PHOTONIC NETWORKS

AND

NETWORK CONTROL THROUGH CHARGING

PhD Thesis

by

MARTIN SABRY

Department of Electronic and Electrical Engineering

UNIVERSITY COLLEGE LONDON

TORRINGTON PLACE LONDON WC1E 7JE
To Far and John
Success should not be measured in heights attained, but by obstacles overcome
Abstract

This thesis will describe a number of concepts and ideas concerned with Advanced Telecommunication Networks.

The thesis starts with an introduction where the convergence of the consumer electronics, media, computing and telecommunication network technology industries is summarised. We discuss the services that a (future) telecommunication infrastructure may offer and some of the many challenges that this could foretell. The remainder of the thesis is made up of three parts.

The first will deal with the implementation of photonic networks. This section includes both a review and summary of the performance characteristics of the enabling technologies as well as a discussion of some of the key aspects of implementing optical transmission systems.

The second part of the thesis concerns itself with optical network architectures. A novel (patented) All Optical Network (AON) architecture is proposed and analysed.

In the final part of the thesis we discuss network control and management. An original means of admission flow control based on dynamic charging is proposed and analysed.

The thesis concludes with a summary of the overall conclusions made from the work and analysis described in each of the three parts of thesis.

The three parts and the individual chapters that constitute them are designed to be individually readable. Each will comprise a section summarising the implications and conclusions made from the analysis and results presented in the various chapters.
# Contents

0.1 List of Abbreviations ................................................................. 23

**Chapter 1 – Introduction** ........................................................................................................... 37

1.1 Convergence of media, consumer electronics, computing and telecommunication networks and services .................................................. 37

1.2 Potential, requirement and possible impact of advanced communication infrastructure ................................................................. 43

1.3 Organisation of the thesis

   1.3.1 Statement of Originality .............................................................. 49

1.4 Bibliography .............................................................................. 51

1.5 References .................................................................................. 55

**Part I Implementation of optical telecommunication systems**

**Chapter 2 - Photonic Network Enabling Technologies** ........................................................................... 57

2.1 Introduction - optics versus electronics ........................................... 57

2.2 Fibre ......................................................................................... 59

2.3 Lasers ....................................................................................... 60

   2.3.1 Wavelength tuneable lasers ..................................................... 61

   2.3.2 Multiwavelength laser arrays .................................................. 64

2.4 Filters ...................................................................................... 64

   2.4.1 Fixed wavelength optical filters ............................................ 64

   2.4.2 Wavelength Tuneable optical filters ..................................... 64

2.5 Amplifiers ................................................................................. 66

   2.5.1 Semiconductor amplifiers ..................................................... 67

   2.5.2 Fibre amplifiers .................................................................. 67

2.6 Bibliography .............................................................................. 69

2.7 References .................................................................................. 71

**Chapter 3 - Optical transmission systems** ................................................................................. 79

3.1 Transmission limitations due to optical fibre nonlinearities in transparent optical networks ................................................................. 79

   3.1.1 Introduction ....................................................................... 79

   3.1.2 Stimulated Raman Scattering ............................................. 81

   3.1.3 Stimulated Brillouin Scattering .......................................... 82

   3.1.4 Optical fibre nonlinearities .................................................. 83

      *Carrier Induced Phase noise* ................................................... 83
<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
</tr>
<tr>
<td>84</td>
</tr>
<tr>
<td>87</td>
</tr>
<tr>
<td>89</td>
</tr>
<tr>
<td>97</td>
</tr>
<tr>
<td>99</td>
</tr>
</tbody>
</table>

**Part II  Towards photonic networks**

**Chapter 4 - Optical networks**

- 4.1 Background and history of optical networks
- 4.2 Application and potential for optics in communication networks
- 4.3 All Optical Network architectures
  - 4.3.1 Modularity and scalability
    - Scalability of network control & management
- 4.4 Summary and conclusions
- 4.5 Bibliography
- 4.6 References

**Chapter 5 - Potential for Wavelength Division Multiplexing in broadband access networks**

- 5.1 General statement of the local access requirement
  - Telephony
  - CATV
  - Video-on-demand
  - Network implications
- 5.1.2 Other related technical changes in the surrounding infra-structure
  - TV display technology
  - Office System Convergence
- 5.1.3 Conclusions on service provision
- 5.2 Roles for WDM in fibre networks
  - Greater Bandwidth
  - Routing
  - Service Growth
  - Mixed Format Services
  - Bi-directional Wideband Service
  - Advanced routing in an All-optical Network
  - General
- 5.3 Optical local access networks architectures
  - 5.3.1 Classes of optical subscriber loop systems
  - 5.3.2 Network topology
6.3.5 Analysis of access network 163
6.3.6 Conclusions 166
6.3.7 References 166
6.4 Summary and conclusions 169
6.5 Bibliography 173
6.6 References 175

Part III  Control and management of broadband networks

Chapter 7 - Resource Control & Management in Broadband Telecommunication Networks

7.1 Introduction 177
7.2 Congestion Control 178
7.2.1 Congestion Control Schemes and Control Theory 180
7.2.2 Timing 182
7.3 Charging 182
   The cost of Service Provision 183
   Incremental Social cost 184
7.3.1 Charging Parameters 184
   Access Network 184
   Core Network 185
7.3.2 Charging Mechanisms 188
   Other Charging mechanisms 189
      Statistical Measurement charging 189
      Envelope charging 190
      Effective Bandwidth charging 190
      'Smart Market' / Dynamic / 'Free Market' charging 191
   Discussion 191
      Usage Sensitivity 191
      Predictability 192
      Complexity & Practicality 192
      Accountability & Auditability 192
      Profitability 193
      Fairness 193
      Knowledge Needed 193
7.4 Summary 194
   Bibliography 197

Chapter 8 - Telecommunication Resource Control through Charging 201
8.1 Introduction 201
B.2 "A Neural Network Implementation Suitable for Correlation Decoding of Some Block Codes".
M. Sabry, D. Grant, J. E. Midwinter and J. T. Taylor

B.2.1 Introduction

B.2.2 Conventional and Previously Proposed Neural Network Decoding Techniques

B.2.3 The BAMNET ANN Prototype

B.2.4 BAMNET ANN Decoder Implementation

B.2.5 Conclusions and Future Work

B.2.6 Acknowledgements

B.2.7 References

B.3 "Telecommunication Network Management & Control Through Charging", Martin Sabry and J. E. Midwinter

B.3.0 Abstract

B.3.1 Introduction

B.3.2 Congestion Control Schemes

B.3.3 Charging Mechanisms

B.3.3.1 Other Charging Schemes

B.3.3.1.1 Effective Bandwidth Charging

B.3.3.1.2 'Smart Market' / Dynamic / 'Free Market' charging

B.3.4 Discussion

B.3.5 References

B.4 "An Analytical Approach to Charging Mechanisms using Control Theory", Martin Sabry, Nina F. Thornhill and John E. Midwinter

B.4.0 Abstract

B.4.1 Introduction

B.4.2 Model

B.4.3 Analysis & Discussion

B.4.4 References

B.5 "Transmission Limitations due to Optical Fibre Nonlinearities in Transparent Optical Networks", Martin Sabry and J. E. Midwinter

B.6 MSc IT projects

B.6.1 "A Review Of Traffic Source Models" by Victor Okoro;
Project Supervisor: Professor John Midwinter & Mr. Martin Sabry

B.6.2 "Free market admission" by A. K. Griffiths;
Supervisors: Professor J. E. Midwinter & Martin Sabry

B.7 Miscellaneous Ph.D. Related Talks & Presentations

B.7.1 "The Future of Telecommunications" by Martin Sabry, Electrical and Electronic Engineering, University College London: Bloomsbury Rotary Club, 2 Jun 1993

B.7.2 "Tomorrow's Telecom Networks and Services" by Martin Sabry (Electrical and Electronic Engineering): UCL Graduate Interdisciplinary Society, 7 Oct. 1997
List of Figures

Fig. 1.1 The IT industry 38
Fig. 1.2 Examples of traffic bitrate versus duration 44
Fig. 1.3 Examples of traffic burstiness versus bitrate 45

Fig. 2.1 Insertion loss of 50 km of single-mode fibre with EDFA gain profile overlaid 67

Fig. 3.1 SF-DWDM spectrum 80
Fig. 3.2 Raman gain profile 82
Fig. 3.3 Mixing efficiency, $\eta$, (dB) versus channel frequency spacing, $\Delta f$, (GHz) for
$f_{i} - \Delta f_{i} + \Delta f_{j}$.

Fig. 3.4 The sum of $\eta f_{W/2} D^{2}$ as a function of $W$ and $f$ for all combinations of $f_{f_{W/2}} = f_{1} + f_{2} - f_{k}$,
where $f_{1}$, $f_{2}$, $f_{k} = f_{1}$, $f_{2}$, $f_{3}$, ..., $f_{f_{W}}$, assuming a 4700GHz (EDFA) transmission window.

Fig. 3.5 Amplifier chain 87

Fig. 3.6 Maximum transmission distance, $N-L_A(N)$, in km versus the number of amplifiers, $N$,
using direct detection, normal fibre, $W=100$, $B_o=50$GHz transmission. 91

Fig. 3.7 Maximum transmission distance, $N-L_A(N)$, in km versus the number of amplifiers, $N$,
using direct detection, dispersion shifted fibre, $W=100$, $B_o=50$GHz transmission. 91

Fig. 3.8 Maximum per channel input power, $P_n$, in dBm versus the number of amplifiers, $N$,
using direct detection, normal fibre, $W=100$, $B_o=50$GHz transmission. 93

Fig. 3.9 Maximum transmission distance, $N-L_A(N)$, in km versus the number of amplifiers, $N$,
using direct detection, normal fibre, $W=100$, $B_o=10$GHz transmission. 95

Fig. 3.10 Maximum transmission distance, $N-L_A(N)$, in km versus the number of amplifiers, $N$,
using direct detection, normal fibre, $W=500$, $B_o=10$GHz transmission. 96

Fig. 3.11 Maximum transmission distance, $N-L_A(N)$, in km versus the number of amplifiers, $N$,
using coherent detection, normal fibre, $W=100$ transmission. 97

Fig. 3.12 Maximum per channel input power, $P_n$, in dBm versus the number of amplifiers, $N$,
using coherent detection, normal fibre, $W=100$ transmission. 97

Fig. 4.1 Broadcast and select optical network 105
Fig. 4.2 Selectively routed optical network 106
Fig. 4.3 Maximum node pair capacity, $C_{max}$ (Bit/s) versus the number of nodes (N) for circuit
switched traffic ($C_{max}.SW$) and individual ATM packets ($C_{max}.ATM$) 109
Fig. 5.1 Generalised local access network
Fig. 5.2 Relationship between switched and distributive services for local access networks
Fig. 5.3 Subscriber loop systems
Fig. 5.4 Network topology
Fig. 5.5 Effective length (km) versus actual length (km)
Fig. 5.6 Hierarchical view of the proposed (optically transparent) network architecture

Fig. 6.2.1 Fibre-Channel-Time space
Fig. 6.2.2 Conceptual view of wavelength space switch
Fig. 6.2.3 Network architecture
Fig. 6.2.4 Success rate and throughput
Fig. 6.3.1 Insertion loss of 50 km of single-mode fibre with EDFA gain profile overlaid
Fig. 6.3.2 Fibre-channel-time space
Fig. 6.3.3 Conceptual view of (3·6·6)x(3·6·6) core node (wavelength space), with arbitrary cross-connect pattern
Fig. 6.3.4 Hierarchical view of proposed (optically transparent) network architecture
Fig. 6.3.5 Functionality of core node
Fig. 6.3.6 Functionality of ring node
Fig. 6.3.7 Optoelectronic interface
Fig. 6.3.8 Core network traffic model
Fig. 6.3.9 Conceptual view of ring-PON interface (where ring consists of two fibres) with arbitrary traffic pattern as seen by PON
Fig. 6.3.10 Mean success rate and throughput against offered load for rings with nonwavelength selective receivers
Fig. 6.3.11 Cell-loss probability against offered load for rings with wavelength selective receivers

Fig. 7.1 Network performance versus offered traffic load
Fig. 7.2 Taxonomy for congestion control algorithms
Fig. 7.3 Classification of existing congestion control schemes
Fig. 7.4 Congestion Control Function Response Time

Fig. 8.1 Demand versus price relationship
Fig. 8.2 Access time bounds due to physical limitations
Fig. 8.3 Future resource allocation versus time
Fig. 8.4 Free Market Access Control Model
Fig. 8.5 Revised Free Market Access Control Model
Fig. 8.6 Detailed Free Market Access Control Model

Fig. 9.1 Feed Forward control loop
Fig. 9.2 Feed Back control loop
Fig. 9.3 Control theory representation of single network resource
Fig. 9.4 Symbolic pole-zero map
Fig. 9.5 Pole-zero map of \( H(z) = \Delta BU/\Delta D \)
Fig. 9.6 Root locus of \( H(z) = \Delta BU/\Delta D \) with respect to \( K \cdot k \cdot \alpha_1 \) term
Fig. 9.7 Root locus of \( H(z) = \Delta BU/\Delta D \) with respect to \( K \cdot k \cdot \alpha_2 \)
Fig. 9.8 Step response of \( H(z) = \Delta BU/\Delta D \)

Fig. B.2.1 Typical transmission system
Fig. B.2.2 Functional diagram of BAMNET ANN
Fig. B.2.3 Levels of correlation generated by first section of WTA layer
Fig. B.2.4 Error Correction as '3' is changed into a '2'
Fig. B.2.5 The Null Response as '4' is changed into a '9'
Fig. B.2.6 Hard-decision correlation decoder
Fig. B.2.7 Transconductance characteristics of present and proposed synapses

Figure B.3.1 Network performance versus offered traffic load

Figure B.4.1 Feed Forward control loop
Figure B.4.2 Feed Back control loop
Figure B.4.3 Control theory representation of single network resource
Figure B.4.4 Symbolic pole-zero map
Figure B.4.5 Pole-zero map of \( H(z) = \Delta BU/\Delta D \)
Figure B.4.6 Root locus of \( H(z) = \Delta BU/\Delta D \) with respect to \( K \cdot k \cdot \alpha_1 \) term
Figure B.4.7 Root locus of \( H(z) = \Delta BU/\Delta D \) with respect to \( \alpha_2 \)
Figure B.4.8 Step response of \( H(z) = \Delta BU/\Delta D \)
List of Tables

Table 2.1 Attributes of photons and electrons 58
Table 2.2 Fibre design types 60
Table 2.3 Fabrication tolerances 61
Table 2.4 Wavelength tuneable lasers 62
Table 2.5 Multiple wavelength laser arrays 64
Table 2.6 Wavelength tuneable filters 66

Table 3.1 System parameters of hypothetical transmission system 94

Table 5.1 Compressed digital video rates 122
Table 5.2 Comparison between directional multiplexing schemes 130
Table 5.3 Comparison of multiple access multiplexing techniques 132
Table 5.4 Implications of multiple access techniques 133
Table 5.5 Maximum per channel input power due to optical fibre nonlinearities 138

Table 7.1 Summary of charging Parameters 187
Table 7.2 Comparison of charging schemes 195

Table 9.1 Definition of control system parameters 223

Table B.4.1 Definition of control system parameters 283
List of Equations

Eqn. (3.1) Worst case attenuation due to SRS 81
Eqn. (3.2) Approximation of Raman gain coefficient 82
Eqn. (3.3) SRS maximum input power condition 82
Eqn. (3.4) SBS maximum input power condition 83
Eqn. (3.5) Phase variation (rad) due to SPM 84
Eqn. (3.6) Phase variation (rad) due to CPM 84
Eqn. (3.7) Total phase variation (rad) due to CIP 84
Eqn. (3.8) CIP maximum input power condition 84
Eqn. (3.9) Power transferred due to FPM 85
Eqn. (3.10) Mixing efficiency 85
Eqn. (3.11) Difference in propagation constant 85
Eqn. (3.12) Power Penalty (PP) due to reduction in extinction ratio 86
Eqn. (3.13) FPM maximum power condition 86
Eqn. (3.14) Optical noise power due to build-up in ASE noise 87
Eqn. (3.15) Spontaneous-spontaneous beat noise 88
Eqn. (3.16) Signal- spontaneous beat noise 88
Eqn. (3.17) Minimum SNR requirement for dominant nsp-sp amplifier noise 88
Eqn. (3.18) Minimum SNR requirement for dominant nsp-sp amplifier noise 88
Eqn. (3.19) Local oscillator-spontaneous beat noise 88
Eqn. (3.20) Coherent detection transmission condition 89
Eqn. (3.21) General transmission criteria regarding build-up of amplifier noise 89
Eqn. (3.22) General transmission criteria for SRS, SBS, CIP & FPM 89
Eqn. (3.23) Effective length of amplifier chain 89
Eqn. (3.24) Maximum amplifier spacing 89

Eqn. (4.1) The time required to transmit the message data, $T_{seg}$ 108
Eqn. (4.2) The time require to control the switch matrix, $T_{ctrl}$ 108
Eqn. (4.3) The maximum capacity of the network between each node pair, $C_{max}$, where $M$ is the message length in bits, $C_{ctrl}$ is the effective capacity of the control channel (Bit/s) and $N$ is the number of nodes. 108

Eqn. (5.1) Gain of probe wave due to optical fibre nonlinearies 137
Eqn. (5.2) The effective length of a section of optical 137
Eqn. (5.3) Worst case blocking probability in proposed AON 141
Eqn. (6.3.1) Highest probability for blocking (in core network) 163
Eqn. (6.3.2) Highest probability for blocking (in core network) 163
Eqn. (6.3.3) Probability that exactly $x$ ATM cells are destined for one PON 164
Eqn. (6.3.4) Probability for making a connection 164
Eqn. (6.3.5) Average fraction of successfully inserted ATM cells that are received 165
Eqn. (6.3.6) Mean number of packets 165
Eqn. (6.3.7) Probability that two or more ATM cells with identical wavelength-time reference are going to the same PON 165
Eqn. (6.3.8) $Q_2, Q_3, \ldots, Q_i$ — where $Q_i$ is the probability that exactly $i$ ATM cells with identical wavelength-time references are destined for same PON 165
Eqn. (6.3.9) Proportion of lost cells 165

Eqn. (7.1) General description of charging cost 188
Eqn. (7.2) Effective Bandwidth charge 190

Eqn. (8.1) Demand versus price relationship 202
Eqn. (8.2) $\Sigma\{Price\}$ charging WP 210
Eqn. (8.3) $\Sigma\{Price\}$ charging 210

Eqn. (9.1.1-3) Time domain characteristic equations of control theory representation of single network resource 222
Eqn. (9.4.4-6) z-domain characteristic equations of control theory representation of single network resource 222
Eqn. (9.7) Transfer function of a FM controlled network resource 222

Eqn. (B.3.1) General expression for charging 271
Eqn. (B.3.2) Effective bandwidth charging 272

Eqn. (B.4.1-3) Time domain characteristic equations of control theory representation of single network resource 282
Eqn. (B.4.4-6) z-domain characteristic equations of control theory representation of single network resource 282
Eqn. (B.4.7) Transfer function of a FM controlled network resource 282
0.1 Abbreviations

2D Two Dimensional
3D Three Dimensional
3R Receive-Regenerate-Retransmit (repeater)
A/D Analogue to Digital (conversion)
A-PIN Avalanche PIN
AAL ATM Adaption Layer
AARC Access And Routing Control
ABC Advanced Broadband Communications
ABR Available Bit Rate
ABT ATM Block Transfer
ABT/DT ATM Block Transfer Delayed Transmission
ABT/IT ATM Block Transfer Immediate Transfer
ACA Amplifier Coupler Absorber (laser structure)
ACE Advanced Communication Experiment
ACG Automatic Call Gapping
ACR Automatic Control Restriction
ACS Advanced Communication Service
ACTS Advanced Communications Technologies and Services
ADCT Adaptive Discrete Cosine Transform Coding (video coding)
ADSL Asymmetrical Digital Subscriber Line
AFNOR Association Française de NORMalisation
AGC Automatic Gain Control
AGWN Additive Gaussian White Noise
AHS Automated Highway System
AI Artificial Intelligence
AIM Advanced Informatics for Medicine (Application Programme of DG XIII/F)
AIP Advanced Infrastructure Planning
AM Advanced Information Processing
AM Amplitude Modulation
AMPS Advanced Mobile telePhone Service (the standard system in the US)
ANN Artificial Neural Network
ANPE Anpassungseinrichtung (German Switching Equipment)
ANSI American National Standards Institute
AOWC All-Optical Wavelength Conversion laser
AP Application Pilot
API Application Programme Interface
APT Application Pilot Transfer (concertation group)
ARC Access and Routing Control
ARG ATM Requirements Group - discuss and find consensus between RACE projects on ATM parameters and contribution to international standardisation bodies
ARTIC Annual Rural Telematics Workshop
ASE Amplified Spontaneous Emission (noise)
ASIC Application Specific Integrated Circuit
AT&T American Telephone and Telegraph Corporation
ATA ATM Traffic Analyser
ATC ATM Transfer Capabilities
ATD Asynchronous Time Division (multiplexing & switching)
ATG ATM Traffic Generator
ATH ATM Transmission Hierarchy
ATM Asynchronous Transfer Mode
Automated Teller Machine
Advanced Testing Mode (see ETSI-ATM)
ATS ATM Transmission System
AU Adaptation Unit
AV Audio Visual
AVT-WG AV Transport - Working Group
B-ISDN Broadband-ISDN
B-NT Broadband-Network Termination
BAMNET Binary Associative Memory NETwork
BAP Broadband Application Part
BB BroadBand
BB-ISDN BroadBand-ISDN
BBA Belgian Broadband Association
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBT</td>
<td>Belgian Broadband Trial</td>
</tr>
<tr>
<td>BCF</td>
<td>Budget Cross Flow Model (analytical tool adapted in RACE I)</td>
</tr>
<tr>
<td>BCPN</td>
<td>Business CPN</td>
</tr>
<tr>
<td>Bellcore</td>
<td>Bell Communications REsearch</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>BERKOM</td>
<td>BERliner KOMunikations system (German optical B-ISDN network trial)</td>
</tr>
<tr>
<td>BICMOS</td>
<td>Bipolar Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>BNR</td>
<td>Bell Northern Research</td>
</tr>
<tr>
<td>BOC</td>
<td>Bell Operating Company</td>
</tr>
<tr>
<td>BPON</td>
<td>Broadband Passive Optical Network</td>
</tr>
<tr>
<td>Brite Euram</td>
<td>Basic Research in Industrial Technologies for Europe — EUropean Research in Advanced Materials</td>
</tr>
<tr>
<td>BSS</td>
<td>Base Station System</td>
</tr>
<tr>
<td>BSU</td>
<td>Broadband Switching Unit</td>
</tr>
<tr>
<td>BT</td>
<td>Burst Tolerance</td>
</tr>
<tr>
<td>British Telecom</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>Conditional Access</td>
</tr>
<tr>
<td>C-BAT</td>
<td>Cost-Benefit Analysis Tool kit (analytical tool adapted in RACE I)</td>
</tr>
<tr>
<td>CAC</td>
<td>Customer Access Connection</td>
</tr>
<tr>
<td>CAD</td>
<td>Connection Admission Control</td>
</tr>
<tr>
<td>CADDIA</td>
<td>Comité d'Action Commerciale (CEPT)</td>
</tr>
<tr>
<td>CAM</td>
<td>Computer-Aided Design</td>
</tr>
<tr>
<td>CADDIA</td>
<td>Cell Assembler/Dissembler</td>
</tr>
<tr>
<td>CAM</td>
<td>Co-operation in Automation of Data and Documentation for Imports/exports and the management of financial control of the Agricultural market (CEC R&amp;D Programme)</td>
</tr>
<tr>
<td>CAT Scan</td>
<td>Computer-Aided Manufacturing</td>
</tr>
<tr>
<td>CATV</td>
<td>Computed Axial Tomodensitometry Scanner</td>
</tr>
<tr>
<td>CATV</td>
<td>Cable TeleVision</td>
</tr>
<tr>
<td>CBDS</td>
<td>Cable Television</td>
</tr>
<tr>
<td>CBTS</td>
<td>Community Access TV</td>
</tr>
<tr>
<td>CBTS</td>
<td>Community Antenna TeleVision</td>
</tr>
<tr>
<td>CBTS</td>
<td>Connectionless Broadband Data Server</td>
</tr>
<tr>
<td>CBO</td>
<td>Continuous Bit Stream Oriented (service)</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant Bit Rate</td>
</tr>
<tr>
<td>CBRA</td>
<td>Constant Bit-Rate Adapter</td>
</tr>
<tr>
<td>CC</td>
<td>Cross-Connect</td>
</tr>
<tr>
<td>CDC</td>
<td>Charge Coupled Device (camera)</td>
</tr>
<tr>
<td>CCCP</td>
<td>Conference Communication Channel Protocol</td>
</tr>
<tr>
<td>CCF</td>
<td>Call Control Function</td>
</tr>
<tr>
<td>CCH</td>
<td>Comité de Coordination pour l'Harmonisation de CEPT</td>
</tr>
<tr>
<td>CCIR</td>
<td>Comité Consultatif International des Radiocommunications (International Radiodiffusion Consultative Committee) of ITU</td>
</tr>
<tr>
<td>CCITT</td>
<td>Comité Consultatif International Télégraphique et Téléphonique (International Telephone and Telegraph Consultative Committee of ITU)</td>
</tr>
<tr>
<td>CCS</td>
<td>Common Channel Signalling</td>
</tr>
<tr>
<td>CCTS</td>
<td>Comité de Coordination pour les Télécoumunications par Satellite</td>
</tr>
<tr>
<td>CD</td>
<td>Compact Disc</td>
</tr>
<tr>
<td>CD ROM</td>
<td>CD Read Only Memory</td>
</tr>
<tr>
<td>CD-I</td>
<td>CD-Interactive</td>
</tr>
<tr>
<td>CDM</td>
<td>Code Division Multiplexing</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CDT</td>
<td>Cell Delay Tolerance</td>
</tr>
<tr>
<td>CDTV</td>
<td>Conventional Definition TeleVision</td>
</tr>
<tr>
<td>CDV</td>
<td>Cell Delay Variation</td>
</tr>
<tr>
<td>CDV</td>
<td>Coherent Detection</td>
</tr>
<tr>
<td>CEC</td>
<td>Commission of the European Communities</td>
</tr>
<tr>
<td>CECC</td>
<td>Cenelec Electronic Components Committee</td>
</tr>
<tr>
<td>CELEX</td>
<td>EC Legislation database in Eurobases databank</td>
</tr>
<tr>
<td>CELP</td>
<td>Code-Excited Linear Prediction</td>
</tr>
<tr>
<td>CEN</td>
<td>Comité Européen de Normalisation (European Standardisation Organisation)</td>
</tr>
<tr>
<td>CENELEC</td>
<td>Comité Européen de Normalisation Electrotechique (European Electrotechnical Standardisation Organisation)</td>
</tr>
<tr>
<td>CEPT</td>
<td>Conférence Européenne des Postes et Télécommunications (European Conference of Post and Telecommunications Administrations)</td>
</tr>
<tr>
<td>CEQ</td>
<td>Customer EQuipment</td>
</tr>
<tr>
<td>CERN</td>
<td>Centre Européen pour la Recherche Nucléaire (European Laboratory for Particle Physics)</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>CERT</td>
<td>Committee on Energy, Science and Technology</td>
</tr>
<tr>
<td>CF</td>
<td>Conventional Fibre</td>
</tr>
<tr>
<td>CFM</td>
<td>Common Functional Model</td>
</tr>
<tr>
<td>CFS</td>
<td>Common Functional Specifications</td>
</tr>
<tr>
<td>CFS</td>
<td>Common Functional Specification</td>
</tr>
<tr>
<td>CIELab</td>
<td>acronym for a standard which defines the colour space of a picture in a device independent manner</td>
</tr>
<tr>
<td>CIM</td>
<td>Computer Integrated Manufacturing</td>
</tr>
<tr>
<td>CIP</td>
<td>Carrier Induced Phase (noise)</td>
</tr>
<tr>
<td>CL</td>
<td>ConnectionLess</td>
</tr>
<tr>
<td>CLI</td>
<td>Calling Line Identification</td>
</tr>
<tr>
<td>CLP</td>
<td>Cell Loss Priority (bit)</td>
</tr>
<tr>
<td>CLR</td>
<td>Cell Loss Ratio</td>
</tr>
<tr>
<td>CLS</td>
<td>CL Server</td>
</tr>
<tr>
<td>CM</td>
<td>Communications Management</td>
</tr>
<tr>
<td>CMC</td>
<td>Coherent Multi-Channel (optical transmission)</td>
</tr>
<tr>
<td>CMP</td>
<td>Common Management Information Protocol</td>
</tr>
<tr>
<td>CMMC</td>
<td>Conference Management and Multiplexing Centre</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal Oxide Semiconductor</td>
</tr>
<tr>
<td>CMTT</td>
<td>Joint CCIR/CCITT Study Group on Transmission of Sound Broadcasting and Television Systems Over Long Distances WP CMTT-AB: Analogue transmission of TV signals including conversion of standards, MAC Systems, and HDTV WP CMTT-AN: Digital or hybrid analogue-digital transmission of TV signals WP CMTT-C: Transmission of sound program signals</td>
</tr>
<tr>
<td>CO</td>
<td>Central Office</td>
</tr>
<tr>
<td>CODEC</td>
<td>Codér/DECoder</td>
</tr>
<tr>
<td>COFDM</td>
<td>Coded Orthogonal Frequency Division Multiplex</td>
</tr>
<tr>
<td>COM(83) 576</td>
<td>Lines of Action of the Community Telecommunication Policy</td>
</tr>
<tr>
<td>COM(84) 277</td>
<td>Progress Report on the Implementation of the Telecommunication Policy</td>
</tr>
<tr>
<td>COM(85) 113</td>
<td>Proposal for the RACE Definition Phase</td>
</tr>
<tr>
<td>COM(85) 145</td>
<td>Communication giving the background and Rationale for RACE</td>
</tr>
<tr>
<td>COM(85) 230</td>
<td>Arrangements for mutual recognition of type approvals</td>
</tr>
<tr>
<td>COM(85) 276</td>
<td>Status of the Community Telecommunications Policy</td>
</tr>
<tr>
<td>COM(85) 325</td>
<td>European Communications Policy</td>
</tr>
<tr>
<td>COMETT</td>
<td>Community Action Programme in Education and Training for Technology</td>
</tr>
<tr>
<td>COSINE</td>
<td>Cooperation for Open Systems Interconnection Networking in Europe (EUREKA-8)</td>
</tr>
<tr>
<td>COST</td>
<td>Co-operation for R&amp;D in Science and Technology</td>
</tr>
<tr>
<td>CP</td>
<td>Common Practices</td>
</tr>
<tr>
<td>CPE</td>
<td>Current Price (see FMAFC)</td>
</tr>
<tr>
<td>CPE</td>
<td>Customer Premises Equipment</td>
</tr>
<tr>
<td>CPFSK</td>
<td>Customer Programmable Environment</td>
</tr>
<tr>
<td>CPM</td>
<td>Continuous Phase Frequency Shift Keying</td>
</tr>
<tr>
<td>CPM</td>
<td>Cross Phase Modulation</td>
</tr>
<tr>
<td>CPN</td>
<td>Customer Premises Network</td>
</tr>
<tr>
<td>CPR</td>
<td>Common Practice Recommendation</td>
</tr>
<tr>
<td>CPU</td>
<td>Computing and Processing Unit</td>
</tr>
<tr>
<td>CRBC</td>
<td>Call Rate Based Control</td>
</tr>
<tr>
<td>CRF(VC)</td>
<td>Connection Related Functions (Virtual Channel)</td>
</tr>
<tr>
<td>CRF(VP)</td>
<td>Connection Related Functions (Virtual Path)</td>
</tr>
<tr>
<td>CRIMP</td>
<td>CRoss IMpact Model (analytical tool adapted in RACE I)</td>
</tr>
<tr>
<td>CRL</td>
<td>Common Representation Language</td>
</tr>
<tr>
<td>CRT</td>
<td>Cathode Ray Tube</td>
</tr>
<tr>
<td>CSCE</td>
<td>Connexion au Réseau Terrestre (French Switching Equipment)</td>
</tr>
<tr>
<td>CSCE</td>
<td>Computer Supported Cooperative Work</td>
</tr>
<tr>
<td>CSDDN</td>
<td>Circuit Switched Data Network</td>
</tr>
<tr>
<td>CSF</td>
<td>Customer Service Function</td>
</tr>
<tr>
<td>CSMA/CD</td>
<td>Carrier Sense Multiple Access with Collision Detection</td>
</tr>
<tr>
<td>CSPDN</td>
<td>Circuit Switched Public Data Network</td>
</tr>
<tr>
<td>CT</td>
<td>Cordless Telephone</td>
</tr>
<tr>
<td>CT2</td>
<td>2nd Generation Cordless Telephone</td>
</tr>
<tr>
<td>CTD</td>
<td>Cell Transfer Delay</td>
</tr>
<tr>
<td>CTI</td>
<td>Computer Telephony Interface</td>
</tr>
<tr>
<td>CTN</td>
<td>Core Transport Network</td>
</tr>
<tr>
<td>CTS</td>
<td>Conformance Testing Service (EC's R&amp;D Programme, DG XIII)</td>
</tr>
<tr>
<td>CW</td>
<td>Continuous Wave</td>
</tr>
<tr>
<td>D/A</td>
<td>Digital to Analogue (converter)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>FDDI</td>
<td>Fibre Distributed Data Interface</td>
</tr>
<tr>
<td>FDM</td>
<td>Frequency Division Multiplexing</td>
</tr>
<tr>
<td>FDMA</td>
<td>Frequency Division Multiple Access</td>
</tr>
<tr>
<td>FDOC</td>
<td>Focus Destination Overload Control</td>
</tr>
<tr>
<td>FE</td>
<td>Functional Entity</td>
</tr>
<tr>
<td>FEOGA</td>
<td>Fond Européen d'Orientation et de Garantie Agricole (European Agricultural Guidance and Guarantee Fund: CEC/DG VI)</td>
</tr>
<tr>
<td>FET</td>
<td>Field Effect Transistor</td>
</tr>
<tr>
<td>FFDI</td>
<td>Fast Fibre Data Interface</td>
</tr>
<tr>
<td>FG</td>
<td>Focus Group</td>
</tr>
<tr>
<td>FIB</td>
<td>Focused Ion Beam</td>
</tr>
<tr>
<td>FIFO</td>
<td>First In First Out</td>
</tr>
<tr>
<td>FITs</td>
<td>Failures in 109 hours</td>
</tr>
<tr>
<td>FF</td>
<td>Feed Forward (compensator, control system)</td>
</tr>
<tr>
<td>FM</td>
<td>Frequency Modulation</td>
</tr>
<tr>
<td>FMAC</td>
<td>Free Market Admission Control</td>
</tr>
<tr>
<td>FMADC</td>
<td>Free Market Admission and Flow Control</td>
</tr>
<tr>
<td>FMFC</td>
<td>Free Market Flow Control</td>
</tr>
<tr>
<td>FP</td>
<td>Fabry-Perot</td>
</tr>
<tr>
<td>Forward Path (transfer function, control theory)</td>
<td></td>
</tr>
<tr>
<td>FPD</td>
<td>Flat Panel Display</td>
</tr>
<tr>
<td>FPLMTS</td>
<td>Future Public Land Mobile Telecommunication Systems</td>
</tr>
<tr>
<td>FPM</td>
<td>Four Photon Mixing</td>
</tr>
<tr>
<td>FRM</td>
<td>Functional Reference Model</td>
</tr>
<tr>
<td>Fast Resource Management</td>
<td></td>
</tr>
<tr>
<td>FSK</td>
<td>Frequency - Shift Keying</td>
</tr>
<tr>
<td>FSP</td>
<td>Flexible Service Profile</td>
</tr>
<tr>
<td>FTAM</td>
<td>File Transfer, Access and Management</td>
</tr>
<tr>
<td>FTP</td>
<td>File Transfer Protocol</td>
</tr>
<tr>
<td>FTTC</td>
<td>Fibre To The Cabinet</td>
</tr>
<tr>
<td>FTTH</td>
<td>Fibre To The Home</td>
</tr>
<tr>
<td>FTTO</td>
<td>Fibre To The Office</td>
</tr>
<tr>
<td>FWM</td>
<td>Four Wave Mixing</td>
</tr>
<tr>
<td>GaAs</td>
<td>Gallium Arsenide</td>
</tr>
<tr>
<td>GAP</td>
<td>Groupe Analyses et prévisions</td>
</tr>
<tr>
<td>GCSR</td>
<td>vertical Grating assisted directional Coupler laser with rear Sampled grating Reflector</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>GEN</td>
<td>General European Network (CEPT)</td>
</tr>
<tr>
<td>GFC</td>
<td>Generic Flow Control</td>
</tr>
<tr>
<td>GIBN</td>
<td>Global Interconnection of Broadband Networks</td>
</tr>
<tr>
<td>GMS</td>
<td>Generic Maintenance System</td>
</tr>
<tr>
<td>GNP</td>
<td>Gross National Product</td>
</tr>
<tr>
<td>GoS</td>
<td>Grade of Service</td>
</tr>
<tr>
<td>GOS</td>
<td>Grade Of Service</td>
</tr>
<tr>
<td>GPM</td>
<td>Gross Potential Market</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GRINSCH-MQW</td>
<td>Graded Refractive Index Separate Confinement Heterostructure — Multi Quantum Well</td>
</tr>
<tr>
<td>GSL</td>
<td>Global Service Logic</td>
</tr>
<tr>
<td>GSLB</td>
<td>Groupe special large bande (CEPT group on broadband telecommunications)</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile communications</td>
</tr>
<tr>
<td>Groupe Special Mobile (CEPT Working Group for a Pan-european Mobile Telecommunication System)</td>
<td></td>
</tr>
<tr>
<td>GSMBE</td>
<td>Gas Source Molecular Beam Epitaxy</td>
</tr>
<tr>
<td>GSN</td>
<td>Group Switching Nodes</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HCI</td>
<td>Human Computer Interaction</td>
</tr>
<tr>
<td>HCM</td>
<td>Human Capital and Mobility Programme (CEC DG XII)</td>
</tr>
<tr>
<td>HD</td>
<td>High Definition</td>
</tr>
<tr>
<td>HD-MAC</td>
<td>High Definition MAC</td>
</tr>
<tr>
<td>HDDV</td>
<td>High Definition Digital Video</td>
</tr>
<tr>
<td>HDI</td>
<td>High Definition Interlaced (Scanned Image)</td>
</tr>
<tr>
<td>HDP</td>
<td>High Definition Progressively (Scanned Image)</td>
</tr>
<tr>
<td>HDTV</td>
<td>High Definition TeleVision</td>
</tr>
<tr>
<td>HDWDM</td>
<td>High Density Wavelength Division Multiplex</td>
</tr>
<tr>
<td>HEMT</td>
<td>High Electron Mobility Transistor</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>HEOS</td>
<td>Highly Elliptical Orbit Satellite</td>
</tr>
<tr>
<td>HF</td>
<td>High Frequency</td>
</tr>
<tr>
<td>HIPERLAN</td>
<td>High Performance Local Area Network</td>
</tr>
<tr>
<td>HIS</td>
<td>Hospital Information System</td>
</tr>
<tr>
<td>HLL</td>
<td>High Level Language</td>
</tr>
<tr>
<td>HOOD</td>
<td>Hierarchical Object Oriented Design</td>
</tr>
<tr>
<td>HPC</td>
<td>High Performance Computing</td>
</tr>
<tr>
<td>HW</td>
<td>HardWare</td>
</tr>
<tr>
<td>I/F</td>
<td>InterFace</td>
</tr>
<tr>
<td>I/O</td>
<td>Input/Output</td>
</tr>
<tr>
<td>IBAG</td>
<td>INFOSEC Business Advisory Group</td>
</tr>
<tr>
<td>IBC</td>
<td>Integrated Broadband Communications</td>
</tr>
<tr>
<td>IBCN</td>
<td>Integrated Broadband Communications Network</td>
</tr>
<tr>
<td>IBN</td>
<td>Integrated Broadband Network</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>ICB</td>
<td>Inter-project Co-ordination Board</td>
</tr>
<tr>
<td>ICOA</td>
<td>Image Communication Open Architecture</td>
</tr>
<tr>
<td>IDR</td>
<td>Intermediate Data Rate</td>
</tr>
<tr>
<td>IEC</td>
<td>International Electrotechnical Commission</td>
</tr>
<tr>
<td>IEE</td>
<td>Institution of Electrical Engineers</td>
</tr>
<tr>
<td>IEN</td>
<td>Individual Electronic Newspapers</td>
</tr>
<tr>
<td>IFEN</td>
<td>Inter company File Exchange Network</td>
</tr>
<tr>
<td>IHS</td>
<td>In-house System</td>
</tr>
<tr>
<td>IM</td>
<td>Intensity Modulation</td>
</tr>
<tr>
<td>IN</td>
<td>Intelligent Network</td>
</tr>
<tr>
<td>INAP</td>
<td>IN Application Protocol</td>
</tr>
<tr>
<td>INFOSEC</td>
<td>INFORMATION SECURITY (CEC R&amp;D Programme; DG XIII)</td>
</tr>
<tr>
<td>INMARSAT</td>
<td>International MARitime SATellite organisation</td>
</tr>
<tr>
<td>InP</td>
<td>Indium Phosphate</td>
</tr>
<tr>
<td>Intelsat</td>
<td>International Telecommunications Satellite Organisation</td>
</tr>
<tr>
<td>IO</td>
<td>Input/Output</td>
</tr>
<tr>
<td>IOC</td>
<td>Integrated Optical Component</td>
</tr>
<tr>
<td>ION</td>
<td>Integrated Optical Network</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IP</td>
<td>Intelligent Peripheral</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>IRE</td>
<td>Infra-Red Emitting Diode</td>
</tr>
<tr>
<td>IRIDIUM</td>
<td>Satellite Mobile Communication System for sparsely populated areas</td>
</tr>
<tr>
<td>IRT</td>
<td>Integrated Rural Telematics</td>
</tr>
<tr>
<td>IRTA</td>
<td>Integrated Regional Telematics Architecture</td>
</tr>
<tr>
<td>IS</td>
<td>Information Security</td>
</tr>
<tr>
<td>IS&amp;N</td>
<td>Intelligence in Services and Networks</td>
</tr>
<tr>
<td>ISDN</td>
<td>Integrated Service Engineering</td>
</tr>
<tr>
<td>ISM</td>
<td>ISDN Standards Management (ETSI Group)</td>
</tr>
<tr>
<td>ISP</td>
<td>Internet SP</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standards Organisation</td>
</tr>
<tr>
<td>ISODE</td>
<td>OSI development environment</td>
</tr>
<tr>
<td>ISPN</td>
<td>Integrated Services Private Network</td>
</tr>
<tr>
<td>ISV</td>
<td>Independent Software Vendor</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>IT&amp;T</td>
<td>Information Technologies &amp; Telecommunications</td>
</tr>
<tr>
<td>ITS</td>
<td>Intelligent Transport System</td>
</tr>
<tr>
<td>ITSEC</td>
<td>Information Technology Security Evaluation Criteria</td>
</tr>
<tr>
<td>ITT</td>
<td>Information Technology and Telecommunications</td>
</tr>
<tr>
<td>ITT&amp;B</td>
<td>Information Technology, Telecommunications &amp; Broadcasting</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>IWP</td>
<td>Interim Working Party of CMTT</td>
</tr>
<tr>
<td>IUW</td>
<td>Inter-Working Unit</td>
</tr>
<tr>
<td>JAT</td>
<td>Java Audio Tool</td>
</tr>
<tr>
<td>JPEG</td>
<td>Joint Photographic Expert Group (ISO/JTC1/SC29)</td>
</tr>
<tr>
<td>JTC</td>
<td>Joint Technical Committee</td>
</tr>
<tr>
<td>JVTOS</td>
<td>Joint Viewing and Tele-operation Service</td>
</tr>
<tr>
<td>Kbps</td>
<td>Kilo bits per second</td>
</tr>
<tr>
<td>KBS</td>
<td>Knowledge-Based System</td>
</tr>
</tbody>
</table>
LAN  Local Area Network
LCD  Liquid Crystal Display
LE  Local Exchange
LEC  Local Exchange Carrier
LED  Liquid Encapsulated Czochralski
LEOS  Low Earth Orbit Satellite
LEX  Local Exchange
LFR  Less-Favoured Region
LiNbO3  Lithium Niobate
LOTOS  Language of Temporal Ordering Specification (method)
LPE  Liquid Phase Epitaxy
LRC  Local Routing Centre
LSI  Large Scale Integration
LT  Line Termination
LWI  Local Work Instruction
MA  Medium Adapter
MAC  Media Access Control
        Multiplexed Analogue Components: a family of standards for the transmission of TV/HDTV
        signals in which the analogue luminance and chrominance components are compressed in
        time and transmitted sequentially. The data channel can be used for multiple sound channels,
        picture enhancement (HD-MAC provides for high definition picture enhancement while
        retaining compatibility with earlier standards)
MACintosh personal computer
MAN  Metropolitan Area Network
MAP  Manufacturing Automation Protocol
Mbone  Multiplexed Analogue Components: a family of standards for the transmission of TV/HDTV
        signals in which the analogue luminance and chrominance components are compressed in
        time and transmitted sequentially. The data channel can be used for multiple sound channels,
        picture enhancement (HD-MAC provides for high definition picture enhancement while
        retaining compatibility with earlier standards)
MACintosh personal computer
MAN  Metropolitan Area Network
MAP  Manufacturing Automation Protocol
Mbone  Multiplexed Analogue Components: a family of standards for the transmission of TV/HDTV
        signals in which the analogue luminance and chrominance components are compressed in
        time and transmitted sequentially. The data channel can be used for multiple sound channels,
        picture enhancement (HD-MAC provides for high definition picture enhancement while
        retaining compatibility with earlier standards)
MACintosh personal computer
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MWTN</td>
<td>Multi Wavelength Transport Network</td>
</tr>
<tr>
<td>N-ISDN</td>
<td>Narrowband ISDN (conventional ISDN)</td>
</tr>
<tr>
<td>NA</td>
<td>Network Adapter</td>
</tr>
<tr>
<td>NB</td>
<td>Narrowband (transmission at less than 2 Mbit/s)</td>
</tr>
<tr>
<td>NB-ISDN</td>
<td>Narrowband ISDN</td>
</tr>
<tr>
<td>NCSA</td>
<td>National Center for Supercomputing Applications</td>
</tr>
<tr>
<td>NDT</td>
<td>Non Destructive Testing</td>
</tr>
<tr>
<td>NE</td>
<td>NT Equipment</td>
</tr>
<tr>
<td>NETZC</td>
<td>existing mobile network in Germany</td>
</tr>
<tr>
<td>NFS</td>
<td>Network File System</td>
</tr>
<tr>
<td>NGoS</td>
<td>Network GoS</td>
</tr>
<tr>
<td>NGOS</td>
<td>Network GOS</td>
</tr>
<tr>
<td>NCSA</td>
<td>A standard for digital encoding of stereo sounds used in TV</td>
</tr>
<tr>
<td>NIP</td>
<td>Network Intelligent Platform</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NMR</td>
<td>Scanner Nuclear Magnetic Resonance Scanner</td>
</tr>
<tr>
<td>NMRRM</td>
<td>Network Management Reference Model</td>
</tr>
<tr>
<td>NMS</td>
<td>Network Management System</td>
</tr>
<tr>
<td>NMT</td>
<td>Nordic Mobile Telephone</td>
</tr>
<tr>
<td>NNI</td>
<td>Network Node Interface</td>
</tr>
<tr>
<td>NNGoS</td>
<td>Node-to-Node GoS</td>
</tr>
<tr>
<td>NNGOS</td>
<td>Node-to-Node GOS</td>
</tr>
<tr>
<td>NO</td>
<td>Network Operator</td>
</tr>
<tr>
<td>NORC</td>
<td>Network Operators RACE Committee</td>
</tr>
<tr>
<td>NP</td>
<td>Network Performance</td>
</tr>
<tr>
<td>NPC</td>
<td>Network Parameter Control</td>
</tr>
<tr>
<td>NRT-VBR</td>
<td>Non RT-VBR</td>
</tr>
<tr>
<td>NRZ</td>
<td>Non Return to Zero (modulation)</td>
</tr>
<tr>
<td>NSA</td>
<td>National Security Agency</td>
</tr>
<tr>
<td>NSC</td>
<td>Network Service Centre</td>
</tr>
<tr>
<td>NT</td>
<td>Network Specialised Centre</td>
</tr>
<tr>
<td>NTE</td>
<td>NT Equipment</td>
</tr>
<tr>
<td>NTSC</td>
<td>National Television Standard Committee</td>
</tr>
<tr>
<td>NTT</td>
<td>Nippon Telegraph &amp; Telephone</td>
</tr>
<tr>
<td>NVoD</td>
<td>Near VoD</td>
</tr>
<tr>
<td>NVOD</td>
<td>Near VOD</td>
</tr>
<tr>
<td>O&amp;M</td>
<td>Operations &amp; Maintenance</td>
</tr>
<tr>
<td>O/E</td>
<td>Opto-Electronic (interface or conversion)</td>
</tr>
<tr>
<td>O&amp;M</td>
<td>Operations and Maintenance</td>
</tr>
<tr>
<td>OAM</td>
<td>Operation And Maintenance</td>
</tr>
<tr>
<td>OB</td>
<td>Outside Broadcast</td>
</tr>
<tr>
<td>OCR</td>
<td>Optical Character Reader</td>
</tr>
<tr>
<td>ODA</td>
<td>Office Document Architecture</td>
</tr>
<tr>
<td>ODMP</td>
<td>Overload Detection and Monitoring Process</td>
</tr>
<tr>
<td>ODP</td>
<td>Open Distributed Processing</td>
</tr>
<tr>
<td>ODSS</td>
<td>Open Data Stream Structure</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Cooperation and Development</td>
</tr>
<tr>
<td>OEIC</td>
<td>Opto-Electronic IC</td>
</tr>
<tr>
<td>OFM</td>
<td>Optical Frequency Multiplexing</td>
</tr>
<tr>
<td>OL</td>
<td>Open Loop (transfer function, control theory)</td>
</tr>
<tr>
<td>OLE</td>
<td>On-line Environment</td>
</tr>
<tr>
<td>OLI</td>
<td>Optical Line Inlet</td>
</tr>
<tr>
<td>OLO</td>
<td>On-Line Interface</td>
</tr>
<tr>
<td>OMC</td>
<td>Operations and Maintenance Centre</td>
</tr>
<tr>
<td>OOA</td>
<td>Open Network Architecture</td>
</tr>
<tr>
<td>ONP</td>
<td>Open Network Provision</td>
</tr>
<tr>
<td>OOC</td>
<td>Originating-Office Control</td>
</tr>
<tr>
<td>OP</td>
<td>Operating Point</td>
</tr>
<tr>
<td>ORA</td>
<td>Opportunities for IT&amp;T in Rural Areas (planning exercise from the Commission DG XIII/F)</td>
</tr>
<tr>
<td>ORA</td>
<td>Opportunities for application of information and communication technologies in Rural Areas (CEC R&amp;D Programme, DG XIII)</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>OSA</td>
<td>Open Service Architecture</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>OSF</td>
<td>Open Software Foundation - international group of computers industrials and users</td>
</tr>
<tr>
<td>OSI</td>
<td>Open System Interconnection (ISO interface specifications)</td>
</tr>
<tr>
<td>OSPF</td>
<td>Open Shortest Path First (protocol)</td>
</tr>
<tr>
<td>OSS</td>
<td>Operations Support Systems</td>
</tr>
<tr>
<td>OTDM</td>
<td>Optical Time Division Multiplex</td>
</tr>
<tr>
<td>OTR 100</td>
<td>Initial Workplan for the RACE Main Programme</td>
</tr>
<tr>
<td>OTR 200</td>
<td>Revised Workplan for the RACE Main Programme</td>
</tr>
<tr>
<td>OTR 300</td>
<td>Draft Workplan for RACE Extension</td>
</tr>
<tr>
<td>PABX</td>
<td>Private Automatic Branch eXchange</td>
</tr>
<tr>
<td>PACE</td>
<td>Perspectives for Advanced Communications in Europe — RACE</td>
</tr>
<tr>
<td>PACS</td>
<td>Picture Archiving and Communication Systems</td>
</tr>
<tr>
<td>PAL</td>
<td>Phase Alternation Line (colour TV standard)</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>PCN</td>
<td>Personal Communications Network</td>
</tr>
<tr>
<td>PCR</td>
<td>Peak Cell Rate</td>
</tr>
<tr>
<td>PCS</td>
<td>Personal Communication Space</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Digital Assistant</td>
</tr>
<tr>
<td>PDH</td>
<td>Plesiochronous Digital Hierarchy</td>
</tr>
<tr>
<td>PDU</td>
<td>Protocol Data Unit</td>
</tr>
<tr>
<td>PEAN</td>
<td>Pan-European ATM Network</td>
</tr>
<tr>
<td>PED</td>
<td>Portable Electronic Device</td>
</tr>
<tr>
<td>PET</td>
<td>Planning Exercise in Telecommunications technologies (forerunner of RACE)</td>
</tr>
<tr>
<td>PHS</td>
<td>Personal Handiphone System</td>
</tr>
<tr>
<td>PI</td>
<td>Programming Infrastructure</td>
</tr>
<tr>
<td>PIC</td>
<td>Photonic IC</td>
</tr>
<tr>
<td>PIN</td>
<td>Positive doped /Insulating / Negative doped (diode)</td>
</tr>
<tr>
<td>PIPO</td>
<td>Parallel In Parallel Out (shift register)</td>
</tr>
<tr>
<td>PISO</td>
<td>Parallel In Serial Out (shift register)</td>
</tr>
<tr>
<td>PL</td>
<td>Project Line</td>
</tr>
<tr>
<td>PMD</td>
<td>Physical Medium Dependent (Layer 1)</td>
</tr>
<tr>
<td>PMMA</td>
<td>Poly-Methyl Meth-Acrylate</td>
</tr>
<tr>
<td>PMN</td>
<td>Private Mobile Network</td>
</tr>
<tr>
<td>PMR</td>
<td>Private Mobile Radio</td>
</tr>
<tr>
<td>PN</td>
<td>Permanent Nucleus created by CEPT for the RACE Definition Phase</td>
</tr>
<tr>
<td>PNO</td>
<td>Public Network Operator</td>
</tr>
<tr>
<td>PON</td>
<td>Passive Optical Network</td>
</tr>
<tr>
<td>POTS</td>
<td>Plain Old Telephone Service</td>
</tr>
<tr>
<td>PP</td>
<td>Power Penalty</td>
</tr>
<tr>
<td>PPV</td>
<td>Pay-Per-View</td>
</tr>
<tr>
<td>PRBS</td>
<td>Pseudo Random Binary Sequence</td>
</tr>
<tr>
<td>PRM</td>
<td>Protocol Reference Model</td>
</tr>
<tr>
<td>PSCS</td>
<td>Personal Service Communication Space</td>
</tr>
<tr>
<td>PSDN</td>
<td>Packet Switched Data Network</td>
</tr>
<tr>
<td>PSN</td>
<td>Public Switched (telephone) Network</td>
</tr>
<tr>
<td>PSTN</td>
<td>Public Switched Telephone Network</td>
</tr>
<tr>
<td>PTI</td>
<td>Payload Type Indicator</td>
</tr>
<tr>
<td>PTO</td>
<td>Public Telecommunications Operator</td>
</tr>
<tr>
<td>PTTs</td>
<td>Post Telegraph and Telephone administrations (Network Operators)</td>
</tr>
<tr>
<td>PV</td>
<td>Process Variable</td>
</tr>
<tr>
<td>PVC</td>
<td>Permanent Virtual Circuits</td>
</tr>
<tr>
<td>PVN</td>
<td>Private Virtual Networks</td>
</tr>
<tr>
<td>PVP</td>
<td>Permanent Virtual Paths</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>QCSE</td>
<td>Quantum Confined Stark Effect</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>QOS</td>
<td>Quality Of Service</td>
</tr>
<tr>
<td>QW</td>
<td>Quantum Well, e.g. advanced QW-based semiconductor lasers</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>Research and Development</td>
</tr>
<tr>
<td>RACE</td>
<td>Research in Advanced Communications technologies in Europe (CEC R&amp;D Programme, DG XIII)</td>
</tr>
<tr>
<td>RAM</td>
<td>Random Access Memory</td>
</tr>
<tr>
<td>RARE</td>
<td>Réseaux Associés pour la Recherche Européenne. (The Association of the European Research Network Organisation)</td>
</tr>
</tbody>
</table>
RAS  Reliability, Availability and Scalability
RATT  R1022 ATM Technology Testbed
RBOC  Regional Bell Operating Companies
RC  Reference Configuration
RCM  RACE Concertation Meeting
RCO  RACE CO (RACE R&D project, DG XIII)
RCPC  Rate Compatible Punctured Convolutional
RCs  Reference Configurations
RD&E  Research, Development and Engineering
RD&T  Research, Development and Technology
RDA  Remote Database Access
RDP  RACE Definition Phase
RDT  Research, Development & Technology
RES  Radio Equipment Systems (ETSI)
RF  Radio Frequency
RGB  Red Green Blue (display)
RIC  RACE Industrial Consortium (prime contractor of projects R1045, R1044 and R2083)
RINC  RACE Integrity Circle
RIP  Routing Information Protocol
RISC  Reduced Instruction Set Computing
RISE  RD&T in Integrated Service Engineering (planning exercise of DG XIII/F)
RM  Resource Management (ATM cells)
RMC  RACE Management Committee
ROM  Read Only Memory
R  Reference Station
RSE/PO  see SE/PO
RSF  Resource Specification Framework
RTC/RTS  Regional Tele Centre/Regional Tele Server
RTD  Research and Technology Development
RTSC  Regional Telematic Service Centre
RT-VBR  Real-Time VBR
RZ  Return to Zero (modulation)
S-PCN  Satellite PCN
SA  Structured Analysis
SADT  Structured Analysis and Design Technique
SAP  Service Access Point
SBR  Statistical Bit Rate
SBS  Stimulated Brillouin Scattering
SCENT  heuristic method for generating and evaluating SCENario Trees (analytical tool adapted in RACE I)
SCF  Service Control Function
SCM  Sub-Carrier Multiplexing
SCMA  Sub-Carrier Multiple Access
SCP  Service Control Point
SCPC  Single Channel Per Carrier
SCR  Sustained Cell Rate
SD-TV  Standard Definition - TV
SDC  Self-correction Distributed Computation (technique)
SDE  Software Development Environment
SDH  Synchronous Digital Hierarchy
SDL  Specification and Description Language (CCITT)
SDM  Space Division Multiplexing (e.g. individual directionally dedicated fibres)
SDMA  Space Division Multiple Access
SDP  Service Data Point
SDU  Service Data Unit
SE/PO  RACE Systems Engineering/Planning Office
SECAM  Séquentiel à mémoire (french colour TV standard)
SERC  Science and Engineering Research Council
SES  Scientific and Engineering Software
SET  Science Engineering and Technology
SF-DWDM  Sparsely Filled DWDM
SF-WDM  Sparsely Filled WDM
SGDBR  Sample Grating DBR
SGML  Standard Generalised Markup Language
SIBs  Service Independent Building blocks
SIMS  Secondary-Ion Mass Spectroscopy
SIPO  Serial In Parallel Out (shift register)
SISO  Serial In Serial Out (shift register)
SM  Single Mode
SMART  Strategy for Mobile Advanced Radio Telecommunications (initiative of DG XIII/D)
SMATV  Satellite Master Antenna TV
SME  Small and Medium size Enterprise
SMRC  Service Management Reference Configuration
SMTP  Simple Mail Transfer Protocol
SNMP  Simple Network Management Protocol
SNR  Signal to Noise Ratio
SOA  Semiconductor Optical Amplifier
SOC  Successive-Office Control
SOG-IS  Senior Officials Group for Information Security
SOG-ITS  Senior Officials Group for Information Technology Standards
SOG-T  Senior Officials Group for Telecommunications
SONET  Synchronous Optical NETwork (standard)
SP  Service Provider
SPAG  Standards Promotion and Application Group
SPC  Stored Programme Control
SPM  Self Phase Modulation
SRS  Stimulated Raman Scattering
SS  Self-Similar (traffic description)
SSF  Steady State (control theory analysis)
SSGDBR  Service Specification Framework
SSDFB  Service Switching Function
SSGDBR  Super SGDBR
SSGDBR  Super Structure DFB
SSP  Service Switching Point
STAR  Science & Technology Programme for less Advanced Regions
STAX  Standardisation Reference Database
STD  Standard TAXonomy (tool developed in RACE 1)
STER  Synchronous Transfer Mode
STORE  Relational database of project results supporting techno-economic tools
STU  Set Top Unit
SUT  Service Usage Trials
SW  Software
SWOT  Strength-Weakness-Opertunity-Threats (analysis)
TA  Terminal Adapter
TAC  Telephony API
TAC  First generation of Mobile Communication System
TCM  Time Compression Multiplexing (ping-pong mode)
TCS  Time Division Multiplexing
TDM  Time Division Multiple Access
TE  Terminal
TEA  Techno-economic Evaluation
TEA  Transit Exchange
TEA  Techno-economic Evaluation & Demand Analysis
TEA  Techno-Economic Ad-hoc Meeting — RACE tools users informal working group
TEA  Tenders Electronic Daily System
TEA  Techno-Economic evaluation and IBC Demand Analysis (RACE)
TEA  Trade Electronic Data Interchange System (CEC R&D Programme)
TEA  Techno-economic oriented Tools
TEA  Techno-Economic Ad-hoc Meeting — RACE tools users informal working group
TEA  Tenders Electronic Daily System
TEA  Techno-Economic evaluation and IBC Demand Analysis (RACE)
TEA  Trade Electronic Data Interchange System (CEC R&D Programme)
TEA  Techno-economic oriented Tools
TEA  Techno-Economic Ad-hoc Meeting — RACE tools users informal working group
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIFF</td>
<td>Tagged Image Format File</td>
</tr>
<tr>
<td>TM</td>
<td>ETSI technical committee for Transmission</td>
</tr>
<tr>
<td>TN</td>
<td>Telecommunication Management Network</td>
</tr>
<tr>
<td>TMOS</td>
<td>Telecommunication Management and Operations Support</td>
</tr>
<tr>
<td>TO</td>
<td>Telecommunications Operator</td>
</tr>
<tr>
<td>TOFF</td>
<td>Tagged Object Format File</td>
</tr>
<tr>
<td>ToR</td>
<td>Term of Reference</td>
</tr>
<tr>
<td>TPN</td>
<td>Telephony PON</td>
</tr>
<tr>
<td>TS</td>
<td>Traffic Station</td>
</tr>
<tr>
<td>TSI</td>
<td>Time Slot Interchange</td>
</tr>
<tr>
<td>TTG</td>
<td>Tuneable Twin Guide (laser structure)</td>
</tr>
<tr>
<td>UBR</td>
<td>Unspecified Bit Rate</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UDT</td>
<td>Usability Design Target</td>
</tr>
<tr>
<td>UIT</td>
<td>Union Internationale des Télécommunications (International Telecommunication Union)</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunication System</td>
</tr>
<tr>
<td>UNI</td>
<td>User Network Interface</td>
</tr>
<tr>
<td>UPC</td>
<td>Usage Parameter Control</td>
</tr>
<tr>
<td>UPT</td>
<td>Universal Personal Telecommunications</td>
</tr>
<tr>
<td>URM</td>
<td>Usage Reference Model</td>
</tr>
<tr>
<td>UTC</td>
<td>Urban Traffic Control</td>
</tr>
<tr>
<td>UTN</td>
<td>Urban Traffic Network</td>
</tr>
<tr>
<td>VA</td>
<td>Value Added (services)</td>
</tr>
<tr>
<td>VAS</td>
<td>Value Added Services</td>
</tr>
<tr>
<td>VBN</td>
<td>Vorläufer Breitband Netz (German advanced broadband network)</td>
</tr>
<tr>
<td>VBN</td>
<td>Vermitteltes Breitband Netz (existing German Broadband Network)</td>
</tr>
<tr>
<td>VBR</td>
<td>Variable Bit Rate</td>
</tr>
<tr>
<td>VBR-NRT</td>
<td>Variable Bit Rate - Non Real Time</td>
</tr>
<tr>
<td>VBR-RT</td>
<td>Variable Bit Rate - Real Time</td>
</tr>
<tr>
<td>VC</td>
<td>Virtual Circuit</td>
</tr>
<tr>
<td>VCC</td>
<td>Virtual Channel Connection</td>
</tr>
<tr>
<td>VCR</td>
<td>Video Cassette Recorder</td>
</tr>
<tr>
<td>VDSL</td>
<td>Very-high-bitrate Digital Subscriber Loop</td>
</tr>
<tr>
<td>VGA</td>
<td>Video Graphics Adapter</td>
</tr>
<tr>
<td>VHSOL</td>
<td>Very High Speed Optical Loop (a testbed used in RACE)</td>
</tr>
<tr>
<td>VICS</td>
<td>Vehicle Information Communication System</td>
</tr>
<tr>
<td>VIP</td>
<td>Virus Insertion Point, i.e. disk drive</td>
</tr>
<tr>
<td>VLSI</td>
<td>Very Large Scale Integration</td>
</tr>
<tr>
<td>VLSIC</td>
<td>Very Large Scale Integrated Circuit</td>
</tr>
<tr>
<td>VM</td>
<td>Voice Mail</td>
</tr>
<tr>
<td>VoD</td>
<td>Video-on-Demand</td>
</tr>
<tr>
<td>VOD</td>
<td>Video-On-Demand</td>
</tr>
<tr>
<td>VP</td>
<td>Virtual Path</td>
</tr>
<tr>
<td>VPI</td>
<td>Virtual Path Identifier</td>
</tr>
<tr>
<td>VPN</td>
<td>Virtual Private Network</td>
</tr>
<tr>
<td>VSAT</td>
<td>Very Small Aperture Terminal</td>
</tr>
<tr>
<td>VSB</td>
<td>Vestigial Side Band</td>
</tr>
<tr>
<td>VT</td>
<td>Video telephony</td>
</tr>
<tr>
<td>VTR</td>
<td>Video Tape Recorder</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
<tr>
<td>WARC</td>
<td>World Administrative Radio Conference</td>
</tr>
<tr>
<td>WAAS</td>
<td>Wide Area Augmentation System</td>
</tr>
<tr>
<td>WDM</td>
<td>Wavelength Division Multiplexing</td>
</tr>
<tr>
<td>WDMA</td>
<td>Wavelength Division Multiple Access</td>
</tr>
<tr>
<td>WI</td>
<td>Wavelength Interchange</td>
</tr>
<tr>
<td>WIXC</td>
<td>Wuling (optical) XC</td>
</tr>
<tr>
<td>WMux/Demux</td>
<td>Wavelength Multiplexer / Demultiplexer</td>
</tr>
<tr>
<td>WP</td>
<td>WorkPackage — usually a sub-division of a RACE project</td>
</tr>
<tr>
<td>Willingness</td>
<td>Willingness to Pay (see FMAFC)</td>
</tr>
<tr>
<td>WS</td>
<td>Wavelength Selective</td>
</tr>
<tr>
<td>WorkStation</td>
<td>WorkStation</td>
</tr>
<tr>
<td>WSX</td>
<td>Wavelength Selective XC</td>
</tr>
<tr>
<td>WSE</td>
<td>Wannier Stark Effect</td>
</tr>
<tr>
<td>WTA</td>
<td>Winner Takes All (circuit)</td>
</tr>
<tr>
<td>WTDM</td>
<td>Wavelength Time Division Multiplex</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Wait ! ! !</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>X.25</td>
<td>A CCITT Recommendation for packet switched data transmission</td>
</tr>
<tr>
<td>XBBS</td>
<td>Xpress Bulletin Board System</td>
</tr>
<tr>
<td>XC</td>
<td>cross-Connect</td>
</tr>
<tr>
<td>XNOR</td>
<td>eXclusive Not (inverted) OR (logical gate)</td>
</tr>
<tr>
<td>XOR</td>
<td>eXclusive OR (logical gate)</td>
</tr>
<tr>
<td>XTP</td>
<td>Xpress Transfer Protocol</td>
</tr>
<tr>
<td>XTPX</td>
<td>XTP eXtended</td>
</tr>
<tr>
<td>ZBLAN</td>
<td>Zirconium-Barium-Lanthanum-Aluminium-Sodium fluoride (glass)</td>
</tr>
</tbody>
</table>
Chapter One

Introduction

In this chapter we present an overall introduction to the whole thesis. We outline the "Convergence of media, consumer electronics, computing and telecommunication network technology and services" (section 1.1) and discuss the "Potential, requirement and possible impact of an advanced communication infrastructure" (section 1.2).

The chapter concludes with a brief summary and outline of the remainder of the thesis (section 1.3 - "Organisation of the thesis").

1.1 Convergence of media, consumer electronics, computing and telecommunication network technology and services

In addition to the rapid technological progress of optical and telecommunication equipment in general, the "external diver" [1] in the field of communication services is the convergence of the media, electronics and computing industries. For instance, the astonishing rise in processing power of silicon ICs and dramatic price drop of VLSI technology has meant that μ-processor based items, such as computers, video game machines, etc. have changed from specialised products to common every-day household consumer goods.
The emergence of home entertainment products based around advanced technology has within the last few years been subject to an extraordinary development. Home 'games stations' (or video consoles), such as the "Sony Playstation" or the "Nintendo 64" are developed around computer hardware with performances and processing power comparable to work-stations of only a few years ago. Entertainment services have also seen tentative development. Per-Per-View (PPV) and sometimes even modest Video on Demand (VoD) trials are offered by many cable company service providers.

Other areas related to, for example, work and working environments are also undergoing a similar change. Continued Professional Development (CPD) services have in the last few years undergone a massive growth and become very visible [9]. In step with technologies rapid advancement and a requirement for any workforce to be equally more 'technologically literate' and skilled, employers (and employees) are realising the necessity for continued re-training, re-education and continued professional evolution and development [12]. Many of these initiatives are based around CBT (Computer Based Teaching or Training) products.

Collectively it seems clear that the computer, media, technology & telecommunication industries are maturing and converging. Each of those four types of industry influence & impact political, commercial and technical aspects of our every day lives in society. Assuming this process can gain enough momentum to overcome the inherent externalities in society, i.e. its antipathy to change, the sum total might be the emergence of an "IT Communication Revolution", Fig. 1.1, which could influence the
fabric of our society as much (or more) than the development of the steam engine — see also [7-9].

The corporate interest is naturally fuelled by the promises of huge revenue. For example, as an indication of the potential turnover for new and extended Information Technology (IT) shopping services, analyst point to the annual $55 billion catalogue and TV shopping in the USA alone [6]. It should also be noted that the mean income of the majority of IT product consumers typically is above the national average. The typical Internet user in the USA earns around 50% more than the national average [6]. This eases selective marketing, increases penetration and uptake rates of the (targeted) market in general.

Governments, politicians as well as NGOs (Non Governmental Organisations) and other organisations concerned with the social infrastructure and welfare of our society have stated their belief that a competitive country in the world (markets) needs to be at the forefront of development and use of IT [10]. For example, both G7 and EU have started programmes, such as e.g. ACTS [8], that are aimed at encouraging the evolution and development for a new (broadband) telecommunication infrastructure, products and services.

In such a world the traditional role of a telco company would most likely have to be redefined. The roles of Network Operator (NO) and Service Provider (SP) which are often today merged will probably become more separated and distinct. The traditionally large telco it seems, due to its inherent flexibility and slow response time, i.e. time to market, is likely to shift its role to become a carrier of information Bits rather than a SP. Smaller more flexible companies would be better suited as SPs.

A telco would in it's new role as "transmission bandwidth broker" generate revenue by charging service providers for the privilege of using a given bandwidth over a given distance with a given QoS for a given time. It might charge per bit, £x/Bit, irrespective of what the bit contains, i.e. which service is being provided. As an example this would mean that hiring a video film from your local "tele-video shop" would cost the equivalent of between 3 and 8, two hour local telephone calls, assuming that the average bitrate of a telephone call and a video source are approximately 25.6 kBit/s and 67.1–221 kBit/s, respectively [2], or between £14.40 and £38.40 given today's BT charges. Naturally this would be quite unacceptable to the customers as a hiring a video film today only costs around £2.95. Conceptually there seems to be two contrary solutions to this problem. The first, to charge per bit irrespective of which service is being

---

1 Of the 4th of Sep. 1996 BT charged £0.04/min for a standard rate local call.
2 Blockbuster video, East Rd., Cambridge, 13 Jun 98
provided while reducing the price per bit to the lowest common denominator. In the above example this would mean that a standard rate local telephone call would only cost £ 0.003–0.008/min compared to £ 0.04/min given the current price of renting a video film. Naturally, we would initially expect a premium to be associated with this ‘replacement’ service; however, an order of magnitude would seem to exceed even an acceptable disparity due to the added convenience of not incurring the ‘cost’ of travelling to and from the video store, in the long run. The alternative would be to extend today’s approach to charging, i.e. a customer would be billed not only per bit but also as a function of what the bit contained. Although the latter would probably allow for greater telco profit margins it is also more open to misuse. Any benefit must therefore be balanced against the additional administrative overhead and policing functions required within, not only a single telco, but the whole of the telco community.

In addition to this threat to revenue generation, telcos face the problem of the acceleration in the development of technology. Even at the time of writing the world is such that most standards are dead upon arrival. In future the telcos will be faced with the dilemma of whether to remain the standard forbearers in the IT industry and be faced with competition from computer industry “free for all’s”, such as the Internet. Alternatively they could take a step back to being “standard transparent” bandwidth providers and just watch the interface war rage between the customer equipment manufacturers and SPs. Optical ‘transparency’ could be a vital ingredient in any such evolution.

At the root of this ongoing argument lies two industries: The computer versus the telecommunication industry. The computer industry has a desire for total freedom and anarchy in communications. The telcos, on the other hand, wish to have switches which are based on strictly defined protocols placed at strategic locations throughout its network so that it can control and manage the overall process of communication. As these two halves of the IT industry develop and move further apart the tension in the IT industry as a whole is increasing. In reality, neither party will win unless some compromise is reached. Any future network must support and tender for both philosophies, as aspect of them both will be required for some time to come.

One of foundations of this debate is the issue of protocols. The telecommunication fraternity is built on slowly agreed standards and common interfaces, as this after all is the only way we all can communicate? The computer industry, on the other hand, is built on lock-in loyalty established through the de facto criterion. This limitation must be removed

3 Cost in this example could be defined objectively in terms of the actual expense associated with travelling as well as subjectively in terms of the convenience of being able to perform the transaction from home.
if we are to enjoy the full benefits of IT. Recent technological trends & developments in form of for instance of the emergence of the Java programming language [11] show promising signs that such a move may be about to happen. In any event, protocol transparency, possibly achieved through transparent optical or “pseudo” transparent optical networks, will be a key requirement.

Based on an examination of sociological patterns, including work and leisure habits etc., correlated with bit-rate predictions of possible future IT services, estimates for the potential volume of data traffic in the next 40-50 years are as much as three to four orders of magnitude for private and business usage, respectively, greater than today [3]. Revenue is not likely to follow the same exponential growth of traffic volume but rather a linear extrapolation. Studies suggest that it might double every ten years [4, 5]. Combined these two estimates would intimate a reduction in the price of carrying a single bit of a further one or two orders of magnitude compared to the figures derived on the previous page.

Whichever proves to be correct, facts and estimates such as these would seem only to add impetus to the belief that we presently are poised at the head of a remarkable IT revolution which not only may impact the technology products wide used and available but change the very perception of usefulness and value of the services provided. This will almost certainly also impact how charging in telecommunications is performed — on volume of data transfer or the (perceived) value of the information communicated.

It seems clear that the structure of many past and present technology research & development programmes, e.g. RACE & ACTS, have been formulated on the basis of ‘technology push’. This probably reflects the telecom industry’s approach. However, in the longer term a much stronger ‘market pull’ will be necessary if the emerging IT industry is to successfully respond to customers’ true needs for products and new services. Paradoxically this appears closer to the computing industry’s ‘route to market’.

It seems strikingly apparent from the minimal uptake and penetration of e.g. ISDN (I Still Don’t kNow) that any move towards broadband telecommunications services requires not only existence of capable technology but also a much stronger ‘market pull’. System demonstrators would certainly seem one possible avenue for facilitating the transition from one to the other and the continued drive therefore. A system demonstrator may set out to realise one or more of a largely varying number of objectives; ranging from developing & formulating the foundation framework around which technology & systems can be developed to showing the feasibility of implementing, assembling and operating such a system in a non-laboratory environment.

A recent example of this would seem to be the WWW (World Wide Web). The underlying framework for the WWW is essentially the integration of several Internet (OSI level 7) application protocols such as SMTP (Simple Mail Transfer Protocol), FTP (File
Transfer Protocol) & HTTP (HyperText Transmission Protocol) as well as the formulation of (basic) HTML (HyperText Markup Language) which defines the layout and interpretation of published web-pages, such as homepages etc. Early in 1989 Tim Berners-Lee of CERN (Centre Européen pour la Recherche Nucléaire) also known as the European Laboratory for Particle Physics proposed the basis of what has since then been adopted as the foundations of data transfer in the WWW. Initially this was envisaged for use exclusively between various high energy physics research institutions. However, due to their immense practicality and the market's readiness for just such a development, these protocols were soon adopted by many organisations outside the original scope of the project and the W3 Consortium was formed to ensure the continued development and evolution of the WWW standards.

In 1993 a “browser”, Mosaic, was released by NCSA (National Center for Supercomputing Applications). It provided an integrated GUI (Graphical User Interfaced) easy-to-use front-end to the underlying protocols of the WWW. Mosaic was released simultaneously for arguably the three most widely used Internet computing platforms; namely X-Windows under UNIX, the Macintosh and Microsoft Windows. This in itself permanently moved the WWW from the domain of ‘obscure’, specialised usage to within reach of the general public and popular demand has ensured that the WWW protocol suite is by far the most popular Internet service in use today. After the release of Mosaic in 1993 and initial demonstration of the potential for such systems, ‘critical mass’ was quickly reached ensuring the ongoing development of the WWW (Internet) browser applications as well as the underlying frame-work of e.g. HTML. Mosaic was quickly superseded in popularity by other privately developed browsers like e.g. Netscape Navigator which not only enhanced the application to better its usability but also extended the underlying framework of HTML to incorporated additional features in order to facility much more complex publications. A survey by The New York Times of more than 72,000 WWW users in 1995 revealed that although Mosaic at the time was probably the most recognisable browser it had been superseded in popularity by the Netscape Navigator and at that time only accounted for approximately 14 % Internet browser usage. The ‘web’ (& the Internet) had truly graduated from a government funded research project exclusive to specialised use and demonstration of new technology into a commercial reality.

Since then the WWW has continued to expand in both proliferation and penetration into a common every-day communication medium and tool used by higher educational establishments, schools, governmental & commercial institutions and private individuals alike. The fact that the WWW is now sometimes known as the World Wide Wait reflects one of the overpowering themes of all current telecommunication services: The need for bandwidth.
1.2 Potential, requirement and possible impact of advanced communication infrastructure

Advanced communications would rely on individual advanced technology consumer products, a communal infrastructure as well as an abundance of diverse of content.

As stated in the previous section telecommunications might in the future be likely to exhibit a more distinct separation between NOs and SPs, i.e. between infrastructure or resource and content or services. In addition, as the range of available services diversifies, customers’ requirements of a given service widen and the elasticity of tolerance decreases, the marketplace is likely to segment and SPs will be ‘forced’ to target the individual customer profiles more exactly. Within telecoms there seems to be evidence to suggest that this is most likely to be achieved in terms of more selective ‘tailoring’ of service packages to the individual needs and requirements of the various customer groups and possibly the individual customers. Comparatively NOs are likely to seek to reduce their cost-margins by consolidating the infrastructure and systems necessary to provide service.

Companies are already beginning to ‘un-bundle’ the services they provide in an effort to make the individual components they offer capable of being more specific to the individual customer’s needs. It is, for example, not uncommon for major operators to run parallel customer support department for residential and business customers. The author believes that this trend is likely to continue and evolve from support to services and possibly eventually within single services.

From the perspective of the telecommunications engineer, such an evolution is likely to be perceived as a requirement for an increasing diversification in types of traffic for which transport is supported. At a later stage, this may be further supplemented with a need for quite varying QoS matrices, such as delay and loss tolerance and priority, for one and the same service.

Fig. 1.2 below plots some examples of both services and general communication infrastructures bit-rates versus typical session durations. It can be seen that the (commercially) available communication infrastructures which are capable of the performance necessary, at least in terms of bit-rate, typically are limited to Local Area Networks (LANs). In other words, although the terminations and terminal equipment potentially are suitable for a future advanced communication infrastructure they are today

---

4 See also section 5.1 “General statement of the local access requirement” in chapter 5 (part II) for further discussion regarding some of the issues introduced in this section (with particular reference to the access network).
only applied in smaller ‘closed’ systems with short call\(^5\) holding-times. Although a large proportion of the reasoning is undoubtedly associated with economic arguments connected to the implementation it also implies that future communication networks and protocols are likely to exhibit higher communication speeds and more complex termination equipment closer to the customers.

\[10^9\] \[10^{12}\] bits

![High performance inter processor communication rates](High Speed LAN)

10^0 bit 10^3 bits 10^6 bits

![Low Speed LAN](A4 page at 300dpi)

10^6 bits

![A4 page by email](Audio)

![HDTV](Video and TV)

Fig. 1.2 - Examples of traffic bitrate versus duration

In addition to the plethora of communication speeds intimated in Fig. 1.2 above, the different services delivered are likely to have varying QoS metrics as well as typical source traffic characteristics. Fig. 1.3 below plots some examples of Burstiness, defined as the ratio of mean to peak, information rate for services and communication rate (capability) for general communication infrastructures (similar to Fig. 1.2). It can be seen that, as in Fig. 1.2, the information rate characteristics of future services is likely to be closer to communication rate profiles used in computer and \(\mu\)-processor systems than current telecommunication infrastructures. This would seem to be yet another indicator that the boundaries between telecommunications and computing are blurring. It would also seem to

\(^5\) Call in this instance denotes the entire session of or data transfer associated with the service provided as perceived at the application layer, i.e. OSI layer 7.
to point to a requirement for future communication protocol suites needing increased diversity of session and QoS metrics. One only has to look to the ongoing evolution of e.g. the ATM and to a lesser extent the IP protocols to realise that this requirement already has been recognised by the IT-fraternity.

As the number of services offered by telecommunication SPs increases and more features, such as e.g. (complex) IN (Intelligent Networking) services, are 'built-in' to the network, it will become increasingly complex to disentangle service and infrastructure, if the current approach to network design is not changed. Certainly 'transparency' of the transport layer will be an absolute requirement and possibly even (partial) transparency of the physical, in the form of e.g. 'optical transparency' would become desirable. Perhaps not surprisingly the larger more established (ex-monopoly) operators are amongst the more staunch critics of 'un-bundling' and whether it should be imposed by law or legislation. However, it is the authors' belief that if even huge (multi-national) operators do not open themselves to 'targeted marketing' and the network, system & technological requirements that come along with that, they may face a real danger of becoming huge immobile and 'dead' dinosaurs.

![Fig. 1.3 - Examples of traffic burstiness versus bitrate](image)

It is clear from the above discussions that one overwhelming issue of tomorrow's IT service provision will be bandwidth or more specifically information capacity of the networks & systems.
However, another debate, which has recently started emerging, is mobility. The advent of mobile communication systems in the form of e.g. cellular and GSM telephones is perhaps arguably the most striking change in telecommunication services in this decade. Although many working in telecommunications assume that mobility poses little or no threat to the existing fixed networks this is almost certainly shortsighted. Presently wired and wireless access telephony services coexist but as Simon Forge of Cambridge Strategic Management Consultants, author of the study: “Near-zero tariff telecommunications”, argues “mobile telephones may in time do to public networks what the railways did to canals”; namely, effectively replace it.

There already is quantitative evidence emerging that there within some user-groups is a shift of preference from fixed- to mobile-line based services: “Worldwide, one new telephone subscriber in six gets a mobile phone” and “In Sweden one person in [every] six has a mobile phone. Indeed, fixed-line connections are now declining: flighty young people choose to have a mobile phone when they leave home, rather than pay a new fixed-line connection every time they move.” [13].

What is missing in the mobile-telecoms market at the moment would seem to be bandwidth. Cellular telephones typically use a large proportion of the available ‘communication space’ to facility robust transfer of data under roaming conditions. The terminal equipment, in this case the customers’ handset, use a large proportion of the available ‘address space’ to switch between alternative transmitters in order to increase its robustness to (transient) fading conditions of one of them.

The PHS (Personal Handiphone System) system does not employ diversity reception and thereby increases the proportion of the address space available for communication. This system will therefore be able to transmit still graphic images (and maybe even low frame-rate video). A ‘price’ of this approach is naturally that a customer would probably not be able to receive data while in transit in e.g. a speeding car or train; but this is also balanced by the likely much reduced ‘cost’ of the NTE (NT (Network Termination) Equipment), i.e. the handsets, due to the lack of advanced switching technology.

The objective of the above comparison systems in not to measure the relative merits of the two technologies, but simply to add weight to the argument that customers, in the future, may prefer to be able to have a single handset capable of terminating signals from what now are separate systems, say satellite, cellular (GSM), PHS, paging and landline systems and automatically switching between them allowing a customer to be reached at any time on a single number; or possibly multiple numbers all terminating on the same piece of NTE. A precursor of the next generation of NTE might be the Nokia 5020 GSM mobile phone which allows the operator to programme different responses to an
incoming call, e.g. ring (& answer), direct to VM (Voice Mail) or redirect to another number, dependent on the pre-programmed categories of CLI (Calling Line Identification) numbers.

The objective of the above examples and discussion is as a general introduction and background to a few of the fields the author believes to be relevant to the general study of telecommunications and the context within which the work & results presented in the remainder of this thesis was performed & derived.

1.3 Organisation of the thesis

In this section we briefly summarise the organisation of the thesis and the work undertaken, results produced & presented as well as highlight the time and context within which they were performed. The studentship started in Oct. '92 and as mentioned in the abstract the falls broadly into three parts.

The first was occupied with the application and possibilities of photonics in various types of network and was performed from '92 to late 93. This work is presented in Part II of the thesis. After an initial introductory chapter (chapter 4), where the author discusses and outlines the potential of optics in switching and some of the key issues, chapter 5 explores the potential for specifically Wavelength Division Multiplexing (WDM) in future broadband access networks. Chapter 6 comprises, in addition to an introduction and conclusion, the two main papers published concerning a developed (and patented) transparent All Optical Network (AON) architecture. As discussed in the introduction, section 6.1, the premise for the network architecture developed and analysed in chapter 6 was an investigation of the potential of the application of optical technology in (future) telecommunication networks with "today's statute quo" (at the time) as a starting point. Although the proposed network architecture is discussed in terms of a national (UK) network, it not to be overlooked that this was done predominantly as an exercise to show the limits to which it might be possible to push the proposed architectural concept, rather than as a serious proposal for the next generation of networks. Indeed, it is the author's opinion that although the analysed concept of a Sparsely Filled WDM (SF-WDM) transmission spectrum was shown to be very powerful it, for many reasons concerning the reality of implementing and maintaining networks, would probably be best utilised within a single network layer or within individual transmission systems. In this case the optoelectronic interface, Fig. 6.3.7, might be situated at the nodes interfacing to the access network (i.e. at the Central Office (CO), refer chapter 5). The approach proposed yielded some novel results and also 8 claims under a patent application filed by the studentship co-
sponsoring organisation, BNR Europe Ltd., with the author (Martin Sabry) and John E. Midwinter as inventors [14]; however, at the time (and even at present) it made some stark assumption of what might be possible wrt implementation of optical switching and transmission systems.

Consequently, the second portion of work, presented in part I of the thesis, undertaken from late ‘93 / early ‘94 to mid ‘94, dealt with the actual implementation of optical networks. Chapter 2 reviews the enabling technologies; including fibre, sources, i.e. lasers, filters and optical amplifiers (at the time). Chapter 3 describes a method for predicting limitations of transmission systems imposed by the nonlinear fibre and amplifier noise constraints. The method is applicable to all transparent networks and specific results are quoted in the context the architecture proposed in part II, (section 5.6 & chapter 6). It must be noted that this work and the results presented were very much a first attempt at quantifying some of the assumptions made in the exploration in of the scope of optical network design summarised in part II in terms of fundamental limits.

The studentship was at this point effectively suspended due to a serious accident for approximately 15 months and once resumed towards the end of ’96 both of these areas had been ‘passed on’ to other Ph.D. students — which both have since graduated. This situation effectively prohibited either of these areas being revisited in the light of subsequent and up to date research and consequently necessitated a change in research direction.

Part III of the thesis is concerned with the work of this period of the studentship and summarises some of the ideas & results, which resulted from a body of research concerning relating charging and management & control mechanisms in future (broadband, multiservice) telecommunication networks. The thrust of the work was to investigate the effects of dynamic charging and using the price or customers’ Willingness to Pay (WP) as a control variable. This area of research has since also been passed on to another (& current) Ph.D. student within the department.

The thesis concludes with a chapter outlining and summarising the overall conclusions the author believes can be drawn from the work outlined in the three parts mentioned above.

Appendix A consists of a short biography of the author and a list of the publications, patents and presentations achieved in the 3 year PhD EPSRC (CASE) studentship.

Appendix B comprises ‘other work’ not directly quoted or referenced in the main body of the thesis carried out during the same 3 year studentship.

As stated in the Abstract, the individual parts, chapters and sections which constitute them have been written, in so far as it has been possible, to be individually readable.
By agreement with Prof. Midwinter, supervisor, and Dr. Mick Flanagan, postgraduate tutor, different sections and chapter in this thesis corresponds to some of the (published) work and results achieved during the PhD EPSRC studentship. Specifically, chapter 5 and a large proportion of chapter 2 is based on a consultancy study undertaken for GPT Ltd. Chapter 3 is based on a body of work concerning (theoretical) limits of linear optical WDM transmission systems. Although initial results had been achieved this work was never completely finalised by the author as the PhD was interrupted due to a serious accident and this direction of research was passed on to other PhD student within the department. Chapter 4 is largely based around the founding thoughts and thinking by the author (and others) in the department regarding photonic networks and two EU consultancy reports produced by Prof. Midwinter regarding the same. (In agreement with Prof. Midwinter, supervisor, and Dr. Mick Flanagan, postgraduate tutor) Chapter 6 consists of reprints of the two main publications (and the patent) of a unique concept concerning optical network architectures as well as introductory and concluding sections, in which the premise and a retrospective critique of the work is summarised. Chapter 7, 8, & 9 summarise the work which was undertaken and the direction of research which started after the PhD was restarted (after interruption due to accident) concerning telecommunication network & resource management and control and charging. Although only performed in the final 13 months (which also includes the time used to ‘write-up’) of the studentship a number of significant results were achieved and a couple of publications (and possibly 3 patents) concerning the work and results are pending. This line of research will naturally also be continued within the department and has already had a studentship (co-sponsored by Nortel Ltd) agreed and assigned to it.

1.3.1 Statement of Originality

In this section we attempt to briefly highlight the areas where we feel that the author’s work has contributed to original work in the area of telecommunications:

- Design of optical networks
  - Sparsely Filed Wavelength Division Multiplexed (SF-WDM) network & transmission systems’ architecture
  - Review of the potential for WDM in local (access) broadband networks
- Implementation of optical transmission systems
  - Detailed review of available component technology (in 1994)
• Proposal of linear method of analysis for the transmission limitations imposed by optical fibre non-linearities

• A novel approach to control & management of telecommunication resources and networks through charging with particular reference to the application of 'free-market' principals

• Detailed work and analysis of dynamic / 'Smart Market' / 'Free Market' charging
  • The '(fast) reflector switch'
  • The maximum cost of resources charging scheme when applying dynamic charging to multiple network resources
  • Price 'lifetime'

• Novel analytical method of analysis of charging mechanisms in telecommunication networks & systems using control theory

In addition to a number of other areas of work not directly connected to the work summarised in this thesis:–

• Architecture and design of novel application of computer based multimedia and telecommunication technology to remote self tuition and learning

• Application of novel Neural Network design and implementation to Forward Error Correcting (FEC) decoders used in telecommunication systems
1.4 Bibliography

FAIR (Forecast and Assessment of socio-economic Impact of Advanced Communications and Recommendations) ACTS project AC093 homepage, http://www.analysys.co.uk/acts/fair


"Like a flock of birds" in The Economist, "The Accidental Superhighway - A survey of the Internet", July 1st, 1995


Keith Ward, "Regulation", 23 pages, University College London, May 7, 1996

Keith Ward, "Manufacturing Industry", 10 pages, University College London, June 12, 1996

Keith Ward, "Global Telecommunications", 12 pages, University College London, May 16, 1996


Keith Ward, “Marketing”, 12 pages, University College London, June 20, 1996
1.5 References


It's easy to cry that you're beaten — and die;
It's easy to crawfish and crawl;
But to fight and to fight when hope's out of sight —
   Why, that's the best game of them all!

And though you come out of each gruelling bout,
All broken and beaten and scarred,
Just have one more try — it's dead easy to die,
It's the keeping-on-living that's hard.

Robert Service
Chapter 2

Photonic Network Enabling Technologies

In this chapter we present a review of the enabling technology available for use in optical networks. After a short introduction where we discuss and compare the nature of electronics and optics, we will review, in turn, optical fibre, sources, filters and amplifiers as some of the key components of optical communications.

2.1 Introduction - optics versus electronics

One could regard optoelectronics as a branch of the broader field of electronics and assume the two are similar. However, it is important to understand the quite opposing nature of these two technologies, some of which are summarised in Table 2.1 below. At the most basic level, one can observe that electrons interact strongly, one with another, while photons do not. Accordingly, it is perhaps not surprising to see light being extremely good for transporting information, where non-interaction is an advantage, and electronics being supremely good at processing information where interaction is all important. Where photons have been made to interact with each other, it is almost always through the medium of electrons in a suitable material (gas, liquid or solid) and as a result, many "optical non-linear devices" prove, on closer inspection, to be electronic (controlled) devices that are optically triggered and read.
<table>
<thead>
<tr>
<th>ATTRIBUTE</th>
<th>ELECTRONS</th>
<th>PHOTONS</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual interaction</td>
<td>Very strong</td>
<td>Weak except through electrons</td>
<td>Implies photons good for transmission, electrons for processing information</td>
</tr>
<tr>
<td>Attenuation in cables</td>
<td>Resistance and skin effect</td>
<td>Very low</td>
<td>Light goes much further</td>
</tr>
<tr>
<td>Dispersion</td>
<td>Finite BW, i.e. high</td>
<td>Very low or zero (solitons)</td>
<td>Optical pulses spread less or not at all</td>
</tr>
<tr>
<td>Carrier frequency</td>
<td>0-300 GHz</td>
<td>more than 200,000 GHz</td>
<td>Optics has spectrum &quot;to burn&quot;, hence good for transmission</td>
</tr>
<tr>
<td>EM Wavelength</td>
<td>&gt;1 μm</td>
<td>0.2-10 μm</td>
<td>Optics good for imaging!</td>
</tr>
<tr>
<td>Electron wavelength</td>
<td>0.001 μm (for 1 volt)</td>
<td>N/A</td>
<td>Implies much smaller electronic devices possible</td>
</tr>
<tr>
<td>Shortest pulse</td>
<td>1-10 ps</td>
<td>0.01 ps</td>
<td>Optics allows ultra-fast probing</td>
</tr>
</tbody>
</table>

Table 2.1 - Attributes of photons and electrons [86]

This observation impinges strongly upon later discussions (in Part II of the thesis) of the possible role of optics in switching and we believe carries with it a strong message on how light should be used in such applications. Moreover, since the electron wavelength even at low energy is very much shorter than the optical wavelength, electronic devices can be and are much smaller than optical devices. This has the important effect that, in general, much higher packing densities of electrical interaction devices are possible than for optical ones.

The radio and microwave spectrum extends from low frequencies in the kHz range to the millimetre wavelength region with a frequency of about 300GHz. The optical carrier frequency for a 1500nm fibre system is about $2 \times 10^{14}$ Hz or 200,000 GHz and accordingly, relatively small fractional bandwidths on optical carriers offer massive equivalent electrical bandwidths. For example, the ~35nm bandwidth of an Erbium Doped Fibre Amplifier (EDFA) centred close to 1500nm centre wavelength corresponds to about 4700 GHz, over ten times the complete radio and microwave spectrum (see Fig. 2.1), yet it is a fractional bandwidth of only $(35/1500) = 2.3\%$. It follows that the blocks of data that can be assembled and processed electronically can only fill a very small part of the potentially available bandwidth in the optical fibre. This points to the optoelectronic interface as the remaining major element controlling what data can be impressed upon the fibre. Thus, in planning future uses for optics, it is as well to recognise that circumventing this bottleneck is likely to feature strongly in system design.
Another key issue that must be faced is that today optoelectronics is largely a discrete device technology, in stark contrast to silicon electronics where a million active devices per chip are not uncommon. The major exceptions to this statement fall primarily in the consumer/professional electronics area where CCD cameras and Flat Panel Active-Matrix LCDs both achieve very high levels of integration although again by building upon the power of silicon technology. Note that both of these examples are concerned with interfacing between the planar 2D world of electronics and the 3D world of imaging optical systems.

We believe that many of the applications we will discuss point clearly to a need to achieve much higher levels of sophistication/integration in optoelectronic device technology than hitherto has been the case. This also underlines the key importance of devices in optoelectronics where the ready availability of a huge range of standard mass produced general purpose yet powerful building blocks do not exist, not to mention the custom component (sub-system!) design and fabrication, such as one takes for granted in silicon devices.

The result of this is two fold. Advanced systems rely heavily on access to small numbers of advanced and key components that are not generally available from multiple vendors and optoelectronics, at least so far as telecommunications is concerned, is characterised at present by a profusion of specialist low volume components. In many markets the move to a small number of high volume components is an essential pre­requisite of cost reduction. Unfortunately, the major mass market for optoelectronic components today is the consumer one and both the expertise and cash flow generated by that, is largely concentrated in Japan. We suggest that an important long term objective for an optoelectronic integrated-circuit programme should be the achievement of a single technology that can provide a wide range of standard components as a potential way of redressing the balance here.

2.2 Fibre

A central component to any photonic network is optical fibre. Typically we categorise single mode fibre as being one of three distinct types: Conventional Fibre (CF), Dispersion Shifted (DS) and Dispersion Compensating (DC). In table 2.2 below we list typical dispersion values for each of these.
<table>
<thead>
<tr>
<th>Fibre type</th>
<th>Wavelength (nms.)</th>
<th>Dispersion (D) (ps/(nm.km))</th>
<th>(dD/d\lambda) (ps/(nm².km))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF</td>
<td>1300</td>
<td>~0</td>
<td>0.06</td>
</tr>
<tr>
<td>CF</td>
<td>1550</td>
<td>17-20</td>
<td>0.05</td>
</tr>
<tr>
<td>DS</td>
<td>1550</td>
<td>~0</td>
<td>0.045</td>
</tr>
<tr>
<td>DC</td>
<td>1550</td>
<td>-50-60</td>
<td>-0.05 to -0.2</td>
</tr>
<tr>
<td>DS+</td>
<td>1550</td>
<td>+2 to +4</td>
<td>-0.05 to -0.2</td>
</tr>
<tr>
<td>DS-</td>
<td>1550</td>
<td>-2 to -4</td>
<td>-0.05 to -0.2</td>
</tr>
</tbody>
</table>

Table 2.2 - Fibre design types

More recently, fibres have been developed with dispersion values close to but not equal to zero in the middle of the EDFA transmission window. We have indicated these above by the nomenclature DS+ and DS-. These will be discussed in more detail in the following chapter.

What is already clear is that the constraints imposed by the need to minimise non-linear fibre effects in very high performance systems will require a more complex link design than has previously been thought necessary when overall balancing the attenuation and dispersion largely sufficed. In a Dense WDM network, one may be interested not in equalising dispersion at a single wavelength but over a range of wavelengths, implying a need to tailor both the 1st and 2nd order dispersion, while the need to control the multiple non-linear effects leads to special requirements to tailor the local dispersion at every point in the link and not just to design for it average to zero overall. For this requirement we find the new types of fibres being discussed (e.g. DS+ and DS-) although how such fibres would be deployed in practice, particularly in inland systems where cable lengths vary according to the location of jointing points, remains very unclear. Probably the first issue to be firmly resolved will be the potential transport capability of CF links using dense WDM transmission and readily installed dispersion compensation elements.

2.3 Lasers

A common theme in many optical networks today is to produce an array of a number of wavelength channels at distinct frequencies/wavelengths that sources and receivers can tune to and thereby establish a unique communications circuit. In this section we review some of the laser devices which have been proposed as potential sources in current and future optical networks.
2.3.1 Wavelength tuneable lasers

In WDMA systems, in particular, the difficulty of matching absolute frequency references production run after production run is of concern. As an example Table 2. below lists the wavelength variation caused by the tolerances of fabrication parameters for a 1.55\textmu m $\lambda/4$ phase shifted DFB laser [41]. It can be seen that the total mean variation in lasing wavelength is almost 15nm. This result suggests that even fixed frequency WDMA systems are likely to require active wavelength tuning either of the filter or the laser.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variation</th>
<th>Wavelength variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active layer width</td>
<td>±20%</td>
<td>±2.8nm</td>
</tr>
<tr>
<td>Active layer thickness</td>
<td>±17%</td>
<td>±6.3nm</td>
</tr>
<tr>
<td>Guide layer thickness</td>
<td>±17%</td>
<td>±1.3nm</td>
</tr>
<tr>
<td>Active layer composition</td>
<td>±0.6%</td>
<td>±2.0nm</td>
</tr>
<tr>
<td>Guide layer composition</td>
<td>±0.9%</td>
<td>±0.5nm</td>
</tr>
<tr>
<td>Grating depth</td>
<td>±22%</td>
<td>±0.8nm</td>
</tr>
<tr>
<td>Grating shape</td>
<td>sine/rect</td>
<td>±0.6nm</td>
</tr>
<tr>
<td>Grating pitch</td>
<td>±0.1nm</td>
<td>±0.7nm</td>
</tr>
<tr>
<td>Total standard deviation</td>
<td></td>
<td>±7.4nm</td>
</tr>
</tbody>
</table>

Table 2.3 - Fabrication tolerances [41].

There are many types of wavelength tuneable lasers. Table 2.4 summarises the physical characteristics of the “state of the art” lasers structure described in the literature.
## Table 2.4 - Wavelength tuneable lasers

**Grating tuned external cavity lasers** - The most conventional method of constructing a tuneable laser is to use an external grating cavity. Tuning is achieved by rotation and axial displacement of the grating. The maximum continuous tuning range obtained is 160nm [2] and 55nm [85] at 1.3µm and 1.5µm respectively. Since tuning involves mechanical movement of the grating, tuning is slow and very susceptible to environmental instability. For these reasons these types of lasers will probably not be suitable for widespread deployment in most types of network.
**Acousto-optical tuneable semiconductor lasers** - Acousto-optically tuned lasers are also external cavity structures. However, instead of a diffraction grating acting as the narrow band filter an acousto-optic filter is inserted between the mirror and the laser diode. Tuning of the laser is achieved by electronically tuning the filter. As mechanical movement is eliminated, a relatively lower requirement of mechanical stability may be enjoyed. (The mechanical stability necessary is still higher than that of monolithically integrated tuneable laser structures as the external cavity alignment accuracy must be upheld.) Due to the wide tuning range of the acousto-optic filter these lasers exhibit a large tuning range with modest power requirements. However, as the filter can only select one of the modes of the external cavity, the lasing wavelength can only be discretely tuned at the resolution of the mode separation. Although this does not conceptually cause a problem in WDMA architectures, it imposes smaller tolerances on any single frequency lasers used.

**Electro-optical tuneable semiconductor lasers** - The electro-optical tuneable laser is similar to that of the acousto-optical laser. Rather than an acousto-optic filter these lasers use an electro-optical filter. For this reason they exhibit a much reduced tuning speed (a few ns rather than ten's of µs) as well as tuning range. In addition, an electro-optic filter typically requires ±50V bias.

**Multiple electrode DFB lasers** - The alternative to external cavity tuneable lasers are monolithically integrated lasers. The two most used laser structures are DFB and DBR lasers. The advantage of DFB lasers is their ease of fabrication, high tuning speeds and narrow linewidth. The disadvantage is the limited tuning range and careful control of the control currents (two or three) is necessary to ensure optimum performance.

**Multiple electrode DBR lasers** - The three sectioned DBR laser has an extended tuning range compared to the DFB laser, particularly when configured for quasi continuous tuning. However, they have a broader linewidth and more complicated control.

In addition to the basic DFB and DBR structures there has recently been much interest in so-called sampled grating and super structure grating structures. These have demonstrated large, > 100nm, quasi continuous tuning ranges.

There have also be a large number of proposals for new types of lasers. One of the more interesting is the DFC laser as it potentially may facilitate 50-100nm tuning requiring only one control current.
2.3.2 Multiwavelength laser arrays

An alternative approach to producing WDM channels is Multi Wavelength (MW) laser arrays. The aim is to produce and array of laser sources at distinct wavelengths as a single Photonic Integrated Circuit (PIC). The main advantage is to reduce the complexity at points in the network where more than one laser source originates, typically the head-end. Table 2.5 summarises current results. Arrays with between 4 and 20 channels separated by 0.5nm to 15nm have demonstrated. Although surface emitting two dimensional arrays seem very promising both in terms of the number of channels and the channel separation, the practicality of coupling the emitted light into optical fibre probably makes these structures impractical.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Channels</th>
<th>Channel spacing</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW linear array</td>
<td>20</td>
<td>1-2nm</td>
<td>[11]</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10-15nm</td>
<td>[16]</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3nm</td>
<td>[17], [18]</td>
</tr>
<tr>
<td>MW PIC linear array with amplifier/combiner</td>
<td>4</td>
<td>2.5nm</td>
<td>[12], [13], [14]</td>
</tr>
<tr>
<td>MW PIC array with individual modulators</td>
<td>4</td>
<td>0.5nm</td>
<td>[15]</td>
</tr>
<tr>
<td>MW surface emitting 2-D array</td>
<td>16</td>
<td>0.9nm</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>0.3nm</td>
<td>[20]</td>
</tr>
</tbody>
</table>

Table 2.5 - Multiple wavelength laser arrays

2.4 Filters

Complementary to being able to produce a wavelength array is the ability to select regions and individual channels within it. This may be required along transmission lines, in optical switches or in terminals and could ultimately be part of the equipment on the customer premises. Filters enable this function.

There are essentially two types of optical filters: fixed wavelength and wavelength tuneable.

2.4.1 Fixed wavelength optical filters

Fixed wavelength filters based on gratings have long been demonstrated. Typical performance is 20 channels spaced 1nm apart [72].

2.4.2 Wavelength tuneable optical filters

The three basic mechanisms of optical filtering can be understood under a single unifying principle [38]. The following sections present a short review of the different
interferometric, mode coupling and active semiconductor based filters shown to be wavelength tuneable. Results are summarised in Table 2.6.

**Interferometric filters** - The two most common types of interferometric filter are the Fabry-Perot and Mach-Zehnder filters. The Farby-Perot filter uses two parallel high reflectivity surfaces separated by a gap filled by a dielectric of refractive index, n, to form a resonant cavity. Depending on the reflectivity such a device will exhibit a number of transmission peaks separated by the free spectral range of the device. Tuning the filter can be done by adjusting the gap length, using piezo-electric or electrostatic control, or by rotating the whole device with respect to the incident light, or by changing the dielectric constant of the material separating the two reflective surfaces. Changing the gap length and rotating the device involve mechanical movement and a reasonable degree of mechanical stability is, therefore, required. Tuning by varying the dielectric constant can be achieved by filling the gap between the two reflective surfaces with liquid crystal. In this case electronic control is possible. Similar tuning ranges for mechanical and electronic tuning have been demonstrated but the bandwidth for the liquid crystal Fabry-Perot is typically an order of magnitude larger.

Tuneable Mach-Zehnder devices using temperature control to adjust the path lengths and, hence, the passband frequencies have also been demonstrated. However, for a large number of channels this method of control becomes quite complicated as the length of each arm in each successive stage must be individually controlled.

**Mode coupling filters** - Mode coupling filters have been demonstrated as both electro- and acousto-optic structures. They show similar characteristics to their laser counter parts. Electro-optical filters offer ns tuning speed and higher resolution but a limited tuning range. Acousto-optical filters have μs tuning speed and a lower resolution but a much increased tuning range and overall a greater number of channels. However, if the transmission window is limited to 35nm (the EDFA transmission window) both types of filter have approximately the same number of channels.

**Active semiconductor filters** - active semiconductor filters are polarisation sensitive and only offer resolution of a few channels. However, they may incorporate gain and the overall insertion loss can be limited to 0dB.
### Table 2.6 - Wavelength tuneable filters

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Tuning range</th>
<th>Tuning speed</th>
<th>Bandwidth</th>
<th>Channels</th>
<th>Loss</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabry-Perot etalon</td>
<td>50nm</td>
<td>ms</td>
<td>&lt;0.01nm</td>
<td>100s</td>
<td>~5dB</td>
<td>[59]</td>
</tr>
<tr>
<td>Fibre Fabry-Perot</td>
<td>50nm</td>
<td>ms</td>
<td>&lt;0.01nm</td>
<td>100s</td>
<td>~5dB</td>
<td>[60]</td>
</tr>
<tr>
<td>Liquid Crystal Fabry-Perot</td>
<td>50nm</td>
<td>ms</td>
<td>0.2nm</td>
<td>50</td>
<td>7dB</td>
<td>[62]</td>
</tr>
<tr>
<td>Mach-Zehnder</td>
<td>4.5nm</td>
<td>ms</td>
<td>0.04</td>
<td>128</td>
<td>7dB</td>
<td>[61]</td>
</tr>
<tr>
<td>Electro-optic</td>
<td>16nm</td>
<td>ns</td>
<td>0.6nm</td>
<td>10</td>
<td>5dB</td>
<td>[63]</td>
</tr>
<tr>
<td>Acousto-optic</td>
<td>400nm</td>
<td>10μs</td>
<td>1nm</td>
<td>100s</td>
<td>5dB</td>
<td>[64]</td>
</tr>
<tr>
<td>DFB</td>
<td>0.4-0.5nm</td>
<td>ns</td>
<td>0.1-0.2nm</td>
<td>2-3</td>
<td>0dB</td>
<td>[65]</td>
</tr>
<tr>
<td>3-section DFB</td>
<td>0.6nm</td>
<td>ns</td>
<td>0.025nm</td>
<td>8</td>
<td>0dB</td>
<td>[66]</td>
</tr>
</tbody>
</table>

#### 2.5 Amplifiers

The development of optical amplifiers over the past few years has to a large extent effectively rendered the traditional 3R (Receive-Regenerate-Retransmit) opto-electronic sub-systems superfluous. One of the major applications of optical amplifiers is as replacements for the in-line opto-electronic repeaters in optical transmission systems. The optical amplifier has several advantages (over the traditional 3R repeaters):–

- The reliability of ‘repeater’ sub-systems may be greatly increased by reducing the volume of electronics in them.
- Because the bandwidth of, in particular, optical fibre amplifiers is quite wide optically transparent systems based around WDM transmission may be established and found to be economically viable.
- This to a certain degree would ensure ‘future proofing’ of transmission systems as they could be upgraded by replacement of only the transmitters and receivers or by the addition of WDM channels. (This statement although ultimately true, is in the author’s opinion only relative, and still remains to be shown in practise.)

The two main types of optical amplifier currently of interest are the semiconductor and doped fibre amplifiers; in particular the Erbium Doped Fibre Amplifier (EDFA). These two types of amplifier do actually exhibit quite different characteristics and will consequently probably find completely distinct application in photonic networks.
2.5.1 Semiconductor amplifiers

The semiconductor amplifier is based around a conventional semiconductor laser structure. The device is biased (just below the lasing threshold) such that light coupled into one facet of the laser structure appears amplified at the other. These devices will naturally exhibit Fabry-Perot resonances and consequently only exhibit gain over a limited bandwidth; maybe as little as \( \sim 5 \text{ GHz} \). However due to their physically small dimensions this type of amplifiers will probably find application as part of Photonic Integrated Circuits (PICs) in larger systems and sub-systems. The limitations of semiconductor amplifiers makes their use as remote in-line amplifiers unlikely.

2.5.2 Fibre amplifiers

In the doped fibre amplifier the gain medium is a length of doped silica or fluoride fibre. Although systems with amplifiers with more than 10km fibre lengths have been demonstrated, the length of this fibre is typically between a few to a couple of tens of metres. With the EDFA gain is achieved in the 1.55\( \mu \text{m} \) (low loss) transmission window as Erbium has an emission line at 1536nm, see schematic depicted in Fig. 2.1 below. Pumping is usually done by a source with a wavelength of between 980nm and 1480nm due to availability of sources.

![Schematic gain spectrum of EDFA](image)

As can be seen from Fig 2.1 gain is attained over \( \sim 35\text{nm} \) and may through equalisation be designed to be relatively flat. Combined with the facts that the time constant associated
with repopulation of the higher energy levels is of the order of milliseconds (compared to nanoseconds for the semiconductor amplifier) and the gain is polarisation insensitive make this type of amplifier ideal for use as an in-line amplifier and also for use in WDM transmission systems.
2.6 Bibliography


Midwinter, J. E., “Towards Photonic Networks”, EC Contract No. 18075.00, August 1995

Midwinter, J. E. and Staff, “Potential for Wavelength Division Multiplexing in Local Broadband networks”, Order reference 94/1/0000793
2.7 References


77


This chapter is based on a body of work which followed the invention and development of a novel transparent optical network architecture (which is discussed in Part II of the thesis) investigating and challenging the limits of implementing AONs. The paper “Transmission Limitations due to Optical Fibre Nonlinearities in Transparent Optical Networks” was, however, never published as its submission was pre-empted due to an interruption in PhD because of a serious accident and post-empted by the topic being ‘passed on’ and a subsequent change in research direction.

Optical transmission systems

3.1 Transmission limitations due to optical fibre nonlinearity in transparent optical networks

3.1.1 Introduction

The concept of transparent optical networks has in the last few years become a rapidly developing research area. Many different network architectures suitable for various types of communication have been proposed. Common to many, in particular the large scale networks, is that they rely on Wavelength Division Multiplexing (WDM) and optical amplifiers, typically Erbium Doped Fibre Amplifiers (EDFAs). Both of these techniques, however, give rise to increased distortion of transmitted signals. EDFAs do not naturally exhibit uniform gain throughout their ~35nm transmission window. Hence, cascading a number of amplifiers together requires either gain flattening of the amplifier or some method of gain equalisation in order to ensure that the shorter wavelength carriers may be
recovered. In addition, an amplifier chain means that Amplified Spontaneous Emission (ASE) noise will build up and eventually also make the carrier message unrecoverable. Finally, densely multiplexing optical carriers within the frequency domain generally means a high spectral power density, which may stimulate non-linear effects within the optical fibre. We assume there is no crosstalk induced by the EDFAs.

In this section, we develop a method for optimising the number of amplifiers as a function of the total transmission distance, the number of multiplexed carriers, the separation between carriers and the parameters of the fibre. We present results which test the assumptions made relating to our own [1-3] and other previously published network architectures.

Fig. 3.1 depicts the spectrum used in the transmission model. It is a Sparsely Filled Densely Wavelength Division Multiplexed (SF-DWDM) spectrum as it has $W$ wavelength slots, each of which accommodate one of $fW$ carriers with probability $f$. In networks where a number of optical channels, distinct in the frequency domain, are used to make transparent optical connections between sets of nodes, the case $f<1$ permits greater flexibility when allocating channels. In general this makes it easier to avoid contention between optical channels holding identical wavelength slots. In terms of architecture and control, this permits a much simplified network. In fact, using this technique we have shown it is possible to design a high performance national transparent optical network without the use of wavelength or time shifting [1-3], refer also Part II. In terms of optical transmission, however, using this technique means that for a given number of optical carriers, some will be spaced closer together. In turn this causes a locally higher spectral power density and an increased susceptibility to optical nonlinearities. The case $f=1$ is equivalent to system which make use of a fully filled transmission spectrum with equally spaced carriers.

![Fig. 3.1 - SF-DWDM spectrum](image)
Optical fibre can, in the context of telecommunications, normally be treated as a linear medium. However, it does in reality exhibit a number of different nonlinear effects. Therefore, a high power density, caused by a large number of frequency multiplexed optical carriers, within a single fibre, might cause degradation of the carried signals due to one or more of these effects. Consequently, we would like to minimise the optical power per carrier so as not to invoke any of these nonlinearities. On the other hand, in order to avoid a fatal build-up in ASE noise, we should maximise the input power of the optical carriers to each segment of amplified fibre and thereby reduce the number of amplifiers required for a given transmission distance. Thus, we have an optimisation problem which is dependent on the number of amplifiers, the number of frequency multiplexed carriers and the length of fibre, as well as all the individual parameters describing each of the nonlinearities for any given system. The succeeding sections each present a short review of the nonlinear effects Stimulated Brillouin Scattering (SBS), Stimulated Raman Scattering (SRS), Self Phase Modulation (SPM), Cross Phase Modulation (CPM), Four Photon Mixing (FPM) and the noise implications of cascading amplifiers, followed by analysis and results.

3.1.2 Stimulated Raman Scattering

SRS is induced by the vibrations of silica molecules in optical fibre modulating the effective dielectric constant. In a WDM system it causes the optical power contained in the channels transmitted on the shorter wavelengths to be transferred to the channels transmitted on the longer wavelengths [4-8]. That is, the higher frequency channels will experience excess attenuation due to SRS. This attenuation will be the greatest for the carrier occupying the shortest wavelength channel slot. The worst case attenuation for this channel in a WDM spectrum, $D$, is approximated in eqn. (3.1), where $\gamma_i$ is the Raman gain coefficient, $P_i$ is the input power and $\lambda_i$ is the wavelength of the $i^{th}$.

$$D = \sum_{i=2}^{w} \frac{\lambda_i}{\lambda_1} \cdot \frac{P_i \gamma_i L_\epsilon}{2A_\epsilon} \quad \text{(3.1)}$$

A sketch of the Raman gain curve versus channel spacing is shown as the dashed line in Fig. 3.2. By approximating it by a sawtooth curve, eqn. (3.2) (shown as the solid line in Fig. 3.2), we can derive the maximum attenuation condition due to SRS, eqn. (3.3), assuming the transmission window is less than 15thz.
3.1.3 Stimulated Brillouin Scattering

SBS is caused by interaction between light and acoustic waves in the fibre and results in backwards scattering of the light. However, unlike SRS it is independent of the number of multiplexed channels regardless of the separation between them. In order to ensure that a systems performance is not reduced due to SBS, the relationship described in eqn. (3.4)
must be upheld for each channel independently, where $\Delta v_L$ is the laser linewidth, $\Delta v_B$ is the optical bandwidth for SBS and $\gamma$ is the SBS gain coefficient \[4,5,8,9\]

$$P_{in} = \frac{42 A_e \cdot v_{L}}{\gamma \cdot \Delta v_B}$$ \hspace{1cm} (3.4)

The exact values for $\Delta v_B$ and $\gamma$ naturally vary depending on the properties of the silica fibres used. However, for the purpose of this report we assume $\gamma = 4 \times 10^{-11} \text{m/W}$ and $\Delta v_B = 20 \text{MHz}$ \[4\]. Eqn. (3.4) assumes a CW source. For a signal with bandwidth $B$ the threshold is increased by approximately $B/\Delta v_B$. Hence, dithering the transmitter laser frequency or employing modulation schemes which produce broad spectra has the effect of increasing the SBS threshold. For example, a 5dB improvement is possible using PSK \[10\]

### 3.1.4 Optical fibre nonlinearities

In addition to SRS and SBS a signal propagating through a fibre will be subject to further deterioration due to the Kerr non-linearities. Although these may be expressed in terms of intensity dependence of the refractive index and analysed collectively through numerical manipulation of Schrödinger’s non-linear wave equation, we shall for the purposes of our discussion in this chapter analyse & discuss them & their effects separately. Namely as Self Phase Modulation (SPM) — how the intensity variations of one channel affects its own phase, Cross Phase Modulation (CPM) — how the intensity variations of all other channels affects the phase in one channel and Four Photon Mixing (FPM) — how combinations of channels produce spurious power at new frequencies.

**Carrier Induced Phase noise**

The refractive index of silica fibres, $n$, is given as $n = n_0 + n_2 I$, where $I$ is the optical intensity. Consequently a signal propagating through a fibre will experience a change in refractive index in relation to the intensity of its pulses. This in turn, will cause the phase of the signal to vary. SPM and CPM are the effects whereby the intensity modulation of a signal modulates the phase of the signal in its own or another channel. Collectively we refer to SPM and CPM as Carrier Induced Phase (CIP) noise. The phase modulation caused by CIP broadens the signal spectrum, which in turn stimulates further pulse broadening due to dispersion (assuming we are operating in the normal dispersion region).

The variation in phase in radians, $\sigma_\phi$, due to SPM is given in eqn. (3.5), where $\sigma_\phi$ is the fluctuation in injected power in militates and $\chi_2$ is the electronic nonlinear susceptibility \[11\].
The phase variation due to CPM, $\sigma_{\phi}$, can likewise be expressed as a function of the fluctuation in input power and effective transmission length eqn. (3.6), where $\chi'$ is the sum of the electronic and Raman nonlinear susceptibilities [11].

$$\sigma_{\phi} = \frac{80 \times 10^4 \pi^2 \omega \chi' L_e \sigma_p}{n^2 c^2}$$

The electronic nonlinear susceptibility, $\chi_e = 3.5 \times 10^{-15}$ esu, is assumed to be independent of wavelength. The Raman nonlinear susceptibility, $\chi_R$, lies in the range $[-0.5 \times 10^{-15}, 0.5 \times 10^{-15}]$ esu dependent on the channel spacing. Hence, the total phase fluctuation in radians is $\sigma_{\phi}$, eqn. (3.7), where $b_i = [1.37 \times 10^{-10}, 1.82 \times 10^{-10}]$.

$$\sigma_{\phi} = \sqrt{\sigma_e^2 + \sigma_R^2} = \left(7.98 \times 10^{-11} \frac{L_e}{A_e} \sigma_p\right)^2 + \sum_{i=2}^{i=n} \left(b_i \frac{L_e}{A_e} \sigma_p\right)^2 \right)^{1/2}$$

We can write the worst case phase variation condition, eqn. (3.8), where $\sigma_{\text{per}}$ is the percentage variation in per channel input power.

$$P_{\text{in}} \leq \frac{10^3 \sigma_e A_e}{L_e \cdot \sigma_{\text{per}} \sqrt{6.36 \times 10^{-21} + 3.32 \times 10^{-20}(fW - 1)}}$$

**Four Photon Mixing**

FPM is a mechanism whereby optical carriers mix and, as a result, generate new copropagating waveforms at wavelengths different from and/or equal to those occupied by the original channels. The frequency $f_{jk}$ can be generated by mixing three waves of frequencies $f_i, f_j, f_k$ that obey the relationship $f_{jk} = f_i + f_j - f_k$.

The power transferred to frequency $f_{jk}$, $P_{jk}$, is given by eqn. (3.9), where $\eta$ is the mixing efficiency, given as eqn. (3.10); $\Delta \beta$ is the difference in propagation constant given by eqn. (3.11); $D$ is the degeneracy factor $= 1, 3, 6$, depending on whether three, two or none of $f_i, f_j, f_k$ are the same; $n$ is the refractive index of the fibre; $c$ is the speed of light in free space; $\chi_{1111}$ is the third order susceptibility; and $D_c$ is the chromatic dispersion of the fibre [12-15].
\[ P_{\text{rk}}(L) = \eta \frac{1024\pi^6}{n^2\lambda^2c^2} (D_s^2) \frac{L^2}{A_r^2} P_r(0)P_f(0)P_r(0)\exp(-\alpha L) \] (3.9)

\[ \eta = \frac{\alpha^2}{\alpha^2 + \Delta \beta^2} \left[ 1 + \frac{4\exp(-\alpha L)\sin^2(\Delta \beta L)}{[1 - \exp(-\alpha L)]^2} \right] \] (3.10)

\[ \Delta \beta = \frac{2\pi\lambda^2}{c} \Delta \lambda \Delta \phi \left\{ \frac{\lambda^2}{2c} \left[ \Delta \lambda + \Delta \phi \lambda \right] \frac{dD_c(\lambda)}{d\lambda} \right\} \] (3.11)

If we assume that the transmission window would be comprised of \( W \) optical carriers spaced multiples of the channel separation, \( \Delta f \), apart, the carriers generated due to FPM, which fall within the transmission window, will always coincide exactly with the frequencies occupied by the original channels. Channel wavelength allocation schemes which make use of non-uniform channel spacings to eliminate this problem have been presented elsewhere in the literature [16]. This technique may indeed offer a practical solution to the problem of FPM. However, it does introduce a requirement for continuously tuneable lasers and receivers in the context of some transparent network architectures.

The degradation due to FPM is strongly dependent on the mixing efficiency, \( \eta \). In our case, as we have assumed uniform channel spacing, the maximum mixing efficiency due any triplet of frequencies will be that of three optical carriers immediately adjacent to one another. If the carriers occupy the frequencies \( f_1, f_2 \) and \( f_3 \) the generated carrier \( f_{132} = f_1 - \Delta f + f_2 = f_2 \) will fall on top of \( f_2 \). Fig. 3.3 plots the mixing efficiency, \( \eta \), for this case as a function of channel separation, \( \Delta f \). In order to ensure a maximum mixing efficiency less than -20 dB, a channel separation of approximately 25 GHz and 100 GHz is required for normal and dispersion shifted fibre respectively.

![Fig. 3.3 - Mixing efficiency, \( \eta \), (dB) versus channel frequency spacing, \( \Delta f \), (GHz) for \( f_{132} = f_1 - \Delta f + f_2 = f_2 \).](image-url)
The total power $P_m$ transferred to frequency $f_m$ is the sum of all $P_{yk}$ for which $f_m = f_i + f_j - f_k$ for $f_i$, $f_j$, $f_k = f_1, f_2, f_3, ..., f_N$. The carrier at frequency $f_{\alpha 2}$ will experience the most degradation due to FPM as it is the centre channel.

We assume there is no correlation between the bit-patterns in different channels and the power transferred to other channels will appear as a mean power level, $P_N$, added to the signal in those channels. Thus, FPM causes a reduction in the extinction ratio of a signal, $r = P_N/P$, which can be equated to a power penalty, $PP$, using eqn. (3.12), [17]. A power penalty of approximately 1dB is incurred with $r=0.01$.

$$PP(dB) = 10 \log \left( \frac{1}{r^2} \right)$$  \hspace{1cm} (3.12)

Hence, for a given maximum $r$ we can write the maximum condition due to FPM, eqn. 3.13.

$$P_m L_c \leq \left( \frac{4.81 \times 10^{24} A_r^2 r}{\sum_{i=1}^{W} \sum_{j=1}^{W} \sum_{k=1}^{W} D^2 \cdot \eta_{yk}, \text{ for } i + j - k = \text{int} \left[ \frac{f_{\alpha 2}}{f_{\alpha 1}} \right]} \right)^{1/2}$$  \hspace{1cm} (3.13)
3.1.5 Optical amplifier chain

Assuming that the gain of each amplifier, $G$, along a span of $N$ amplifiers is equal to the loss due the fibre segment of length $L_A$ between each amplifier the input signal power, $P_s$, will be equal to the output signal power, $P_\text{in}$, fig. 3.5. The optical noise power, $P_{sp}$, due to the build-up in ASE noise is given by eqn. (3.14).

$$P_{sp} = h f B_w n_{sp} N(G - 1) \quad (3.14)$$

$$P_{\text{in}} = P_s + P_{sp}$$

$$P_{\text{out}} = P_{\text{in}}$$

Fig. 3.4 - The sum of $\eta_{f, 2}\cdot D^2$ as a function of $W$ and $f$ for all combinations of $f_{j, 2} = f_1 + f_j - f_k$, where $f_n, f_j, f_k = f_1, f_2, f_3, ..., f_n$, assuming a 4700GHz (EDFA) transmission window

Fig. 3.5 - Amplifier chain
If the receiver is thermal noise dominated, which is typically the case for direct detection, the principle amplifier noise will be either spontaneous-spontaneous, $n_{sp-sp}$, or signal-spontaneous, $n_{s-sp}$, beat noise. If $L$ denotes the fractional power loss between the Nth (last) amplifier and the receiver eqn. (3.15 & 3.16) describe, $n_{sp-sp}$ and $n_{s-sp}$ respectively [18, 19].

$$n_{sp-sp} = \frac{\left( I_{sp}\right)^2 B_s \left( 2B_e - B_s \right)}{B_o^2} \quad (3.15)$$

$$n_{s-sp} = \frac{4GI_s I_{sp}L^2 B_e}{B_o} \quad (3.16)$$

By introducing the requirement that the receiver noise be dominant over the principle source of amplifier noise, we derive transmission criteria in terms of a minimum Signal to Noise Ratio (SNR) requirement for dominant $n_{sp-sp}$ or $n_{s-sp}$ amplifier noise, eqn. (3.17 & 3.18), respectively, where $P_s$ is the input signal power.

$$\frac{N(G-1)}{P_s} \leq \frac{1}{h\nu \sqrt{2SNR \cdot B_eB_o}} \quad (3.17)$$

$$\frac{N(G-1)}{P_s} \leq \frac{1}{4B_eSNR \cdot n_{sp}h\nu} \quad (3.18)$$

These two criteria are in fact identical when $B_s = 8B_eSNR$. When $B_s > 8B_eSNR$ spontaneous-spontaneous beat noise, $n_{sp-sp}$, dominates and when $B_s < 8B_eSNR$ signal-spontaneous beat noise, $n_{s-sp}$, dominates. For example, if $B_s = 4700GHz$, $B_e = 2.4GHz$ and SNR = 144, i.e. $B_s < 8B_eSNR$, we use the $n_{sp-sp}$ criteria, eqn. (3.19), as this is the strictest. Alternatively, if there is a wavelength selective element, such as a wavelength filter, between the last amplifier and the receiver we may find that the $n_{s-sp}$ noise is dominant. For example, if a wavelength filter with an optical bandwidth of 50GHz is inserted immediately before the receiver, $B_s < 8B_eSNR$ and we therefore use the $n_{s-sp}$ criterion.

In the case of coherent detection, the dominant receiver noise is the local oscillator shot noise and the amplifier induced noise is the local oscillator-spontaneous beat noise, $n_{lo-sp}$, eqn. (3.19), [19].

$$n_{lo-sp} = 4I_{lo}L(G-1)n_{sp}qB_e \quad (3.19)$$

By applying the same principle as above we are able to express the transmission criteria for coherent detection in terms of receiver sensitivity, $P_{sen} = P_sL$, eqn. (3.20).
3.1.6 Analysis and results

The transmission criteria regarding the build-up of amplifier noise, for both direct and coherent detection, all fit the general format of eqn. (3.21), where \( \text{SNR}_{\text{Term}} \) is a constant dependent on receiver and system parameters.

\[
\frac{N(G-1)}{P_s} \leq \frac{1}{2n_{up}P_{\text{sen}}} \quad (3.20)
\]

Similarly, the criteria derived for SRS, SBS, CIP and FPM, eqn. (3.3, 3.4, 3.8 & 3.13) respectively, can all be expressed in the format shown in eqn. (3.22), where \( \text{NL}_{\text{Term}} \) is also a constant depending on the fibre parameters.

\[
P_{in}L_e \leq \text{NL}_{\text{Term}} \quad (3.22)
\]

The effective length, \( L_e \), of an amplifier chain is the sum of the effective lengths of each amplified segment of fibre, eqn. (3.23), [4].

\[
L_e(N) = N \left( 1 - \frac{\exp(-\alpha L_A)}{\alpha} \right) \quad (3.23)
\]

Assuming that the input power to each stage is the sum of the signal power and the accumulated ASE noise power, \( P_{in} = P_s + P_{sp} \), we can combine eqn. (3.22 & 3.23) to eliminate \( P_s \). If the gain, \( G=(G-1) \), matches the loss experienced by a signal between each amplifier, the maximum amplifier spacing, \( L_A(N) \), for a given number of amplifiers, \( N \), can be expressed as a function of \( \text{NL}_{\text{Term}}, \text{SNR}_{\text{Term}} \) and the optical bandwidth, \( B_o \), eqn. (3.24).

\[
L_A(N) \leq \frac{\ln \left[ \frac{\text{NL}_{\text{Term}} \alpha}{\ln B_o P_{\text{sp}} + \frac{1}{\text{SNR}_{\text{Term}}}} \right] + 1}{\alpha} \quad (3.24)
\]

Using eqn. (24) we investigate the limitations of a transmission system. Initially we shall consider direct detection. Table 3.1 details the relevant parameters of a hypothetical transmission system. The minimum SNR is chosen as 144, corresponding to a BER of \( 10^{-9} \). The remaining parameters are selected such that a maximum 1dB power penalty due to optical fibre nonlinearities is incurred. Fig. 3.6 & 3.7 plot the maximum transmission
distance, $N \cdot L_A(N)$, in kilometres versus the number of amplifiers, $N$, for a direct detection transmission system with 100 wavelength slots, 50GHz optical receiver bandwidth, using normal and dispersion shifted fibre respectively. As expected the transmission distance initially increases with the number of amplifiers. However as the effective transmission length, $L_e$, also increases the maximum input power per channel must be decreased in order to maintain the systems performance. This in turn requires the amplifier spacing to decrease to maintain the SNR. Hence, we expect a maximum transmission distance, corresponding to optimal combination of amplifier spacing and the total number of amplifiers. For a fixed transmission distance, the minimum number of amplifiers is given as the smallest integer no less than the equivalent value of $N$. Alternatively, the optimal amplifier spacing is the value of $L_A$, where $N \cdot L_A(N)$ exhibits a maximum.
Fig. 3.6 - Maximum transmission distance, $N \cdot L_A(N)$, in km versus the number of amplifiers, $N$, using direct detection, normal fibre, $W=100$, $B_0=50$GHz transmission.

Fig. 3.7 - Maximum transmission distance, $N \cdot L_A(N)$, in km versus the number of amplifiers, $N$, using direct detection, dispersion shifted fibre, $W=100$, $B_0=50$GHz transmission.

For normal and dispersion shifted fibre respectively, SRS and FPM cause the strictest restriction on input power. Hence, these constraints give rise to the minimum transmission...
distances. In the case of normal fibre with 50 active channels, \( f=0.5 \), and a transmission distance of, say, 500km 4 amplifiers are required. The maximum transmission distance under the SRS constraint is 1291km for \( N=32 \), with \( f=1.0 \) and 1493km for \( N=42 \), with \( f=0.5 \). Comparatively, dispersion shifted fibre requires 5 amplifiers for 500km and only offers a 1044km with \( N=29 \), \( f=0.5 \) or 981km with \( N=28 \), \( f=1.0 \). The decrease is due to the difference in effective fibre core areas and the dispersion coefficients.

These figures indicate that the optimal amplifier spacing, \( L_{\text{A}} \), is \( \sim 35.5 \)km, giving an amplifier gain \( G=7.1 \)dB. However, the total input powers, \( P_{\text{in}} \), for these examples range from 5.4dBm (DS, FPM, 0ps/km.nm, \( f=1.0 \), \( W=100 \)) to 7.3dBm (CF, SRS, \( f=0.5 \), \( W=100 \)). That is, although the amplifier spacings for the different transmission system may be the same, the greater the haul length the greater the required signal power and hence the total input power to each stage. Consequently, the saturation power of the amplifiers used must also be greater. For instance, a transmission system which aims to fully exploit the CF, SRS, \( f=0.5 \), \( W=100 \) limit requires optical amplifiers with a saturation power in excess of 7.5dBm, whereas a DS, FPM, 0ps/km.nm, \( f=50 \), \( W=100 \) limited system only requires 5.5dBm. In this respect, it may be desirable to reduce the input signal power further than the SRS or FPM limits, in order to reduce the specification on the optical amplifiers, if a particular network is known not to exceed a lesser transmission length. Higher signal powers, however, have the added advantage that the loss succeeding the last amplifier, \( L \), can be greater for a given receiver sensitivity, \( P_{\text{sen}} \). This could be of particular importance in network architectures where the last link between two terminals is a “broadcast and select” Passive Optical Network. For example, the per channel input signal powers, \( P_s \), for the two above mentioned transmission systems are -14.7dBm to -12.8dBm. That is, SRS limited transmission could address approximately 1.5 times as many far-end terminals as the same network architecture limited by the FPM criterion.

From \( L_{\text{A}}(N) \) is possible to determine \( L_{\text{c}} \) and hence the maximum per channel input signal power, \( P_{\text{max}} \), for SBS, SRS, CIP and FPM individually. The overall \( P_{\text{max}} \) is given as \( \min\{P_{\text{max, SBS}}, P_{\text{max, SRS}}, P_{\text{max, CIP}}, P_{\text{max, FPM}}\} \). Fig. 3.8 plots the maximum per channel input signal power in dBm versus the number of amplifiers, \( N \), for the direct detection, normal fibre, 100 wavelength slot transmission system. We have already seen that applying the SRS \( f=0.5 \) criterion at the transmission distance limit leads to \( P_s=-12.8 \)dBm. If we assume that the receiver sensitivity, at a desired transmission and Bit Error Rate, is -30dBm, then \( L \) must be less than 17.2dB. However, if a routed signal is passively split between 128 customers within a 5km radius after the last amplifier, \( L=22.1 \)dBm. The minimum per channel input signal power tolerable in this case would be -7.9dBm. Applying the SRS

\[^{1}\text{min\{A, B, ..., Z\} returns the minimum value from the numbers \{A, B, ..., Z\}.}\]
criterion, the maximum transmission distance is reduced to 995km + 5km = 1000km (N=11).
Comparatively the FPM, 0ps/km nm, W=100, f=1.0 limit dispersion shifted fibre is 471km (N=4).

Fig. 3.8 - Maximum per channel input power, $P_s$, in dBm versus the number of amplifiers, N, using direct detection, normal fibre, $W=100$, $B_s=50\,\text{GHz}$ transmission
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>Wavelength slots</td>
<td>100, 500</td>
</tr>
<tr>
<td>(B_o)</td>
<td>Optical receiver bandwidth</td>
<td>10 GHz (W=500), 50 GHz (W=100)</td>
</tr>
<tr>
<td>(B_e)</td>
<td>Electronic bandwidth</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>Wavelength</td>
<td>1.55 (\mu)m</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>Fibre attenuation</td>
<td>0.2 dB/km</td>
</tr>
<tr>
<td>(A_e)</td>
<td>Effective fibre core area</td>
<td>80 (\mu)m(^2) CF(^2), 50 (\mu)m(^2) DS (^3)</td>
</tr>
<tr>
<td>(D_c)</td>
<td>Chromatic dispersion</td>
<td>16 ps/km nm CF, 1 ps/km nm DS, 0 ps/km nm DS</td>
</tr>
<tr>
<td>(dD_c(\lambda)/d\lambda)</td>
<td>2nd order chromatic dispersion</td>
<td>0.080 ps/km nm(^2) CF, 0.055 ps/km nm(^2) DS [20]</td>
</tr>
<tr>
<td>D</td>
<td>Maximum attenuation</td>
<td>1 dB</td>
</tr>
<tr>
<td>(\Delta\nu_L)</td>
<td>Laser linewidth</td>
<td>20 MHz, 100 MHz</td>
</tr>
<tr>
<td>(P_{sn})</td>
<td>Receiver sensitivity</td>
<td>-30 dBm DD(^4), -53 dBm CD(^5)</td>
</tr>
<tr>
<td>(\sigma_t)</td>
<td>Maximum phase error</td>
<td>0.15 rad</td>
</tr>
<tr>
<td>(\sigma_{%P})</td>
<td>Maximum rms. input power percentage variation</td>
<td>20 %</td>
</tr>
<tr>
<td>n</td>
<td>Refractive index</td>
<td>1.52</td>
</tr>
<tr>
<td>r</td>
<td>Extinction ratio</td>
<td>0.01</td>
</tr>
<tr>
<td>(n_{sp})</td>
<td>Spontaneous noise coefficient</td>
<td>1.4</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
<td>144</td>
</tr>
</tbody>
</table>

**Table 3.1 - System parameters of hypothetical transmission system**

An increased transmission distance can be achieved by manipulating the system parameters to increase the minimum \(NLTerm\) and \(SNRTerm\) values. A simple way of increasing \(SNRTerm\) is reducing the optical bandwidth of the receiver, \(B_e\). Fig. 3.9 plots \(N\cdot L_A(N)\)

\(^2\)Conventional Fibre.

\(^3\)Dispersion Shifted fibre.

\(^4\)Direct Detection.

\(^5\)Coherent Detection.
versus N for the direct detection, normal fibre, 100 wavelength slots transmission system, where the optical bandwidth is now 10GHz rather than 50GHz. SRS $f=1.0$, followed by SRS $f=0.5$, is still the most limiting factor. The average increase in maximum transmission distance is 1.3%. Although, decreasing the optical bandwidth in this example only causes a modest improvement, it has the advantage that it comes at no additional cost in terms of increased requirement on the amplifier saturation power nor on the expense of decreased signal power. Therefore, reducing the optical bandwidth further than strictly necessary in terms of the WDM partition is a potent method for improving performance, particularly, in systems with few channels. It also has the additional advantage, that it can be applied individually and unevenly to terminals that are extraordinarily distant from the core of the network or require superior transmission performance.

Fig. 3.9 - Maximum transmission distance, $N \cdot L_A(N)$, in km versus the number of amplifiers, N, using direct detection, normal fibre, $W=100$, $B_o=10$GHz transmission

Increasing the number of wavelength slots, $W$, within the EDFA window does not only reduce the optical bandwidth occupied by each one, $B_o$, it increases the susceptibility to nonlinearities. Hence, we expect to see a compounded effect consisting of a reduction in $B_o$, giving rise to an improvement, and an increase in the number of channels, causing a deterioration. Fig. 3.10 plots $N \cdot L_A(N)$ versus N for the direct detection, normal fibre, $W=500$ transmission system. In this system SRS gives rise to, by far, the most dominant criteria. The maximum transmission distances under the SRS $f=0.5$ & $f=1.0$ limits are 657km (N=19, $P_s=-16.5$dBm) and 569km (N=16, $P_s=-17.1$dBm) respectively, i.e. more than a halving from corresponding $W=100$ system.
Common to all the direct detection systems is a strong relationship between the signal power and the maximum transmission length. We have seen that longer transmission lengths are best achieved by increasing the signal power. This, however, also means higher average power in the fibre. The alternative to direct detection is coherent detection. Coherent detection has the advantage that the receiver is generally shot, rather than thermal, noise limited. Hence, we expect a much improved receiver sensitivity, while the average input signal remains the same as the equivalent direct detection system. As in the case of minimal optical bandwidth, the advantage is that increased performance does not require an elevated specification on the optical amplifiers. In the case of a coherent system we also expect to enjoy a higher tolerance on L due the superior receiver sensitivity.

Fig. 3.10 - Maximum transmission distance, $N-L_A(N)$, in km versus the number of amplifiers, $N$, using direct detection, normal fibre, $W=500$, $B_o=10\text{GHz}$ transmission

Fig. 3.11 & 3.12 plot $N-L_A$ and $P$, versus $N$ for a normal fibre, 100 wavelength slot, coherent detection transmission system, assuming $P_{\text{sen}}=-53\text{dBm}$. The CIP noise effect is included, as random phase variation will affect the performance in, for instance, Phase Shift Keyed (PSK) systems. Indeed, with 20% residual AM and a maximum tolerable phase error of 0.15rad (corresponding to a 0.5dB power penalty [4]) CIP noise becomes nearly as limiting as SRS. The maximum transmission distance with 100 active channels is 4362km (SRS) compared to 1291km (SRS, $f=1.0$, $B_o=50\text{GHz}$) for the equivalent direct detection system. Although $P_r=-18.7\text{dBm}$ is 5.2dB lower than direct detection, $L$ can still be increased by more than 17dB. Coherent detection with 50 active channels promises transmission distances in excess of 5000km (SRS limited).
3.1.7 Conclusions

The optimisation method proposed in this paper may be used to minimise the use of amplifiers in a transparent amplified WDM transmission system. We have also shown how
it can be used to estimate maximum transmission distances and signal and total powers. We have presented results for both direct and coherent detection as well as normal and dispersion shifted fibre.

The maximum transmission distances are achieved by the transmission systems with the maximum values of the $NLTerm$ and the $SNRTerm$. Excluding inherent dispersion limitations, normal fibre generally offers the better performance as the effective core area and, hence, the $NLTerm$'s are larger than those of dispersion shifted fibre. The signal and total input powers are, however, also proportional to $NLTerm$ and the increased transmission distances, therefore, require increased transmission powers. The much reduced dispersion coefficients of dispersion shifted fibre may overcome dispersion limitations but, when used in the context of D-WDM transmission, also results in an increased susceptibility to FPM. Indeed, the results shown in the previous section indicate that FPM limits transmission distances to an extent where a low chromatic dispersion coefficient is no longer essential. One possible solution to this problem could be to intermix sections of normal fibre and fibre with a negative dispersion coefficient, thereby achieving a near zero overall dispersion coefficient while maintaining low FPM mixing efficiency.

Increasing the $SNRTerm$, rather than the $NLTerm$, is a desirable way to improve system performance, without increasing power levels. For direct detection, however, the only real option is to reduce the optical bandwidth which will generally only facilitate an improvement proportional to $\ln(\sqrt{B_o})$.

Finally, in some network architectures the tolerable attenuation after the last amplifier, $L$, is critical. When a high value for $L$ is required, it may seriously reduced the maximum transmission distance of a system. The only ways to increase $L$ is either to allow higher signal powers or improve upon the receiver power sensitivity.

Overall, the results obtained seem to suggest that optical network architectures requiring between a few tens to a hundred transparent channels per fibre are at best limited to transmission distances ranging from a few hundred to a couple of thousand kilometres, i.e. regional and (small) national networks.
3.2 References


Optical networks

In this chapter we present a brief history of optical technology leading to the ongoing debate regarding All Optical Networks (AONs). With reference to the condensed review of the available optical technology presented in chapter 2 we outline the potential for the application of optics in communication networks, discuss some of the potential impact it might have and summarise a few of the fundamental considerations regarding aspects of the network architecture and control when considering how to design (or implement) an optical network.

4.1 Background and history of optical networks

Although the concept of using optical fibre for communication is generally considered to originate from a paper by Kao and Hockham in 1966, it was not until the beginning of the 1970s that the first optical fibre with losses less than 20dB/km was reported. By the end of the decade prototype silica fibre was being installed in realistic working environments. The first systems installed were based upon multi-mode, graded index fibres operating at wavelengths between 850 and 900nm using GaAlAs semiconductor laser sources and PIN or A-PIN detectors. Single mode systems looked too difficult for real systems.
By 1980 the first graded-index multi-mode systems were installed for point to point transmission at bit rates between 8 and 140Mbit/s over repeater separations up to about 10km. However, in research it had been recognised that superior performance was possible using single mode fibres and longer wavelength (1300nm) sources.

By the middle of the 80s repeater spacing had been increased to 60km and bit rates to 650Mbit/s; the first transoceanic 1550nm system was deployed and single mode fibre had become de rigueur for inland transmission systems.

The ever falling cost of optical communication devices attracted interest in fibre based LANs and research into how to economically push the optical systems towards business and the individual customer premises, e.g. Passive Optical Networks (PONs) etc.

The development of the Erbium Doped Fibre Amplifier (EDFA) at Southampton by the end of the 80s and the possibility for optical amplification, fuelled renewed thinking about optical fibre and transmission systems and their potential as "loss-less data pipes".

The further development of ultra-fast soliton systems combined with the realisation that single mode fibre exhibits a low loss transmission window of more than 25 THz around approximately 1.5μm (or 'only' 4700GHz if limited to the EDFA passband) and the potential for portioning and accessing individual sections of the available transmission spectrum directly within the wavelength domain in the form of Wavelength Division Multiplexed (WDM) channels, redefined the premises of optical networks. It became the 'building blocks' of an on-going debate regarding the potential for optically routed overlay networks possibly leading to transparent AONs.

4.2 Application and potential for optics in communication networks

Over the past few decades optical fibre has emerged as one of the principal transmission mediums in telecommunications. A single fibre has a usable bandwidth of several Terahertz in the 1.3μm and 1.55μm transmission windows. World-wide the telecommunication companies have invested heavily in fibre, as have the TV cable and alternative access carrier companies. Currently fibre is almost universally used as a substitute for copper. However, the role of fibre in a telecommunications network is predominantly merely to facilitate point to point links between the existing electronic nodes. With the arrival of broadband services, it will be necessary for the service providing companies to update their networks, to provide both additional bandwidth and greater flexibility.

Fibre relatively easily yields the increased bandwidth required. With the present and predicted future capacity requirements, refer chapter 1, an extremely moderate exploitation of the potential fibre bandwidth is all that is necessary. In fact, once the fibre reach is
reach is extended to the individual subscribers, the operators will have access to far more bandwidth than they can probably conceive of using. The present network architectures rely on a relatively complicated control systems. Historically, this was justified as bandwidth was a precious commodity. It was therefore, preferable to construct a complicated control system, which minimised the wastage of available capacity. With the introduction of optical fibres, bandwidth is almost “free”. Hence, one progression could be to trade bandwidth for electronics to reduce the complexity of network nodes and the consequent operation & management of a network.

One possible way of achieving this could be to make the network ‘transparent’. At the extreme this is totally optical transmission from customer to customer, without intermediate stages of opto-electronic conversion and electronic signal processing. One objective of the work presented in this part of the thesis was to investigate to what extent transparent optical networks are feasible and whether they in fact offer reduced complexity compared to their opto-electronic counterparts.

With reference to the vision of a possible future broadband multi-service communication network built around the IT, media, silicon as well as telecommunication industries outlined in chapter 1 and our comparison of optics and electronics in section 2.1 - “optics versus electronics” - it seems clear that introducing and extending the way in which we use of optics in telecommunication networks may be at least one possible avenue of exploration.

Some of the consequences might be as follows: An increase in the carrying distance of optics over electronics will cause a “flattening” of the hierarchical network structure and a reduction of the number of active components such as regenerators, switches etc. Due to this reduction in the number of active network elements reliability will increase.

The introduction of optical transparency will cause the control within networks to move towards the periphery, i.e. closer to the customer premises. This in turn will tend to cause a reduction in the amount of centralised control software required, e.g. at switches. It has been estimated that a completely optical network would have a fault-rate of less than one fifth of today's telecommunication networks [7]. There will, however, be a “crossover” point in terms of reliability at some point in this network evolution: the consolidation/reduction of transmission hardware and software causes an increase in reliability but it also causes an increase in the severity of any single component outage. Consequently it is the opinion of the author that a move towards greater optical transparency within networks, i.e. fewer active network components etc. should be done with great caution, particularly in the name of ‘greater reliability’.

The reduction in the number of active network components will cause a reduction in the volume of data needed to be processed for the network management. That is, the
complexity and therefore the cost of maintaining and running a network is likely to decrease in step with any move towards greater optical transparency.

In addition, the increased variety and complexity of services expected from future networks combined with the fact that more than one quarter of all reported faults in some way may be attributed to the actions of the telco's own craft people carrying out routine maintenance, [7], make a very strong case for the replacement of manual distribution point and wiring frames with a network which may be serviced and maintained remotely via e.g. PONs. Combined with the increased hardware reliability of optical networks it has been estimated that the saving on running cost may as great as 40%, [7].

If the above figures of reduction in equipment volume, increased device reliability, the possibility of remote equipment diagnostics & configuration and the expected increase in traffic volume is correlated with the magnitude of the work force of a telco, a reduction in the required number of employees from a nominal number of 100,000 employees per 10 million customers (of a 1960's telecommunication service provider) may possibly be reduced to around 30,000 or less, [7]. It should be noted, however, that although the work force necessary to manage and maintain such a future network would be smaller, the demands upon it would be far more diverse. The author believes, therefore, that although the work force may be smaller in numbers it would need to be correspondingly more highly skilled and trained.

It is clear that as well as a "flattening" of the hierarchical logic structure of networks and a consolidation of the network transmission and switching devices there is a rapid migration of intelligence towards the customer premises in networks. Present day examples of this are desk & laptop computers linked to each other via the existing telecommunication network to provide extended working environments. Given this trend continues, it is hard to envisage any service a telco in future could uniquely place within its network. The bad news for the telcos is therefore that they in future will probably be closed out of anything but mainstream service provision, such as national television, due to their large corporate size and inherent inflexibility. The good news for the telcos is that the increase in traffic resulting from such an evolution would result in a tenable business of data transport. This will mean a shifting of the role of the national telcos from service providers to 'bit carriers'.
4.3 All Optical Network architectures

At the other end of list of considerations regarding AONs is a host of issues regarding any architecture itself (and the implementation of it, refer part I chapter 2 & 3). Optical network architectures may be classified as either of two general types:

- Broadcast and select networks
- Selectively routed networks

These are diagrammatically illustrated in Fig. 4.1 & 4.2, respectively. The broadcast and select type of network relies upon the huge capacity available within optical networks and trades it for relative simplicity in the network control by broadcasting all the transmitted information to all the destinations so that actively manipulating the transmission spectrum through e.g. switching (which in optics tends to be non-trial) is avoided. Selection of the desired information is only performed at the receiving station by simply ‘tuning-in’ to only a portion of the available information, e.g. in the time or the wavelength domain.

This type of optical network architecture is used mainly in networks topologically limited to shorter transmission distances, where the implementation is very cost sensitive, e.g. in the access network, see chapter 5, and/or where the service requirement is such that network functionality is not so critical such as e.g. IBM’s Rainbow network [5].

![Fig. 4.1 - Broadcast and select optical network](image)

In the cases where the volume of communicated information is such that even the enormous capacity of optical fibre is not enough (this is typically emphasised when combined with restrictions imposed by topology and implementation) it is necessary to manipulate and select, i.e. switch, all or proportions of the traffic. If this is done within the optical domain this may be referred to as the other type of optical network, namely a selectively routed optical network. This type of network is diagrammatically represented in Fig. 4.2. The access nodes each transmit or receive the data destined for them directly in pre-defined allocations of e.g. the wavelength and/or time spectrum. The individual
proportions of the optical transmission space are manipulated directly in the optical domain by optical cross-connects (or add/drop multiplexers). This type of optical network architecture is typically used in extended networks or where the required functionality warrants the increased technological complexity and cost, see chapter 6.

![Selectively routed optical network](image)

**Fig. 4.2 - Selectively routed optical network**

### 4.3.1 Modularity and scalability

In the author’s opinion, when considering any (all) optical network, two issues, which often seem to be either overlooked or diminished, are of key importance as they concern the fundamental assumptions upon which any telecommunication network is based. Namely, modularity and scalability. By a network’s modularity we mean to which extent it is singularly extendible. In other words, can we extend the network, i.e. add nodes, one node at a time?

Several of the optical network architectures initially proposed, such as the wavelength routed shuffle-net, Manhattan Street Network (MSN) or hyper cube were
centred around the available wave guide optical technology. These networks are not intrinsically\textsuperscript{1} modular.

This is not necessary a disadvantage but the accompanying limitations and restrictions must be acknowledged and recognised. For example, it seems nonsensical to suggest and discuss, for instance, a national UK network based around e.g. the hyper cube topology as anything other than a logical exercise. Although the 'requirement' for 64 nodes would seem to map quite closely onto e.g. the UK core network the fact that they would have to be topologically arranged so that they conform to the hyper-cube structure over distances up to 1000km (or more) would seem to be too restrictive, certainly within the context of a UK core network. That is of course not to say that within the right framework, such as e.g. a distributed multi-processor environment, this type of network might potentially be very powerful.

Associated with the question of modularity is also the issue of scalability. To what extent can the network be extended. In our example of a hyper-cube network, can we keep on adding nodes? Yes, but only at the cost of additional wavelengths assuming that larger guided wave devices could also be constructed (and only in factors of 2, see above). Hence, the hyper cube is not scalable (or modular).

Ultimately the importance of these issues depends on whether the proposed network is scalable enough for the envisaged application now or in case of future expansion.

\textbf{Scalability of network control & management}

A 'subordinate' and often even more neglected problem is the scalability of the control and management of the investigated network. For instance, if we consider a single STM-16 bit stream (2.4Gbit/s) made up from multiplexed ISDN channels it would consist of approximately 30000 circuit switched 'calls', e.g. telephone calls. If the average holding time for these is 2 minutes this corresponds to a call arrival rate of about 300 (calls) per second. Given that telecoms switches are based around a 8kHz cycle rate this implies that a switch's connection pattern should be re-programmed about 2.5 million (= 300\times8000) times per second, where each new call (within the context of a UK network) would be selected from one in 6.25\times10^{14} (= 25 million terminations\textsuperscript{2}) different possibilities. If the STM-16 channel was composed of ATM cells a switch would need to responsively change its connection pattern about 5.5 million times per second although be it from a hugely

\textsuperscript{1}The author is aware that work has been published which discusses how these types of networks may be extended, e.g. [6], but they are not inherently modular (or scalable).
smaller 'menu'. In either case, these extremely simplified 'back of the envelope' calculations\(^2\) imply control data flow and processing overheads that are non-trivial.

For example, if we are to consider communication between \(N\) nodes we find that there will be \(N \times N\) possible connections and the minimum number of control bits (in a binary switch matrix\(^3\)) required to describe the connection is \(N \log_2 N\). If the effective capacity\(^4\) of the control channel is \(C_{\text{ctrl}}\) and we assume that the network should spend at least the same proportion of time transmitting, \(T_{\text{msg}}\), Eqn. (4.1), as controlling (the transmission of) information, \(T_{\text{ctrl}}\), Eqn. (4.2), i.e. \(T_{\text{msg}} \leq T_{\text{ctrl}}\), we find that the maximum capacity of the network between each node pair, \(C_{\text{max}}\), is given by Eqn. (4.3), where \(M\) is the message length in bits.

\[
T_{\text{msg}} = \frac{M}{C} \tag{4.1}
\]

\[
T_{\text{ctrl}} = \frac{N \log_2 N}{C_{\text{ctrl}}} \tag{4.2}
\]

\[
C_{\text{max}} = \frac{M \cdot C_{\text{ctrl}}}{N \log_2 N} \tag{4.3}
\]

We plot the maximum communication capacity between any node pair, \(C_{\text{max}}\), if the traffic carried is units of 2 minutes worth of circuit switched ISDN traffic (\(C_{\text{max}}\).SW) and individually switched ATM cells (\(C_{\text{max}}\).ATM) versus the number of nodes, \(N\), for a network with an effective control channel capacity, \(C_{\text{ctrl}}\), of 1GBit/s in Fig. 4.3 below.

For example, in the case of a network of proportions equivalent to the UK, i.e. \(N = 25 \times 10^6\), the effective maximum communication capacity between any node pair would be in the order of 12MBit/s or 600Bit/s for SW or ATM traffic respectively, assuming an effective control channel capacity of 1GBit/s.

\(^2\)For example, in the case of our ISDN traffic example, the author is aware that the number of connection patterns in practise is much less than \(6.25 \times 10^{14}\) due to the 'call stripping' and resource aggregation [1] which place in any real network.

\(^3\)By a binary switch matrix we mean a switch fabric composed of (or represented by) 2\(\times\)2 switches. Each of the 2\(\times\)2 (pass-through or cross-over) switches require a binary control signal, i.e. one control bit.

\(^4\)By effective capacity we mean the actual end to end capacity including all transmission & processing delays etc.
Although the above ‘analysis’ is rather stylised it does seem to give some clear indications regarding any potential application of photonics in switching. Namely, that transparent optical networks due to their relative non-responsive control would seem to be better suited for circuit switched or trunk-type traffic rather than distributed processing networks based around packet switched traffic; particularly when communicating between a large number of entities, i.e. nodes.

It would further more seem to suggest that optical networks would lend themselves towards aggregating or bundling of the information transmitted into larger units which then in turn may be switched so that the number of processed units is reduced. This would though mean that the utilisation of the available information space would not be maximised. However, we might not find this to be critical, particularly in view of the huge information space potentially available in optical fibre. In other words, the design of transparent AONs might be best served by “wasting” the huge information space available in such a way that the processing load is reduced. One possibility, which has been used in several AONs, e.g. LAMBDANET and others [2-4], might be to partition the available information space and designate each portion, e.g. wavelength channel, to a certain (transmitting) node. For a node to receive information from another node it is then only necessary to transmit the information in question on the allocated wavelength and the receiving node to tune to that wavelength [2-4]. Thus the information space has become address space. See chapter 5 & 6 for further development of this concept with application to the access and national network, respectively.
4.4 Summary and conclusions

In this chapter we have discussed and analysed a few of the fundamental issues associated with photonic networks and AON architectures. In particular we have discussed the relative merits of the two types of optical network; namely, broadcast & select and selectively routed optical networks.

Although substantial device and systems work still remains it seems clear that photonic technology is rapid reaching a level of maturity where it may be applied within telecommunication networks not only as direct equivalents of the current electronic components but also in a manner to enhance the network’s functionality. Using wavelength as an additional dimension in (optical) networks has been shown to be feasible and may prove to be very profitable.

In summary, it seems clear to the author that one may apply the use of photonic technology according to one of several principles:-

- Transmission
- Back-plane or interconnects, e.g. switches etc.
- Closed local networks, self routing networks or ‘Super LANs’
- The Utopian ‘Optical Ether’
- A new type of network based around the available photonic technology and the strengths of optics — with no direct (opto)-electronic equivalent

Optics is already the transmission technology of choice for many communication systems. This is due to its low cost and high reliability. It has also been applied as interconnective back-planes both between electronic sub-systems and in reconfiguring systems such as telecommunication switches. Although the latter application has been shown to be feasible, it is the author’s opinion that it may not be making the best use of optics’ strength. As also discussed in section 2.1 - “optics versus electronics” - due to the non-interaction of optical signals (waves), photonic processing systems require at the very least electronic control. Consequently, so called ‘optical’ devices which rely on electronic control must be developed and used with caution as a comparable exclusively electronic device may ultimately be more versatile and/or cheaper.

Indeed it seems clear to the author that this in turn leads to the conclusion that optics is best put to use in transmission systems unless a new type of network can be developed that makes use of the advantages of photonic technology and minimises the functions that optics is not so good at, i.e. processing. Such a network would probably have no direct electronic equivalent.
The above statement are very closely related to a networks controllability which we (superficially) investigated in the final section of this chapter. It was shown that even if equivalent optical technology was available, how it should be controlled is a non-trivial issue.

In the remaining two chapters in this section, chapter 5 & 6 we discuss the potential and relative merits of optical network systems and architectures with reference to the access network and a hypothetical national transparent optical network, respectively. We do briefly outline the control functions necessary although we do not discuss any specific means of implementation. In part III of the thesis we propose a generic multi-service broadband network management & control scheme. Although we analyse that in the broadest terms it certainly is also applicable to optical networks. Especially the analysis of access or individual network resource control.
4.5 Bibliography


4.6 References


Chapter 5

This chapter is based upon a consultancy study undertaken for GPT LTD., “Potential for Wavelength Division Multiplexing in Local Broadband networks” in 1994 following a meeting between BT Labs, Siemens, GPT, GMMT & UCL, Wednesday 18th May 1994 at the Heathrow Excelsior hotel under the Optical Switching in Networks Collaboration. It later resulted in further collaboration between the Electronic and Electrical Engineering Department, UCL and GPT LTD, under a EPSRC (CASE) PhD award.

Potential for Wavelength Division Multiplexing in broadband access networks

5.1 General statement of the local access requirement

Popular discussion on broadband networks today seems to centre on two separate and distinct types. Most people would think first of CATV, operating as a distributive network with a future "switched broadband" network superseding it at some indeterminate time in the future. The distinction between these is often presented as absolute, one or the other but not both and the former not upgradeable to the latter. In the discussion below, we examine very briefly some of the underlying assumptions inherent in this classification.

In the most general terms, the design problem is presentable in the form shown below in Fig. 5.1, in which the names have been chosen to deliberately break some of the standard knee-jerk connotations. The problem has initially been presented in this form to completely avoid the need to make any statements about the network format, multiplexing employed etc.
At the customer end, we have recipients who wish to benefit from their connection to a network in various ways. The network provides a means of delivering services, by metal or optical cable, radio or some combination of these. The network connects the recipients to some location where access to services exists. By way of illustration, let us consider the provision of entertainment video and telephony to provide three different cases for dimensioning the network.

**Telephony** - In this case, the menu of possibilities includes all the telephone users in the world and is thus vastly greater than the number of recipients likely to be connected to any particular local access network. The conclusion is clearly that the menu selection process must be done close to the menu itself; it makes no sense to connect all possible users through to every user and have the selection done at the recipients end. Indeed, in this case, the selection is far more distributed than the diagram illustrates. However, since the expected utilisation of any one connection is low and the probability is that only a small number of users among those connected to the local access network will actually be seeking service simultaneously, there is scope for providing some selection close to the recipient (e.g. concentration) to minimise the bandwidth demands made on the delivery medium.

**CATV** - Cable TV as currently practised is characterised by a very finite menu of offerings, in the range 10-100 channels, over a single local access network that might well serve 100-1000 recipients. Utilisation of the service is likely to be high at peak viewing hours and each user may wish to access more than one menu offering at a time. As a result, it makes good sense to consider placing the service selection at the recipient end and piping all services to all recipients.

**Video-on-demand** - In this case, the situation is much less clear. The size of the menu offering is unknown but presumably, unless it is very significantly larger than that offered on a network-TV/CATV service, it will have no attraction. Hence one might consider a target of 1000-10,000 offerings on any one day? This places the service in a region on the
diagram, Fig. 5.2, whereby the number of menu items offered is comparable to or greater than the number of recipients. Assuming that these numbers are correct, then in the terms of this discussion, it has little effect on the dimensioning of the network where the selection is carried out, unless the menu offered is very large. However, if one believes that this service might prove attractive, then it would be natural to allow it to grow to provide an ever greater number of menu offerings and that would very rapidly favour selection at the menu end rather than the recipients'. Moreover, considerations of ease of maintenance would also lead to the same conclusion, since more of the capital equipment is placed at the service-providers premises.

Note also another aspect of this type of service. It is clear the video-on-demand service assumes the existence of a very large data storage capacity in electronically accessible form. Perhaps 1000 VHS players might be feasible in the local delivery centre but a 100,000 seems clearly not to be. Hence, it seems clear that thinking must be directed to more all-electronic and less electro-mechanical storage media and this points directly towards a need for large amounts of digital data compression, hence opening up the use of digital storage technology. Typically it appears that 2Mbit/s compressed video is adequate for entertainment purposes using the PAL or NTSC display standards although 4 MBit/s might be preferred. This then injects another question into the generalised debate as offered here. Where should the digital decompression and D-to-A conversion be located, at source or destination?

Transmitting the compressed signal clearly reduces the demands upon the delivery medium but implies decompression and A-D conversion at the recipients. Since the latter has to be done somewhere, the number of such modules required does not depend upon their location unless the expected peak-utilisation of the Video-on-Demand is low, which seems unlikely. Assuming network channels are provided over the same bearer and in the same transmission format, then this seems unlikely to be the case. Hence, their placement is more likely to be governed by considerations of housing and maintenance than initial capital cost.

Developing a wideband network around the delivery of compressed video seems to point to an important convergence between the development of 2Mbit/s switched broadband circuits for business conference, high quality conference video-phone etc. and the provision of uni-directional entertainment services building on the CD-I format or similar standards as well as CD-quality uncompressed audio, all of which require that order of transmission capacity.
5.1.1 Network Implications

It follows from the above discussion that some services lend themselves to implementation over a switched network whilst others fit better onto a distribution network. However, from the point of view of the bearer medium, the distinction is rather small and the major implications fall on the interfaces at which the services offered are multiplexed onto the bearer and demultiplexed of it. The situation is summarised in Fig. 5.2 below.

![Diagram showing the relationship between switched and distributive services for local access networks.]

Fig. 5.2 - Relationship between switched and distributive services for local access networks.

Notice that a CATV network can in principle always be converted into a switched service network for broadband distribution by restricting the number of customers served over any one bearer so that the number of available (transmitted) channels equals or exceeds the number of customers. Then by changing the way services are multiplexed and demultiplexed on-to and off-of the transmission medium, allocating transmission channels to customers rather to service offerings, or in terms of our Fig. 5.1 moving the selection from the customer to the service provider end of the network.

5.1.2 Other related technical changes in the surrounding infra-structure

*TV display technology* - There is a noticeable convergence starting to appear between the display technology used for Personal Computers and that for TV. On the one hand, we have the rapid development of dramatically improved computer graphics, extending to real-time video capability, for use in multi-media PCs such as the Apple or IBM ranges. This is increasingly linked to the very large digital storage abilities of CD-ROM and the rapidly decreasing cost of silicon RAM frame-stores. Its impact is seen in the wide spread growth of new video games as well as more serious education and other information systems products. In entertainment, we have the emergence of technologies such as CD-I
which again combine compressed video storage but utilise standard PAL or HD-TV display. The growth of digital signal processing capability with digital frame-storage in advanced domestic TV receivers is further pointing to convergence into an all-digital processing system.

Office System Convergence - Compatibility between computer based systems in the office has been a notorious problem area ever since they started to become widespread and needed to converse! The UNIX operating system started a process whereby major applications software packages could be ported to many different machines. More recently, we have seen the growth of applications packages for PCs, such as MS-Word, which, whilst individually tailored to a given PC family, could at least accept files produced by their brothers running on other families. The recent announcement of the Power-PC range by Apple, to be followed by an IBM Power PC range, heralds a further convergence since it is claimed that these machine can run both Apple Finder and MS-Windows software packages simultaneously and patch results from one directly into the other. Then the introduction of Windows-NT by Microsoft appears to usher in an era when the application software becomes hardware independent on most machines and finally, the recent announcement of Microsoft-at-Work appears to open the way for a much broader level of integration of office type equipment, embracing not just easy interchange between PC families but also the integration of telephones, FAXs, modems, personal assistants, OCRs, printers and copiers and of course, their interaction remotely. It is tempting to believe that the "electronic office" may actually be arriving!

The common theme across all the applications is digital technology, from storage through transmission and processing to display and with the emergence of operating systems and standards designed to enable interworking. Furthermore, the barriers between the previously completely separate business sectors of information systems and communications are fast disappearing with the interesting emergence for transmission purposes of the ISDN 64 and 128 kBit/s levels and the T1 or CEPT1 rates of 1.5 and 2 MBit/s as prime carriers for many information services. The US and BT approach using ADSL over a local copper network is clearly targeted at exactly this scenario. What is unclear is the extent to which such a technology can be deployed; is it to 100% of the existing Local Access Network or 50% or less, and if not 100%, then how does one identify which parts? It seems likely that even a compressed video offering will require fibre for a major part of its travel towards the customer and that at best ADSL will provide a vehicle for launching unidirectional services at low penetration into the user community.

The digital storage of video is driven by the existence and growing influence of the Motion Picture Experts Group (MPEG) and the standards they are developing (Table 5.1
below). These already pervade both home and office and it seems difficult to avoid the conclusion that they will also be influential in the development of a switched local broadband network. The H.261 standard forms the basis of the latest video-phone, draws upon the MPEG standards and allows heavily compressed moving pictures to be carried over an ISDN line. Furthermore, multi-media PCs operating to H.261 also allow the machines to view and interact with other application windows displayed on their screen in addition to the video phone window, a development of great potential for teleworkers.

<table>
<thead>
<tr>
<th>Title</th>
<th>Subject &amp; typical bit rate.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPEG-1</td>
<td>Digital video into CD-ROM format up to 1.5 Mbit/s</td>
</tr>
<tr>
<td>MPEG-2</td>
<td>Higher rates, from 2 up to 20Mbit/s also embraces HDTV</td>
</tr>
<tr>
<td>MPEG-4</td>
<td>Very low bit rate audio and image data</td>
</tr>
<tr>
<td>H261</td>
<td>Standard for video and multimedia phone</td>
</tr>
</tbody>
</table>

Table 5.1 - Compressed digital video rates

These parallel developments in consumer and office electronic technology seem to point very strongly to the widespread deployment of MPEG standard video decoders in both markets, greatly enhancing the case for developing broadband access networks around digital rather than analogue standards.

The developments in compression technology point to another change in desired traffic characteristics. Telephony is seen as a steady information flow whereas a heavily compressed video signal generates little traffic when the picture is changing slowly but generates bursts of data when the picture changes substantially. The variation in bit rate can only be smoothed out using buffer memory which introduces delay. This is of little consequence for entertainment services but may critical in video phones and other interactive video services.

5.1.3 Conclusions on service provision.

It seems universally agreed that for the private subscriber, there are two revenue earning services today. The first is switched telephony with limited provision of other services over the same channel, perhaps evolving to ISDN provision over the coming decade in response to a need for videophone or some other heavily compressed "wideband" service. The second is the distributive CATV type service, currently based upon service selection at the customer terminal.

Looking to the future, we can imagine that video-on-demand might become attractive and that this would generate a market for a switched inward broadband channel,
at a data rate in the range 2-20 MBit/s. Whether this would be carried in analogue or
digital format is an open question. Favouring digital are the several trends noted above:-

- the likely necessity of digital compressed storage
- the rapidly falling cost of decompression hardware
- the deployment of compressed digital in consumer electronics (e.g. PCs, CD-ROM & CD-I)
- the increasing amounts of digital picture processing in advanced TV receivers

Favouring analogue distribution is the status-quo, existing TV receivers, existing CATV
equipment etc.

The requirement for a wideband bi-directional (2-20Mbit/s) channel) seems unclear
since only an advanced conference-quality video-phone seems to require this provision and
unless the price is very competitive, this will have no attraction to the private consumer in
the next decade or two.

In the business area, there appears to be a much stronger case for wideband
interactive links and channels in the 2-20Mbit/s range are already used. Penetration will be
dependent upon the service (customer premises equipment + software) offered and its
price.

**NOTE:** We are aware that cost effectiveness is an issue of critical importance and this
work has presented no discussion of it. However, we have no reliable data upon which to
base such a discussion and therefore have not attempted to do so.

## 5.2 Roles for WDM in fibre networks

The use of wavelength as a additional dimension in network planning seems to have been
considered in a variety of different contexts :-

**Greater Bandwidth** - By adding additional wavelength channels over a fibre link,
additional total transmission bandwidth is made available. In the local-network context,
this seems the least persuasive reason for using it, particularly when a dedicated fibre is
used to each customer since a large bandwidth is already available by straight electrical
time or frequency multiplexing relative to the service demands of the single customer.

**Routing** - Because discrete wavelength channels can be readily split off a wavelength
multiplex into separate fibres, WDM technology offers the possibility of simple fixed or
slowly changing routing of channels within an all-optical network. We envisage that this
might provide a cost effective way of delivering into a B-PON network different services emanating from different physically located sources such as the telephony switch, CATV-Satellite Head-End and the Video-on-Demand Server. It could also be used to deliver broadcast services by power and wavelength division into a switched fibre network at the switch point.

The same approach could be extended to routing at the delivery end of a PON by using wavelengths to address sub-communities within the overall community served by the PON, i.e. by replacing the passive splitters normally proposed for PONs with Wavelength Multiplexers. Such an approach might allow the provision of a set of business related services on one wavelength or wavelength group with private services on another group of wavelength, splitting on to two dedicated networks at the entrances to the residential development and the business park. It might also provide a means for upgrading a distributive CATV network into a virtual fully-switched-star network by defining much smaller service communities within the total community served on the whole PON.

Service Growth - Wavelength multiplexing also offers an elegant means of grafting on to an existing installation additional services targeted at selected customers. This is the TPON-BPON upgrade scenario. For example, wavelength multiplexing offers the possibility of overlaying mobile-radio-access channels on an existing fibre network for fixed terminal access. A further attraction of this approach is that it results in a capital expenditure that largely follows service-take-up since the more sophisticated terminal equipment is not installed unless the customer takes the service it provides.

Mixed Format Services - The use of separate wavelength channels offers an elegant way of offering over a single network, services of varying types carried using different electrical multiplexing techniques. Thus, broadcast CATV might be carried using Microwave Sub-Carrier multiplexing on one wavelength, telephony and data using the TPON digital bi-directional multiplexing using another carrier with Video-on-Demand on a wideband digital multiplex on a third wavelength.

Bi-directional Wideband Service - Use of different wavelengths for downstream and upstream services over fibre in the local loop offers some advantages. The problems of near-end-crosstalk can be greatly reduced, thus making the system much less susceptible to bad joints or connectors, and hence more robust. In addition, for very long local loops, there may be some marginal advantage in exploiting the lower attenuation at 1500nm to support either a wider bandwidth service or more power efficient transmission.
Advanced routing in an All-Optical Network - The possibility seems clearly to be emerging for a much more sophisticated routing capability exploiting the combination of time and wavelength multiplexing that is so clearly available and becoming accessible in an optical network. We have shown (see Appendix A) that in principle it can allow nationwide addressing in its most ambitious implementation. The ideas are developed further in the section discussing the future potential for WDM. We see little likelihood of deploying this technology in the local network in the near future since it presupposes the existence of sophisticated sources, receivers and WDM switches which are unlikely to be cheap until low and medium volume market needs have been met and production development costs written off. However, there seems nothing intrinsically expensive about the proposals so that, given a network requirement, the technology could be taken seriously. The case for its deployment seems to rest upon the desirability of devolving network control to the network periphery, opening up access to greater bandwidth to the customer and offloading some of the massive complexity now being proposed for electronic ATM switches into the optical domain.

General - In all cases other than those concerned specifically with routing, the major attraction of WDM seems to be the flexibility it offers to upgrade a network during its lifetime without having to completely scrap much of the existing installation. By contrast, an all-digital system of the B-ISDN type seems to presuppose that the complete system is fully specified at the time of installation and that later changes are thus likely to involve extensive changes to the multiplex and de-multiplex equipment.

5.3 Optical local access network architectures
In this section we investigate the various aspects of optical local access network architectures: Classes of optical subscriber loop systems, network topology, transmission multiplexing techniques and access methods. We present a short summary of the different techniques used in each of the areas and compare the implications some solutions in one area may have on other areas. Finally, we briefly discuss five actual local access network architectures.

5.3.1 Classes of optical subscriber loop systems
The majority of proposed subscriber loop systems can be classified into one of three categories: Fibre To The Curb (FTTC), Fibre To The Office or building (FTTO) and Fibre
To The Home (FTTH). Fig. 5.3 illustrates the different systems. FTTC and FTTO carry multiplexed data from a Central Office (CO) to a Remote Terminal (RT) over optical fibre, where it is distributed between the individual users. The drop from the RT is currently envisaged as being by means of metallic cables, either twisted copper pair or coaxial cable, but could also include optical fibre or a combination of the three. FTTO systems transmit multiplexed data over optical fibre between a CO and a RT, where the latter is housed in a building with a high concentration of users, e.g., an office or a block of flats. As in the case of FTTC, the users can be connected to the RT via fibre, copper wire and/or optical fibre. FTTH, as the name indicates, implies that individual users are connected directly to a CO using only optical fibres.

Recent experiments (ADSL) indicate that it may be possible to supply moderate data rates (1-4MBit/s) suitable for short and medium term broadband services [1] over existing copper wires provided that the feeder point (RT) is relatively close (<1km) to the subscriber. If this is found to be the case, FTTC and FTTO offer a potentially very attractive copper to fibre upgrade [2]. FTTC and FTTO avoid replacing a substantial and expensive part of the existing copper network while sharing the cost of the optical equipment required between many users. Thus, they offer a gradual and economic [3] introduction method for broadband services, while minimising complex engineering problems such as remote power feeding [4] but without inhibiting full FTTH deployment in future. We note, however, that as ADSL is limited to downstream transmission (due to the near end crosstalk) it is probably most suited for distributive services rather than (high bit rate) interactive services.
5.3.2 Network topology

Optical fibre has several advantages over conventional metal cables. It has low loss and a huge bandwidth (several terahertz in the 1.3μm and 1.55μm transmission windows), allowing high speed transmission with a high degree of electronic multiplexing, as well as WDM in the optical domain. Conversely, optics has very poor processing capabilities when compared to digital electronics. Based on these observations many different access network topologies have been proposed. Fig. 5.4 summarises the most common network topology primitives.

![Network Topologies Diagram]

The single star is a direct analogue of most contemporary copper networks. It has direct dedicated links between the CO and each fibre termination, which can either be at the customer premises, i.e. FTTH, or a RT, i.e. FTTC or FTTO. It is very simple in design and is the least restrictive in terms of upgradeability, transmission capacity and transmission length. The large amounts of fibre involved, however, make it especially sensitive to civil engineering costs and potentially very expensive to repair in case of a cable cut, as many
fibres may use the same ducts. The need for dedicated lasers and receivers means that cost sharing at the head-end, i.e. the CO, is restricted but may allow a lower component specification.

The passive multiple star shares the head-end and fibre resources by sharing it between several users. Each fibre emanating from the CO carries a multiplexed signal which is distributed between the users by passive optical splitting of the signal at strategically located points. Each opto-electronic termination is thus required to demultiplex the signal and select out only the information relevant to itself. The advantage of the multiple star, compared to the single star, is that it reduces the CO and fibre laying/repair costs by spreading them over many users. Its disadvantages are reduced transmission length, additional cost and reduced reliability due to the passive splitters and a need for demultiplexing at the user.

The active multiple star is similar to the passive multiple star, as it aims to reduce cost by sharing resources. By introducing an active RT, which actively multiplexes and demultiplexes the incoming signals, the need for multiplexing in each user terminal vanishes.

Bus and loop arrangements use either asymmetric optical taps or active add/drop multiplexes at each user. In the case of passive taps they assume essentially the same topology as passive stars, where the only difference is the physical layout of the fibre. Active add/drop multiplexes at each terminal gives rise to a high component per user cost but are linear and singularly extendible.

Rings are the same as buses, except they have no starting or end point. The CO acts as one of the nodes in the ring. They offer high performance and reliability but also suffer from high per user costs due to the high complexity and intelligence required at each add/drop point. Finally, they are badly suited to the physical layout of most urban and residential areas.

The topology chosen for a given network architecture depends on many factors. A more detailed presentation and discussion of the primitives is available elsewhere in the literature [5]. It is, however, generally accepted that all the different topologies have various advantages and that two or more different network topologies may be needed to cover all subscriber loop scenarios [6]. Consequently, a network operator may, indeed, decide to deploy several network architectures. If this is the case, it is desirable that there be as large an overlap in terminal equipment as possible, even though the topologies may be different.
5.3.3 Transmission multiplexing techniques

The access network requires bi-directional communication. We shall consider six different directional multiplexing schemes, namely Space Division Multiplexing (SDM), Directional Division Multiplexing (DDM), WDM, Time Compression Multiplexing (TCM), Code Division Multiplexing (CDM) and Subcarrier Multiplexing (SCM).

In SDM a single fibre is used for services travelling to a single destination. Thus, in a single star network, Fig. 5.4, two fibres per customer are needed. It requires simple terminal equipment, both optical and electronic, but large amounts of fibre.

Since to a large extent, two waves travelling in opposite directions on the same fibre do not interact, bi-directional transmission can also be used and this is DDM. In DDM the number of fibres is reduced by a factor of 2, by transmitting in both directions on each fibre. This type of transmission requires careful design of the optical components in order to minimise reflections within the fibre, particularly near the ends where the disparity in the inward and outward signal levels is the greatest, i.e. NEXT. Unfortunately, this is where one is most likely to place connectors. In these cases bi-directional DDM transmission differs from SDM in using a single instead of two fibres per customer.

In WDM the optical spectrum is divided into wavelength regions, each accommodating a data stream. The individual channels are received by filtering out all but the desired wavelength region. This calls for fairly complicated optics but has the benefit of reducing the electronic bandwidth to the lowest necessary for the service provided.

TCM again uses a single wavelength to carry both the information streams and operates bi-directionally on a single fibre as did DDM. The data from the two channels is sent sequentially in short bursts, where the line rate is proportional to the total information rate of the channels in the up and down stream directions combined. To achieve bi-directional transmission over a single fibre, the transmission at any given time takes place in one direction only but alternates in time. This is often referred to as "ping-pong" transmission. The disadvantage of TCM "ping-pong" transmission is that it requires a high level of timing precession to maintain efficiency and buffer memory at each end to smooth the service. Moreover, "ping-pong" systems also impose restrictions on the transmission format, i.e. the line rate to information rate, in relation to the total transmission distance between the CO and the users or the RT [7]. The longer the transmission distance the longer the burst duration necessary, to maintain efficiency. In other words, a given "ping-pong" architecture is either increasingly inefficient over the longer transmission distances, or else it needs to be self optimising. A self optimising architecture poses many complex engineering problems and implies a non-standard transmission format, with whatever consequences that may have on the customer terminal equipment.
In CDM systems the individual data channels are modulated by different codes with little or no correlation before they are transmitted. Once detected, the individual channels may be recovered by multiplying the received data stream by the appropriate code-words. The disadvantages of CDM systems is that the coding process results in higher bit rates than that of the two data streams combined.

In SCM the different channels modulate distinct microwave subcarriers which can then be modulated onto a single optical carrier. The effect of crosstalk due to reflected light may then be eliminated by electronically filtering the demodulated signal. Although the data channels may be digital, SCM is essentially analogue. The required bandwidth is the sum of the bandwidth occupied by each channel, including guardbands.

<table>
<thead>
<tr>
<th></th>
<th>SDM</th>
<th>DDM</th>
<th>WDM</th>
<th>TCM</th>
<th>CDM</th>
<th>SCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibres</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Line rate</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>&gt;2-B</td>
<td>&gt;&gt;2-B</td>
<td>analogue</td>
</tr>
<tr>
<td>Wavelengths</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Optics</td>
<td>simple</td>
<td>complex (reflection)</td>
<td>complex (filtering)</td>
<td>simple</td>
<td>simple</td>
<td>simple</td>
</tr>
<tr>
<td>Electronics</td>
<td>simple</td>
<td>simple</td>
<td>simple</td>
<td>complex</td>
<td>complex</td>
<td>complex</td>
</tr>
</tbody>
</table>

Table 5.2 - Comparison between directional multiplexing schemes

One much debated topic is the scope for designing transparent optical networks. Predominantly, this is a concept explored within larger interoffice, regional and national networks [8, 9, 10] but it remains a very persuasive argument within the access network also. The idea of a transparent optical network is to make use of the “transparent” properties of optics to design a network which to as large an extent as possible is oblivious to the format of the services it provides and thus, “future proof” the network. In the case of the subscriber loop, for example, the single network may provide services in both analogue and digital format. SDM, DDM, WDM and SCM all grant this level of transparency.

DDM, WDM, TCM, CDM and SCM all reduce the fibre count by a factor of two compared to SDM. Hence, DDM, WDM and SCM are the only multiplexing schemes which offer both minimal fibre count and network transparency.

The case for WDM for bi-directional transmission is attractive due to its simplicity and flexibility. Only single wavelength transmitters and receivers are required as every access point transmits and receives no more than one wavelength. The complexity of
terminal electronics is only that required by the services provided and is not affected unduly by the network size or architecture. Therefore, directional WDM is robust to topology changes, such as an upgrade from FTTC to FTTH, and the introduction of additional services. The increased complexity of the optics is also minimal. As every receiver sees only one wavelength channel, filtering based on purely passive devices is adequate. Moreover, the optical complexity appears at a single point only, namely the receiver, as opposed to every splice and connector for DDM.

In terms of network topology SDM, WDM, CDM and SCM as means of bi-directional transmission have no major advantages or disadvantages compared to one another. In addition to the limitations of TCM already discussed, DDM also gives rise to additional considerations. Since DDM is sensitive to reflections due to fibre splices and connectors, it may be found less suitable for architectures where a large number of connectors is an inherent part of the network topology, such as multiple stars networks with a large number of splits, bus and ring networks.

### 5.3.4 Access methods

Sharing resources requires multiplexing. Even if only one service is provided to each customer at any given time, all network topologies in Fig. 5.4, apart from the single star, imply multiplexing in some form or another. A single star also requires multiplexing if it is to provide a more than a single service. The multiplexing in these cases is needed to establish access. Thus, to distinguish them from the directional multiplexing techniques discussed in the previous section we shall refer to them as multiple access methods.

The same multiplexing techniques used to establish bi-directional transmission can be used to facilitate multiple access for unidirectional transmission. The techniques discussed here are Space Division Multiple Access (SDMA), Wavelength Division Multiple Access (WDMA), Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) and Sub Carrier Multiple Access (SCMA).

WDMA, SCMA and CDMA all use essentially the same transmission principles as their directional multiplexing counter parts, but in a single direction. In WDMA systems, for example, the information delivered by the network is partitioned in the wavelength domain. The optical network units access a desired subset of information by selecting to receive the corresponding wavelength region only. The individual wavelength channels can be used to address single or groups of customers, provide different services or simply extend the capacity of the network.

Strictly speaking this also applies to SCMA and CDMA. However, as they rely on processing in the electronic domain to demultiplex the channels, they do not lend themselves well to routing of signals. Routing a single channel involves detecting the
optical signal, demodulating the electronic signal, remodulating it and retransmitting it. We therefore suggest that SCMA and CDMA are best suited for point to point delivery of multiple services. This effectively excludes CDMA and SCMA for use in network topologies which rely on active switching such as the active multiple star and ring or bus networks with active add/drop multiplexers. In addition, CDMA is only suitable for digital services and here the overall line bit rate increases super-linearly with the number of multiplexed channels.

In the subscriber loop SDMA amounts to installing separate fibre networks for each service or data stream provided to a customer. Although this is certainly a possibility, it is clearly a huge waste of the fibre resource and will probably only find application in the subscriber loop feeder to overcome power budget problems, if at all.

TDMA is analogous to TCM in that it partitions the available communication space with respect to time. The successive time slots carry packets or cells which contain data to each of the customers. Like TCM, TDMA is unsuitable for analogue services. Table 5.3 summarises the properties of these access techniques.

<table>
<thead>
<tr>
<th>SDMA</th>
<th>WDMA</th>
<th>TDMA</th>
<th>CDMA</th>
<th>SCMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibres/cables</td>
<td>W</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Line rate</td>
<td>B</td>
<td>B</td>
<td>W-B</td>
<td>&gt;&gt;W-B</td>
</tr>
<tr>
<td>Wavelengths</td>
<td>1</td>
<td>W</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Optics</td>
<td>simple</td>
<td>complex</td>
<td>simple</td>
<td>simple</td>
</tr>
<tr>
<td>Electronics</td>
<td>simple</td>
<td>complex</td>
<td>complex</td>
<td>complex</td>
</tr>
</tbody>
</table>

Note: W - no of channels, B - bit rate of a channel.

* The required bandwidth is W times the bandwidth of a single channel.

Table 5.3- Comparison of multiple access multiplexing techniques.

Although the multiplexing techniques are essentially the same for directional transmission and multiple access, the implications for the network architecture are quite different. As the multiplexing is no longer a question of just point to point transmission where and how the information channels are accessed in relation to the network topology and the services carried must be considered when developing a network architecture. As has already been demonstrated, SCMA and CDMA are unsuitable as anything more than vehicles for multiple services or increasing capacity. Table 5.4 summarises the implications of WDMA, TDMA, CDMA and SCMA.
WDMA and TDMA may both be performed optically and therefore they do not impede the transparency of any of the optical network topology primitives in Fig. 5.4 any further than is inherent to them. (The transparency of WDMA and TDMA are limited by optical bandwidth and burst duration respectively.) Active star, active bus and active ring networks are therefore best suited to these methods of access.

TDMA relies on the optical network interface to extract and insert information into a high bit rate data stream. This requires electronic buffer memory and timing precision. In multiple star networks timing information is not directly available to the users as they are removed from the multiplexing point, i.e. a RT or a passive splitter. A ranging mechanism [12] and high timing precision is therefore necessary to avoid collisions of upstream traffic from different users. In bus and ring networks a ranging protocol is not necessary as timing information is available directly at all network interface points. Given that the difficulty of implementing a ranging protocol is proportional to the overall line rate, and the overall line rate is proportional to the number of channels/users, TDMA star networks may only be suitable for carrying low to medium bit rate services to a limited number of users. It should be noted that high bit rate add/drop multiplexing, even in ring and bus networks, is non-trivial. WDMA does not suffer from this problem although contention must be avoided within the wavelength domain. This may require wavelength locking of filters and lasers.

The one feature TDMA has over all the other multiplexing techniques is that a single receiver sees all incoming information whether it receives it or not. Comparatively, a WDMA receiver will only see the signal carried on one of its wavelength channels at any one time. Consequently it may miss information, where a pure TDMA terminal would not (provided the arrival rate of cells does not exceed the size of the buffer memory). This

<table>
<thead>
<tr>
<th>Access domain</th>
<th>WDMA</th>
<th>TDMA</th>
<th>CDMA</th>
<th>SCMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Services</td>
<td>Analogue &amp; Digital</td>
<td>Digital</td>
<td>Digital &amp; Digital</td>
<td></td>
</tr>
<tr>
<td>Access control complexity</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Access complexity</td>
<td>Low (freq. referencing)</td>
<td>High (timing)</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>

Table 5.4 - Implications of multiple access techniques.
lend themselves to Time Division Multiplexing, e.g. connection less and varying bit rate services, than WDMA. For this reason overlaying TDMA channels by WDMA as a means of extrapolating capacity requires careful consideration of the control and arrangement of services within the wavelength channels. Conversely, the fact that a TDMA terminal inherently receives data which is not intended for it may represent an increased security risk.

The conventional approach to the problem of controlling a WDMA network is to introduce a protocol at the CO which signals a user terminal, on a predetermined wavelength, what frequency to tune its receiver to every time it is about to receive data. However, the longer the distance between the CO and the user and the greater the bandwidth, the “shorter” the transmitted message compared to the time of signalling and tuning the receiver. It has been found that if the bandwidth distance product exceeds 1 km-GBit/s, advanced warning of individual cells becomes inefficient [13]. The alternative, is to arrange the services in the wavelength channels so that there is a minimum of conflict between them. An example of such an arrangement could be two digital downstream wavelength channels, one which provides TDM broadband services and the other bursty and connection less services as well as control. This way the receiver when inactive and waiting to receive would need to monitor only one wavelength channel, the second, without risking missing any data. Additional wavelength channels could be overlaid for, say, distributive CATV and video.

In such a network in the future it unlikely that the various services will all originate at the CO or even in the same place. It seems more reasonable to assume that they will be transmitted from a number of central data stores located in various places around the interoffice or even the national network, to the local COs where the information will be redirected. In this case WDMA combined with transparent optical routing at the COs makes for a very attractive and elegant system. If the individual services are transmitted on distinct wavelengths the interoffice network can route them to and through the CO nodes, directly to the users, using wavelength routing [10]. By minimising electronic switching in this way we remove potential bottle necks and “future proof” the network to changes in both services and data formats. Section 6 and the following chapter discuss this aspect of WDM in greater detail.

5.3.5 Examples of optical access networks

This section briefly describes five optical fibre based access networks in terms of the architectural concepts discussed above. All the networks have been demonstrated in either
in laboratory or field trials and together seem to form a representative cross-section of the current efforts in optical subscriber loop trials.

**Totally Transparent Optical Subscriber System**

The Totally Transparent Optical Subscriber System (TTOSS) is a single star FTTH network developed by Philips Research Laboratories. It is based on either SDM or DDM, depending on the transmission distance, and WDMA services. The multiplexing technology used is either passive wavelength routers or optical splitters with optical filters, depending on the available power budget.

Demonstration systems include a system with a single downstream wavelength carrying 140Mbit/s video with upstream traffic providing channel selection; and another system carrying 512kBit/s TDM services to the customer.

**AT&T SLC Series 5 carrier system**

The AT&T SLC Series 5 carrier system is available as a commercial product and is now being deployed in several fibre trials. It has an active double star architecture which provides either FTTH or FTTC. It operates at 1.3μm and uses PIN diode receivers.

**Raynet Fibre Bus System**

Like the AT&T SLC Series 5 carrier system, the Raynet fibre bus system network is also a commercial product. It provides FTTC/FTTO over a SDM bus with active add/drop nodes. Each node serves up to 8 customers over copper drops, using TDMA to provide voice traffic and control.

**Telephony Passive Optical Network**

The Telephony Passive Optical Network (TPON) is a triple passive star, 128 way split, FTTH network developed by BT Research Laboratories. It is initially supposed to provide telephony services only but has been designed to allow the upgrade to a Broadband Passive Optical Network (BPON). Although, the TPON is based on TCM and TDMA, each the user terminal have an optical filter built in so terminal equipment will not have to be replaced if WDMA is introduced. The system has been demonstrated, including the automatic ranging protocol, for up to 2MBit/s byte interleaved upstream, over 6km operating at 1.3μm.

**TPON to BPON upgrade**

The BPON-TPON upgrade introduces WDMA to the TPON architecture. A modest number of wavelengths, typically 2-6, can be used to provide additional services such a
video signals. 32 TDM 69MBit/s video channels will, however, only have a reach of 32 customers due to the power budget.

**Passive Photonic Loop**

The Passive Photonic Loop (PPL) is a high capacity passive double FTTH star developed by Bellcore. It uses WDM directional transmission and WDMA with passive routing at the RN with a different wavelength for each customer in both the up and downstream direction. As each customer terminal only receives one wavelength channel, tuneable receivers are not required. 10 customer systems have been demonstrated at 1.5μm over distances up to 13km, using DFB laser, providing both 600MBit/s and 1.2GBit/s per customer. A second trial based on Light Emitting Diode (LED) transmitters combined with spectral slicing [14] has demonstrated 384kBit/s and 1.5MBit/s per customer.

Overall the proposed uses of WDM for both access and transmission are quite consistent. However, apart from the PPL architecture, it serves only as a capacity or service enhancing feature. Even in the PPL, the usage of the wavelength channels is relatively limited and possibly not all that effective; it only substitutes higher component complexity for fibre.

### 5.4 Impact of optical fibre nonlinearities

Optical fibre can, in the context of telecommunications, normally be treated as a linear medium. However, it does in reality exhibit a number of different nonlinear effects. These are Stimulated Brillouin Scattering (SBS), Stimulated Raman Scattering (SRS), Self Phase Modulation (SPM), Cross Phase Modulation (CPM) and Four Photon Mixing (FPM). A high power density, caused by a number of frequency multiplexed optical carriers within a single fibre or high semiconductor laser powers used to overcome splitting losses, might cause degradation of the carried signals due to one or more of these effects. Consequently, we must limit the optical power per carrier so as not to invoke any of these nonlinearities. This may influence the choice of access network architecture, including network topology, multiplexing and access techniques. The succeeding section presents a short summary of the limits these effects impose on optical access networks. For a more in depth review of each of the above nonlinear effects please refer to chapter 3.

#### 5.4.1 Optical fibre nonlinearities

Assuming that there is no depletion of the pump wave, P₂, due to fibre nonlinearities, the gain of the probe due to a particular nonlinear effect is described generally by eqn. (5.1),
where $\gamma$ is the gain coefficient, $L_e$ is the effective length given by eqn (5.2) and $A_e$ is the effective cross section area of the fibre core. Typical values for $A_e$ are $80\mu m^2$ and $50\mu m^2$ for normal single mode and dispersion shifted fibre respectively.

$$P_i(L) = P_i(0) \exp\left(\frac{\gamma P_i(0)L_e}{A_e}\right)$$  \hspace{1cm} (5.1)

$$L_e = \frac{1 - \exp(-\alpha L)}{\alpha}$$  \hspace{1cm} (5.2)

In an amplified transmission system the effective length is the sum of the effective lengths of each segment of amplified fibre. Assuming the attenuation of the fibre, $\alpha$, is 0.2dB/km, Fig. 5.5 plots the effective length versus the real length. From Fig. 5.5 and eqn (5.2) we see that the effective length, $L_e$, is always less than the real length, $L$. As $L$ tends to infinity $L_e$ tends to $1/\alpha$, in this case 22km.

From eqn. (5.1) it is possible to derive the limiting condition for the individual nonlinearities, see chapter 2. For the purpose of this report, however, we shall simply state them.

### 5.4.2 Results

Using the above equations it is possible to calculate the maximum input power for each of the non-linear effects. Table 5.5 below summarises the results for dispersion shifted fibre.
with 128 channels spaced 5GHz apart as this appears to be the most limiting transmission configuration proposed.

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>Max. per channel input power, L=1km</th>
<th>Max. per channel input power, L=5km</th>
<th>Max. per channel input power, L=10km</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS</td>
<td>27.6dBm</td>
<td>21.0dBm</td>
<td>12.8dBm</td>
</tr>
<tr>
<td>SBS</td>
<td>17.2dBm</td>
<td>10.7dBm</td>
<td>9.6dBm</td>
</tr>
<tr>
<td>CIP</td>
<td>12.6dBm</td>
<td>6.1dBm</td>
<td>5.0dBm</td>
</tr>
<tr>
<td>FPM</td>
<td>-4.0dBm</td>
<td>-2.4dBm</td>
<td>-2.5dBm</td>
</tr>
</tbody>
</table>

Table 5.5 - Maximum per channel input power due to optical fibre nonlinearities

The greatest limitation is clearly caused by FPM. However, a maximum input power of -4dBm and -2.5dBm for transmission distances of 1km and 10km respectively seem to leave room for a reasonable power budget.

5.5 Technology

The preceding discussion did not take into account the availability and complexity of the optical devices themselves. However, this is clearly a very important area particularly in the access network where the per unit cost is most critical. In this section we present a brief discussion of the available technologies and their performance with respect to, in particular, WDMA.

When proposing WDMA the technology emphasis is on producing an array of laser light sources and optical filters which exhibits high enough stability to maintain the low error rates required. Consequently, we shall focus our discussion on these two areas. WDMA systems are based on one of the four configurations listed below:-

1. Fixed wavelength lasers and filters.
2. Tuneable lasers and fixed filters.
3. Tuneable filters and fixed lasers.
4. Tuneable filters and lasers.

In the case of wavelength tuneable lasers or filters, connections are based on dynamic allocation of wavelengths by tuning either the receiver or transmitter or both. In WDMA networks the tuning speed is therefore of interest. This is especially the case when applying a packet based protocol in the wavelength as well as time domain. For example, a tuning speed of 1ms in conjunction with 155MBit/s ATM cell based transmission implies a maximum cell arrival rate of 997cells/s and a maximum communication rate of 3.8MBit/s. In circuit switched systems the tuning speeds are not as critical.
A large number of WDM channels implies either a large tuning range and/or wavelength discrimination. As local access networks are typically single hop (or two hop if the RT incorporates wavelength routing) an extinction ratio ≤ 20dB is typically sufficient to ensure the power penalty due to inter channel crosstalk is less than 1dB.

5.5.1 Lasers

When investigating various laser diode configurations the physical characteristics such as the output power, linewidth, tuning range and tuning speed are normally compared. Naturally, these parameters are of primary importance as they directly determine the feasibility of a given network architecture. In addition, however, there are a large number of “secondary” parameters which are related to the actual practicality of using a specific laser. Ultimately, this is determined by the cost but put in terms of more tangible parameters they include tuning method, operational stability, operational difficulty and production difficulty.

5.5.2 Filters

Similar to lasers we generally classify filters by bandwidth, cross-talk, tuning range, tuning speed and loss. Typically the minimum channel spacing is 3-10 times the filter bandwidth for acceptable cross-talk levels. Therefore the WDMA systems only allow around 30% utilisation of the available bandwidth, compared to nearly 100% for TDMA or SCM. However, as the bandwidth of optical fibre by far exceeds any realistic capacity prognoses in the access network, efficient utilisation of the optical bandwidth is not of importance.

5.5.3 Amplifiers

The transmission distances associated with current local access networks is typically a few kilometres. The low transmission losses of optical fibre may make it viable to extend this to tens of kilometres. However, the greater the range of this type of network, the more likely it will be that it is shared between a greater number of users. Therefore, the total loss experienced by a signal may be several tens of dBs even though the attenuation due to the fibre is only a couple of dB’s. For this reason it may be that, in particular 2nd generation, fibre based local access networks will use optical amplifiers to overcome large splitting and splicing losses.
Optical amplifiers have three major advantages:

1. By reducing the amount of electronics in the CO the reliability of the network can be increased.

2. Because the amplifiers are transparent to the transmission format of the information carried, upgrading the network requires the minimal replacement of electronics, and this can be done gradually.

3. The large bandwidth of optical amplifiers allows WDMA without any additional complexity to the transmission system itself.

The two main types of amplifiers of current interest are semiconductor amplifiers and the Erbium Doped Fibre Amplifiers (EDFAs). The advantage of semiconductor amplifiers is their small dimensions and that they may be integrated into PICs. This makes them of particular interest as pre- and postamplifiers. The superior characteristics of the EDFA, especially in multiple wavelength networks, means that it is well suited for inline amplification in WDMA systems. In this case the transmission window is reduced to the passband of the amplifier, namely ~35nm. In conjunction with, for instance, acousto-optical filters this severely limits the channel capacity.

5.6 Future Potential for WDM in optically routed networks

Despite the spectacular advances in the development of fibre communication systems, it is obvious that much capability remains unexploited. Some simple numbers make the point. The fibre low-loss transmission window centred around 1550nm is of order 100nm in width. This corresponds to a spectral window of about 13,000 GHz, to be compared with the total radio and microwave spectrum of about 300 GHz. Transmitting "high data rates" of 10 GBit/s thus exploits much less than 1% of the available communications space. Even restricting the transmission window to that covered by the Erbium Doped Fibre Amplifier (EDFA) hardly alters the situation. Its gain region extends over about 35nm or 5000 GHz.

Several factors have blocked the use of this communications space. The difficulty of extending electronic modulation techniques into the multi-GBit/s region suggests that access to the space will not be gained by means of ever higher digital electronic multiplexing data-rates. Even using microwave sub-carrier techniques, one would be hard-pushed to access but a small part. The clear message is that to access this communication space, we must follow the example of the radio and microwave engineer in accessing his spectrum space and deploy many (optical) carriers each carrying a conveniently (electrically) accessible information rate. Our ideas for optical routing using WDM
technology in this way in an all-optical network are discussed in detail in the following chapter which contains a copy of paper published in IEE Proc.-Optoelectron. An outline of the concept(s) (which have also been patented) is given below.

Noting that the EDFA transmission window extends across about 35nm or 5THz (5000 GHz) of optical spectrum, and for reasons of avoiding unwanted non-linear interactions in a long fibre link, it is probably desirable to separate carriers by at least about 50GHz or 0.35 nm leaving room for about 100 carriers. We then examine how to exploit this potential to address the UK National Network. Consider an extended route passing through a network in the form of an NxM rectangular grid of nodes. For the UK, one might associate each node with a major trunk node of which there are about 50 and hence we might, for thinking purposes, consider N = 5 and M = 10. In traversing the longest route in such a network, assuming one takes a direct path, then one must pass through at most (N+M-3) nodes at each of which channel regrouping must occur. If W wavelength slots are used with a filling factor f, and there are p fibres accessible along each of the (N+M-1) arms. When we come to set up the last route through this network, assuming that fW-1 channels are already established on a random basis throughout the route, then the probability of not being able to connect a given channel that is empty at the outset all the way through is given by eqn. (5.3).

\[
P_{\text{blocking}}(f,p) = \left[1 - \left(1 - \frac{(f - 1)}{W}\right)^{(N+M-3)}\right]^{(pW+1)}
\]

Here we have assumed that we search all the (pW+1) unused wavelengths to find one that is free all through the route. Evaluating this result for N+M-3 = 12, p = 3 and W = 100 as a function of f, the filling factor, we find that with f = 0.5, the probability of not being able to make that final links is already below 3x10^-6. In other words, by exploiting the spectrum space that is available in the fibre by using the unfilled space, complex routing operations through large networks are theoretically possible.

If we then ask ourselves how complex an address space do we actually need to serve the UK National Network, again some simple calculations give a useful pointer. We note that there are of order 300 local exchanges in the UK. If one examines the traffic flow from any one such area to each of the others, one typically finds a 10 to 1 variation in the flow between the most and least frequently accessed areas. These traffic flows, plotted as a histogram, typically form a sub-linear plot. In other words, if the least flow is one unit of traffic, the most frequently accessed has about 10 units and the mean is < 5 units/area to all areas. This points to a requirement for an address space that can be broken up into approximately 5x300 units of traffic flow so that frequently accessed areas can be allocated proportionately more communication space than less frequently accessed areas.
This observation then leads directly to our network proposal which envisages a mesh network for the core, optically connected to fibre rings for the "local exchange areas" which in turn might ultimately connect to PONs as local access networks. (See Fig. 5.6)

**Fig. 5.6 - Hierarchical view of the proposed (optically transparent) network architecture.**

We consider, at the local access level, a minimum of two fibres in any access ring each capable of carrying a basic data rate equivalent to STM-16 per active wavelength carrier but with each 125µsec frame sub-divided into 16 sub-frames, each of approximately 125/16 = 7.8µsecs duration. With order of 100 wavelength slots available for access and a 50% filling factor, this then provides an address space for communication purposes containing 0.5x100x2x16 = 1600 units of routeable communications space which maps well onto the UK local (exchange or ring) address requirement. The switching requirements in such a network postulate relatively slowly reconfiguring wavelength multiplexers, demultiplexers and regroupers or space switches, operating to a very slowly varying control sequence that responds only to variations in mean traffic flow.

We have also considered how traffic emanating from the customer terminals attached to a Passive Optical Network attached to such a ring might be entered into the network and extracted from it, both in ATM format. Clearly this implies the existence of fast switching at some point, since a single ATM cell at the STM-16 rate of 2.5Gbit/s lasts only about 170ns. However, given our network assumption of a pre-partitioned address space at each entrance and exit exchange, only precision timing at the PON level and its entrance to the Ring is required. Once a cell has been injected into the correct point in the address space, it will be routed to its destination using the slowly cycling switching of the core network. We also envisage the cell header being used in a smart-pixel switch to drop it from an incoming ring onto the destination PON. In this way, many of the most extreme problems associated with fast routing of ATM cells seem to be completely side-stepped and the only decision is the simple local one to be taken at the launch time of what time co-ordinates in the available ($\lambda, t$) space should be allocated to the cell. Buffer memory in
such a network is entirely banished to the opto-electronic interfaces at its periphery and control complexity is minimised. The penalty for this simplification is wasteful use of "bandwidth" or transmission capacity. It is axiomatic in such a network that individual wavelength bearers will be less than 100% loaded since by design we utilise less than half and in practice probably much less than that. Given the claim that "bandwidth is free" in optics, perhaps this represents a good trade.

We believe that our proposals demonstrate the enormous unexploited scope that exists to use the wavelength domain for sophisticated routing of data. Our discussion of components highlights the fact that many of the component requirements are now being studied in research laboratories and whilst not well-engineered, certainly do not appear to be unrealisable dreams.

5.7 Summary & Conclusions

During the course of the study, a number of factors have come to our attention as being of major significance. We are acutely aware that we do not have access to proper costing data for the various network possibilities and we are thus left "backing hunches". These are highlighted below :-

1. There appears to be a growing opinion (consensus?) that FTTH will not cost-in in the immediate future and that a broad-band service for the private customer must be launched via FTTC and/or the Fibre/Coax-Hybrid (FCH) approach.

2. There is a rapidly growing interest in Video on Demand (VoD) which in turn concentrates attention on the growing emergence of a digital data rate in the region of 2Mbit/s as a potential point of convergence since it appears to be common to :-
   - compressed video storage for electronic file servers for VoD
   - ADSL for VoD
   - the range of Compact Disc products (CD, CD-ROM, CD-I and CD-V as well as Photo-CD)

3. Taken together, these call into question the ambitious proposals that appeared common with programmes such as EC-RACE for 155Mbit/s provision to the home (CCITT recommendation I.413).
4. Taken together, these do not immediately appear to strengthen the case for the use of WDM!

5. In terms of transmission to and from the local loop area, if the economics favour the use of bi-directional operation on a single fibre, then simple WDM seems likely to be strongly advantageous as a way of minimising Near End Cross Talk problems by using different wavelengths for up and down stream signals.

6. The next obvious potential use for WDM is in service differentiation. For example, one might envisage feeding a PON for 64 homes with (order) 70 Mbit/s VoD channels on one wavelength and using 2 other wavelengths for up and down stream TPON telephony. We suggest the advantages of this approach would be:
   - ease of service fusion at the PON entry (using a W.MUX)
   - ease of service separation at the Curb or Customer Terminal
   - ease of billing and service management

7. WDM technology also seems ideally placed for service upgrading in that a new service can be overlaid on an existing network with additional receivers only being added to those customers choosing to take it. Existing customers installations remain unchanged.

8. Note in of the cases 6 & 7 one is balancing the cost of additional optoelectronics against the cost of higher data rate TDM transmit and receiver electronics/optoelectronics and software/management complexity. We are not able to assess that trade-off.

9. WDM technology could also be used to address different geographic communities over a single PON. For example, we envisage that one (group of) wavelength(s) might be used to service businesses and another for private users, separated by W.MUX at a suitable downstream distance. In practice however we suspect that it will be preferred to provide separate feeder fibres.
10. WDM technology could be used to upgrade an installed PON set up for the delivery of "broadcast TV" services by splitting the final service area into a series of sub-areas, each served via a different wavelength carrier. In this way the balance of capacity could be shifted to provide a number of carriers exceeding the number of customers so that switched star operation becomes possible.

11. Looking further into the future, we have noted that among the key telecommunications technologies, it is fibre that today has by far the greatest unexploited capability. (e.g., the EDFA window is circa 5000GHz wide yet we use at most a few GHz of spectrum). Since there is no expectation of a 1000-fold increase in traffic, this raises the philosophical question of whether there are other ways of exploiting this space.

12. The components to access this empty communications space are rapidly becoming available in the form of tuneable DFB type lasers, tuneable filters etc. An opportunity that exists.

13. We offer our own "vision" of the Sparsely-Filled DWDM National Network as one possibility that inexorably leads to a number of new issues :-

   - completely new ways of multiplexing data into and out of the network
   - substantially different control characteristics
   - different transmission characteristics (greater transparency, lower delay etc.)
   - new ways of trading-off between transmission and switching.

14. We do not suggest that SF-DWDM is the only such possibility or that it is likely to become commercially attractive in the very short term. We simply offer it as an illustration of the radically different alternative routes for future development that are now opening up.
5.8 Bibliography

Proceedings The X International Symposium on Subscriber Loops & Services, issls 93, “Communications at a Crossroads”, September 27 - October 1, 1993, Vancouver, British Columbia, Canada


Optical Switching in Networks Collaboration, “Notes from meeting at Heathrow Excelsior”, Wednesday 18th May 1994

John E. Midwinter & Staff, “Potential for Wavelength Division Multiplexing in Local Broadband networks”, 1994
5.9 References


Chapter Six

Transparent All Optical Networks

As outlined in chapter 4 this chapter concerns itself with the investigation of routing within a complete network entirely within the optical domain: from a customer, through the access, core and back through the access network to the receiver. The chapter is (in agreement with Prof. J. E. Midwinter, supervisor and Dr. Mick Flanagan, post-graduate tutor) composed of the two main papers published by the author regarding this part of the author's PhD work, “Introduction” and “Summary & conclusions” sections. Initially the premise and context within which the work was performed is outlined, the results that were published (and patented) follow and a retrospective comment on the work and ideas in the light of subsequent & current research and results conclude the chapter.

6.1 Introduction

With reference to the comparison of optics and electronics, “optics versus electronics”, (section 2.1) and our discussion of the “Application and potential for optics in communication networks” (section 4.2) in the previous chapter, we aim to design a ‘hypothetical’ transparent optical network which would allow us to transmit information from one user to another entire within the optical domain in a manner which would allow us to take full advantage of the potentials of the emerging photonic technologies, refer chapter 2 (part I).
Would it at all be possible to envisage an AON architecture which was modular and scalable to a level required in the next generation of broadband multi-service networks. If so, is there a gradual evolutionary sequence of steps which might lead to it from a starting point of today's networks?

The work (and results) presented in the following two sections (papers) should not necessarily be taken as a literal suggestion for a future national UK network but rather as an exercise, the purpose of which was to explore a hypothetical objective from an 'impossible' set of starting assumptions.

The two sections (papers) each investigate the same Sparsely Filled WDM (SF-WDM) architecture from the premise of a national (UK) telecommunication network. However, they each affirm their respective positions and emphases. Section 6.2 (“A Modular a Scalable Transparent Optical Network”, [1]) discusses the proposed (non-)utilisation of the available transmission optical spectrum from the perspective of modularity and scalability of the proposed network architecture/concept, refer section 4.3.1. Section 6.3 (“Towards an optical ether”, [2]) deals with the same concept, i.e. SF-WDM, and discusses the same architecture but this time with an emphasis on the evolutionary sequence, starting from today's networks and concluding in an Utopian 'optical ether'.
A Modular and Scalable Transparent Optical Network

M. Sabry and J.E. Midwinter
Department of Electronics and Electrical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

Abstract • In this paper we propose a transparent optical telecommunications network architecture. Such a network, typically having the dimensions of a national UK network, (25x 10^6 subscribers) is shown to provide an aggregate capacity of 2MBit/s per subscriber.

Introduction

Recently, there has been much discussion of the concept of an optical ether for national or global telecommunications use. However, the various proposals published seem fundamentally to require simultaneous implementation across a whole network to reap significant benefits. In this paper we propose a modular and scalable network architecture which allows signals between subscribers to be routed entirely within the optical domain in a manner that permits graceful evolution of the network starting from today’s position of optical fibre transmission on some routes only.

Opportunities, constraints and assumptions

A critical evaluation of what optics brings to telecommunications increasingly focuses on its interconnect capability, low attenuation and high bandwidth and de-emphasises its processing capabilities when compared to digital electronics. Consequently, it would appear that any solution must somehow exploit these strengths to overcome the growing hurdles currently caused by processing bottlenecks in telecommunication transport networks.

Fibre is already widely used for point to point transmission. Regenerators, along these routes dictate a rigid data format in the time domain. However, the continued development of the Erbium Doped Fibre Amplifier (EDFA) indicates that regenerators will be replaced by amplifiers and fibre links will become transparent “loss less” data pipes with a transmission window of about 35nm or 4700GHz in spectral extent. Furthermore, this communication space is the same in any fibre in the network, whether it is part of a trunk line or the last 10m from the curb to a house. If Wavelength Division Multiplexed (WDM) STM-16 channels were spaced by 10-50GHz, full loading would imply up to 400x2.4GBit/s = 0.96TBit/s per fibre. Even though gradually more of the bandwidth will undoubtedly be exploited as the demand for capacity increases, it seems improbable that we could contemplate fully filling this potentially vast capacity.

The grey shaded plane in fig. 1 represents the communication space in a single fibre. As has already been pointed out, it is possible to wavelength multiplex and one can therefore think of the frequency dimension as divided into bands each representing a transparent optical channel. In addition, time may also be partitioned analogous to conventional Time Division Multiplexing (TDM). Combined, the two result in a matrix of "pixels", each of which may be described by a unique (λ, T) reference within the fibre. Thus if each time frame is made up of 16 sub-frames (F=16) then each pixel within the communication space might correspond to the equivalent capacity of one optical STM-1 channel. A third dimension is added to accommodate the possibility that more than one fibre may exist between any two points in a network. In contrast to the dimensions of the time-channel plane, which are the same at any point in any optical network, the extent of the fibre dimension may vary through a network.

In order to maintain strictly all-optical transmission paths, the only operation that can be permitted when switching is to shift the time-wavelength plane (or parts of it, say one pixel) within this fibre dimension, i.e. space switching.

Fig. 2 uses the fibre-channel-time space to show an all-optical node switch with 3 input and 3 output fibres, q=3. The communication space in each fibre is divided into 6 wavelength slots, W=6, which in turn each contain 6 sub-frames, F=6, but only half of the available channels are filled, f=0.5, (indicated by the shaded pixels). The switch has the ability to separate the individual wavelength slots on each of the input fibres and reassemble them in a different configuration on the output fibres [1]. However, as we are proposing all optical switching the pixels must maintain their position within the time-
wavelength planes of the fibres, from the input to the output of the node. Nonetheless, given that nodes generally have several input and output fibres originating at and destined for a number of different nodes the channels contained within each of the pixels in the communication space (e.g. STM-1) may be routed through an extended network.

Evolving towards a national optical network

Given that we do not wish to use all of the communication space simultaneously, we have previously shown that it is possible, using the above approach, to route through a complex multi-node mesh network, typically having the dimensions of the UK core network [1]. It was found that as long as half of the wavelength slots in any one fibre are left unused at any given time, the probability for being able to establish a sub-pixel, e.g. to STM-1, transparent optical point to point link between two nodes along any particular route is greater than $10^6:1$. If the EDFA transmission window is partitioned into $10^6$ wavelength slots, i.e. 47GHz channel spacing, a network of this type would provide an aggregate capacity of $50 \times 2.4 \text{ GBit/s} = 120 \text{ GBit/s}$ per fibre.

The discussion below examines the implications of extending the optical domain outwards towards the customer, via rings and Passive Optical Networks (PONs). Fully developed we believe such a network could ultimately facilitate routing of signals from customer to customer entirely within the optical domain.

A schematic representation of a network of this type is shown in fig. 3. The inner core (the highest hierarchical level) of the network is a highly interconnected mesh. Each of the core nodes provides routing within the inner core, as well as access to one or more ring-type networks, which make up the outer core. The ring and mesh fibres entering and leaving a core node all have equal status in the switch. In turn the rings are optically connected to the peripheral terminals via PONs.

It seems desirable that our approach should start from a position of having the maximum overlap with today’s network. In the first stage of evolution, we therefore envisage existing transmission links within the inner core network being upgraded to use Dense WDM transmission but with each carrier operating within the Synchronous Digital Hierarchy (SDH) to standard interfaces. Later some interfaces would be replaced by transparent wavelength switches and thereby facilitate transmission between rings entirely in the optical domain. Ultimately, individual customers could be serviced directly through PONs optically connected to the rings.

Throughout the development of such a network the component set remains the same, namely, the wavelength space switch, the optical ring PON interface and the opto-electronic interface. Consequently, the evolution can proceed as desired and unevenly across the network. For example, in an area with mainly business subscribers we might imagine the network being developed to ring or even to PON level at the same time as it was being developed to the core level in a rural area.

A strong argument for a gradual evolution is economies of scale. It is axiomatic that opto-electronic interfaces will always be required at the periphery of the optical ether. Although they remain essentially the same as they move closer to the customer, they will be required in increasingly greater volume at lower cost.

In a high capacity network carrying traffic multiplexed from many users, one would not expect the mean traffic flow along routes to fluctuate by large amounts in short times. Accordingly, we propose that the traffic carrying ability in the inner and outer core networks (mesh and rings) should be allocated on a slowly time varying basis (hours), initially in blocks of one optical carrier at STM-16. However, as the network is extended to include rings and the opto-electronic boundary is brought closer to the subscribers, we can introduce a limited degree of time division (albeit without time shifting buffers at nodes) in order to further sub-divide the communication space. Recognising that telecommunications is based around an 8kHz sampling cycle, corresponding to a 125μs frame duration, we propose dividing it into 16 sub-frames, each of about 7.8μs. This has the effect of reducing the minimum capacity building block to 155MBit/s and increasing the total address space in any one fibre by 16 times.

In order to synchronise the switching of sub-frames in the core network all sections of fibre
should be made integer lengths of the sub-frame interval. An 8kHz frame rate divided into 16 sub-frames would on average imply an additional 0.75km of fibre in each cable termination. In addition, a guard-band is necessary between information contained in consecutive sub-frames in order to allow for differences in propagation delay caused by material dispersion.

Although capacity may be allocated in blocks of 155Mbit/s within and between rings, if the optoelectronic border is to be brought out to the individual customers, the network must realistically be capable of providing as little capacity as 64kbit/s channels. We therefore propose that the sub-frames be composed of time slots of a duration sufficient for the transmission of standard ATM cells (~170ns at STM-16) which could be dropped into or extracted from a PON individually. As a result, it would be possible to transmit information from customer to customer entirely within the optical domain in denominations ranging from an ATM cell upwards.

Connections between subscribers are then established by injecting ATM cells into vacant slots in a sub-frame channel that connects the relevant rings. Since the routing pattern of the sub-frames is predetermined at the inner core level, the data channels (pixels within the fibre communication space of fig. 1 and fig. 2) may be viewed as set of "postal pigeon holes" each addressing a different destination ring. Consequently, the transmission of information from source to destination is reduced to the insertion and removal of ATM cells at the two involved ring-PON interfaces. Moreover, as these are both strictly localised processes, maintaining precise timing is required only within the individual PONs.

Traffic analysis of the network

As previously stated, the function of the core network is to establish high capacity "transparent" data channels between rings corresponding to \( R \) geographically localised communities of interest. Based on the requirement that the network must be able to accommodate traffic flows varying by a factor of 10 between the least and most popular nodes addressed by any one node, \( R \) was calculated. Within the context of a national UK network we typically envisage \( R=320 \). This would imply rings with an average radius of roughly 30km assuming a uniform population density across the country. Alternatively, with \( 25\times10^6 \) network subscribers, it implies ~78,000 subscribers per ring.

Although, we have postulated that only part of the time-wavelength-fibre space will be used for transmission at any one time, we also assumed that the connection pattern in the core would be (albeit slowly) dynamically varying. Consequently, there will be a finite probability that a new channel along a given route can not be accommodated. In order to calculate this probability we use the equations developed in[1]. For example, if there are 8 and 2 fibres per cable in the mesh and ring networks respectively, with 100 wavelength slots and a 50% filling factor in each fibre, then the probability for not making a connection between two rings separated by 11 core nodes (longest path in 7x7 rectangular mesh) would be \( 5\times10^{-28} \) on any particular chosen route.

In addition to the blocking which takes place within the core network, there exists an additional set of blocking probabilities within the access network which arises from contention between ATM cells. In general there can exist three different types of contention within the access network. When an ATM cell is injected into the network by a terminal it will cause a collision if the given ATM slot already contains another cell. Accordingly, the network control should keep track of which of the slots in the various data channels in a ring are in use at any given point in time and only make those that are empty available for "new" connections. This implies that before a connection can be established between two terminals, the one which initiates the request must signal the local network control to determine the time and wavelength references of the appropriate channels which are available as well as which ATM slots are unused within them. Given that a ring consists of two or more fibres the control algorithm must also verify that two terminals in the same PON are not competing for identical ATM slots on different ring fibres as this will result in the cells colliding within the PON. In such cases, only one of the available slots may be allocated. The control issues associated with these two types of collision may be administered by the local hardware which oversees the dialogues between the PONs (terminals) and their rings. These localised control centres would not store sufficient information about the whole network for a terminal to determine, at the time of initiating a call, whether the receiving terminal is available or in fact if the time- wavelength reference chosen for the transmission of ATM cells will cause contention with already existing connections between other PONs in other rings and the receiving PON. Thus, the third type contention arises from potential collisions in the receiving PON. Viewed from the position of a ring receiving a number of calls (or request for calls), however, this new set of administrative control issues are again localised. We therefore suggest that this network architecture lends itself to distributed control concentrated, for example, in each ring. Communication between terminals would then only require the involvement of the control centres directly affected, i.e. generally only two.

We assume that at the time of making a connection ATM cells are selected and allocated at
random (provided they do not cause collisions within the transmitting PON), i.e. a non-intelligent control algorithm. Hence, the probability that two or more indistinguishable ATM cells do not arrive in the same PON at the same time, and thus cause a collision, may be approximated as the probability, \( P_c = \exp(-\alpha) \), that no ATM cells arrive in a given PON during a time period equal to the transmission time of an ATM cell, where \( \alpha \) is the arrival rate of contentious ATM cells.

In the case where the receivers in customer terminals are not wavelength selective, \( \alpha \) is equal to the ratio of used ATM cells to the number of PONs in a ring. For example if a ring has the dimensions used above and the ATM cells transmitted in a certain slot are distributed between 500 PONs (which in the context of a network with \( 25 \times 10^6 \) customers and 520 rings implies \( 150 \) subscribers per PON) \( \alpha = 100/500 \), where \( f \) is offered load expressed as a fraction of the available capacity. Similarly, if the receivers are made wavelength selective \( \alpha = 0.25/500 \). The success rate, \( P_s \), and the throughput, given as \( \text{offered load} \times P_s \), is plotted in fig. 4 versus \( f \) for both of the above examples.

![Fig. 4 - Success rate and throughput](image)

The throughput graph plotted in fig. 4 represents the total fraction of traffic passing through the network and arriving uncorrupted at its destination. Accordingly \((1-\text{Throughput})\) describes the proportion of data which is lost or corrupted. Therefore, depending on the protocol, this may be equated to either a Bit Error Rate (BER), or to the proportion of the gross offered load which is retransmitted data. If the basic protocol does not incorporate retransmission, the net throughput will be equal to the gross throughput but with a Packet Error Rate imposed by the protocol. In the case where the terminals are wavelength selective this equates to a maximum BER of \( 5 \times 10^{-4} \) at 100% PON load or a mean traffic of 3MBit/s per customer. In comparison, if all corrupted packets are retransmitted the control itself will not give rise to a BER, although the net throughput will be reduced by a factor equal to the original success rate, i.e. \((\text{net throughput} = \text{offered load} \times \text{success rate})^2\). So, similarly to the hardware vs. performance trade-off found in fig. 4, we find that there exists a relationship between protocol issues and the performance/characteristics of the network. It should be noted, however, that even without wavelength selective receivers and this reduced throughput the network would still provide an aggregate capacity of 2MBit/s per subscriber.

Moreover, provided \( \alpha^1 < 1 \), i.e. the arrival rate of contentious ATM cells does not exceed the number of PONs in a ring, the throughput does not exhibit a maximum (for \( f \leq 1 \)). This implies that the throughput will increase continuously in step with the load. In other words, we would not expect it to be necessary to build safeguards into the network control to cope with congestion as is currently proving to be necessary with conventional ATM [2].

**Summary**

The wavelength routing strategies in conjunction with the network architecture described in this paper represent a scalable transparent optical network suitable as a national telecommunications network. The network dimensions used as examples were chosen to coincide with networks of a size comparable to that required to cover the UK.

The three components found necessary to evolve such a network are a wavelength space switch, an all optical Ring-PON interface, and an opto-electronic "line-card". The main constituents of these three components are already available today although not in the configuration required here. However, as the architecture is modular and hierarchical it can benefit from economies of scale given that a standard specification for the components can be reached.

Although the analysis was presented has mainly focused on the fully implemented network (including mesh, ring and PON), the expressions developed are equally valid and pertinent for the individual parts of the architecture.

In conclusion, we note that although the basic premise of the design has been based around STM-16 and the EDFA transmission window there are no fundamental restrictions in the architecture that prevents it from being applied to other technologies.

**Acknowledgements**

The authors would like to thank members of the Networks Group at UCL and Dr. P. Bayvel at BNR Europe Ltd. (Networks Technology) for their support and useful discussions. This work was carried out with the support of SERC and BNR Europe Ltd.

**References**


Towards an optical ether

M. Sabry
J.E. Midwinter

Indexing terms: Optical routing, Passive optical networks, Wavelength switches, Broadband network

Abstract: The authors propose a novel technique to exploit the unused bandwidth in optical fibres to develop an optically routed broadband network that is potentially deployable on a scale matching that of the UK national network. Furthermore, the techniques proposed allow such a network to evolve gradually from a few dense wavelength division multiplexed (DWDM) transmission links through sections of optically routed core network to reach finally a fully optically connected 'house to core network to house' ATM connection. The design provides for 25 million customers with in excess of 3 Mbit/s capacity per customer. The technology proposed relies on DWDM transmission and optical space switching without time or wavelength shifting.

1 Introduction

In the last decade, there has been much discussion of photonic switching aimed either at replacing part of an electronic switch, such as the matrix, by optical components or more ambitiously, building a network in which the whole transmission path is entirely in the optical domain [1]. The last-mentioned have predominantly been closed network architectures of moderate size and as a result been restricted in their application to areas such as high capacity data networks or computer interconnects between multiprocessor machines [2–4]. More recently, there has been much discussion of the concept of an optical ether for national or global telecommunications use [5–7]. However, most of the proposals published seem to require fundamentally simultaneous implementation across a whole network to reap significant benefits. For example, LAMDNAET-type transparent optical networks [2] require the ultimate number of nodes in the network to be fixed at the beginning of the design. In this paper, we explore a different approach to the same problem, namely, to examine how signals may be routed entirely within the optical domain through an extensive multinode network in a manner that allows graceful evolution starting from today’s position of point-to-point optical fibre transmission on some routes only.

In the first stage of the evolution described here, erbium-doped fibre amplifiers (EDFAs) will replace regenerators. Later, existing transmission links within an inner core mesh network are upgraded to use dense wavelength division multiplexed (DWDMD) transmission with each carrier operating within the synchronous digital hierarchy (SDH) to standard interfaces. Next, these interfaces are replaced by transparent wavelength switches and thereby reconfigure the transmission between nodes entirely in the optical domain. Hereafter, the core network can be extended to include rings optically interfaced to the mesh nodes. Ultimately, individual customers could be served directly through passive optical networks (PONs) optically connected to the rings.

2 Opportunities, constraints and assumptions

A critical evaluation of what optics brings to telecommunications increasingly focuses on its interconnect capability, low attenuation and high bandwidth and de-emphasises its processing capabilities when compared with digital electronics [8]. Consequently, it would appear that any solution must somehow exploit these strengths to overcome the growing hurdles currently caused by processing bottlenecks in telecommunication transport networks.

Fibre is already widely used for point-to-point transmission. Regenerators along these routes dictate a rigid data format in the time domain. However, the continued development of the EDFAs indicates that regenerators will be replaced by amplifiers and fibre links will become transparent lossless data pipes. Fig. 1 reveals that single-
mode fibre in conjunction with the EDFA would provide a transport layer with a transmission window of about 35 nm or 4700 GHz in spectral extent, within which it is possible to wavelength division multiplex (WDM) optical carriers. Each of these channels could conceivably carry an STM-4 or STM-16 signal. Even though gradually more of the bandwidth will undoubtedly be exploited as the demand for capacity increases, it seems improbable that one could contemplate fully filling this potentially vast capacity. For example, if multiplexed STM-4 channels were spaced by 10–50 GHz, full loading would imply up to 400 × 600 Mbit/s = 0.24 Tbit/s per fibre.

Furthermore, this communication space is the same in any fibre in the network, whether it is part of a trunk line or the last 10 m from the curb to a house. This contrasts very strongly with the networks of today, although its benefit is less obvious. The grey shaded plane in Fig. 2 represents the communication space in a single fibre. It is possible to wavelength multiplex and one can therefore think of the frequency dimension as divided into bands each representing a transparent optical channel. In addition, time may also be partitioned analogous to conventional time division multiplexing (TDM). Combined, the two result in a matrix of pixels, each of which may be described by a unique (A, T) reference within the fibre. Thus each pixel within the communication space might correspond to one optical STM-1 channel. A third dimension is added to accommodate the possibility that more than one fibre may exist between any two points in a network. In contrast to the dimensions of the time-channel plane, which are the same at any point in any optical network, the extent of the fibre dimension may vary through a network. For instance, the feeder in the local access part of the network may only use a single fibre, whereas there may be ten or more fibres between pairs of core nodes. Moreover, to maintain all optical transmission paths which require a minimum of electronic control and supervision, the only type of operation that can be permitted when switching is to shift the time-wavelength plane (or parts of it, say one pixel) within this dimension, i.e. cyclic space switching of the individual wavelength channels.

A node in a mesh network will generally have several input and output fibres connecting it to a number of different nodes. Given that nodes have the ability to separate the individual wavelength slots on each of their input fibres and reassemble them in a different configuration on their output fibres, we use the fact that the output fibres lead to different nodes to route wavelength channels through the network. Assuming further that the amount of time required by the node to reconfigure the mapping of wavelength slots is relatively small compared to the degree of time division, each of the pixels in Fig. 2 may be thought of as a temporary transparent data channel that can be used to make an arbitrary point-to-point connection in the network. As we are proposing all-optical switching the pixels must maintain their position within the time-wavelength planes of the fibres, from the input to the output of a node and indeed throughout the whole multinode network. Fig. 3 shows a hypothetical all optical mapping for a node with three input and three output fibres each filled (indicated by the shaded pixels) to half of the theoretically available capacity.

Underpinning all our proposals is the observation that the communication space at any point in an all-fibre network is the same. It is therefore viable to inject an optical signal directly into the wavelength region of the communications space at the line rate chosen for long-haul transmission be it equivalent to a communication rate of 64 kbit/s or 2.4 Gbit/s, so that conventional multiplexing is completely avoided.

3 Towards a national transparent optical network

Given that one does not wish to use all of the communication space simultaneously, we have previously shown [9] that it is possible using the approach outlined to route through a complex multinode network, typically having the dimensions of the UK core network. The following discussion examines the implications of extending the optical domain outwards towards the customer, using networks structures assuming ring and PON topologies. Once fully developed, we believe such a network could ultimately facilitate routing of signals from customer to customer entirely within the optical domain [10].

A schematic representation of a network of this type is shown in Fig. 4. The inner core (the highest hierarchical level) of the network is a highly interconnected mesh. Each of the core nodes provides routing within the inner core, as well as access to one or more ring-type networks.
which make up the outer core. These are in turn optically connected to the peripheral terminals via PONs. Naturally, the rings and PONs we are proposing operate with very different protocols to conventional ring and PON networks, such as FDDI rings and T-PONs. This architecture allows the network to be developed in incremental steps from the top down using a common set of components throughout. In general, this will ensure that maximum benefit can be derived by spreading costs among users at the higher network levels, and reductions in component costs, owing to mass production of standardised modules at the subsequent lower levels. Although the following discussion and analysis assume a network architecture including mesh, ring and PON networks arranged in the configuration described, the underlying principles for the three proposed types of network remain the same for them individually. Hence, any combination of the three is possible.

It seems desirable that our approach should start from a position of having the maximum overlap with today’s network. In the first stage of evolution, we therefore envisage existing transmission links within the inner core network being upgraded to use D-WDM transmission but with each carrier operating within SDH to standard interfaces. Later, these interfaces would be replaced by transparent wavelength switches which, once rings are added, would facilitate transmission between rings entirely in the optical domain. Eventually, individual customers could be serviced directly through PONs optically connected to the rings. We emphasise again that as our proposed evolution of the whole network relies on a single modular set of building blocks the development can stop at any stage as well as proceed unevenly without affecting the possibility of interfacing it to today’s networks at any point. For example, in the process of redeveloping an existing telecommunication network according to the strategy outlined, some parts of the network can have been extended to PON level while others have only just introduced D-WDM transmission and still others remain totally unchanged.

In a high capacity network carrying traffic multiplexed from many users, one would not expect the mean traffic along routes to fluctuate by large amounts in short times. Accordingly, we propose that the traffic carrying ability in the inner and outer core networks (mesh and rings) should be allocated on a slowly time-varying basis (hours), initially in blocks of one optical carrier at STM-4. The choice of STM-4 may seem low, at a frame rate divided into 16 subframes would on average imply an additional 0.75 km of fibre in each cable termination. Moreover, a 35 nm x 20 ps/nm km x 1000 km = 700 ns (=420 bits or ≈1 ATM cell at STM-4) guard band is necessary between information contained in consecutive subframes to allow for differences in propagation delay caused by material dispersion.

To synchronise the switching of subframes in the core network the propagation time of all sections of fibre should be made integer lengths of the subframe interval. An 8 kHz frame rate divided into 16 subframes would on average imply an additional 0.75 km of fibre in each cable termination. Moreover, a 35 nm x 20 ps/nm km x 1000 km = 700 ns (=420 bits or ≈1 ATM cell at STM-4) guard band is necessary between information contained in consecutive subframes to allow for differences in propagation delay caused by material dispersion.

The functionality of the core nodes (represented by bullets in Fig. 4) is shown in Fig. 5. Each is a reconfigurable wavelength space switch (without wavelength or time shifting) of the type described with reference to Fig. 3 and Reference 9. Each thin line indicates the available paths for each wavelength channel independent of the others. Switching between rings and core nodes, as well as from a ring to the core and vice versa are all required.

If $M$ is the total number of fibres entering (and leaving) each node ($#_e + #_m$ of mesh fibres), $w$ the number of active/used wavelength channels per fibre and $F$ the number of subframes at the basic building block rate of 38 or 155 Mbit/s, the node can be seen to be a switch receiving $MwF$ input channels and cross connecting them to $MwF$ output channels. Typically, we envisage $M = 10$, $w = 50$ and $F = 16$ giving a throughput of between 0.3 and 1.24 Tbit/s made up of 8000 independently routed basic broadband channels. As the capacity in both the mesh and the rings is allocated in SFs the individual fibres have equal status at the core node irrespective of whether they are part of a ring or the mesh network. Hence, only one type of core node is specified and it may be used either in conjunction with rings or purely as part of the inner core network switching fabric. As a consequence, rings must also be integer multiples of subframes in circumference as well as unidirectional, to maintain both the timing within the inner core network and their self healing properties [11].

Although capacity may be allocated in blocks of 38 or 155 Mbit/s to a single distant node within and between
I extracted from a PON individually. As a result, it would be possible to transmit information from customer to customer entirely within the optical domain in denominations from an ATM cell upwards.

Consequently, the routing which takes place is based on the individual customers, the network must realistically be capable of providing as little capacity as 64 kbit/s channels. We therefore propose that the subframes be composed of time slots of a duration sufficient for the transmission of standard ATM cells (~680 ns at STM-4 or ~170 ns at STM-16) which could be injected into or extracted from a PON individually. As a result, it would be possible to transmit information from customer to customer entirely within the optical domain in denominations from an ATM cell upwards.

The ring–PON interface is shown in Fig. 6. Similar to the core node, it accesses all the incoming wavelength channels on each of the input fibres, and each of the thin lines indicate the possible paths each of the wavelengths can take. Several input and output fibres may exist on the ring level. However, as all these fibres by definition pass through the same sequence of nodes, no SF → SF switching is required. The real purpose of these switches is to establish access to one or more PONs from the ring. Consequently, the routing which takes place is based on a subframe to ATM cell (C)transformation, or the inverse. The nodes continuously examine optical preheaders which precede every ATM cell (indicating the cells' final destination PON on the ring) and then direct them individually, either to the appropriate PON, or to the ring level output fibre for further transmission down the line of PONs. Depending on the exact configuration of the ring–PON interface, the preheader may be lost in the process, but as no further routing decisions are required before the customer terminal this is of no consequence.

In the cases where a ring is made up of several fibres, the node also ensures that packets injected from the customers in a PON are directed to the correct ring fibre. Connections between customers on different rings are then established by injecting ATM cells into vacant slots in a subframe channel that connects the relevant rings. Since the routing pattern of the subframes is predetermined at the core level, these data channels (pixels within the fibre communication space of Figs. 2 and 3) may be viewed as set of 'postage pigeon holes' each addressing a different destination ring. Consequently, the transmission of information from source to destination is reduced to the insertion and removal of ATM cells at the two involved ring–PON interfaces. Moreover, as these are both strictly localised processes, maintaining precise timing is required only within the individual PONs.

Communication between customers on the same ring is achieved in a similar fashion by allocating additional subframe channels used for local traffic only. If the two customers happen to be on the same PON or on PONs connected to the ring at the same point, the ATM cells simply make one complete round trip before they are dropped back into the PON. Alternatively, an additional level of switching may be introduced, indicated by the dashed lines in Fig. 6, that would allow ATM cells belonging to this category of traffic to be routed directly within the ring–PON node. This would mean that two terminals in the same group of PONs would communicate directly through the ring–PON switch without using any of the communication capacity provided in the ring for local traffic. Indeed, this would not require the introduction of a new level of functionality, as the processing necessary is a subset of that which has already been specified for the other kinds of traffic managed by this type of node.

It is axiomatic that electronic interfaces will always be required at the periphery of the optical ether. Ultimately, the electro-optic conversion might be performed at the subscriber terminal at the end of the PON. However, during the network evolution the wavelength channels seen at the core and ring–PON nodes can also be routed to local optoelectronic receivers so the underlying signals from that point onwards can be carried over today's networks. In line with the development strategy we have outlined, the specification of these optoelectronic interfaces must remain the same throughout the expansion of the network. As shown in Fig. 7, the optics consists of two main components: a tunable narrow line-width laser and a photodetector. A tunable wavelength discriminatory device may also be inserted into the receiving path if one desires the terminal to be able to single out a unique wavelength channel if several appear on the input fibre simultaneously. The two electronic modules are the control circuitry required for the correct operation of the terminal and some high-speed buffer memory. The buffer memory ensures that when an ATM cell destined for the terminal is detected by the optical receiver it may be stored temporarily to allow the data to be read out at the lower bit rate anticipated on the electronic side of the interface. In this way, the only high-speed electronics...
channels available for external traffic in a ring, the condition that \( wqF \geq (10/2) \times R \) must be maintained. For example, if there are a minimum of two fibres in the rings, each carrying 50 active wavelength channels, each partitioned into 16 subframes dedicated to core traffic, the network can accommodate a minimum of \( 50 \times 2 \times 16 = 320 \) rings. Within the context of a national \( 200 \times 200 \) km network this implies rings with an average radius of roughly 30 km assuming a uniform population density across the country or alternatively, with \( 25 \times 10^6 \) network customers, it implies ~78,000 customers per fibre.

Although we postulated that only part of the time–wavelength–fibre space would be used for transmission at any one time, we also assumed that the connection pattern in the core would be dynamically varying, albeit slowly. Consequently, there is a finite probability that a new channel along a given route cannot be accommodated and may have to be denied when requested. If the network control algorithms do not possess any intelligence and requests for new channels are made and allocated utterly at random, the highest probability for blocking, \( P_B \), arises when trying to make a connection along the longest route where the transmission space, on 7 stages, is one channel short of being utilised to its maximum permitted level. If \( f \) denotes the maximum fraction of the total number of wavelength slots allowed to be active in each fibre (e.g. \( f = 0.5 \) if only half of the wavelength slots are used), and \( q \) the number of fibres in the ith cable along the route, then \( P_B \) is given by eqn. 1, where \( p \) is number of wavelength slots tested [9].

\[
P_B = \left[ 1 - \left( 1 - f^q \right) \right] \times \cdots \times \left( 1 - f^q \right) \times \left( 1 - f^q \right) = \left( 1 - f^q \right)^p \tag{1}
\]

Assuming that the filling factor \( f \) and the number of the fibres \( q \) is the same throughout the whole of a route for a given connection and all \( (1-f)W \) vacant channels are tested, eqn. 1 may be rewritten as eqn. 2, where \( q, i \) and \( W \) are as defined and \( W \) is the total number of channel slots in a fibre (used and unused).

\[
P_B = \left[ 1 - \left( 1 - f^q \right) \right] \times \left( 1 - f^q \right)^{W-1} \tag{2}
\]

For example, if there are eight and two fibres per cable in the core and ring networks respectively, with 100 wavelength slots and a 50% filling factor in each fibre, the probability for not making a connection between two rings separated by 11 core nodes (longest path in 7 x 7 rectangular mesh) would be \( 5 \times 10^{-28} \) on any particular chosen route.

5 Analysis of access network

The blocking probability above is only relevant when setting up a new broadband core network connection. However, in addition to that, there exists a set of blocking probabilities within the access network which arise from contention between individual ATM cells. Fig. 9 depicts a conceptual view of the interaction between a ring consisting of two fibres and a PON. The pixels resulting from the original partition of the communication space, i.e. the basic core communication channels associated with that particular ring, are delimited by the
bold lines. In accordance with the proposed architecture, these have been further divided into smaller areas within the time dimension to represent the individual ATM cells. As was the case for the capacity available in the core network, these carry a varying amount of traffic (indicated by the shaded areas) depending on the offered load and the existing traffic patterns.

The upper part of Fig. 9 shows a hypothetical switching pattern for the ring-PON interface. The node does not perform any SF → SF subframe switching, nor does it facilitate C → C ATM cell switching between the ring input and output fibres. Its sole purpose is to drop or insert into the predefined subframe structure, single ATM cells. The first type of contention arises as a result of the PON delivering an ATM cell intended for an ATM slot in the ring which already contains another cell. The second potential type of contention occurs when dropping two or more cells overlapping each other in the time domain from the ring into the same PON. If the dropped cells have identical carrier frequencies or if the receivers in the terminals are not wavelength selective the cells will be indistinguishable and their content therefore lost.

The third type of collision is illustrated by example in lower part of Fig. 9. It also arises from the difference in the extent of the fibre dimension in the time–channel–fibre space which will generally exist between a PON and a ring in a practical network. Since a PON is normally understood to interface to the ring via a single fibre, two or more ATM slots overlapping in the time domain cannot be filled by terminals in the same PON if they occupy identical wavelength channels, even when they are on separate ring fibres.

To further quantify the impact these three general types of contention have on the network, we will present a general analysis of how potential collisions effect the performance as well as the network control required. To avoid injecting ATM cells into already filled slots, the network control must keep track of which of the slots in the various data channels in a ring are in use at any given point in time and only make those that are empty available for injected cells. Given that a ring consists of two or more fibres the control algorithm must also verify that two terminals in the same PON are not competing for identical ATM slots on different fibres as this will result in the cells colliding within the PON.

The control issues associated with both of these two types of cell loss may be administered by the local hardware which oversees the dialogues between the individual PONs (terminals) and a ring. However, these control centres would not store sufficient information about the whole network for a terminal transmitting ATM cells to determine whether the receiving terminal is available, or if they will be in contention with other ATM cells originating from terminals in other rings when they are dropped into the PON. The control required to minimise and register packet losses due to these far-end contentions can also be implemented as processes localised at the individual destination ring–PON interfaces. However, if any potential contention is not predicted at the time of initiating a call, an unduly high cell loss of this type could lead to high retransmission rates and consequently much diminished network performance.

The probability that exactly \( x \) ATM cells selected at random from \( W/f_{q,R} \) are destined for one of \( N \) PONs \( P_x \), eqn. 3, is given as the probability that \( x \) cells are going to a single PON times the probability that the other \( (W/f_{q,R} - x) \) cells go to the remaining \( (N - 1) \) PONs, where the 

\[
P_x = \left( \frac{1}{N} \right)^x \left( 1 - \frac{1}{N} \right)^{W/f_{q,R} - x}
\]

(3)

For example, if a ring has the dimensions used above and the ATM cells transmitted in a certain time slot are distributed between 500 PONs (which in the context of a network with 25 \( \times 10^6 \) customers and 320 rings implies ~150 customers per PON), the likelihood that a PON does not receive an ATM cell when all ATM slots are filled, i.e. \( f_R = 1 \), is \( P_0 = 0.819 \). In other words, if the receivers are not wavelength selective, an ATM cell transmitted with a random wavelength–time reference has only an 82% chance of arriving.

Clearly, this is unacceptable as it stands. However, if the terminals are configured in a way such that any message consisting of more than a few ATM cells would all be transmitted with a consistent wavelength–time reference it would be possible to avoid cell losses by polling the receiving PON for a noncontentious reference before transmitting any message cells. Communication between terminals would then only require the involvement of the control centres directly affected, i.e. generally only two. Moreover, since the contention control procedure remains localised to the individual PONs the network would lend itself to distributed control concentrated, for example, in each ring–PON node. In this case, the probability for making a connection is given by eqn. 4, where \( f_{q,R} = (q - 1)/N \) expresses the probability of choosing a contentious ATM slot in the transmitting PON and \( q \) is the number of different ATM slots tested:

\[
K = \left[ 1 - \frac{1}{N} f_{q,R} (q - 1) \right] \left[ 1 - P_0 q \right]
\]

(4)

For example, consider two rings that each have 500 PONs and two fibres, with 50 active channels per fibre. If they have only one broadband (38 Mbit/s) channel, operating at 99.6% of full capacity, connecting them, a new 64 kbit/s line may on average be established 96.7% of the time. Alternatively, if the same rings only had 100 or 50 PONs, the number of customers per PON would increase to approximately 750 and 1500 and the probability for making a given connection would reduce to 59.8% or 24.8%, respectively. It is understood that, if one measures hardware in terms of the number of PONs (and hence size and number of ring–PON interfaces) and performance in terms of the likelihood of being able to make a given connection, one can generally trade one for the other.

To determine the overall blocking of the traffic which arrives in a ring it is necessary to take the traffic arriving from all the other rings in the whole network into account. Ignoring any correlation in the choice of wavelength–time slots that may exist owing to relatively long connection times (which is effectively a worst-case assumption), the distribution of ATM cells appears as a localised random process. Initially, we consider the case where receivers are not wavelength selective. Such terminals are unable to distinguish arriving ATM cells which overlap in the time domain even if they are on different wavelength channels. Therefore the average fraction of successfully inserted ATM cells that are received is given by V eqn. 5, against the offered traffic load can be expressed as the probability that a PON receives exactly zero or one cell with respect to the fraction of requested.
It reveals that, for a ring with two fibres, 50 active wavelengths per fibre and 500 PONs, $V = 0.982$ when $q = 1$. In other words, when 100% of the available ATM slots are requested a nonintelligent control system will on average be able to allocate 98.2% of them without causing any cell losses. For the hypothetical national telecommunications network used as the example previously, this equates to a capacity of (0.982 x 100 x 2.4 x 10^7) / 3 x 10^9 = 3 Mbit/s per customers for 25 x 10^8 customers simultaneously. By comparison, the same network with 50 and 500 PONs per ring, respectively, promise capacities of 2.3 and 1.2 Mbit/s per customer for $f_R = 1$.

The throughput curves $(V \times f_R)$ calculated for the examples with 500 and 100 PONs per ring, do not exhibit maxima. This implies that the throughput will increase continuously in step with the load. Hence, we should not expect it to be necessary to build safeguards to the network control to cope with congestion as is currently proving to be required with conventional ATM networks. However, in the example where the ring only had 100 PONs, the throughput shows a flat maximum around $V = 0.8$. This result suggests that when the number of PONs in a ring is small compared to the address space stored in the inner core (which is a function of the total number of rings and the respective traffic distributions) we should be taken to limit the injected traffic to the maximum level.

Alternatively, contention can be reduced by increasing the usable address space within the PONs by making the receivers wavelength selective. If terminals have the ability to distinguish individual wavelength channels appearing simultaneously on the input fibre, we can remove the restriction of a single ATM cell per PON per ATM slot. Instead, the only state the PON control must avoid is two (or more) ATM cells transmitted in identical wavelength-time slots on different ring fibres (presupposes $q \geq 2$), being dropped into the same PON at the same time. Hence, for every $i$ ATM cells ($i - 1$), must be rejected. Consequently, the mean number of packets lost is given as $L_q$, eqn. 6, where $Q_q$ is the probability that exactly $i$ ATM cells with identical wavelength–time references are destined for the same PON and $q$ is the number of ring fibres:

$$L = Q_2 + 2Q_3 + \cdots + (i-1)Q_i + \cdots + (q-1)Q_q$$

(6)

$$a_q = \frac{WfR}{fR} fR C_q^q$$

(7)

We can now calculate $Q_2, Q_3, \ldots, Q_q$, eqn. 8:

$$Q_2 = \frac{WfR}{fR} \frac{fR}{N} C_q^q$$

$$Q_3 = \frac{WfR}{fR} \frac{fR}{N} C_q^q \times C_q^q - Q_2 C_q^q - Q_4 C_q^q - \cdots - Q_q C_q^q$$

$$Q_i = \frac{WfR}{fR} \frac{fR}{N} C_q^q \times C_q^q - Q_{i+1} C_q^{i+1} - Q_{i+2} C_q^{i+2} - \cdots - Q_q C_q^q$$

(8)

By substituting eqn. 8 into eqn. 6 and simplifying one obtains a new expression $PL = LWfR/q$, eqn. 9, expressing the proportion of lost cells:

$$PL = \frac{1}{WfR/q} \times \frac{q}{\sum_{i=2}^{q} \left\{ (i-1) \left( \frac{WfR}{fR} \frac{fR}{N} C_q^q \right)^{(i-1)} \times C_q^q - \sum_{k=i+1}^{q} Q_k C_q^k \right\}}$$

(9)

The three uppermost graphs in Fig. 11 plot the probability of cell loss $PL$ against the offered load $f_R$ for rings with 100 wavelengths slots per fibre, two fibres, a generic
justify the additional complexity and cost required at each receiver per time slot. The new cell-loss probability may be exhibited a maximum cell loss probability ranging from $10^{-6}$ to $2 \times 10^{-7}$. As the three lowermost terminals and their individual ring-PON interfaces. Cells which would normally be discarded at these ring-PON nodes could then be reduced the probability of cell losses simply by reducing the probability of 1.8% blocking), it seems difficult to justify the additional complexity and cost required at each receiver if the sole objective is to increase throughput by 1.8%. However, a 'far-end' cell loss probability of $10^{-6}$ approaches the $[10^{-8}, 10^{-12}]$ minimum requirement normally quoted for ATM [13]. A cell-loss probability around or below this limit would mean that it would be viable to configure the network as a pseudo ATM-type network, where the only pre-emptive content control necessary would be between transmitting terminals and their individual ring-PON interfaces.

One method of reducing the blocking probability would be to have additional receivers at each concentration of ring-PON interfaces. Cells which would normally be discarded at these ring-PON nodes could then be received and later retransmitted. In this case, the cell loss probability is reduced to the probability of having to discard no more than one ATM cell per additional receiver per time slot. The new cell-loss probability may be calculated using the Poisson distribution as the probability of the additional terminals not receiving more than one ATM cell per time slot [14]. As the three lowermost graphs in Fig. 11 show, with only one additional terminal situated at each of 50 ring-PON nodes, the network would exhibit a maximum cell loss probability ranging from $2 \times 10^{-6}$ to $2 \times 10^{-12}$ depending on the number of attached PONs.

Moreover, a comparison of $V$, eqn. 5, and $L$, eqn. 9, reveals that $V$ depends on the total number of used ATM slots and the number of PONs, whereas $L$ only depends on the filling factor within each fibre, the number of fibres and the number of PONs. This means that in the case of wavelength selective receivers one is able to reduce the probability of cell losses simply by reducing the filling factor $f$, even if at the same time one increases the number of wavelength slots $W$ to maintain the same overall capacity. For instance, if a ring has 500 PONs, $f = 0.1$ and $W = 500$, i.e. ~10 GHz channel spacing, we would maintain 50 active channels per fibre overall and we would expect the maximum cell loss probability to drop from $10^{-6}$ to $2 \times 10^{-7}$.

6 Conclusions

The wavelength routing strategies in conjunction with the evolutionary considerations described in this paper represent a scalable transparent optical network suitable as a national telecommunications network. The network dimensions used as examples were chosen to coincide with networks of a size comparable to that required to cover the UK. It has been discussed how the proposed architecture permits users to reap full benefits of all optical transmission but at the same time allows gradual implementation starting from a point comparable to many of the networks existing today. The three components found necessary to evolve such a network are:

(i) a reconfigurable optical wavelength space switch
(ii) an all optical ring-PON interface
(iii) a standard optoelectronic line-card.

The main constituents of these three components are already available today although not in the configuration presented here. However, as the architecture is modular and hierarchical it can benefit from economies of scale, given that a standard specification for the components can be reached. Whether to include wavelength selectivity in the receiver specification or indeed retransmission facilities etc. requires an analysis of the size and characteristics of the individual network. The examples given in the previous Section indicate that for different networks there exist strong arguments for the different options. However, with respect to the technical and control implications, different solutions can coexist within one network and these are therefore issues which can be resolved on a PON for PON (ring for ring) basis. Since it is likely that a given PON (ring) will provide access to users, all with much the same demands, e.g. business against privateor telephony against data, in practice one might expect different solutions to be employed within the same network, depending on the types of customers in the individual communities.

The analysis presented has mainly focused on the fully implemented network (including mesh, ring and PON). However, the expressions developed are equally valid and pertinent for the individual parts of the architecture. At such, the configuration explored is the worst-case scenario since it requires the integration of all three types of network. Alternative architectures that only make use of a subset of the features discussed would in general be expected to show relatively better performance and lower complexity. In conclusion, we note that although the basic premise of the design has been based around STM-4 (or STM-16) and the EDFA transmission window there are no fundamental restrictions in the architecture that prevents it from being applied to other technologies should they mature in the future.

The underlying principle of the proposed network architecture could also be applied to a communication space with a single-wavelength channel with 1600 subframes, although this would imply processing and transmission rates 100 times faster than proposed (private communication, Prof. N.J. Doran, University of Aston).

7 References


IEE Proc.-Optoelectron., Vol. 141, No. 5, October 1994


SONET add-drop multiplex equipment (SONET ADM) generic criteria for a self-healing ring implementation', Bellcore Technical Advisory, TR-TSY-00496, September 1991, issue 2, supplement I


As regards potential implementation (see part 1), it is the author's opinion that the advent and maturity of optical devices is such that a national 'optical fibre' would not be unfeasible, perhaps within the next generation of networks, say 30–50 years (two to three decades)...

The implementation of optical transmission systems, however, although already reached great maturity and are still rapidly advancing might be hard pushed to stretch to a national network, certainly within a country comparable to the U.K. In chap. 1 we presented a method of analysis and optimization of optical WDM transmission systems. The results showed that for a network where the distances are separated by a few km, transmission up to 3 Gb/s and km is very much possible in theory, although this would leave no margin of specification or actual implementation. Continued work within the department [4–6] seems to have indicated that a channel spacing around 70 GHz separation would be more practicable for transmission up to about 3–4000 km [7].
6.4 Summary and conclusions

As mentioned in the “Introduction” (section 6.1) the objective of the work presented in this chapter was to examine the potential of optical networks. We aimed to investigate without prejudice whether it was possible to utilise the strength of optics, i.e. transmission and vast bandwidth, to design a network which overcame photonic devices antipathy to processing. The concept(s) developed, SF-WDM, would seem to do exactly that, namely, trade part of the (enormous) potential transmission space for reduced processing while maintaining a high QoS, i.e. low blocking probability. In section 6.2 we examine the modularity and scalability of the proposed approach and in section 6.3 a possible evolution towards an ‘optical ether’ with a starting point of today’s networks and technology. Although both these sections discuss the offered concept and ideas within the context of a national network the author suggests that a shorter term possibility for actual implementation might be within only a single network layer such as a Metropolitan Area Network (MAN) or a regional transport (or core) network. The examination of a national network is more appropriately viewed as an exercise. As already stated, the results presented naturally apply to each network layer individually but perhaps more importantly the analysis illustrates how optical networks could be connected across network layers (or hierarchies) such as form the national to a metropolitan or a metropolitan to an access network. This could also include a boundary due to different network topologies such as mesh to ring or ring to PON.

As regards potential implementation (refer part I), it is the author’s opinion that the diversity and maturity of optical devices is such that a national ‘optical ether’ would not be unthinkable, possibly within the next generation of networks, say 30-50 years (see chapter 2).

The implementation of optical transmission systems, however, although have already reached great maturity and are still rapidly advancing might be hard pushed to stretch to a national network, certainly within a country comparable to the UK. In chapter 3 we presented a method of analysis/optimisation of optical WDM transmission systems. The results showed that for a system where the channels are separated by approximately 50 GHz, transmission up to about 1000 km certainly was possible in theory; although this would leave no margin of specification or actual implementation. Continued work within the department [4-6] seems to have indicated that a channel spacing around 70 GHz separation might be shown to be more practicable for transmission up to about 3-4000km [3].
Indeed, the recommendation published by the ITU (in 1995) is 200 GHz separation although a *de facto* standard based around 100 GHz separation seems to be emerging within the telco industry and purchasable systems.

Consequently, it is the author's opinion that the concept of SF-WDM transmission systems might (at least for the moment) be most readily applicable to regional and (small) national trunk transmission networks.

Subsequent work within the department has as a matter of fact confirmed that optically routed WDM network have a tendency towards being generically sparsely filled. From the work underlying [7, 8] it can be shown that random network graphs fully connected through WDM routed Wavelength Selective (WS), i.e. no wavelength translation and tandem nodes, optical overlay networks tend towards a natural sparsely filled requirement (to avoid wavelength collisions). Indeed it has been shown that in a large number of randomly generated network graphs, the trend exhibited for a 'natural requirement' for sparse filling tended towards a (limited\(^1\)) normal distribution with a mean wavelength utilisation of 50%, corresponding to the suggested filling factor value of \(f = 0.5\), see section 6.2 & 6.3. It should be noted, however, that the quoted work [7] 'ignores' the possibility for fibre multiplicity, which in fact has been shown [15] to be one of the domineering factors when comparing the performance of WDM routed AON architectures. Nonetheless, this result would seem to indicate that there naturally within all (but the singular) WS WDM routed AONs is some degree of wavelength non-utilisation (or sparse filling) necessary in order to avoid contention between wavelengths.

Furthermore, the author would like to draw the reader's attention to the fact that similar & additional work published subsequent to the two papers presented in sections 6.2 & 6.3 analysing blocking probability [9-11] and examining the benefit of introducing the capability of wavelength translation at intermediary wavelength routing nodes in AONs [12-17] seem to demonstrate only limited benefit in networks with Wavelength Interchanging (WI) capability. Typically only around 5-15%. Indeed the author would like to suggest that if normalised by the additional *economic* cost of implementation, WI would seem to only offer advantages in cases where the network implementation or topology is restricted by other, e.g. natural or evolutionary, circumstances.

\(^1\) By a *limited* normal distribution we mean that wavelength utilistion, \(f\), naturally can not be less or more than 0 or 100%, i.e. \(0 \leq f \leq 1\) (rather than \(-\infty & +\infty\), as required by the definition of a 'real' normal distribution).
The work presented in sections 6.2 & 6.3 resulted in 8 claims under a patent application filed by the studentship co-sponsoring organisation, BNR Europe Ltd., with the author (Martin Sabry) and John E. Midwinter as inventors [18].
6.5 Bibliography


6.6 References


[3] Derek Rothnie, private communication


Chapter 7

Resource Control & Management in Broadband Telecommunication Networks

7.1 Introduction

Traditional telecommunication networks, such as the terrestrial telephony network, are designed to support users with homogeneous and simple Quality of Service (QoS) requirements. It provides essentially only a Constant Bit Rate (CBR) circuit switched service to typically two types of customers, i.e. business and residential. This type of session and each of the classes of customer conform to distinct but well established patterns. Future integrated (broadband) communication networks, however, are expected to support a multitude of communication services from a broad selection of session and QoS metrics. Consequently, it is expected that more advanced session and network control & management functions will be necessary.

In tomorrow's telecommunication world we expect a multitude of Network Operators (NOs) and Service Providers (SPs) serving the customers. The customers will in turn demand a multitude of distinct services, which will require many different types of session as well as different Qualities of Service (QoS) characteristics, such as delay and data loss probabilities. In the context of, for instance, an ATM network these may be specified by e.g. the Cell Delay Tolerance (CDT) and Cell Loss Ratio (CLR) parameters. Such a network will require a plethora of diverse control & management mechanisms and functions dealing with issues as diverse as traffic session establishment to physical maintenance and updating of the network.
In general telecommunication resources' performance is a (typically non-linear) function of their utilisation. In the context of a network, resources could be the basic network elements such as switches, communication channels and buffer memory or refer to the aggregation of devices in for instance the administrative domain, such as QoS parameters of individual sessions. Components of data transmission in network communications can also be potential bottlenecks that may cause congestion in the network. They, therefore, demand control not only individually but also as a whole in order to ensure optimal network performance.

### 7.2 Congestion Control

One important area will be the allocation and re-allocation of resources in conjunction with admission as well as flow control. Although the aim of these mechanisms should be to maximise the throughput of data, they should do so without causing excessive delay or ultimately loss of information. Intuitively, it seems quite reasonable to state that this is best achieved by distributing the demand on resources so that the traffic load is spread evenly. This must be done either through restricting access, i.e. admission control, or choosing to only use resources with partly or completely unused capacity, i.e. flow control.

If this is not achieved the performance of the individual devices and therefore in turn the network as a whole will deteriorate as the load is increased. Figure 7.1 on the facing page depicts the typical performance of a communication device or sub-system such as a network queue versus the traffic load which it is offered.

Initially the Throughput — the proportion of the arriving traffic which is delivered, i.e. served — increases linearly in step with the offered traffic. However, as the traffic Load is increased still further the controller part of the queue will be spending an increasing proportion of its time processing data, to serving it. Hence, the gradient of the Throughput (with respect to the offered Load) will decrease and eventually tend towards zero. This is the point of 'on-set' of congestion, i.e. when the controller spends an equal proportion of its time processing and serving data. If the offered Load is increased still further, the Throughput will not increase; the Throughput has reached a maximum and any increase in the volume of offered data will not cause any more data to be served. Depending on the (congestion) control mechanisms in place it will be possible to operate the queue at this point of maximum Throughput for a range of offered Load, until the point of network collapse.
is reached. The point of maximum *Throughput* is when the controller is spending and equal amount of time serving and rejecting traffic. Any further increase in the offered *Load* beyond this point will cause the *Throughput*, i.e. the volume of served traffic, to drop as the controller will then spend a greater proportion of time rejecting traffic than serving it.

The *Round-trip delay* (or *response time*) follows a similar pattern. At first the response time increases linearly (at a low rate) until the ‘on-set’ of congestion at which point it still continues to increase but at an increased rate.

Telecommunication network and device control mechanisms typically aim to minimise the *Round-trip delay* (or response time) and maximise the *Throughput*. To get a quantitative measure of resource/network performance we define the resource *Power* as the ratio of the *Throughput* to the response time. The resource/network *Power* reaches a maximum at the ‘on-set’ of congestion. In other words, a network’s performance is optimal when it’s devices are operated at (or just below) the ‘on-set’ of congestion. This should thus be the objective of any network (device) control & management scheme.

![Figure 7.1: Network performance versus offered traffic load](image)

Traditional congestion control schemes can be categorised as either congestion *avoidance* or congestion *recovery* strategies [1]. Congestion avoidance schemes are supposed to be preventive in nature. Their aim is to keep the network operating at
(or as near as possible to) the point of maximum Power. As their name indicates, congestion recovery strategies 'only' aim to restore the network (or device) to the normal range of operation, i.e. where \( \frac{\partial \text{Throughput}}{\partial \text{Load}} > 0 \), after congestion has occurred.

Conceptually avoidance is preferable to recovery. However, even the best avoidance system requires a mechanism for recovery in the event an un-predictable disturbance occurs. Otherwise the network might not be able to recover from, say, a sub-system collapse caused by e.g. equipment failure. In practise whether to employ avoidance or recovery is strongly dependent on the details of how the individual schemes are implemented and ultimately how close to the \( \frac{\partial \text{Power}}{\partial \text{Load}} = 0 \) condition it is possible to operate the system (or sub-system). For instance, it may be preferable to operate a congestion recovery system, to an avoidance system, if the control mechanism is very responsive and it is, therefore, possible to operate the system closer to the set point (sp) of maximum Power even if the error signal is negative, due to the nature of the control mechanism — recovery rather than avoidance.

### 7.2.1 Congestion Control Schemes and Control Theory

From the above discussion it is clear that there are many similarities between conventional control systems and congestion recovery schemes. In [1], the taxonomy based on control theory terminology depicted in figure 7.2 is proposed for the classification of congestion management schemes.

![Figure 7.2: Taxonomy for congestion control algorithms](image)

Figure 7.2 on the facing page classifies a number of congestion management schemes proposed in the literature according to the taxonomy summarised in figure 7.2.
Figure 7.3: Classification of existing congestion control schemes
7.2.2 Timing

As also can be understood from figure 7.1 on page 179, another aspect of congestion control will be timing issues. In the first instance there will be a wide spectrum of response times. Figure 7.4 is based on the ITU-T recommendations [24], and depicts the traffic management & control functions in relation to the time frames of information flow. As an example, we can see that it is suggested that the time frame for Feedback controls, such as the reduction of the information volume offered by a source, in e.g. an ATM ABR session or a dynamic charging system, should be the same order of magnitude as the Round-trip propagation time. In the case of dynamic charging, this implies that the price will probably tend to vary on a time frame comparable to the round-trip propagation (including processing and control) time.

![Figure 7.4: Congestion Control Function Response Time](image)

7.3 Charging

Inherent in telecommunication networks is the possibility that there may be times and places when capacity is scarce. A major role of prices\(^1\) is to present information

---

\(^1\)A price is the amount the entity is charged for the ‘privilege’ of (attempting to) complete an action.

182
to the market players or entities, i.e. NOs, SPs & customers, about the implications, i.e. the cost\(^2\), of their actions. When the bandwidth available far exceeds the demand, there is little or no role for economics. However, in the more typical situation, where this is not the case, if prices accurately reflect cost, a user can assess and compare the benefits of her actions to the cost of them and make informed decisions about their (potential) value\(^3\).

In addition to, the more conventional association, of recovering the costs (and making a profit\(^4\)) of the services provided, prices can also be used to prioritise usage of resources, if congested; and thereby be used to allocate different classes of service for different actions.

A major (and often neglected) issue associated with charging is the accounting and administrative cost of operating the mechanism(s) in place. The cost of a poor charging mechanism could be so great as to impede the technical innovation, usage and evolution of the network. The value of any particular charging mechanism depends not only on how well it is designed and functions & operates but also on whether the benefits it provides exceed the accounting and administrative costs, i.e. whether profit > 0.

**The cost of Service Provision**

The cost of service provision in telecommunication networks is dominated by the capital costs tied up in the infrastructural resources. For example, 80\% of the budget of NFSNET is allocated to pay for the rental of the transmission trunks and routers, whereas only approximately 7\% of the budget is used for on-going costs of keeping the network running [25].

Although the NFSNET is different from many other networks, in that the NO does not own the infrastructure used (but rents it), the above statement is equally valid for other networks [26]. By far, the majority of the budget for most telecommunication networks is allocated to paying for the cost of the infrastructure; if not directly, through capital depreciation. The incremental cost of operating for servicing additional traffic is negligible (up to the capacity of the network).

However, care should be taken that costs are not increased disproportionately due to the cost associated with the efforts spent accounting and billing. On the other...

\(^2\)The cost of an action is the effort required in order to (attempt to) complete it.

\(^3\)value is a measure of the 'worth'/benefit of any action taken.

\(^4\)profit = price – cost
hand, costs may decrease because the charging mechanism increases the efficiency of the network functioning beyond the price of doing so, i.e. profit is increased.

**Incremental Social cost**

When a network resource is operated near its maximum capacity, new entities who wish to use the resource will inconvenience and increase the cost of the existing users through greater delay and/or data loss, see figure 7.1 on page 179 and section 7.2.

For example, if the time it takes Mr. X to transfer a file is increased, i.e. delayed, by one minute because Mr. Y is also using (and congesting) the network, then Mr. Y’s cost should also include the value (and possibly the price) of one minute of Mr. X’s time. Congestion cost of this type should be counted as part of the social cost of increasing network traffic.

If the industry is competitive (or effectively regulated), then revenues will approximately equal costs, which in turn will maximise profits and social welfare and consequently the evolution and performance of the telecommunication industry.

### 7.3.1 Charging Parameters

'Next generation' networks will support a wide range of telecommunication services, a number of different types of session, such as NRT-VBR and CBR, and a large number of administrative and maintenance functions. We consequently expect a greater number of system parameters will be available for inspection. We believe that it is not unlikely that many or several of these will be used for charging.

We divide them into two main groups: those associated with the access network and with the core network.

**Access Network**

By the access network we mean the proportion of the network between the network termination element at, or immediately before, the customer premises and the first telecommunication network resource.

The charging parameters available in this part of the network are related to the characteristics of the equipment located between the customer and the first telecommunication network resource controlled by the NO. They include type &
is reached. The point of maximum Throughput is when the controller is spending and equal amount of time serving and rejecting traffic. Any further increase in the offered Load beyond this point will cause the Throughput, i.e. the volume of served traffic, to drop as the controller will then spend a greater proportion of time rejecting traffic than serving it.

The Round-trip delay (or response time) follows a similar pattern. At first the response time increases linearly (at a low rate) until the 'on-set' of congestion at which point it still continues to increase but at an increased rate.

Telecommunication network and device control mechanisms typically aim to minimise the Round-trip delay (or response time) and maximise the the Throughput. To get a quantitative measure of resource/network performance we define the resource Power as the ratio of the Throughput to the response time. The resource/network Power reaches a maximum at the 'on-set' of congestion. In other words, a network's performance is optimal when it's devices are operated at (or just below) the 'on-set' of congestion. This should thus be the objective of any network (device) control & management scheme.

![Network performance versus offered traffic load](image)

**Figure 7.1:** Network performance versus offered traffic load

Traditional congestion control schemes can be categorised as either congestion avoidance or congestion recovery strategies [1]. Congestion avoidance schemes are supposed to be preventive in nature. Their aim is to keep the network operating at
(or as near as possible to) the point of maximum *Power*. As their name indicates, congestion recovery strategies 'only' aim to restore the network (or device) to the normal range of operation, i.e. where $\frac{\partial \text{Throughput}}{\partial \text{Load}} > 0$, after congestion has occurred.

Conceptually *avoidance* is preferable to *recovery*. However, even the best avoidance system requires a mechanism for recovery in the event an un-predictable *disturbance* occurs. Otherwise the network might not be able to recover from, say, a sub-system collapse caused by e.g. equipment failure. In practise whether to employ avoidance or recovery is strongly dependent on the details of how the individual schemes are implemented and ultimately how close to the $\frac{\partial \text{Power}}{\partial \text{Load}} = 0$ condition it is possible to operate the system (or sub-system). For instance, it may be preferable to operate a congestion recovery system, to an avoidance system, if the control mechanism is very responsive and it is, therefore, possible to operate the system closer to the *set point* (*sp*) of maximum *Power* even if the *error signal* is negative, due to the nature of the control mechanism — recovery rather than avoidance.

### 7.2.1 Congestion Control Schemes and Control Theory

From the above discussion it is clear that there are many similarities between conventional control systems and congestion recovery schemes. In [1], the taxonomy based on control theory terminology depicted in figure 7.2 is proposed for the classification of congestion management schemes.

![Figure 7.2: Taxonomy for congestion control algorithms](image)

Figure 7.3 on the facing page classifies a number of congestion management schemes proposed in the literature according to the taxonomy summarised in figure 7.2.
Figure 7.3: Classification of existing congestion control schemes
7.2.2 Timing

As also can be understood from figure 7.1 on page 179, another aspect of congestion control will be timing issues. In the first instance there will be a wide spectrum of response times. Figure 7.4 is based on the ITU-T recommendations [24], and depicts the traffic management & control functions in relation to the time frames of information flow. As an example, we can see that it is suggested that the time frame for Feedback controls, such as the reduction of the information volume offered by a source, in e.g. an ATM ABR session or a dynamic charging system, should be the same order of magnitude as the Round-trip propagation time. In the case of dynamic charging, this implies that the price will probably tend to vary on a time frame comparable to the round-trip propagation (including processing and control) time.

![Figure 7.4: Congestion Control Function Response Time](image)

7.3 Charging

Inherent in telecommunication networks is the possibility that there may be times and places when capacity is scarce. A major role of *prices*\(^1\) is to present information

\(^1\)A *price* is the amount the entity is charged for the 'privilege' of (attempting to) complete an action.
to the market players or entities, i.e. NOs, SPs & customers, about the implications, i.e. the cost\(^2\), of their actions. When the bandwidth available far exceeds the demand, there is little or no role for economics. However, in the more typical situation, where this is not the case, if prices accurately reflect cost, a user can assess and compare the benefits of her actions to the cost of them and make informed decisions about their (potential) value\(^3\).

In addition to, the more conventional association, of recovering the costs (and making a profit\(^4\)) of the services provided, prices can also be used to prioritise usage of resources, if congested; and thereby be used to allocate different classes of service for different actions.

A major (and often neglected) issue associated with charging is the accounting and administrative cost of operating the mechanism(s) in place. The cost of a poor charging mechanism could be so great as to impede the technical innovation, usage and evolution of the network. The value of any particular charging mechanism depends not only on how well it is designed and functions & operates but also on whether the benefits it provides exceed the accounting and administrative costs, i.e. whether profit > 0.

**The cost of Service Provision**

The cost of service provision in telecommunication networks is dominated by the capital costs tied up in the infrastructural resources. For example, 80% of the budget of NFSNET is allocated to pay for the rental of the transmission trunks and routers, whereas only approximately 7% of the budget is used for on-going costs of keeping the network running [25].

Although the NFSNET is different from many other networks, in that the NO does not own the infrastructure used (but rents it), the above statement is equally valid for other networks [26]. By far, the majority of the budget for most telecommunication networks is allocated to paying for the cost of the infrastructure; if not directly, through capital depreciation. The incremental cost of operating for servicing additional traffic is negligible (up to the capacity of the network).

However, care should be taken that costs are not increased disproportionately due to the cost associated with the efforts spent accounting and billing. On the other

\(^2\)The cost of an action is the effort required in order to (attempt to) complete it.

\(^3\)value is a measure of the 'worth'/benefit of any action taken.

\(^4\)profit = price - cost
hand, *costs* may decrease because the charging mechanism increases the efficiency of the network functioning beyond the *price* of doing so, i.e. *profit* is increased.

**Incremental Social cost**

When a network resource is operated near its maximum capacity, new entities who wish to use the resource will inconvenience and increase the *cost* of the existing users through greater delay and/or data loss, see figure 7.1 on page 179 and section 7.2.

For example, if the time it takes Mr. X to transfer a file is increased, i.e. delayed, by one minute because Mr. Y is also using (and congesting) the network, then Mr. Y’s *cost* should also include the *value* (and possibly the *price*) of one minute of Mr. X’s time. *Congestion cost* of this type should be counted as part of the *social cost* of increasing network traffic.

If the industry is competitive (or effectively regulated), then revenues will approximately equal costs, which in turn will maximise *profits* and *social welfare* and consequently the evolution and performance of the telecommunication industry.

### 7.3.1 Charging Parameters

‘Next generation’ networks will support a wide range of telecommunication services, a number of different types of session, such as NRT-VBR and CBR, and a large number of administrative and maintenance functions. We consequently expect a greater number of system parameters will be available for inspection. We believe that it is not unlikely that many or several of these will be used for charging.

We divide them into two main groups: those associated with the *access network* and with the *core network*.

**Access Network**

By the access network we mean the proportion of the network between the network termination element at, or immediately before, the customer premises and the first telecommunication network resource.

The charging parameters available in this part of the network are related to the characteristics of the equipment located between the customer and the first telecommunication network resource controlled by the NO. They include *type* &
length of the physical access and the maximum bandwidth & the maximum number of simultaneous connections.

The access network charging parameters are important for the subscription charge and in particular the "one time setup charge" component, not the session charge component, refer section 7.3.2 on page 188.

Core Network

The core network denotes the remainder of the network devices and telecommunication resources. The charging parameters in this part of the network are related to the characteristics of the network elements and their performance, rather than their physical attributes. We divide the core network into three planes:

- user plane
- control plane
- management plane

These planes are based on the same set of underlying variables but each of them process them differently to derive their own parameters from them.

The user plane interprets the available charging parameters from the perspective of the individual user. The resulting parameters consequently include: the Time period of e.g. the day, such as working-day or mother's day, week or year, the Duration of the 'call' or traffic session, the geographical Distance separating the entities communicating, the Volume of data communicated, the Trafficdescriptor, including e.g. the peak data rate, mean data rate & the data rate PDF\(^5\), the Quality of Service parameters, such as delay and data loss, the Proportion of data not conforming to the declared Trafficdescriptor.

The control and management planes deal with the administrative signalling between the Operating System (OS) of each device within and between the planes and management of all the traffic sessions taking place within the network, as a whole, respectively. The control plane would, for example, perform the measurement or computation of the QoS parameters, whereas the management plane would decide the response to a change in the WP (Willingness to Pay).

\(\text{PDF} \sim \text{Probability Density Function}\)
In table 7.1 on the facing page we summarise the above discussion with the addition of *entry* and *charged* price in the *management plane*.
Table 7.1: Summary of charging parameters

<table>
<thead>
<tr>
<th>Part</th>
<th>Parameter</th>
<th>Entity Responsible</th>
<th>Special Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access network</td>
<td>Type of physical access</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Length of access</td>
<td>OS</td>
<td>Admission control functions</td>
</tr>
<tr>
<td></td>
<td>Maximum bandwidth allowed</td>
<td>OS, user plane</td>
<td>Limiting the number of on demand connections (control plane)</td>
</tr>
<tr>
<td></td>
<td>Maximum number of simultaneous connections</td>
<td>OS, control plane</td>
<td></td>
</tr>
<tr>
<td>Core network (user plane)</td>
<td>Time period</td>
<td>OS</td>
<td>Event notifications from the different planes to the OS</td>
</tr>
<tr>
<td></td>
<td>Duration of the session</td>
<td>OS, control plane, user plane, management plane</td>
<td>Event notifications from the different planes to the OS</td>
</tr>
<tr>
<td></td>
<td>Distance (location)</td>
<td>OS</td>
<td>Location of involved parties</td>
</tr>
<tr>
<td></td>
<td>Traffic volume sent or received*</td>
<td>OS, user plane</td>
<td>Data volume counting per connection at output links</td>
</tr>
<tr>
<td></td>
<td>Traffic descriptor</td>
<td>OS, user plane</td>
<td>Admission control functions</td>
</tr>
<tr>
<td></td>
<td>QoS</td>
<td>OS, management plane</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proportion of non-conforming data</td>
<td>OS, user plane</td>
<td>Tagged and non-conforming data volume counting at admission controller</td>
</tr>
<tr>
<td></td>
<td>Willingness to Pay</td>
<td>OS, user plane</td>
<td></td>
</tr>
<tr>
<td>Core network (control plane)</td>
<td>Aspects related to the number of bandwidth modification requests</td>
<td>OS, control plane, management plane</td>
<td>Bandwidth request notification from control and management plane</td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of QoS class modification requests</td>
<td>OS, control plane, management plane</td>
<td>QoS request notification from control to management plane</td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of add/drop parties requested</td>
<td>OS, control plane, management plane</td>
<td>Request notification from control to management plane</td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of add/drop connections requested</td>
<td>OS, control plane, management plane</td>
<td>Request notification from control to management plane</td>
</tr>
<tr>
<td></td>
<td>current (instantaneous), mean, std^price</td>
<td>OS, control plane, management plane</td>
<td>The time interval over which the mean &amp; std^price are computed must be relevant to the rate of change in demand</td>
</tr>
<tr>
<td>Core network (management plane)</td>
<td>Establishment of connection</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of bandwidth modification requests</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of QoS class modification requests</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of add/drop parties requested</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspects related to the number of add/drop connections requested</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Re-routeing/protection switching</td>
<td>OS, management plane</td>
<td>Ad hoc for protection switching</td>
</tr>
<tr>
<td></td>
<td>Current entry^price</td>
<td>OS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current charged^price</td>
<td>OS</td>
<td></td>
</tr>
</tbody>
</table>

*It should be noted that there is an issue whether the value of traffic volume used in the charging function should be only the data volume received rather than transmitted, as there may be a (significant) difference between the two due to data-loss, refer section 7.3.2 on page 192.

1Publication of the instantaneous or current price would seem to be of little value, see section 7.3, (especially if we assume that there is a rapid variation in demand) but should be performed to maintain auditability, refer section 7.3.2.

2std ~ standard deviation

3The entry bid and the amount charged is not necessarily the same.
In addition to the 'basic' charging parameters discussed so far, the NO could offer 'extra' functionality, such as e.g. protection switching. This will probably be at extra cost and of additional value and therefore justifiably price?

7.3.2 Charging Mechanisms

In the most generic terms any charging scheme is based on one or both of the following components:-- subscription or/and session charges. The subscription charge may be further divided into a one time setup charge, which is usually fixed and dependent on the infrastructure laid in, e.g. a "normal" telephone line versus an ISDN line; and possibly a recurring rental charge. The latter, although typically thought of as static, is strictly speaking a function of time — although be it an extremely slowly varying function. The session charge, if any, may also be further sub-divided into a once per call access charge and a usage charge which may be based on either volume or duration or a combination of both.

This description, in turn, leads us to a general expression for charging, equation (7.1). The cost or charge of a given 'call', \( C \), is composed of a constant component, \( \alpha \), which denotes the subscription charge and any shorter term access charge, a duration dependent term, \( \beta_1 \cdot t \), where \( \beta_1 \) is the charge per unit time and \( t \) is the lifetime/duration of the specific session and a volume dependent term, \( \beta_2 \cdot v \), where \( \beta_2 \) is the cost per unit bandwidth and \( v \) the volume of data transmitted.

\[
C = \alpha + \beta_1 \cdot t + \beta_2 \cdot v
\]  

(7.1)

This general expression for charge readily lends itself to describing what may be termed the three most basic charging schemes. Namely, (flat rate) subscription only charging, as used in e.g. the NFSNET network, subscription + duration, as is currently the case in most telephony networks, and subscription + volume.

Although, these, by the premise that they charge customers, may be thought of as methods of macro-economic admission & flow control, they are so, only in the most primitive sense. For instance, although (flat rate) subscription charging may be argued to restrict access to a network, i.e. admission control, it does so only at the very highest level. A (potential) customer will decide whether he finds subscribing to the service provided worth the cost or not and the Service Provider (SP) will consequently be faced with either all or none of the customer's traffic. Subscription + duration and subscription + volume charging have the same effect but with the
further refinement that the customer typically will choose to restrict and effectively minimise the duration of their calls or volume of the data they transmit/receive as every unit time or volume of data costs them typically significant extra.

In many conventional telephony networks, where subscription + duration charging is applied, the charging scheme is further refined by introducing time varying rates. The concept here is that as these networks provide essentially only a single service, i.e. telephone calls, to a limited number of (typically two) different types of customer, the traffic patterns will be static and follow a relatively set pattern. In future telecommunication networks, however, where there will be many more different types of services, we might not necessarily expect these traffic patterns to hold true [27]. The implication is that we will probably need other more sophisticated methods of charging to deal with the increased diversity of types of traffic session and services provided.

Other Charging Schemes

Usage-based charging schemes are from an economical stance preferable as they "force" the customer to weigh (at least their own) cost against benefit of usage. From a telecommunication engineering viewpoint, they also increase the network efficiency and functioning. Several usage based charging schemes have been proposed in the literature. Amongst these are:-

- Statistical Measurement charging
- Envelope charging
- Effective Bandwidth charging
- 'Smart Market'/Dynamic/'Free Market' charging

The effective bandwidth and 'Smart Market' / dynamic / 'Free Market' charging schemes are particularly interesting, as they both encourage co-operative sharing within the network.

Statistical Measurement charging In future multi-service communication networks where transmission is based on multiplexed packets or cells it may be administratively intractable to count each and every cell. In [28] Vieroe proposes that the burden of accounting may be reduced by counting only every $T$ cells or packets. Naturally care must be taken when choosing $T$ in order to ensure that the counted data represent a true picture of the 'sampled' source. If $T$ is too big the possibility of
missing a data burst (in a VBR session) becomes a distinct possibility. Furthermore, the auditability or the NO’s accountability to any customer query in this method of charging is not obvious.

**Envelope charging** In the envelope charging scheme [29], the user is given a traffic description statistical envelope, which she must not exceed, and charged according to session duration (and possibly contract violations). A Poissonian statistical envelope is proposed due to its wide applicability, that they are linearly summable and the relative ease of auditability.

**Effective Bandwidth charging** This charging scheme is, as the name indicates, a proposed method of admission control charging based around the concept of effective bandwidth. A signal’s effective bandwidth [30] is a measure of the bandwidth required to ensure, that for a source with a given set of traffic characteristics, the probability of data loss and delay does not exceed a specified threshold. In addition to the required delay and loss probability QoS requirements, the effective bandwidth of a signal depends on the traffic source characteristics, such as peak rate, mean rate and burstiness. For example, the effective bandwidth of a bursty source with strict delay and loss probability QoS requirements may be ten times (i.e. 1000%) the mean bit-rate, whereas it may only be 110% (of the mean bit-rate) in a constant bit-rate source with lower delay and data loss probability requirements.

In the effective bandwidth charging scheme proposed by Kelly [31] the user would be asked to declare the expected mean rate of the source, \( m \), at the time of call admission and will then be charged according to \( f(m, M) \), equation (7.2), where \( M \) is the measured mean data rate and \( a(m) \) & \( b(m) \) are the charge per unit time & charge per unit of traffic carried, respectively.

\[
f(m, M) = a(m) + b(m) \cdot M
\]  

(7.2)

\( b(m) \) is chosen as the gradient of the effective bandwidth plotted against the measured mean data rate, \( M \), at the point where \( M = m \), i.e. where the measured equals the declared mean data rate. This has the effect of encouraging the user to cooperate with the network, share information and be honest about the characteristics of his traffic source as the total charge for a ‘call’ will be minimised the more accurate the user’s declaration of expected mean data rate is.
‘Smart Market’/Dynamic/‘Free Market’ charging  The concept of dynamic charging, also known as ‘Smart Market’ and ‘Free Market’ charging, has been proposed and worked with in the literature, as a mechanism of charging in telecommunication networks. This scheme allows/requires the charge levied on the use of a device to vary in proportion to the degree of utilisation of the resource. That is, as the use of a network resource, such as a switch or transmission trunk, is increased beyond the sp of utilisation (for maximum throughput while maintaining the desired QoS), the price is also increased (and conversely decreased) in order to discourage (or encourage) traffic from entering the network.

This charging scheme may be thought of as a finely tuned macro-economic means of admission and flow control [32, 33] and may be used in conjunction with conventional telecommunication network resource management procedures.

Another strong argument in favour of dynamic / 'Smart Market' / 'Free Market' charging is that the price internalises the externality by making users face the cost and/or burden he or she imposes on the network and its resources and thereby on the other users (and himself). In contrast, under non-dynamic charging a user may consider and weigh only their own costs and benefits from usage. This phenomenon is known by economists as congestion externality (and by ecologists as the problem of the commons). It has been argued, [34] and others, that dynamic pricing is (the most) favourable when considering how the pricing scheme used affects the industry structure, performance and evolution.

Discussion

When evaluating and comparing charging schemes it is possible to do so at many different levels. These include sensitivity, practicality, fairness, predictability, profitability, accountability, complexity and auditability. Each of these issues may be considered from the perspective of the Network Operator (NO), the Service Provider (SP) and the customer or user.

Below we briefly summarise each of these and in table 7.2 on page 195 compare them against the Subscribe (flatrate), Subscribe + duration, Subscribe + volume, Statistical measurement, Effective bandwidth, Dynamic & Envelope charging schemes discussed above, respectively.

Usage Sensitivity  The usage sensitivity is the measure of how sensitive the charging mechanism is to the utilisation of the network resource(s).
Predictability  Is the complete tariff know before a 'call' is commenced. Does the predictability depend on knowing some or all of the call statistics, such as the duration, peak data rate, mean data rate and data rate PDF\(^6\). Is there a penalty associated with under- or over-estimating the source characteristics; and if so, who does it penalise or benefit:— the user, the SP or the NO. Is the charge penalty itself predictable and known when the traffic contract is agreed.

Complexity & Practicality  Complexity and practicality are naturally very closely linked attributes. This is also the case when comparing and evaluating charging schemes. How easy or obvious is it to implement a charging mechanism and the infrastructure required to administer it. What are the administrative and computational implications of realising it. How much memory and processing power is required at each resource; what proportion of the network capacity is required for signalling and how do they scale with e.g. the average node connectivity, the number of nodes and the capacity of the trunks.

Ultimately, what is the cost of implementing a charging mechanism.

Accountability & Auditability  A charging mechanism's accountability & auditability and the ease & accuracy with which this is achieved are also key features of the network management. These aspect of a charging mechanism are also linked to how easy it is to understand & predict and the complexity of establishing & accounting a 'call' session.

A related issue for usage based charging schemes which rely on monitoring the volume of data transmitted, is where to do so — at the source or at the destination. In networks built around packet (or cell) based transmission protocols there may be a significant discrepancy between the transmitted and received volume of data due to packet losses. Consequently, for services, where the instigator, i.e. the subject of the eventual bill, is co-located with the main data source; we are faced with the dilemma of whether to monitor the traffic transmitted, which normally is easier, i.e. less complex, or the volume of data received, which is fairer but comes at a higher administrative cost.

A converse argument against counting only the successful traffic, i.e. the data volume which is not lost and arrives at its intended destination, is the economist's one. Namely, that the cost imposed on the network by traffic that only has a high

\(^6\)PDF ~ Probability Density Function
enough WP to get part of the way to its destination is not zero. In a purely 'Free telecom Market' there may consequently be a case for charging for all traffic, whether it reaches its destination or not, for the cost it imposes on the network. If a data packet (or cell) does not get to its destination, it could simply be viewed as the sender's bad luck, due only to his own 'tight-fistedness'.

**Profitability**  A charging scheme's *profitability* is the measure of how well it allows the NO(s), SP(s) and users alike to maximise their respective goals.

**Fairness**  Fairness is a subjective concept and consequently difficult to quantify or measure. Fundamentally, however, charging schemes' *fairness* might be said to depend on whether they favour the NOs, SPs or customers.

Users' motivation when evaluating charging schemes is probably no more profound than 'getting the most for the least' possible or 'getting what you pay for'. If this is the case, until bandwidth becomes abundant and the need for charging schemes superfluous, usage based billing will probably be preferred as it will ensure users feel that they 'get what they pay for'.

If the user is required to state or predict some unknown traffic source characteristic or parameters this will also affect fairness. Fairness, from the users' perspective, is closely linked to *predictability*, discussed above.

Fairness from the perspective of the NO(s) or SP(s) is also appreciable. The NO(s) and SP(s) will also prefer charging schemes that will allow the revenue to be *predictable*, see above, and offer 'no surprises' but will also give them the possibility to match the *price* charged for the services offered to the *cost* of providing them, see figure 7.1 on page 179 in section 7.2.

In many ways a charging scheme's *fairness* is the combined subjective measure of all the above properties.

**Knowledge Needed**  The *knowledge* needed by the various entities operating within the telecommunications market, e.g. the NOs, SPs and customers, to maximise the *total profit* (or social welfare, see above).

*Knowledge needed* is marked out of:–

- none
- low
- moderate
7.4 Summary

In this chapter we have reviewed the likely required charging & management functions of future (broadband) integrated telecommunication networks. We have discussed the basis of charging mechanisms in general and with reference to a few particular schemes, i.e. the statistical measurement, envelope, effective bandwidth and dynamic (also known as 'Free Market' or 'Smart Market') pricing mechanisms.

The two later mechanisms encourage co-operative sharing. In both schemes the user is rewarded for supplying information about his traffic source to the network management so that it, in turn, will be able to, more efficiently, perform its task and maximise profit (or social welfare). It may be argued that both of these charging schemes 'internalise' the congestion externality. However, in the case of effective bandwidth charging this process is established only for each customer individually. In a truly dynamic 'Free telecom Market', however, this property is maximised as each user is "forced" to value her own requirements within the context of all the other customers.
Table 7.2: Comparison of charging schemes
Bibliography


[26] Figures and info on capital depreciation etc.


8.1 Introduction

From the review and brief discussion of network resource (congestion) control and charging schemes & parameters presented in the previous chapter it seems clear to us that several of the underlying drivers for both of these aspects of network control & management have many points in common. Consequently, we found ourselves (in collaboration with some on-going work at Nortel, Harlow Labs by Paul Kirkby) challenged to consider how the principles of free-market pricing might be used to provide sole control of flow and access to a telecommunications network. In studying this, we have not set our work in a context of ATM or SDH but have considered the problem more broadly, rather more from the point of view of starting with a clean sheet of paper and asking what would limit such an approach from a fairly fundamental point of view. Our preliminary conclusions are set out below.

It is important to recognise at the outset that we have taken the market analogy literally, namely that prices are free to rise and fall without constraint in respond the interplay of demand and availability and clearly this is in dramatic contrast to today’s pricing mechanisms and particularly also ‘dynamic pricing’ as it typically is discussed.
However, by taking such an approach, we believe that it draws attention to a number of fundamental issues.

![Figure 8.1 — Demand versus price relationship](image)

We have assumed that once such an approach were established, the market for telecommunications capacity would develop in much the same way as the market for other commodities. For markets with large numbers of purchasers and commodities, there tends to be a log-log relationship between demand, \( D \), being the number of purchasers willing to pay a given price, and the price of the item, \( P \) such that:

\[
\log(D) = -K \log(P) + M \tag{8.1}
\]

where \( K \) & \( M \) are constants that characterise the given product or service and customer (group) demand-price elasticity.

From the outset and as briefly touched upon in section 7.2, we have been very conscious of the critical issue of response time for the control mechanism. Simple calculation shows that there are some key times that must be taken into account. These are illustrated in Figure 8.2 below and are calculated on the basis that the speed-of-light in glass is about \( 2 \times 10^8 \) m/s, that the round-trip distance within the local access loop is unlikely to be less than 2 km and that the longest round-trip distance by landline is likely to be 40,000 km although involving satellite circuits could extend that still further. Thus we define three regions for a global network as seen from the viewpoint of any given customer.
It is our assumption that a price will be associated with the use of any given resource and that that price should reflect, on a rapidly changing basis, the loading of that resource. However, if this price is to control the access to the resource, it must be known to the sender of the data. Figure 8.2 therefore defines several operating regions for access control by pricing. At the very shortest times, (e.g. sub 1E-4 secs), it is likely to be physically impossible for a customer to have even local 'current price' data. Thus if control is to be exerted at the source of the traffic, it can only be informed by the most recent price data available and that will always be somewhat out of date. The control mechanism must at least take account of this unless constraints are applied to the accepted traffic that severely constrain its temporal rate of change.

At the other extreme, we see that for very long haul circuits (London-Auckland, NZ) the round-trip time is necessarily very much longer and hence the ability for London to know the 'current' congestion status via price or any other metric is very severely constrained. In between these extremes lie a range intermediate situations.

Thinking more generally about the operation of markets, we can observe a number of broad issues that we believe have relevance to the telecomm situation.
In Figure 8.3, time is measured forward from the present. If there were no pre-booking system for allocating bandwidth, it seems logical to assume that the network operator must introduce one simply as part of his forward planning in establishing a working starting-price. However, it is also clear that some users will wish to exploit their own knowledge of future requirements to purchase capacity ahead of the time they wish to use it and to obtain preferential rates as a result. The diagram thus alludes to an ‘options’ / ‘futures-market’ existing in capacity. Users are assumed to plan either no time ahead or a short time ahead while Futures Traders or the Network Operators themselves study the market characteristics, predict ahead likely demand, set forward prices and possibly purchase forward capacity to sell onward nearer the time of use. Capacity could even be ‘oversold’ on the basis that not all of it is likely to be used similar to ‘options-markets’ of more conventional commodities. In the following sections we will suggest some ways in which such pricing structures might be introduced into the access and flow control scenario.

If we wished to extend the allocation of capacity to allow for a greater fraction to be allocated on a spot (instantaneous) basis, then we would clearly need to leave the region labelled “capacity purchased by users” at a lower level for the longer times (sec to minute) leaving a significant fraction available for very short response time purchase. Whether this is practical is another issue.

### 8.2 Free Market Access Control

In trying to formulate the Free Market Admission Control (FMAC) problem more precisely, we have focused initially upon access control in a local loop. The model is thus assumed to be as shown in Figure 8.4 below.
As stated above, the immediate problem with this model is that the measurement of resource utilisation or congestion can only be made at the resource, that the price must therefore be set there to reflect the instantaneous demand versus the utilisation and that knowledge of that price will take a finite time to reach the customer, who even if they were perfectly responsive, could only respond a finite time after the congestion problem was identified. To overcome this objection, we have introduced a number of what we believe to be new proposals.

The control model by the introduction of another element, shown in Figure 8.5 below and labelled a "Fast Reflector Switch" that is colocated with the congestible resource. The elements within the large right hand box now comprise a local and fast acting control loop.

The operation of the control system is customers are flagged with 'most-recent' Price and they launch traffic in the light of that information. The fast reflector switch, being
colocated with the resource, can monitor the status of the resource and accept or reject the traffic on a virtually instantaneous basis. This immediately provides the basis for very effective flow control through the resource but carries with it some other less attractive implications.

If traffic selection is to be done on the basis of Willingness to Pay (WP), which we see as a way for the user to express the priority he attaches to the traffic, then the units of traffic must carry with them an expression of each sender’s WP. However, if a single “most recent price” is quoted, how will such a distribution be generated? In the absence of it, traffic would be selected arbitrarily which seems to serve no purpose.

It must be a characteristic of the control loop above that the price could fluctuate as rapidly as the traffic load or the (2nd, ref. section 9.2) integral of it. If we postulate traffic demand being randomly varying, then we must also assume that the price generated by the control loop will be similar. This immediately suggests an answer to the question posed above. A much better description of price in the above situation would be given by flagging to the user recent Mean Price (MP) and recent Standard Deviation of Price (SDP). The customer would then have a natural mechanism for signalling priority in terms of \( WP = MP + n \cdot SDP \) where \( n > 0 \) would signify increasing priority and \( n < 0 \) would signify low priority. At the same time, we might expect this to generate a distribution of quoted WPs.

The operation of the access control loop now involves the customer deciding on his priority for transmission, setting a WP value using the \( WP = MP + n \cdot SDP \) formula, launching his traffic unannounced into the network and taking pot-luck as to whether or not it is accepted. However, he can materially affect his chances of being accepted by his choice of \( n \). Clearly if all customers choose large \( n \) values, then the MP will rise steadily and in the meantime, they will have data (or a message of non-acceptance) returned to them by the fast reflector switch. Part of the work presented in section 9.2 discusses how we have separately examined the conditions for stability of this control loop.

In considering what data should be exchanged by a customer and the network resource controller, it appears five items of information may be of interest: The type of traffic session initiated, e.g. packet or circuit-switched, Constant Bit Rate (CBR) or Variable Bit Rate (VBR), (& possibly a description of the expected traffic statistics, such as mean & peak bit rate etc.); the Current Price (CP); MP & SDP and an identifier to allow the data that went astray or traffic session that was queried / ‘poled’ for, to be identified and relaunched or initiated, if desired; with the required WP. This would also

---

1 It should be noted that the control theory model proposed and analysed in section 9.2 is for the resource & the utilisation of it only, i.e. ‘after having passed’ the “Fast Reflector Switch”. In other words, it examines the demand on & utilisation of a network resource by customers who have declared a WP priority high enough to not be reflected, i.e. higher than the CP.
allow a customer to determine the current going rate at any point in time for a given service (imminent or booked for the future) by launching a empty ‘test header’ of or ‘resource / service request’ containing the details of the traffic session, e.g. VBR or CBR & QoS metric etc., but with zero WP, which would then be ‘reflected’ with the current appropriate CP, MP & SDP.

One could easily imagine that under some circumstances these ‘poling messages’ (& returned data) could easily occupy a significant proportion of the available communication capacity. Consequently, we believe that a ‘poling overload’ mechanism is essential part of this congestion / charging scheme. The ‘returned data’ would seem to present a smaller challenge in that the network or a resource could decide to omit returning these if it was determined that they were impacting the overall network / resource performance. The launching of ‘test messages’ by the customers would seem to be slightly more tricky as there, as discussed above, at the ‘customers’ side’ of the NE (NT (Network Termination) Equipment) fundamentally appears to no way of locally determining a resource or network utilisation or congestion. However, two solutions to this aspect of the proposed strategy of network control spring to mind: Namely, that the fundamental protocol has built into it a resource utilisation sensitive ‘exponential back-off’ time-delay requirement similar to that of say the CSMA/CD (Carrier Sense Multiple Access with Collision Detection) protocol, so that a customer for a given would be resource or service would be required to wait for a specified length of time before relaunching a failed ‘resource / service request’; or that perhaps quite suitably in this type of network launching a ‘resource / service request’ has associated with a (fixed, predetermined) charge comparable to the cost to the Network Operator (NO) of reflecting it.

Building on the ideas enumerated above, a number of embellishments seem to flow rather readily and these are illustrated in Figure 8.6 below.
The diagram recognises the existence of other criteria that have not been discussed. For example, in measuring resource loading, this must clearly be done against some target Set Point (SP). The choice of SP may be set fundamentally or left adjustable in response to other criteria, perhaps, for example, some Quality of Service (QoS) measures, in a similar manner to the calculation of 'effective bandwidth' [1-3]. For example, depending on the characteristics of the traffic source(s), such as peak & mean bit rate (i.e. burstiness), and the Desired QoS metric the target SP for (mean) network Resource Utilisation, such as buffer occupancy of a given server, could be set at say 50 or 80%. This is alluded to in the lower (and assumed much slower) control loop whose purpose would be to slowly change the SP target loading for the resource in response to other criteria than instantaneous demand.

Following on directly from that comes the Feed Forward (FF) compensator aspect of the control loop. In the first instance, we envisage that this would recognise in advance the approach of, say, busy hour, and start to off-load traffic or increase price in anticipation.

The existence of a FF control system opens up wider possibilities. It is axiomatic that such a loop is driven by prior knowledge of traffic demand. That can include both prediction on the basis of previous experience (e.g. busy hour), data on known events (such as a TV ‘phone-in’ or ‘Mother’s day’) and customer bookings of telecommunication resource(s), e.g. capacity, for use (‘futures’) or availability (‘options’) in the future. It thus opens up a bridge between the very fast but statistical access control we have postulated above that operates on the basis of no insight and no prior knowledge.
and an access control process much closer to that of circuit switching. There seems no obvious reason why each resource should not maintain its own Futures data base and accept reservations against firmly quoted prices on any time frame so long as it is longer than the network response time. Such capacity would thus be sold at a pre-agreed price and would take priority over other traffic. It follows that as the resource utilisation became large, latecomers buying capacity would pay a higher price for the same capacity than early-birds. Such an approach also opens up the potential for telecomm. traders to buy in advance blocks of capacity and to gamble on re-selling closer to the time of use at a profit, opening up a **Telecomm Futures market** opportunity, as also discussed in the previous section and depicted in Figure 8.3. The existence of such a market would have the benefit of setting well ahead of actual utilisation a “market price” for the capacity, albeit a price that would be refined usually upwards nearer to usage.

### 8.3 Provisional thoughts on Free Market Charging in Extended Networks

In the previous section we have developed & proposed how this method of charging and telecommunication network resource management might be applied to the access network; or more precisely, to a *single* service or group of network resources which for the purpose of network control and management is considered one entity. This would seem to a reasonable ‘first stab’ at developing this method of FM charging as CAC (Connection Admission Control) in say an ATM network is likely to be negotiated on a session for session basis, i.e., e.g. for the whole of a VC (Virtual Circuit), even though several individual network resources or components such as switches and transmission lines are likely to be involved.

However, recognising that extended networks are necessarily involved, we believe that some advance may be derived from extending the proposed method of FM charging to control of *multiple* resources, as this might also be used as an ‘internal’ means of resource allocation & usage for a given Service Provider (SP) and/or Network Operator (NO) or NOs.

In considering this method of network management in ‘extended networks’, we have looked at a number of different approaches. Briefly these are stated as follows:-

- Assemble at the entry point(s) to the network a complete array of all most recent prices so that the customer terminal can pre-compute a cost effective route and likely cost. The problem with this approach is that it implies a fairly massive data-flow to assemble such price databases and keep them updated and we believe represents an overkill to solving the problem.
• An even more sophisticated option would be to examine total traffic flow through the network and seek to optimise overall throughput by adjusting prices according to some global pricing algorithm over and above the local pricing algorithm. This appears to be completely impractical in terms of computational complexity.

Accordingly, we have initially homed in on two much simpler pricing algorithms which appear to fulfil many of our requirements. However, it remains to be established whether stable flow control can be achieved in a complex network using them. They are:

8.3.1 $\Sigma${Price} Charging

Here we assume that the user or customer sets her WP according to a formula roughly of the form of that in Eqn. (8.2); namely, that the overall WP as the sum of the WP for the $m$ individual network resources utilised.

\[
WP = \sum_{i=1}^{m} WP_i + n \cdot \sum_{i=1}^{m} SDP_i
\]  

On entry to the network a user would be charged according to his declared WP ('admission cost') and the message data or call circuit would be tagged with a value of $\Delta = n \cdot \sum_{i=1}^{m} SDP_i$ and on transit through the network, this amount would be reduced at each resource to become, at the $m^{th}$ resource transited:

\[
\Delta_m = n \cdot \sum_{i=1}^{m} SDP_i - \sum_{i=1}^{m} (P_i - WP_i)^2
\]  

The reason for doing this is to leave only a small amount of credit available for the message to be used to overcome rising price from an emerging congestion point and to operate this in conjunction with the fast-reflector switch that one only reflects if $\Delta_m$ is negative.

8.3.2 $\max\{Price\}$ Charging

Here we assume that the user is charged according to his WP but that he sets that on the basis of the Maximum value of $P$ on his chosen route and its associated SDP. This effectively says that 'price paid' would be dominated by the price of the most congested element.
8.4 Summary and conclusions

In this chapter we have explored an extended the proposed concept of 'Free Market' charging as applied to telecommunication network services & resources. As stated in the introduction we have done so on a premise of control of generic broadband multi-service integrated network rather than specific to any particular protocol, such as IP (Internet Protocol) or ATM, or charging mechanism, such as dynamic charging.

Furthermore, we believe we have proposed a couple of unique concepts: Namely the “Fast Reflector Switch”, ref. Figure 8.5, and the like emergence of a “Telecomm Futures” market, ref. Figure 8.3, in section 8.8.1.

In the following chapter, we will present some of the initial investigative work undertaken on the proposed method of charging and resource & network control; in particular, we will present (we believe) a unique method of *analytical* analysis of charging and network control & management schemes using control theory.
8.5 References


Chapter 9

Provisional Investigation of Free Market Charging

9.1 Introduction

In this chapter we summarise some of the provisional investigative work undertaken in order to analyse the 'Free Market' (FM) charging & management control scheme proposed and developed in the previous two chapters. The main emphasis of the work & results presented was to evolve the already recognised similarity between network resource management and classical control theory in an endeavour to probe the scope for developing an (to the authors knowledge original) technique for original technique of investigation of control & charging mechanisms in telecommunication networks. In the case of the proposed FM charging scheme, particular attention was paid to formalising and quantifying the intuitive dampening effect of resource utilisation and the criteria of stability. This work was also published at an IEE Colloquium on “Charging for ATM — The Reality Arrives”, 20 Nov 1997, at Savoy Place, London, ref. Section B.4 in Appendix B.

The chapter concludes in section summarising the initial findings and possible avenues for future investigative and analytical work of the proposed network management rhetoric. This area of work was also 'handed over' to another Nortel sponsored Ph.D. student within the department.
A control system based on FB is responsive in nature. In such a system the controlling signal is the *error signal* generated as the difference between the current and desired output, i.e. the $PV$ and $SP$, response. This type of control system are normally depicted by a schematic comparable to that in figure 9.2. It can be seen that the input to the this type of system is still the SP but that the control signal is the difference in the actual output and desired output, i.e. the difference between PV and SP.

**9.2.2 Model**

In this section we propose a control theory model of a single FM operated device. In principle, this could be any type of telecommunication network device such as e.g. a router, a transmission line, a buffer, a switch or any sub-system consisting of two or more individual devices considered as *one* entity from the perspective of control or charging (and the modelling thereof).

Figure 9.3 is a flow diagram representation of the proposed control theory model of a single FM operated device. As discussed above this controlled resource could be
any network device or sub-system but for the sake of explaining and analysing the model we have chosen it to be a memory buffer. One possible aim of network management might be to operate the network resources as close as possible to maximum throughput while maintaining the desired QoS. Consequently the through variable(s) of our control system should be the QoS parameters we would like to optimise for. In an ATM network these could, for example be Cell Loss Ratio (CLR) and Cell Delay Variance (CDV). In addition to these there will be one or more parameters which describe the attributes of the resource. In the case of a memory module this might be the buffer size and the line rate of the server. These, the QoS through variables and the resource parameters, are combined in the Controller into a single (local) control variable, i.e. the OP, which, in this case, could be specified in terms of the percentage or mean number of cells buffer occupancy.

The OP is in turn compared to the actual buffer utilisation, where the difference signal feeds the Pricing function, i.e. the Actuator. For example, if the target buffer occupancy OP has been set at 80% but the actual buffer utilisation is 85% the price or the MV should be increased until 5.9% (= 85−80/85) of the customers in the system change their Intentions and delay making their calls to a later (cheaper) time. Although this may initially seem like a rather Draconian approach it can be shown that it is ultimately the most advantageous for the users, i.e. the customers, the service providers and the network operators alike [37, 34].

The Pricing function may also be fed by a FF compensator, which could anticipate, e.g. busy hour, traffic patterns. Even though, the proposed charging-flow control mechanism is aimed at ‘next generation’ multi-service networks where the current conventional understanding of traffic patterns will not necessarily hold true [27], we might still expect a ‘busy hour’ phenomenon. One only has to look at the Internet where there is no (major) admission or flow control techniques, economic\(^1\) or otherwise, in place but there still exists a busy hour. For instance, an ‘Internet-regular’, living in the UK, would never intentionally access a web site in the States to download a large file in the afternoon. He would do this in the morning (UK time) when the majority of the US population is asleep and there consequently is less congestion and delay in the US proportion of the network.

\(^1\)One could argue that as a noticeable proportion of Internet users access it over telephone lines and PPP or SLIP connections via ISPs the charging rates of some telephone network operators is a form of macro economic admission control. However, as this is not meant as a dedicated method of Internet access and the telephone networks are designed to fulfil a function in their of right we shall discount this minor exception.
Intuitively, we might therefore expect a combined charging mechanism [38] with a slowly time varying mean component based on the expectation of (busy hour) traffic patterns and a rapidly varying dynamic component which would accommodate (and smooth) the variations within an expected pseudo-static time period.

The controlled network, sub-network or, in our case, the memory module would then deliver an actual QoS or one or more PVs which could be measured and compared against the Desired QoS. The Buffer occupancy function could react and possibly modify the OP, i.e. the coveted buffer occupancy, in response to any significant difference between the Desired and Delivered QoS.

With the measurement of the Delivered QoS we would expect a possibly significant time delay, $\Delta \tau$. Consequently an alternative model for an actual implementation of a memory module might be a hierarchical control system [39], with a fast reacting inner loop and a slower outer loop.

### 9.2.3 Analysing Model

For the purpose of analysing the control system, however, we initially propose the (single) feedback loop depicted in figure 9.3. In theory, this could represent either the short time constant buffer utilisation, OP feedback loop or the long time constant feedback loop from the PVs, i.e. the Delivered QoS, to the SP Desired QoS. In the description and analysis following we shall though only refer to the case of the buffer utilisation loop. It should be noted, that in the case of the long time constant feedback loop the through variables, the output or PVs and the SP will be specified in terms of the QoS parameters being controlled (rather than the buffer utilisation). In addition, the control loop’s characteristic equation should incorporate the time delay, associated with measuring the Delivered QoS, between the output (or the PVs) and the Price function.

It can be seen from figure 9.3 that the desired Buffer Utilisation (BU) SP is compared to the actual BU, where the Price function regulates the current price, $P_t$, according to the expression below, where $K$ is the relative weighting of the error signal, $E = SP - BU$, and $\alpha_2$, is the weighting of the price of the previous time slot.

---

^2For a more in depth discussion of the time frames involved with the various levels of this charging mechanism and network control functions in general, the reader is referred to [33] and sections 7.2 & 8.1.
\[ P_t = \alpha_2 P_{t-1} + K (SP - BU) \]

\( BU \) is given by the equation below. \( LR \) and \( ND \) denote the Line Rate and the Net Demand respectively,

\[ BU_t = BU_{t-1} + (ND_t - LR) \]

where \( ND \) is given by the expression below. \( D \) is the control system disturbance, i.e. the Demand, \( k \) is the 'pounds' to traffic data demand converting ratio, e.g. specified in number of ATM cells/£, and \( \alpha_1 \) is the SS ratio of accepted traffic. That is where the receiver or service is busy or traffic is (maliciously) sent with too low a Willingness to Pay (WP) or 'price priority'. Using current telephony networks as a guide we would probably expect \( \alpha_1 = \sim [0.85, 0.975] \). Furthermore, if we were to use British Telecom's local off-peak telephone call rates\(^3\) as guide for the price of a nominal 64 kBit/s information rate, \( k = \sim 4 \times 10^{-6} \text{£/pkt} \). This is in good agreement with [25], where 1/600 cent per packet (or \( \sim 10^{-6} \text{£/pkt} \)) based on the turnover and traffic volume of the NFSNET, is suggested as a guideline price.

\[ ND_t = \alpha_1 (D_t + kP_t) \]

Normally one would use superposition, with \( LR = 0 \), to determine the transfer function of this type of system. It would though be nonsensical to set \( LR = 0 \) and this would indeed result in an improper transfer function. Hence, we re-specify \( D \), \( ND \), \( P \), \( LR \) and \( BU \) in terms of the change in Demand, Net Demand, Price, Line Rate and Buffer Utilisation, \( \Delta D \), \( \Delta ND \), \( \Delta LR \) and \( \Delta BU \), respectively. Now, \( \Delta LR \equiv 0 \) by definition and we can derive the system's transfer function (by superposition).

The resulting expressions relating \( \Delta D \), \( \Delta ND \), \( \Delta P \) and \( \Delta BU \) are given by (9.1), (9.2) & (9.3) and (9.4), (9.5) & (9.6) in the time and z-domain respectively. Table 9.1 summarises the definition of the parameters and suggests the typical range of values.

\[ \Delta P_t = \alpha_2 \Delta P_{t-1} + K \Delta E \]

\(^3\)Of the 4\(^{th}\) of Sep. 1996 BT charged £0.04/min for a standard rate local telephone call.
\(^4\)The exchange rate at the close of business the 12\(^{th}\) of May 1997 was 1.623.
Figure 9.3: Control theory representation of single network resource

\[ \Delta BU = \alpha_2 \Delta P_{t-1} + K (\Delta BU - \Delta SP) \]  
(9.1)

\[ \Delta BU_t = \Delta BU_{t-1} + (\Delta ND_t - \Delta LR) \]  
(9.2)

\[ \Delta ND_t = \alpha_1 (\Delta D_t + k \Delta P_t) \]  
(9.3)

\[ \frac{\Delta P_t}{\Delta BU} (z) = \frac{K}{1 - \alpha_2 z^{-1}} \]  
(9.4)

\[ \frac{\Delta BU}{\Delta ND} (z) = \frac{1}{1 - z^{-1}} \]  
(9.5)

\[ ND (z) = \alpha_1 (\Delta D (z) + k \Delta P (z)) \]  
(9.6)

We note also that as well as \( \Delta LR \equiv 0 \), \( \Delta P = 0 \) in the Steady State (SS) as \( BU \) has been the changed to \( \Delta BU \) and the desired SS \( \Delta BU \) therefore tends towards zero, \( \Delta BU \rightarrow 0 \). Combining (9.4), (9.5) and (9.6) we derive the transfer function of our control system model \( H (z) = \frac{FP}{1 + OL} \), (9.7), as the Forward Path (FP) over one plus the Open Loop (OL) transfer function.

\[ H (z) = \frac{FP}{1 + OL} = \frac{\frac{\Delta BU}{\Delta ND} \cdot \frac{\Delta ND}{\Delta D} \cdot \frac{\Delta P}{\Delta BU} \cdot k}{\frac{\alpha_1 (1 - \alpha_2 z^{-1})}{(1 - z^{-1})(1 - \alpha_2 z^{-1}) + \alpha_1 Kk}} \]  
(9.7)
Parameter | Definition | Possible Range | Typical Range | Comments |
--- | --- | --- | --- | --- |
\(\alpha_1\) | SS accepted traffic ratio | \(0 \leq \alpha_1 \leq 1\) | 0.850–0.975 |
\(\alpha_2\) | weighting of price of previous time-slot | \(\alpha_2 \in \mathbb{R}\) | \(0 \leq \alpha_2,\) simple system \(\alpha_2 = 1\) | \(\alpha_2 = 1,\) denotes the case where the price is stationary |
\(K\) | weighting of (change in) error signal, \(\Delta E = (\Delta BU - \Delta SP)\) | \(K \in \mathbb{R}\) | \(0 < K,\) simple system \(K = 1\) | \(K = 0,\) denotes a system where charging is not dynamic |
\(k\) | '£' (pounds) to traffic, e.g. packets, converting ratio | \(k \in \mathbb{R}\) | \(0 < k\) | \(k = 0,\) denotes a system where there is no usage charge |

Table 9.1: Definition of control system parameters

### 9.2.4 Results & Discussion

By examining the model transfer function \(H(z), (9.7)\), we see that our control system model is in fact a **double-integrator** with two poles and two zeros. Hence, we would expect the output of system to converge towards the desired response with a zero SS error in response to a step or ramp disturbance and a non-zero but finite SS error for an 'accelerating' disturbance signal. In other words, as long as the (change in) demand does not accelerate but 'only' changes in steps (which is probably the most likely case) or linearly, the admission controlling effect of dynamic charging should ensure that the offered traffic volume will tend towards and (eventually) equal the decided SP. Consequently we conclude that according to the model developed in the previous section dynamic charging, given a certain infrastructure and specified social environment, would seem to offer the facilities for maximising throughput, profits and social welfare — provided naturally that the benefits are not consumed by the cost of administration.

Figure 9.4 depicts the system's **pole-zero map** where the position of the poles and zeros are specified symbolically in terms of the parameters of the control system model.
If we chose values for $\alpha_1$, $\alpha_2$, $K$, & $k$ in the Typical Range suggested in table 9.1 with $\alpha_1 = 0.9$, $\alpha_2 = 1$ & $K \cdot k = 1$ we find, see figure(s) 9.5 (& 9.4), that the poles are a complex conjugate pair in the first & fourth quadrant inside the unit circle and the poles are placed at the origin and intersection of the unit circle & real axis, respectively. In other words, dynamic charging (in this configuration) is a stable pricing mechanism and it will exhibit a dampening effect on variation in resource utilisation, $BU$, with respect to the (gross) demand, $D$. That is, we have proven that dynamic (FM or SM) charging is a generic means of (macro-economic) admission control — much as one might hope and/or expect.

Let us now investigate how the performance/characteristics of the system changes as we vary the parameters of the charging mechanism. Figure 9.6 is a root locus diagram of the system transfer function with respect to the $K \cdot k \cdot \alpha_1$ term. That is, it is a diagrammatic representation of the movement of the poles as $K \cdot k \cdot \alpha_1$ changes from 0 to $\infty$. It can be seen that at $K \cdot k \cdot \alpha_1 = 0$ the poles coincide on the zero on
the unit circle and as the term increases they move (as a complex conjugate pair) through the first (and fourth) quadrant, shown by the solid line, until they again coincide at the origin for $K \cdot k \cdot \alpha_1 \to \infty$. Hence, for any value of $K \cdot k \cdot \alpha_1 = [0, \infty[$ the system is stable.

The root locus diagram depicted in figure 9.6 further seems to suggest there for a telecom network resource exposed to a given distribution and frequency, i.e. rate of variation, of demand may seem to be an analytically derivable and quantifiable price per volume of traffic optimal for the dampening of the (expected) oscillation in resource utilisation (or demand), $BU$.

In figure 9.7 we plot the movement of the poles of the characteristic equation in response to variations in the $\alpha_2$ or the weighting of the price of the previous time-slot, see table 9.1, for $\alpha_1 = 0.9$ & $K \cdot k = 1$. We see from the root locus that the system is only stable for part of the range of values of $\alpha_2$. Through numerical evaluation we establish that the system is stable, i.e. the roots are located inside the unit circle, up to $\alpha_2 \to 2$; but that as $\alpha_2$ increases beyond 2, the poles move outside the unit circle and the dynamic charging becomes unstable. In other words, if the weighting of the price in the previous time-slot is greater than twice the weighting of the error signal the charging scheme may cause the utilisation of the network resource to increasingly oscillate and exhibit unstable system, i.e. network, behaviour.
Figure 9.6: Root locus of $H(z) = \frac{A_{BU}}{A_D}$ with respect to $K \cdot k \cdot \alpha_1$ term

We can also see from figure 9.7 that as $\alpha_2$ increases (beyond 0) the poles of the transfer function move along the real axis until they coincide and then split into a complementary complex conjugate pair. The point where they coincide, $\alpha_2 = \sim 0.32$, corresponds to a critically damped system. This is where the admission controlling aspect of dynamic charging exhibits the most effective dampening in resource utilisation without causing it to oscillate.

Finally, in figure 9.8 we plot the fractional amplitude of the change in resource utilisation, $\Delta BU$, in response to a unit step in change of demand, $\Delta D$, for $\alpha_1 = 0.9$, $\alpha_2 = 1$ & $K \cdot k = 1$. It can be seen that the SS error does indeed converge towards zero as we predicted (due to the double-integrator transfer function) but that dynamic charging under these conditions would cause the (change in) resource utilisation to oscillate. That is, the system is not critically damped.

In our model we have assumed that the relationship between $ND$, $D$ and $P$ is linear, refer equation (9.6). This may indeed be a valid assumption and as there does not really exist any data which describes how customers might react to the proposed charging scheme it seems to be a reasonable first guess/approximation. It further has the advantage that it maintains the requirements of linearity for 'standard' control theory and therefore may be modelled using normal techniques.

However, as also discussed in section 8.1, the relationship between price and
Figure 9.7: Root locus of $H(z) = \frac{BU}{AD}$ with respect to $\alpha_2$

demand as typically seen and understood by economists to be the case in free markets is indeed linear but only in a log vs log relationship. The equation below describes the relationship between consumption, price and income as it has been modelled by economists regarding the consumption of products [40], where $C$ is the consumption, $Y$ the real income per capita, $P$ is the price of the product in question and $\beta_0 (> 0)$ and $\beta_1 (< 0)$ are coefficients specific to the market described.

$$\log C = \beta_0 + \beta_1 \log Y + \beta_2 \log P$$

Although this relationship generally is used to describe markets under very different circumstances it is not unlikely that a log vs log price-consumption relationship would prove to be the best for a dynamically charged telecommunication services market. If this was to be proven to be the case it would naturally impact our control system model and influence the proposed method of analysis.

9.2.5 Summary

In this section we have proposed a new and novel method of analysis for charging schemes in telecommunication systems. Namely, we have shown how classical control theory may be applied to the analysis of the proposed FM (SM or dynamic) charging
mechanism to derive analytical and proved results. Specifically we showed how the model was developed and analysed. We investigated the short & long term effects (transient & SS responses in control theory terminology) of the charging mechanism. We also evaluated conditions of stability.

It should be noted that although this model was derived to enable us to describe and analyse the performance of the proposed dynamic 'Free Market' charging mechanism, it is a generic model for all telecommunication resources and charging mechanisms. It does indeed describe how any network device behaves at the macro level in response to any charging scheme depending on the specific values of the individual coefficients. For instance, if $K = 0$, i.e. the weighting of the error signal is zero, we are modelling a charging mechanism which is not dynamic, such as most charging schemes for telephony services today.

### 9.3 Other avenues of investigation

In addition to the analytical analysis presented in the previous section the author also co-supervised (in collaboration with Prof. John E. Midwinter two M.Sc. IT projects; both of which aimed to to further analyse and quantify the behaviour of the proposed charging and control mechanism through simulation using the OPNET
simulation suite. The projects respectively aimed to implement and analyse models for integrated (broadband) network traffic sources and control for a single resource and were entitled "A Review of Traffic Source Models" and "Free Market Admission", refer section B.6 in Appendix B. Both project received passing final grades and the later also with a distinction.

The observed behaviour and results derived from the simulation models realised in OPNET agreed with those derived from the analytical model in the previous section.

9.4 Summary, Conclusions and thoughts on Possible Future Work

As well as further investigation of the proposed and demonstrated method of control theory analysis, future work could include the application of this analysis method to other charging schemes or the development of extended models for the analysis of multiple resources. One could also aim to probe and quantify the levels of agreement between the this method of analysis and the more traditional simulation further.

Indeed the author believes that building extended models for characterisation of behaviour of multiple resources in for instance extended networks probably is beyond this method of analysis. However, it might prove very powerful means of ‘proving’ (by demonstration) simulation models of individual resources which then in turn could be combined in larger more complex systems as a means of extending the investigative work of the proposed FM charging scheme to ‘flow control’.

Another possible avenue of exploration might be to investigate how FMAFC relates to for instance the framework of the ITU-T TMN standards and recommendations. Limited benefit would probably be gained from working towards a formal protocol definition but the author believes that substantial insight could be gained from formalising how FMAFC ‘fits in’ to the generic hierarchy defined by the TMN recommendations.

This might also be a step towards applying how FMAFC could be used in for instance ATM or IP networks. Discussions with Paul Kirkby, Nortel would also seem to indicate that there might be substantial scope for a control system of this nature in future satellite communication systems.

Finally the author also believes that additional gain could be derived from further economic analysis of the proposed control mechanism in the light of a more
concrete protocol definition. Particularly it seems desirable to evaluate performance of FMAFC from the perspective of individual agents a marketplace.
Bibliography


[26] Figures and info on capital deperesiation etc. i.


233


Chapter Ten

Overall conclusions

10.1 Introduction
In this thesis the work which was undertaken by the author under a three year EPSRC (CASE) award studentship with then BNR Europe, now Nortel Technologies Plc. as co-sponsor from Oct '92 till Dec '96 has been summarised. The reason for the prolonged duration of the studentship was a serious accident had by author mid '94 which effectively suspended it for one & a half years. As stated both in the abstract and the introduction, section 1.3 “Organisation of the thesis”, it contains work from three areas all related to advanced telecommunications. These have been presented, in as far as it has been possible, individually with separate sections summarising the work and results presented with suggestions for possible further work.

All of the three areas of work have been passed on to other PhD students within the Electronic & Electrical Engineering Department of UCL and two of these have since graduated.
10.2 Summary and conclusions

The first area of research was AONs. The emphasis of the work in this huge area of study was predominantly network architecture and was performed Oct '92 till mid/late '93. The results, which emanated from this work, are presented in Part II of the thesis. The approach proposed resulted in a number of papers, ref. chapter 6, and eight claims under a patent application filed by BNR Europe Ltd., with the author (Martin Sabry) and John E. Midwinter as inventors [1]. The proposed & analysed concept of a Sparsely Filled WDM (SF-WDM) transmission spectrum was shown to be very powerful indeed and would seem to be applicable to transmission and switching systems of all magnitudes of wavelength-time-space dimensions, ref. chapter 6. However, for many reasons concerning the reality of implementing and maintaining networks, it would probably be best utilised within a single network layer or within individual transmission systems. In this case the optoelectronic interface, Fig. 6.3.7, might e.g. be situated at the nodes interfacing to the access network (i.e. at the Central Office (CO), refer chapter 5). Evidence, that minimum wavelength allocation in randomly connected non-wavelength shifting AONs generically result in sparely filled transmission spectra has also since emerged [2].

As stated already the SF-WDM concept, at the time, certainly, and even at present, make some stark demands of the physical transmission and/or switching systems wrt. implementation. Consequently, the second part of the thesis, which seemed to follow on quite naturally from the first, was an attempt to quantify some of the assumption made during the development and investigation of the SF-WDM concept in terms of limitations imposed by implementation. This work is presented in part I of the thesis. The approach taken was to attempt to quantify these constraints in terms of fundamental theoretical limits imposed by e.g. fibre non-linearities and cross-talk\(^1\) in transmission & switching systems.

The studentship was at this point effectively suspended due to the serious accident had by the author mentioned above. When the studentship was resumed both of these areas had been ‘passed on’ to other Ph.D. students — which both have since graduated. This situation effectively prohibited either of these areas being revisited and consequently necessitated a change in research direction.

Part III of the thesis is concerned with the work of this period of the studentship and summarises some of the ideas & results, which resulted from a body of research concerning relating charging and management & control mechanisms in future

\(^1\) NB No results regarding cross-talk and switching systems have been presented.
(broadband, multi-service) telecommunication networks. The thrust of the work was to investigate the effects of dynamic charging and using the price or customers' Willingness to Pay (WP) as a control variable. As stated in section 1.3.1 “Statement of Originality”, pp. 49-50, the approach pursued in this part of the thesis was to investigate, develop & analyse in detail how this method of charging and/or network resource management & control could be applied to a single telecommunication process or hardware resource such as CAC (Connection Admission Control) or a server. The results presented include an analytical method of analysis of charging mechanisms in telecommunication networks & systems using control theory, the proposed ‘(fast) reflector switch’, the concept of price ‘lifetime’ and minimisation of computational complexity & signalling overhead through the maximum cost of resources charging scheme when applying dynamic charging to multiple network resources. Patents for the latter three have been proposed and are pending possible pursuit by Nortel.

This area of research has since also been passed on to another (& current) Ph.D. student within the department.

10.2.1 Further Work

As stated above, each part, chapter & section have been written as far as possible to be self-contained and include individual summaries of the findings presented and any conclusions and possible suggestions for further work. Naturally as two thirds of the work has not been evolved since it was suddenly interrupted in '94 due to the circumstances mentioned above some of the conclusions and suggestions have already been superseded by subsequent research & results. This is particularly evident in part I of the thesis which reviews the (at the time) state of the art enabling technologies and attempts to analyse the limitations of implementation. However, to the author's knowledge, no work has since been produced which invalidates any of the fundamental concepts presented.
10.3 References


Appendix A

About the author

A.1 Martin Sabry Homepage
A.2 List of Publications
A.3 List of Patents
Martin Sabry

Department of Electronic and Electrical Engineering
University College London
Torrington Place
London WC1E 7JE - United Kingdom
e-mail: m.sabry@eleceng.ucl.ac.uk
phone: +44 (0)171 - 419 3843
fax: +44 (0)171 - 387 4350

Martin Sabry was born in Edinburgh, 1972. He received his education at Daniel Stewart & Mellville College, Edinurgh, The Perse School, Cambridge and Holte Gymnasium, Copenhagen; He graduated with a B. Eng. (Hons.) degree from the University of Essex, Dept. of Electronic Systems Engineering, 1992. He received the final year project prize for his work on "A Multimedia Help Sytem". He joined UCL in 1992.

Besides English Martin speaks fluently & writes German, French and Danish; He has travelled extensively round the globe.

Martin's main research interests include:-

- Network control and management, including methods of charging, flow & addmission control
- All Optical Network (AON) architectures
- Remote & distributed computer based teaching services within telecommunication networks
- Transmission and switching limitations within AONs
- Application of Neural Networks within telecommunication systems

Martin's published work includes:-

- Sparsely Filled Wavelength Division Multiplexed (SF-WDM) All Optical Network (AON) architectures
- Evolution of transparent AONs
- Transmission limitations of WDM transmission systems
- Application of Neural Networks as telecommunication decoders

He also holds registered patents dealing with the invention of SF-WDM network architectures.

Martin Sabry is an IEEE Associate Member and an IEEE Member.

You are the person to visit this page

http://www.ee.ucl.ac.uk/~msabry/ 26/09/98
A.2 List of Publications


Contributions made to:-

EC Contract: “Towards Transparent Photonic Networks”

GPT Ltd. consultancy: “Potential for Wavelength Division Multiplexing in Local Broadband Networks”

EC Contract: “Optical Communications Evolution”

3 August 1994 - 1 November 1996 PhD interrupted due to serious accident


A.3 List of Patents

M. Sabry, J. E. Midwinter (inventors) & Northern Telecom, “Optical communications network”, UK patent application 931 897 9.3

M. Sabry, J. E. Midwinter (inventors) & Northern Telecom, “‘Free Market’ Charging”, Patent application pending

- The “(fast) reflector switch”
- The “maximum cost of resources charging scheme” when applying dynamic charging to multiple network resources
- “Price ‘lifetime’ ”
Appendix B

Other work

B.1 "Bi-directional wavelength routing of optical communication ring networks", M. Sabry and J. E. Midwinter ................................................................. 251
   Abstract ................................................................................................................................. 251
   Text ........................................................................................................................................ 250
   B.1.1 Acknowledgements ....................................................................................................... 255
   B.1.2 References ................................................................................................................... 256
   B.1.3 Figures .......................................................................................................................... 256

B.2 "A Neural Network Implementation Suitable for Correlation Decoding of Some Block Codes", M. Sabry, D. Grant, J. E. Midwinter and J. T. Taylor ................................................................. 259
   B.2.1 Introduction .................................................................................................................. 259
   B.2.2 Conventional and Previously Proposed Neural Network Decoding Techniques .......... 259
   B.2.3 The BAMNET ANN Prototype .................................................................................... 261
   B.2.4 BAMNET ANN Decoder Implementation ................................................................. 264
   B.2.5 Conclusions and Future Work ..................................................................................... 266
   B.2.6 Acknowledgements ...................................................................................................... 267
   B.2.7 References .................................................................................................................... 267

B.3 "Telecommunication Network Management & Control Through Charging", Martin Sabry and J. E. Midwinter ................................................................. 269
   B.3.0 Abstract .......................................................................................................................... 269
   B.3.1 Introduction .................................................................................................................. 269
   B.3.2 Congestion Control Schemes ....................................................................................... 269
   B.3.3 Charging Mechanisms .................................................................................................. 270
   B.3.3.1 Other Charging Schemes .......................................................................................... 271
      B.3.3.1.1 Effective Bandwidth Charging .............................................................................. 271
      B.3.3.1.2 'Smart Market' / Dynamic / 'Free Market' charging ........................................... 272
   B.3.4 Discussion .................................................................................................................... 272
   B.3.5 References .................................................................................................................... 273
B.4 "An Analytical Approach to Charging Mechanisms using Control Theory", Martin Sabry, Nina F. Thornhill and John E. Midwinter

B.4.0 Abstract ..........................................................................................................................................279
B.4.1 Introduction ....................................................................................................................................279
B.4.2 Model .............................................................................................................................................280
B.4.3 Analysis & Discussion ...................................................................................................................283
B.4.4 References ......................................................................................................................................286

B.5 "Transmission Limitations due to Optical Fibre Nonlinearities in Transparent Optical Networks", Martin Sabry and J. E. Midwinter

B.6 MSc IT projects

B.6.1 "A Review Of Traffic Source Models" by Victor Okoro; ...............................................................291
Project Supervisor: Professor. John Midwinter & Mr. Martin Sabry
B.6.2 "Free market admission" by A. K. Griffiths; .................................................................................293
Supervisors: Professor J. E. Midwinter & Martin Sabry

B.7 Miscellaneous Ph.D. Related Talks & Presentations

B.7.1 “The Future of Telecommunications” by Martin Sabry, Electrical and Electronic Engineering, University College London: Bloomsbury Rotary Club, 2 Jun 1993
B.7.2 “Tomorrow’s Telecom Networks and Services” by Martin Sabry (Electrical and Electronic Engineering): UCL Graduate Interdisciplinary Society, 7 Oct. 1997 ..............................................................................295
Bi-directional wavelength routing of optical communication ring networks

M. Sabry and J. E. Midwinter
University College London, Torrington Place, London WC1E 7JE

A new approach to wavelength routing optical ring networks is proposed which is shown to have properties that typically give rise to a 3 fold improvement in capacity compared to today's FDDI networks and previously proposed wavelength routing schemes of optical rings.

Communication networks generally facilitate full connectivity for a set of nodes. An M node network thus requires M(M-1) links. These links can either be actual or virtual. The latter is generally the case in real systems, since the number of nodes is such that installing physical links is impractical.

WDM however, allows a single optical fibre to carry many individual channels, which, if utilised in a wavelength routing approach, may be thought of as individual physical node to node connections. When routing round a ring network structure, a given wavelength may always be used twice, from node A to node B and node B to node A. A fully interconnected wavelength routed ring network is therefore understood to require M(M-1)/2 wavelengths [1].

However, this number is only correct if routing takes place in one direction round a ring. If one assumes that it is possible to route in both clockwise and anticlockwise directions round the same ring, one may considerably reduce the number of wavelengths, because A and B are always connected in both directions via the
shortest route, which permits the reuse of the same wavelength over the longer path. The bi-directionality could either be achieved simply by using dedicated fibres for each direction or using duplex transmission [2, 3]. It will be shown that both give rise to an equal improvement in capacity. However, since duplex transmission potentially allows twice as efficient use of the transmission window, one may reduce the number of fibres required to facilitate all the necessary channels even further than in the case of uni-directional transmission. This would naturally be at the expense of a more stringent specification of the termination of each length of fibre.

In this letter we propose a new approach to wavelength routing rings, where individual wavelengths are used to make as many connections as possible of a given distance. To illustrate this we consider a six node ring, which is shown as fig. 1. Wavelength $\lambda_1$ forms 6 connections of distance 1, $\lambda_2$ forms 3 connections of distance 2, etc. It can be seen that full connectivity can be achieved with a total of 5 \textit{(or 9)}$^1$ where conventional routing would require 15 wavelengths. The improvement naturally arises from always being able to connect nodes via the shortest route. In the case where $M \rightarrow \infty$, the number of wavelengths required to connect all nodes fully will be \((2 \times) \sum_{i=1}^{\infty} \frac{1}{i} = (2 \times) (M/4)(1+M/2)\). At the limit where $M \rightarrow \infty$, this is 8 \textit{(4)} times less than the number of links made, i.e. each wavelength is on average reused 8 \textit{(4)} times.

In a real network the use of the wavelengths can not be quite as efficient, since the number of nodes through which a connection is passively routed will not always be a factor of $M$. In those cases it is necessary to use additional wavelengths to make the remaining connections. Eqn. (1 & 2), express the total number of wavelengths

\begin{footnotesize}
\textsuperscript{1}Uni-directional transmission is described in \textit{bold italics.}
\end{footnotesize}
required to fully connect an M node ring bi-directionally for uni-directional and
duplex transmission respectively, where the \( i^{th} \) term in the summations is the number
of wavelengths required to make all connections of distance \( i \).

\[
\sum_{i=1}^{\frac{M-1}{2}} \left\{ 2 \cdot i + M \mod i \right\} + \frac{\pi}{2} \cos^2 \left( \frac{\pi i}{2} \right) \tag{1}
\]

\[
\sum_{i=1}^{\frac{M-1}{2}} \left\{ i + (M \mod i) \right\} \div 2 + (M \mod i) \mod 2 \\
+ \left( \frac{\pi}{2} \div 2 + \frac{\pi}{2} \mod 2 \right) \cos^2 \left( \frac{\pi i}{2} \right) \tag{2}
\]

Fig. 2, illustrates the difference in the number of wavelengths required when
wavelength routing a ring network by the method proposed by Hill [1] and the
routing scheme proposed above, versus the size of the ring. If 510 wavelengths slots
are available, one can fully interconnect 58 (41) nodes adopting bi-directional routing
versus 32 using uni-directional routing. Comparatively, fully connecting 58 and 41
nodes in one direction only, requires 1653 and 820 wavelengths respectively.
Moreover, for \( M = 3, 4, \ldots, 41 \), the mean wavelength reuse is 6.21, with a standard
deviation of 0.39 (3.22, with a standard deviation of 0.13) compared to 2.

Eqn. (1 & 2) and fig. 2, indicate that some sizes of networks are more efficient than
others in their use of wavelengths. The reason for this is that a connection of a given
distance may require more wavelengths in some rings than in others. For example,
connecting all nodes to those 2 nodes removed from them requires 3 (5) wavelengths
in a ring with an odd number of nodes; but only 2 (4) if the ring has an even number
of nodes. This phenomenon becomes increasingly more noticeable as the size of the
structure increases and the connections become more distant. If \( M = 41 \), all
connections of terminals separated by more than 13 nodes demand 41 wavelengths,


Figures

![Diagram of a 6 node wavelength routed ring. Full connectivity achieved with 4 and 5 wavelengths in the counter clockwise and clockwise direction respectively, i.e. a total of either 5 or 9 depending upon whether duplex transmission is used or not.]

Note: In the case where duplex transmission is not used lambda i in the clockwise (solid line) and anticlockwise direction (dashed line) are different wavelengths.

Fig. 1 6 node wavelength routed ring. Full connectivity achieved with 4 and 5 wavelengths in the counter clockwise and clockwise direction respectively, i.e. a total of either 5 or 9 depending upon whether duplex transmission is used or not.
Uni-directional routing

Bi-directional routing and uni-directional transmission

Bi-directional routing and duplex transmission

Fig. 2 The number of wavelengths required to fully connect a wavelength routed ring.

Fig. 3 The number of wavelengths required to make the connections versus the distance of the connection (in nodes) for M=5, 8, 11, ..., 41
A conceptually simple method of hard-decision decoding of a linear \((n, k)\) block code is minimum-distance/maximum-likelihood decoding. The \(n\)-bit received vector is compared with each of the \(2^k\) possible code words and the one found to be the closest in terms of Hamming distance is selected. Minimum distance decoding is in fact optimum in the sense that it results in a minimum probability of code word error \( [2] \). One possible algorithm is to add (modulo-2) the received vector to all \(2^k\) possible code words to obtain the error vectors, \( e \), and hence select \(\min \{ \text{weight}(e_i) \} \), i.e. the code word corresponding to the error vector with the fewest 1's. It can be seen that this is extremely computationally intensive even for small code word sets.

An alternative approach is to compute an error syndrome for the received vector, then, using a lookup table, find the most probable error pattern for the particular error syndrome and hence obtain what is most likely to be the corrected code word by addition of the received code word and the error pattern. The error syndrome is obtained by multiplying the received vector by the parity check matrix for the code used. Block codes may be divided into categories, e.g. linear, non-linear and cyclic codes, according to their properties. When a code is cyclic, the syndrome computation may be implemented by a shift register with number of feedback paths. This method of decoding requires the shift register to work at a speed comparable to that of the transmission rate, whereas the algorithm proposed for minimum distance decoding would need to operate many times faster in order not to cause a bottleneck.

In 1986 Platt and Hopfield proposed that a Hopfield ANN could be used to implement an error correction decoder \([3]\). They argued that due the inherent parallelism of ANNs, very high decoding rates could be achieved. In addition it was pointed out that this type of decoder could be realised with relatively simple analogue circuitry. However, since a Hopfield ANN can only store a maximum of \(n\) memories for an \(n\)-dimensional input vector \([6]\), they were forced to specially design what was essentially "1 in \(n\)" codes to limit the number of code words so as not to exceed the storage capacity of the Hopfield ANN.

This restriction was also cited by Yuan and Chen \([4, 5]\). The fact that the memory capacity of a Hopfield ANN is at best \(n\), seems to present a fundamental problem as an ordinary \((n, k)\) block code will have \(2^k\) code words, where generally \(2^k \gg n\). They showed how this obstacle may be overcome by exploiting a number of structural properties of some block codes and illustrated this for the \((8, 4)\) Hamming and \((24, 12)\) Golay codes. They showed that the extended Golay code can be divided into eight subcode sets and a further 64 sub-subcodes which may be expressed in terms of a 3-dimensional universe code. This allowed them to firstly correlation decode the received vector with
the individual subsets and subsequently determine the element of the resultant set of code words with the maximum overall correlation using a number of Hopfield n-flops

Based on this principle, a theoretical implementation of a pseudo ideal (24,12) Golay code correlation decoder, capable of both soft and hard decision decoding, was simulated, successfully illustrating the feasibility of ANN based correlation decoders. A real implementation of the proposed decoder would, however, require a total of 16896 resistors, 576 op-amps, 8 64-flops and 1 8-flop. One final point to note is that a Hopfield network, as well as limiting the size of the code word set, also demands that the individual codes be of a format to prevent the circuit oscillating indeterminately [5].

The BAMNET ANN Prototype

As already mentioned a BAMNET ANN prototype has been developed and experimentally verified as being an optimum minimum error binary pattern classifier. This ANN is implemented in a modular analogue VLSI architecture using standard CMOS 2.4µm Mietec technology [7]. A functional schematic of the BAMNET ANN is shown in fig. 2.

![Functional diagram of BAMNET ANN](image)

The structure consists of two distinct parts: a correlation layer and a Winner-Takes-All (WTA) layers. The correlation module circuitry is an N×M array of analogue XNOR transconductors which generates matching-scores for M stored exemplar patterns for a given N-dimensional input vector, where M \( \leq 2^N \). The WTA circuits consists of two sections. The first controllably quantizes the current levels into voltage levels which represent the degree of correlation. Five such measured levels are shown in fig. 3 with a 10MHz control clock cycle. The second section selects the node with the highest voltage and suppresses activity in all others. Overall the WTA layers select the

---

1 Where a Hopfield n-flop basically implements a winner takes all function.

2 The only deviation from the ideal occurred when an input vector is equidistant from two or more stored code words.
exemplar which generates the highest matching score, i.e. the one which most resembles the input pattern.

Using a 3MHz clock cycle, a typical scenario is presented in fig. 4. Initially the applied vector exactly matches the digit three and is corrupted bit by bit until it becomes closest to the exemplar pattern representing the digit two. With reference to fig. 4, the upper signal of each pair monitors the output response from node Vo3 and the lower signal node Vo2. For the first pair, the applied vector in (a), best represents the exemplar 3, so the node Vo3 'fires' and no signal from Vo2 or any other node is observed. In the second (middle) pair the applied vector in (b) still best represents the 3 and the situation remains unchanged. Finally another bit of the applied pattern is changed in (c), making it closest to the exemplar pattern 2, consequently Vo2 fires and the signal in Vo3 disappears. Note also that the input vector is being classified in under 200ns for a 3 MHz clock cycle, in one pass.

In the case where the received code word is so corrupted that it is impossible to determine the original code word, and hence the data content, a decoder so indicate this. In relation to the BAMNET ANN a "null response" is given, when an input vector is either equidistant from the two closest matching code words, or further removed than a pre-adjusted Hamming distance from all of the code words.

Fig. 3 - Levels of correlation generated by the first section of the WTA layer.
A demonstration of a null response is shown in fig. 5 (with a 3MHz cycle). The upper signal of each pair represents Vo4 and the lower signal Vo9. Initially the applied vector (a), best represents a four so output node Vo4 generates a digital signal and no signal is produced in Vo9, or any other node. Then a key bit is corrupted such that the applied vector (b) is equidistant from the exemplar patterns representing nine and four. Since the difference in voltage generated for the corresponding correlation levels in the WTA first section is insignificant, the WTA second section cannot be activated and so no output responds, as shown in the middle pair. Finally, another bit is changed making the applied vector exactly match the 9, such that Vo9 now responds and Vo4 remains off. When the applied vector is significantly different from all the stored memories the correlation levels produced by the WTA first layer saturate to a common threshold and thus the BAMNET ANN will also produce a null response.

### Fig. 4 - Error Correction as '3' is changed into a '2'

<table>
<thead>
<tr>
<th>Channel 1</th>
<th>Channel 2</th>
<th>Memory 1</th>
<th>Memory 2</th>
<th>Memory 3</th>
<th>Memory 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>V/div</td>
<td>Signal</td>
<td>V/div</td>
<td>Signal</td>
<td>V/div</td>
</tr>
<tr>
<td>Channel 1</td>
<td>5.00</td>
<td>Channel 2</td>
<td>5.00</td>
<td>Memory 1</td>
<td>5.00</td>
</tr>
<tr>
<td>Memory 1</td>
<td>5.00</td>
<td>Memory 2</td>
<td>5.00</td>
<td>Memory 2</td>
<td>5.00</td>
</tr>
<tr>
<td>Memory 3</td>
<td>5.00</td>
<td>Memory 4</td>
<td>5.00</td>
<td>Memory 3</td>
<td>5.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Timebase</th>
<th>100 ns/div</th>
<th>Delay/Pos</th>
<th>Reference Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Offset</td>
<td>Probe</td>
<td>Coupling</td>
</tr>
<tr>
<td>Channel 1</td>
<td>15.0000 V</td>
<td>10.00 : 1 dc BW 11m (1m amp)</td>
<td></td>
</tr>
<tr>
<td>Channel 2</td>
<td>-12.5000 V</td>
<td>10.00 : 1 dc BW 11m (1m amp)</td>
<td></td>
</tr>
<tr>
<td>Memory 1</td>
<td>-11.2500 V</td>
<td>MemSweep = 100 ns/div</td>
<td></td>
</tr>
<tr>
<td>Memory 2</td>
<td>0.0000 V</td>
<td>MemSweep = 100 ns/div</td>
<td></td>
</tr>
<tr>
<td>Memory 3</td>
<td>-1.25000 V</td>
<td>MemSweep = 100 ns/div</td>
<td></td>
</tr>
<tr>
<td>Memory 4</td>
<td>-1.25000 V</td>
<td>MemSweep = 100 ns/div</td>
<td></td>
</tr>
</tbody>
</table>

Trigger mode: Edge
On Positive Edge of Chan2
Trigger Level
Change = 3.00000 V (noise reject OFF)
holdoff = 40.000 ns
BAMNET ANN Decoder Implementation

In relation to the above discussion these attributes appear attractive for the practical implementation of a neural network correlation decoder. Since the maximum number of memories is not limited to N, as is the case for a Hopfield network [6], its use is not restricted to communication systems which employ '1 in n' codes [3], or codes featuring structural properties which permit partitioned correlation decoding [5]. In addition, the BAMNET ANN implements direct parallel memory access to the stored code words without the occurrence of indeterminate oscillation. Finally, it exhibits 100% associativity regardless of the size of the store code word set, thus eliminating the possibility of convergence to an incorrect solution, i.e. faulty decoding of the received data.

For these reasons the BAMNET ANN will be ideal as a correlation decoder, suitable for the type of system depicted in fig. 1. If a return channel was available the request for re-transmission signal could be supplied by the null response flag. In its present form the prototype is limited to 100ns classification regardless of the size of the code word (n) and the code word set (M). However, this is
a restriction caused by the fabrication technology rather than the architecture. The 2.4\mu m CMOS Mietec technology also means that the present size of each neuron is approximately 100\mu m x 80\mu m. This represents a limitation for practical decoder implementations, as it may not be possible to store large code word sets on a single die. However, the modular design of the BAMNET ANN permits large code word sets to be split over multiple dies. Alternatively, superior technologies could facilitate miniaturisation of the constituent circuit modules and consequently larger code words on an equivalent die area would be made possible. A possible \((n, k)\) BAMNET ANN based hard-decision decoder is depicted in fig. 6.

Fig. 6 - Hard-decision correlation decoder.

In addition to the BAMNET ANN it would make use of a Serial In Parallel Out (SIPO) & Parallel In Parallel Out (PIPO) shift register circuit module, a decoder module and some control logic. The SIPO would load the serial data stream received from the decision circuitry in blocks of \(n\) bits. These would then be parallel loaded into the PIPO which in turn would hold the entire code block for the duration required by the BAMNET ANN for correct classification, \(\sim 100\) ns. Hence, the maximum bit-rate \((B_{\text{max}})\) imposed by the BAMNET ANN is \(n/100\) ns, e.g. for the \((24, 12)\) Golay code \(B_{\text{max}} = 2.40\) MBit/s. The output decoder would be a ANN reconfiguration layer that would supply the original code word corresponding to the best matching block code, or a null response flag. The control logic would generate the sequence of the signals required to synchronise the operation of the various modules from the clock recovery circuitry.
Conclusions and Future Work

In summary, the principle features of our BAMNET ANN, which make it suitable for the implementation of an ideal hard-decision correlation decoder, are:

- Storage of the maximum theoretical number of memories, i.e. $M \leq 2^n$.

- The classification of a partially complete or corrupt input vector with 100% associativity, i.e. the capability to take arbitrarily chosen erroneous input vectors and correct the bits in error as dictated by the nearest stored code word.

- Direct parallel memory access, resulting in real time pattern classification in about 100ns.

- Production of a "null response", when an input vector is either equidistant from the two closest matching code words, or further removed than a given (adjustable) Hamming distance from all of the code words.

- Permits both single and multiple die/chip implementation.

A further improvement to the BAMNET ANN would be to modify the existing synapse structure to include programmability [9]. This would enable the decoder to process different code word sets.

Although, the present ANN prototype is a binary pattern classifier a future modification could be to realise analogue valued correlation by the implementation of continuous value transconductor XNOR synapses. Specifically, this would require the re-scaling of two of the transistors in each synapse. Fig. 7 shows the transconductance characteristic of each synapse for both the present BAMNET ANN and after the proposed modification.

Such a structural alteration would mean that the ANN would lend itself to soft- as well as hard-decision decoding. Under these circumstances, the SIPO & PIPO module in the decoder proposed in fig. 6 and the decision circuitry of fig. 1 would have to be replaced by a sample and hold circuit module so the individual bits of the code block may be held for the duration required for classification. It is as a soft-decision correlation decoder that the type of ANN presented in this paper may potentially have its greatest advantage over conventional decoding techniques. It is generally accepted that soft-decision decoding can offer significant improvements in BER [1] but they are computationally very expensive to implement, e.g. decoding the Golay (24, 12) code using the direct search algorithm requires 98304 computational steps and even the improved Conway and Sloane algorithm requires 1584 [8].

---

3The decision module could still be included in the receiver system if either quantized level soft-decision decoding or hard-decision decoding is desired.
Fig. 7 - Transconductance characteristic of present and proposed synapses.

Acknowledgements

This work was carried out with the support of SERC and BNR Europe Ltd.

References

Telecommunication Network Management & Control
Through Charging

Martin Sabry and J. E. Midwinter

Department of Electronic and Electrical Engineering
University College London
Torrington Place, London WC1E 7JE

msabry@eleceng.ucl.ac.uk

Abstract

This paper summarises some of the work which has taken place as part of a body of research concerning relating charging and management & control mechanisms in future telecommunication networks. The thrust of the work has been to investigate the effects of dynamic charging and using the price or customers' Willingness to Pay (WP) as a control variable.

We review the typical performance of telecommunication network resources versus loading, summarise some of the properties of a number of charging mechanisms and discuss how each of them may be used to control and manage resources in telecommunication networks.

1 Introduction

Traditional telecommunication networks, such as the terrestrial telephony network, are designed to support users with homogeneous and simple Quality of Service (QoS) requirements. The telephony network provides essentially only a Constant Bit Rate (CBR) circuit switched service to typically two types of customers, i.e. business and residential. This type of session and each of the classes of customer conform to distinct but well established patterns.

Future integrated (broadband) communication networks, however, are expected to support a multitude of communication services from a wide selection of session and QoS metrics. Consequently, it is expected that more advanced session and network control & management functions will be necessary.

2 Congestion Control Schemes

One important area will be the allocation and re-allocation of resources in conjunction with admission as well as flow control. Although the aim of these mechanisms should be to maximise the throughput of data, they should do so without causing excessive delay or ultimately loss of information. Intuitively, it seems quite reasonable to state that this is best achieved by distributing the demand on resources so that the traffic load is spread evenly. This must be done either through restricting access, i.e. admission control, or choosing to only use resources with partly or completely unused capacity, i.e. flow control.

If this is not achieved the performance of the individual devices and therefore in turn the network as a whole will deteriorate as the load is increased. Figure 1 on the following page depicts the typical performance of a communication device or sub-system such as a network queue versus the traffic load which it is offered.

Initially the Throughput — the proportion of the arriving traffic which is delivered, i.e. served — increases linearly in step with the offered traffic. However, as the traffic Load is increased still further the controller part of the queue will be spending an increasing proportion of its time processing data, to serving it. Hence, the gradient of the Throughput (with respect to the offered Load) will decrease and eventually tend towards zero. This is the point of 'on-set' of congestion, i.e. when the controller spends an equal proportion of its time processing and serving data. If the offered Load is increased still further, the Throughput will not increase; the Throughput has reached a maximum and any increase in the volume of offered data will not cause any more data to be served. Depending on the (congestion) control mechanisms in place it will be possible to operate the queue at this point...
be ten times (i.e. 1000\%) the mean bit-rate, whereas it may only be 110\% (of the mean bit-rate) in a constant bit-rate source with lower delay and data loss probability requirements.

In the effective bandwidth charging scheme proposed by Kelly [2] the user would be asked to declare the expected mean rate of the source, \( m \), at the time of call admission and will then be charged according to \( f(m, M) \), equation (2), where \( M \) is the measured mean rate and \( a(m) \) & \( b(m) \) are the charge per unit time & charge per unit of traffic carried, respectively.

\[
f(m, M) = a(m) + b(m) \cdot M
\]  

(2)

\( b(m) \) is chosen as the gradient of the effective bandwidth plotted against the measured mean data rate, \( M \), at the point where \( M = m \), i.e. where the measured equals the declared mean data rate. This has the effect of encouraging the user to co-operate with the network, share information and be honest about the characteristics of his traffic source as the total charge for a 'call' will be minimised the more accurate the user's declaration of expected mean data rate is.

3.1.2 ‘Smart Market’/Dynamic/‘Free Market’ charging

The concept of dynamic charging, also known as ‘Smart Market’ and ‘Free Market’ charging, has been proposed and worked with in the literature, as a mechanism of charging in telecommunication networks. This scheme allows/requires the charge levied on the use of a device to vary in proportion to the degree of utilisation of the resource. That is, as the use of a network resource, such as a switch or transmission trunk, is increased beyond the sp of utilisation (for maximum throughput while maintaining the desired QoS), the price is also increased (and conversely decreased) in order to discourage (or encourage) traffic from entering the network.

This charging scheme may be thought of as a macro-economic means of admission and flow control [6, 5] and may consequently be used in conjunction with conventional telecommunication network resource management procedures.

Another strong argument in favour of dynamic / ‘Smart Market’ / ‘Free Market’ charging is that the price internalises the externality by making users face the cost and/or burden he or she imposes on the network and its resources and thereby on the other users (and himself). In contrast, under non-dynamic charging a user may consider and weigh only their own costs and benefits from usage. This phenomenon is known by economists as congestion externality (and by ecologists as the problem of the commons). It has been argued, [4] and others, that dynamic pricing is (the most) favourable when considering how the pricing scheme used affects the industry structure, performance and evolution.

4 Discussion

In this paper we have reviewed the likely required QoS and control & management functions of future (broadband) integrated telecommunication networks; we have shown the typical performance of a network resource versus the load it is offered; and we have discussed the basis of charging mechanisms in general and with reference to two particular schemes, i.e. the effective bandwidth and dynamic (also known as ‘Free Market’ or ‘Smart Market’) pricing mechanisms, which encourage co-operative sharing.

In both schemes the user is rewarded for supplying information about his traffic source to the network management so that it, in turn, will be able to, more efficiently, perform its task and maximise social welfare. It may be argued that both of these charging schemes 'internalise' the congestion externality. However, in the case of effective bandwidth charging this process is established only for each customer individually. In a truly dynamic 'Free' telecom 'Market', however, this property is maximised as each user is "forced" to value her own requirements within the context of all the other customers.
References


Colloquium on

CHARGING FOR ATM -
THE REALITY ARRIVES

Savoy Place, London

Thursday, 20 November 1997

Organised by Professional Group E7
(Telecommunication networks and services)
and co-sponsored by Professional Groups
C3 (Information systems and networks) and
E14 (Television, radio and data broadcasting)

Reference No: 1997/328
CHARGING FOR ATM - THE REALITY ARRIVES

Thursday, 20 November 1997

Colloquium organised by Professional Group E7 (Telecommunication networks and services) and co-sponsored by Professional Groups C3 (Information systems and networks) and E14 (Television, radio and data broadcasting).

held at Savoy Place, London

PROGRAMME

9.00  
REGISTRATION AND COFFEE

Chairman:  
J Griffiths (Queen Mary & Westfield College)

SESSION 1:  
THEORETICAL BASIS

9.30  
Simple models for cost prediction over DAVIC based architectures:  
I Stergiou, C Smythe, S Cvethovic and P Tzerefos (University of Sheffield)

9.50  
Comparing usage-based pricing schemes for broadband networks:  
C Courcoubetis, V A Siris and G D Stamoulis (ICS-FORTH)

10.10  
An analytical approach to charging mechanisms using control theory:  
M Sabry, N F Thornhill and J E Midwinter (University College London)

10.30  
COFFEE

SESSION 2:  
THE MARKET

10.50  
How do users expectations in ATM match with what telecommunications operators offer: A study of European SMES using ATM and the particular case of Finland:  
V Cartuyvels and C Delhaye (University of Liege, Belgium)

11.10  
Charging mechanism and policy: adapting to the commercial environment:  
D J Songhurst (Lyndewode Research)

11.30  
Business implications for ATM charging in a multimedia environment:  
T Ludwig (Health On Line Service Technology) and K Keil (O.tel. O Communications)

11.50  
A techno-economic analysis of selected ATM charging schemes for aggregate LAN interconnect:  
P McEntee, J McGibney, D Botvich, D Morris and T Curran (Teltec Ireland)

12.10  
LUNCH (Available at the IEE)
**Chairman:**
L Anania (European Commission DGXII/B)

### SESSION 3: REALITY

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.40</td>
<td><strong>Keynote Speaker:</strong> TBC</td>
</tr>
<tr>
<td>14.00</td>
<td><strong>Experiments with user understanding of usage based charging of ATM:</strong></td>
</tr>
<tr>
<td></td>
<td>J G Snip and N van Foreest (KPN Research)</td>
</tr>
<tr>
<td>14.20</td>
<td><strong>Charging ATM internet access, an experiment of usage based ATM charging:</strong></td>
</tr>
<tr>
<td></td>
<td>R Andreassen, M Stoer and O Osterbo (Telenor R &amp; D)</td>
</tr>
<tr>
<td>14.40</td>
<td><strong>Charging ABR from an operator's and from a user's view:</strong></td>
</tr>
<tr>
<td></td>
<td>K van der Wal and J van Lierop (KPN Research)</td>
</tr>
<tr>
<td>15.00</td>
<td>TEA</td>
</tr>
</tbody>
</table>

### SESSION 4: REALISATION

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.20</td>
<td><strong>Business models and system architectures for future QOS guaranteed internet services:</strong> P Kirkby (Nortel)</td>
</tr>
<tr>
<td>15.40</td>
<td><strong>A new OAM function for accounting: exchanging ATM charging data:</strong></td>
</tr>
<tr>
<td></td>
<td>A Kuiper (Cap Gemini) and K Scrupps (Queen Mary &amp; Westfield College)</td>
</tr>
<tr>
<td>16.00</td>
<td><strong>Dynamic charging for information services:</strong></td>
</tr>
<tr>
<td></td>
<td>C Redmond and V Wade (Trinity College Dublin)</td>
</tr>
<tr>
<td>16.30</td>
<td>CLOSE</td>
</tr>
</tbody>
</table>

The IEE is not, as a body, responsible for the views expressed by individual authors or speakers
© 1997: The Institution of Electrical Engineers
UK ISSN 0963-3308

This event qualifies for 2½ technical IEE PDU points

REFERENCE NO: 1997/328
Analytical Approach to Charging Mechanisms using Control Theory

Martin Sabry¹, Nina F. Thornhill and John E. Midwinter

msabry@eleceng.ucl.ac.uk

Abstract

In this paper we suggest how control theory may be used to model and analytically analyse telecommunication network charging mechanisms. We propose a control theory model which was used to analyse the dynamic (also known as 'Smart Market' or 'Free Market') charging scheme. We show how the analytical model was formalised and analysed to determine the long term or Steady State (SS), in control theory terminology, and short term or transient responses of the mechanism. We show how the system's performance varies as the parameters change. We also discuss the limitations of this method of analysis and in particular how this affects the results of the dynamic charging scheme analysis.

Introduction

Charging mechanisms are part of the control, management and administrative procedures of any system, is also the case in telecommunication networks. As is the case with most administrative procedures, actual procedure itself will influence the system being managed, in this case the network. It is frequently interesting to quantify how a control, i.e. charging mechanism, affects the system(s) that controlling. Building a representative model and analysing it may be quite challenging. One widely option is certainly simulation. However, simulations are virtually impossible to prove and non-trivial verify. Classical control theory analysis, on the other hand, is analytical and any results derived using it consequently proven. Furthermore, control theory has been used to model economic systems and either as an adjunct to or a means of verifying simulation work, or as a method of analysis in right.

In this paper we suggest and show how control theory may be used to model and analytically analyse telecommunication network charging and management mechanisms. In particular, we propose a control model which was used to analyse the dynamic (also known as 'Smart Market' (SM) or 'Free Market' (FM)) charging scheme. We will show how the analytical model was formalised and how it was determined the long term or Steady State (SS), in control theory terminology, and short term or transient responses of the system. We will further show how a stability criterion was formalised and how response of the system varies as the parameters change. Examples of the parameters include the ring of the previous price, the SS accepted traffic ratio and the price per cell.

The motivation for this method of analysis has been the desire for an, at least, pseudo-analytical characterising the proposed method of charging; although, be it presently within a rather limited boundary of assumptions. Not only could this be used to verify simulation results but also as an independent avenue of investigation which would lead to a clearer understanding of the issues which this method of charging.

Additionally control systems are described as either Feed Forward (FF) or Feed Back (FB) systems. They are respectively depicted in figures 1 and 2, where SP, OP, MV and PV denote the Set Point, Ring Point, the Manipulated Variable and the Process Variable. A FF compensator or control figure 1, is similar to an open loop control system in that the Controller and Actuator's functions and responses are constant and fixed. FF control systems do, however, rely on knowledge of the disturbance and may change their response with time (following a pre-determined pattern) in anticipa- tion of the expected behaviour of the disturbance. In a telecom context the expected behaviour could be on knowledge of, for instance, the busy-hour traffic patterns in a traditional telephony network.

Controller's anticipation could be to switch in additional capacity in preparation for the expected
We see from the root locus that the system is only stable for part of the range of values of $\alpha_2$. Through numerical evaluation we establish that the system is stable, i.e. the roots are located inside the unit circle, up to $\alpha_2 \rightarrow 2$; but that as $\alpha_2$ increases beyond 2 the poles move outside the unit circle and the dynamic charging becomes unstable. In other words, if the weighting of the price in the previous time-slot is greater than twice the weighting of the error signal the charging scheme may cause the utilisation of the network resource to increasingly oscillate and exhibit unstable system, i.e. network, behaviour.

We can also see from figure 7 that as $\alpha_2$ increases (beyond 0) the poles of the transfer function move along the real axis until the coincide and then split into a complementary complex conjugate.
Figure 6: Root locus of $H(z) = \Delta B_U / \Delta D$ with respect to $K \cdot k \cdot \alpha_1$ term

Figure 7: Root locus of $H(z) = \Delta B_U / \Delta D$ with respect to $\alpha_2$

pair. The point where they coincide, $\alpha_2 \approx 0.32$, corresponds to a critically damped system. This is where admission controlling aspect of dynamic charging exhibits the most effective dampening in resource utilisation without causing it to oscillate.

Finally, in figure 8 we plot the fractional amplitude of the change in resource utilisation, $\Delta B_U$, in response to a unit step in change of demand, $\Delta D$, for $\alpha_1 = 0.9, \alpha_2 = 1 & K \cdot k = 1$. It can be seen that the SS error does indeed converge towards zero as we predicted (due to the double-integrator transfer function) but that dynamic charging under these conditions would cause the (change in) resource utilisation to oscillate. That is, the system is not critically damped.

In our model we have assumed that the relationship between $ND, D$ and $P$ is linear, refer equation (6). This may indeed be a valid assumption and as there does not really exist any data which describes how customers might react to the proposed charging scheme it seems to be a reasonable first guess/approximation. It further has the advantage that it maintains the requirements of linearity for 'standard' control theory and therefore may be modelled using normal techniques.

However, the relationship between price and demand as typically seen and understood by economists to be the case in free markets is indeed linear but only in a log vs log relationship. Equation (8) describes
the relationship between consumption, price and income as it has been modelled by economists regarding the consumption of products [7], where \( C \) is the consumption, \( Y \) the real income per capita, \( P \) is the price of the product in question and \( \beta_0 (> 0) \) and \( \beta_1 (< 0) \) are coefficients specific to the market described.

\[
\log C = \beta_0 + \beta_1 \log Y + \beta_2 \log P
\]  

Although this relationship generally is used to describe markets under very different circumstances it is not unlikely that a log vs log price-consumption relationship would prove to be the best for a dynamically charged telecommunication services market. If this was to be proven to be the case it would naturally impact our control system model and influence the proposed method of analysis.

It should be noted that although this model was derived to enable us to describe and analyse the performance of the proposed dynamic 'Free Market' charging mechanism, it is a generic model for all telecommunication resources and charging mechanisms. It does indeed describe how any network device behaves at the macro level in response to any charging scheme depending on the specific values of the individual coefficients. For instance, if \( K = 0 \), i.e. the weighting of the error signal is zero, we are modelling a charging mechanism which is not dynamic, such as most charging schemes for telephony services today.

In this paper we have proposed a new and novel method of analysis for charging schemes in telecommunication systems. Namely, we have shown how classical control theory may be applied to the analysis of the proposed dynamic (SM or FM) charging mechanism to derive analytical and proved results. Specifically we showed how the model was developed and analysed. We investigated the short & long term effects (transient & SS responses in control theory terminology) of the charging mechanism. We also evaluated conditions of stability.

As well as further investigation of the proposed model, future work could include the application of the analysis method to other charging schemes or the development of extended models for the analysis of multiple resources. We also aim to probe the levels of agreement between the proposed method of analysis and the more traditional simulation.

References


A Review Of Traffic Source Models

Author: Victor Okoro

A Masters in Information Technology Thesis

Date: September 1996

Project Supervisor: Professor John Midwinter.

Disclaimer
This report is submitted as part requirement for the Msc Degree Information Technology at University College London. It is substantially the result of my own work except where indicated in the text. The report maybe freely copied and distributed provided the source is explicitly acknowledged.
Information Technology Thesis
Free market admission

A.K.Griffiths
Department of Electrical and Electronic Engineering,
University College London

Supervisors: Professor J.E.Midwinter & Martin Sabry

September 6th, 1996

This report is submitted as part requirement for the MSc Degree in Information Technology at University College London. It is substantially the result of my own work except where explicitly indicated in the text. This report may be freely copied and distributed provided the source is explicitly acknowledged.
GIS, a society run by postgraduates in association with the Graduate School, is a regular event for UCL graduate students from a wide range of departments. It provides a forum for students from different disciplines to meet and explore topics of common academic interest. Most meetings are centred around a presentation followed by formal discussion and informal chat over cheese and wine.

Meetings in term time
1st Tuesday of the Month
5.30 - 7.30
Council Room, Registry Corridor
or
Haldane Room, North Cloisters

Next meeting

Tuesday, 7th Oct, 5.30 to 7.30
‘Tomorrow’s Telecom Network and Services’
- Martin Sabry (Electrical and Electronic Engineering)

free Cheese and wine!

All UCL Postgraduate Students and Guests Welcome

For further details or to be included on the G.I.S. mailing list, please contact via internal mail or e-mail:

David Scott (Department of Psychology) D.Scott@ucl.ac.uk
Programme, 1997-98

7th October  ‘Tomorrow’s Telecom Network and Services’
- Martin Sabry  (Electrical and Electronic Engineering)

4th November  ‘Postmodernism and virtual worlds, the case of cyberspace’
- Stephen Cummaford  (Psychology)

2nd December  ‘Money and Language’
- Gina Wenzel  (German)

6th January  ‘Changing Images - Changing Laws’
- Marion Reichart  (Law)

3rd February  ‘Perceptions of Privacy’
- Anne Adams  (Computer Science / Psychology)

3rd March  T.B.A.

5th May  T.B.A.

2nd June  T.B.A.

Tuesdays  Council Room, Registry Corridor
5.30 - 7.30  or
Haldane Room, North Cloisters

If you might be interested in presenting a paper or helping to organise future events, please contact:

David Scott  (Department of Psychology)  D.Scott@ucl.ac.uk