of entire palaeolandscapes (Fitch *et al.*, 2007; Chapman *et al.*, 2009). The use of multiple data sources, combined through weighted algorithms (Kvamme 2006), is likely to increase processing times further still and therefore now is the time to ensure the necessary access to the resources to cope with this are in place.

Although in the past it may have been preferable to process large datasets in a series of smaller areas to enable grid matching (to compensate for variations resulting from data collection over an extended period [Ernenwein and Kvamme, 2008]), with cart systems and multichannel arrays it is easier to collect large areas of data in a short time (Francesse *et al.*, 2009) and, or, multiple frequencies (Daniels, 2004). The inefficiency of repeating processing steps for individual sub-

* GSB Prospection Ltd, Cowburn Farm, 21 Market Street, Bradford BD13 3HW. (www.gsbprospection.com) (jimmy.adcock@gsbprospection.com)

** In association with VISTA Visual and Spatial Technology Centre, Institute of Archaeology and Antiquity, University of Birmingham, Edgbaston, Birmingham, B15 2TT. (www.vista.bham.ac.uk)

Figure 1: Volumetric models of (from left to right) Medieval bell-tower foundations at Salisbury cathedral (UK), mausolea structures at Binchester Roman Fort (UK), underside of Norman keep foundations with central pillar footings from Oxfordshire (UK).

grids is thus negated but the level of processing power required to deal with the survey data *en-masse* becomes an issue.

One obvious step would be to move away from standard “off the peg” desktop computers towards the more high-end machines, typically used for large scale rendering projects and gaming development (the primary driving force behind computing hardware development). Within the University of Birmingham’s VISTA centre such machines (small mixed architecture render farms featuring 32 processor cores and over 80GB of Physical RAM) are used for just this – for example the processing and subsequent rendering of laser-scanned point clouds to produce realistic, virtual-world reconstructions of sites. This paper demonstrates the gains that may be made through using these high specification computers by looking at the processing times of a series of test datasets. The test models comprise a relatively compact data set (only 30m by 85m) over a Roman villa, which has then been repeated to create ever larger (2x, 4x, 8x, 16x etc.) datasets (Fig. 2).

However, there are further, far more sophisticated systems available such as the BlueBEAR (Birmingham Environment for Academic Research); this is a cluster of CPUs comprising 384 dual-processor, dual-core (i.e. 4 cores/node), 64-bit 2.6 GHz AMD Opteron 2218 worker nodes giving a total of 1536 effective processor cores plus 3 quad-processor, dual-core (8 cores/node), 64-bit 2.6 GHz AMD Opteron 8218 nodes. The majority of the nodes have 8 GB of memory with 16 having 16 GB, whilst the quad core units have 32 GB of memory. There is also over 150 TB of user disk space available within the system. This ‘super-computer’ has seen applications as diverse as virtual crash-testing of motorcycle helmets, computational fluid dynamics, palaeoclimate modelling [http://www.bear.bham.ac.uk/documents/BEAR_case_studies_2.pdf] and could handle the processing

of geophysical datasets with ease. However, harnessing the immense processing power is not without initial difficulties; the software generally needs to be specifically written, or at least adapted, to take advantage of the multiple cores (processing needs to be split between the nodes) and this paper looks at how this may be achieved and the potential time gains that could be made.

The final part of the paper suggests how these systems may be made available to the wider research community; it is now obvious that the hardware is in place to deal with very large datasets and the associated multiple processing iterations required to get the best results, but accessing them is the difficulty. However, the use of remote access grids may prove to be the answer. Remote access is nothing new, the ability to control one computer from another is how many IT support companies manage to keep your desktop running without the need for costly site visits. That said, this same technology could also allow small research groups or units to tap into the kind of resources that an institution such as Birmingham has to offer. Booking time and processing power on either high-end machines or BEAR-type systems would allow for multiple iterations, batch data processing and the temporary storage of large volumes of meta-data, prior to the final output, from anywhere in the world (Fig. 3). Not only this, but provided one can find a means of sending the data back, long term projects in remote areas can take advantage of the system from the field – all that is required is a dial-up connection as very little information is being transferred over the net, merely a graphical display; the hard work and processing is all done on a node local to the processors.

Figure 2: Multiplying a GPR dataset from over a Roman villa (Caerwent, UK) to test processing speed.

Figure 3: Birmingham University's BEAR environment.

References

- CHAPMAN, H., ADCOCK, J., WOOD, E. and GATER, J., (under review). Mapping Buried Prehistoric Land Surfaces Using GPR and GIS and the Implications for Heritage Management. *Journal of Archaeological Science*.
- DANIELS, D., 2004. *Ground Penetrating Radar (2nd edition)*. The Institution of Electrical Engineers: London, 164-169.
- ERNENWEIN, E. and KVAMME, K., 2008. Data Processing Issues in Large-Area GPR Surveys: Correcting Trace Misalignments, Edge Discontinuities and Striping. *Archaeological Prospection*, 15: 133-149.
- FITCH, S., GAFFNEY, V. and THOMSON, K., 2007. In sight of Doggerland: From speculative survey to Landscape Exploration. *Internet Archaeology*, 22. Mesolithic Archaeology [http://intarch.ac.uk/journal/issue22/fitch_index.html]
- FRANCESE, R., FINZI, E. and MORELLI, G., 2009. 3-D High-Resolution Multi-Channel Radar Investigation of a Roman Village in Northern Italy. *Journal of Applied Geophysics*, 67: 44-51.
- KVAMME, K., 2006. Integrating Multidimensional Geophysical Data. *Archaeological Prospection*, 13: 57-72.
- LECKEBUSCH, J., 2003. Ground Penetrating Radar: A Modern Three-Dimensional Prospection Method. *Archaeological Prospection*, 10: 213-240.