

# Future Trends in SoC Interconnect

Steve Furber  
School of Computer Science  
The University of Manchester  
Manchester M13 9PL, UK  
[steve.furber@manchester.ac.uk](mailto:steve.furber@manchester.ac.uk)

**Abstract**—Self-timed packet-switched networks are poised to take a major role in addressing the complex system design and timing closure problems of future complex Systems-on-Chip. The robust, correct-by-construction characteristics of self-timed communications enables each IP block on the SoC to operate in its own isolated timing domain, greatly simplifying the problems of timing verification. Design automation software can remove the need for expertise in self-timed design, enabling the on-chip interconnect to be treated as an additional IP block within a conventional (synchronous) design flow.

The paradigm shift from viewing the SoC design problem as a matter of organizing complex hierarchies of buses with multiple coupled timing domains, where every interface between timing domains must be verified carefully, to viewing the SoC as a problem in network design where those timing issues are automatically isolated, promises significant improvements in designer productivity, component reuse and SoC functionality.

## INTRODUCTION

The System-on-Chip (SoC) industry has developed rapidly over the last fifteen years from producing VLSI devices that integrated a processor and a few memory and peripheral components onto a single chip to today's high-performance SoCs that incorporate hundreds of IP blocks. This progress is a consequence of Moore's Law (which enables ever-higher levels of integration) and of market economics (where consumers demand ever-more functionality in smaller, lower-cost products with better battery life).

Complex systems have always been hard to build and debug. Systems-on-chip are no different in this respect, but the economics of integrated circuit manufacture do not allow debugging by trial and error; the design must be 'right first time'. As a result, SoC designers adopt disciplines that minimize the risk of design error. One such discipline is to be very systematic about the way interconnect is used on a complex chip. Any communication failure, whether due to noise or an error in timing or protocol, is likely to require a design iteration that will be expensive in both mask charges and time to market.

## SoC INTERCONNECT EVOLUTION

Early SoCs used an interconnect paradigm inspired by the rack-based microprocessor systems of earlier days. In those rack systems, a backplane of parallel connections formed a 'bus' into which all manner of cards could be plugged. A system designer could select cards from a catalogue and simply plug them into the rack to yield a customized system with the processor, memory and interfaces required for any given application.

In a similar way, a designer of an early SoC could select IP blocks, place them onto the silicon, and connect them together with a standard on-chip bus (see Figure 1). The backplane might not be apparent as a set of parallel wires on the chip, but logically the solution is the same.

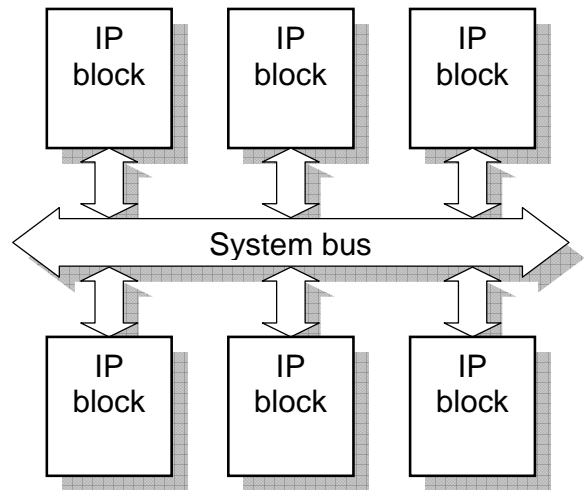


FIGURE 1: EARLY (CIRCA 1995) SoC STRUCTURE BASED UPON A SINGLE SHARED BUS.

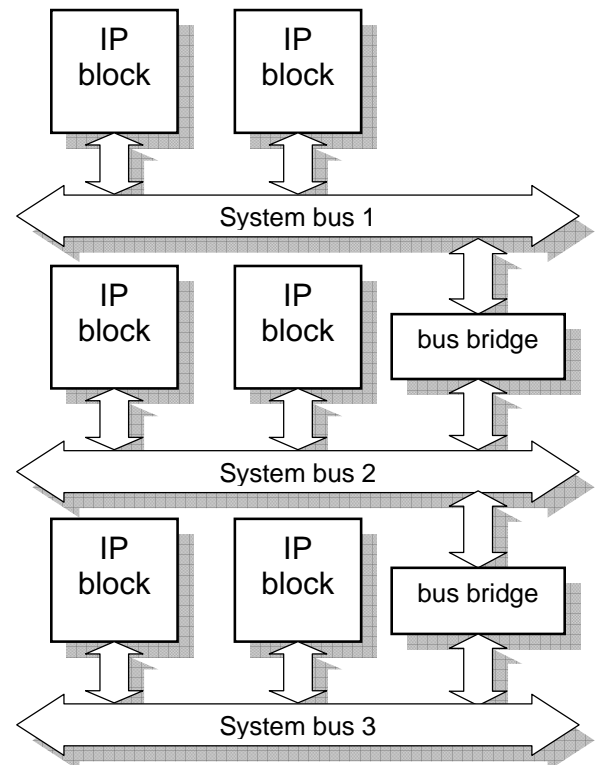


FIGURE 2: MORE RECENT (CIRCA 2000) SoC STRUCTURE BASED UPON A HIERARCHY OF BUSES.

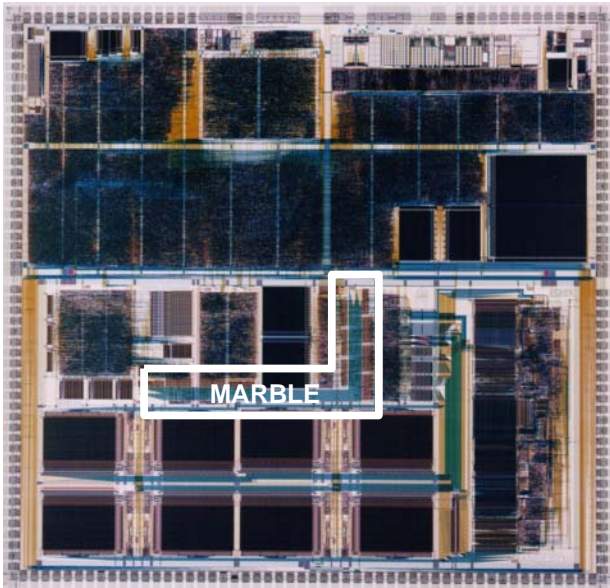


FIGURE 3: DRACO CHIP PLOT SHOWING THE ASYNCHRONOUS MARBLE BUS.

However, buses do not scale well. With the rapid rise in the number of blocks to be connected and the increase in performance demands, today's SoCs cannot be built around a single bus. Instead, complex hierarchies of buses are used (as illustrated in Figure 2), with sophisticated protocols and multiple bridges between them. Communication between two remote blocks can go via several buses, and every section of every path must be carefully verified. Timing closure is a growing problem because there is so much that must be checked. Bus-based interconnect is being stretched to its limit, and as the limit is approached the risk of errors increases rapidly. A new interconnect strategy is required to bring these risks back under control.

An example of a complex system-on-chip that employs the hierarchical bus solution is the DRACO DECT-ISDN controller chip (see Figure 3) which was based around the Amulet3i asynchronous processing subsystem [1]. This chip incorporates two local processor buses serving the instruction and data requirements of the Harvard architecture Amulet3 processor core, the asynchronous multi-master MARBLE bus with production test support [2], and a bridge to a synchronous on-chip bus (SOCB in Figure 4) that serves the clocked telecommunication peripherals.

### NETWORKS-ON-CHIP

Where bus-based solutions reach their limit, packet-switched networks are poised to take over [3]. A packet switched network offers flexibility in topology (see Figure 5) and trade-offs in the allocation of resources to clients. Such a network has been developed at the University of Manchester [4] and demonstrated on a prototype smart card chip [5]. The layout of the smart card chip is shown in Figure 6, and the CHAIN interconnect can be seen as the top-level wiring and the small logic blocks associated with it.

This Network-on-Chip (NoC) employs self-timed logic techniques to deliver a robust, correct-by-construction interconnection fabric that allows each client block to operate in its own fully-decoupled timing domain, thereby addressing system-level timing-closure issues.

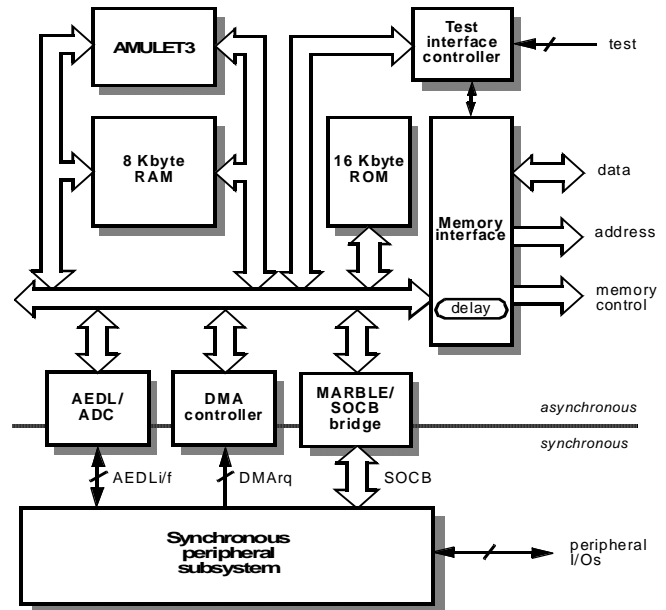


FIGURE 4: THE DRACO SYSTEM-ON-CHIP HIERARCHICAL BUS STRUCTURE.

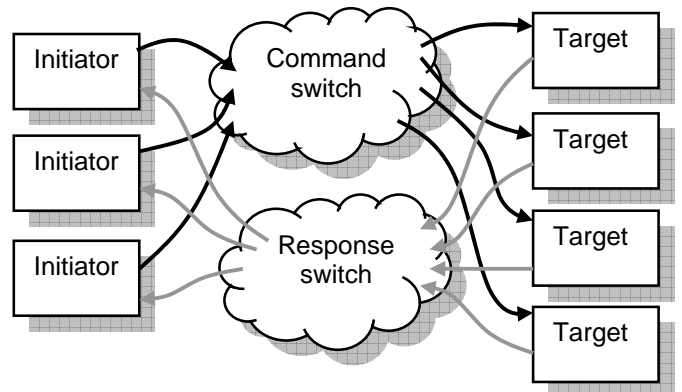


FIGURE 5: A PACKET-SWITCHED NETWORK-ON-CHIP OFFERS THE MOST FLEXIBLE INTERCONNECT STRUCTURE.

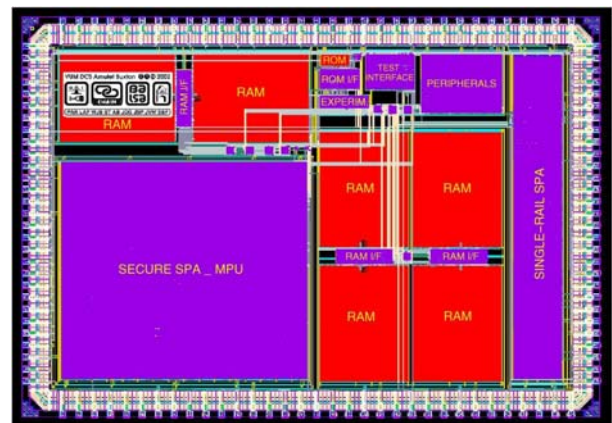


FIGURE 6: THE SMART CARD CHIP INCORPORATING THE PROTOTYPE CHAIN ON-CHIP INTERCONNECT.

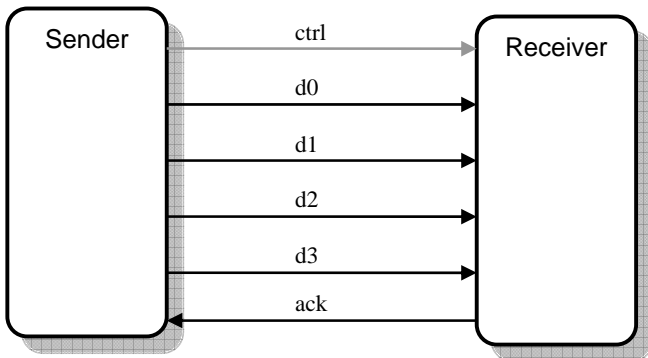


FIGURE 7: A CHAIN POINT-TO-POINT SELF-TIMED LINK.

The principles of the CHAIN self-timed point-to-point interconnect are illustrated in Figure 7. The link employs a 1-of-4 data encoding to send 2 bits of data by switching one of the 4 data wires using a return-to-zero protocol.

Every transition is acknowledged, thereby ensuring that variations in transmission delay (due, for example, to crosstalk) are accommodated automatically. A 5<sup>th</sup> data wire is used to indicate end-of packet, resulting in a 6-wire data link. This link has been demonstrated operating at a 500 MHz symbol rate (where each symbol is 2 bits of data) on a 0.18 micron CMOS process, yielding a data throughput of 1 Gbit per 6-wire link.

Where a single link provides insufficient performance, multiple links can be deployed in parallel and individual links can be pipelined, as shown in Figure 8. An example of a pipeline repeater circuit is shown in Figure 9. Five Muller C gates form the synchronizing latches, holding the 1-of-4 data and the control bit. The 5-input NOR gate is a ‘completion detection’ circuit, indicating when one of the latches has been set. The repeater forms one stage of a Muller pipeline and can be used to divide long point-to-point links into shorter sections, thereby counteracting the throughput problems created by the delays inherent in driving long wires. Of course, long wires still introduce end-to-end latency problems, but these are inherent in any on-chip communication system.

The switching technology in CHAIN comprises 2-to-1 multiplexers and 1-to-2 demultiplexers. With these components a range of standard network topologies can be implemented, such as a basic mux-demux topology, cross-bar mesh structures, and even an on-chip ring (see Figure 10). Although technically feasible, rings do not seem well-suited to on-chip interconnect as their efficiency depends on multiple packets being in transit simultaneously within the ring fabric, which does not occur with the latencies that arise on chip.

### NETWORK CONFIGURATION

Configuring an NoC to support the functionality and performance requirements of a complex SoC is still a daunting task. However, this is a task that can be addressed by design automation software. Silistix Limited, a company spun-out from the University of Manchester and formed around the NoC team, is developing sophisticated tools that will enable the interconnect fabric to be treated as just another synthesizable IP block integrated into the standard design flow (see Figure 11). This will eliminate the risks inherent in current bus-based interconnect methodologies and bring timing-closure problems down to manageable size.

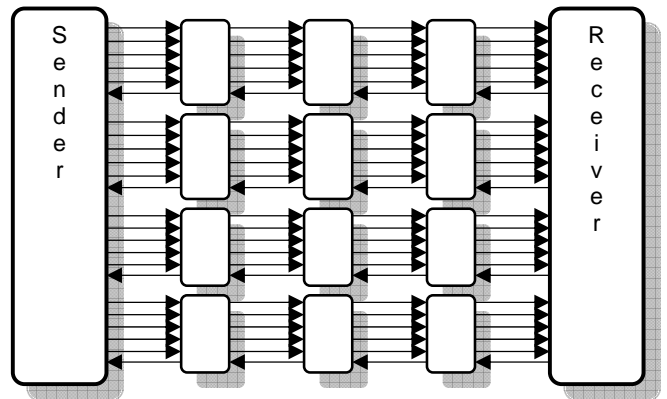


FIGURE 8: A WIDE, LONG CHAIN CONNECTION WITH PARALLEL LINKS AND PIPELINE REPEATERS.

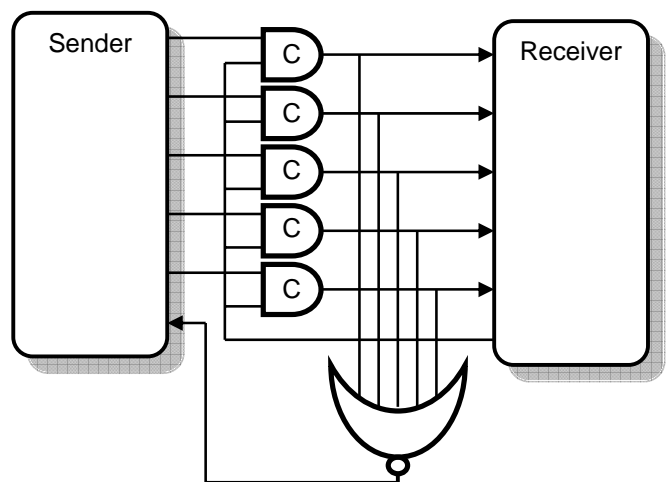


FIGURE 9: A CHAIN PIPELINE REPEATER.

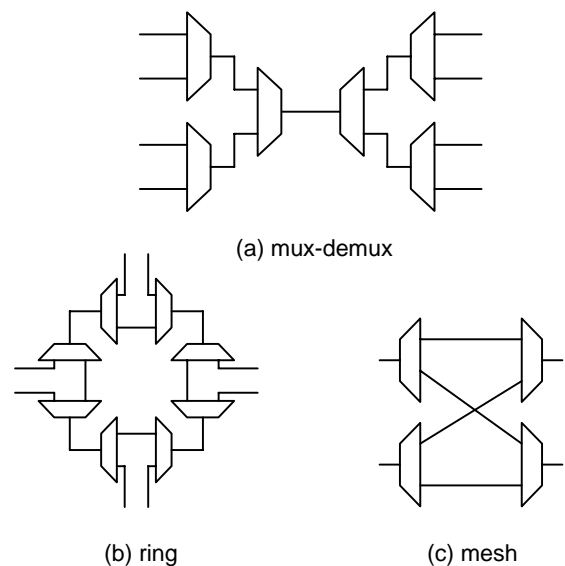


FIGURE 10: ALTERNATIVE CHAIN NETWORK TOPOLOGIES.

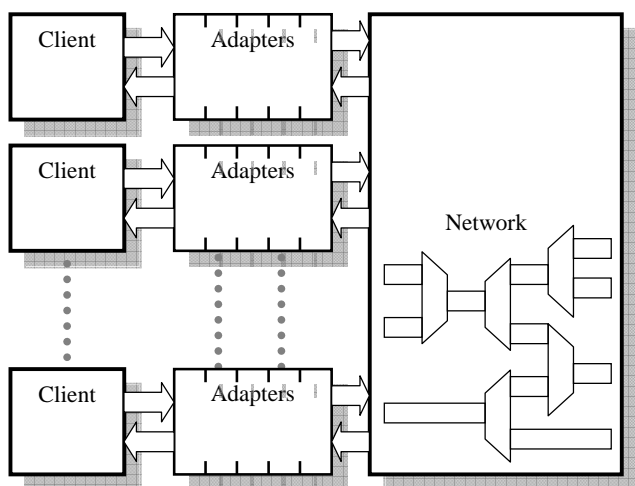


FIGURE 11: THE SILISTIX SELF-TIMED NETWORK-ON-CHIP DESIGN APPROACH.

Once the full SoC interconnect requirements have been specified, the Silistix tools will generate the appropriate network structure, selecting link widths, pipeline repeaters and switching topology to meet those requirements at minimum cost. The tools fit into a conventional SoC design flow as shown in Figure 12, removing any need for the designer to have any knowledge or expertise in self-timed design.

### FUTURE REQUIREMENTS

The Silistix tools will provide designer-friendly support for self-timed networks-on-chip that will deliver a robust and cost-effective solution to the requirement for a systematic approach to on-chip interconnect for complex systems-on-chip. Beyond this, the future presents many design challenges for on-chip interconnect.

Quality-of-Service (QoS) support is an obvious next-step, and the feasibility of providing QoS support on an asynchronous NoC has already been demonstrated [6]. At present the cost of implementing QoS is high compared with a best-effort network such as CHAIN, where it is always possible to provide dedicated links for specific connections that require guaranteed performance.

The current CHAIN technology assumes that on-chip logic is reliable. It is tolerant to delay variations, but it is not designed to recover from logic faults or failures. It may become necessary to design on the basis that on-chip logic is fallible, so that fault-tolerance becomes a requirement for all on-chip functions, including interconnect. Interconnect could be made fault-tolerant by building redundancy into the fabric, but it is likely to be more cost-effective (at the low expected error rates) to implement error detection and retry capabilities into the interfaces, effectively adding fault-tolerance as a layer on top of an unreliable fabric (as is the case with off-chip networks).

### CONCLUSIONS

Packet-switched networks-on-chip are the clear solution to the problem of complex SoC interconnect, and future developments will see advances in these networks to improve their performance, flexibility, power-efficiency and functionality. Support for Quality-of-Service protocols, fault-tolerance, secure communication and

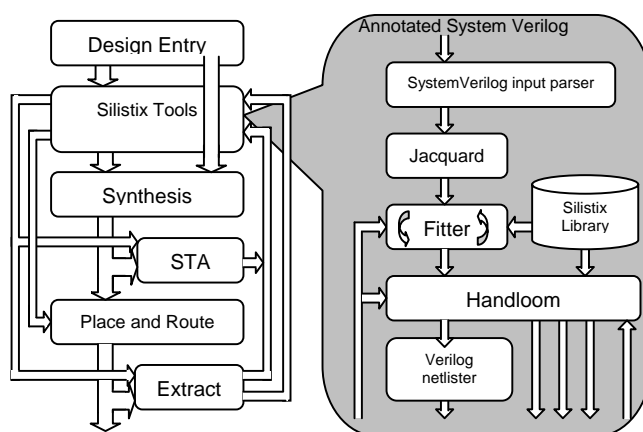


FIGURE 12: THE SILISTIX SELF-TIMED NETWORK-ON-CHIP DESIGN FLOW.

other similar high-level functions will emerge over the next few years to establish the NoC as the de-facto on-chip interconnect technology.

### ACKNOWLEDGEMENTS

The CHAIN self-timed network-on-chip technology that forms the basis of this paper was developed by Dr John Bainbridge at the University of Manchester. John was funded by EPSRC (the UK Engineering and Physical Sciences Research Council) through a PhD studentship and subsequently as a research assistant and research fellow. EPSRC's support is gratefully acknowledged.

The CHAIN technology is now owned by Silistix Limited who provided several of the diagrams in this paper.

### REFERENCES

- [1] J.D. Garside, W.J. Bainbridge, A. Bardsley, D.M. Clark, D.A. Edwards, S.B. Furber, J. Liu, D.W. Lloyd, S. Mohammadi, J.S. Pepper, O. Petlin, S. Temple, & J.V. Woods, "AMULET3i - an Asynchronous System-on-Chip", *Proc. Async 2000*, Eilat, Israel, 4-6 April 2000, pp 162-175.
- [2] W.J. Bainbridge & S.B. Furber, "MARBLE: An Asynchronous On-Chip Macrocell Bus", *Microprocessors and Microsystems*, **24**(4), 1 August 2000, pp. 213-222.
- [3] W.J. Dally & B. Towles, "Route Packets, not Wires: On-Chip Interconnection Networks", *Proc. DAC 2001*, ACM Press, New York, 2001, pp. 684-689.
- [4] W.J. Bainbridge & S.B. Furber, "Chain: A Delay-Insensitive Chip Area Interconnect", *IEEE Micro*, special issue on the Design and Test of System-on-Chip **22**(5), September/October 2002, pp. 16-23.
- [5] W.J. Bainbridge, L.A. Plana & S.B. Furber, "The Design and Test of a Smartcard Chip Using a CHAIN Self-timed Network-on-Chip", *Proc. DATE'04*, Vol. 3, Paris, Feb 2004, p. 274.
- [6] T. Felicijan & S.B. Furber, "An Asynchronous On-Chip Network Router with Quality-of-Service (QoS) Support", *Proc. IEEE International SOC Conference*, Santa Clara, CA, Sept. 2004, pp. 274-277.