

Fast Packet Switching for Integrated Services

Peter Newman

Wolfson College
University of Cambridge

A dissertation submitted for the degree of
Doctor of Philosophy

December 1988

Abstract

As the communications industry continues to expand two current trends are becoming apparent: the desire to support an increasing diversity of communications services (voice, video, image, text, etc.) and the consequent requirement for increased network capacity to handle the expected growth in such multi-service traffic. This dissertation describes the design, performance and implementation of a high capacity switch which uses fast packet switching to offer the integrated support of multi-service traffic. Applications for this switch are considered within the public network, in the emerging metropolitan area network and within local area networks.

The Cambridge Fast Packet Switch is based upon a non-buffered, multi-path switch fabric with packet buffers situated at the input ports of the switch. This results in a very simple implementation suitable for construction in current gate array technology. A simulation study of the throughput at saturation of the switch is first presented to select the most appropriate switch parameters. Then follows an investigation of the switch performance for multi-service traffic. It is shown, for example, that for an implementation in current CMOS technology, operating at 50 MHz, switches with a total traffic capacity of up to 150 Gbits/sec may be constructed. Furthermore, if the high priority traffic load is limited on each input port to a maximum of 80% of switch port saturation, then a maximum delay across the switch of the order of 100 μ secs may be guaranteed, for 99% of the high priority traffic, regardless of the lower priority traffic load.

An investigation of the implementation of the switch by the construction of the two fundamental components of the design in 3 μ m HCMOS gate arrays is presented and close agreement is demonstrated between the performance of the hardware implementation and the simulation model. It is concluded that the most likely area of application of this design is as a high capacity multi-service local area network or in the interconnection of such networks.

Preface

I wish to thank my supervisor, Roger Needham, for his encouragement and valuable advice during the course of my research. I would also like to thank Ian Leslie, David Wheeler, David Tennenhouse and David Milway for many useful discussions.

I am also indebted to the following people who have read and commented upon this dissertation: Roger Needham, Ian Leslie, David Wheeler and Brian Robertson.

I would like to make a special note of thanks to my parents and to Dr. Elizabeth Bates for the support and encouragement that they have given me during the period of my research.

Throughout my research I have been supported by a grant from the Science and Engineering Research Council for which I am also grateful.

Except where otherwise stated in the text, this dissertation is the result of my own work and is not the outcome of any work done in collaboration. Furthermore, this dissertation is not substantially the same as any that I have submitted for a degree, diploma or any other qualification at any other university. No part of this dissertation has already been or is being concurrently submitted for any such degree, diploma or any other qualification.

Contents

List of Figures	xi
List of Tables	xiv
Glossary of Terms	xv
1 Introduction	1
1.1 Objectives	2
1.2 Outline	2
1.3 Growth of Communications Networks	3
1.4 Multi-Service Traffic	6
2 Statistical Switching Mechanisms	9
2.1 Multiplexing	9
2.2 Switching Mechanisms	11
2.3 Evolution of the Packet Switch	16
2.4 Fundamentals of Fast Packet Switch Design	18
2.5 Summary	20
3 Fast Packet Switch Architecture	23
3.1 A Simple Classification of Switch Designs	23
3.2 Input Buffered Switches	25
3.3 Output Buffered Switches	28
3.4 Internally Buffered Switches	32
3.5 Performance Comparison	37
3.6 Summary	38
4 Multi-Stage Interconnection Networks	41

4.1	An Introduction to Interconnection Networks	41
4.2	The Crossbar Network	46
4.3	Banyan Networks	47
4.4	The Clos Network	50
4.5	The Beneš Network	50
4.6	The Batcher Sorting Network	52
4.7	Summary	53
5	Design of the Cambridge Fast Packet Switch	55
5.1	Binary Routing Networks	55
5.2	Design Issues	57
5.3	The Switching Mechanism	59
5.4	The Switch Fabric	61
5.5	The Multi-Plane Switch Structure	66
5.6	Summary	67
6	Switch Fabric Performance	69
6.1	Traffic Models	69
6.2	The Simulation Model	70
6.3	The Crossbar Switch Fabric	71
6.4	The Delta Network	75
6.5	The Beneš Network	80
6.6	Summary	83
7	Performance for Multi-Service Traffic	85
7.1	Multi-Service Traffic Requirements	85
7.2	Extensions to the Switch	87
7.3	Traffic Models	87
7.4	Poisson Traffic	88
7.5	Talkspurt Voice	91
7.6	Packet Length	93
7.7	Buffer Overflow	95
7.8	Discussion	96
7.9	Summary	98

8	Implementation of the Fast Packet Switch	99
8.1	An Experimental Implementation	99
8.2	The Switching Element	100
8.3	The Input Port Controller	104
8.4	Performance Measurements	105
8.5	Towards a Full-Scale Switch Implementation	106
8.6	Summary	112
9	Conclusion	113
9.1	Summary	113
9.2	Discussion	117
9.3	Multicast Operation	119
9.4	Network Aspects	121
	Appendix: Simulation results for throughput at saturation	123
A	Simulation Results for Throughput at Saturation	123
	References	131

List of Figures

1.1	Relationship between bit rate and holding time for various classes of traffic and switching mechanisms.	8
2.1	An example of time division multiplexing (TDM).	10
2.2	The spectrum of switching mechanisms.	12
2.3	A single path decentralised packet switch.	17
2.4	General structure of a fast packet switch.	19
3.1	A simple classification of fast packet switch design.	24
3.2	Basic structure of the three phase Batcher-banyan switch.	26
3.3	A two-plane switch structure.	27
3.4	The Starlite fast packet switch.	28
3.5	Structure of the Knockout Switch.	30
3.6	The bus interface of the Knockout Switch.	31
3.7	Structure of the Prelude switching element.	33
3.8	Structure of the Bus Matrix switching element.	34
3.9	Structure of the IBM switching element.	35
3.10	The output queueing segment of the IBM switching element.	35
3.11	Structure of the TDM Bus switching element.	36
4.1	A simple classification of interconnection networks.	42
4.2	Examples of regular static network topologies.	43
4.3	A single stage 8×8 shuffle exchange.	44
4.4	A single sided Clos network.	44
4.5	The general representation of a crossbar network.	46
4.6	A self-routing crossbar switch.	47
4.7	The general structure of a delta network.	48

4.8	Examples of 8×8 delta networks constructed from 2×2 switching elements.	49
4.9	A square three stage Clos network.	51
4.10	An 8×8 Beneš network constructed from 2×2 switching elements.	51
4.11	An 8×8 bitonic sorter.	52
4.12	A 16×16 full sorter.	53
5.1	The structure of a buffered binary routing node.	56
5.2	The basic structure of the Cambridge Fast Packet Switch.	59
5.3	A 64×64 delta network of 8×8 switching elements.	62
5.4	A 16×16 modified delta network of 8×8 switching elements.	63
5.5	A 64×64 Beneš network of 8×8 switching elements.	65
5.6	A 32×32 sub-equipped Beneš network of 8×8 switching elements.	66
5.7	A two-plane switch structure.	67
6.1	Throughput at saturation for the crossbar switch fabric.	72
6.2	Analysis and simulation of mean delay performance for slotted traffic.	74
6.3	Mean delay performance of crossbar switch structures for slotted traffic.	74
6.4	Throughput at saturation for single plane input buffered flooding delta networks.	75
6.5	Throughput at saturation for multiple delta networks in parallel.	77
6.6	Comparison of algorithms to select a free path across the network.	78
6.7	Throughput at saturation for two-plane pure input buffered delta networks.	78
6.8	Comparison of mean delay performance for slotted traffic of various switch structures of size 64×64 .	79
6.9	Throughput at saturation for flooding Beneš structures.	81
6.10	Comparison of mean delay performance for slotted traffic of 64×64 sub-equipped Beneš networks against other structures.	82
7.1	Throughput performance for the Poisson reserved service + saturated unreserved service traffic model.	89
7.2	Maximum reserved service packet delay for the Poisson reserved service traffic model with and without saturated unreserved service traffic.	89
7.3	Unreserved service throughput performance for the Poisson reserved service + Poisson unreserved service traffic model.	90
7.4	Mean unreserved service packet delay for the Poisson reserved service + Poisson unreserved service traffic model.	90

7.5	Comparison of maximum delay performance of various switch designs of size 64×64 for Poisson traffic with and without saturated unreserved service traffic.	91
7.6	A comparison of maximum reserved service packet delay for Poisson, talkspurt and TDM voice models both with and without saturated unreserved service traffic.	93
7.7	Effect of unreserved service packet length on throughput performance for the Poisson reserved service + saturated unreserved service traffic model.	94
7.8	Effect of unreserved service packet length, constant and exponentially distributed, on maximum reserved service packet delay.	94
7.9	Buffer overflow probability for the input buffered crossbar switch.	96
8.1	Structure of the 4×4 crossbar switching element.	100
8.2	Implementation of a 1 to 4 selector.	102
8.3	Implementation of a 4 to 1 arbiter.	103
8.4	The experimental input port controller.	105
8.5	Structure of a fast packet switch implementation.	107
8.6	The I/O port controller.	108
9.1	Switch structure for multicast operation.	120
9.2	General model of protocol structure for a network of fast packet switches.	122

List of Tables

1.1	Multi-service traffic characteristics.	7
6.1	Switch fabric design parameters.	70
6.2	Percentage error in throughput at saturation of simple model for delta networks with switching elements of degree 8.	79
6.3	Percentage error in throughput at saturation of simple model for single plane delta networks.	80
7.1	Comparison of maximum delay performance of various 64×64 switch designs at maximum reserved service traffic load.	92
8.1	Estimated complexity of crossbar switching elements.	109
8.2	Approximate maximum bandwidth per switch port for various implementation technologies.	110
A.1	Throughput at saturation for crossbar switch fabrics.	123
A.2	Throughput at saturation of delta networks with switching elements of degree 2 for a searching algorithm.	124
A.3	Throughput at saturation of delta networks with switching elements of degree 4 for a searching algorithm.	124
A.4	Throughput at saturation of delta networks with switching elements of degree 8 for a searching algorithm.	125
A.5	Throughput at saturation of delta networks with switching elements of degree 16 for a searching algorithm.	125
A.6	Throughput at saturation of delta networks with switching elements of degree 2 for a flood-planes algorithm.	126
A.7	Throughput at saturation of delta networks with switching elements of degree 4 for a flood-planes algorithm.	126
A.8	Throughput at saturation of delta networks with switching elements of degree 8 for a flood-planes algorithm.	127

A.9 Throughput at saturation of delta networks with switching elements of degree 16 for a flood-planes algorithm. 127

A.10 Throughput at saturation of sub-equipped Beneš networks with switching elements of degree 2. 128

A.11 Throughput at saturation of sub-equipped Beneš networks with switching elements of degree 4. 128

A.12 Throughput at saturation of sub-equipped Beneš networks with switching elements of degree 8. 129

A.13 Throughput at saturation of sub-equipped Beneš networks with switching elements of degree 16. 129

Glossary of Terms

ATD:	Asynchronous Time Division, (see page 5).
ATM:	Asynchronous Transfer Mode, (see page 5).
BiCMOS:	A high speed implementation technology integrating both bipolar and CMOS devices on the same integrated circuit, (see page 110 and table 8.2).
B-ISDN:	Broadband ISDN, (see page 5).
CAD:	Computer Aided Design.
CAM:	Computer Aided Manufacture.
CATV:	Community Antenna Television, i.e. cable TV.
CCITT:	The International Telegraph and Telephone Consultative Committee.
CMOS:	Complementary Metal Oxide Semiconductor — An implementation technology, (see page 110 and table 8.2).
DTDM:	Dynamic TDM, (see page 11).
ECL:	Emitter Coupled Logic — A high speed implementation technology, (see page 110 and table 8.2).
FDDI:	Fiber Distributed Data Interface — A high speed local area network.
FIFO:	First In First Out — A queueing discipline.
GaAs:	Gallium Arsenide — A very high speed implementation technology, (see page 110 and table 8.2).
HC MOS:	High speed CMOS — An implementation technology, (see page 99).
HDLC:	High-level Data Link Control — A popular data link layer protocol.
I/O:	Input/Output.
ISDN:	Integrated Services Digital Network, (see page 5).
LAN:	Local Area Network, (see page 3).
MAN:	Metropolitan Area Network, (see page 4).
PABX:	Private Automatic Branch Exchange — A private telephone exchange.
STM:	Synchronous Transfer Mode, (see page 5).

- TASI:** Time Assignment Speech Interpolation, (see page 13).
- TDM:** Time Division Multiplexing, (see page 9).
- TTL:** Transistor Transistor Logic — An implementation technology.
- VCI:** Virtual Circuit Indicator, (see page 107).
- VLSI:** Very Large Scale Integration — An integrated circuit containing a large number of active devices.