Compulsory

Key words to describe the work:
Optoelectronics, Smart-Pixels, Parallel Computing

Key Results:
A model of parallel computing suitable for analysing the properties of a high bandwidth optoelectronic interconnection

How does the work advance the state-of-the-art?:
Construction of a model linking the BSP computational model parameters with the physical parameters of the system.

Motivation (Problems addressed):
Determination of the best way to utilisation the large bandwidth available optically.

MODELLING OF OPTICAL INTERCONNECTS FOR PARALLEL PROCESSING

G. Russell , J. Snowdon, T. Lim, I. Gourlay, P. Dew

Abstract

As the use of computers in simulation and data gathering and analysis become more important tools in science and engineering parallel processing machines are required to tackle the ever more complicated problems. The original interest in this field was in large custom built parallel machines but, as commodity components have become cheaper and more powerful, interest has moved to computational clusters built from PC type components.

In all areas of parallel computing the advances in available technology have made it possible to consider systems with more and more processors. Critical to large systems is the performance of the interconnect network. However, physical limits on both the data rate and connectivity of electronic interconnects are likely to limit performance. Therefore, it is useful to consider an optical interconnection scheme as the high data rates available and 3D connectivity make this an attractive proposal.

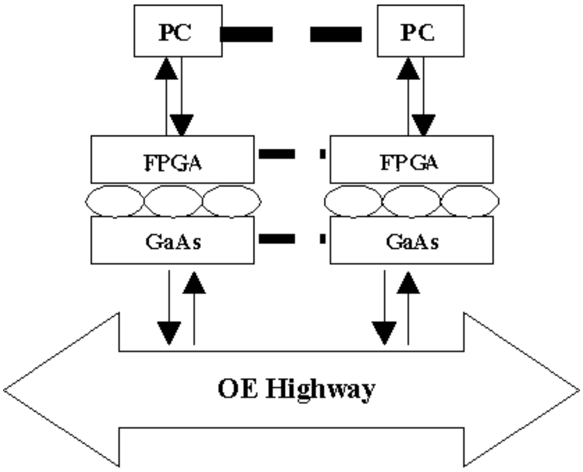


Figure 1 – Architecture for an optically connected PC cluster with FPGA programmable logic chips solder bump bonded to a GaAs optical chip.

Based on the above ideas a PC cluster with an optical interconnect system will be considered[1]. To interface a PC cluster with an optical communications backbone, a smart-pixel[2] based layer is need. This layer can not only provide the electronic to optical conversion but can also perform computation or data control. This architectural model is shown in Figure 1.

Several models have been developed to describe parallel algorithms for example, the Bulk Synchronous Parallel (BSP) model [3]. BSP describes the computational and communication costs in terms of parameters which describes the computer system. It separates an algorithm into a number of supersteps that contain as much processing as possible then a barrier synchronisation at the end of each superstep during which all communications between nodes occur.

Gordon Russell is a first year PhD student in the Department of Physics at Heriot-Watt University, Edinburgh
John Snowdon is a Lecturer in the Department of Physics at Heriot-Watt University, Edinburgh.
Theo Lim is a Research Associate in the Department of Physics at Heriot-Watt University, Edinburgh
Iain Gourlay is a Research Fellow in the Informatics Research Institute, University of Leeds.
Peter Dew is a Professor of Computer Science at the University of Leeds and Director of the Informatics Research Institute within the School of Computing.

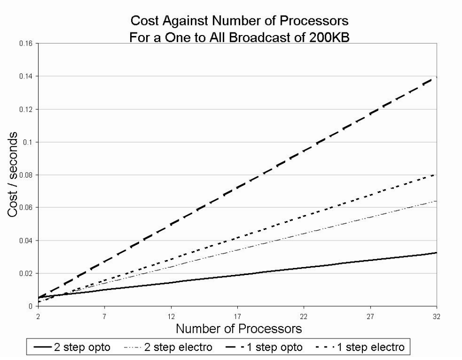
The aim of this work is to map the BSP parameters to physical parameters in order to investigate the performance of different interconnect topologies and functionality in the SPA layer can be investigated.

This has led to the formulation of a cost model in terms of the bandwidth, latency, connectivity of the interconnection system and the processor speed. The model considers the PC and its associated SPA to be a BSP node with data transfer allowed been them during the a BSP superstep and SPA synchronising at the barrier stage.

Currently, the proposed model suggests a substantial performance gain in problems where the PC can process small blocks of the total data on the SPA while waiting for the next block to be sent or received by it. If the processing time is greater than the PC to SPA communications time then the bottleneck can be totally removed. Examples of such problems are sample sort algorithms, weather forecasting and graphics applications.

If the SPAs have some simple functionality a performance can be enhanced if the operation on the SPA results in a small data-flow from the PC becoming a large data-flow on the optical highway and back before returning to the PC. This approach can often reduce latency at the cost of increasing the require bandwidth. An example of this would be the one to all broadcast.

Graph 1 shows the modelled cost for a medium sized one to all broadcast using two algorithms. The 1-step method is simply one machine sending all the data to each machine in turn. This method is generally poor as only one node is doing all the work. A better method is to split the message up into blocks, send a block to each node, and then have each node send its block to all others. In the optical case the blocks can be sent to the SPAs which can be communicated amongst them without involving the PCs. The calculations were carried out using data from an existing PC cluster and the optical parameters were taken from previous optical demonstrator systems such as SPOEC [4].



Graph 1 – Graph showing the modelled time cost for a 200KB One to All broadcast using one- or two-step algorithms on a conventional fast Ethernet PC cluster and the same cluster with a fast optical highway.

Clearly the graph indicates that a 2-step optically connected broadcast would out perform the equivalent electronic broadcast. Note that the 1-step optical broadcast is the slowest way of doing the broadcast. This is due to the time cost involved in carrying out the electrical to optical conversion and back again.

Although the 2-step optical system has to pass more data around the network than the electrical 2-step but transfers latency costs to the SPAs which for large numbers of processors are likely to be lower. As the bandwidth of the optical interconnect is so high and the PC latency is so big the extra data transmission time is much less than the time saved in latency costs.

Future work includes analysing various applications and designing a possible demonstrator system.

References

[1] B. Layet, J.F. Snowdon. Comparison Of Two Approaches For Implementing Free-Space Optical Interconnection Networks. Accepted for publication. January 2001.

[2] A.C. Walker, T.-Y. Yang, J. Gourlay, J.A.B. Dines, M.G. Forbes, S.M. Prince, D.A. Baillie. D.T. Neilson, R. Williams, L.C. Wilkinson, G.R. Smith, M.P.Y.Desmulliez, G.S. Buller, M.R. Taghizadeh, A. Waddie, I. Underwood, C.R. Stanley, F. Pottier, B. Vogele, and W. Sibbett. Optoelectronic Systems Based On InGaAs-Ccomplementary-Metal-Oxide-Semiconductor Smart-Pixel Arrays And Free-Space Optical Interconnects. *Applied Optics*, Vol. 37, No. 14, 10 May 1998.

[3] D.B. Skillicorn, J.M.D. Hill and W.F. McColl. Questions and Answers About BSP. Technical Report PRG-TR-15-96, Oxford University Computer Laboratory, 1996.

[4] A.C. Walker, M.P.Y. Desmulliez, M.G. Forbes, S.J. Fancey, G.S. Buller, M.R. Taghizadeh, J.A.B. Dines, C.R. Stanley, G. Pennelli, A.R. Boyd, P. Horan, D. Byrne, J. Hegarty, S. Eitel, H.-P. Gauggel, K.-H. Gulden, A. Gauthier, P. Benabes, J.-L. Gutzwiller and M. Goetz. Design And Construction Of An Optoelectronic Crossbar Switch Containing A Terabit Per Second Free-Space Optical Interconnect.

IEEE Journal Of Selected Topics In Quantum Electronics, Vol.5, No.2, 236-249, March/ April 1999.