inmos

The STC104 Asynchronous Packet Switch

Peter Thompson and Julian Lewis

INMOS Limited, Bristol, England

thompson@inmos.co.uk jdl@inmos.co.uk

SGS-THOMSON MICROELECTRONICS

inmos

Talk-Overview

1. System context (why? where?)
2. Chip architecture (what?)
3. Methodology and implementation (how?)
4. Performance (how much?)
5. Conclusion (how long?)

SGS-THOMSON MICROELECTRONICS

inmos

1. System Context

The ST C104 is a system-building chip. We will look at the following aspects of how it works in a system:

- the basic concept
- flexibility
- packets and routing
- the routing algorithm
- adaptive and universal routing

SGS-THOMSON MICROELECTRONICS

inmos

System Context: Basic Concept

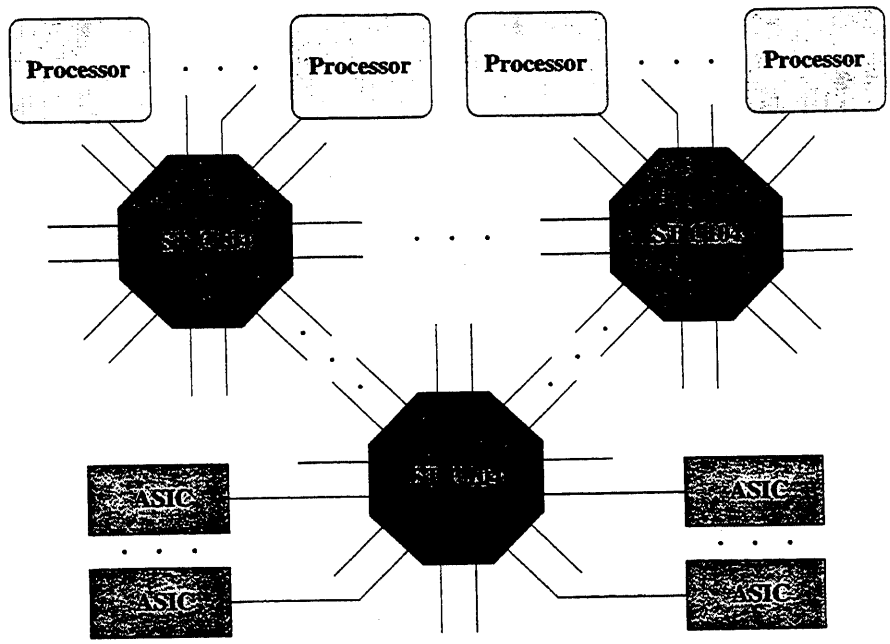
The ST C104 is a chip for constructing parallel systems of any size. It interconnects 32 bi-directional ports.

Some ports may be connected to other ST C104s in larger systems. Data can be transferred between different pairs of ports at the same time. Thus the system bandwidth increases with the number of ports.

Many ports can be active at the same time so each port need not be excessively fast. Thus they can be serial and asynchronous for ease of engineering. Each bi-directional serial connection is called a DS-Link.

SGS-THOMSON MICROELECTRONICS

inmos



SGS-THOMSON MICROELECTRONICS

inmos

System Context: Flexibility

For maximum applicability, the chip is as 'policy free' as possible:

- completely variable packet size
- small packets routed efficiently for fine-grain systems
- programable routing for different topologies
- data is not discarded except in error conditions
- extra features are programable options

SGS-THOMSON MICROELECTRONICS

inmos

System Context: Packet Routing

The ST C104 transfers data in *packets.*

The first one or two bytes of each packet are used as the header, which determines what the chip does with the packet.

Usually a path is opened across the internal crossbar, and the packet is pipelined across the chip until a termination marker is received ("wormhole routing").

This means:

- the low-load latency does not depend on the packet length
- packets of any length can be routed

If the required output is busy, the packet is *stalled.* The ST C104 arbitrates between requests for the same outputs, serving them in turn.

SGS-THOMSON MICROELECTRONICS

inmos

System Context: Routing Algorithm

The ST C104 compares the header of each packet with a set of programable ranges, one for each output port.

The packet is routed to the output in whose range the header falls ("interval routing").

This is like using a compressed routing table, which means:

- addresses are absolute, not relative, and are very short
- any network topology can be used
- any small *or* regular topology can be used optimally

The ST C104 optionally allows the packet header to be discarded for certain outputs. This allows high-level 'source routing' through multiple/hierarchical networks.

SGS-THOMSON MICROELECTRONICS

System Context: Adaptive and Universal Routing

The base routing is deterministic. This can be relaxed in two ways:

- several output ports can be programed to be equivalent, so the ST C104 uses whichever is free.

  This can be used to load-balance between links and for fault-tolerance.

- to prevent the formation of 'hot spots' in large networks, a two-phase algorithm called 'universal routing' can be used.

  This sends every packet first to a randomly-selected ST C104, and then from there to the original destination.

  This reduces 'difficult' traffic patterns to the average case, so the performance is uniform.

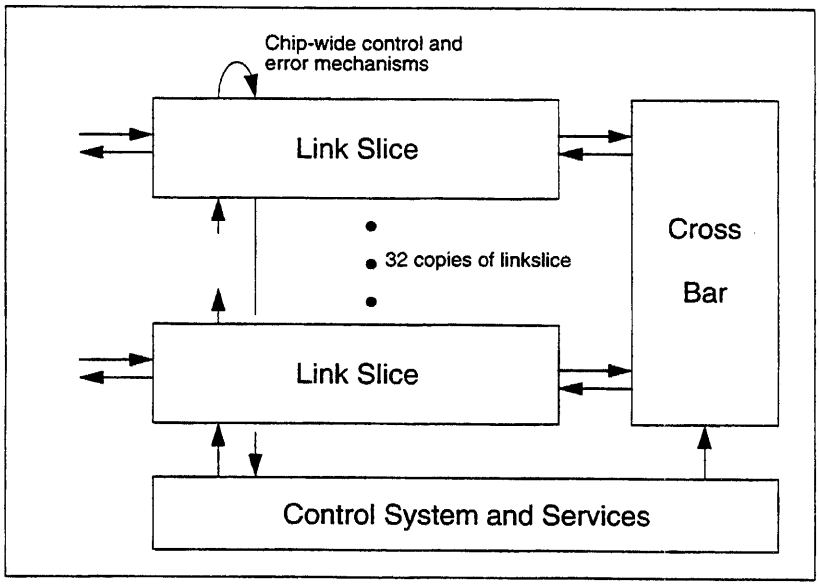
2. Chip Architecture

We will look at the following aspects of the ST C104 architecture:

- overall architecture
- concurrency and pipelining
- the crossbar
- the linkslice

inmos

Chip Architecture: Overall



SGS-THOMSON MICROELECTRONICS

inmos

Chip Architecture: Concurrency and Pipelining

The IMS C104 is a highly parallel machine. It consists of an array of 32 linkslices operating concurrently, each containing a pipeline of functional blocks.

There are *no* shared resources. This maximizes the benefit of replicating large functional blocks.

Each data path across the chip has 9 pipeline stages in the core clock domain. These pipelines include the crossbar, so they are created and dissolved dynamically.

The ST C104 contains in excess of 500 parallel state machines, containing over 1k bits of state. Since most of the state machines operate independently, the whole chip has more than 10^{300} states.

SGS-THOMSON MICROELECTRONICS

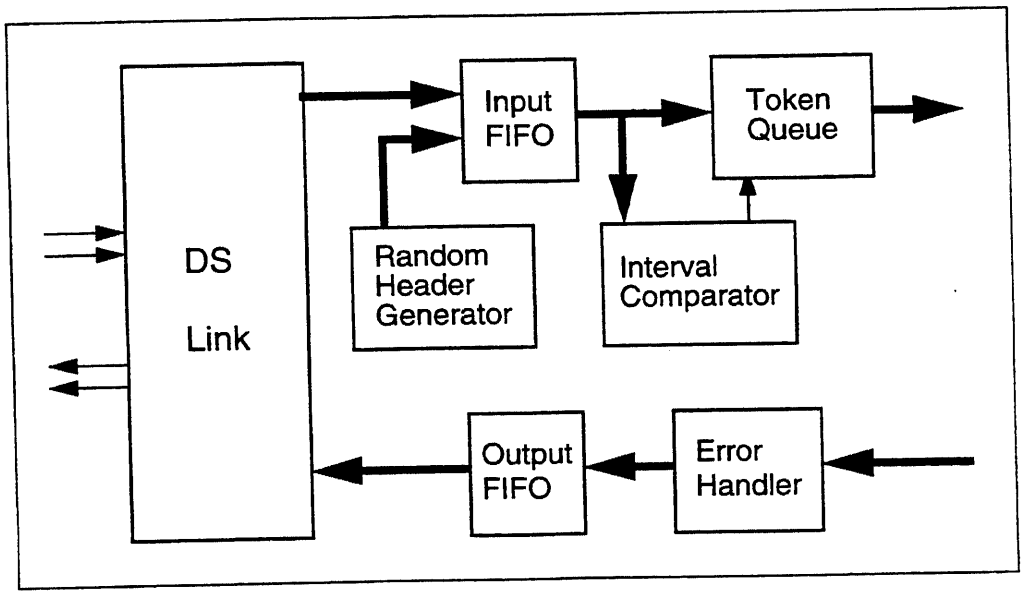
Chip Architecture: Crossbar Switch

The heart of the ST C104 is a full crossbar switch:

- 32 11-bit input ports (11 = 8 + 1 + 2)
- 32 11-bit output ports
- 32 round-robbin arbiters (one per output)
- 1024 11-way crosspoint switches
- peak bandwidth of 1.6 GBytes/s

(This needed full-custom layout!)

Chip Architecture: The Linkslice



inmos

3. Methodology and Implementation

- design style and effort
- chip details
- the link protocol
- the control system

SGS-THOMSON MICROELECTRONICS

inmos

Methodology and Implementation: Design Style and Effort

The ST C104 was initially specified in a natural language document.

Modelling was done in occam and VHDL. The micro-architectural decomposition was formally checked using a CSP algebra checker.

Automatic coverage analysis was used to develop a complete set of test vectors for each block.

Logic synthesis was used wherever possible, and full custom layout where necessary.

The total design time was approximately 10 engineer-years including all architectural, design and verification work.

SGS-THOMSON MICROELECTRONICS

inmos

Methodology and Implementation: Chip Details

The ST C104 is implemented in a 0.7 micron 3-level-metal CMOS process with stacked contacts.

It contains 1.875 million transistors and has a die-size of 16.5mm x 12.4mm (204.6 mm^2).

It is provided in a 208 pin CQFP package.

It uses TTL rails and signal levels.

External clock: 5MHz. Internal clocks: 50MHz, and 50-200MHz (programable), generated by PLLs. 34 main clock domains in total.

Power consumption: 7W

SGS-THOMSON

inmos

Chip Die Photo

(insert chip photo here)

SGS-THOMSON

inmos

Methodology and Implementation: The Link Protocol

Each bi-directional DS-Link uses a simple layered protocol:

- output signals are impedance-controlled and TTL-compatible;
- a serial bit-stream and a clock are transmitted on a pair of wires ("Data" and "Strobe") using a convention which allows a full bit-time of skew tolerance;
- the bit-stream is interpreted as a series of parity-checked tokens carrying data bytes and control information, including flow-control;
- the flow-controlled token streams are interpreted as a sequence of packets, beginning with a one or two byte header and ending with a distinguished control token.

This protocol is included in the IEEE P1355 draft standard.

SGS-THOMSON MICROELECTRONICS

inmos

Methodology and Implementation: Control System

The ST C104 contains a 'control unit' which deals with remote reset, error reporting and programing of the device.

The control unit is accessed through a DS-Link using a special protocol on top of the packet layer.

Another link is provided for 'daisy-chaining'. This allows any number of ST C104s to be individually controlled through a single DS-Link.

SGS-THOMSON MICROELECTRONICS

inmos

Peak Performance

A DS-Link operating at 200MBits/s can transfer a total of 38MBytes/s bi-directionally.

Counting each data stream only once, gives a peak aggregate chip bandwidth of just over 600 MBytes/s.

With a random address distribution, we get 60% of peak bandwidth on average, i.e. a sustainable bandwidth of 365 MBytes/s.

Since packets can be as short as 14 bits, the peak packet processing rate is 4.3×10^8 per second. The actual packet processing rate will depend on the length of the packets.

Under low load (or perfect non-contention for outputs), the latency across the chip is 475ns. With contention for outputs the latency is proportional to the length of packets.

SGS-THOMSON MICROELECTRONICS

inmos

Measured Performance

(Measurements not complete when going to press)

SGS-THOMSON MICROELECTRONICS

inmos

Current Status

The ST C104 is in production, with full documentation and software support.

Current chips are engineering samples which:

- have 100 MBit/s links (test limitation)
- run at 30MHz core clock speed (test limitation)
- are functional in prototype systems

SGS-THOMSON MICROELECTRONICS

inmos

Conclusion

The ST C104 is the first chip dedicated to providing reliable, concurrent, serial interconnect for digital system construction.

The combination of simple protocols on high-speed serial connections, wormhole routing, and interval routing, enables a high valency device to be constructed on a single CMOS chip.

Its features have been chosen to cover a wide range of system contexts, in particular to provide scalable interconnect bandwidth.

Its architecture is simple, but highly parallel for high performance.

SGS-THOMSON MICROELECTRONICS