

Hardware for transputing without transputers

C P H (Paul) Walker

*4Links for technical help, P O Box 816, Milton Keynes, UK
+44 1908 566253, paul@walker.demon.co.uk*

Abstract. Transputing without transputers is mostly concerned with porting software to make other processors behave like transputers. One key aspect of the hardware is to provide these other processors with links. In the board described in this paper, IEEE 1355 DS-Links are chosen to plug in to PCI, which offers good performance, is inexpensive, and pervasive. The board uses a PCI chip, two C101 parallel link adaptors for DS-Links, and the buffering according to IEEE 1355. Prototyping space is also provided for developing applications. Potential applications are described, including: hosting T9000 systems; building Virtual Transputers; building servers or database systems with commodity PCs in a Redundant Array of Inexpensive Servers (RAIS); building a small-area ATM office network; and developing an In Home Digital Entertainment Network (IH DEN).

1. Introduction

In the debates and discussions at the last WoTUG conference [1], there was an almost unanimous wish to make the benefits of the transputer independent of transputer silicon from SGS-THOMSON. The bulk of such work is in software, to enable processors other than transputers to behave like transputers. The one key aspect of hardware that other processors need to make them behave like transputers is the transputer's links, so that it is as easy to communicate between processes in different processors as between processes running on the same processor.

A number of ideas were discussed informally at WoTUG - 18, and the development of a PCI [2] board interfacing to IEEE 1355 DS-Links [3] has been a direct outcome of these discussions. In particular, Dr B M Cook of Keele University offered to take on the necessary software aspects of the work. He has actually done far more than that and the development would not have been possible without him — this paper really ought to be a joint paper.

It is hoped that the work (and the resulting board) is of benefit to the community served by WoTUG, and that it may enable this community to broaden its appeal by using commodity computers and IEEE standard connections instead of proprietary hardware. SGS-THOMSON themselves can only benefit from such developments — where links are used, transputers inevitably follow.

2. Why DS-Links?

To an audience of transputer experts, I don't need to extol the benefits of transputer links. The reasons for choosing DS-Links rather than OS-Links or HS-Links [4] are that:

- At 100 Mbits/s to 200 Mbits/s full duplex, DS-Links offer good performance compared with OS-Links and other possible competing interfaces
- Both DS-Links and HS-Links have been standardised in IEEE 1355 [4]
- The C101 [3] and particularly the C104 [5] routing switch exist and are available for DS-Links, whereas no comparable routing switch exists for HS-Links
- In some respects, HS-Links, at 1 Gbit/s full-duplex, are too fast for PCI, unless, perhaps, bandwidth is wasted by inefficient network topologies
- It is possible to build both PCI and DS-Links in programmable logic, and so to evolve via proven steps towards a single-chip interface.

3. Why PCI?

The ideal bus is one that is pervasive, cheap, and fast. This combination does not leave much option, as shown in the table below:

Bus	Pervasion	Cost	Performance	Processors	Examples
ISA	Ubiquitous	Cheap	Slow	x86 only	PCs
EISA	Rare	Expensive	Reasonable	x86 only	Servers
MCA	Rare	Moderate	Reasonable		IBM only
Sbus	Moderate	Moderate	Reasonable	SPARC only	SUNs
Nubus	Moderate	Moderate	Reasonable	68K only	Macs only
VME	Moderate	Expensive	Moderate to fast	Any	Industrial
PCI	Ubiquitous	Cheap	Fast	Any	Any

Apart from being the only bus that appears to satisfy the requirement of pervasive, cheap, and fast, PCI's transfers of address followed by a few tens of bytes of data align closely to the packets transferred on DS-Links.

4. What is IEEE 1355

DS-Links were introduced by INMOS/SGS-THOMSON as the links originally for the T9000 family of transputers. The concept of fast virtual channels, however, is far broader than one family of processors, and the DS-Links and C104 routing switch chip were recognised as having much wider application. Interfaces tend only to be adopted if they are part of standards, and so SGS-THOMSON did considerable work on (initially proprietary) standards for these links and then submitted a standardisation proposal to the IEEE.

The IEEE assigned project 1355 to this effort as “Standard for Heterogeneous InterConnect (HIC) (Low Cost Low Latency Scalable Serial Interconnect for Parallel System Construction)” An ESPRIT project was also set up with interested parties, including Bull who contributed the HS-Links, and CERN, who did much of the proving work on the differential (buffered) and optical fibre versions of DS-Links..

The standard includes DS-Links, with physical implementations as single-ended TTL (unbuffered), differential pseudo-ECL/LVDS (buffered), and optical fibre, and HS-Links

with physical implementations for coax and fibre. It specifies DS-Links up to 200MBaud, and HS-Links from 700MBaud to 1GBaud.

The 1355 project ran through the many hurdles of the standardisation process with great success, and it became a fully ratified IEEE standard in September 1995.

5. What is PCI?

A good introduction to PCI is given by one of the manufacturers who makes chips for interfacing to the PCI bus:[6]

"The local bus concept was developed to break the PC data bottleneck. Traditional PC bus architectures are inadequate to meet the demands of today's graphics oriented systems and large application sizes. A local bus moves peripherals off the I/O bus and places them closer to the system's processor bus, providing faster data transfer between the processor and peripherals.

"The PCI Local Bus addresses the industry's need for a local bus standard that is not directly dependent on the speed and size of the processor bus, and that is both reliable and expandable. It represents the first time in the history of the PC industry that a common bus, independent of microprocessor design and manufacturer, has been adopted and used by rival computer architectures. PCI offers simple 'plug and play' capability for the end user, and its performance is more than adequate for the most demanding applications, such as full-motion video."

5.1 Performance

The basic PCI bus is 32 bits wide, and can transfer a 32-bit word of data every clock period of a 33MHz clock, to give a peak data rate of 132Mbytes/s.

Faster operation is specified both by a bus width of 64 bits, and by doubling the clock speed to 66MHz. Either on their own gives a peak bandwidth of 264Mbytes/s, both together give a peak bandwidth of 528Mbytes/s. While chips and systems have been designed for 64-bits and 66MHz, the bulk of current implementations use 32-bits, 33MHz.

As well as these headline speeds, performance is also optimised by allowing the bus to operate concurrently with the processor and memory, and by hiding the arbitration.

5.2 Cost

Multiplexing address and data onto the same pins means that for a 32-bit bus, only 49 signal pins are required, cutting down on both connector size and chip size.

The electrical specification has been designed around standard ASIC processes, making it comparatively easy to produce integrated PCI peripheral chips, without the glue logic required by most other buses.

Unlike other reasonable performance buses, PCI is being made in huge volume, and the economies of scale result in low costs.

5.3 Pervasion

Although originally designed by Intel, PCI has been adopted not only for PCs but also for Macintoshes and PowerPCs, and DEC Alphas. Sun are expected to introduce it on Sparcs, and there is already at least one 'embedded' processor (i960) which is integrated with PCI.

While in early 1995, only the 'premium' PCs were being sold with PCI, as this is being written in early 1996, most of the PCs being advertised come with PCI.

Whereas there used to be a sharp distinction between the commercial buses used for PCs and the industrial buses such as VME, the PCI bus is blurring this distinction. If a particular form-factor is popular for a board, there is a PCI standard being developed for that form-factor. Examples are the PMC mezzanine board, the Cardbus PCI version of the PCMCIA form-factor, and a compact PCI which uses 3U x 160mm Eurocards. The highest volume, and so the lowest cost, is likely to remain the PC form-factor, which also fits PowerPCs and physically compatible motherboards made with DEC Alphas.

5.4 Some technical details

PCI has three separate address mechanisms, one for memory, one for I/O, and one for configuration. The configuration space is used by the BIOS to ask all the boards in the system how much address-space they need. The BIOS then allocates the base addresses of each address-space and so eliminates the jumpers on ISA boards which determine the address the board responds to.

The basic type of transfer on the bus is a 'burst', of a 32-bit address, followed by multiple words of data:



Figure 1 Burst transfers on PCI

The burst may be a single word of data, or to I/O space may even be a single byte of data. On the other hand it may be several tens of bytes: the PCI specification quotes examples of transfers of 32 bytes up to 256 bytes. (This format works well with the DS-Link packet size of from 32 bytes for the T9000, through 48 bytes for ATM, and 192 bytes for MPEG2.)

Of course not all chips or boards can keep up with a word every clock period, and the specification allows a latency between the address and the first data of 16 clock periods, and, for subsequent data, of 8 clock periods. Where each access is to a single word, and all of this latency is used, performance plummets to a peak of 4Mbytes/s — still a substantial improvement on ISA.

6. What is on the PCI-1355 boards?

In some respects the boards [7] are so simple they need hardly any description. They use a standard PCI interface chip, one or two C101s, a single glue chip between the PCI chip and the C101s, and the link buffering needed for IEEE 1355. This does not take up all the space of the standard PCI form-factor 'small' board, so the unused space is filled with useful footprints for prototyping new applications with DS-Links.

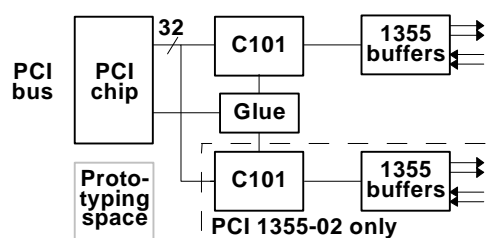


Figure 2: Block diagram of PCI-1355 boards

6.1 PCI chip

The PCI chip used is the AMCC S5933 [6]. This chip has been specifically designed for boards like the PCI-1355 which do not have processors, although it does not prevent the use of a processor.

In slave (target) mode, the S5933 presents five address spaces to the PCI bus, one of which is used for its internal operation registers; the other four address spaces are 'passed through' to the application logic. Two of these address spaces are used to access the two C101s.

A useful by-product of a set of otherwise unused mailbox registers is a simple hardware mechanism to provide interrupts from the status signals from each of the C101s.

The S5933 includes logic for master (initiator) mode access to the PCI bus, including short FIFOs for each direction of transfer. This logic is not used on the PCI-1355 boards, but may be used on future products.

A small 8-pin E²ROM is used to configure the S5933 and to provide information for the BIOS of what type of board this is and how much space is required for each address space. AMCC provide a device ID unique to each user of the S5933, and 4Links has a Device ID for the PCI-1355 boards.

6.2 C101s

It is possible to implement DS-Links in programmable logic, and indeed for many applications it will be appropriate to do so. The C101 is useful, however, in providing for processing the headers while data transfer is occurring. The also C101 provides the mux/demux between eight bits and 32 bits, and provides the various header and control registers, all of which would be considerably more expensive in a programmable implementation.

Each location in the register map of the C101 is seen from PCI as a 32-bit word, with a clean mapping of these registers onto the PCI bus.

6.3 Glue

A small state machine is needed to collect addresses from the PCI chip, latch them in the C101, and then transfer the data between them. For the two C101s used in slave (target) mode, this logic fits into a single 44-pin EPLD. The Altera (ex Intel) EPX740 was chosen for this because its SRAM configuration via JTAG makes development particularly cost-effective. Additional features of this logic which have proven to be useful compared with other logic families are the flexible clocking and output-enabling.

6.4 Buffering for IEEE 1355

The buffering used is the AT&T 41 series transceivers, in their SOIC (1141) packaging. Pads are present for using discrete resistor terminations, but the 1141 MMs with all the resistors integrated appear ideal. In particular the 1141 MMs have a higher ESD breakdown voltage than those without the integrated resistors. Further ESD protection is provided by catch diodes, which should make the boards proof to any (mis)handling of the cables and external connector pins.

The 41 series buffers have been shown to be excellent in EMC emissions [8]. Work on one of the OMI projects, however, found that with poor connections between the

connector shielding and the equipment chassis, they could be susceptible to common noise sources such as commutated AC motors.

The PCI-1355 boards have excellent connection between the connector shielding and the rear panel of the PC board. Not all PCs, however, and particularly those sold before January 1996, have such excellent connection between the rear panel and the equipment chassis. So the PCI-1355 boards include common-mode chokes on the output links, which were found greatly to reduce the susceptibility to radiated noise. Of course the chokes also attenuate any residual EMC emissions from the 1141 transceivers.

6.5 Prototyping space

One of the advantages of using the S5933 PCI chip was that AMCC provided a low-cost evaluation kit for it, which had the S5933 on a PCI board and some breadboard space for prototyping. This kit was used to check out the feasibility of the PCI-1355 design before committing to PCB layout. When the layout was complete, there was considerable empty space on the board.

As a number of users of the PCI-1355 board want to develop applications based on IEEE 1355, beyond exactly what was on the board, it seemed sensible to use the empty space to provide for bread-boarding in the same way as AMCC had done. As well as the 0.1" grid pattern of holes, the footprints for 208-pin and 100-pin QFPs provide for a wide range of transputer products (including the C104) or for FPGAs, EPLDs, or 32-bit-wide RAMs. Pads are also provided for a couple of HTRAM slots.

7. What can the PCI-1355 boards be used for?

Anyone who knows transputers knows that the possible applications are unlimited. In many respects, the same applies to any computer with link interfaces such as the PCI-1355 board provides, and all the applications in "Networks, Routers and Transputers" [9] are relevant. None of the following applications has, at the time of writing, been developed, and they all require further work. They are selected, however, as applications that are likely to gain substantially from the use of IEEE 1355 links with commodity computers and communications devices.

Although they all require further work, a great deal of the ground-work has already been done, and the majority of it by members of WoTUG, to bring these applications to fruition.

7.1 Fast host interface for T9 system

While the board has deliberately been designed without a T9000 transputer [10], it is attractive as a host interface for systems using T9000s and C104s.

The host interface for the 20Mbit/s OS-links of the T2xx, T4xx, T8xx families of transputers is grossly limited by the ISA bus, and so there appears little to be gained by perpetuating that limitation with DS-Links running five or ten times as fast as OS-Links. Almost all PCs bought now have PCI as well as an ISA bus, so the PC cost of the extra performance is negligible.

7.2 Occam For All Virtual Transputer

Occam For All [11] makes it possible to run your Occam programs on hardware other than transputers. Recent developments in the Occam For All project enable these

computers to talk to each other via Ethernet or Internet [12], which is a very useful capability. But, for good reason, transputers do not use Ethernet and Internet to talk to each other, they use links.

Putting links on the PCI interface makes it possible for these computers running Occam For All to communicate both internally and externally just like transputers, so they become "virtual transputers".

Of course the combination of using links, Ethernet, and Internet, each where appropriate, is far more useful than any of them alone. It may indeed become possible to think of the global network of internetworked computers as being a giant virtual transputer system.

7.3 RAIS (Redundant Array of Inexpensive Servers)

Servers are an area of computing where there is never enough performance. Indeed it has been suggested that in a few years time, parallel processing will be commonplace, not in conventional high performance computing, but in servers and database systems.

Evidence for this can be seen from the success of N-Cube both with Oracle and with their Video-On-Demand server — success which has been achieved similarly but perhaps less publicly by Meiko and Parsys. Further evidence is that many computers now sold as servers have at least two processors, whether Sparcs, Pentiums, or Alphas.

The actual disk memory of these server systems is often organised as a RAID (redundant array of inexpensive disks). The servers need to be connected to a variety of networks, and so are most likely connected via a big comms hub, which may provide switching, bridging, or routing to external networks.

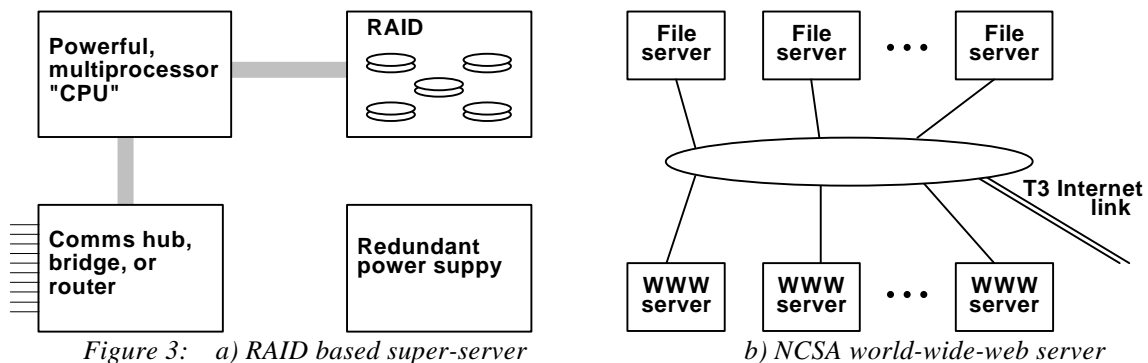


Figure 3: a) RAID based super-server

b) NCSA world-wide-web server

An alternative way to build servers is to use conventional computers, and to hook them together in a network. NCSA[13] realised that they would become a popular site for the world-wide-web (www), and conventional approaches to providing a server were not scalable to the likely demand. So they have built their 'server' out of a collection of HP workstations, connected together by what they claim is a scaleable interface network (a single FDDI ring). The advantage of this architecture is that servers can be added seamlessly, as the demand increases, and maintenance is also easier as the 'field replaceable unit' is a complete workstation instead of some part of the specialised super-server.

A further alternative would be to use cheap commodity PCs instead of the workstations, and to use a genuinely scaleable interconnect such as DS-Links and C104 routing switches. For small systems, the cost advantage is clear. For large systems the cost advantage is maintained, while the interconnect bandwidth and latency are superb.

Consider a Terabyte system based on PCs, each with four 1Gbyte IDE drives, so the system would have around 256 PCs. Each PC would have one or two external network

connections, such as ATM, Ethernet, or one of the fast Ethernets. One PCI board would connect two links to a CLOS network of routing switches. 48 C104s is enough to give this network a diameter of only three, a bisection bandwidth of around 50Gbits/s, and very low contention leading to minimal latency. With such a network connecting the PCs, it is possible to think of all the disks as being logically one big RAID, and all the external network connections as being part of a big hub/router, while physically the disks and network I/O is distributed around the system.

7.4 Small, Cheap, High-bandwidth "ATM" LAN

ATM uses 'cells' of 48 bytes plus a five byte header, which are extraordinarily similar to the packets of a DS-Link. So similar indeed that the ATM Forum has suggested DS-Links as a possible 'Serial Utopia'[14] (Utopia is a standardised byte-wide communication between ATM chipsets).

If DS-Links can be used, therefore, between chips in an ATM terminal or switch, it must also be possible to use them over short distances outside the terminal or switch. This produces a network like the 'Small Area Network' which INMOS described in 1991 [15]. The network is fully tolerant of single point failures in cables, hubs, and (if the servers are mirrored) servers. With the simple protocol, the small packet size, the low cost, and the 100Mbits/s to 200Mbits/s full-duplex performance of DS-Links, this makes a very attractive office network.

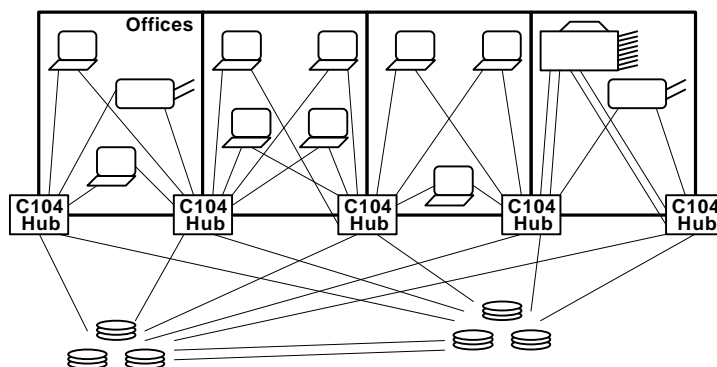


Figure 4: Fault-tolerant 'ATM' small-area office network based on IEEE 1355

It is possible indeed that cost/performance is so good with this network that the computers in the offices do not need disks, and that they access a 1355 board in the PC as if it was the disk board. SGS-THOMSON have built a example of such a system as part of their video-server demo.

7.5 Home multimedia networks

SGS-THOMSON have combined with National Semiconductor in proposing the use of a derivative of IEEE 1355 for an 'In Home Digital Entertainment Network' (IHDEN) [16].

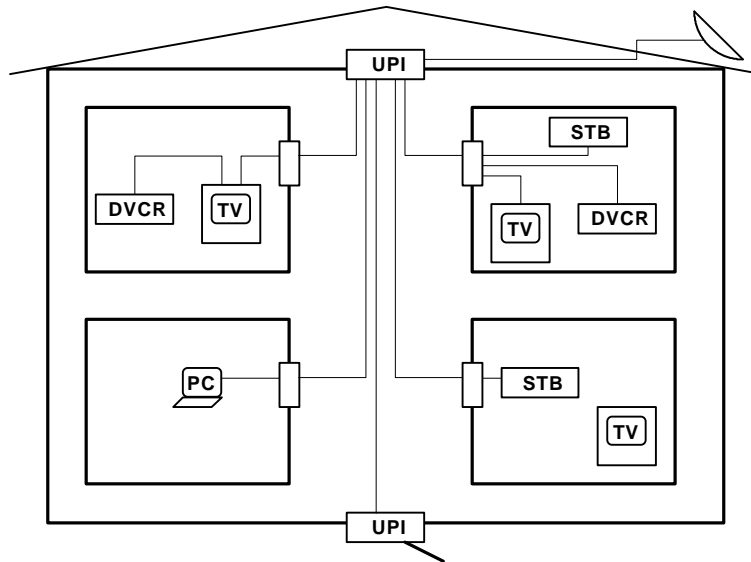


Figure 5: Proposed 'In-Home Digital Entertainment Network' (IH DEN)

The bulk of the traffic in such a network will be MPEG frames, and there may be additional traffic in ATM. The ability of DS-Links to multiplex these two seamlessly on the same physical link, without additional segmentation and reassembly, is a major advantage for these networks.

The SGS/NS proposal includes integrating the 1355 buffers into a single CMOS chip, and adds a simple additional physical routing protocol, so that it is possible to build daisy-chained networks as well as hub-based networks.

Clearly, for the volume required for such a consumer application, the whole MPEG/ATM/1355 function needs to be integrated onto a single chip. But plenty of system development and concept-proving can be done with the C101 on boards such as the PCI-1355.

8. Does it do what you want?

With any new product, initial enquiries often suggest a need for slightly different products. In particular, while the performance is much better than we are used to from ISA interfaces, there is still plenty of opportunity to enhance the performance. An obvious further product is one with a Virtual Channel Processor on the board, but even starting from a proven implementation of virtual channels in software, designing a VCP is a large piece of work.

With the prototyping space, if you want something extra, you at least have the opportunity to add it yourself. But for an interface from PCI, and for the applications of transputing without transputers, we have to start somewhere. The important thing is to start at all, and I hope that this board is a reasonable place to start.

9. Acknowledgements

Apart from Barry Cook, without whom this development would not have been done, I am grateful to many people over many years for their help, encouragement, and feedback concerning links and interfacing to them. The discussions at WoTUG 18 which triggered the development were based on previous work done with Michael Poole (now at Kent University) and with Flemming Christensen of Sundance, which in turn evolved

from discussions with many of my former colleagues at INMOS. Particular substance was given to the value of PCI as a useful standard for links by Andre Bakkers of Twente. And the IEEE 1355 standard was the work of many people, but particularly of Colin Whitby-Strevens of SGS-THOMSON.

10. References

- [1] P Nixon (ed.), *WoTUG - 18: Transputer and Occam Developments*, IOS Press Amsterdam,, 1995
- [2] PCI Special Interest Group, *PCI Specification Rev. 2.1*, Portland OR, USA, 1995.
- [3] SGS-THOMSON, *IMS C101 parallel DS-Link adaptor, Preliminary Data*, April. 1995, 42 1593 03.
- [4] IEEE *1355 Standard for Heterogeneous Interconnect (HIC) 1995*, IEEE Standards
- [5] SGS-THOMSON, *IMS C104 packet routing switch, Preliminary Data*,. June 1994, 42 1470 04
- [6] AMCC, *S5930-S5933 "Matchmaker" PCI controllers*, 1995 /PCI-0495
- [7] 4Links for technical help, "*PCI-1355-1, -2, Product Outline*", December 1995, available via <ftp.inmos.co.uk/inmos/projects/iee/hic/data/pci-rdme.txt>, or from author.
- [8] AT&T Microelectronics, *Achieving EMI Compliance Using the AT&T 41 series of Line Drivers and Receivers*, Application Note, Nov. 1992, AP92-024DBIP
- [9] May, Thompson & Welch (eds.), *Networks, Routers and Transputers*, ISBN 90 5199 129 0, IOS Press, Amsterdam, 1993.
- [10] INMOS/SGS-THOMSON *T9000 Hardware Reference Manual*, DBT000RMST/1.
- [11] URL <http://www.hensa.ac.uk/parallel/occam/projects/occam-for-all>
- [12] B M Cook, *Technical Report TR96-01, "Occam Virtual Channels over Ethernet"*, ISSN 1353-7776, Dept. Computer Science, Keele University, Keele, Staffs, UK.
- [13] Kwan, McGrath, Reed, *NCSA's World Wide Web Server: Design and Performance*, IEEE Computer, V 28, # 11, Nov. 1995, pp 68-74.
- [14] The ATM Forum Technical Committee, *Utopia, An ATM-PHY Interface Specification, Level 2, Version 0.95, Letter Ballot*, June 1995, ATM Forum
- [15] Paul Walker, *Connecting 100 MBaud HLinks*, Poster paper presented at Transputer Applications '91, Glasgow, August 1991
- [16] Phil Hughes, Colin Whitby-Strevens, *IHDEN: In Home Digital Entertainment Network*, proposal to EIA in response to EIA R-4.1 RFP, April 1995, National Semiconductor and SGS-THOMSON.