The Cambridge Backbone Ring

David J. Greaves[*] Dimitris Lioupis[†] & Andy Hopper[‡]

February 23, 1994

Abstract

The Cambridge Backbone Ring (CBR) is a collaborative project between Olivetti Research Limited and the University of Cambridge Computer Laboratory. The project has designed and built an experimental ring communication network for integrated computer data, voice and real-time video applications. The network is currently aimed as a backbone facility for interconnecting groups of CFR (Cambridge Fast Ring) networks in the Cambridge area. Media access control is through the empty slot technique, with transmitting stations filling multiple slots each ring revolution. The network operates on monomode optical fibre and is designed for an eventual line rate of 1000 Mbit/sec

In the Backbone Ring architecture [1], the bandwidth of the fibre optic channel can be partitioned into a number (currently four) of TDM channels. This enables stations of varying cost and bandwidth to be attached to one network, parameterised by the number of channels a station can use concurrently. Unlike several other TDM ring projects in the literature, the Backbone Ring architecture achieves full connectivity, even with the simplest type of station.

1 Background and Requirement

One motive for the Cambridge Backbone Ring project is the provision of high bandwidth interconnection for clusters of Cambridge Fast Rings. The Cambridge Fast Ring (CFR) [2] is a slotted ring local-area network which operates at 75 Mbit/second over dual twisted pair and multimode optical fibres. The Backbone Ring uses an empty slot access protocol which is similar to that used for the CFR. Consequently,

[*]University of Cambridge Computer Laboratory. Email: djg@cl.cam.ac.uk

[†]Olivetti Research Ltd

[‡]Olivetti Research Ltd and University of Cambridge Computer Laboratory.

both rings are suitable for multi-media applications and can carry real-time traffic with very low delays.

The Backbone Ring is designed to cover an area of up to 50 km diameter which accords with the IEEE proposed specifications for a Metropolitan Area Network (MAN) [3]. The initial hardware was aimed to operate somewhere between 500 and 1000 Mbit/second, although the Backbone Ring architecture has been optimised for the 1 GHz frequency and above. Our basic station can achieve up to 200 Mbit/second on a point-to-point basis, but more typically we would envisage an active set of 25 such stations, each averaging a throughput of 25 Mbit/second. Higher bandwidth stations are also possible and we would envisage using these for Backbone Ring bridges, video file servers or in other cases where the cost is justified.

The Backbone Ring design makes use of a monolithic VLSI access device in order to avoid any digital interconnections operating at the clock rate. This device reduces the speed from 1 GHz serial to 100 MHz eight bits wide in silicon ECL. Further semi-custom devices increase the word width and reduce the word rate, but these can also provide selective access to the ring channel which provides for greater ratios of ring speed to host attachment speed. By varying this ratio, we can construct stations of various bandwidths. Another advantage is that when higher frequency VLSI devices become available, in principle we should be able to increase the ring operating rate while using the same station control logic and offering the same host interfaces.

The Backbone Ring forms a switch for short mini-packets. In many respects, these are no different from the 'cells' which are being used in ATM fast packet switching and proposed for broadband ISDN [4]. In the future, when bandwidth requirements are even higher than in our current research environment, a network of the Backbone Ring type can also serve as a local distribution loop for ATM traffic; we would hope to reduce the size of our current Backbone Ring stations so that we can also use them in the local area.

2 Choice of Topology and Access Method.

The ring topology was chosen for the Backbone Ring project partly because of the previous experience with rings at Cambridge. The benefits of the ring topology are well proven for local-area applications. These are primarily the use of point-to-point links and the low entry cost and ease of expansion which result from the absence of a central switch. The advances in electro-optical and VLSI technology have made it possible to extend the local-area techniques by multiplying both the clock rate and the length of cable by nearly one hundredfold. The Backbone Ring project therefore aims to examine whether the benefits accrue at the larger geometries. The ring topology does not suffer the headend asymmetry of the unidirectional bus topologies. Instead, its inherent symmetry forms a natural basis for fairness guarantee mechanisms, such as pass-on-free, and these do not become less fair as the geometry is increased.

A slotted ring has the electrical delay of the stations and cables permanently formatted into a number of fixed size slots which rotate. Stations place data into a slot as it passes, marking a full/empty flag full. When a receiving station detects its own address in the destination field of a slot, it copies the data from the slot into a receive buffer. The receiver marks a response bit at the end of the slot and when the slot has made its way back to the transmitter, the full/empty flag is cleared and the response read. In order to avoid hogging, the station must then pass the slot free to the next station. This reduces the maximum utilisation for N stations to about N/(N+1).

The empty slot media access technique was chosen for the Backbone Ring because our own simulations and those of others [1, 5, 6] have shown this to be suitable for real-time traffic. In order to provide fine grain sharing and for compatibility with the CFR at the lowest data-link level, known as the M-service access level [7], the Backbone Ring slot size was chosen to be the same size as that of the CFR. A slot holds a mini-packet which contains 16 bit source and destination addresses and a 256 bit data field.

As with the original Cambridge Ring, the standard CFR stations limit themselves to one transmission into a slot per ring revolution. This provides acceptable point-to-point throughput for a local-area network since there are typically only a few slots. One transmission per revolution also provides for an effective hardware re-transmission mechanism using the response field at the end of the slot. Re-transmission in hardware is valuable for local-area networks since it provides speed matching between different types of machines. In addition, With only one transmission per revolution, re-transmissions cannot result in out-of-order reception.

When the line rate is increased to one gigahertz and the cable length is also increased to around 200 kilometres, the number of slots increases such that the throughput of the CFR protocol would not be acceptable. To overcome this, the Backbone Ring relaxes the one slot per revolution rule, enabling stations to transmit multiple mini-packets in each revolution. (They are still obliged to pass on used slots free.) Two disadvantages of this are that re-transmissions can result in out-of-order reception and stations can monopolise successive slots yielding token-like behaviour. We have not attempted to solve the former problem, relying on the reassembly engines at the receiving stations, but the token like behaviour, which causes increased delays for high priority traffic, has been ameliorated by banning excessive successive slot fillings. This is inherent to the stations described in section 7 where stations do not transmit when they are receiving or have just received. It is also avoided in the enhanced architecture described in section 4.

3 Frame Format and Access Control

Header	Full Monitor Type	Data area for four mini-packets. Each contains nine 32 bit words including the routing tags.	Response and Qualifier	CRC
(4)	(4+4+4)	($4 \times 9 \times 32 = 1152$ bits)	(4+4)	(12)

Figure 1: Backbone Ring frame. It contains four CFR size slots.

The format of a Backbone Ring frame is shown in figure 3. The Backbone Ring slots are aggregated together into groups of four to form a frame. A 50 km diameter ring will have about 500 frames each separated by a number of 'syn' characters. Syn is a non-data codebook symbol (five bits) which is used for synchronisation. Syn characters are inserted or deleted by the elastic buffers in order to equalise clock rate differences. Another function of the elastic buffers is to extend the ring electrical delay to an integral number of code symbols.

Each frame starts with a header block. The header block is chosen to have a large Hamming distance from syn so that start-of-frame can be detected in the presence of a bit error. The frame then has the full/empty (FE), monitor passed (M) and type (T) flag bits of each slot. These are bit interleaved. The address and data fields of the four slots are interleaved on an 8 bit basis. The frame trailer contains response (R) and qualifier (Q) bits for each slot and a 12 bit CRC that covers all fields from FE to Q inclusive. The access control flags are bit interleaved so that they appear as a single word on the output of a station's 4B5B decoder. This enables protocol decisions for the frame to be made very quickly in parallel. The data fields of the four slots are byte interleaved since this is the basic unit of data for the demultiplexer devices which read and write the data fields. These are described in section 7.

On the outward direction the R and T single bit fields of a slot contain attribute information which can be used to specify quality of service if a mini-packet is going to traverse a further network. This is an extension to the CFR architecture where there are no additional bits available in the outward direction. In the Backbone Ring, the R bit can be used to specify whether an automatic re-transmission in case of data corruption or congestion at the receiver is required. For instance, it may be required for data traffic but not for voice. At the receiver the R bit is overwritten and in the return direction, the R bit indicates whether the slot was received correctly. The T bit carries priority information which can be used for queue jumping within a station. Alternative R and T bit protocols may be implemented in the station logic if required.

The Backbone Ring frame format is set up and maintained by one of the stations which is the active monitor. Any station can adopt this role without interrupting its normal operation. The monitor functions include initially clearing the ring and ensuring that all stations have gained correct bit synchronisation, then writing out a frame structure of the correct length. The maintenance functions are clearing permanently full slots through the use of the monitor passed (M) bit and passively monitoring the ring operation in case the format becomes corrupted.

The Backbone Ring uses an MSR (multiple slotted ring) MAC protocol. A station with data to transmit waits for a slot with a clear FE bit. It sets the FE, M, and Q bits, overwrites the T and R bits according to the data to be transmitted, and writes into the addressing and data fields. Transmitters are responsible for clearing the FE bits of slots that they used in the previous revolution so that they may be used by the next station in a fair manner. The transmitter also reads the response information from the receiver at this time. As described, transmitters do not limit themselves in the number of slots they use in a revolution. Receivers have a bit map held in a dedicated RAM of addresses on which they can receive. The destination fields of full slots are looked-up in the map as they pass. If the RAM entry is a one, then the contents of the slot are copied and the R bit updated.

At all stations, the CRC of received frames is checked and if bad, the Q bit of all four slots within the frame is cleared. Since the Q bit is set when the slot data is written, it indicates if the contents could have been corrupted during transmission. All stations recalculate a correct CRC for outgoing frames. Since the CRC is shared over four slots, a single error can result in the loss of as many as four mini-packets. However, since errors are very rare in fibre optic channels, this is more than compensated for by the increased channel efficiency resulting from sharing the CRC overhead.

Permanently circulating full slots are prevented by the monitor passed protocol. At the active monitor, the M flag of each slot is cleared and the old value copied to the FE flag. If the M flag were clear before, then the slot could not have been in use and the slot has been released. The livelock error condition is also corrected by this action, since then the FE flag would have been erroneously clear and is now set [8]. When the ring format is first written, all slots are marked full in order to prevent any immediate transmissions which might fail if the monitor changes the ring structure. The active monitor frees the slots after one successful revolution and transmissions commence.

4 Stations of Different Bandwidths

The Backbone Ring is designed to carry simultaneously traffic from about 10 to 50 active transmitting stations. Accordingly, the throughput at many of the stations will tend to be between 1/50 th and 1/10 th of the total ring bandwidth. Other stations, such as those at gateways and file servers, will require access to a greater proportion of the ring bandwidth. In order to meet these varying requirements, the Backbone Ring architecture enables stations of various bandwidths and complexities to be constructed. Relatively expensive, high bandwidth stations therefore need only be provided where their throughput is required; most stations can be of one of the simpler, lower throughput designs.

The provision of a relatively simple station which has access to only one quarter of the ring bandwidth is a primary motivation behind the frame containing four slots. This enables stations with bandwidths between one quarter and the full ring bandwidth to be constructed, while retaining full connectivity between the different types of station. When necessary, each of the four slots within a frame can be considered to be part of a separate TDM channel. The simplest type of receiving station is permanently assigned to one particular channel and then only needs the capacity to handle one quarter of the network bandwidth. This means it need only examine the address field of every fourth slot and the bandwidth requirement of the mini-packet storage buffers is similarly reduced. Transmitters must use an appropriate channel for a given receiver. This translates into knowledge of which slots within a frame can be used for a particular mini-packet. This information is held in a RAM at each station which is maintained by the software which maintains the receiving look-up tables. The least complex Backbone Ring stations therefore have a maximum host bandwidth of about 1/4 of the total ring bandwidth. This facilitates a worthwhile reduction in station complexity, and therefore, station cost.

The next most complicated type of station consists of two simplex sections. One half of the station transmits while the other half receives. Simultaneous transmission and reception into separate slots of the same frame is possible. As described later in section 6, a fully equipped Backbone Ring station has four half-duplex sections and, if necessary, is able to transmit into or receive from all of the slots in a frame.

Enhanced Backbone Ring Architecture: Fixed assignment of receivers to physical channels can result in one channel being overloaded while free bandwidth remains on another. This is overcome in the Enhanced Backbone Ring architecture, although we have not implemented this in hardware yet. In the enhanced architecture, receivers are assigned to one of 32 logical channels instead of one of the 4 physical channels. This assignment imposes no further management overheads and it does not matter if some logical channels remain unused. The logical-physical mapping is changed by a well known, pseudo-random rule each frame time. The mapping is complete, meaning that every logical channel is given a physical channel in each frame. An important property of the rule is that two logical channels which map to the same physical channel in one frame, map to distinct physical channels in subsequent frames. Transmitters evaluate the mapping to determine which channel to use for a mini-packet. The mapping can be generated by a simple PLA or held in a 1024 by 2 PROM. The mapping causes traffic to be spread evenly over the physical channels, avoiding overload. Another advantage is that bulk arrivals of mini-packets, resulting from higher-level message fragmentation, are split over the physical channels. This removes a main cause of the full slot clustering which can often cause token-like behaviour of a slotted ring. As a result, the delay for priority traffic is reduced. This is demonstrated in the next section.

5 Backbone Ring Simulations

Frames	100	
Slots	400	
Frame time	1.2	microseconds
Latency	128	microseconds
Fibre length	25.6	kilometres
Data rate	800	megabit/second
Stations	18	
Addressing overhead	0	bits

Table 1: Physical parameters for networks simulated in this paper.

Name	Priority	Arrival process	Inter-arrival	Bulk size	Bandwidth Kbits/sec
Voice	4	Synchronous	2 ms	1	128
Music	3	Synchronous	87 us	1	2900
Low video	3	Synchronous	87 us	1	2900
File	1	Poisson	500 ms	45	23
RPC	2	Reciprocal	2 ms	3	2 × 750
Total					7451

Table 2: Traffic source model for a multi-media user.

Priority level	4	3	2	1
Throughput (Mbit/sec)	9.41	424	88	2
Backbone Ring with pipeline blocking	69 130	74 150	92 385	245 305
Backbone Ring without pipeline blocking	69 130	72 145	86 225	247 310
Enhanced Backbone Ring without pipeline blocking	66 127	67 126	78 147	307 335
Source-release slotted ring without channels	62 120	62 123	67 127	120 161
Early-release token ring with exhaustive service	199 444	243 454	247 467	197 369

Table 3: Simulation results for five different types of ring with the stated dimensions and 72 multi-media users. The upper figure is the mean message delay and the lower is the 99^{th} percentile, both in microseconds.

Some brief traffic simulation results are presented in this section. The physical parameters shown in table 1 were used for the simulations. The overheads of addressing and control fields are not included in the simulations since their effects are easily applied to the final results, although this has not been done here. In addition, the decision of which overheads should be included is often somewhat arbitrary. Applicable overheads are the data-link level header of four bytes taken from the 32 byte mini-packet [9, 10], the addressing overhead of four bytes external to the mini-packet and the MAC flags and CRC overhead of five bytes per four mini-packets.

The Backbone Ring simulations reported here employ the concept of a multi-media user. The main simulation parameter is the number of users supported by the network. The user is defined in table 2. At the start of the simulation, each user establishes a set of virtual connections across the network. The set consists of a 64 kilobit, voice channel containing 2 milliseconds of speech per mini-packet (the amount which can be lost without subjective channel degradation), a forward-error corrected music stream (of the type used on compact discs), a real-time video channel (of the type used on our 'Pandora' workstations), some bulk file transfers and some remote procedure call traffic. The last is interesting since the RPC packets are queued for transmission an exponentially distributed time after receipt of the last RPC message from a partner on another station. The total load per user is nearly one percent of the simulated network capacity.

The results in table 5 show the mean and 99^{th} pecentile of delay for the last mini-packet of a message to arrive at the destination station from queuing at the source station. The first line of results is for our basic half-duplex station which can transmit or receive from one slot within each frame. Both the mean and 99^{th} percentile delays are very acceptable. The higher priority traffic receives lower delay, showing that priority through expediting queues within a station is effective. This is a result of the short mini-packet and the fine-grain sharing intrinsic to a slotted ring.

As to be described in section 6, this basic station cannot transmit in a frame if it received a mini-packet from the previous frame. The second line of results shows the performance of an equivalent station which does not suffer from this. There is very little difference.

The third line of results shows the performance of the enhanced architecture. This makes the priority mechanism even more effective, resulting in delays for the voice traffic which consist nearly entirely of the propagation delay to the destination.

The last two lines of results are included for comparison purposes. The slotted ring without channels offers slightly better performance than the Backbone Ring architecture owing to the possibility of transmitting into several slots which would otherwise be in the same frame. The early-release token ring offers greater delays and does not properly respect the priority requests of the traffic. The token performance might be considered acceptable with the fairly short length of fibre assumed for these results, but it degrades much more rapidly than the slotted systems as the length

Figure 2: Block diagram of a simple Backbone Ring station.

of cable is increased [1].

6 Backbone Ring Hardware Design

The Backbone Ring uses 4B5B/NRZI modulation. Monomode optical fibre is the physical medium. Each station uses a local, crystal-locked transmit clock and re-times the received data using an elastic buffer. These techniques add about 50 bits of delay per station, but this is of no consequence when compared with the intrinsic cable delay at these dimensions.

Figure 2 shows the hardware configuration of a simple half-duplex station. The received data from the optical fibre is passed to an ECL access chip which performs two types of function. The first concerns the transformation of the serial data to eight bits wide and provision of clock and byte synchronisation. This requires the following functions:

- Decide whether the current input voltage is a one or zero using a D-type flip-flop,
- Convert from NRZI to NRZ,
- Decode 5 bit blocks into 4 bit data word or special code 'syn' used for synchronisation,
- Gain bit synchronisation using a slip method,
- Signal a code error on codebook violation,
- Gain frame synchronisation on frame header, and
- Encode output stream into 4B5B and thence NRZI.

The second type of function provided by the access chip is at the frame level and includes:

- Check CRC of received frame,
- Examine and update F/E bits to decide which slot is to be read or written,
- Generate data strobes for reading or writing the desired channel,
- Check and update response and qualifier bits of outgoing frame and
- Write new valid CRC to outgoing frame.

The 8 bit data streams from the ECL access chip are connected to the RAM packet buffer through the semi-custom demultiplexer devices. Two similar devices are used in the basic station. Their purpose is to widen the data from 8 to 32 bits. This data widening results in a clock rate reduction of 4 to 1 and there is a further 4 to 1 reduction within these devices since they give access to only one of the four TDM channels at a time. The rate of the 32 bit words is then 6.25 MHz for a 1 GHz implementation, a rate and bus width which are fairly convenient for direct connection to computer backplane buses.

Two demultiplexer devices are used in the basic station to artificially increase the station delay. This enables a single bank of buffer memories to be used, both for receiving and transmitting. The first device is used to look-up the address field of the station's receive channel slot in advance of data transfer to or from the ring through the second device. Since there is only one set of buffer memories, a transmission cannot be started before it is known whether a receive is to take place from the current frame. The additional station delay is sufficient for this to be known before the start of the frame containing the flag control bits has left the station. Owing to pipeline effects when using the same buffer for receive and transmit, with this type of station it is not possible to transmit in a frame if a reception was made in the previous frame. As shown in the last section, this does not result in a significant degradation of transmission throughput.

The packet buffer at the basic station currently consists of four, byte-wide CMOS devices and they can hold 2048 mini-packets. There is no fixed partitioning between the receive and transmit queues. Instead the control logic implements FIFO chained lists of mini-packets. Separate queues are required for receive and for each priority level of each transmit channel. The transmit side of the station logic can interface to the host device as the consumer process for a shared cyclic buffer in host address space. It can therefore handle long messages consisting of multiple mini-packets with very little fragmentation overhead for the host. The receive side of a station is potentially more complex, requiring the simultaneous reassembly of several messages. Fortunately this function is not required within our first generation CFR to Backbone Ring bridges.

Two priority levels are supported using simple expedited service of the higher level within a channel. There is no access layer priority using the MSR protocol.

One ECL access device at a station remains sufficient for the higher bandwidth configurations, but the number of demultiplexer devices can be increased. In the dual simplex station, which has twice the bandwidth of the basic station, two deserialiser devices remain sufficient, but now one is dedicated to receive and one is dedicated to transmit. The additional hardware requirement consists of another set of buffer memories and buffer address generators. Access to a greater number of channels is gained by stacking further demultiplexer chips in parallel. The outline of a fully equipped station is shown in figure 3.

Inevitably, the ring data finds itself on a 128 bit wide bus, the provision of which was avoided in the simpler stations. Owing to the cost of looking-up the destination addresses of four slots at once, the complexity of the fully equipped station is probably greater than could be achieved without the channelised architecture. It is our contention that the additional cost of this configuration is justified by the low entry cost of the basic station configuration.

The physical medium carries a maintenance sub-carrier in addition to the main data channel [11]. The sub-carrier uses mean-power modulation of the optical sources and so can operate even when the ring frame structure has failed. The maintenance channel carries 1200 baud BFSK HDLC frames from station to station in the same direction as the ring data. At each station an eight bit microprocessor monitors various status and alarm signals and sends the information as telemetry in the forward direction. The received telemetry is also forwarded until it has made one

Figure 3: One configuration for a fully equipped Backbone Ring station.

ring revolution. The telemetry is used to isolate failed ring sections and detect the need for a reframe. A terminal may be attached to any station for engineering maintenance, where measured error statistics can be observed and faults accurately pin pointed. The telemetry system is responsible for electing the active ring monitor.

7 Current Implementation

The hardware design of the Backbone Ring consists of a set of chips which can be used for nodes and network interfaces of various types. The chips consist of the ECL access chip, the demultiplexer devices and a CMOS gate array for the station control logic.

The ECL device consists of about 1000 gates and has been implemented as a silicon gate array with gate delays of about 180 ps. This dissipates about 5 watts and can be clocked above 600 MHz. A follow on version is envisaged which will achieve the target speed of 1 GHz.

The demultiplexer devices have been fabricated using inexpensive CMOS gate arrays which were directly drawn by E-beam. These will also have to undergo a second revision to achieve the 1 GHz rate where they will then be clocked at 100 MHz. We envisage using Bi-CMOS gate arrays.

The station logic mainly consists of the map RAMs and the associated control logic.

This control logic is being made using a CMOS gate array which operates at 6.25 MHz in the full speed system.

The fibre optic modules have been constructed using commercial 1300 nanometre laser diode and pin-fet components with 8/125 micron fibre. Monomode optical fibre cables have been laid between the Olivetti Research Laboratories and the University Computer Laboratory in Cambridge. It is our hope that very soon we will be able to send real-time voice and video over the Backbone Ring between the multi-media workstations connected to the CFRs at each site. In future, it would be nice to connect other sites to the experimental Backbone Ring.

Acknowledgements

Thanks are due to the following people who have contributed to networking research in Cambridge and towards the Backbone Ring project: Ian Leslie, Derek (Mac) McAuley, David Milway, Roger Needham, Peter Newman, Cosmos Nicoloau, John Porter, Brian Robertson, David Tennenhouse, Roy Want, Ian Wilson and many more.

References

[1] 'Multi-access metropolitan area networks.' DJ Greaves. University of Cambridge PhD dissertation, October 1989.

[2] 'The Cambridge Fast Ring Networking System.' A Hopper and RM Needham. IEEE transactions on computers, Vol 37 number 10. October 1988.

[3] 'Overview of metropolitan area networks.' RW Klessig. IEEE Communications Magazine Vol 24 No 1. Jan 86.

[4] 'CCITT (draft) series I.121 recommendations for broadband ISDN.' CCITT 1988.

[5] 'Performance analysis of Slotted Ring Protocols in HSLANs.' M Zafirovic-Vukotic, IG Niemegeers and D Valk. University of Twente, The Netherlands. IEEE J-SAC Vol6 No 6, July 1988.

[6] 'Analysis of channel access schemes for high-speed LANs.' NN Pedersen and R Sharp. SIGCOMM 84 Symposium on communications architectures and protocols, Montreal, Canada pp 164-171.

[7] 'CFR M-Access Service Definition.' AM Chambers. Project Unison document reference UA010. 5th Nov 86.

[8] ‘The livelock-free protocol of the Cambridge Ring.’ Prof DJ Wheeler. The Computer Journal, Vol 32 No 1, page 95. 1989.

[9] ‘The CFR Unison Data Link Service Specification.’ DL Tennenhouse. Project Unison document reference UC021. 24th Oct 86.

[10] ‘Cambridge HSLAN protocol review.’ DJ Greaves, ID Wilson. Proceedings of IFIP TC6 International Workshop on ‘Protocols for high-speed networks’ edited by H Rudin and R Williamson, held at IBM Ruschlikon 1989. Elsevier 1989.

[11] ‘An optical fibre supervisory system.’ AJ Sharland (et al). Br Tel Tech J Vol 1 No 1. July 1983.