Modeling and Simulation of Cache Access to Utilize a High Bandwidth Optical Processor-Memory Bus

G.A. Russell*, K.J. Symington, T. Lim, and J.F. Snowdon

Department of Physics, Heriot-Watt University, Edinburgh, EH14 4AS, Scotland, UK
*Email: g.a.russell@hw.ac.uk

A planar optical interconnect will be proposed as a method of implementing an interleaved memory bus. Analysis and simulation will show the design of caches can be improved due to the higher bandwidth and connectivity.

1. INTRODUCTION

Current processors, both commodity and specialized, offer impressive computational performance. However, the growing disparity between processor data requirements and memory throughput is limiting their potential. A way to alleviate this is to use interleaved banks of memory with the data partitioned across it. The data can then be read in parallel and the aggregate bandwidth is utilized. Such a system is not scalable electronically due to the high pin densities required and relatively high channel bandwidths [1].

The use of the 3D connectivity in free-space optics is well documented [2] and allows for the high pin density, high channel bandwidth interconnect required to make large interleaved banks possible. Such interconnects, although interesting for research, are not suitable for a commercial system due to alignment problems. It has been suggested that using a thin glass plate as a planar interconnect medium [3] may alleviate the problem.

In planar waveguides, light can be guided by internal reflection within a thin glass plate. This acts as a free-space optical system but folded up on itself. This gives the same geometric advantages of free-space but in a more mechanically stable package. Figure 1 shows schematically how a 4-f imaging system would be implemented.

Hybrid Smart Pixel Arrays (SPA) are a likely technology to provide the optoelectronic interface for such systems [4]. These integrate silicon CMOS for processing and switching and GaAs optically active elements such as VCSELs. Such devices have been proven to 32×32 pixel arrays [5] and are soon expected to reach greater than 128×128 arrays.

This paper envisages the use of a thin, glass, waveguide optical interconnect to implement an interleaved memory system and presents the results of computer simulations of memory access patterns.
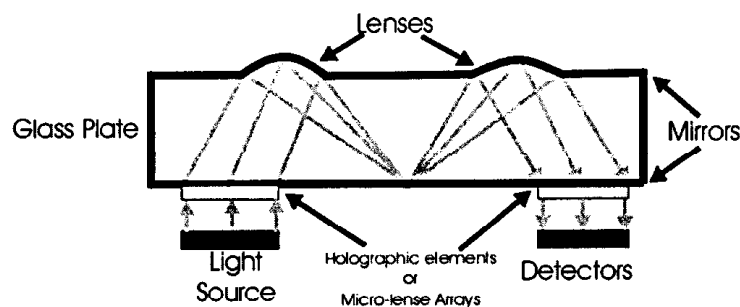

Fig. 1. A schematic of a planar implementation of a 4-f imaging system. In a real system the air-glass interface would act as the mirrors and the lenses would be defractive not reflective elements.
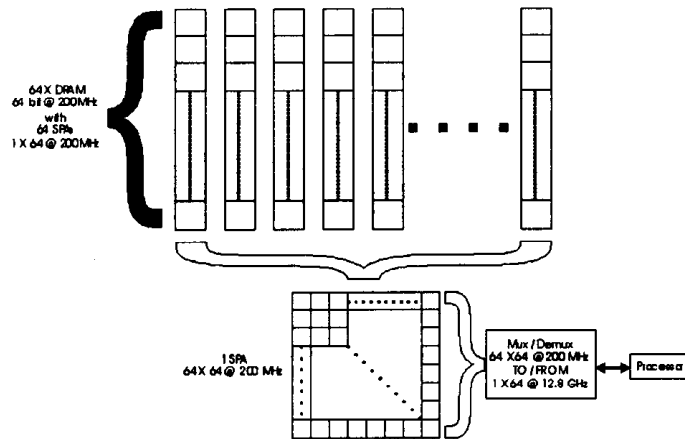
Fig. 2. Schematic of optical memory bus. 64 banks of standard memory (64 bit wide SDRAM) are connected in parallel to a single 64 × 64 SPA chip which carries out the necessary mux/demux functions to provide the processor with a high bandwidth memory access at 64 bits wide.

2. THE SYSTEM

Potentially high density interconnects can be realized via optics. In such a system it would be feasible to envisage up to 64 banks of memory running at 64 bits wide and 200 MHz (DDR-SDRAM for example) as shown in Fig. 2.

Each memory bank in Fig. 2 would have an associated 1 × 64 SPA that would route the data from each bank out of the plane and back down to a 64 × 64 SPA on or near the processor. If 1 bit of each 64-bit-wide word is on each bank, the data will be imaged onto the SPA in such a way that the data can be transferred electronically to the processor in parallel with a simple shift register similar to that implemented in the SCIOS sorter [6]. A similar scheme would be used to write the data back to memory. This would provide a memory bandwidth of 820 gigabits per second. This is greater that any current processor's internal bandwidth, so the number of banks can be decreased or multi-processor machines can be considered. In this paper, it is assumed that the SPA layer has an adequate number of high-speed buffers to allow the processor to utilize a large percentage of the possible peak bandwidth.

3. RESULTS

The following results were obtained by using data from current commodity computer components and previous optoelectronic demonstrator systems as the input to a custom computer simulation. The simulation assumes a simplified memory hierarchy of processor registers, level 1 cache and main memory. No processing is assumed, i.e. the data is only loaded to the registers and ignored. The results are scaled to the clock-speed of a 1 GHz machine for ease of presentation.

Graph 1 shows the time in clock cycles for data to be transferred. The bottom line is for level 1 cache and the top a is standard single bank DRAM. The middle line shows the interleaved memory with 64 banks. It can be seen that the transfer time is almost constant at the start-up latency for the interleaved case. Even considering a conservative penalty for electrical to optical to electrical conversions of 10%, the interleaved memory greatly outperforms standard memory, and at large transfers starts to be better than cache.

It is interesting to note that due to the almost constant nature of the data transfer time on-chip caches can be designed differently. Normally caches are divided into a number of cache-lines or blocks of memory. When a memory request is made and the location is not in cache, it is recovered from the main memory with a whole cache-line of data. The time for this is the miss penalty. It is accepted [7] that increasing cache-line size increases processor memory performance, however the larger the cache-line the bigger the miss penalty and a balance must be reached. If the data transfer time is constant, the miss penalty does not change with cache-line size and so they can be made as large as required.

Graph 2 shows a simulated memory to processor access in a processor with level 1 cache with 128 byte and 256 byte cache-lines for single bank and interleaved memories. The problem considered is loading a 5 by 5 matrix
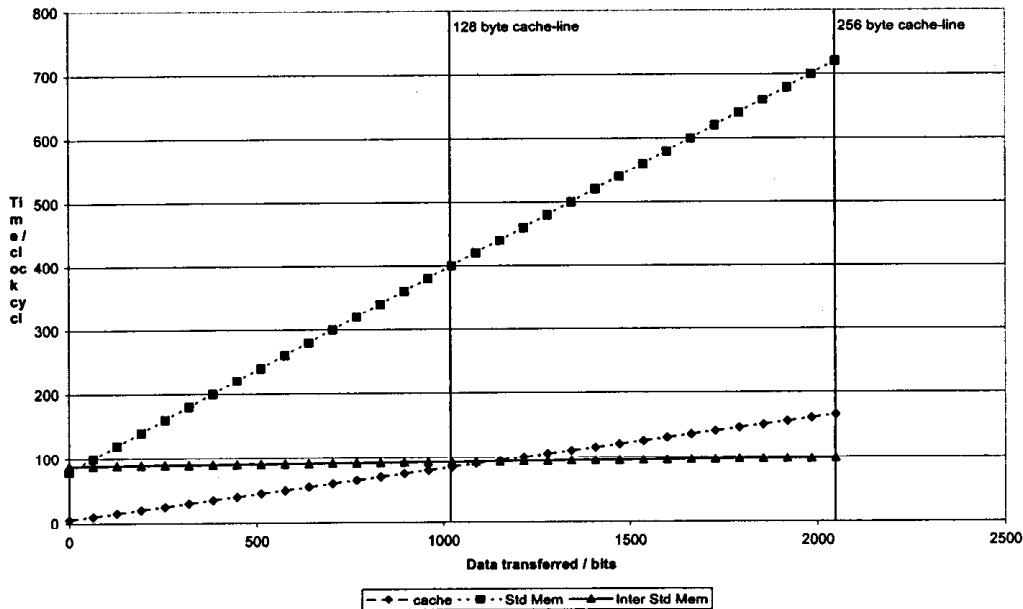
Graph 1. Simulated times for block data transfers for level 1 cache, standard memory and interleaved memory. The vertical lines show the transfer times for blocks of 128 and 256 bytes, i.e., cache-line sizes.
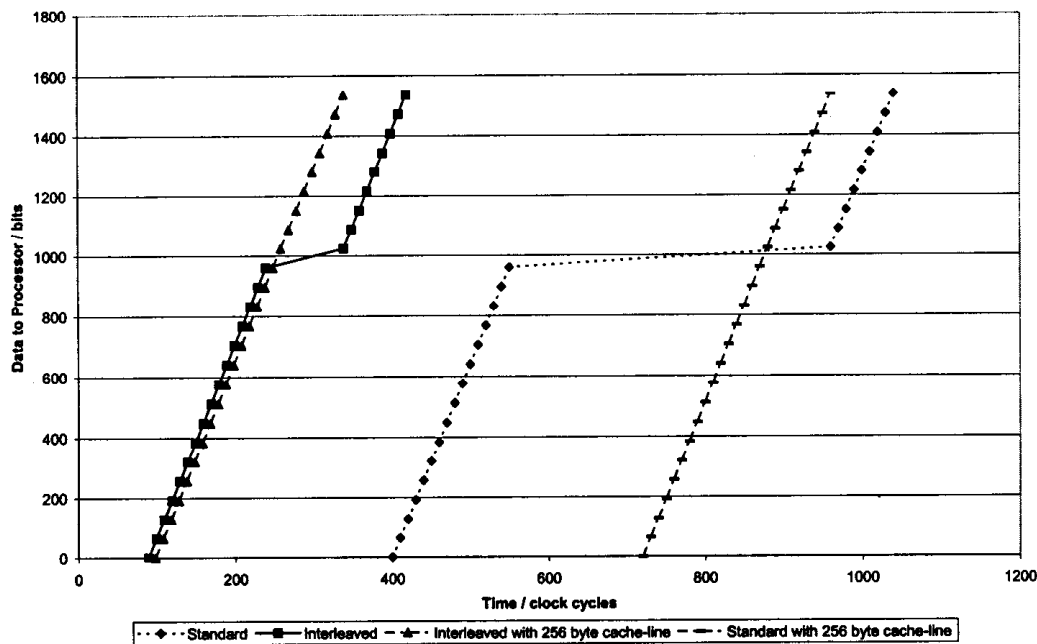
Graph 2. Graph of simulated memory access times for standard and interleaved memory with 128 and 256 byte cache lines for reading to the processor the rows of a 5 × 5 matrix partitioned row-wise. The step-like behavior is due to the data requested by the processor not being in the cache and the penalty from retrieving the data from main memory.

of double floating point numbers into the processor registers from main memory. It is assumed the matrix is not in the cache and there is no level 2 cache. The step-like features are due to the data being retrieved from the main memory (the plateaux) before being transferred internally by the cache. The optical solution, with faster main memory access has smaller plateaux and, therefore, completes the whole memory access in less time.

The previous example has maximum data coherence, i.e., cache is at maximum efficiency as data is in a continuous block. If the problem is to access 25 noncoherent double floats, i.e. a miss on every request, the total access time is shown in Table 1.

Table 1. Simulated results for a load to processor of 25 nonlocal double float (64 bit) numbers, i.e., totally noncoherent.

System Type	Cache-line/bytes	Total Access Time/clock cycles
Single Bank	128	10250
Single Bank	256	18350
Optically Interleaved	128	2475
Optically Interleaved	256	2700

This clearly shows the penalty due to a large cache line. When data is not coherent in the standard memory hierarchy, doubling the cache-line size almost doubles the penalty (79% increase). The interleaved memory only has a penalty of 225 cycles (9.1% increase) because the start-up latency, not the time for transfer, dominates.

Real applications have memory access patterns between the two extremes examined above. A good example is in matrix-matrix multiplication. As matrices are usually partitioned row-wise in memory reading in the rows of the first matrix is the same as the first case above. Reading the columns will depend on the size of the matrix. If the columns are shorter than the cache-line, multiple hits are realized for each miss penalty paid. However, if the columns are longer only one hit will be achieved per miss and this is the same as the second case. If the two matrices are larger than the total cache size, then the matrices will never be wholly cached with the penalties being paid multiple times. The processing time should also be considered because overlapping data retrieval with processing can provide gain. These gains decrease as processor speeds increase.

The analysis and simulation above describes how the use of optical connectivity can be used for a many-banked, interleaved memory hierarchy which can support a very large processor memory bus. It also shows how such a high bandwidth will effect the design of the underlying electronics in the form of the cache-line sizes. An improvement is shown in coherent data transfers and how, by changing the cache-line sizes, both coherent and non-coherent data transfers can be greatly improved.

4. CONCLUSION AND FUTURE WORK

When a commodity interconnect such as the optical memory bus described above is being deployed for services, some general aspects need to be considered with regard to how it evolves in time:

1. Compatibility — how it interworks within various networks and systems already in existence.
2. Scalability/Survivability — its ability to expand itself for more capabilities such as providing higher bit-rate capacity, supporting more nodes/users, and interconnecting to one another.
3. Reliability — the degree of fault-tolerance (protection), both accidental and deliberate.
4. Deployment — this is the strategy for the deployment of these interconnects. Considerations include capital investment, pricing of services, and market penetration.
5. Parallel architectures — increasing the use of parallel processing, both in the traditional multi-processor sense and in multiple pipelines in single processors, will increase the requirement for high bandwidth, parallel data transfer.

In order to exploit the advantages of optics versus electronics in relation to interconnecting networks, the underlying electronic processor has to be considered as part of the system. There is a high possibility that processor design, especially cache design, will be changed dramatically as the constraints on bandwidth are removed. There are moves [8] to replace the computational cost-based design inherited from the early days of computing when wires were cheaper than valves with a communications-based model since gates are now much cheaper than the wiring between them.

Finally, the increasing use of parallel computers in the industry and research will drive higher bandwidths in both the processor-memory bus and the processor-processor bus.

In this paper, we have shown how an optical interconnect can be used within conventional silicon processor-memory architectures. It does not fully exploit the bandwidth available in the optical domain; each channel was only driven at 200 MHz, but this demonstrates the use of high interconnectivity. Even so, this was only an example of how a one-to-one interconnect can be utilized to compliment silicon electronics.

REFERENCES

[1] D.A.B. Miller, "Rationale and Challenges for Optical Interconnects to Electronic Chips," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 728-749, 2000.

[2] M.W. Haney and M.P. Christensen, "Fundamental Geometric Advantages of Free-Space Optical Interconnects," *Proceedings of MPPOI '96*, pp. 16-23, 1996.

[3] J. Jahns, "Planar packaging of free-space optical interconnections," *Proceedings of the IEEE*, vol. 82, no. 11, pp. 1623-1631, 1994.

[4] 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 Interconnect," *Applied Optics*, vol. 37, no. 14, p. 10, 1998.

[5] M.H. Ayliffe, D.R Rolston, A.E.L Chuah, E. Bernier, F.S.J Michael, D. Kabal, A.G. Kirk, D.V. Plant, "Design and testing of a kinematic package supporting a 32 × 32 array of GaAs MQW modulators flip-chip bonded to a CMOS chip," *Journal of Lightwave Technology*, Oct. 2001, pp. 1 –17.

[6] D.T. Neilson, S.M. Prince, D.A. Baillie and F.A.P. Tooley, "Optical design of a 1024-channel free-space sorting demonstrator," *Applied Optics*, vol. 36, no. 35, p. 10, 1997.

[7] J.L. Hennessy and D.A. Patterson, *Computer Architecture: A Quantitative Approach*, Palo Alto, California, Morgan Kaufmann Publishers, 1996.

[8] W.S. Coates, J.K. Lexau, I.W. Jones, S.M. Fairbanks and I.E. Sutherland, "FLEETzero: An Asynchronous Switching Experiment," *Technical report SML# TR-2000-0768*, Sun Microsystems, 901 San Antonio Road, Palo Alto, CA 94303-4900, 2000.

[9] I. Gourlay, P.M. Dew, K. Djemame, J.F. Snowdon, G.A. Russell, T. Lim, and B. Layet, "Supporting Highly Parallel Computing With a High Bandwidth Optical Interconnect," *University of Leeds School of Computing Research Report Series*, Report 2001.19, October 2001.