

Extending FPGA based Teaching Boards into the area of Distributed Memory Multiprocessors

Michael Manzke and Ross Brennan

Trinity College Dublin

Ireland

michael.manzke@cs.tcd.ie,ross.brennan@cs.tcd.ie

Abstract

Reconfigurable hardware, in conjunction with soft-CPU's, has increasingly established itself in computer architecture education. In this paper we expand this approach into the area of distributed memory multiprocessor systems.

Arguments that supported the introduction of reconfigurable hardware as a substitute for commodity CPUs on educational computer architecture boards are equally applicable to teaching hardware that facilitates the construction and configuration of multiprocessor systems.

The IEEE Standard for the Scalable Coherent Interface (SCI) was chosen as the interconnect technology because it enables the demonstration of the most important architecture concepts in this context. This interconnect exhibits high bandwidth and low latencies and not only specifies a hardware Distributed Shared Memory (DSM) architecture, but also defines cache coherence protocols. Consequently an implementation of this standard allows the design of Non-Uniform Memory Access (NUMA) and cache-coherent NUMA (ccNUMA) multiprocessor systems.

1 Introduction

The initial design objective for the next generation of computer architecture teaching boards was driven by the desire to design a system that would provide *reconfigurable hardware* resources for a range of *soft-CPU's*. So far the following three *soft-CPU's* are available for the system:

- Open-source SPARC LEON P-1754 [1]
- Op-code compatible MC68008 [2]
- Teaching Instruction Set Processor [3]

Current final year projects will provide a *Java Virtual Machine* and an *Advanced Teaching Instruction*

Set Processor. The latter processor will include *Instruction Level Parallelism (ILP)* concepts.

The design of these boards promises a number of advantages over the use of commodity CPUs. Most importantly it enables undergraduate students to test their synthesisable hardware description language (HDL) models on real hardware. The current generation of teaching boards is based on the Motorola 68008 processors. Second year undergraduate student are asked to construct a complete microprocessor system by wire-wrapping integrated circuits (ICs) such as memory and input-output (IO) devices. The next generation of hardware will also allow the students to experiment with a microcoded multi-cycle instruction set processor that they design in VHDL in the second semester of the course. This is one example that demonstrates how the introduction of *reconfigurable hardware* has significantly widened the scope of the boards to now cover several subjects throughout the Computer Science and Computer Engineering undergraduate degree courses at TCD, ranging from the simple design of an instruction set processor to the design of high performance floating point pipelines.

Our successful prototype, in conjunction with three *soft-CPU's* that were either adapted to the board's architecture or newly developed, has motivated us to take the design one step further by integrating hardware that allows for the construction of closely coupled multiprocessor systems through the interconnection of these boards. We chose the *Scalable Coherent Interface (SCI)* as the most suitable interconnect technology [4]. The interconnect architecture is implemented through a hybrid design of commodity *LC3 Link Controllers* from Dolphin Interconnect Solutions Inc.¹ and Field Programmable Gate Arrays (FPGA) [5]. The *LC3 Link Controller* implements link-level aspects of the SCI standard whereby higher level parts of the SCI specification,

¹This work is supported by Dolphin Interconnect Solutions Inc.

such as the cache-coherence protocol, are implemented through *reconfigurable hardware* to guarantee access to these functions.

This system architecture enables the configuration of *Non-Uniform Memory Access (NUMA)* and *cache-coherent NUMA (ccNUMA)* multiprocessor systems. Furthermore, SCI switches may be used to demonstrate a direct interconnection network. As an alternative, indirect networks can be constructed through multidimensional meshes. Implicit communication is provided by the shared address space of the *NUMA* or *ccNUMA* architecture and explicit communication may be implemented on top of shared address space architectures.

2 The Interconnect

The *Scalable Coherent Interface (SCI)* as defined in the IEEE standard 1596-1992 provides bus-like services by replacing a system-bus with 16 bit parallel point-to-point unidirectional links. This approach overcomes electrical limitations of system buses that are characteristic for *Symmetric Multiprocessor (SMP)* systems. Multiprocessor systems can be constructed with this technology that scale up to several thousand nodes.

2.1 SCI in Commodity Systems

Defined in 1992, SCI is a well established technology and many high performance cluster implementations employ this interconnect (e.g, PC2 University of Paderborn Germany, University Of Delaware - The Bartol Research Institute USA and National Supercomputer Centre in Sweden) [6]. Subsets of the SCI standards have been implemented and are available as commodity components. In particular, Dolphin [7] have implemented PCI cards that bridge PCI bus transactions to SCI transactions. Compute nodes with PCI slots may be interconnected through PCI-SCI bridges together with a suitable SCI fabric topology, thus bridging their PCI buses. References made by one of these nodes into its own PCI address space are translated into a SCI transaction and transported to the correct remote node. The remote node translates this transaction into a memory access, thus providing a hardware DSM implementation. Programmed IO (PIO) and Direct Memory Access (DMA) may be performed without the need for system calls.

In a commodity SCI card a PCI-SCI bridge translates between PCI transactions and SCI transactions and forwards them onto the PCI bus or the BLink bus. The SCI BLink bus interconnects the PCI-

SCI bridge with up to seven SCI Link Controllers (LC) or alternative components. SCI cards with two SCI Link Controllers attached to the BLink are consequently suitable for the construction of a 2-dimensional torus. Systems with more than one LC can route packets over the BLink to the correct LC according to a routing table. This enables distributed routing of SCI packets between individual SCI rings without an expensive central SCI switch. Routing is configured during SCI fabric initialisation. Every LC has an input and output port and the output port of one LC component is connected via a cable to the input port of another LC component. These links are 16 bit parallel and unidirectional with a bandwidth of 667Mbytes/s.

2.2 SCI and Cache Coherence

The connection of SCI link controllers to the IO bus via a bridge was not intended during initial specification of the SCI standard but it allows commodity component manufacturers to offer SCI subsystems that may be attached to a diverse set of computer architectures as long as they provide a standard IO bus. This therefore enables the construction of Non-Uniform Memory Access (NUMA) machines with commodity PCs. This approach prohibits the implementation of cache coherence as defined in the SCI IEEE 1596-1992 standard [4].

2.3 SCI and Real Time Constrains

One possible application of the teaching board is in the context of embedded distributed control systems that must meet real-time constraints. The SCI technology with its low latencies and deterministic behaviour would assist such architecture to meet these constraints. SCI is already being used in mission critical real-time applications. For example, Thales Airborne System employs SCI for backplane communication in their EMTI unit (Data Processing Modular Equipment). This scalable unit is integrated into the Mirage F1, 2000 and Rafale combat aircraft, NH-90 helicopters, Leclerc tanks, submarines, the Charles de Gaulle aircraft carrier and strategic missiles [8]. The application of SCI technology in airborne systems provides ample evidence concerning its suitability for real time applications. Some research on the SCI fabric's suitability for real time application was conducted at Trinity College Dublin [9, 10].

3 The Teaching Board

The design of the teaching board avoids the PCI-bus as the interface to the SCI interconnect. Two commodity *LC3 Link Controllers* from Dolphin Interconnect Solutions Inc. [5] are directly connected to a *Memory Bridge Field Programmable Gate Ar-*

ray (FPGA). This interconnection is implemented through a *64 bit Backside Link (B-Link)* bus [11] . A *soft-CPU FPGA* is also connected to the *Memory Bridge FPGA* via the CPU's system-bus. The *Memory Bridge FPGA* implements the *SCI upper level protocol management* and bridges bus transactions between the *soft-CPU's* system-bus and the

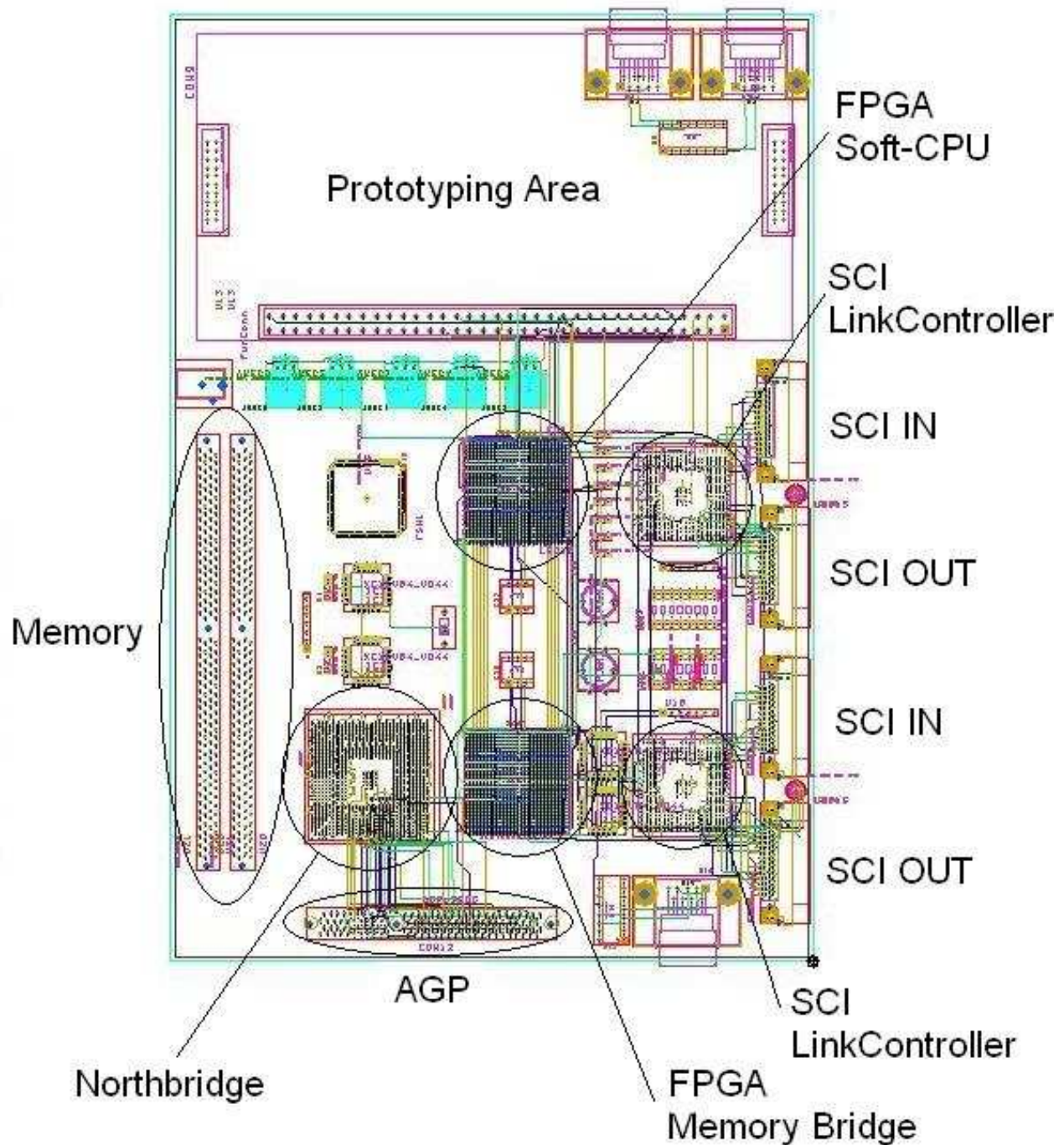


Figure 1: Printed Circuit Board

SCI B-Link. Figure 1 shows the layout of the *Printed Circuit Board* and annotates the main components of the system. The *Memory Bridge FPGA* is also connected to a *Northbridge* to enable access to the local memory of the system. The function of the *Memory Bridge FPGA* is to route the *soft-CPU's* memory references on its system-bus to either the local memory, via the *Northbridge*, or to remote memory by translating the memory reference into a SCI transaction. These SCI transactions are routed via the SCI interconnect fabric to the correct remote node. The SCI transactions will then be converted to a local memory reference by the remote *Memory Bridge FPGA*, therefore implementing a hardware *Non-Uniform Memory Access (NUMA)* multiprocessor systems. This approach does not require *Operating System Calls* to execute load and store transactions on remote memory but it does require software intervention to implement cache coherence.

The SCI IEEE 1596-1992 standard [4] defines

Cache-coherence Protocols as an optional implementation. This directory based protocol enables processors to cache data from remote memory locations while maintaining the coherence of the multiple copies. There are also commercial products that implement the SCI *Cache-coherence Protocols* in hardware. A good example of a *cache-coherent NUMA (ccNUMA)* machine that implements SCI including the *Cache-coherence Protocols* is the *NUMA-Q* [12]. Figure 2 assumes that the soft-CPU FPGA holds an Open-source SPARC LEON P-1754 [1] and the *Memory Bridge FPGA* implements in addition to the SCI upper level protocol management a SCI Cache Coherence Protocol. This figure highlights the components of the teaching board that may be used by the students to implement their design solutions through VHDL. This area is labeled *Reconfigurable Hardware or Playground*. The remaining components on the board are commodity ICs that can only be configured by the students through the modification of registers.

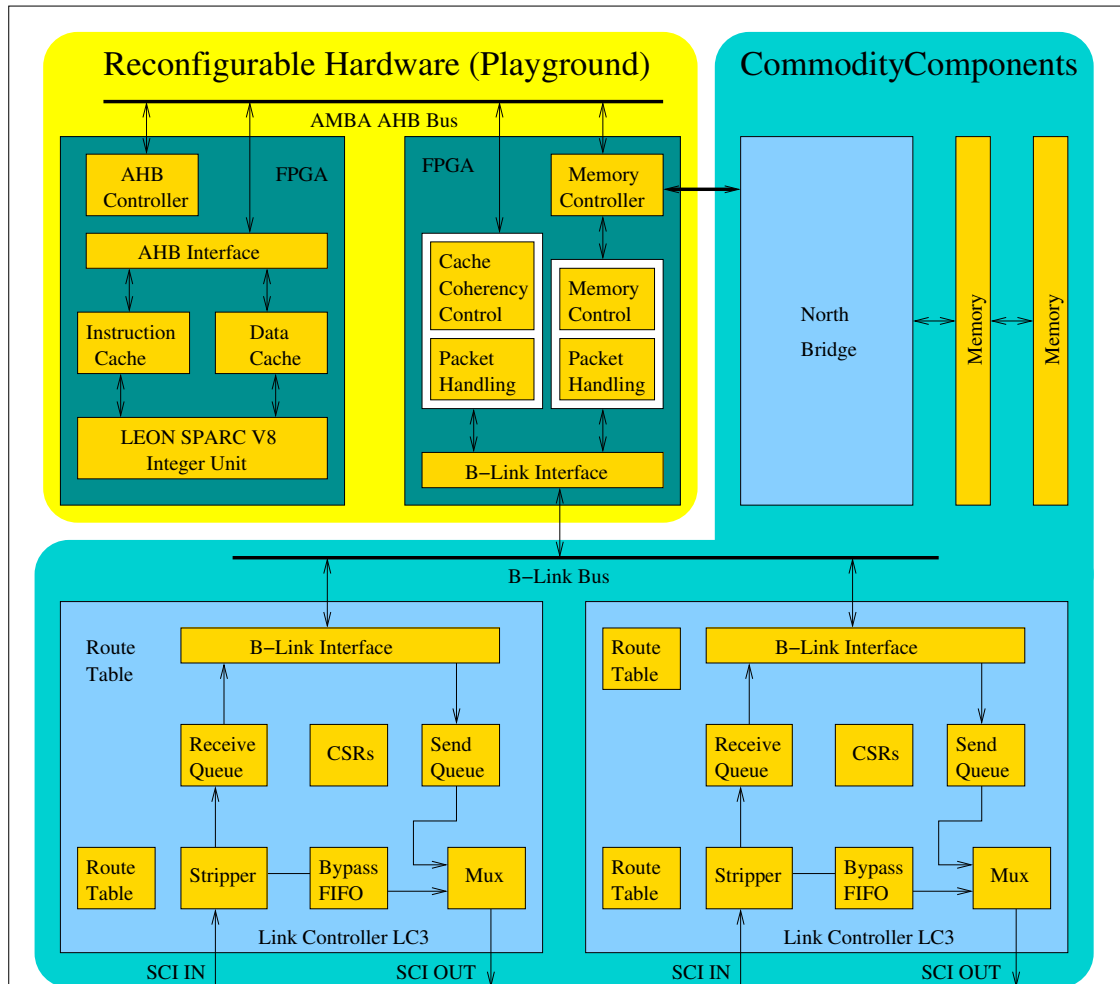


Figure 2: System configuration that implements a LEON CPU with Cache Coherency Protocol

For example the two *Route Tables* in the *LC3 Link Controllers* can be manipulated by the students to implement the desired routing on the *64 bit Backside Link (B-Link)* bus [11] or the SCI fabric.

Figure 3 shows how individual boards could be interconnected as 2D torus to build a *cache-coherent NUMA (ccNUMA)* multiprocessor system. Again assuming that the soft-CPU FPGA holds a Open-source SPARC LEON P-1754 [1] and the FPGA Memory Bridge implements in addition to the SCI upper level protocol management a *SCI Cache-coherence Protocols*.

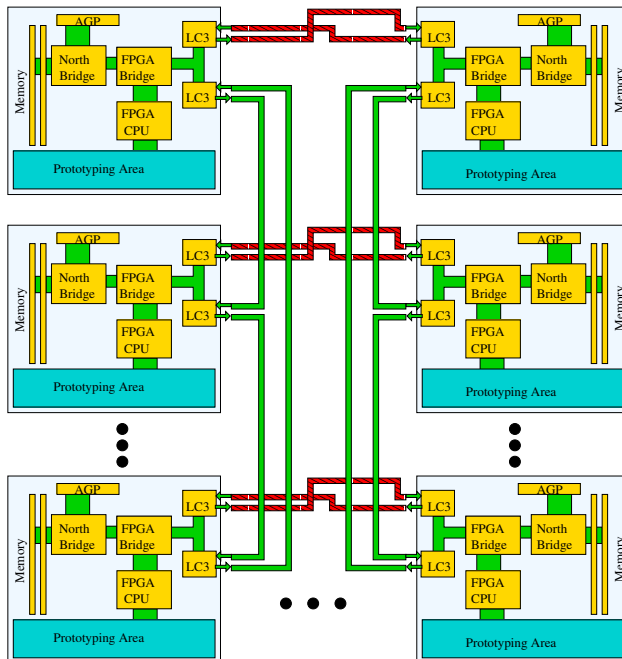


Figure 3: 2D Torus Multiprocessor

The *Memory Bridge FPGA* is responsible for acting as a bridge between the different commodity components on the board and so interfaces with both the *soft-CPU FPGA* and the *Northbridge*, an Intel Graphics and Memory Controller (GMCH) [13], in addition to the two *SCI LC3 Link Controllers*.

The board is designed to be compatible with any of the XC2V-FF896 range of FPGAs [14]. This gives the option to install different density FPGAs on the board. An XC2V2000 FPGA is required for the *Memory Bridge FPGA*. The *soft-CPU FPGA* can be either an XC2V1000/1500/2000 chip depending on board requirements. The XC2V2000 FPGA would be suitable for a SMP solution. This option of components has direct consequences on the final cost of the boards as the FPGAs are the most expensive parts of the system and their pricing is determined

by their memory density and speed grades.

3.1 Hard and Soft Hardware

The *Northbridge* is included in the design as it provides access to DDRRAM. It connects the *Memory Bridge FPGA* via a *Front-side Bus (FSB)* interface and also provides the possibility to attach a gigabit ethernet communications link to the system using the dedicated CSA bus on the *Northbridge*.

The bus connecting the *Memory Bridge FPGA* to the *soft-CPU FPGA* is a combination of the *AMBA Advanced Highspeed Bus (AHB)* and *Advanced Peripheral Bus (APB)* [15], allowing direct memory mapping between modules in both FPGAs. Finally, there is also a dedicated bus interface between the *soft-CPU FPGA* and the *Prototyping Area* on the board, which enables students to connect components (such as ROM and RAM) directly to the soft-CPU core in the FPGA, bypassing the *Memory Bridge FPGA*.

Although there are several soft-CPU core options available for the board, the LEON P-1754 SPARCv8 certified processor HDL model [16, 17] should be used if an open source operating systems with a supporting tool-chain is required. The CPU offers many features including a configurable cache, dedicated debugging link, floating point and memory management units, which may be individually configured or disabled. The main internal bus for the Leon2 core is the AMBA bus, which will be modified to communicate directly with the AMBA bus present in the *Memory Bridge FPGA*. When the board is configured with the full version of the Leon2 core, it will be able to communicate with all of the peripheral devices on the board. In order to reduce the complexity of the system, where necessary for teaching, a second version of the Leon2 core, with most of the internal components disabled, can be uploaded onto the board or alternatively one of the other soft-CPU cores, which have been modified for the board, can be used.

3.2 Operating Systems and Software

As the Leon2 core is fully SPARCv8 compliant, there is a choice of several different operating systems for the board. RTEMS will be used where it is necessary to have a hard-real-time operating system and Linux will be used as a more general purpose operating system. RTEMS is an open source real-time-operating-system (RTOS) designed for embedded systems [18]. It has multitasking capabilities and is POSIX 1003.1b compliant. It also has support for TCP/IP and the GNU toolset chain (includ-

ing ISO/ANSI C and ISO/ANSI C++). It supports several different filesystems including FAT32/FAT16 and NFS and can connect to GDB (Gnu Debug) over ethernet or serial port.

Several different ports of Linux were made for the Leon2 core and these are now being merged into the main linux kernel tree [19]. As such, it is possible to configure and compile a linux kernel suitable for use on the board. It also has full support of the GNU tool-chain.

3.3 Educational Scope

The teaching boards provide students with three *Playgrounds* for their experimental work:

- Prototyping Area - For wire-wrapping IC components to interface logic in the *soft-CPU FPGA*.
- Reconfigurable Hardware - For soft-CPU's, Logic, Memory Management, Cache-coherence Protocols and much more.
- Interconnecting Boards with SCI - Different SCI fabric configurations in conjunction with logic in the Reconfigurable Hardware.

In an effort to increase the synergy of various hardware related subjects of the Computer Science and Computer Engineering syllabi we incorporated as many feature as necessary to enable students to incrementally build on previous experimental experience. It remains to be seen to what extend the academic staff in the departments will adapt the features provided by teaching boards in their courses. The first set of prototype PCBs will be manufactured and populated with integrated circuits in October 2004. The subsequent academic year will be used to debug the hardware and develop VHDL models for the reconfigurable hardware. The academic year 2005/2006 will introduce the board to undergraduate student in the second year Computer Architecture course.

The following list gives an example of the potential applications of the board in the Computer Science undergraduate degree:

- Introduction to Computing
- Digital Logic Design
- Systems Programming
- Computer Architecture I - Microprocessor Systems

- Computer Architecture I - Computer Architecture
- Computer Architecture II - Workstations
- Computer Engineering
- Systems Software - Operating Systems
- Compiler Design II

4 Conclusions

This paper argues for a single multi-purpose FPGA based lab-board which provides features suitable for the entire hardware related subjects of Computer Science and Computer Engineering undergraduate and postgraduate students. It was demonstrated that reconfigurable hardware not only allows the system to operate under a range of soft-CPU's, it also provides the means to let students experiment with their own synthesised HDL models. Furthermore the hardware implementation of the link level part of the *Scalable Coherent Interface (SCI)* standard through commodity components in conjunction with an FPGA implementation of higher level protocol management and cache coherency protocols enables students to build and experiment with *Non-Uniform Memory Access (NUMA)* and *cache-coherent NUMA (ccNUMA)* multiprocessor systems. The Leon soft-CPU allows the system to execute Linux and RTEMS, therefore providing a Unix like general purpose operating system and a hard-real-time operating system. All these three components Leon, Linux and RTEMS are open source which reduces the cost of the system significantly and more importantly gives students access to the hard and soft source code. Last but not least students may wire-wrap a microprocessor system on the prototyping area and operate the IC components with a soft-CPU of their choice by switching the soft-CPU's system bus interface from the Bridge-FPGA to the prototyping area. It should be emphasised that many of the functions provided by the board have been initially developed or evaluated through final year project. This is to point out that the complexity of the various teaching objectives is well suitable for Computer Science and Computer Engineering undergraduate students.

References

- [1] R. Brennan and M. Manzke, "On the introduction of reconfigurable hardware into computer architecture education," in *Workshop on*

- Computer Architecture Education WCAE 2003* (E. F. Gehring, ed.), pp. 96–103, June 2003.
- [2] D. Lynch, “A motorola 68008 op-code compatible vhdl cpu,” April 2004. <http://www.cs.tcd.ie/Michael.Manzke/fyp2003-2004/DavidLynch.pdf>.
 - [3] L. Redmond, “Design of a teaching instruction set processor in vhdl,” April 2004. <http://www.cs.tcd.ie/Michael.Manzke/fyp2003-2004/LauraRedmond.pdf>.
 - [4] P. M. Kelty, *IEEE Standard for Scalable Coherent Interface*. IEEE, iee std 1596-1992 ed., March 1992.
 - [5] Dolphin Interconnect Solutions Inc., *LC3- SCI Link Controller for System Area Networks*. <http://www.dolphinics.com/products/hardware/lc3.html>.
 - [6] “Clusters @ top500,” May 2004. <http://clusters.top500.org/>.
 - [7] “Dolphin interconnect solutions inc.,” May 2004. <http://www.dolphinics.com>.
 - [8] “Dolphin interconnect solutions inc.,” June 2000. <http://www.dolphinics.com/news/2000/june020-2000.html>.
 - [9] M. Manzke and B. Coghlan, “Non-intrusive deep tracing of sci interconnect traffic,” in *SCI-Europe* (W. Karl and G. Horn, eds.), pp. 53–58, September 1999.
 - [10] B. C. O. L. Michael Manzke, Stuard Kenny, “Tuning and verification of simulation models for high speed interconnection,” in *PDPTA* (H. Arabnia, ed.), pp. 1087–1093, June 2001.
 - [11] Dolphin Interconnect Solutions Inc., *A back-side link (B-Link) for scalable coherent interface (SCI nodes)*, May 1996.
 - [12] H. Hellwagner, *SCI: Scalable Coherent Interface*, vol. 1734 of *Lecture Notes in Computer Science*, ch. 1. The SCI Standard and Applications of SCI, pp. 26–30. Springer, 1999.
 - [13] Intel, *Intel 865G/865GV/865PE/865P Chipset*. Intel, March 2004. <http://developer.intel.com/design/chipsets/>.
 - [14] Xilinx, *Introduction to the VirtexII Product Family*. Xilinx Inc, December 2001. <http://www.xilinx.com>.
 - [15] ARM, *AMBA Specification V2.0*. ARM Limited, May 1999. <http://www.arm.com>.
 - [16] J. Gaisler, *Leon2-1.0.10 Users Guide*. Gaisler Research, December 2002. <http://www.gaisler.com>.
 - [17] SPARC, *SPARC V8 Manual*. SPARC International Inc, January 1992. <http://www.sparc.org>.
 - [18] “Rtems is the real-time operating system for multiprocessor systems,” May 2004. <http://www.rtems.com/>.
 - [19] “Linux for leon2 processor,” May 2004. <http://www.gaisler.com/linux.html>.