Computing for the future of the planet

Andy Hopper, Andrew Rice and Alastair Beresford

Computers have transformed western society and look likely to do the same in the developing world. The challenge for the future is to harness the enormous power and potential of computing technology to generate a better understanding of the Earth and its environment and to provide low-impact alternatives to humankind's current activities. Simultaneously, both academia and industry must seek to minimise the environmental impact of computing so that the sheer abundance and energy consumption of technology does not threaten the goals of sustainable development.

The human species is living an unsustainable existence. The scientific consensus is that our planet will be unable to provide long-term support for us if we continue depleting the natural environment as we have over the past century or so. And since the global population is expected to grow from 6 billion today to 9 billion by 2050, and living standards are predicted to increase, managing our environmental impact is going to be vital to the future of society.

Over the last 60 years computers have transformed many aspects of our lives, and it seems likely that this transformation will continue, particularly in the developing world. Computing confers numerous benefits, but consumes a lot of natural resources.

Here we describe how computing could become be a positive force for reducing the environmental impact of humankind. First we consider the environmental impact of computing itself and the challenge of reducing resource consumption and waste. Secondly, advances in computing and networking will enable us to collect the sensor data needed to understand our world - providing crucial information for managing our planet in a sustainable manner. Thirdly, better models of the world will allow us to make more accurate predictions about our future. Finally, we believe that computing has the potential to offer less environmentally damaging alternatives to many of today's profligate consumer activities.

An optimal digital infrastructure

The power-inefficient nature of modern computers is apparent in their generation of waste heat and the costs of dispersing it. A large data centre must typically expend an additional watt of power on cooling for every watt of power supplied to a machine. Furthermore, around half the energy used over the typical three-year lifetime of a desktop computer is consumed in its manufacture, even if it is running 24 hours a day for its entire working life. The energy used by computers must be minimised throughout the lifecycle of manufacture, operation and disposal.[1] Chip manufacturers are already adopting green computing initiatives which seek to reduce the resource consumption and pollution of chip manufacture, and many charities now refurbish computers and mobile phones for reuse in the developing world.

Servers in large data centres typically operate at utilisation levels of 10-50%,[2] so for much of the time a machine has significant unused capacity. Some of this wastage is deliberate, enabling the system to cope with a sudden peak in load. Wastage also occurs because mission-critical software systems may interact in unusual ways and are therefore installed on separate physical machines. Virtualisation technologies offer a solution to both these problems by dividing a single physical machine into a number of virtual machines.[3] Multiplexing a number of existing servers onto a single physical machine leaves operators free to switch off unneeded hardware, while applications that have variable workloads can be scaled dynamically so that the resources an application uses depend on the current service load. Virtualisation of the popular types of server machines has only recently become viable and further hardware and software facilities are needed before it is ready for the desktop. However, industry is rapidly moving to adopt this technique in server infrastructure wherever appropriate.

Not all tasks require high-performance computing. A huge range of consumer electronics devices now exist - mobile phones, personal digital assistants, MP3 players - many of which are equipped with small, embedded processors. Whilst these processors are very efficient at performing the kinds of applications we typically use today, they are not very adaptable to possible applications of the future. They would not, for example, be suitable for the huge numbers of sensors needed to gather comprehensive data about the health of our planet. This type of application requires highly parallel processing - where lots of data can be processed independently and concurrently. Meeting these goals will require new computer architectures that can efficiently and dependably collate, process and distribute sensor data.

All computers need power. Currently the demand on our electricity grid varies massively depending on time of day, weather conditions, and even what is on television: a nationwide boiling of kettles often occurs during the adverts in popular soap operas. In addition, renewable energy sources often have variable supply levels, wind power being a particularly good example. Due to their significant energy demands, modern data centres are often built in areas of cheap electricity supply, such as near hydroelectric dams, rather than in areas of good digital connectivity. This reflects the fact that it is more efficient to transmit digital data than power over long distances. We believe that placing data centres next to renewable energy sources can exploit the flexibility of power control and virtualisation to smooth an increasingly variable supply and demand. Recently there has been interest in constructing data centres in Iceland due to its plentiful geothermal energy and naturally cold climate. Perhaps in the future every windmill will have a server farm attached, and every village in sunny parts of the world will provide low-cost hosting services.

Charging mobile phones at the village power centre in Kiangombe using hybrid renewable energy

There has been a huge uptake in broadband internet connectivity in developed nations. Yet, while 66% of the UK population has internet access, globally the figure is only 19% .[4] Efforts are being made to reduce this gap. Mobile phone technology has emerged as a valuable technology for many countries with low levels of access. In these regions fixed-line infrastructure is often inadequate, and electricity supply is unreliable, so battery-powered mobile phones are the only feasible means of long-distance communication. Street businesses that recharge phones from car batteries have been quick to spring up to support this new activity. Fixed broadband wireless systems have the potential to deliver high bandwidth data connections over large distances without the need for expensive and environmentally damaging laying of cable - an ideal technology for connecting remote and isolated regions. Research is ongoing to improve data rates, reduce power consumption and overcome interference between radio signals.

Nigeria: Phone vendor

Delay-tolerant networks are another way to improve access to digital data.[5] Such networks operate under the assumption that communicating nodes are mobile, unreliable and are frequently disconnected - as mobile phones are in weakly provisioned areas. Delay-tolerant networks could be used to deliver email to destinations with unreliable network connections. Alternatively, vehicles could be used as carriers. With delay-tolerant networks, deliveries to outlying areas can include both physical and digital resources - the local bus service can now deliver passengers, email and web pages. One form of this idea is already operated in rural India, Rwanda, Cambodia and Paraguay, where buses cache website data and relay it to computers with no direct internet connection.

Another example of the challenges of delivering computing to people in developing countries concerns the management of pictures from camera phones. Many people in Africa have a mobile phone with a camera, but no means of printing or displaying these pictures. In response, researchers at the University of Cape Town developed a distributed photo presentation application which allows sharing of photos between phones. In addition to being highly popular, this might be useful for people meeting in public places in developed countries. Similarly, 'ruggedised' low-power phones or computers designed for Africa have also generated significant interest in other continents. As researchers strive to minimise the power and resource consumption of our digital infrastructure, its very likely that technology and solutions inspired by the challenges present in the developing world will eventually be deployed more pervasively, helping us all do more with less.

Sensing the planet

We live in complex societies constructed within complex ecological systems. A detailed understanding of these systems can help us make informed choices about how to protect or improve them. Crucial to this understanding is the collection, dissemination and analysis of sensor information about the physical world.

One small-scale application of this approach is seen in 'sentient computing', which attempts to improve human-computer interactions by providing machines with information about people and their surroundings.[6] This can help the computer model the outside world and anticipate to some extent what will happen in the future. Sentient computing applications often rely on an internal model containing details about the current state of the world: in an office environment this might include the positions of walls, doors, desks and computers. Keeping this model up to date is increasingly difficult as its size and fidelity increases. Privacy can also be an issue: for example, location data collected from GPS receivers attached to vehicles can be used to assess road congestion, but can also reveal which hospital clinic you have visited or where you went after work.

Testing a G3 Communication device, Khartoum

Transportation is another area that can benefit from remotely collected data. Prototype intelligent vehicles incorporate numerous communications mechanisms, low-level sensors, cameras and touch-screen displays. These have the potential to improve in-vehicle satellite navigation systems. Mapmakers face an ever-increasing struggle to maintain the quality of their data: incorporating new routes, deleting closed roads and providing information about road quality and size. Appropriate processing of GPS position traces from augmented vehicles may make it possible to build these maps autonomously: raw journey traces from many journeys (and vehicles) can be aggregated to form a working map of the road network.[7]

Sensing has the potential to play a huge role in the developing world. Widely distributed autonomous sensors could collect and relay detailed information about water or soil quality, or use of such resources, to support effective planning. Sensors could be used to provide a clearer picture of illegal logging, or even a way to locate felled timber. Researchers have tackled diverse projects ranging from tracking the habits and movements of wild animals to enhancing water management using satellite imagery.

These projects are a step along the way to sensing the entire planet,[8] something which would enable ecologically minded consumers to find out the true cost of their purchases, the impact of emissions-changing initiatives, and improve our general understanding of the world in which we live immeasurably. Online mapping services already include real-time congestion information of our road networks and research projects such as TIME at the University of Cambridge and SenseWeb at Microsoft Research aim to integrate a plethora of additional sensed data. In the future, global sensor information might provide new data layers for these services, such as an infrared layer showing heat loss from buildings and a pollution layer showing current air quality conditions.

Nevertheless, there are substantial technical obstacles to global deployments of sensor technology, ranging from the environmental impact of sensor construction and deployment to power requirements and recovery of data. It will also be a challenge to exploit this technology without compromising privacy or individual rights.[9]

Modelling the planet

While it is now generally accepted that increasing levels of greenhouse gas emissions are altering the climate, determining the consequences of emissions is a difficult task reliant on large-scale computer models of the climate. These models are only destined to get more complex as they are refined and the amount of sensor data collected increases.

An important and largely unexplored question is whether the implementation of the model correctly captures the intent of the engineer. The impact of some nuance in the model could be masked by a programming error. Theoretical computer science has long investigated techniques for improving the reliability of software and for ensuring that it meets the intent of its designers. Examples include model checking, in which one specifies certain properties about a program and then verifies that they always hold. In the context of physical models, a programmer might wish to check that the program can never generate a state known to be impossible or that global properties such as conservation of energy are preserved by the system. Resolving these issues is a current research problem. We aim to explore how these techniques can be applied to the implementation of physical models to make sure that we can discover the effects of natural processes rather than the effect of a programming error.

Digital alternatives

The rapid evolution of computing has produced a huge change in our society, particularly with the emergence of the internet. It has done so principally by competing with traditional activities and providing compelling new alternatives. Digital activities potentially have a much smaller environmental impact than similar activities in the physical world.[10] At the same time, a shift from physical to digital is often preferred by consumers. This is a win-win situation: our environmental impact is reduced and our convenience increases.

Consider music sales from a conventional store or via the internet. With the latter, there is no physical storefront to maintain and no physical media to manufacture or transport. Online music is also undeniably popular, with an estimated US$2 billion spent on downloadable music online in 2006.[11] Online grocery shopping also compares favourably to visits to the supermarket in a car: deliveries to the same neighbourhood may be consolidated and the shop itself no longer requires any facilities to support the consumer. Digital alternatives are already emerging in the developing world too, where mobile phone networks are erected instead of fixed-line services, and air-time is used as the basis of trade in lieu of traditional currency.

Morocco: Cyber Café

Of course, the digital infrastructure has an environmental and economic cost, but this should become less significant with improvements in efficiency and as the number of online activities increases. The interesting fact about online shopping is that it is rapidly growing in popularity without environmental legislation. For many people, the online option is simply more compelling than the conventional alternative - the environmental benefit is an unexpected bonus. However, a new technology can also have a negative impact on the environment. For example, many companies have seen the use of teleconferencing increase the demand for travel because people have a desire to meet their co-workers in person. Computing is a tool which can reduce our impact or allow us to increase it. We must be vigilant, and examine in detail how computing is used and measure the emergent effect.

The shift from physical to digital is not restricted to goods, but can also be applied to people and services. Online virtual worlds have recently experienced explosive growth - by April 2008 Second Life had 13 million registered residents. Large corporations are exploring the use of virtual worlds to help employees collaborate and communicate without leaving the office. Some run their own worlds within their private networks, others rent space in worlds such as Second Life. The virtual worlds of the future may well provide an immersive environment in which we shop, socialise, travel and create wealth without significant burden on our planetary resources. A virtual world might even be able to support virtual tourism - for example, a local guide could take a mobile computer together with a high-resolution video camera and give a personalised tour of a distant place of interest.

Conclusion

The Cambridge EDSAC (electronic delay storage automatic calculator), one of the world's first computers, ran its first program in 1949. Since then, computing has changed itself and society beyond all recognition. Our vision is of another revolution, which we hope will benefit every person on the planet: with an optimal digital infrastructure we seek to improve the efficiency of our existing computing platform but also to manage the demands of computing to meet the constraints of our future energy supply. We envisage global sensor networks that help us optimise our use of resources. Theoretical analysis of computer models should provide scientists with a more accurate and trustworthy view of the future. And compelling digital alternatives to current activities could reduce mankind's environmental impact while providing equally rewarding experiences.

Computing provides the opportunity for an unbounded upside. If everyone can shift their activities to a virtual world then there is huge potential for growth and innovation whilst reducing environmental impact. To achieve this we must ensure that the world's computing platform is efficient, effective and appropriately applied.

Andy Hopper, Andrew Rice and Alastair Beresford are in the University of Cambridge Computer Laboratory, Cambridge, UK. Andy Hopper was the founder of Acorn Computers and has launched numerous high-tech start-up companies. He is a Fellow of The Royal Academy of Engineering and was made a Fellow of the Royal Society in 2006.

References

1. Williams E. Energy intensity of computer manufacturing: Hybrid assessment combining process and economic input-output methods. *Environmental Science and Technology* 2004; 38(22): 6166-74
2. Blazek M, Chong H, Loh W, Koomey JG. Data centers revisited: assessment of the energy impact of retrofits and technology trends in a high-density computing facility. *Journal of Infrastructure Systems* 2004; 10(3): 98-104
3. Barroso L A, Hölzle U. The case for energy-proportional computing. *Computer 2007*; 40(12): 33-37
4. www.internetworldstats.com/stats.htm [accessed 12 August 2008]
5. Pelusi L, Passarella A, Conti M. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *IEEE Communications Magazine* 2006; 44(11): 134-41
6. Hopper A. Sentient Computing: The Clifford Paterson Lecture 1999. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences* 2000; 358(1773): 2349-58
7. Davies J J, Beresford A R. Distributed, vehicular computation for map generation. Presentation to the Annual Meeting of the Association of American Geographers 2007, April 2007
8. Gibbons P B, Karp B, Ke Y, Nath S, Seshan S. IrisNet: An architecture for a world-wide sensor web. *IEEE Pervasive Computing* 2003; 2(4): 22-33
9. The Royal Academy of Engineering. *Dilemmas of Privacy and Surveillance: Challenges of technological change*. London: The Royal Academy of Engineering; 2007
10. Reichart I, Hischier R. The environmental impact of getting the news: A comparison of on-line, television, and newspaper information delivery. *Journal of Industrial Ecology* 2002; 6(3-4): 185-200
11. International Federation of the Phonographic Industry. *Digital Music Report 2007*. London: International Federation of the Phonographic Industry; 2007