

RN1: Low-Latency, Dilated, Crossbar Router (Extended Abstract)

Henry Minsky Tom Knight André DeHon
(hqm@ai.mit.edu) (tk@ai.mit.edu) (andre@ai.mit.edu)

MIT AI Lab
545 Technology Square, Cambridge, MA 02139
fax: 617.253.5060 voice: 617.253.7807 August, 1991

The RN1 chip (Figures 1 and 2) is a self-routing crossbar switch which forms the building block for a scalable, low-latency, fault-tolerant processor-to-processor interconnection network. The chip was designed to provide high enough performance so that the round trip delay for a remote processor-cache reference is on the order of today's single-processor main memory reference delays.

A multistage self-routing interprocessor communication network [Min91] [DeH90] can be built entirely from cascaded RN1 chips with no additional active or passive components. Routers communicate with each other via the RNP routing protocol [DKM91b]. The RN1 chip switches synchronously clocked byte-wide data channels with an associated control bit. The chips operate together to establish pipelined virtual circuits through the routing network. Once established, a circuit allows half-duplex bidirectional data transmission at the system clock rate. Data is not buffered at any place within the network. Information about the status of a circuit is automatically returned to the originator in otherwise unused pipeline cycles when the direction of data transmission is reversed.

The chip has eight byte-wide input and output ports, each having an associated control bit for out-of-band signalling (see Figure 2). The chip can be operated as either an 8x4 (dilation 2) crossbar or two independent 4x4 (dilation 1) crossbars (see Figure 1).

In a conventional crossbar, if an output channel in a given direction is in use by a circuit connection, any other message wishing to route in that direction will be blocked. With dilation, several independent messages can use the same logical output port. This improves the performance of the network under congestion. If there is a choice of more than one free channel in a logical direction, one is picked by the chip line-allocation circuits using a pseudo-random number generator. The dilation feature of the RN1, combined with randomized redundant wiring in the network, provides fault-tolerance by ensuring that there are multiple paths through the network for any source and destination node [DKM91a].

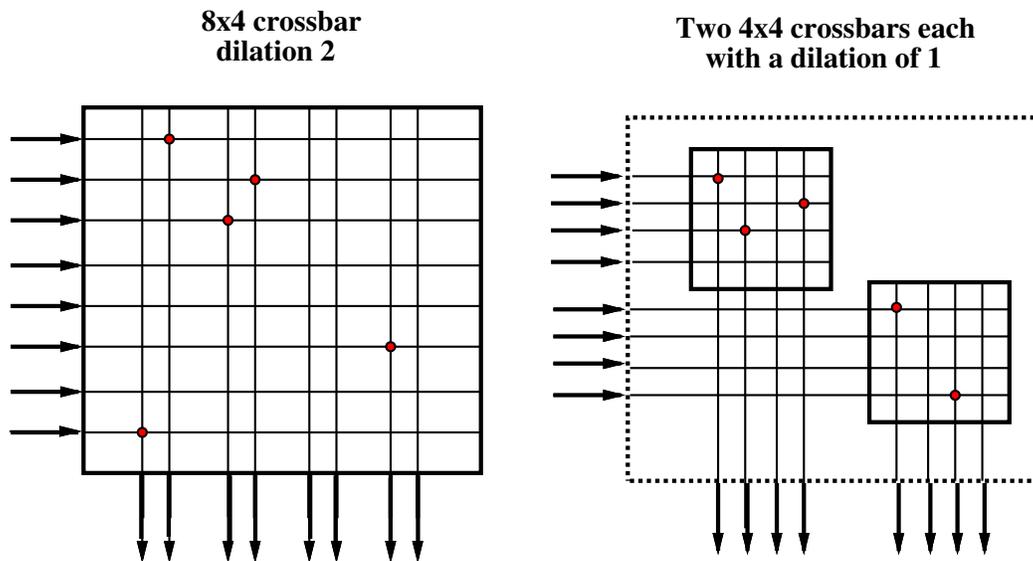
When several input ports attempt to open a connection to a logical output port, an 8 way arbitration for access to the output channel occurs. The arbitration is complicated by the dilation factor, in which the input ports are potentially competing for two resources rather than just one.

In order to speed up this critical path, a novel dynamic logic circuit is used, consisting of a dual manchester style carry-circuit with crossover shunts at each stage.

In order to save time, decoding of the destination of a new connection request is done simultaneously with the precharge of the dynamic logic circuits, thus efficiently utilizing both halves of the two-phase clock for computation.

The custom packaging and connectors provide dense three-dimensional stacking and short wire length interconnect between routing and processor boards by using the chip carrier itself as the board-to-board connector Figure 3. This eliminates the need for signals to be routed on and off a system backplane to go between boards.

The initial prototype, implemented in Hewlett-Packard's CMOS HP34 process, can route data at clock rates in excess of 50MHz. Latency through the five-volt I/O pads on the prototype account for a significant fraction of the component's total latency. For the next generation part, we intend to use custom ECL-compatible one-volt, controlled impedance pads. From our experience with RN1, we believe clock rates in the 100MHz to 200MHz range are achievable.

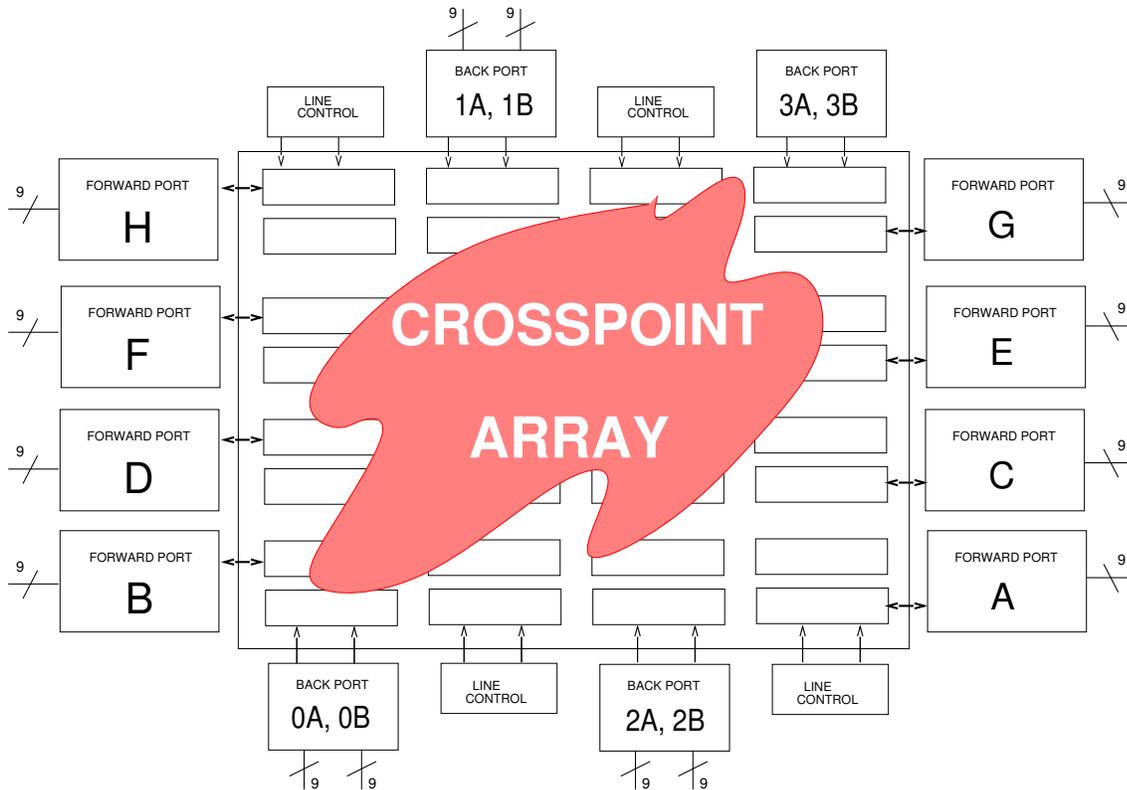


RN1 can be configured either to act as a single 8 input, 8 output, radix 4, dilation 2 router or to act as a pair of independent 4 input, 4 output, radix 4, dilation 1 routers.

Figure 1: RN1 Logical Configurations

References

- [DeH90] André DeHon. [Fat-Tree Routing For Transit](#). AI Technical Report 1224, MIT Artificial Intelligence Laboratory, April 1990. <http://www.cs.caltech.edu/~andre/abstracts/dehon_sb.html>.
- [DKM91a] André DeHon, Thomas F. Knight Jr., and Henry Minsky. [Fault-Tolerant Design for Multistage Routing Networks](#). In *International Symposium on Shared*



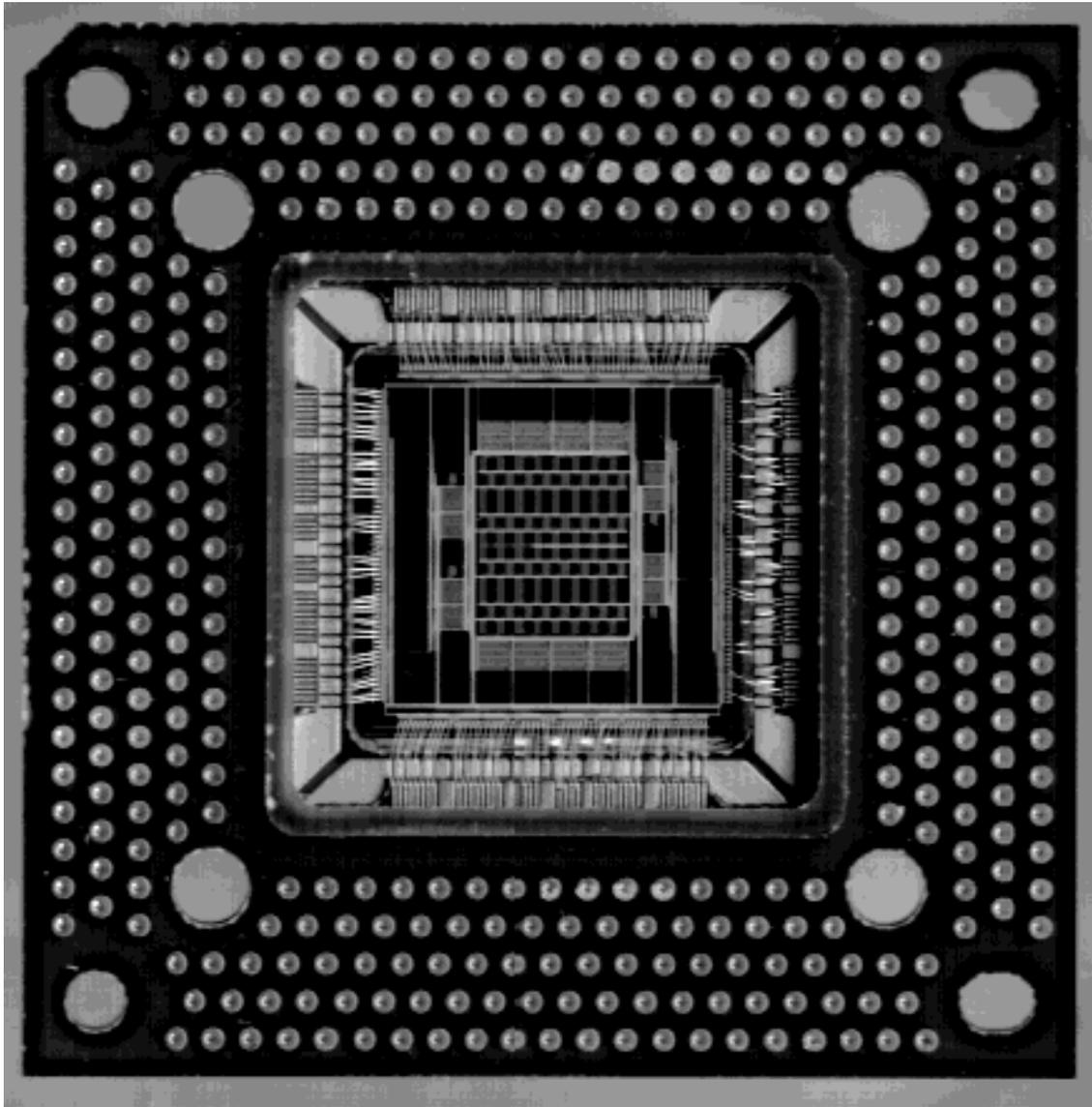
RN1 is constructed from 8 byte-wide input (forward) ports, 8 byte-wide output (back) ports, a crosspoint array for switching, and line-control modules for selecting between logically equivalent outputs. A ninth bit associated with each input/output port is used for out of band signalling between routers.

Figure 2: **RN1** Internal Logic Composition

Memory Multiprocessing, pages 60–71. Information Processing Society of Japan, April 1991. <http://www.cs.caltech.edu/~andre/ps/multipath_issmm91.ps>.

[DKM91b] André DeHon, Thomas F. Knight Jr., and Henry Minsky. RNP: Fault Tolerant Routing Protocol. Transit Note 41, MIT Artificial Intelligence Laboratory, March 1991. <<http://www.cs.caltech.edu/research/ic/transit/tn41/tn41.html>>.

[Min91] Henry Q. Minsky. A Parallel Crossbar Routing Chip for a Shared Memory Multiprocessor. Master's thesis, MIT, 545 Technology Sq., Cambridge, MA 02139, January 1991. Anonymous FTP <ftp://publications.ai.mit.edu/ai-publications/pdf/AITR-1284.pdf>.



RN1 is packaged in a 1.4 inch square dual sided pad grid array packaged. Land grids exposed on the top and bottom of the package are connected to printed circuit boards through mating button board connectors.

Figure 3: Packaged RN1 Component