Universitatea POLITEHNICA Bucureşti

Facultatea de Electronică, Telecomunicaţii şi Tehnologia Informaţiei

Teză de doctorat

Contribuţii la proiectarea interconexiunilor multi-Gigabit cu aplicaţii la comutatoare în tehnologie 10 Gigabit Ethernet

Design and Qualification of the Interconnect for a High-Performance Modular 10 Gigabit Ethernet Switch

Doctorand: **ing. Alexandra Dana Oltean**
Conducător ştiinţific: **prof. dr. ing. Vasile Buzuloiu**

Ianuarie 2006

Acknowledgments

This thesis is the result of three years of work at CERN, Geneva, Switzerland. I would like to take this opportunity to thank some of the people from whom I have received help and guidance along the way.

Firstly, I would like to thank Professor Vasile Buzuloiu from Politehnica University Bucuresti for persuading me to start a PhD and proposing me to come to CERN, for his continuous confidence in difficult times and for his contagious optimism which has pushed me forward many times. I also would like to take this opportunity to thank Prof. Teodor Petrescu, Prof. Alexandru Serbanescu, Prof. Iuliu Szekely and Prof. Florin Sandu for accepting to be the reviewers of my thesis.

I am very grateful to Professor Bob Dobinson (who so sadly and unexpectedly passed away) and to my direct supervisor, Brian Martin, who accepted me into their group at CERN. I particularly express my gratitude to Brian, for being the person who has supported me the most during the last three years. I greatly appreciate that his door was open for discussions and questions at any time of the day, for helping me to focus in the right direction and for proofreading my thesis.

With respect to the project I was involved in, I want to thank Dr. Alexis Lestra for all the help during the AdvancedTCA backplane tester design and deployment. A particular appreciation to Matei Ciobotaru, for sharing his programming knowledge with me during the software development. I would also like to mention my gratitude to the other group members – Dr. Razvan Beuran, Mihail Ivanovici, Dr. Catalin Meirosu and Dr. Stefan Stancu - who were always there in time of need.

Last but not least, I would like to express my deepest gratitude to my family - my parents, who sustained me from a far-away distance, and my dear Johan - because of their beliefs in me and their constant encouragements. I could never have done it without you!

Abstract

The 10 Gigabit Ethernet standard, IEEE802.3ae, was adopted in 2002 and represented a ten-fold increase in performance over the previous 1Gigabit standard IEEE 802.3z. The standard itself covers only the details of the individual point-to-point links and does not address any system issues of how Ethernet switches may be built or deployed. Designing a switched system involves choosing an appropriate architecture, selecting the best available silicon technology and determining suitable electronic packaging.

The EU funded ESTA project was established to perform the basic research and development that would define the specifications for building an Ethernet switch to meet 10Gbps switching requirements. The part of the project carried out at CERN, within which this thesis work was done, involved the development, construction and validation of the backplane support for this 10Gbps Ethernet switch. In addition the program required that there be developed an understanding of where the limits of performance lay in terms of materials, manufacture and packaging technologies.

The most performant switches are built in a modular form consisting of a chassis and a number of plug-in modules. These modules plug into a common backplane which provides power, control, management and data interconnects. The functionality of such a system depends critically on the underlying interconnect even for such relatively low speed applications such as VME or Gigabit Ethernet. The backplane constitutes one of the weakest links of switching systems because it is subject to bandwidth limitations due to a limited number of connections and is also the source of signal degradation in the forms of amplitude and frequency loss, crosstalk, noise and reflections.

To meet the project requirements it would be necessary to systematically identify and quantify each aspect of the design process. This had to be done in terms of the effects on the system and its performance at 10Gbps as well for higher speeds. In addition to the theoretical study there was also going to have to be a practical method developed that would show that the required performance was not only met but exceeded. This was necessary to validate to what extend the theory converged with practice and also to ensure that the final constructed backplane was free from any possible system errors when running under maximum design conditions.

Content

1. Introduction ... 1
1.1 Context ... 1
1.2 Approach ... 1
1.3 Outline of the Thesis ... 2
2. Networking over Ethernet ... 4
2.1 Ethernet Networks ... 4
2.2 The Evolution of Ethernet ... 5
2.3 The 10 Gigabit Ethernet Standard ... 7
2.3.1 Specification ... 7
2.3.2 Physical Layer ... 8
2.3.3 Range of Media ... 8
2.3.4 Physical and Data Link Interfaces ... 9
2.3.5 10 Gigabit Ethernet Patterns ... 10
2.4 Summary ... 11
3. Design and Considerations for a 10 Gigabit Ethernet Switch ... 12
3.1 High-Performance 10 Gigabit Ethernet Switches ... 12
3.1.1 Architectures ... 12
3.1.2 High-Speed Backplanes ... 14
3.2 High-Speed Design Challenges ... 15
3.2.1 PCB Loss ... 15
3.2.2 Inter-Symbol Interference ... 16
3.2.3 Impedance Discontinuities ... 17
3.2.4 Crosstalk ... 19
3.3 Approaches in Achieving 10 Gigabit ... 20
3.3.1 Serial versus Parallel ... 20
3.3.2 Single Ended versus Differential Signaling ... 21
3.3.3 Silicon Technology ... 21
3.3.3.1 SerDes Devices ... 21
3.3.3.2 Output Drive ... 22
3.3.3.3 Signal Conditioning by Using Pre-Emphasis ... 24
3.3.3.4 Pre-Emphasis Implementation in the Transmitter ... 25
3.3.3.5 Ethernet SerDes Functionality for Testing ... 27
3.3.4 High-Speed Connectors ... 28
3.4 Summary ... 29
4. A Choice of Technologies for the High-Speed Backplane System ... 30
4.1 Connectors ... 30
4.1.1 Connector Technology ... 30
4.1.2 Connector Selection ... 32
4.1.2.1 Tyco HM-Zd or ERNI ErmetZD ... 33
4.1.2.2 Teradyne GBX ... 34
4.1.3 Connector Models ... 34
4.2 PCB Traces ... 35

4.2.1 Differential Trace Configurations 35
4.2.2 Controlled-Impedance PCB Traces 36
4.2.2.1 Dielectric Material Influence 36
4.2.2.2 Trace Geometry Influence 38
4.2.2.3 Manufacturing Tolerance Influence 39
4.2.3 PCB Trace Models 40
4.3 I/O Performance 42
4.3.1 I/O Solutions and Our Choice 42
4.3.2 Marvell Alaska 88X2040 42
4.3.3 Study and Validation of the Marvell I/O Characteristics 43
4.3.3.1 Marvell I/O Model 43
4.3.3.2 Simulation Environment 43
4.3.3.3 Validation 44
4.4 Summary 46
5 The Design, Modeling and Simulation of a 10Gbps Backplane 48
5.1 Backplane Design 48
5.1.1 Routing Constraints on the Backplane 48
5.1.2 PCB Trace Constraints 50
5.1.3 Backplane Stack-up 53
5.2 Via Hole Terminology 55
5.2.1 Via Hole Diameter 56
5.2.2 Via Pad 56
5.2.3 Via Anti-Pad 57
5.3 Connector Pin-field Modeling 58
5.3.1 The Concerns 58
5.3.2 Defining the Problem 59
5.3.3 Establish Simulation Environment – Quasi-static vs. Full-wave 60
5.3.3 Calibrating the Simulation Environment 61
5.3.3.1 Edge-Coupled Striplines 62
5.3.3.2 Single VIA 63
5.3.3.3 Coupled VIAs 65
5.3.3.4 Conclusions on the Calibration Study 66
5.3.4 Full-wave Connector Pin-field Modeling 66
5.3.4.1 The Approach 66
5.3.4.2 Crosstalk in the Connector Pin-field Area 67
5.4 Simulating a Realistic Backplane Environment 69
5.4.1 The Realistic Backplane Environment 69
5.4.2 Full-path Transmission Channel on the Backplane 70
5.4.3 Study of the Crosstalk Effect 71
5.4.3.1 Influence of the Routing Topology 72
5.4.3.2 Influence of the Aggressing Signals 72
5.4.3.3 Conclusions on the Study of the Crosstalk Effects 73
5.4.4 Simulating the 10 Gbps Transmission Channel 73
5.4.4.1 Influence of the Connector's Pin-fields Area 74
5.4.4.2 Influence of the Marvell Receiver and its Package 74
5.4.4.3 Influence of the Routing Topology 75

5.4.4.4 Influence of the PCB Trace Lengths ... 76
5.4.4.5 The Influence of Pre-Emphasis ... 76
5.4.4.6 Conclusions on 10Gbps backplane simulation ... 78
5.5 Limits to Performance ... 79
5.5.1 Introduction ... 79
5.5.2 Current Channel Performance ... 79
5.5.3 Future Channel Performance ... 80
5.5.3.1 Silicon ... 81
5.5.3.2 Multi-level Signaling ... 81
5.5.3.3 PCB Technologies ... 81
5.5.4 Performance Conclusions ... 83
5.6 Summary ... 83
6 Advanced Telecom Computing Architecture ... 84
6.1 AdvancedTCA Compatibility Study ... 84
6.1.1 AdvancedTCA ... 84
6.1.2 Comparison Study ... 85
6.1.2.1 The Study ... 85
6.1.2.2 Connectivity ... 86
6.1.2.3 Conclusions on the Compatibility Study ... 87
6.2 AdvancedTCA Backplane Architecture ... 87
6.2.1 AdvancedTCA Architecture ... 87
6.2.2 Fabric Interface ... 88
6.2.3 Base Interface ... 89
6.3 Summary ... 90
7 AdvancedTCA Backplane Tester ... 91
7.1 Context and Motivation ... 91
7.2 AdvancedTCA Backplane Tester Hardware Design ... 92
7.2.1 Traffic Generation on Fabric Channels ... 92
7.2.1.1 Fabric Interface Backplane Routing ... 92
7.2.1.2 SerDes for Traffic Generation ... 93
7.2.1.3 SerDes Implementation in the Tester ... 94
7.2.2 AdvancedTCA Backplane Tester Control ... 96
7.2.2.1 MDIO Interface ... 96
7.2.2.2 Point of Control ... 98
7.2.2.3 Backplane ... 99
7.2.2.4 Control Signals ... 100
7.2.2.5 Control Architecture ... 102
7.2.2.6 Microcontroller Implementation ... 103
7.3 AdvancedTCA Backplane Tester Boards ... 104
7.4 AdvancedTCA Backplane Tester Control Software Design ... 106
7.4.1 Software Architecture Model ... 106
7.4.2 Software Architecture Design Issues ... 107
7.4.3 User Interface ... 108
7.4.4 Client Application ... 110
7.4.5 Server Application ... 111
7.5 Summary ... 112

8 Backplane Tester System Integration 113
8.1 Control Software Validation 113
8.2 Hub1 Board Testing 114
8.2.1 Communication with the Micro-Controller 114
8.2.2 Testing the Client-Server Communication 114
8.2.3 Traffic Generation 115
8.3 Backplane Tester Exploitation 115
8.3.1 Commissioning Problems 115
8.3.2 Clock Fan-Out Board 117
8.3.3 Final Successful Integration 119
8.3.4 BER Measurements 120
8.3.5 Signal Integrity Evaluation 121
8.3.5.1 High-Speed Measurements 121
8.3.5.2 Test Conditions 121
8.3.5.3 Effect of Pre/De-Emphasis 122
8.3.5.4 Effect of Track Length 127
8.3.5.5 Crosstalk 127
8.3.6 Simulations versus Measurements 128
8.3.6.1 Effect of Pre/De-Emphasis 128
8.3.6.2 Effect of Track Length 130
8.3.6.3 Conclusions on the Simulation-Measurement Correlation 130
8.3.7 Effect of Adding Further Loss 131
8.4 Summary 132
9 Conclusions and the Author's Contributions 134
9.1 Conclusions 134
9.2 The Author's Contributions 137
9.3 Future Work 138
Appendix. Modeling with HFSS 139
References 143

1. Introduction

1.1 Context

An unprecedented demand for high-volume data switching at higher data rates is part of the recent innovation in telecommunication technology. For achieving the new technology requirements of ever increasing data transfer rates, two main directions of evolution have been taken in the industry. A first step was the move from previous shared-bus architectures, suffering from the bottleneck in the bandwidth and the limited interconnectivity available, to point-to-point switch fabric architectures. The second step was to increase the operating speed of the serial links, supported by newer and faster communication protocols and standards. One of those standards answering to the industry-wide requirements is 10 Gigabit Ethernet [IEEE-02], and much of the infrastructure of tomorrow's new telecommunication systems for high speed data transfer is based on 10 Gigabit Ethernet switch equipment.

The work presented in this thesis was carried out at CERN within the framework of the European Commission funded ESTA project [ESTA]. The ESTA project started while the 10 Gigabit standard was still being finalized, as collaboration between 3 partners from European Research Institutions and industry based in Israel (BATM), Denmark (Niels Bohr Institute) and Switzerland (CERN). The objectives of the project were to promote and validate the emerging 10 Gigabit (Gbps) Ethernet standard and evaluate its impact. To achieve this goal, an architectural study of a prototype 10 Gigabit Ethernet switch was undertaken within the ESTA project. This prototype switch architecture uses a fabric infrastructure to interconnect a large number of 10Gbps ports/modules over point-to-point serial links across a passive copper backplane. The work part of this thesis carried out at CERN included the high-speed signal integrity aspects of the 10Gbps Ethernet switch backplane (in its first part), and the design, the construction and the exploitation of a tester system, named later on throughout the thesis as the AdvancedTCA backplane tester (in its second part).

1.2 Approach

The present generation of telecom systems is based on modular solutions and rely on high-speed point-to-point interconnections which are running across a passive backplane. This thesis provides a detailed study of the performances of high-speed interconnects which are encountered by the backplane system of the prototype 10 Gigabit Ethernet switch designed and built within the ESTA project. This backplane is loaded with a few hundreds of bi-directional, point-to-point serial interconnects, each of them operating at 3.125Gbps as they are sent over the XAUI serial interface defined in the 10 Gigabit Ethernet standard. The objective of this thesis was to provide a design proposal for the Ethernet switching backplane and to demonstrate that this design would meet the high-speed transmission requirements.

The performance and reliability of the basic interconnect technology over printed circuit board (PCB) transmission media on the backplane was examined. The accent was to study the impact of the signal integrity aspects and to quantify their influence on the overall transmission across the switching backplane. To validate the backplane design proposal, we established a methodology which allowed us, through extensive modeling and simulation, to characterize the design in a pre-layout phase and to establish that the interconnect across the 10 Gigabit backplane can sustain error-free operation.

Advanced Telecom Computing Architecture (AdvancedTCA) standard was realised more or less in the same time with the finalization of our study and after a careful evaluation we determined that the standardized backplane design was very close to our proposal. It appears to be less expensive to follow the already defined architecture and therefore the AdvancedTCA standard has been adopted for the 10 Gigabit Ethernet switch. From the beginning of adopting this solution however, we feared about possible errors in the final prototype switch which could be caused by a low quality AdvancedTCA backplane.

A 10 Gigabit Ethernet traffic generator based on the AdvancedTCA architecture has been designed and constructed in order to test, validate and completely characterize the backplane defined by the standard. This system is intended to test the AdvancedTCA backplane by completely populating with 10 Gigabit Ethernet controllable traffic every single interconnect existent on the backplane and determining its status in terms of Bit-Error Rate measurements. The AdvancedTCA backplane tester allows the interconnect performance to be studied as a function of different traffic type and high-speed signal parameters, included those expected to be used in high-performance telecom systems. At the time when this system was developed it is believed it was unique in its ability to measure AdvancedTCA backplane performances under well controlled and well defined conditions. Furthermore, we have evaluated the integrity of the signals transmitted across the AdvancedTCA backplane by the mean of high-speed measurements. These measurements are the link between the two parts of this thesis since they can be compared with the simulation results obtained by following the described methodology in the first part.

1.3 Outline of the Thesis

Following the introduction, chapter 2 gives a brief overview of the networking over Ethernet and introduces the 10 Gigabit Ethernet standard.

Chapter 3 presents some of the theoretical aspects appearing later on throughout the thesis. It describes the importance of the 10Gbps systems in today's telecommunication equipments and the signal integrity challenges in the transmission of multi-Gigabit signals. In addition, the chapter presents the available solutions for achieving data rates of 10Gbps and the facilities for testing their integrity.

Chapter 4 reports and justifies the choices made for the I/O components, the high-speed connectors, the PCB trace geometry and the routing scheme of the high-speed signals passing across the 10Gbps backplane. For the chosen I/O 10Gbps transceivers, special

attention is given to the validation through our own simulations of the companies published results. For the high-speed connectors, several specifications are analyzed and the choice made for the 10Gbps backplane is justified.

Chapter 5 presents the modeling and simulation approach which have been undertaken for the design of the 10Gbps Ethernet backplane. A routing proposal is made for the backplane and it is checked for its integrity by including in the simulation chain the most critical aspects of the interconnect for which corresponding full-wave models have been created previously. The modeling environment is calibrated and the full-wave modeling procedure of the connector pin-field regions is described in detail. Different design criteria are simulated and as a result, design rules are provided which can sustain error-free operation across the 10Gbps switching backplane. This chapter also includes a study of the impact of the next generation technology beyond 10Gbps, towards a possible increase to 40Gbps bit rate. Chapter 5 concludes the first part of the thesis by providing an engineered solution of a suitable design proposal for the 10Gbps Ethernet switching backplane.

Chapter 6 is the start of the second part of the thesis. It introduces the emerging AdvancedTCA architecture and shows a comparative study with our proposed solution for the backplane design. The conclusion is that the 10Gbps backplane will follow the AdvancedTCA standard definition.

Chapter 7 presents the context and the motivation of building a tester for the AdvancedTCA standard backplane. The architecture, the hardware design and the software architecture model are all described here.

Chapter 8 deals with the backplane tester system integration, automatically performed Bit Error Rate (BER) tests on the AdvancedTCA backplane and the signal integrity measurements of the high-speed signal across the backplane. This chapter is closing the loop between the two parts of the thesis, by making a comparison between the simulations performed previously and the practical measurements of the signals.

Finally, chapter 9 gives a summary of the conclusions and author's contributions presented throughout this thesis.

2. Networking over Ethernet

2.1 Ethernet Networks

Today, Ethernet is the de-facto standard for local area network (LAN) connectivity. Ethernet owes its success to its cost effectiveness, simple management, and continuous evolution while maintaining backward compatibility with previous versions. Ethernet has continuously evolved to faster speeds, by a factor of ten each time, and greater distances. Ethernet also added more functionality which allowed it to extend beyond the LAN and find applications in the Metropolitan Area Networks (MAN) at city scale and even Wide Area Networks (WAN) at country to intercontinental scales, as it is the case of 10 Gigabit Ethernet.

The International Standards Organisation defined a model for interconnecting computing systems, known as the Open System Interconnect (OSI) model [OSI-94]. The OSI model divides the space between the user interface and the transmission medium into seven vertical layers. The data is passed between layers top-down and bottom-up in such way that a layer communicates directly only with the layer above. The OSI model is presented in Figure 2.1.

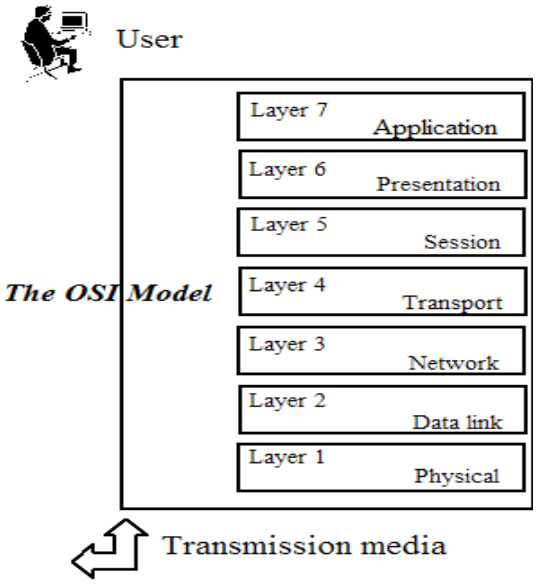


Figure 2.1: The OSI model

The physical layer interfaces directly the computer to the transmission medium. It converts the bit stream received from the upper layer into signals adapted to an efficient transmission over the physical medium. The data link layer provides a managed virtual communication channel. It defines the way the transmission medium is accessed by the connected devices, detects and possibly corrects errors introduced by the lower layer and handles the control messages and protocol conversions.

2.2 The Evolution of Ethernet

Ethernet is a technology that spans the first two layers of the OSI model: the physical layer and the data link layer.

Invented in the research laboratories of Xerox in 1973, Ethernet was first standardized by IEEE in 1985 [ETH-85] for the transmission speed of 10 Mbps. The official denomination was 10BASE5 (10 Mbps data rate, base band transmission, network segments having a maximum length of 500 meters). Each segment consisted of a single co-axial cable that was shared by all nodes on that segment. Ethernet defines a mechanism known as Carrier Sense Multiple Access with Collision Detection (CSMA/CD) [TAN-96] used for the arbitration in the transmission between two or more end-stations. For this mechanism to work the segment length must be limited, as imposed by the CSMA/CD, to about 500meters at 10 Mbps.

It is well known that the electrical signals on the segment are attenuated and degraded by cable losses and external signal interference while propagating along the cable. To extend a LAN beyond the 500m limit, devices called repeaters were introduced in order to clean and amplify the signal, and prevent arbitration signals from propagating between segments. Figure 2.2 presents the result of the rules that directed the design of the first Ethernet networks.

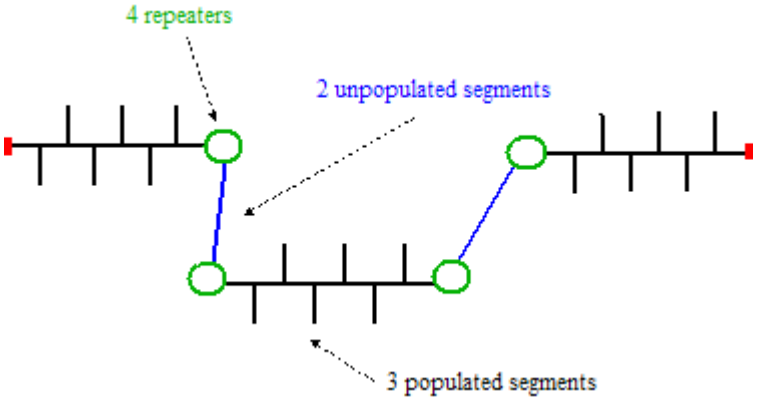


Figure 2.2: Illustration of the 5-4-3-2-1 Ethernet rule

The design of the first Ethernet networks was governed by five simple principles, known as the "5-4-3-2-1":

- 5 segments, each segment spanning maximum 500 meters;
- 4 repeaters to interconnect the segments;
- 3 segments populated with nodes for a maximum of 100 nodes per segment;
- 2 segments cannot be populated with nodes;
- All the above form a 'collision domain' or logical segment.

The number of connected devices increased as personal computers became more affordable for companies. However, the limitations due to the CSMA/CD operation imposed serious performance limitations, further limiting the achievable transfer rate on Ethernet segments with a large number of nodes.

One approach for reducing the number of nodes in a segment was the introduction of active network devices called bridges. The bridge interconnected two segments, thus effectively doubling the span of the network (Figure 2.3).

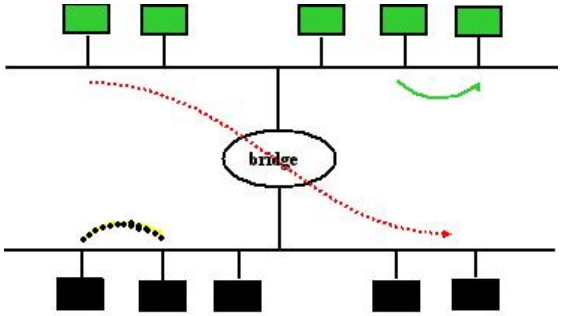


Figure 2.3: Bridged Ethernet network

The bridges were first defined by IEEE in 1990 with the IEEE 802.1D standard [BRI-90]. In the same year with the bridging standard, the IEEE introduced the 802.3i standard [UTP-90] to define Ethernet transmission over the Unshielded Twisted Pair cable (UTP). The UTP cable could be used for full-duplex transmission. The topology of the network segment changed from a shared bus to a star and the device positioned at the centre of the network was either a hub or a switch.

The hub was a special type of multi-port repeater that had a single network node connected to each one of its ports. It just replaced the shared coaxial cable but brought no additional functionality. From the logical point of view, the entire network was still a shared bus. The communication on a segment was still made in half-duplex mode, hence limited to 10 Mbps shared between all connected devices.

The switch (Figure 2.4) was a multi-port bridge that interconnected any port to any other port. The presence of a switch enabled full-duplex connections, hence practically doubled the bandwidth at each node and greatly increased the total bandwidth available on the network. All the modern Ethernet networks are built using switches.

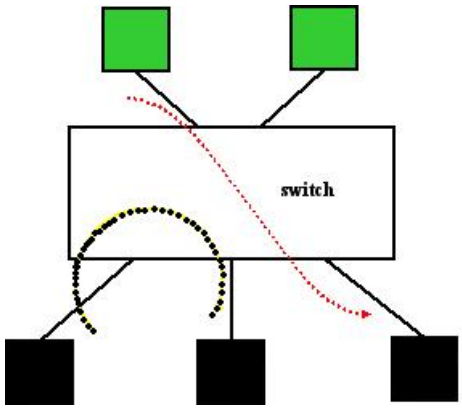


Figure 2.4: Switched Ethernet network

Transmission of Ethernet over optical fiber was introduced in 1993 by the IEEE 802.3j standard [FIB-93]. An upgrade in speed followed in 1995, increasing the available

bandwidth to 100 Mbps (Fast Ethernet) [FAST-95]. Initially, 100Mbps Ethernet was with half-duplex operation.

The biggest step forward in the Ethernet evolution was the transition from half duplex CSMA/CD to full duplex in 1997 [FRA-99]. Since then, Ethernet has proven an evolutionary upgrade path to high speed. The transmission speed was further increased in 1998 to 1 Gbps by the IEEE 802.3z specification (known as Gigabit Ethernet) [GIG-98]. At the time of adoption, Gigabit Ethernet was introduced over optical fiber in 1998, and subsequently over standard UTP copper cable in 1999 (802.3ad standard). The Gigabit Ethernet standard was the first Ethernet standard that did not target workstations as the primary user. The backbones of the data centers and small campus networks were addressed instead. The optical components defined by the 802.3z standard allowed for a maximum of 5 km distance between the two endpoints of a point-to-point connection. However, non-standard components allowed for up to 10 km and, in 2003, for up to 100 km distances. The Gigabit Ethernet standard kept the legacy of the half-duplex operation and CSMA/CD to provide backward-compatibility support for devices operating in half duplex mode, but in practice nobody ever used half-duplex Gigabit Ethernet. Starting with the Gigabit Ethernet standard, Ethernet practically became a technology that only uses full-duplex point-to-point connections and a switch-centered star topology for the network.

10 Gigabit Ethernet, adopted in 2002 under IEEE Draft P802.3ae [IEEE-02], continues the natural evolution of the original Ethernet model in speed and distance. The 10 Gigabit Ethernet only supports full-duplex operation having no longer an inherent distance limitation. Henceforth, the distance of a point-to-point connection will only be limited by the optical components used for transmitting and propagating the signal.

2.3 The 10 Gigabit Ethernet Standard

2.3.1 Specification

The architectural positioning of 10 Gigabit Ethernet within the OSI model is shown in Figure 2.5. Under the OSI model, Ethernet is fundamentally a Layer 2 protocol, but actually it addresses the two lower layers of the model: the Physical Layer (Layer 1) and the Data Link Layer (Layer 2).

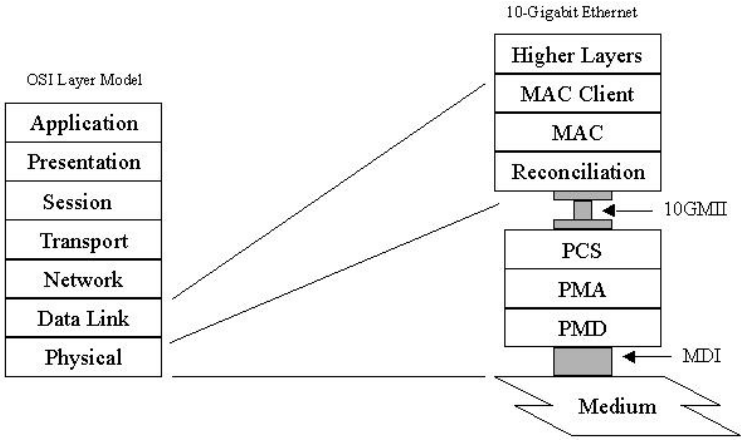


Figure 2.5: Architectural Positioning of 10 Gigabit Ethernet

2.3.2 Physical Layer

At the Physical layer, an Ethernet physical layer device (PHY) connects the optical or copper medium to the Data Link layer, represented by the Medium Attachement Controller (MAC device). Ethernet architecture further divides the physical layer into three sub-layers: Physical Medium Dependent (PMD), Physical Medium Attachment (PMA), and Physical Coding Sublayer (PCS).

The Physical Coding Sublayer (PCS) encodes 8-bit data bytes received from the MAC into 10-bit code groups, using a patented block-coding technique known as 8B/10B encoding. By using 8B/10B encoding, each 8-bit word is converted into a 10-bit word, which gives 20% of more codes to choose from than are actually needed. The choice is used to select specific codes that limit the number of consecutive ‘zeroes’ or ‘ones’ (known as the run length limit), and balances the total number of ‘zeroes’ and ‘ones’ in any given time frame (DC balance).

The 10-bit code-groups produced by the PCS are then transmitted one bit at a time in a serial fashion by the Physical PMA using a simple non-return-to-zero (NRZ) line coding. The NRZ waveforms are inherently time-domain two-level signals which assign a unique time slot duration to each transmitted bit. Each bit in a random NRZ sequence has a 50% probability of being a ‘one’ or a ‘zero’. The corresponding two-level signal is either high, representing a ‘one’, or low, representing a ‘zero’, during the entire bit period. A logic ‘one’ is defined as the high voltage state (on copper media) or high-intensity (on optical fiber), while a logic ‘zero’ is defined as the low voltage state or low intensity.
PMD sub-layers provide the physical connection and signaling to the medium.

2.3.3 Range of Media

Several physical layers were defined in the 10 Gigabit Ethernet standard. The only transmission medium supported by the original version of the 802.3ae standard, issued in June 2002, was optical fiber. Two types of optical fiber - multimode and single mode

fiber - are currently used as specified by 10GBASE-(X)(Y). 'X' denotes different wavelengths of the system and 'Y' specifies the encoding used. The distance supported for transmission at 10Gbps across the fiber vary based on the type of fiber and wavelength (nm) implemented in the application, as it is specified in the 10 Gigabit Ethernet standard.

The adoption of a standard for transmitting 10 Gigabit Ethernet over copper cable is currently scheduled for 2006. Here there are two directions defined by two different task forces: 10GBASE-CX4 and 10GBASE-T [INT-x].

10GBASE-CX4 specifies a solution for those copper cabling defined in the Infiniband Architecture Specification, by using signaling provided by the 10 Gigabit Ethernet attachment unit interface (XAUI) (more details about XAUI in paragraph 2.3.4). The Infiniband copper cables enable signaling over 8 shielded copper differential pairs (4 per direction) at 2.5 Gbps and are designated as (4X). The encoding scheme used is 8B/10B. The Infiniband copper cable interconnects are specified to 1.25 GHz. presently, the 10GBASE-CX4 standard is specified to operate over a physical distance of 15m.

10GBASE-T is competing against the fiber technology and 10GBASE-CX4. The potential advantage that 10GBASE-T will have over 10GBASE Fiber will be a relative cost reduction. The potential advantage that 10GBASE-T will have over 10GBASE-CX4 will be the ability to transmit over a greater physical distance. 10GBASE-T standard supports copper media that is representative of either class E (also referred to as category 6) or class F (category 7) UTP or FTP cabling. Category 6 is expected to achieve physical distances beyond 55m and up to 100m, while category 7 cabling can easily support maximum channel length requirements of 100m. The greater distance supported for transmission comes from the fact that the 10GBASE-T standard uses Pulse Amplitude Modulation (see paragraph 5.5.3.2) encoding, which allows a maximum bandwidth of 250MHz for category 6, respectively 600MHz for category 7 cabling.

The topic of this thesis, concentrates on the transmission of 10Gbps signals across copper media within a PCB and not on cabling assemblies outside the box. The transmission across copper is achieved on the PCB by using stripline or microstrip traces which transport the signals provided by the XAUI interface.

2.3.4 Physical and Data Link Interfaces

Ethernet has included an enhanced Media-Independent Interface (MII) below the MAC sublayer which is designated as 10 Gigabit MII (XGMII). XGMII is a 32-bits parallel transmit-and-receive synchronous data interface, which was intended to provide a chip-to-chip convenient demarcation between the data layer's MAC, and the physical layer's PHY.

The 10 Gigabit Ethernet standard also includes a specification for a serial interface, which is called the 10 Gigabit Ethernet Attachment Unit Interface and it is generally abbreviated as XAUI (pronounced "Zowie"). The 'AUI' portion is borrowed from the Ethernet Attachment Unit Interface. The "X" represents the Roman numeral for ten and implies ten Gbps.

The XAUI interface corresponds to the PCS layer of the OSI model and it was designed as a serial interface extender of the parallel XGMII. Similar to the XGMII interface, XAUI interface describes 10 Gbps Ethernet connections between the PHY and the MAC. The XAUI is multi-channel interface which supports four high-speed serial channels at 3.125Gbps each of them with 8B/10B encoding technique. A 3.125Gbps signal path in XAUI interface is a point-to-point differential pair connection. There are four differential paths in each direction - or sixteen connections full-duplex– for achieving the PHY-MAC connectivity at a raw throughput of 12.5GBaud, respectively 10Gbps total effective throughput when using 8B/10B encoding.

2.3.5 10 Gigabit Ethernet Patterns

The encoding scheme of any serial data stream results in a known, and bounded set of sequential bit patterns that can occur [RED-04]. These define the spectrum of the signal and also the bandpass characteristics of the medium required to carry them. Every medium will have loss and if the rules for establishing the cabling interconnect are followed strictly then there should be no problem with end-to-end communication. However it is possible that cable or connector losses are affected or modified during the lifetime of the installation and it is useful to be able to test the media for conformance with requirements. For signal transmission testing purposes [LAQ-02], the 10Gbps standard specifies diverse test patterns which allow stressing various aspects of system or component performance in different applications. The 10 Gigabit Ethernet standard defines two categories of patterns: pseudo-random bit stream (PRBS) and Jitter patterns [WAT-02].

PRBS patterns are defined as a family of general test sequences for NRZ applications, used to simulate random data for transmission across the link. Common practice denotes pseudorandom patterns as $2^X - 1$ PRBS, where the power X indicates the length of the shift register that creates that pattern. Each $2^X - 1$ PRBS contains every possible combination of X number of bits.

There are also five different jitter patterns defined in the 10 Gigabit standard for addressing different aspects of the system performance. Three different types of Jitter patterns (48A.1, 48A.2 and 48A.3) are testing random and deterministic components of the jitter at low, high and mixed frequency by transmitting repetitively code-groups corresponding to K28.7, D21.5 and K28.5. K28.7 is the low frequency test pattern which consists of alternating groups of five ones and zeros. This pattern includes the standard ‘comma’ sequence (0011111) that the receiver uses to frame the incoming serial stream. D21.5 is the high frequency test pattern, including the alternating 0-1 sequence. It is intended for testing random jitter and asymmetry of transition times. K28.5 is the mixed frequency pattern and also contains the standard ‘comma’. It is intended to test the combination of random and deterministic jitter. A summary of the existent patterns is given in Table 2.1.

Pattern Name	Pattern Bits	Notes
K28.7	0011111000	Low Frequency, Comma
D21.5	0101010101	High Frequency, Alternating
K28.5	0011111010110000010 1	Mixed Frequency, Comma
PRBS	User Defined	Pseudo-Random

Table 2.1: Summary of 10 Gigabit Ethernet Patterns

From a signal integrity point of view, K28.5 is the most interesting pattern since the sequence includes isolated ‘ones’ and ‘zeroes’ which stress the low-pass filter capability of the transmission channel.

2.4 Summary

Chapter 2 briefly introduces the evolution of Ethernet and places 10 Gigabit Ethernet in the context of preceding technologies. In particular it draws attention to the thinking and methods foreseen to determine if the transmission media is adequate for error-free operation at 10Gbps. This reflects concerns by the standards committee about how to validate a working transmission link.

3. Design and Considerations for a 10 Gigabit Ethernet Switch

The 10 Gigabit Ethernet standard presented in chapter 2 only addresses the point-to-point technology. How packets may be switched from one point to another inside the switch is left as an implementation issue to the switch manufacturer. Europe was perceived as a major market growth in this area but there was virtually no European presence among Ethernet product vendors. The objectives of the ESTA project were to deliver a know-how that makes the difference between the theory of the 10 Gigabit Ethernet standard and the practice upon which products can be developed. The most important practical goal of the ESTA project was to build a first prototype 10 Gigabit Ethernet switch.

3.1 High-Performance 10 Gigabit Ethernet Switches

Developing an architecture for a prototype 10Gbps Ethernet switch called for an understanding of current switching and transmission technologies. It also called for an appreciation of the engineering and manufacturing constraints to develop a scheme that could be reproduced at a commercially acceptable cost. An inappropriate switch design may well have the required non-blocking characteristics and manageable bandwidth but simply not work because the cabling, connector, signal integrity and timing constraints could not be met. Innovative design practices will be necessary to achieve a match of performance, cost and manufacturability ([MAD-x], [MCM-03]). The present thesis presents the design research and development needed to produce the 10Gbps backplane of the 10Gbps Ethernet switch.

3.1.1 Architectures

The typical way to package high-performance Ethernet switches is to use modular solutions within a chassis infrastructure. The chassis is a one off cost and the customer can then add modules to meet bandwidth requirements as the load increases over time. The 10Gbps Ethernet switch is developed within a chassis where a series of line cards together with one or several switch cards are plugged in and interconnected via a common backplane. The line cards are used to transfer system data to and from the outside world and the Switching Card is the central resource which provides the data distribution for the whole system.

Switch architecture design is a subject of much academic and commercial study and is not the topic of this thesis. A basic understanding of the problem however was required so that we could appropriately dimension the needs of the 10Gbps switch fabric infrastructure. The real key to the modular solutions is a point-to-point switch fabric architecture that is engineered for fault tolerance and scalability. The way that data is

switched between different line cards depends on the chipset(s) available. There are two main classes of switch depending on the choice of chipset. In the first case the switch functionality is split between what is port specific and that which is system specific. This leads to a star architecture (Figure 3.1), where part of the switching function is done on the line card and the rest in the central core switch which may be made up of one or several chips. The core switch or the switching card is in this case the central resource which provides the data distribution for the whole system.

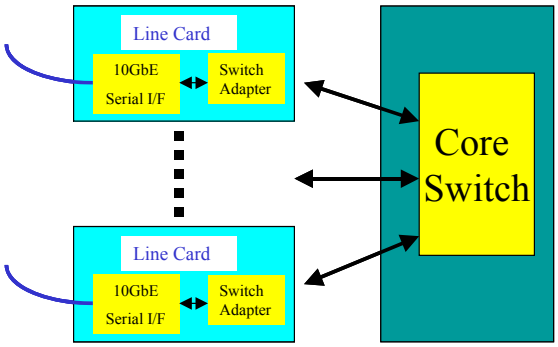


Figure 3.1: Star topology for a 10Gbps switch

One major drawback of this architecture is that the core switch card constitutes a single point of failure and so the preferred solution is the redundant dual-star topology (Figure 3.2). The dual-star provides the system reliability, so if one core switch fails then the switch can quickly be reconfigured to use the other one.

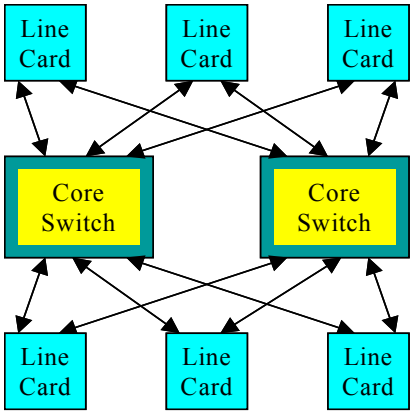


Figure 3.2: The dual-star topology

In addition to the star and the dual-star topology, there are also fully meshed interconnects (Figure 3.3) considered as a superset of the star topologies since it offers all–to-all connectivity at every node instead of needing to route through some intermediate node.

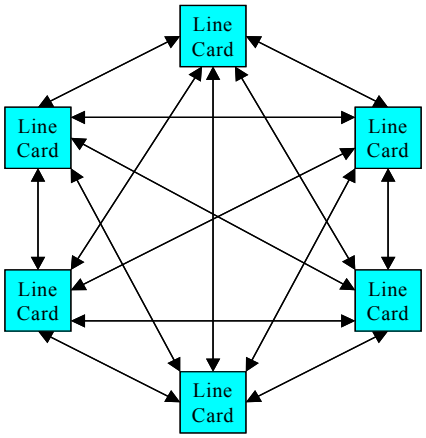


Figure 3.3: Mesh topology

In the mesh architecture, all of the switching functionality can be accommodated on the line card. Each line card populates a point-to-point connection until it builds up the complete switch where all cards have connections to all other nodes.

3.1.2 High-Speed Backplanes

The topology of choice for the 10Gbps Ethernet switch is the dual-star architecture because it achieves the required redundancy at a lower cost than using a full mesh where every line card has also to carry a part of the switching fabric. A switch however has to do more than just provide the nominal line speed between the switch fabric and a line card. The act of switching takes time to identify and route data between any two ports. This could lead to congestion if the bandwidth through the switch fabric was only equal to the bandwidth through the aggregate of the switch ports. The simplest and most effective way to avoid this problem is to over-provision the core switch so that it could run much faster than the subscribing ports could deliver. There are two ways to do this: by increasing the line speed or by providing more links for transmission across the backplane. Today's technology is already line speed limited, the only effective choice is to increase the number of links required to switch the data over the backplane, which in turn leads to a problem in routing the number of traces required. As the number of traces increases, the difficulties of routing those traces and maintaining their signal integrity also increases and the question must be asked, at what point does the backplane become the limiting factor of performance or reliability in the system as a whole?

The present thesis contains the work performed to design the backplane of a 10 Gigabit Ethernet switch by taking into consideration all the aspects having a direct impact on interconnect performance. The work takes into account the continuous trade-off between what is theoretically the best solution and which is the most cost-effective practical approach in building the 10Gbps Ethernet switch backplane.

3.2 High-Speed Design Challenges

The propagation of multi-gigabit signals through PCB interconnects comes together with a series of challenges [IEEE-10G_BP], discussed in more details in the present paragraph. The signal integrity of the copper transmission is affected by the PCB loss due to the skin effect and the dielectric loss, increased noise due to the crosstalk, impedance discontinuities and inter-symbol interference (ISI).

3.2.1 PCB Loss

One key problem, which applies to all copper transmission media are frequency dependant losses. The two electrical characteristics primarily responsible for frequency-dependent losses in PCB traces are skin effect and dielectric losses ([DEU-01], [Johnson-03]). Figure 3.4 shows the transmitted signal and the corresponding received signal, affected by the losses during its transmission.

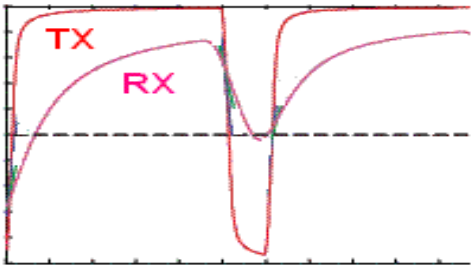


Figure 3.4: PCB Loss affecting the signal transmission

As the frequency of a propagating signal increases the current is no longer uniformly distributed through the conductor but is increasingly constrained to the outer surface of it. With less conductor cross-section now carrying the current the effective resistance increases. This increase is proportional to the square root of the signal frequency and causes attenuation of the signal as it propagates along the conductor. One approach to compensate for skin-effect losses is to simply increase the surface area by increasing the trace width. However, this simplistic solution is bounded by physical constraints. One is that wider traces require more board space to route them. As the space between component (connector) pins shrinks, the available wiring channels become narrower, reducing the number of traces through each channel. This results in additional wiring channels and probably additional board layers. Another limiting factor is that wider traces will require thicker boards to achieve the required characteristic impedance. Thicker boards in turn require, wider via holes (see paragraph 5.2) which increase signal loss and cost more to produce.

The other type of loss experienced in the PCB is due to the energy dissipation in the board dielectric, caused by the heat created at the molecular level during signal propagation in the PCB medium. This kind of energy dissipation manifests itself as a loss of signal magnitude and the degradations of the rising and falling edges of the signal. The parameter defining the energy dissipated is called the loss tangent. The signal attenuation

due to the surrounding insulating material is computed as the product of the loss tangent, the frequency, and the relative dielectric constant. The dielectric loss is increasing proportionally with frequency, and is thus the dominant factor at high-frequencies.

Both, dielectric loss and skin effect are directly proportional to the length of the transmission path and both increase with frequency. The Figure 3.5 shows the PCB losses encountered by 40 inches of stripline trace.

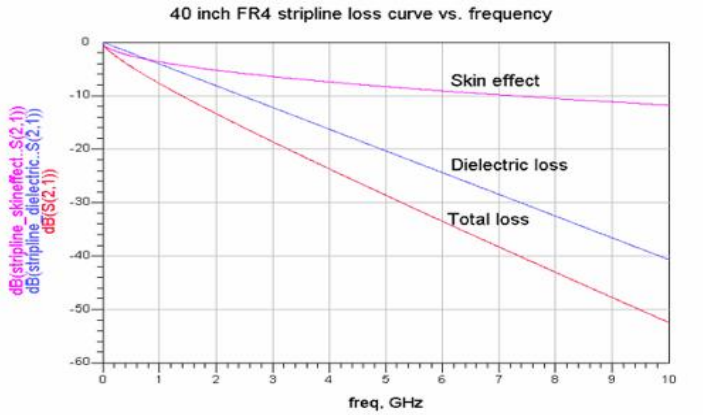


Figure 3.5: PCB Losses encountered in a 40inch of FR4 stripline (reproduced from [ALT-03])

The amount of the dielectric loss, very important at high-frequencies, is highly dependent on the quality of the dielectric material used to build the PCB layout. The classical dielectric material used in the PCB manufacturing process, FR4, has a significant loss tangent value, which can result in large signal deterioration. Many materials that are used as PCB laminates have better loss tangents ([WAL-x], [HUG-05]), but even in volume production, these materials are still cost-prohibitive. The object of this thesis is to study and understand the impact on the performance of the transmission channel of the PCB losses when classical FR4 dielectric material is used for building the backplane of the 10Gbps Ethernet switch.

3.2.2 Inter-Symbol Interference

As a multi-Gigabit signal propagates through a lossy dielectric transmission line, the frequency components of the wave front travel at different velocities – low frequency components travels largely unaffected while the high frequency components are attenuated by the transmitted channel. Due to the effects of the dielectric loss combined with the low pass filter characteristics of the transmission channel, the edges of the signal are slowed down, behaviour which results in a smearing of the wave front of the digital signal, a phenomenon which is called dispersion. The dispersion results in inter-symbol interference (ISI) which can be defined in the time-domain by analyzing the response of a single-bit passing through a transmission channel (Figure 3.6).

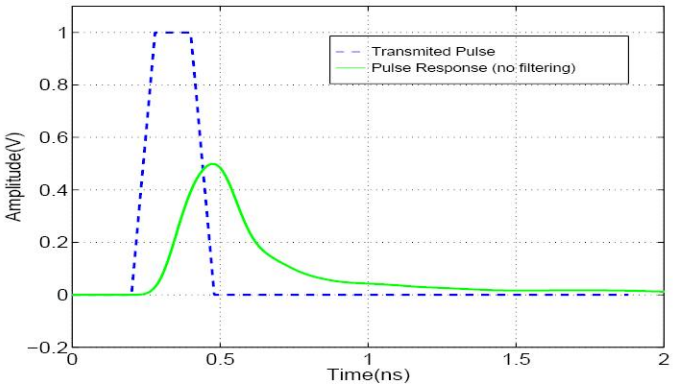


Figure 3.6: ISI impact on a signal pulse (reproduced from [WOO-02])

The received pulse waveform is both attenuated and has a long tail due to the spreading in the time section corresponding to the next bit. The signal spreading in the next-bit slice has destructive effect especially when it is part of a bit pattern. The Figure 3.7 demonstrated the destructive effect of ISI on a simple '101' pattern transmitted down a lossy channel to the receiver.

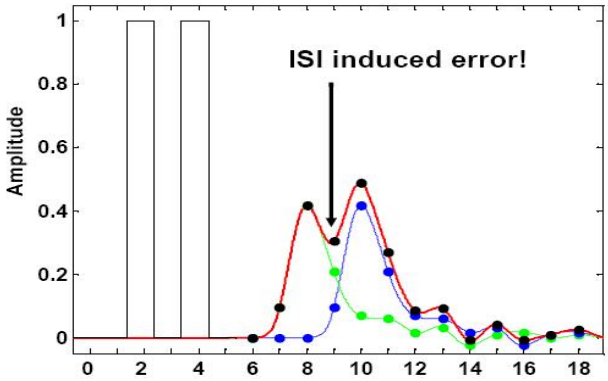


Figure 3.7: ISI effect on the '101' pattern (reproduced from [ZER-04])

Due to the dispersion phenomenon, the voltage level at the receiver is not only determined by the value (0 or 1) of the transmitted bit, but also by the value of the previous bits. The resulting error is induced in this case by 'pre-cursor' ISI (blue) added to the 'post-cursor' ISI (green). Their sum will produce a voltage for the '0' bit significantly above the 0 voltage threshold.

A way to compensate the ISI effects is to cancel the tail of the pulse by applying at the transmitter an inverse characteristic of the transmission channel itself. The implementation is done with a FIR filter at the signal pulse implemented in the pre-emphasis driver (see paragraph 3.3.3.4).

3.2.3 Impedance Discontinuities

A perfect passive channel would exhibit zero loss or noise. However, that is not reality. Aside from high-frequency losses and ISI, the Bit Error Rate (see paragraph 3.3.3.5) of a transmission channel is significantly affected by impedance mismatches. Such

mismatches can be seen at any point along the path between the transmitter and the receiver, where the signal encounters a geometric structure change in the media. The causes for impedance mismatches are mainly the manufacturing problems, design problems and passive components (such as connectors and vias) tolerances. A complex, high-density design such as the Ethernet switch system is likely to have the potential for many of these discontinuities because of the number of PCBs and connector systems through which the signals must transition. Connectors, plated-through-holes for connector pins, vias for transitioning layers and differing board impedances, all present impedance discontinuities that can have varying effects on high-speed signals' integrity.

The primary effect of an impedance discontinuity is the reflection of the signal energy back to the source, which adds to or subtracts from the incident signal. The magnitude of the reflection is directly proportional to the impedance mismatch; i.e., the larger discontinuity, the more damaging the effect on the signal, and more difficult for the transceiver device to counteract the results.

The driving parameter for end-to-end impedance is typically given by a constant differential trace impedance of 100 Ω ± 10%, which must be supported by all high-speed PCBs in the transmission path from driver to receiver. Any mismatched above or below 10% of 100Ω results in ringing and reflections which are added to the nominal value of the impedance. A common technique used to study the impedance discontinuities across a transmission channel and to determine their source is the Time Domain Reflectometry (TDR) [OLT-02]. A rapid rise-time pulse is injected into a transmission line and reflections from any discontinuities and changes in impedance are probed close to the point of injection and displayed on an oscilloscope. Figure 3.8 illustrates a typical TDR measurement for a transmission channel across a backplane.

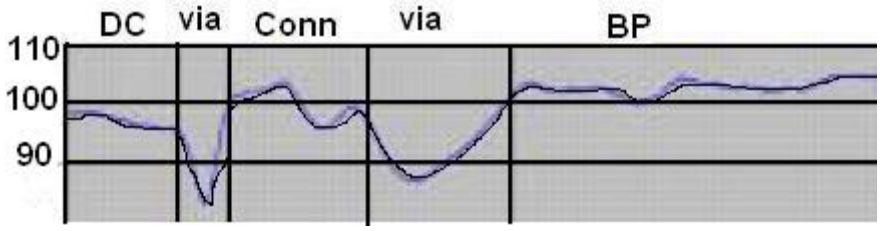


Figure 3.8: TDR measurements for a backplane channel

Even though the backplane itself seems to meet the controlled-impedance requirements, the impedance discontinuities exercised in the via region exceed the maximum admissible limit.

An estimation of the limitation of the channel response limitation due to the connectors, vias and devices along the path, in addition to the PCB losses, is shown in Figure 3.9.

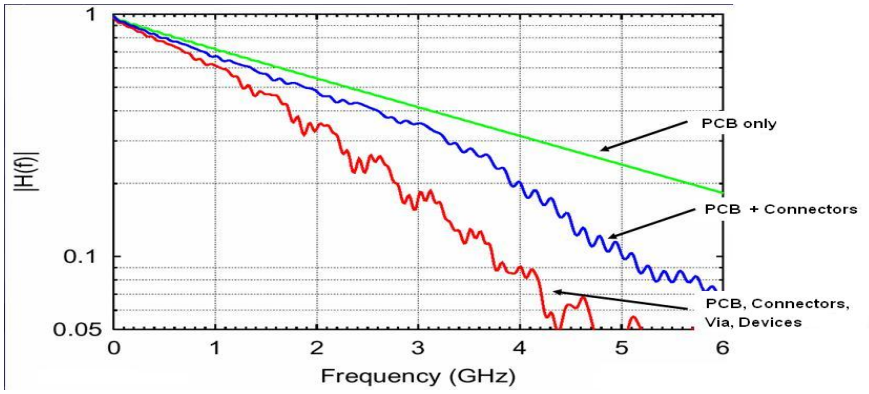


Figure 3.9: Estimation for a transmission channel limitation (reproduced from [ZER-04)

The response of the channel is limited due to the discontinuities and impedance mismatches, mainly created by the connector components and the vias along the transmission path. These mismatches are getting more important at higher frequencies and they affect the overall quality of signal transmission. Therefore, they must be address and minimized as much as possible in the design of high-speed applications.

3.2.4 Crosstalk

Crosstalk occurs in a passive transmission channel when a signal on one copper trace couples with a signal on another copper trace. This generally happens when two single or differential pairs of traces run in close proximity parallel to each other for an appreciable distance. The causes of the crosstalk effects are the inductive and capacitive coupling, determined by the electromagnetic fields which appear between two adjacent traces.

Crosstalk from an 'aggressor' trace onto a 'victim' trace propagates in both directions away from the point of coupling: that which follows the direction of the aggressor signal is known as forward crosstalk (FEXT) and that which flows in the opposite direction is called near end crosstalk (NEXT) [DEU-98]. Both these two components are heavily dependent on the spacing between the pairs and on the aggressor signal switching rate [OLT-02].

In addition to the crosstalk effect discussed so far, which appeared between adjacent traces, some other possible sources of crosstalk in a modular backplane-based system (Figure 3.10) can be the PCB traces which couples with connector pins on the backplane or the connector itself.

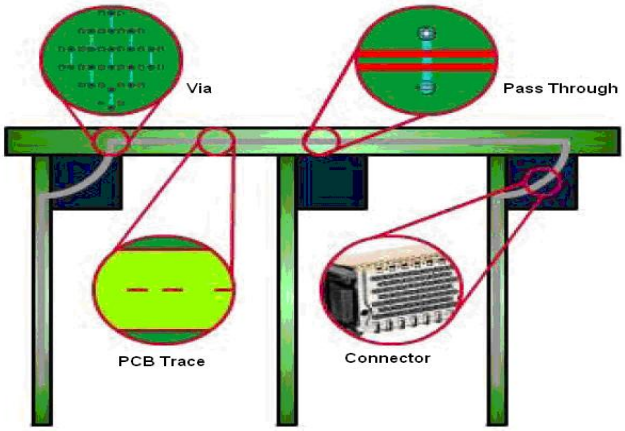


Figure 3.10: Sources of crosstalk

Crosstalk from adjacent traces on the PCB, can be avoided or reduced by close attention to conductor layout. However the crosstalk generated inside a connector is entirely a function of the selected component, since it is the connector geometry that determines the position of adjacent differential pairs relative to each others. The new generation of high-speed connector have low intrinsic crosstalk, as their design considerably reduces the crosstalk effect by surrounding each pair of signal pins with a ground shield which is both present in the board section and in the mating interface to the backplane, providing this way continuous shielding across the connector body (for details see paragraph 4.1).

Regarding the amount of crosstalk introduced in the signal traces at the breakout of traces from the connector to the PCB and in the traces which passes across the connector pins on the backplane, there are still great challenges and unresolved issues which would need further study. This thesis will consider and study the amount of crosstalk introduced in the connector pin-field region on the backplane (see chapter 5).

3.3 Approaches in Achieving 10 Gigabit

The problems described so far are well known and different approaches in the silicon, the connector technology, and in the signaling methods have been developed in order to enhance the performance of high-speed transmission across standard copper structures.

3.3.1 Serial versus Parallel

The current generation of 10 Gigabit Ethernet silicon transmission and switching components uses the XGMII parallel interface. This has a high pin count and is useful only for transmission over short distances for several reasons. Firstly it is difficult to maintain equal length conductors for every member of a parallel bus. Small differences in conductor length translate into skew between the arrival time of signals and the clock rate has to be reduced to compensate for this. Secondly it is impossible to avoid cross-talk

effects that affect one or more members of a bus from sources such as adjacent clock lines or other busses, unless the bus is constrained to areas of the printed circuit board or backplane that excludes any such 'aggressor' signals. This rapidly becomes untenable in terms of available routing space. Finally the high pin count connectors needed per link will occupy more backplane area than can conveniently be built. While acceptable at lower speeds, a parallel architecture fails to keep up with the demands of a 10 Gigabit Ethernet design.

The solution of choice is the use of a high-speed serial interface. Serial protocols have the clock embedded in the data stream so a single transmission path can run faster and over longer distances without the skew problems inherent with parallel busses. A 5Gbps serial bus can easily drive up to 50cm of transmission line whereas the XGMII parallel bus delivering the same bandwidth is limited to only 7cm. The upper limit for transmission distance is now more a function of the mediums loss than any routing concerns.

3.3.2 Single Ended versus Differential Signaling

A single ended transmission line is subject to crosstalk effects (see paragraph 3.2.4). A single ended receiver compares the incoming signal level with an expected logical zero or one threshold. This makes it vulnerable to noise that can add or subtract from the signal and cause a false output.

Differential transmission across the PCB is performed using two identical matched trace conductors transporting two complementary signals on each line of the differential pair. The differential receiver looks for the differences between these two signals and therefore any noise common to both signals is rejected. For Gigabit high-speed applications, differential transmission is mandatory. The differential transmission line designs however, have strong requirements for controlled-impedance geometry accuracy since any difference between the two lines is itself a source of differential noise.

3.3.3 Silicon Technology

3.3.3.1 SerDes Devices

To keep up with the market requirements for increased performance, chip-to-chip and board-to-board interconnections increasingly employ high-speed serial solutions. So, parallel-form data (coming from XGMII) has to be converted into a serial form and transmitted using high speed serial links (corresponding to XAUI) over backplanes. The connection operates by taking parallel signals in on the 'local' side of PCB and converting them to a serial bit stream on the backplane side. The XAUI interface employs high-speed serial transceivers, known as Serializer-Deserializer (SerDes) devices, which perform the serial-to-parallel conversion to support four high-speed serial channels at 3.125Gbps [NEL-02].

A SerDes device consists of a transmit data path and a receive data path. The basic components of a SerDes device are illustrated in Figure 3.11.

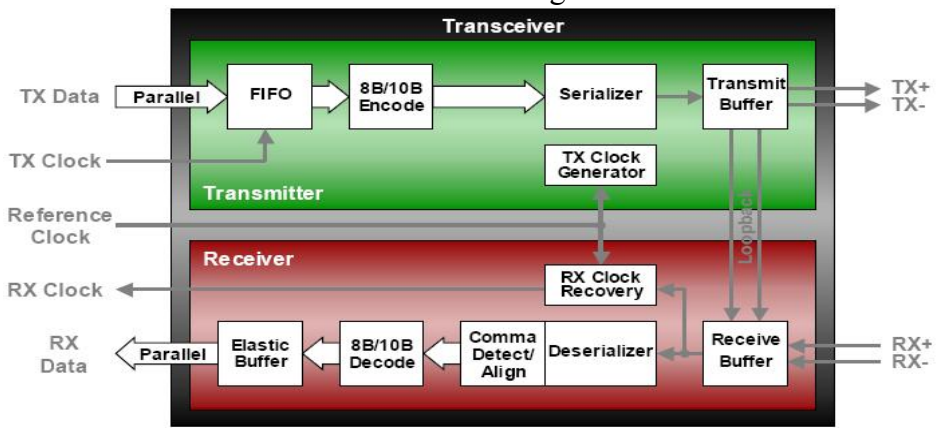


Figure 3.11: Basic components of a SerDes device

In the transmit portion of the device, a data-input buffer accepts parallel data, which are encoded, serialized, synchronized with a transmission clock which is embedded into the serial data signal that is transmitted down on the line. A characteristic of high-performance SerDes devices is encoding, used within data transmission because of the ways it facilitates clock synchronization and error correction. There are many encoding methods, the one used by the XAUI interface protocol in the Ethernet standard is 8B/10B.

The receive data path in the SerDes takes the high-speed serial stream of data from the backplane side and does exactly the complementary operation as the transmitter: it deserializes the input and presents it in parallel form for the 'local' side of the PCB. The receiver part of the SerDes contains a deserializer together with clock-and-data-recovery (CDR) circuitry to reconstruct the encoded parallel data. The CDR function is necessary because the receiver has to recover the clock embedded in the incoming data stream. The clock extracted from the data by the CDR unit can then be fed into the deserializer allowing for the data to be recovered in its original parallel form.

In addition to the transmit and the receive section, the SerDes includes a common block, which provides the support functions for the transmitter and the receiver blocks. The common block generates the bias voltages and currents, the clock and operating mode control.

3.3.3.2 Output Drive

Once the initial digital patterns have been properly encoded, synchronized, NRZ-modulated and this way prepared for transmission, the resulting analog signal is ready to be sent through the physical medium over the transmission line. The signal is transmitted in differential form with an output swing of, typically a few hundred millivolts (Figure 3.12).

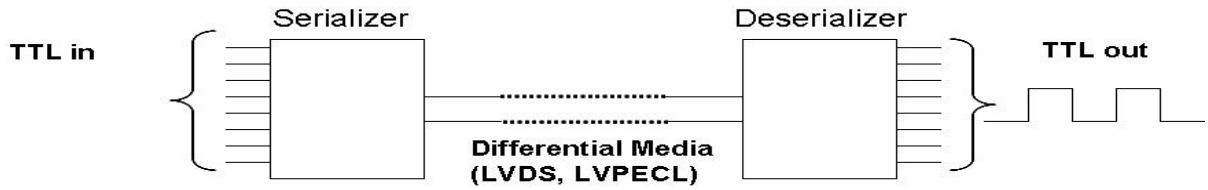


Figure 3.12: Flow diagram for a serializer/deserializer pair

Among the most common technologies to transmit signals across differential media at high-speed are Low-Voltage Differential Signaling (LVDS), Emitter-Coupled Logic (ECL) and the Current Mode Logic (CML) [BOL-02].

LVDS is a high-speed and low-power differential interface for generic applications, which was standardized as an electrical layer standard by the Telecommunications Industry Association (TIA) and is published as ANSI/TIA/EIA-644-A. An LVDS driver provides a typical 350mV differential output voltage centered at about +1.25V. A typical LVDS receiver is specified with a 100mV threshold over the receiver's input range of ground to +2.4V. This allows for the nominal active signal to shift up or down 1V in common-mode due to ground potential differences or coupled noise. The LVDS driver is intended to be used with 100Ω interconnects. LVDS strength consists in delivering high-speed operation while consuming little power. Power is minimized by the LVDS technology by limiting the load current to 3.5mA and by minimizing the static current by the use of sub-micron CMOS processes. Data rates supported by LVDS tend to be in the DC to 2Gbps range.

ECL or LVPECL is used place of LVDS in some high-performance SerDes devices when higher drive and higher speeds are required. ECL technology has been standardized in TIA's ANSI/TIA/EIA-612 and TIA's ANSI/TIA/EIA-613. The ECL drivers are low impedance open emitter outputs that generate a typical 700 to 800mV output voltage. ECL parts are usually powered between ground and -5.2V. For compatibility with other popular devices (logic, ASICs, processors etc.), it is possible to shift the operating point so that the circuit is driven between +5V and ground. This positive rail operation of ECL is called PECL. There also exists a low (3.3V) version of the technology known as LVPECL - Low Voltage PECL The output stage for an ECL driver is operated in the active region, saturation is prevented, thus the edge rates obtained for the ECL are fast. Typically, ECL operates from DC to more than 10Gbps.

CML is a high-speed point-to-point interface. Similar to LVDS, differential CML also has a relatively low signal swing. The single-ended swing for most differential CML drivers is between 500mV and 800mV, but can be set to almost any level by simply adjusting the current source. Typical receivers have a common-mode range of approximately 1.0V. A unique feature of CML is that it typically does not require any external resistors as termination is provided internally by both the driver and the receiver devices. CML supports data rates above 10Gbps depending upon the process for the drivers and receivers. A Differential CML output buffer can transmit data over relatively long distances. For example it is possible to transmit 1Gbps Ethernet over 25m twinax at 1.25 GBd and achieve 0.5m transmission of 10Gbps Ethernet in an FR4 trace when using CML drivers.

3.3.3.3 Signal Conditioning by Using Pre-Emphasis

To compensate for the undesired signal integrity effects which appear at high-speeds and are described in the paragraph 3.2, signal conditioning is the enabling technology used to achieve reliable high-speed transmission in the copper interconnects. There are two common forms of signal conditioning techniques which are added to the SerDes design: transmit pre-emphasis (hereby being considered as well the transmit de-emphasis) and receiver equalization. These methods require extra circuitry in the transmitter and receiver to compensate for high frequency losses, which appear during signal propagation through copper PCB medium.

The transmit pre-emphasis and receive post-equalization essentially accomplish a similar function, but they are implemented differently. With receive equalization, the signal is transmitted without any special processing and is 'equalized' at the receiver using filtering techniques. Pre-emphasis on the other hand, is a process performed in the transmitter and offers many advantages and better results in backplane applications. Therefore the topic of this thesis will make use of the pre-emphasis techniques to overcome challenges imposed by the high-speed design.

Pre-emphasis operates on the observation that losses and distortion in a transmitted signal are experienced more severely by the transitions from one state to another, which contains the most of the high-frequency components of a signal. These components are needed to provide the fast rise time required to reach the intended signal value within the bit period and they are removed from the signal due to the low pass nature of the transmission channel. Pre-emphasis logic operates inside the chip by increasing the amplitude of the first symbol to follow a state transition, while subsequent symbols in the sequence are transmitted at the nominal amplitude. Figure 3.13 shows a bit stream transmitted with (in red) and without pre-emphasis (in black).

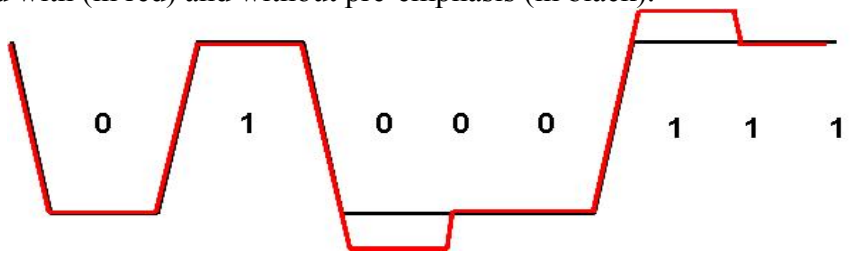


Figure 3.13: Waveform definition of pre-emphasis

Boosting the level of bits following a transition results in an amplification of the power at the higher frequencies, which compensates for the higher frequency losses exercised during the signal propagation through the interconnect. The extra signal energy is removed by those losses, so that the receiver receives a close approximation to the intended transmitted signal.

A pre-emphasis driver uses more power than a driver without pre-emphasis; large signal swings are difficult to implement and therefore in practice a maximum amount of amplification (boost) that can be achieved is limited ([RAD-99], [MAX-02]). All these problems associated with a pre-emphasis driver can be overcome by using de-emphasis. With the de-emphasis the peak-to-peak voltage is the same as with a regular driver, but

the level of bits not following a transition is attenuated. Therefore from a practical point of view, the de-emphasis effect is preferred for being implemented in the driver due to its lower power necessities. Figure 3.14 shows a short bit stream transmitted by a regular driver (in black) and by a de-emphasis driver (in red).

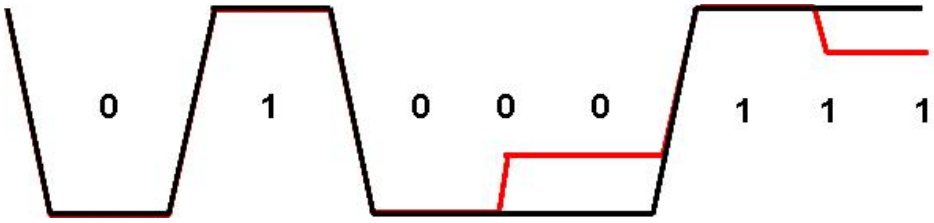


Figure 3.14: Waveform definition of de-emphasis

The voltage swings in the case of de-emphasis are the same as with a regular driver, but power consumption is less than with a normal driver since the power at the lower frequencies is attenuated. Even though the amplitude of the de-emphasis driver is lower that in the case of a regular driver, the received signal at the receiver when de-emphasis is considered has better quality than that with a regular driver. That is because the de-emphasis attenuates lower frequencies in the transmitted signal in a similar/equal amount as the high-frequency components are reduced from the signal during the propagation through the copper medium. So, the signal at the reception has not only less higher-frequency components (as the case of a normal driver) but less high and low frequency components, which gives a better approximation of the original transmitted signal. Which ever technique we use, pre-emphasis or de-emphasis, the final goal is the same and implies the improvement of the quality of the transmission across PCB interconnects.

3.3.3.4 Pre-Emphasis Implementation in the Transmitter

Most vendors of SerDes components use the transmit pre-emphasis or de-emphasis method to improve the channel performance by compensating for erroneous signal at the receiver. Transmit pre-emphasis is dynamically programmable inside the SerDes device to different levels so that optimum signal transmission can be achieved through appropriate programming. Programming the pre-emphasis inside the transmitter requires the specification of the signal amplitude and the percentage of amplification desired to be achieved. Figure 3.15 is used for understanding how the pre-emphasis measurement is made because that determines the corresponding setup in the driver chip.

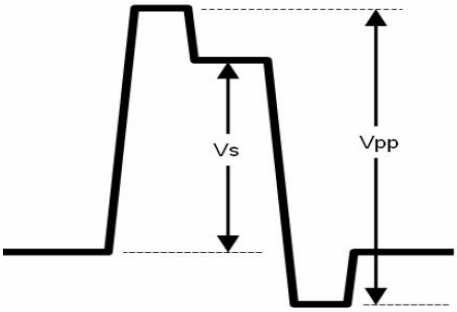


Figure 3.15: Waveform definition of pre-emphasis

The pre-emphasis level ([IEEE-02], [TAI-x]) is defined in equation (3.1):

$$\Pr e - Emhasis[\%] = \frac{V_{pp}}{V_s} - 1 \qquad (3.1)$$

Pre-emphasis can be effectively implemented by a finite impulse response (FIR) filter ([FAR-99], [LI-x]). The FIR filter reduces inter-symbol interference by inverting the transmission channel low-pass characteristic. The structure of a generic FIR filter is presented in Figure 3.16.

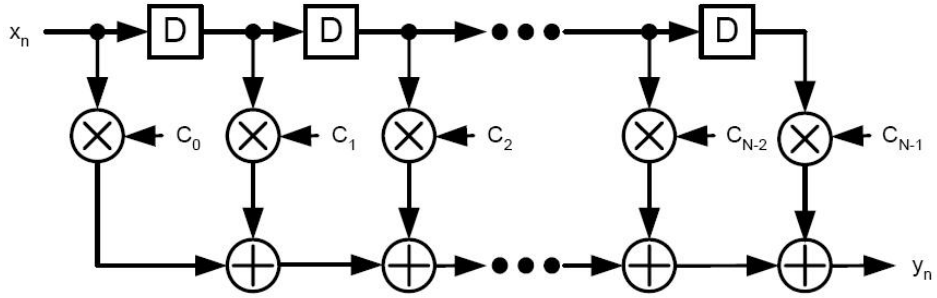


Figure 3.16: FIR filter structure

The transfer function implemented in z-domain by a FIR filter with n-taps is

$$H(z) = \sum_{n=0}^{N-1} c_n \cdot z^{-n} \qquad (3.2)$$

where $z = e^{-j\omega T}$ and T is the inverse of the sampling frequency of the signal. The filter coefficients are represented by c_n.

A practical implementation of a two-tap digital FIR filter for the pre-emphasis is presented in [GEE-03]. The coefficients of this filter are chosen $c_0 = 1 + PE$ and $c_1 = -PE$, where PE is the boost of pre-emphasis which is defined as $PE = \frac{Vpp - Vs}{2}$. So, for a two-tap FIR filter with these coefficients, the frequency response is $H(z) = (1 + PE) - PE \cdot e^{-j \cdot \omega \cdot T}$.

The implementation of pre-emphasis in the FIR filter works by applying a delayed and inverted pulse to be added to the received signal (pulse) at the other end of the channel, at the next cycle corresponding to the first bit within the subsequent bits following a transition ([ZHA-02]). The delayed added waveform (blue) and the received (compensated) pulse for the case with pre-emphasis (green) are shown in Figure 3.17.

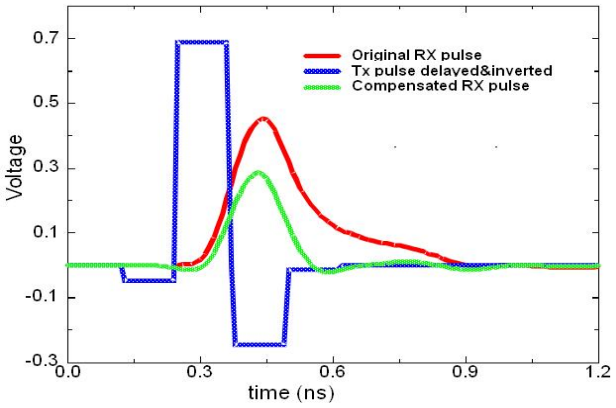


Figure 3.17: Using pre-emphasis inside the driver

The trace representing the original pulse (represented in red) shows what the received signal would have looked like without the benefit of pre-emphasis. This amount to which the long tail of the original received pulse is cancelled depends on the pre-emphasis boost. So, the higher the pre-emphasis given by the driver and implemented in the filter coefficients, the better performances (better shape) will be observed on the signal at the receiver. A simplified driver with pre-emphasis implemented is shown in Figure 3.18.

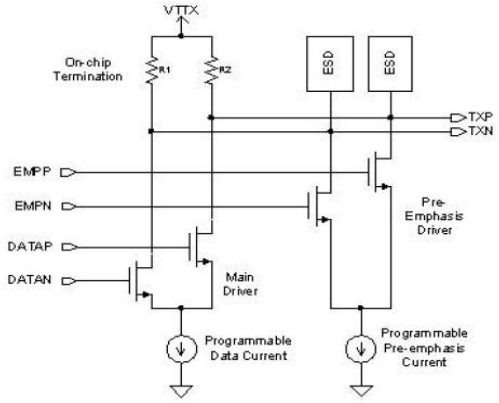


Figure 3.18: Simplified driver with pre-emphasis

The Driver with pre-emphasis is implemented with parallel output buffers. One of them is driven with the normal signal output and the other is driven with the calculated output from the FIR filter.

The optimal amount of pre-emphasis depends on the losses present in any given transmission line and has to be empirically determined [TAI-04].

3.3.3.5 Ethernet SerDes Functionality for Testing

Testing the quality of the signal transmission at 10Gbps is very important when a reliable backplane design is required. In addition to the transmitter and receive section, a SerDes chip will usually have addition control logic and testability blocks to enhance test and characterization. Using testability facilities implemented inside the SerDes, the designer has the possibility to test the SerDes design and the quality of the transmission channel on the PCB with no need of other external active components.

The ultimate figure of merit for the transmission is the amount of measured Bit Error Rate (BER) at the SerDes interface. BER is an indication of the quality of the transmission directly influenced by the SerDes device and by the PCB transmission channel environment. BER is the number of bits errors transmitted/received divided by the total number of bits transmitted/received. The 10 Gigabit standards specify that the XAUI receiver shall operate with a BER of less than 10E-12, which translates in not allowing more than one error bit per a total of 10E12 bits transmitted.

To be able to performing this kind of BER tests in the situation when the design under test exceeds the performance of the measurement equipment, the designer has to make use of the SerDes functionality for testing purposes. A way of adding testability functionality is to include a pattern generator in the SerDes transmitter design, generator that can internally supply the parallel data to the transmitter in terms of a variety of patterns, as defined by the 10Gbps Ethernet standard (see paragraph 2.3.5.). The data patterns, internally generated by the SerDes are sent down the transmission channel and the quality of the transmission is estimated from the measurements performed at the receiving point.

A plot of a single instance of a sequential bit pattern is not sufficient to characterize signal quality. The common method used for measuring the signal quality degradation along the transmission channel is the eye-diagram. An eye-diagram results from the waveforms superposition of all bits contained within a given bit pattern. Serial protocols have standards that are defined on the use of eye diagram templates, which specifies the limits of the eye-opening. The eye-opening is defined as the boundary within which no waveform trajectories lie. A large eye-opening means that the system has a greater tolerance for amplitude and timing variations, and therefore potentially a lower BER. In a bandwidth limited communication channel, the pulses of consecutive bits will overlap, thereby effectively reducing the eye-opening.

The net effect due to a poor transmission quality results in the eye-closure of the signal measured at the receiver end. A smaller eye-opening would make signal reception and recovery difficult for the receiver and would result in high BER. A high BER for instance would causes packet losses and reduced effective bandwidth in the case of the 10Gbps Ethernet switch. The second part of the present thesis will provide a design intended to test the quality of transmission across the 10Gbps backplane, making use by the SerDes functionality for test.

3.3.4 High-Speed Connectors

Connectors represent a substantial part of a backplane system and they have to be extremely carefully selected since the signal quality in a backplane application is highly dependent on the performances of the high-speed connectors used. For the study of the impact of the connectors and the available solutions for our 10Gbps backplane application we dedicated an entire paragraph in the chapter 4 of this thesis.

3.4 Summary

This chapter presents the most important aspects and challenges related to transmission of high speed signals across copper interconnects. It also describes some of the techniques employed to achieve the data rates required to implement a 10 Gigabit Ethernet switch.

4. A Choice of Technologies for the High-Speed Backplane System

4.1 Connectors

4.1.1 Connector Technology

For any high speed board-to-board interconnection over a backplane the connectors have to be carefully selected and extremely well designed, so that they have as little an effect as possible on the overall performance of the interconnection system.

At the time when the 10 Gigabit Ethernet standardization committee specified the XAUI protocol for 10Gbps data transmission over copper, this was at the limit of what was possible with the pre-existing connector technology. Several manufacturers were preparing to provide a new and better generation of connectors which could address the challenges imposed by the increased data rates. Compared with the previous connectors, this new generation had to be able to support higher data rates and to provide better signal isolation and impedance matching for maintaining signal integrity through the whole transmission path. In addition to these electrical prerequisites, the connectors had to address the mechanical issues by providing a robust mechanical structure and reliable attachment methodology at the PCB interface [CLI-03]. This attempt to provide improved performance at higher bandwidth has led to two general connector solutions.

To increase bandwidth and meet the high-speed design challenge the manufacturer's first solution was to increase the pin count in the connector [ROO-02], by either adding rows or columns of contacts. In order to maintain the required levels of signal integrity, additional pins in the connector are assigned to ground, and special attention is given to designing controlled impedance transmission lines which should meet the 100Ω +/- 10Ω requirements. The second solution focuses on the minimization of the crosstalk effect. To deal with this, the manufacturers had to shift to differential signal technology and to provide better solutions for shielding and grounding to substantially reduce the crosstalk effects inside and outside the connector body. Each manufacturer was tackling the shielding issue differently ([KOL-03], [KOL-04]), the solutions here including extra ground pins, extra shielding or offset of the contact rows.

Whichever were the solutions adopted, a new generation of connectors with better performances for high-speeds signal transmission has been provided on the market. Various connector design schemes have been proposed and several manufacturers brought out new products on the market to meet the emerging 10 Gigabit standard. An article in Electronics Weekly - Electronic News from 4/19/2002 provided a useful overview of the available choices we had:

"Teradyne is proposing its GbX platform connector as a speed upgrade to its existing 2mm connectors. With less than 2 percent crosstalk, the GbX connector is available as a 4-pair platform, which offers 55 differential pairs per linear inch, while the 5-pair configuration offers 69 differential pairs per linear inch.

Tyco Electronics Corp. is working with at least one chip supplier, Gennum Corp., to demonstrate that its Z-PACK HM-Zd connector system using FR-4 board backplanes will run serial data at 10Gbit/sec in a 100ohm differential signaling environment. The Z-PACK HM-Zd product, first introduced in 2001, extends Tyco's existing IEC 61076-4-101 hard metric connector family with differential signaling that supports 3.5Gbit/sec data rates.

FCI Electronics is also offering a 100ohm differential version of its 2mm HM family, Metral 4000. Its design uses a stripline structure in both the receptacle and header and modified contact geometry better impedance match and lower insertion loss. The aim says FCI is to allow designers to mix different data rates on the same daughtercard connection.

Harting Group has a multi-source agreement with 3M for its high-speed hard metric (HSHM) family of connectors, while Molex and Teradyne offer compatible high-speed differential versions of the VHDM 2mm connector system, called VHDM-HSD, designed for data rates up to 5Gbit/sec in commonly used daughtercard-to-backplane applications."

Given the wide variety of high-speed connectors on the market, we had to decide which is the one that would allow the best interconnect solution for the 10Gbps Ethernet backplane application. The selection of the most appropriate connector is highly important for multi-Gigabit transmission and it has been addressed by many authors ([SHO-02], [NAD-x]). In this selection, the most important criteria to consider are electrical performance, mechanical robustness and cost.

Figure 4.1 below shows the footprints of a backplane connector (left) and its matching daughter card connector (right) with the main parameters of interest which appears in Table 4.1. The table lists the parameters for each of the competing products described in the article above.

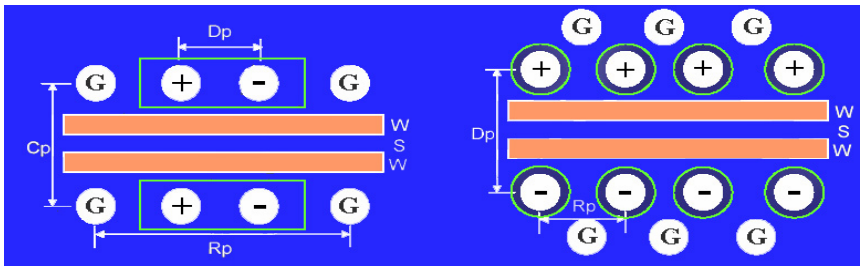


Figure 4.1: Critical connector footprint dimensions for motherboard and daughtercards

Manufacturers	Teradyne & Molex	Teradyne	Tyco & Erni	FCI & ITT-Cannon	3M & Harting
Connector type	**HSD**	**GbX**	**HM-Zd & ERmet ZD**	**Metral 4000**	**MetPak HSHM**
Intra-pair spacing on backplane (BP), Dp [mm]	2.25	1.85	1.5	2.0 mm	2.0
Intra-pair spacing on linecard (LC), Dp [mm]	2.0	1.5	2.5	2.0 mm	2.0
Pair pitch in column direction, Cp [mm]	2.0	1.85	2.5	2.0 mm	4.0
Recommended through hole drill size [mm]	0.66	0.57	0.7	0.6 mm	0.6
Recommended through hole pad size [mm]	1.0	0.9	1.02	0.9	0.86
BP routing channel [mm]	1.34	1.28	1.8	1.4	1.4
LC routing channel [mm]	1.34	0.93	1.8	1.4	1.4
Pairs per inch for 4 pair columns	38	55	40	38	38
Worst case differential crosstalk [%]	3.2	<5	3.1	<5	6

Table 4.1: Comparative figures for connector footprints

The performance of these connectors, according to their published specifications, show crosstalk figures in the region of -25dB to -30dB (3% to 6%). There are however two connectors manufacturers, 3M and FCI, which have apparently retrofitted extra shielding onto their connectors and that shows up in their comparatively worse crosstalk figures. However, the overall performance of the new generation of connectors [MOE-03] is indeed very good, showing considerable improvements in the supported speed compared to the previous generation.

4.1.2 Connector Selection

Assuming that all the vendors' claims of achievable bandwidth and crosstalk were credible then making a choice was divided between technical and commercial. From the technical point of view we weighed the comparative advantages of the connectors electrical specifications, the pin density versus the difficulties of routing [SHA-03], and the possibility of mixing different connector types (for power and local control) on the same backplane. Commercial considerations included cost and multiple source

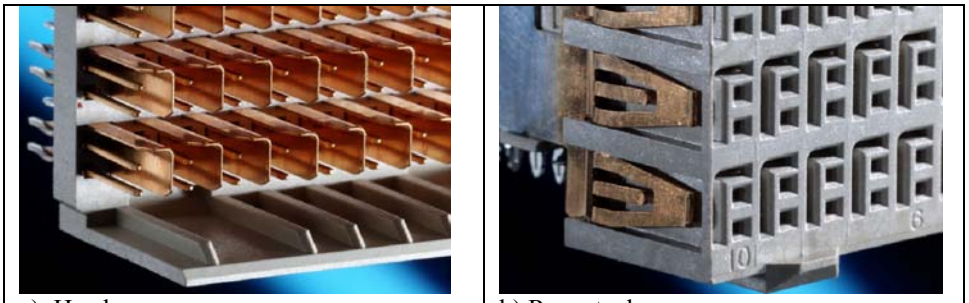
availability, as the connectors themselves could represent a substantial cost of a backplane system.

After studying the electrical and mechanical specifications of the connectors presented in Table 4.1, we choose as a first option Tyco HM-Zd / Erni ERmet ZD connector because it offers an acceptable number of pins (40 pins/inch) and provides good crosstalk immunity. As a second choice, if we would be forced into a pin count problem because of the architectural design, then the Teradyne GbX would be a good candidate, although the narrowness of the routing channel on the backplane is cause for concern.

Both the HM-Zd (or ERmet ZD) and GbX connectors are carefully thought out and well engineered parts, which show good crosstalk figures and answer the electrical and mechanical requirements of the 10Gbps Ethernet backplane.

4.1.2.1 Tyco HM-Zd or ERNI ErmetZD

Tyco and Erni manufacturers united their efforts in developing and bringing on the market a differential high-speed connector addressed as Z-PACK HM-Zd or ERmet ZD. Tyco HM-ZD connector is a member of Z-PACK 2mm HM family [AMP-01], while ERmet ZD connector is a certified second source for the Z-PACK HM-Zd from Tyco [HM-ZD]. Figure 4.2 shows the footprint and the main characteristics of this connector.



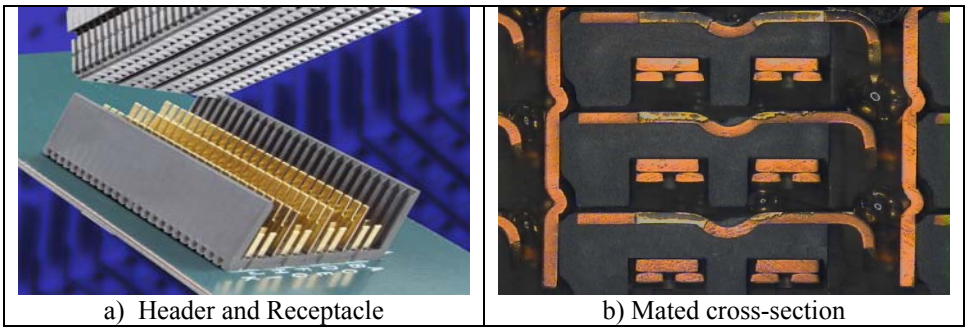
a) Header b) Receptacle

Figure 4.2: HM-Zd (or ERmet ZD) connector

To reduce crosstalk between adjacent differential pairs, HM-Zd connectors have a fully shielded structure that becomes an electrical extension of the PCB environment. To achieve that, the header assigned to the backplane (Figure 4.2 a.) has a wrap round L-shaped section whose matching other half is in the receptacle in the daughter-card (Figure 4.2 b.). When mated to the header, the receptacle has a spring leaf that makes contact with the headers screen to ensure ground continuity through the connector. This results in an optimized crosstalk behaviour. In addition, the spacing of conductors and the integrated ground shields are very closely controlled to maintain the target impedance to a typical value of 50Ω signal to ground in single-ended, and 100 Ω signal-to-signal in differential applications.

4.1.2.2 Teradyne GBX

The GbX connector from Teradyne has been studied in respect to the specified parameters in its datasheet and considered as a possible candidate for the interconnection between the backplane and the daughter-cards of the 10 Gbps Ethernet switch. Figure 4.3 shows the footprint and the main features of the GBX connector.



a) Header and Receptacle b) Mated cross-section

Figure 4.3: GbX connector

The crosstalk is reduced in the GbX connector by surrounding each signal pair with a ground shield. These shields are presented both in the daughter-card section (the header) and in the mating interface on the backplane (the receptacle), providing continuous shields across the connector. The ground shields in the header can be seen in the left hand picture (Figure 4.3 a.) and shows up as well as the vertical strips in the right hand picture (Figure 4.3 b.).

Every signal pair in the connector is thus completely shielded from the other to provide good crosstalk immunity. Figure 4.3 b.) shows this where the vertical shielding plates are those on the male connector and the horizontal shielding plates are built into the female connector. The pin pair thus propagates through an impedance controlled waveguide fully shielded from its neighbours.

The GbX connectors were later used by Xilinx to promote high speed data rates across copper backplanes within the ‘Serial Tsunami’ initiative [SHA-03].

4.1.3 Connector Models

Each manufacturer provides, either publicly, or under Non-Disclosure-Agreement (NDA), SPICE models of their connectors’ behaviour.

In order to obtain a good characterization of their connectors over a wide frequency bandwidth and hence a good model, the manufacturers have to address a difficult problem [SIV-97]. They have to use three-dimensional (3D) electromagnetic field solvers which translate the physical characteristics of the connector into equivalent electrical models. Traditionally, the electrical parameters for a connector are extracted in the form of coupled multiple Resistance (R), Inductance (L), Capacitance (C) and Conductance (G) sections (RLCG), or in form of Scattering-parameters [POZ-98], depending if a quasi-static or a full-wave analysis is employed [OLT-R1]. At the time we were looking

into that issue, the most common connector representation was the SPICE compatible model including several cascaded section of R, L, C and G elements.

Tyco and ERNI provided us with RLCG SPICE compatible models for their HM-Zd and ERmet ZD connectors. They offered two levels of models for their connectors: the single pair model and the multi-pair model, both of which would be accurate for signal rise time of 60ps or slower. The very compact single-pair model has distributed R, L and C elements for the pair's, along with characteristic impedance and propagation delay for the pair's equivalent circuit and is intended for analyzing the reflection behaviour and impedance matching of the connector. The multi-pairs model represents multiple connectors' pairs (two, three or four signal pairs) and includes inter-contact, contact-to-shield and inter-pad coupling and can fully represent the connector behaviour by including impedance matching, reflections and crosstalk effects.

Similarly, Teradyne provided us, under an NDA, the multi-pairs SPICE compatible model for their GbX connector. The model contains a number of 29 cascaded RLCG sections and Teradyne claimed it can accurately support and simulate signal rise times of up to 60ps.

All connector manufacturers warned about the fact that the models they provide are only for the connector itself, in isolation from any backplane or daughter-board they are connected to. This is understandable as the vendors have no control over how their connectors are used. It is important however to understand the implications arising from the insertion of the connector pins in the PCB structure and this will be studied with the aid of full-wave field solvers by ourselves in chapter 5.

4.2 PCB Traces

4.2.1 Differential Trace Configurations

The preferred method for high-speed transmission is differential signaling (see paragraph 3.3.2). A differential pair in a PCB is formed by two traces that carry complementary voltages, have the same characteristic impedances to the reference planes and have equal propagation delays. The most popular configurations which satisfy these requirements are the differential microstrip, the edge-coupled stripline and the broadside-coupled stripline shown in Figure 4.4.

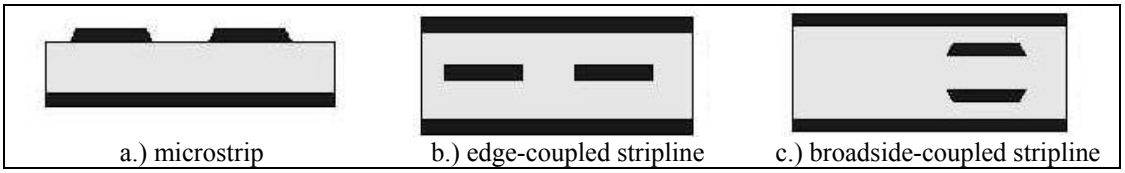


Figure 4.4: PCB trace configurations

In comparison with the microstrip configuration, situated at the top or bottom of the PCB, the striplines situated between internal layers of the board, are a non-dispersive transmission media providing an excellent isolation between adjacent traces. Therefore among the configurations presented in Figure 4.4, differential striplines have been

considered more suitable for backplane applications where both the connector assembly and the compactness require the routing of traces in internal layers. In addition, edge-coupled striplines were preferred for the 10Gbps backplane application rather than broadside-coupled because of less sensitivity to impedance variations caused by manufacturing process deviations ([BEY-02], [LES-02]).

4.2.2 Controlled-Impedance PCB Traces

The accepted standard in 10 Gigabit Ethernet for the controlled-impedance PCB traces is a characteristic impedance (Z_0) of 50Ω and a differential impedance (Z_{diff}) of 100Ω. When designing such controlled-impedance traces used for the transmission of a multi-Gigabit signal on a PCB, there is a series of issues that should be considered, including the trace configuration, the dielectric material of the board, the influence of the trace geometry and the manufacturing tolerances.

4.2.2.1 Dielectric Material Influence

From the beginning we made the assumption that we do not need to resort to exotic and expensive dielectrics, such as low loss materials, in order to achieve 50Ω single-ended and 100Ω differential impedance of the traces on the PCB. Many studies and practical built systems have demonstrated that the classical FR4 dielectric can be still used to transmit signals with achievements of reliable reception up to 5Gbps ([DER-x], [SHO-02]). This limit has been upwardly revised over the past years essentially as a result of the more advanced techniques of pre-emphasis and equalization (see paragraph 3.3.3.3) used in modern backplane transceivers, as well as improved design, simulation and measurements of the system [SAY-x].

We therefore decided to use FR4 dielectric material in building the backplane of the 10Gbps Ethernet switch. This material has a relatively low-cost while still offering good mechanical and thermal properties. However, the dielectric constant (εr) of FR4 is anything but constant. For standard FR4, the dielectric constant varies as function of temperature, frequency and resin content. Typical values can range from 4.0 to 4.9 **(**[AME-97], [COO-79]) and tests on one particular sample are shown in Figure 4.5.

The dielectric constant directly influences the impedance value of the PCB trace which passes over a plane separated by that dielectric material. Equations (4.1) and (4.2) show how the impedance value of a microstrip, respectively of a stripline is calculated, function of the dielectric thickness (h), track width (w), track thickness (t), respectively separation between grounds (b) [Johnson-93].

$$Z_0 = \frac{87}{\sqrt{\varepsilon r + 1.41}} \cdot \ln\left(\frac{5.98 \cdot h}{0.8 \cdot w + t}\right) \qquad (4.1)$$

$$Z_0 = \frac{60}{\sqrt{\varepsilon r}} \cdot \ln\left(\frac{1.9 \cdot b}{0.8 \cdot w + t}\right) \qquad (4.2)$$

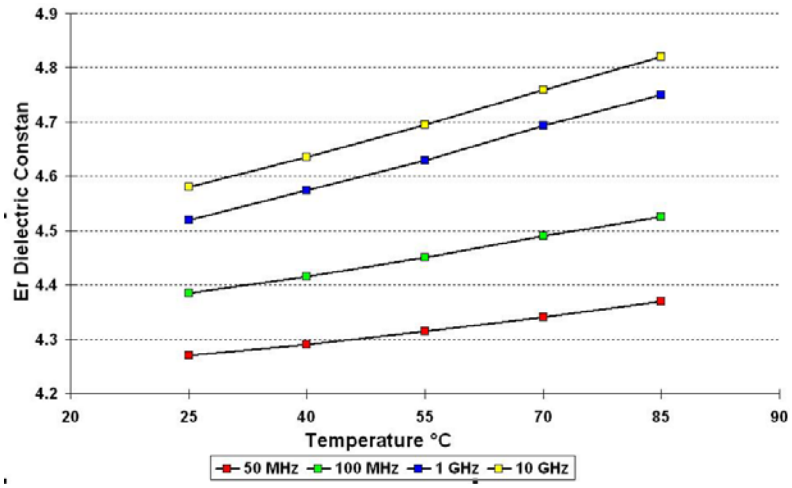


Figure 4.5: Variations of Permittivity with Temperature and Frequency

The graph in Figure 4.6 shows the effect of the range of dielectric constant's values from Figure 4.5 on the differential impedance of a stripline built in FR4.

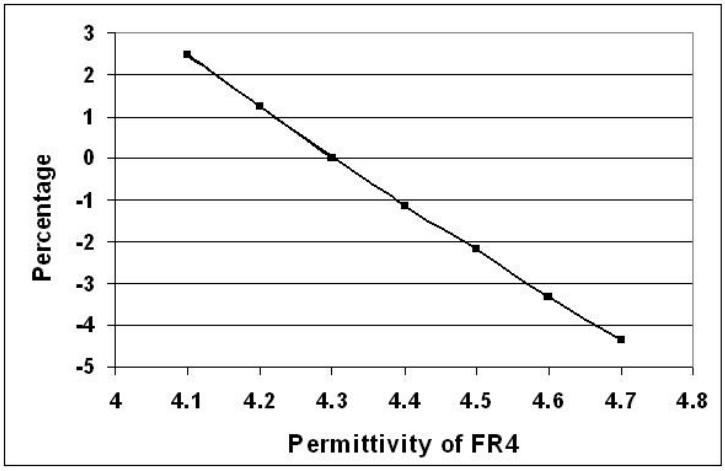


Figure 4.6: Variation of Z_{diff} with Permittivity

The best way to determine an indicative value of the impedance of a PCB transmission line is the use of a two-dimensional (2D) field solver or a 2D stack-up calculator, which can predict the trace impedance for a given trace configuration when the trace width, thickness, separation, dielectric constant and height are specified. A stack-up calculator bases its computation on the configuration and the cross-section geometry we supply for the transmission line. The problem is that a simple 2D stack-up calculator does not take into account factors which affect the traces impedance, such as variations of the dielectric stack-up due to the frequency of operation and temperature. As a result, the final manufactured impedance could be anywhere within +2.5% to -4.5% of nominal expected value, as shown in Figure 4.6. The observation from this is that for obtaining the desired target impedance, the final PCB should be built in conjunction with the board manufacturer, as discussed later in section 4.2.2.3.

4.2.2.2 Trace Geometry Influence

Other important factors in achieving a controlled impedance design are the parameters defining the geometry of the trace such as the trace width, trace separation and the thickness of the board dielectric.

By choosing the corresponding parameters for the trace geometry, 50Ω characteristic impedance can easily be achieved. Running the two traces, each one 50Ω, as far apart as possible one from the other give a theoretical 100Ω differential impedance. The main problem with that however, is the difficulty of maintaining equal length traces and managing the changes in impedance that occurs as the traces are separated away after the transmitter and then brought closer together again at the receiving device. Therefore, it is preferable to define a fixed minimum spacing between traces and maintain it from source to destination. In addition, another advantage of maintaining traces close together is the minimization of their electromagnetic interference effect since their radiated fields cancel out.

By bringing 50Ω characteristic impedance traces closer together, the coupling coefficient increases (as the differential coupling is not just between the traces and their reference plane but also between adjacent traces), and the differential impedance is lowered. 100Ω differential impedance is restored by altering the PCB geometry and increasing the mismatch in the single-ended impedance component. For example, 100Ω differential impedance is given by 60Ω single-ended impedance, by narrowing the trace widths and maintaining the same trace-to-trace separation. Figure 4.7 shows the range of values chosen for the trace width and separation, to give the available options for achieving 100Ω differential impedance.

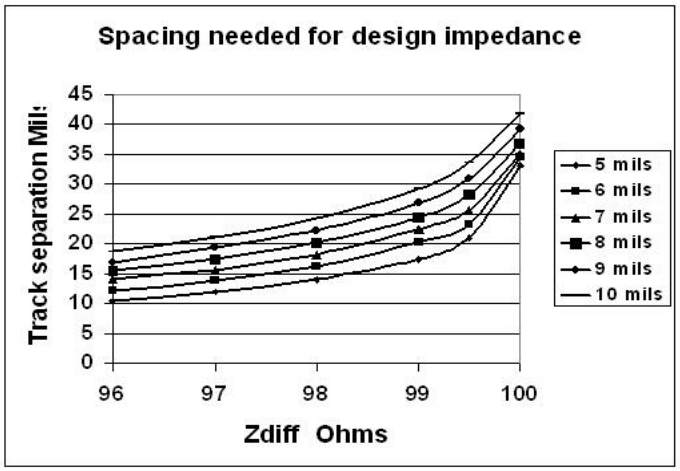


Figure 4.7: Variation of Zdiff with trace width and trace separation

For the same differential impedance, the spacing between narrower traces is less than for wider traces. In addition, the final thickness of the PCB is much less when narrow traces are used than for wide traces. On the other hand, wider traces are considered to be better for signal integrity due to lower skin-effect losses, but they result in a thicker board with more layers [WAL-x].

At this point, the choices for the trace widths and their separation comes down to a question of personal preference. Commonly, PCB designers with great faith in their analysis tools will favour the narrower traces to achieve the design target of 100Ω, while those with production cost experience will favour the wider traces that are cheaper and can more reliably be reproduced.

4.2.2.3 Manufacturing Tolerance Influence

One of the key PCB technologies employed to optimize performance for multi-gigabit transmission is a tight control within the board manufacturability process [MIT-04]. Typically, the PCB manufacturer adjusts the parameters supplied for the controlled-impedance design to account for etching, non-regular shape of the finished trace cross-section or to accommodate the standard dielectric thickness [KOL-03]. We studied some of these concerns and illustrated their order of magnitude effect in PCB manufacturing.

Etching traces on copper is a well controlled lithographic process. Nonetheless it is not possible to achieve micron accuracy and there is a trade-off between cost and tolerance both in trace width and plane-to-trace spacing. Just taking one manufacturer reference at random [REF_pnc], they offer tolerances of 0.5mil per side of an etch in ½ ounce copper which results in a trace width tolerance of 1mil. Their overall board thickness is within 10% for the 'regular' price or 5% for a premium. These figures are typical for the industry and we studied what they mean and how they influence our design choices. Figure 4.8 shows the effect of varying trace width on the impedance.

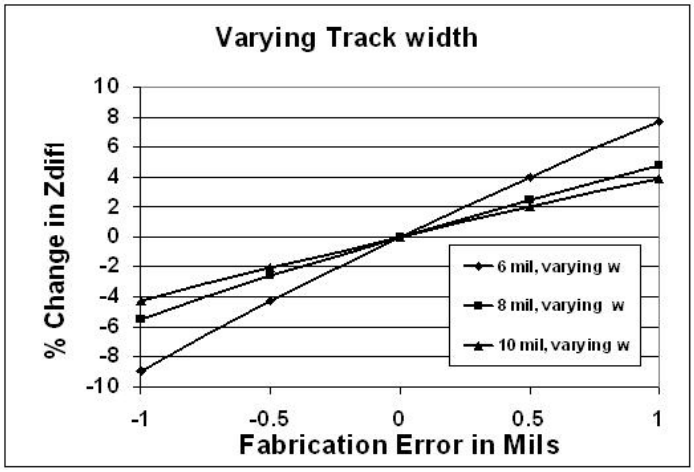


Figure 4.8: Effect of trace width tolerance on Zdiff

The etch tolerance is an absolute dimension, not a percentage, so its effect is more pronounced for smaller geometries. So, while there appears to be little differentiation between a width of 8 and 10mil (as illustrated in Figure 4.8), for a 6mil trace width there is a much greater risk. In support of this, the results of an independent study ([SHO-02]) showed that up to 5Gbps there is no significant impact or differences on overall system performance if 8mil or 10mil traces are used. For our application at 3.125Gbps, we therefore considered values somewhere from 8 and 10mil for the trace width, but the final decision is taken after an extensive analysis presented in the paragraph 5.1.2.

Subsequently, we addressed the issue of board thickness tolerance. FR4 dielectric materials have better tolerance than the prepreg which is used to laminate them together. Figure 4.9 shows the effect of dielectric thickness errors on the impedance.

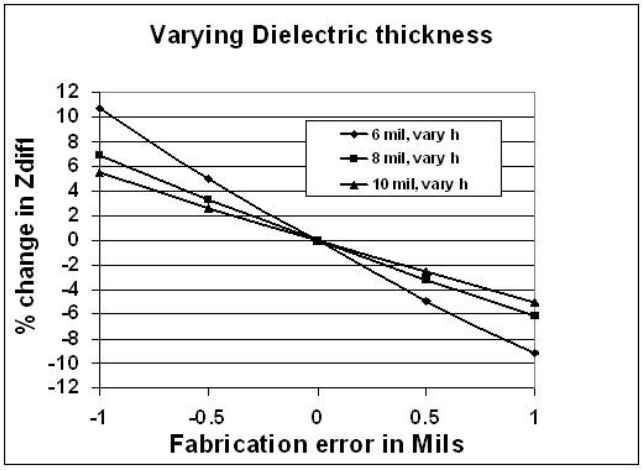


Figure 4.9: Effect of dielectric thickness tolerance on Zdiff

We considered a tolerance of 10%, which easily translates into errors of the order of magnitude of 1mil.

The overall magnitude of the manufacturing errors is alarmingly high but in practice the two effects (trace width and dielectric thickness) are cancelled out in the production stage. The PCB manufacturer plays a very important role in achieving the target controlled impedance for a given design, as to him are given the parameters required to achieve the target controlled impedance. He has to adapt these dimensions and first build a test coupon for the board on which the finished results can be measured. Once this first trial build has been done the manufacturer can iterate between etch times, lamination process times and pressure to achieve the final target design while at the same time keeping the manufacturing tolerances within a reasonable limit.

Reaching an optimum dimension for impedance controlled traces is not a trivial issue since it involves making continual trade-offs between what is ‘desirable’ and what is ‘necessary’. But at the end, no matter how far we push the theoretical analysis for obtaining the optimum dimensions for the trace design with the PCB stack-up, the final design must be built in conjunction with the PCB manufacturers who can adapt the nominal dimensions we supply to achieve the desired results within an acceptable degree of manufacturing tolerance.

4.2.3 PCB Trace Models

Meeting the challenges imposed by the high-speed backplane design (as presented in paragraph 3.2) requires the ability to simulate signal integrity issues and accurately predict potential problems of a PCB design before manufacturing. An accurate simulation however requires in addition of accurate models of the system’s components (e.g. I/O

devices, connectors), accurate models for the interconnect elements such as the PCB traces ([KOL-x], [LIP-02]).

The frequency dependent losses in FR4 and the crosstalk effect are the most important challenges in transmitting signals at multi-gigabit rates. To account for these effects, the PCB traces should be modelled as coupled- and lossy-transmission lines. There exist several types of representations used for modeling PCB traces, such as parametric models, frequency dependent RLCG tables and behaviour models [SMO-03].

In our design we made the assumption that there are straight traces with uniform cross-sectional geometries passing along on the daughter-cards and on the backplane. These traces are then treated as transmission lines and the best suited for modeling them appears to be the parametric model, which will produce a characterization of the interconnection system in terms of distributed RLCG matrices. An example of such parametric model is the W-element which contains a HSPICE transmission line model that can be applied to simulate from a simple lossless line to complex frequency-dependent loss coupled lines [KUZ-02].

The HSPICE bi-dimensional field solver provides two constant matrices for Capacitance (C) and Inductance (L) [HSPICE-04] and in order to better model the frequency-dependent losses, extremely significant at 3.125Gbps, the W-element model assumes square root dependency of the skin effect and linear dependence of the dielectric loss with the frequency [MAY-03]. For this to happen, two Resistance matrices (R_0 and R_S) and two Conductance matrices (G_0 and G_d) are given with the following rules [ELC-02]:

$$R(f) = R_0 + \sqrt{f} \cdot (1 + j) \cdot R_S \qquad (4.3)$$

$$G(f) = G_0 + \frac{f}{\sqrt{(1 + (\frac{f}{f_{gd}})^2) \cdot}} \cdot G_d \qquad (4.4)$$

where f is the operating frequency and f_{gd} is the cut-off frequency for the G_d.

Even though the W-element parametric model enables the simulation of the frequency dependent losses that appears in the PCB traces based on some assumptions, these methods tend to be accurate only within a certain frequency range. For example, the cut-off frequency for the G_d terms (f_{gd}) is by default of 1GHz in HSPICE, but it can be defined to a higher value to account for higher frequencies.

Finally in order to conclude about the W-element model we underline the major drawbacks and advantages they might have. The main advantages of this model are represented essentially by the accuracy (at least up to 1GHz) and by the reduced simulation time that the HSPICE field solver offers if compared to a 3D parameter extractors [WED-04]. Moreover, we can retrieve a complete RLCG characterization that can be easily imported in more complex HSPICE netlists thanks to the W-element. The limits of this model are represented by the bi-dimensional nature of the HSPICE field solver, which allows dealing only with traces geometries invariant along one axis.

However, as we intend using the W-element for modeling uniform transmission lines, this does not appear as a disadvantage as far as the frequency-dependent losses are taken into account with a reasonable amount of accuracy.

More accurate methods, such as 3D (quasi-static and full-wave) electromagnetic modeling are possible, but their simulation takes much longer and can easily produce unstable, non-causal, non passive results. We concluded that such an approach is needed only for modeling more complicated geometries, such as vias and connector pins, as will be shown in chapter 5. For simple PCB trace modeling as presented in this chapter, we stayed with the W-element model which we extracted from HSPICE by specifying the trace geometry and the dielectric characteristics.

4.3 I/O Performance

4.3.1 I/O Solutions and Our Choice

Transmitting higher data rates over a backplane relies greatly on high-speed transceiver technology. At the beginning of the ESTA project several competing companies were gearing up to provide chips for the anticipated rapid take up of the 10 Gigabit Ethernet market following the adoption of the standard. Several transceiver devices available on the market by 2002-2003 at data rates up to 10Gbps were primarily optical, while a small number of companies were offering 10Gbps for copper transmission.

We did study the market for the 10Gbps electrical transceivers with the intention of tracking competing chip products and to separate publicity information from reality. The collapse of the telco market in 2002 affected several companies to a greater or lesser degree. This left a severely restricted set of choices and ultimately the decision was made, on both technical and commercial grounds, in favour of the Marvell 10Gbps transceiver (Alaska 88X2040).

4.3.2 Marvell Alaska 88X2040

The Marvell Alaska 88X2040 [MAR-03] was specifically designed for 10 Gigabit Ethernet and serial backplane applications and was one of the industry's first 0.15 micron CMOS with advertised low power consumption, superior jitter performance and smaller form factor solution on the market [DAL-01]. The Alaska 88X2040 is a SerDes device which incorporates four transceivers operating up to 3.125Gbps, each with a selectable 8B/10B encoder/decoder. This device performs clock and data recovery for the received path, as well as pre-emphasis, serialisation and clock generation for the transmit path.

The programmable output amplitude and transmit pre-emphasis make it possible to support high-speed copper transmission. Marvell claimed more than adequate signal integrity when using their device for the transmission over copper traces in excess of 60 inches with multiple connectors over a FR4 backplane. They were claiming a BER of

10E-15 [DAL-02], which is several orders of magnitude better than the IEEE requirement of 10E-12 BER for 10 Gigabit links.

4.3.3 Study and Validation of the Marvell I/O Characteristics

4.3.3.1 Marvell I/O Model

Given that the Marvel Alaska came to be recognised as one of the industries benchmark products and the published results of their tests were showing an excellent behaviour, the suitability of using this device for our backplane application was hardly likely to be negative. However, in order to independently validate those claims in the context of our own design, we decided to validate those published results and to predict Marvell's device behaviour for our particular design.

To make this possible we had first to obtain under a NDA a model of the Alaska 88X2040 driver and receiver. This was not the full transistor model but rather a behavioural version that was claimed to offer comparable results, with the caveat that their technology was constantly improving and therefore we should not make any negative decisions based on the current version.

The model was defined in HSPICE and it included two sub-circuits, the transceiver, the receiver, and an I/O bond wire pair for their package. The bond wire was noted in the models documentation to be asymmetric which we considered to be an important feature since packaging has a major effect on signal integrity and each bond path is different. An admitted asymmetry was therefore likely to give a clearer picture of true performance than the sanitised marketing results that companies usually portray.

We went then further on to establish the corresponding simulation environment, and to compare the results of our simulations with the measurements published by Marvell to establish a degree of confidence that the tools we used and methods we employed produced credible results [ESTA-R1].

4.3.3.2 Simulation Environment

The model definition in HSPICE caused some problems since HSPICE import was not part of the suite of Computer Aided Design (CAD) tools setup at CERN.

The existing Cadence Spectraquest [CAD_SQ] signal integrity analysis tool uses a Cadence proprietary SPICE-based language (DML) and a proprietary simulation engine (tlsim) to model complex I/O devices, which include programmable drive buffers, dynamic pullup/pulldown I/O buffers and dynamic clamp diodes. To model the I/O elements, SpecctraQuest accepts behavioural IBIS models embedded in a standard 'wrapper' in form of a larger SPICE macro model. As Marvell's models were HSPICE compatible, the workaround was to install HSPICE as a separate simulation engine in SpecctraQuest and embed the models we received in the IBIS like interface. This entailed a fair amount of manual work since now the graphical interface of the tool was severely

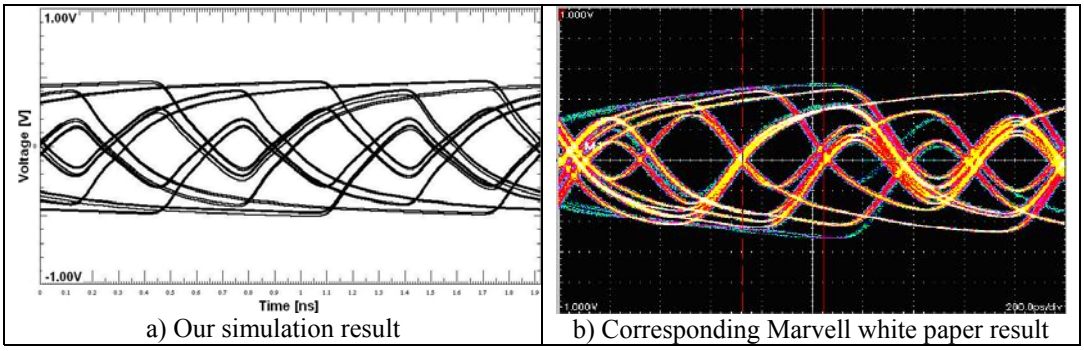
limited and we had to edit, modify and check the netlists to be simulated 'by hand'. Nonetheless, the appropriate simulation environment, as Cadence SpecctraQuest with HSPICE simulation engine, has been successfully setup and we could continue further our evaluation work.

4.3.3.3 Validation

As well as the model, Marvell has produced a white paper promoting the performance of their Alaska SerDes technology across standard FR4 backplanes. Having established the environment we then set out to reproduce the Marvell's published results to establish a degree of confidence in our simulation and in the actual Alaska SerDes performance.

The Marvell white paper showed the results of their measurements with a 40 inches differential pair across FR4 passing through two connectors and four SMA coaxial sockets to launch and receive the signals between their transceivers. It was not clear from the document which connectors they used, where they were placed along the 40 inches of trace and which was the configuration of the trace, so we made a model that was intended to be representative of a typical application. This model included two microstrip traces of five inches each on a daughter board spaced thirty inches of striplines apart on a backplane. This sets an (extremely) upper bound for a proprietary rack. Obviously the distance was chosen to highlight the power of the transceiver, the message being that if Marvell can resolve this extreme case, then all the simpler cases can also be resolved with considerable headroom.

Figure 4.10 a.) shows our simulated waveform compared to Figure 4.10 b.) showing Marvell's reported results. The representations are in terms of eye-diagrams (defined in the paragraph 3.3.3.5).

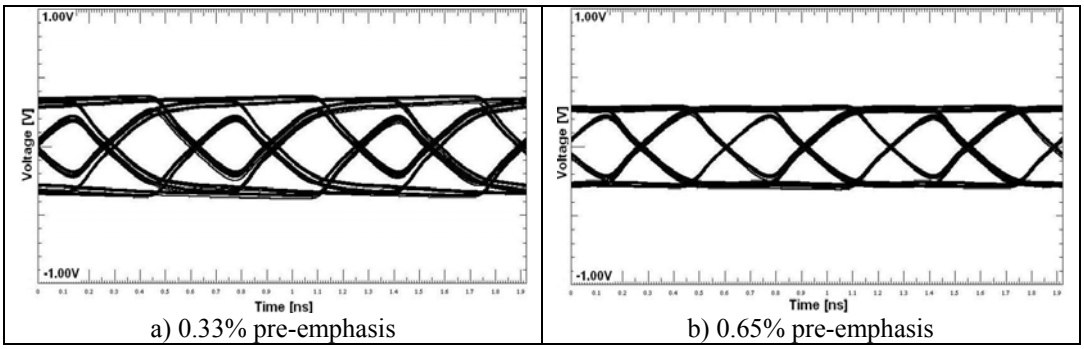


a) Our simulation result b) Corresponding Marvell white paper result

Figure 4.10: Reception after 40 in. FR4, no pre-emphasis

While Marvell's results have a slightly less open eye, due most probably to more lossy PCB traces on the daughter-cards and the backplane, the jitter effect is comparable. It was immediately clear however that the present conditions (when no pre-emphasis was used in the transmit signal) could not be relied for transmitting data at 3.125Gbps over 40 inches of FR4 trace. For achieving a reliable transmission, we needed to use signal conditioning techniques and switch on the pre-emphasis circuitry of the Alaska transmitter.

Marvell's white paper does not specify either which is the level of pre-emphasis they use to improve the signal quality. We then performed simulations for two different levels of pre-emphasis (0.33% and 0.65% pre-emphasis) and Figures 4.11 a.) and b.) show our results for these cases.



a) 0.33% pre-emphasis b) 0.65% pre-emphasis

Figure 4.11: Reception after 40 in. FR4

Either case would now be a good candidate for data transmission. They compare favourably with Marvell's results shown below in Figure 4.12.

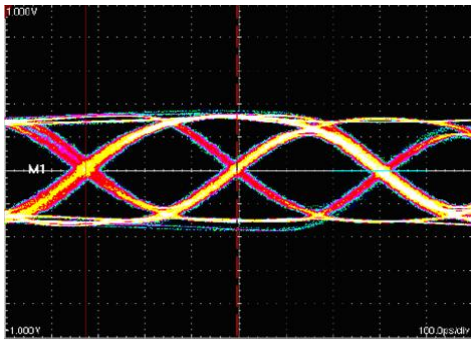
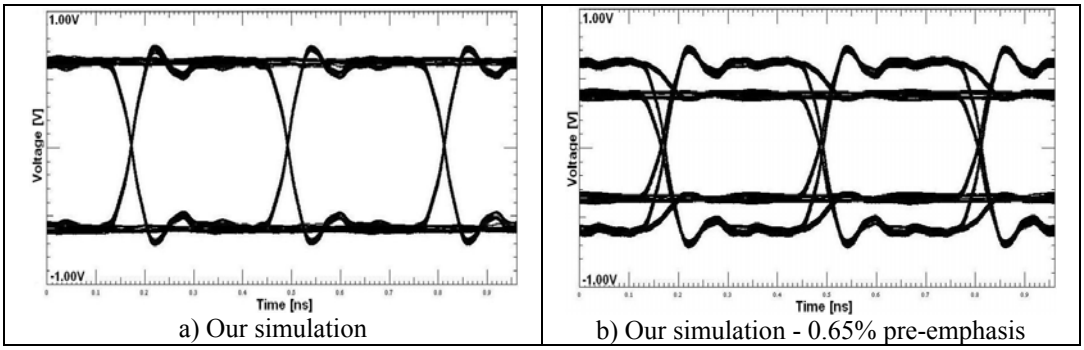


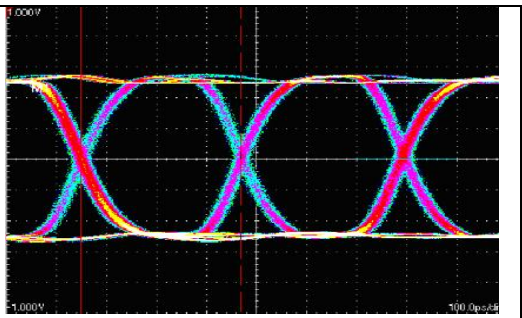
Figure 4.12: Marvell white paper result - Reception after 40 in., with pre-emphasis

It was interesting to note the effect that pre-emphasis has on the transmitted signal, although in their paper Marvell do not stipulate the amount of pre-emphasis used to obtain this plot. To see how pre-emphasis affects the transmitted signal consider Figures 4.13 and 4.14 a.) and b.) which show our and Marvell's simulated transmit signal with and without pre-emphasis.

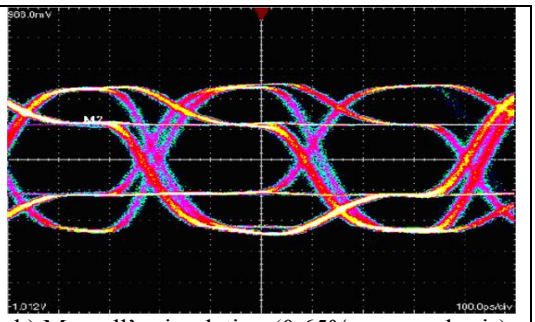


a) Our simulation b) Our simulation - 0.65% pre-emphasis

Figure 4.13: Signal Transmission without pre-emphasis



a) Marvell's simulation



b) Marvell's simulation (0.65% pre-emphasis)

Figure 4.14: Signal Transmission with pre-emphasis

The signal pair simulated by Marvel has a greater level of jitter than shown in our simulation although it does not exhibit the ringing that we observe after each transition. We found that this ringing is directly attributable to the mismatch between the impedance of the PCB trace and the termination impedance inside the chip package.

There is also a good correlation between our and Marvell's simulations. Note that the scale of the plots in both cases is the same and that the voltage levels, respectively the eye openings for the emphasised and non-emphasised bits are very close.

We conclude that the behavioural model supplied to us by Marvell corresponds to their own publicity. The differences we observed between our simulations and the Marvell's published results are most probably due to the differences encountered between our and their PCB trace and connectors models. Our conclusions however after this validation study were that Marvell Alaska 88X2040 is a highly performant I/O circuit that should meet our requirements for achieving error-free backplane transmission even over a maximum length of 40 inches of FR4 PCB trace.

It is important to note however that what has been simulated here, and the comparisons that are made, correspond to an electrically clean environment in which there is no simultaneous switching noise from other outputs in the same package, no crosstalk from adjacent PCB traces, no other signals passing through the same connector and no reflections from deep vias in thick backplanes. These issues are treated separately in chapter 5, by extending the presented methodology to carry out the simulations.

4.4 Summary

The prototype 10 Gigabit Ethernet switch built within ESTA uses state of the art pre-standard 10Gbps Ethernet components as the basis for its design. This chapter covers the main aspects and the factors involved in the selection of those components, starting with the appropriate interconnects (hereby included the high-speed connectors and the PCB traces) and finishing with the packaging techniques and chip technologies required by the engineering aspects of the Ethernet switching backplane.

Choosing the right components and determining their limits of operation was a necessary input to the design process. Therefore we have reported here and justified every component choice we made. In addition, we have double-checked the companies

published results for the chosen 10Gbps Ethernet transceiver device, the Marvell Alaska 88X2040 and we established that way a degree of confidence in our own simulation setup.

5 The Design, Modeling and Simulation of a 10Gbps Backplane

Once we have established a good knowledge of the available PCB interconnect, connector and line driving technologies we are able to put the different elements together in a design for a backplane that meets the connectivity and signal integrity requirements of the projected 10Gbps Ethernet switch.

5.1 Backplane Design

We consider the routing requirements of the backplane to determine how many layers will be necessary and then the stack-up that will meet that need. The stack-up of a board defines the PCB cross-section, the total number of layers, the overall thickness of the board, the thickness of the individual conducting layers and the dielectric properties used in the fabrication of each layer.

The routing requirements derive from the total number of traces that need to be routed over the backplane, their routing constraints and their impedance control requirements. We studied these requirements, the available possibilities and the issues related to the signal transmission on the 10Gbps Ethernet switching backplane. Following that study we derived a proposal for the backplane architecture of the 10Gbps Ethernet switch and for the backplane board stack-up ([OLT-R2], [ESTA-R3]).

5.1.1 Routing Constraints on the Backplane

We have already established that the current Serdes technology is limited to 3.125Gbps. To reach a bandwidth of 10Gbps four such channels are required. This gives a total of 12.5Gbps but this is data encoded with the 8B/10B scheme. Removing the encoding overhead yields a real payload of 10Gbps.

This is adequate for point-to-point connections but it is not sufficient for transport to and from the switch fabric. The act of deciding the routing path and then dis-assembling Ethernet frames into cells small enough to make optimum use of the switching fabric, running a transfer protocol between the port interface and the switch fabric, and then re-assembling the frame once switched to the correct output port is all pure overhead which takes time. To compensate for this it is common practice to run the switching fabric circuitry at a higher bandwidth than the sum of the ports that it services, typically between 30 to 60% [BRO-03]. For the switch fabric in our 10Gbps switch design a 50% increase in bandwidth is therefore required and to account for that the number of links between each board and the switch fabric must rise from 4 to 6.

The topology chosen for the data switching over the 10Gbps Ethernet switching backplane was the redundant dual-star architecture. Each switching plane is designed for 16 ports and a simple calculation which takes into account the number of ports, links per port and differential pairs per link, yields a total of 384 pairs of differential traces (16ports * 2switching planes * 6links/port * 2diff.pairs/ link/direction) available on the whole backplane. This is not counting at least one and possibly two PCI busses for control, as well as other interconnects for power monitoring.

In these conditions, one of the main problems was given by the routing of this number of traces within a limited available area existent on the backplane. We came out however with a proposal for the backplane architecture and lay it out, as shown in Figure 5.1.

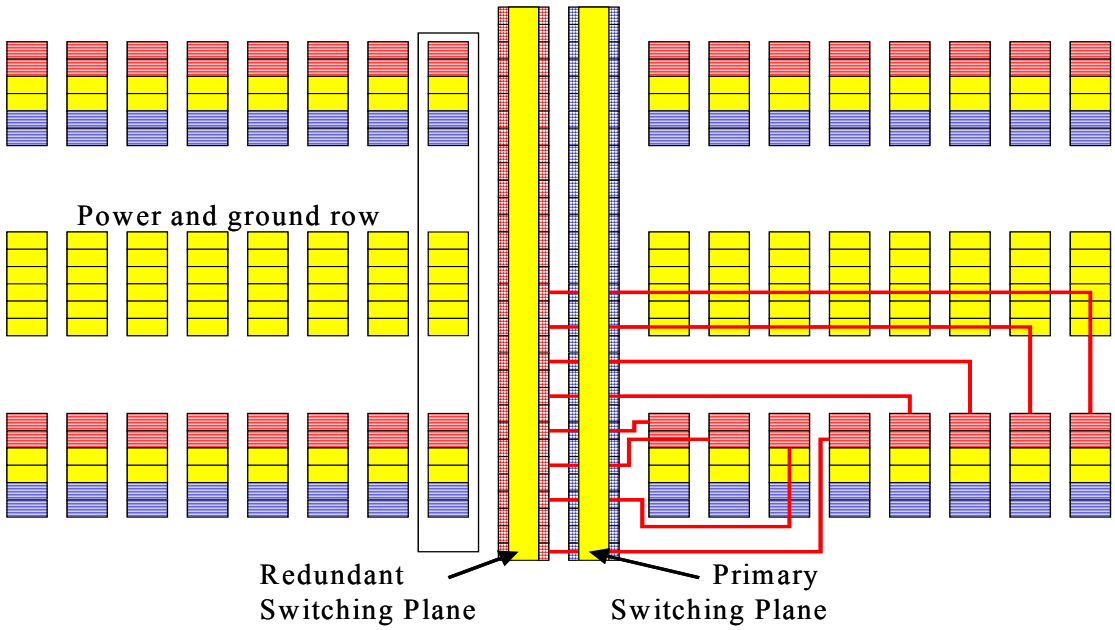


Figure 5.1: Proposal for the Backplane Architecture

We have placed the two switch fabric planes in the centre of the backplane board with eight ports on each side. Each port is equipped with three connectors: the top and bottom connectors carry interconnects and the middle one distributed the power and extra ground. No provision was yet made for anything other than the switching fabric and the connectors, as we first had to see if that will fit in the available space.

The lines in the bottom right quadrant indicate where groups of 6 links (times four traces) need to be routed between the ports and the connector of the primary switching plane. This is the most complicated route since it has to pass through the primary switching plane before reaching the secondary one. It was clear from the outset that in the area of the switching planes connectors there is not going to be any free routing path that avoids the connectors. At least one and probably more connector footprint have to be crossed by traces that don't make connections there. The cross-over area between these two switching planes represents an obvious bottleneck for signal routing.

The Figure 5.2 shows a horizontal slice through the switching fabric connectors (Tyco HM-Zd or Erni ERmet ZD). These connectors contain four differential signal pairs per each row and their ground pins are not shown for clarity.

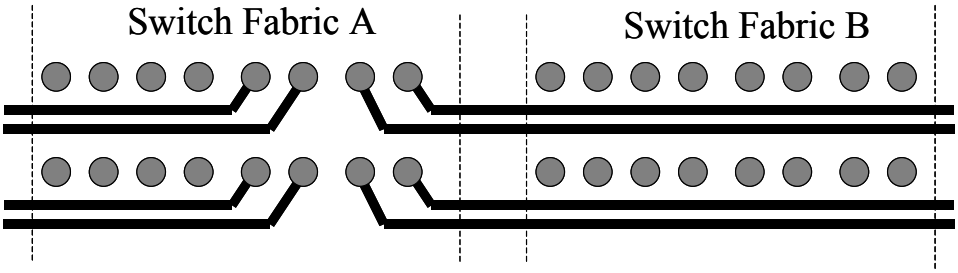


Figure 5.2: Routing through the Connectors on the Backplane

As the connectors are one of the most critical area for routing on the backplane, we consider how many routing planes will be needed for this. We choose to only route one pair between adjacent pin rows so it is obvious that four layers will be needed to route the escapes from each row of HM-Zd connector.

We considered then if the available space on the backplane is sufficient to accommodate all the connectors and all the traces required to achieve the desired connectivity between all the slots. The HM-ZD connector can supply 40 signal pairs/inch. Thus each switching plane with 384 differential traces (192 signal pairs) requires 4.8 inches of connector to carry the interconnect. As a 6U board has just less than 9 inches of connector space, there is no major problem with providing enough connectors pins for routing the signals corresponding to each of the switching plane. There will also be routing requirements for the other services mentioned (e.g. control and power monitoring), but no need is foreseen for making the backplane taller or for using a higher density connector than the already chosen four pairs HM-Zd (or ERmet ZD).

5.1.2 PCB Trace Constraints

Here we address the concerns of maintaining well-defined target impedance design for PCB traces that have to pass through space constrained areas. From this perspective, one of the most geometrically constrained areas is the one where the traces must navigate between the rows of connector pins, as shown in Figure 5.2. A typical differential signal pair routing through a slice of a generic connector footprint with its corresponding parameters is shown in Figure 5.3.

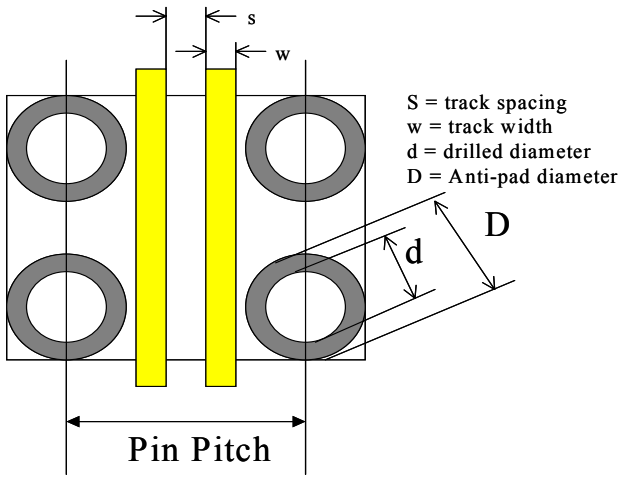


Figure 5.3: Trace Routing Through a Generic Connector Footprint

As those parameters which are directly affecting the controlled impedance design were already discussed in the paragraph 4.2.2, we could directly come out with a list of constraints needed to be satisfied by a pair of differential traces routed on the backplane through the connector footprint. In order to achieve a 50Ω single-ended, respectively a 100Ω differential impedance design, the following, conflicting, constraints should be accomplished:

- The two traces shall fit between the connector pins without overlapping the anti-pad footprints;
- The traces shall be as far from adjacent vias as possible to reduce the crosstalk effects;
- The traces shall be as wide as possible to reduce skin effect losses over long distances;
- The traces shall be as narrow as possible to minimise the distance between the reference planes and thus the overall board thickness;
- The traces shall be as close together as possible to ensure equal trace lengths over the whole routed path;
- The traces shall be as far apart as possible to maintain a differential impedance twice that of the single-ended impedance.

The goal was to find the optimum parameter values for the PCB trace configuration among those conflicting constraints and to use them to achieve the proposed impedance target for the design of the 10Gbps Ethernet backplane.

The PCB manufacturing has its own limitations which have to be considered when making the design choice. For instance, the smallest trace width for volume production is 5mil and the minimum trace-to-trace separation is 5mil as well. In addition, for any given connector the degrees of freedom are also quite limited. Taking the case of Tyco HM-Zd, its critical dimensions are presented in Table 5.1.

Pin Pitch	98.425mil
Drill Diameter	27.6mil
Anti-Pad Diameter	50mil
Max Routing Channel	48.425mil

Table 5.1: HM-Zd Connector's Critical Dimensions

Figure 5.4 shows graphically the remaining options within the available routing space for the HM-Zd connector.

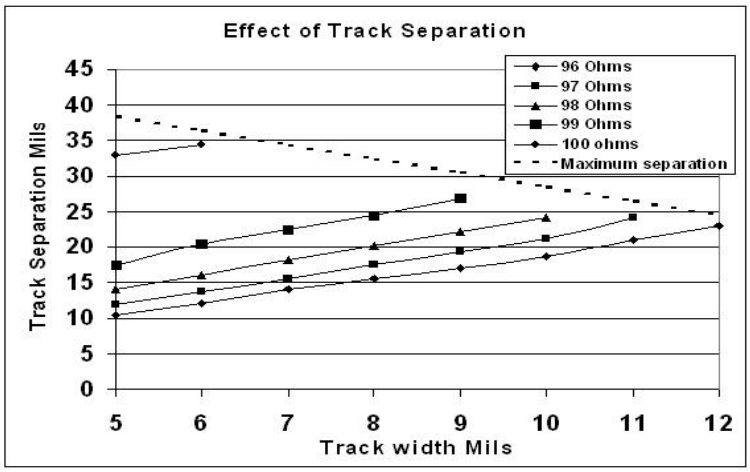


Figure 5.4: Possible Trace Width/Spacing Options within the Connector Routing Channel

It can be seen that true 100Ω differential impedance could only be possible with 5 or 6mil trace width and for more than 30mil of trace separation distance. These dimensions however are not feasible for the HM-Zd connector under study. We could still achieve an impedance value within 1% for traces up to 9mil wide, 2% for 10mil, 3% for 11mil and 4% for 12mil. How then could we choose the optimum between these? We based again our decision on the previous study (see paragraph 4.2.2.3) where conclusions were leading to the consensus of using values between 8 and 10mil for the trace width. In addition, that study showed that it was not much difference in the manufacturing tolerances and in the signal transmission quality when 8mil and 10mil width traces were used. As we were constrained by a limited space within the connector routing channel, we then decided for using the minimum trace width of 8mil in our design.

Now, depending how close we want to be to the desired 100Ω differential impedance, we have to adjust the space between the traces and the space from the traces to the reference plane. When choosing the right distance between the two traces, we had to consider that a minimum of 5mil trace-to-anti-pad spacing is required at the manufacturing. This then leaved us options for the trace-to-trace separation from 5 up to 22.4mil (48.425-2*5-2*8), which corresponded to a trace-to-(via) anti-pad distance (see paragraph 5.2.3) from 5 to 13.7mil ((48.425-2*8-5)/2).

Taking into account all the considerations described above, such as connector routing channel limitation and the goal of achieving an impedance-controlled design, we proposed the following geometry dimensions for the pair of edge-coupled striplines to be routed over the backplane:

- Trace width = 8mil

- Trace-to-trace distance = 18mil
- Trace-to-anti-pad distance = 7.2mil
- Trace-to-reference plane height = 10mil

The stack-up calculator of Cadence SpecctraQuest has been used to obtain the impedance values of the edge-coupled stripline configuration with the dimensions given above. For FR4 dielectric material (electric permittivity εr =4.3 and loss tangent lt=0.022), the calculated values are 51.6Ω for the single-ended impedance and 100.6Ω for the differential impedance.

Later on in this chapter (see paragraph 5.3), extensive studies will be presented concerning the influence of the connector pins on signal traces routed through the connector rows. If this influence, which is directly proportional to the trace-to-anti-pad distance, will be too high for the chosen geometry distances, then a new configuration having a more relaxed position of trace related to the anti-pad will be proposed.

5.1.3 Backplane Stack-up

Knowing the requirements of the backplane design, after having proposed the backplane architecture (in paragraph 5.1) and the optimum dimensions of the edge-coupled stripline traces (in paragraph 5.1.2), we can now define the backplane PCB stack-up. We assumed that the following number of layers needs to be considered for the stack-up of the backplane:

- Four controlled impedance routing layers for the switch fabric;
- Two controlled impedance routing layers for the control bus (conservative);
- Four power planes;
- Top layer with no signals, just extra ground and power pin via pads;
- Bottom layer, place for decoupling and termination components;
- Two extra signal layers without controlled impedance.

Defining the layers that make up the PCB is driven by manufacturing requirements as well as by the need for optimal precision. The printed circuit is made up of a succession of layers of double sided copper clad FR4 dielectric bonded together with a fiber mesh pre-impregnated with epoxy (pre-preg). The ‘sandwich’ of layers is baked under pressure to cure the epoxy which forms a dielectric with a permittivity very close to that of normal FR4 although it varies with the amount of resin in the epoxy mixture. The plasticity of the pre-preg under pressure results is a much lower dimensional stability that for the pre-cured FR4. Therefore any impedance controlled PCB track must have at least one reference plane ‘seen’ through FR4 and must never be between two layers of pre-preg.

Another constraint is that the stack of layers must be symmetric about the center otherwise the finished board will be at risk of warping during the curing process. Passive decoupling is achievable in PCBs if the dielectric between power and ground planes is

reduced to a minimum. Where possible then, power and ground planes should be placed face to face separated by a single, thin dielectric. Several picofarad per square centimeter are possible without the dielectric becoming so thin that it risks violating the security requirement of ensuring that power and ground never come into contact.

Consider then the final stack-up shown in Figure 5.5. The center of the stack is layer 11, a ground plane, around which layers 1 to 21 are arranged symmetrically. Layers 10 (VCC2) and 12 (VCC3) are separated from this ground by only 2mil FR4. Moving outwards to layers 9 and 13 (VCC1 and VCC4) are also separated by 2mil FR4 although this is to reduce the maximum thickness of the build rather than for any decoupling advantage since neither of these planes is closely referenced to ground. Placing all the power planes at the core means that none of the controlled impedance layers needs to be referenced to anything other than ground. Nominally ground and power are considered equal for referencing purposes but in practice the best results are obtained by referencing to ground. Layers 8 and 14 therefore are used for the two non-impedance controlled track layers. The controlled impedance layers Signal 1 to 4 are all encapsulated between one FR4 and one pre-preg layer between full ground planes. Finally layers 2 and 20 contain the two controlled impedance tracks for the control bus. There are no signal tracks on the top or bottom layers. This gives a total of 9 Ground, 4 Power, 6 controlled and 2 no-controlled impedance signal layers.

Including all the layers, the total thickness of the backplane PCB results equal to 183.7mil. The press-fit pins of the connector can be delivered with a length of up to 2.5mm or 100mil long. Press-fit pins exercise horizontal stress on the inside of the via walls into which they are inserted to generate a gas tight connection. This stress is therefore roughly in the center of the board which will not cause undue bending when all the connectors are in place. At most however we could add another two signal layers or further power planes before causing such problems.

#	Layer	Material	Mils
1	TOP	Cu	0.7
		FR4	12
2	Control 1	Cu	0.7
		pre-preg	12
3	GND	Cu	0.7
		FR4	12.6
4	Sig 1	Cu	0.7
		pre-preg	18
5	Gnd	Cu	0.7
		FR4	12.6
6	Sig 2	Cu	0.7
		pre-preg	12.6
7	GND	Cu	0.7
		FR4	4
8	X layer1	Cu	0.7
		pre-preg	4
9	VCC 1	Cu	1.4
		FR4	2
10	VCC2	Cu	1.4
		pre-preg	2
11	GND	Cu	1.4
		FR4	2
12	VCC3	Cu	1.4
		pre-preg	2
13	VCC4	Cu	1.4
		FR4	4
14	X layer 2	Cu	0.7
		pre-preg	4
15	Gnd	Cu	0.7
		FR4	12.6
16	Sig 3	Cu	0.7
		pre-preg	12.6
17	Gnd	Cu	0.7
		FR4	12.6
18	Sig 4	Cu	0.7
		pre-preg	12.6
19	GND	Cu	0.7
		FR4	12
20	Control 2	Cu	0.7
		pre-preg	12
21	Bottom	Cu	0.7
		Total	183.7

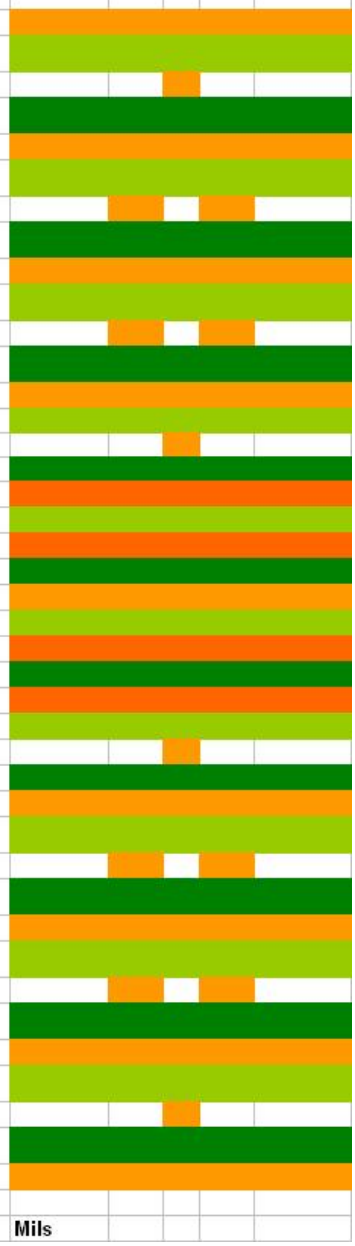


Figure 5.5: Possible Stack-up for 10 Gbps Backplane

5.2 Via Hole Terminology

A plated through hole (PTH) or 'via' is typically used to receive connector pins and/or route signals between different layers of a PCB. Figure 5.6 shows a via hole which was manufactured by drilling through all the PCB layers. It is used to explain some of the terms which appear in the present chapter.

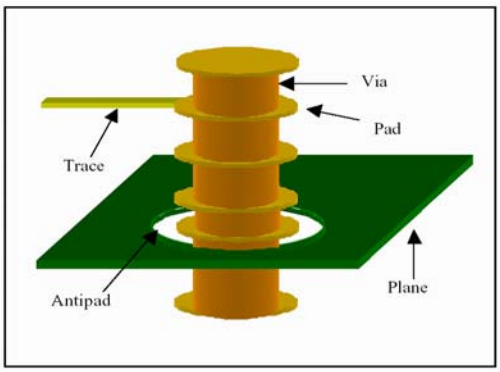


Figure 5.6: View of a Signal Via through a PCB

5.2.1 Via Hole Diameter

Ideally all signals would be routed between source and destination on a single layer. Even for simple designs a minimum of two layers are need to cope with the fact that some signals will run in the 'x' direction and some in the 'y' direction and mutually block each other. As the track density rises and there is less 'real estate' available to route traces with sufficient spacing between them, more layers are added and signals will often be moved between them to find a suitable route. Moving between layers is done through a 'via' which is a hole drilled through the PCB stack and plated for conductivity. However the act of drilling the hole removes routing real estate from each layer of the PCB and reduces the routing options on every layer. Therefore it should be drilled with as narrow a diameter as possible. There are limits to how narrow this can be however. As the thickness of the PCB increases for any given via diameter it becomes more difficult for the plating process to evenly coat the inner walls of the hole. The practical limit to the aspect ratio (height divided by diameter) that can be successfully plated is about 5. For backplane applications however, the barrel diameter limit is set by the requirements of the connector pin that is inserted into it. This is much larger than is needed for a signal via. Typically, signals can be routed through 6mil diameter vias whereas a connector can require from 20 to 40mil in via diameter. This increase in dimension of via diameter has direct repercussions on the capacitance of the interconnect and hence on the signal integrity.

5.2.2 Via Pad

The via pad is a circular pad of copper, somewhat wider than the drilled via diameter, and used to make a good connection between the surface trace and the conducting via barrel. They are placed on the top and bottom layers of the PCB at each end of the via barrel. Similar pads are used to make the connection to traces on internal layers of the board, irrespective of if there is a connection there to be made or not. Originally, this was to assist in the take-up of copper that was deposited in the barrel. With modern deposition technologies this is no longer necessary although some manufacturers still insist on it.

Typically the pad diameter is between 10 and 18mil wider than the barrel diameter. This is to ensure that even if the drill is off-centre within its allowed tolerance there will still be sufficient copper to ensure the connection as shown in the Figure 5.7.

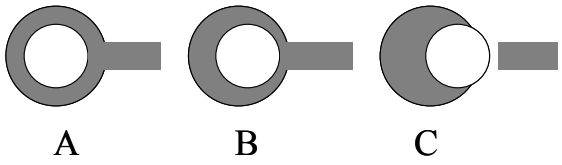


Figure 5.7: Drilling Tolerance Requirements

Case A is a well-centered barrel, B is off centre but within tolerance, and C is an out of tolerance drilling.

5.2.3 Via Anti-Pad

When a via passes through a power or ground plane there must be sufficient clearance around the barrel and its pad so that it cannot make accidental contact with the plane it passes through. This clearance is called the anti-pad. Typically an anti-pad will offer 10mil of clearance around the drill hole. The anti-pad dimensions causes problems for the routing of the traces since the anti-pad occupies the area of copper that the traces are referenced to for their controlled impedance.

A first issue is related to the return current distribution in the anti-pad area. Consider the case in Figure 5.8 where two traces, A and B, are asymmetrically disposed about a via.

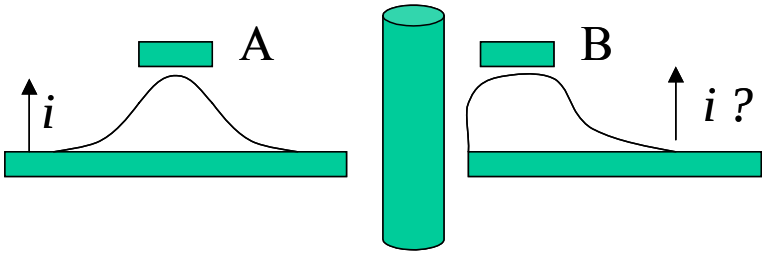


Figure 5.8: Reduction in Current Return Path due to the Anti-Pad

The trace A is sufficiently far from the anti-pad gap in the ground plane and the return current is symmetrically distributed in the ground plane along the axis of the trace and spreading over about three times the trace width. In the other case, the trace B is as close as it can be to the via barrel without overlapping the anti-pad. The return current from trace B has some unknown distribution, which by definition will have some possibly negative affect on signal integrity.

A second issue of the anti-pad dimension is the fringe capacitance. Consider the case in Figure 5.9 where the non-connected pad has been removed.

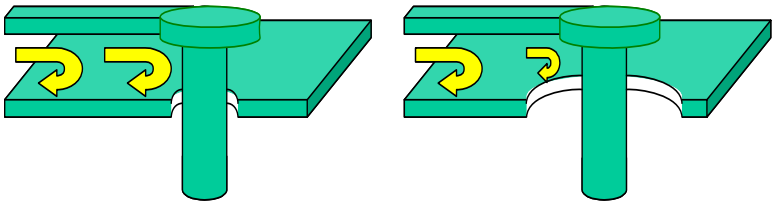


Figure 5.9: Anti-Pad Effects

If we opt for a minimal mechanical tolerance (case on the left), then there will be significant capacitance between the ground plane and the via which affects the signal integrity. If we expand the anti-pad size (case on the right) the capacitance is reduced but with the ground plane now further away from the via barrel, there is less continuity for the return signal path (represented by the arrows) and a greater discontinuity for the signal using that via.

These issues of the anti-pad dimensions and the position of a trace in the proximity of a via are closed related to the routing of the traces and determine the optimum routing scenario on the backplane PCB.

5.3 Connector Pin-field Modeling

5.3.1 The Concerns

In a backplane based system, such the one used to host the prototype 10Gbps Ethernet switch, the signal path from transmitter to the receiver which are situated on two different line-cards has to pass through two connectors before reaching the backplane. The connector structure at the daughter-card interface with the backplane is referred as the 'connector launch' area and constitutes one of the most important impedance discontinuities on the transmission path. The effect of the connector launch has been considered by some authors ([VOG-x], [CLI-03]) more or less in the same time with the study hereby presented.

In addition, we have shown that, on its way to the receiver on the backplane, the signal trace has to pass through the connector rows corresponding to the primary switching plane. This area on the backplane where a pair of differential traces passed through the connector pins (see Figure 5.2) is defined by us as the 'connector pin-field' region. Due to the proximity of the traces to the connector pins within the constrained routing channel, the connector pin-field can constitute one of the most important factors of crosstalk on the transmission channel. Connector companies were beginning to study these issues but there were no published reports available until 2003/2004.

The specifications of the studied connectors showed good crosstalk figures (see paragraph 4.1.1) but all these are true only inside the body of the connector in isolation from any PCB. Unfortunately, the connector pin entering the PCB no longer has proper ground isolation and there is where the main crosstalk effects occur. AMP did some typical measurements of a 'real' environment [AMP-x]. The following extract from their study is highly significant:

"At 3.125 Gbps the connector has a significant impact on system performance. Typically when reviewing the performance of a connector, little consideration is given to the footprint of the connector in the PCB. ……Using the conditions specified by the XAUI Channel Working Group the worst case simulated noise of the Z–PACK HM-ZD was 3.1%. This meets the ……..proposed connector noise budget of 4%.

The inclusion of the footprint significantly changed the overall shape and magnitude of the near end and far end noise pulses. For all three pin-outs considered the amount of near-end noise increased by at least double. The impact on far-end noise was pin-out dependent, ranging from a slight decrease (Pin-out 2) to a significant increase (10.5 times for Pin-out 1 and 2 times for Pin-out 3"

AMP study confirmed our expressed concerns about the crosstalk in the connector. It showed that the admitted effect of the connector can be almost discounted since the act of plugging it into a backplane can increase the crosstalk noise by anything or up to a factor of ten, depending on how carefully the footprint was designed. We considered then that an extensive study of the connector pin-field area on the backplane would be a prudent way to go before sending the backplane PCB proposal for manufacturing.

5.3.2 Defining the Problem

Considering again the subset of a connector pin-field in Figure 5.2, the traces carrying differential signals on the backplane were passing between the connector pins which carry other differential signals. The traces have a crosstalk influence on the connector pins (vias) and vice versa. Our crosstalk concern was not particularly for the victim pair of vias since the zone of interaction was so short and noise coupled to one via is very close to the noise coupled to the other via in the differential pair. In addition most of this coupling was in common mode and the receiver was therefore largely insensitive to it. However, noise coupled from the connector's pins to the nearest trace of the differential pair would be more than the noise coupled to the farther trace. Thus differential noise could be induced into the trace pair, which carries signal along the backplane. In the case of the backplane environment for the 10Gbps Ethernet switch for instance, the signal traces pass through at least eight pairs of differential vias (4 pairs/connector row on each side of the traces), and that only corresponds to the primary switching plane. The suspicion is that the induced crosstalk noise from all these connector pins would accumulate up to the point of creating errors in the signal arriving at the receiver.

To address these concerns, we had to resort to modeling and simulation techniques and further on, we will describe how 3D electromagnetic modeling has been used for extracting an accurate equivalent model for the connector pin-field structure.

5.3.3 Establish Simulation Environment – Quasi-static vs. Full-wave

Complex PCB interconnects are usually modelled with either the 3D quasi-static or the full-wave modeling approach ([SWI-99], [WIL-x], [WAL-00], [GEE-01], [LAE-02]). The choice of modeling approach depends on the size, the complexity together and the maximum frequency of interest of the signal which passes through that interconnect. Then a quasi-static approach represents the interconnect by cascaded sub-sections of lumped or distributed R, L, C and G circuit elements, as it was the case of the models for the connector obtained from the manufacturer (see paragraph 4.1.3). By contrast, the full-wave modeling approach completely solves Maxwell's equations inside the volume of the interconnect, and outputs its model in terms of multi-port scattering-parameters (S-parameters) representation [POZ-98]. A good summary of the interconnect modeling with an overview on the quasi-static and full-wave modeling approach is presented in [OLT-R1].

The advantage of quasi-static modeling is that it is fast and computationally inexpensive compared with full-wave modeling. It has however practical limitations that should be understood and considered. The quasi-static representation loses accuracy once the frequency gets to the point that the largest dimensions of the interconnect becomes less than one tenth of the electrical wavelength. This is because some assumptions in the propagation of the electromagnetic field made by the quasi-static calculations are no longer valid for small complex PCB structures at higher frequencies.

For modeling the connector pin-field region on the backplane, we employed first the quasi-static approach by using Maxwell Q3D, the quasi-static field solver from Ansoft [AN-Q3D-1]. To meet the requirements of the reduced dimensions of the interconnect to be quasi-statically modelled, we have split the connector pin-field into several sections and model only such a small section of the whole area. The section chosen for modeling was about 7mm long and contained two pairs of connector pins and their adjacent ground on each side of a differential pair of traces. The section geometry was then entered using the drawing tool of Maxwell Q3D and the excitation points for the whole structure were defined according to the settings requirements of the tool. Maxwell Q3D uses the Finite Element Method and Boundary Element Method to compute frequency dependent R, L, C and G matrices for the given connector pin-field section [AN-Q3D-2]. When the solution process was complete, the data (RLCG representation) was exported from Maxwell Q3D as lumped and distributed models in SPICE.

The transient simulations in SPICE with the connector pin-field model obtained from the quasi-static field solver showed that this model loses accuracy once the frequencies increase to the GHz range [OLT-R1]. A more careful look at the model geometry will show why. An interconnect of 7mm becomes electrically 'long' once its electrical wavelength is greater than 70mm. The frequency value which corresponds to a 70mm wavelength for a FR4 dielectric result around 2GHz. So, the theoretical validity of the quasi-static approach for the modelled section of 7mm connector pin-field goes up to 2 GHz. If we take into consideration that the validity of an accurate model representation must go up to the 5^{th} frequency harmonic, which is about 8GHz for our 3.125Gbps signal,

we rapidly realized that a quasi-static representation of the connector pin-field section would not be accurate enough for our requirements.

Some other similar studies were showing as well that the quasi-static approach was not accurate enough for small geometries above certain frequencies and recommended the use of the full-wave modeling for obtaining an accurate model, valid over a wide range of frequency, for the complex PCB interconnects ([CAM-x], [PLA-02]). When compared to computationally inexpensive quasi-static approach, full-wave modeling has a series of advantages, as it accounts for frequency-dependent losses, higher-order mode propagation and other parasitic electromagnetic effects which appear during signal propagation in the interconnect [POZ-98]. The solution however is computationally demanding and very slow.

For the full-wave modeling approach we used HFSS (High Frequency Structure Simulator) [AN-HFSS] which is the Ansoft's full-wave field solver. HFSS takes as input the 3D geometry of the connector pin-field section within the PCB stack-up, the electromagnetic characteristics of the materials inside its geometry volume and the boundary conditions required for the computation of Maxwell's equations, which are solved by HFSS using the Finite Element Methods. The output of interest for us from HFSS for the connector pin-field section is represented by the frequency-domain S-parameters ([AN-HFSS], [AGI-00]), which can be later exported into a full-wave model for time-domain SPICE simulation.

5.3.3 Calibrating the Simulation Environment

A full-wave simulation can give good results and provide accurate models for multi-Gigabit applications, as long as the simulator is well-guided by providing a good problem definition and setup. HFSS has a very steep learning curve, but this was accepted at the cost of having trustworthy results. Modeling in HFSS involves many variables regarding the optimum size of the geometry, the correct boundary conditions to be chosen and other issues related to the best meshing strategy and the proper frequency bandwidth setup. Details about the simulation process in HFSS and explanation of most of the terms and concepts appearing later in this paragraph is given in Appendix.

We found an optimum setup for HFSS by starting from a 'bottom up' approach where we first modelled the simplest structures (such as PCB traces, single-ended via and coupled differential vias) whose results we could calibrate against the literature ([AN-01], [ARO-02] and [XFP-MSA]). Some other publications addressing studies and modeling of different PCB structures in HFSS have been of great help [LAE-01], [SER-02], [WIL-02], [SZI-03]. From this calibration exercise, we extracted our own setup rules for the simulation in HFSS which we could further used in the modeling of the connector pin-field region of the 10Gbps Ethernet switching backplane.

5.3.3.1 Edge-Coupled Striplines

Figure 5.10 shows the edge-coupled stripline, geometry used as a first check of the HFSS simulator to see if we could obtain the expected results. The edge-coupled stripline is a typical configuration for backplane routing and its geometry parameters have the dimensions which resulted from our previous study in the paragraph 5.1.2. The modeling approach in HFSS requires to include in the analysis a certain part of the PCB around the stripline traces. The geometry introduced in HFSS for computation is shown in Figure 5.10. We had to find which are the optimum dimensions to consider in HFSS for the edge-coupled striplines so that this structure can be computed in a reasonable amount of time and converge quickly to a stable solution. In addition, we have also found the correct boundary conditions needed by HFSS for solving the Maxwell's equations within the volume of this geometry.

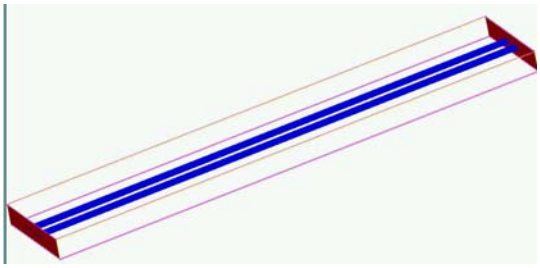
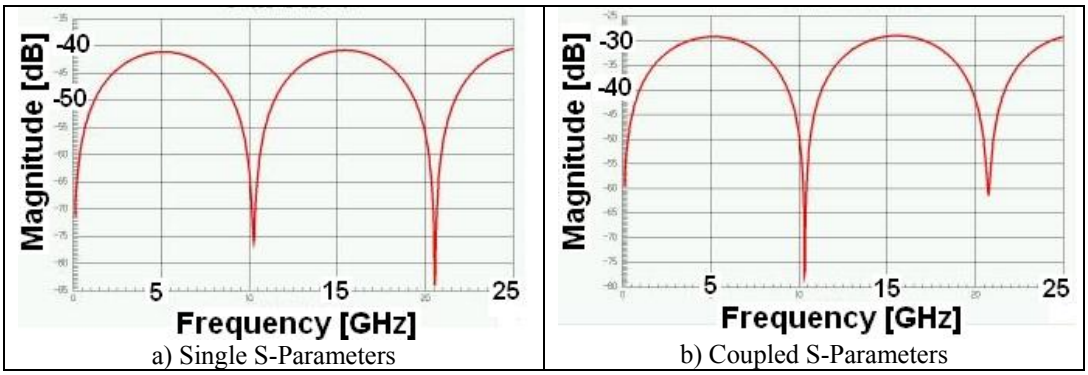


Figure 5.10: The Edge-Coupled Striplines configuration in HFSS

To reduce the calculation time in HFSS, we defined the ground planes within the PCB stack-up as having no thickness and as being made from an electrically perfect material (having an infinite conductivity). On the lateral sides of the volume box in Figure 5.10, we assigned Perfect-E boundary conditions to provide a low-impedance path between power and ground planes. This would be the real case in a PCB environment where the ground vias and the decoupling capacitors situated in the proximity of the stripline traces determine a low-impedance path. The electromagnetic wave excitations from HFSS to this structure is applied via two 'driven-terminal wave-ports' (see Appendix) placed at each end of the striplines.

We requested the solution process over a frequency range from 100MHz to 25GHz. HFSS had taken about ninety minutes on a 1GHz processor to compute the electromagnetic fields for this simple structure. The final mesh had 40000 tetrahedra, created and refined through an adaptive process for the maximum frequency of 25GHz. The final solution, in terms of modal S-parameters is provided through an interpolation of 1000 points over the chosen frequency range.

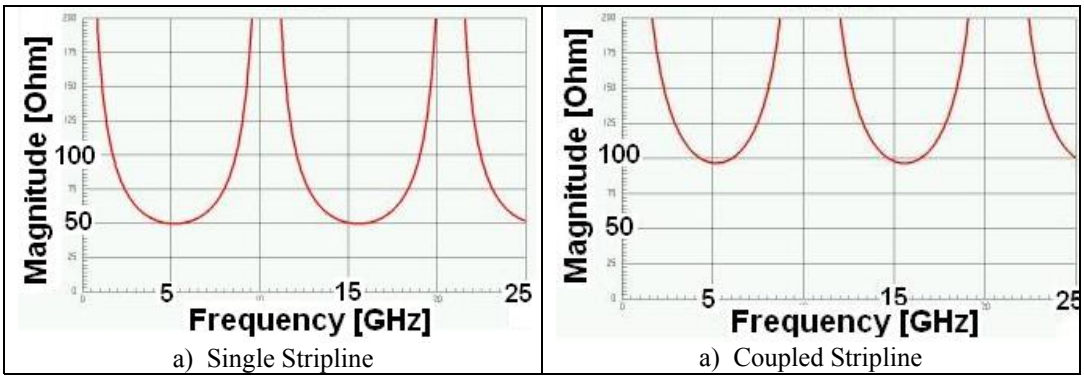
Figure 5.11 shows the single and differential return loss (S11) for the edge-coupled pair of striplines obtained from the simulation with HFSS. For both, single and coupled S-parameters, there was a nice clean signal insertion (S12) with no undesirable high frequency resonances.



a) Single S-Parameters b) Coupled S-Parameters

Figure 5.11: S-Parameters for the Edge-Coupled Stripline

For the particular case we simulated, the reflected energy was at the –30dB and below level, down in the numerical noise. The Figure 5.12 shows the impedance curve for the single and the differential stripline pair.



a) Single Stripline a) Coupled Stripline

Figure 5.12: Impedance Profile for the Edge-Coupled Stripline

The 'U' shape of the impedance curve is defined at its minimum by the characteristic impedance. The asymptotic maxima are the result of quarter wave-length resonances due to the shortness of the line compared to the wavelength of the exciting frequency. The characteristic impedance given by HFSS was within 0.5% of the theoretical value obtained from SpecctraQuest the stack-up calculator.

5.3.3.2 Single VIA

Figure 5.13 shows the geometry of a single via that was simulated in HFSS. It is a very short via hole (26mil high) which carries the signal from top to the bottom layer of a PCB. The anti-pads and the internal pad are taken into account around the via hole. Within the PCB stack-up, two ground planes and an internal signal layer are included. The microstrip lines on the top and the bottom layer are designed for 50Ω impedance and each of them is referred to a ground plane.

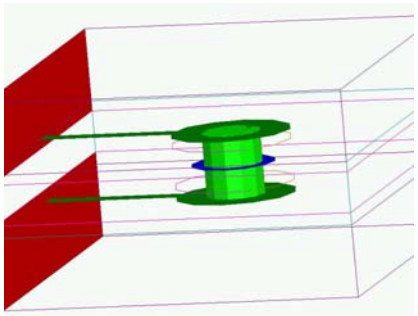
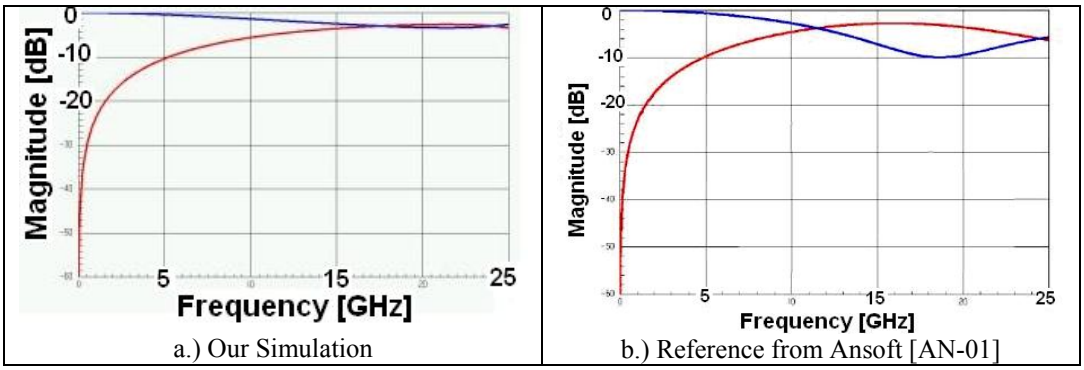


Figure 5.13: Single Via through a PCB

For the HFSS computation of the single via geometry we kept the same simulation setup as in the previous paragraph 5.3.3.1. HFSS provided the final solution with a mesh oft 70000 tetrahedra in about 3 hours.

For the single via case, the Figure 5.14 a.) shows our simulated result, and b.) shows the results presented by Ansoft in a reference design for a via ([AN-01]).



a.) Our Simulation b.) Reference from Ansoft [AN-01]

Figure 5.14: S-Parameters for a Single Via

The results we obtained from the simulation of a short via presents a 'generic' reflective curve rising up to about –5dB reflection for frequencies around 15 to 25GHz and we accepted this as sufficient proof of concept. In Ansoft's analysis, the signal through their via passes from the top layer to an internal layer of the PCB making a short via stub to the bottom. The resonance seen around 12GHz in their graph (Figure 5.14 b.) is due to that.

We had to accept that a short signal via between one or two ground planes has very little effect on the transmitted signals that pass through it. According to some authors [VIC-04], no undesired resonances appear for the signal passing through a single via from its top to the PCB bottom, until the thickness of the backplane board is 200mil or more. However, even though our via doesn't have the same dimensions as the Ansoft's geometry, the shape and the order of magnitude obtained for the insertion and return loss appear to indicate that our results are going in the right direction.

5.3.3.3 Coupled VIAs

Figure 5.15 shows a pair of differential coupled vias connecting differential traces on the top and bottom layers of a PCB. The two other vias shown to the left and the right of the differential pair provide a well-controlled current path to the ground. This geometry configuration is known in literature under the name 'ground-signal-signal-ground' (GSSG) and it is considered as an improvement to the 'traditional' differential via ('signal-signal' geometry) [XFP-MSA].

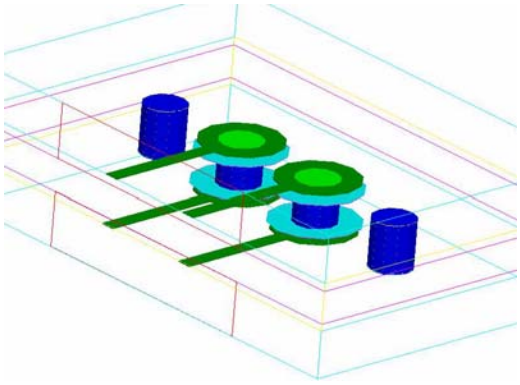
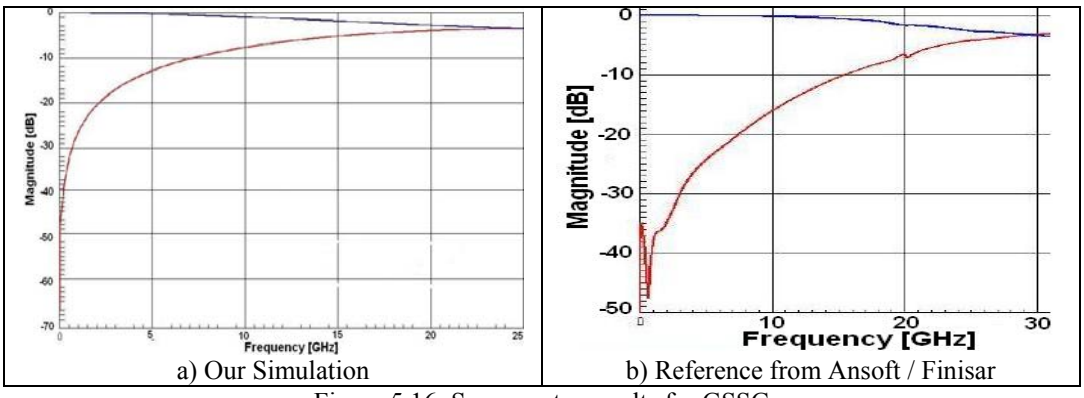


Figure 5.15: Differentially Pair of Via's (GSSG)

The GSSG geometry is a subset of a typical high speed connector footprint and the critical dimensions for the via diameter, pad and anti-pad corresponds in the case of the geometry shown in Figure 5.15 to the dimensions of the HM-Zd connector. A similar GSSG geometry configuration, but with different values for the via parameters and characteristics (such as oval anti-pad) has been studied by Finisar ([ARO-02], [XFP-MSA]). The results of our simulation in HFSS together with Finisar results are shown in Figure 5.16.



a) Our Simulation b) Reference from Ansoft / Finisar

Figure 5.16: S-parameter results for GSSG

The differences in the S-parameters magnitude (return and insertion loss) found the explanation in differences of dimensions between the two geometries (our simulation versus Finisar reference design). However, our simulation presents a resonance free increasing return loss up to –5dB above 15GHz and we considered that as being an acceptable good result.

5.3.3.4 Conclusions on the Calibration Study

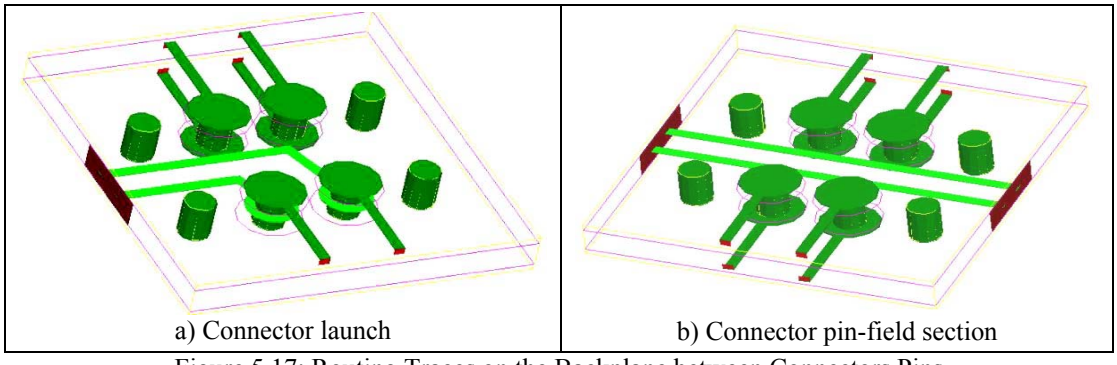
Three simple PCB geometries – edge-coupled striplines, single via and coupled vias - have been analysed as an exercise for the calibration of the simulation environment, in order to find the right inputs and settings in HFSS. We had to make some simplifications in the definition and the input of the given interconnect to give a tractable level of complexity, in order to achieve the final result in HFSS within a manageable amount of time and with a limited amount of computation power.

The simulation results we obtained have been compared with available published results which have been presented for similar geometries. These results were not 100% identical, as neither the geometries under study were not exactly the same. However, a relatively good correlation has been obtained and this gave us the sign that the modeling approach we considered was going to the right direction. The experience achieved during this successful calibration exercise, will be further applied and used for the modeling of the connector pin-field region in HFSS.

5.3.4 Full-wave Connector Pin-field Modeling

5.3.4.1 The Approach

It is not possible to simulate in a full-wave electromagnetic field solver, such as HFSS, a complete PCB or even a complete connector pin-field region corresponding to the primary or secondary switching plane on the backplane This is because large aspect rations (usually larger than 1000) are to be avoided, as the simulation might easily fail during mesh creation and might take many adaptive passes to converge to a correct solution. The observation above leads to the conclusion that the geometry to input in HFSS needs to include only a small section of the connector pin-field. The same approach as already considered in the quasi-static modeling was overtaken i.e. sub-divide the connector pin-field region into several sections. These sections are represented by parts of the HM-Zd connector with a pair of differential 8mil stripline traces passing between two rows of pins. Figure 5.17 a.) shows a section of the connector launch area at the backplane interface with the daughter-card, while Figure 5.17 b.) shows a section of the connector pin-field on the backplane.



a) Connector launch b) Connector pin-field section

Figure 5.17: Routing Traces on the Backplane between Connectors Pins

For each of these geometries, frequency-domain results in terms of S-parameters were computed in HFSS. Once the analysis in HFSS was successfully finished, we performed the data export into time-domain where we analyse the crosstalk behaviour in the connector pin-field region and its influence on the overall performance of the full-path transmission channel from the driver to the receiver across the backplane.

5.3.4.2 Crosstalk in the Connector Pin-field Area

By modeling the connector pin-field region from Figure 5.17 b.) in HFSS we could see the electromagnetic field behaviour in the geometry. Figure 5.18 highlights one frame of the current density vector animation showing the differential stimulation of the two vias and the currents induced in the adjacent traces due to the coupling effects. It could be seen that the adjacent trace to the aggressor pair of vias was picking up slightly more noise than the trace running further away. It was a relatively small effect we could see in this animation, but it effectively confirmed our concern from the beginning that some amount of differential noise was induced due to the routing topology in the adjacent PCB traces.

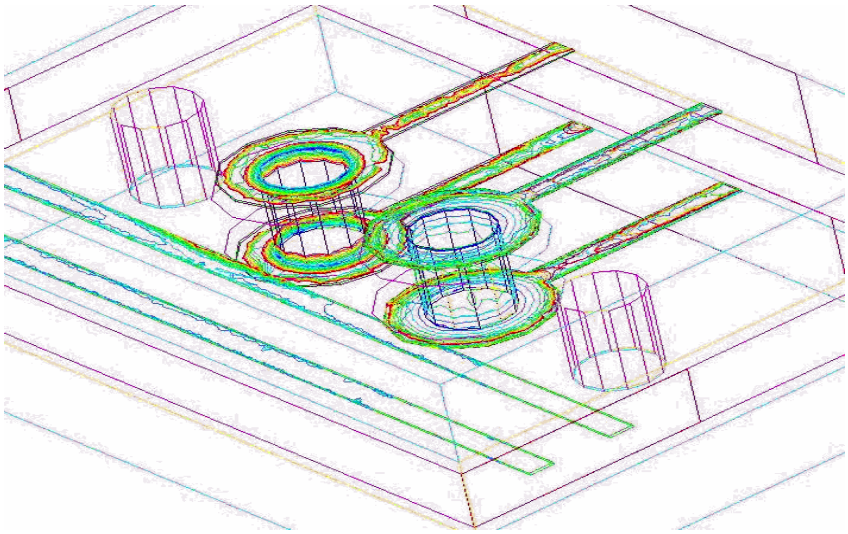


Figure 5.18: Differential driven vias adjacent to differential traces

In order to have an estimate of the amount of induced noise, we proposed to model in HFSS the connector pin-field section in two extreme situations (Figure 5.19 a. and b.). One case was placing the stripline trace unrealistically far-away (inside the constrained connector's routing channel) from a differential pair of vias (19.6mil), while the second situation considered the pair of traces situated as close as the PCB fabrication design rules allowed (5mil).

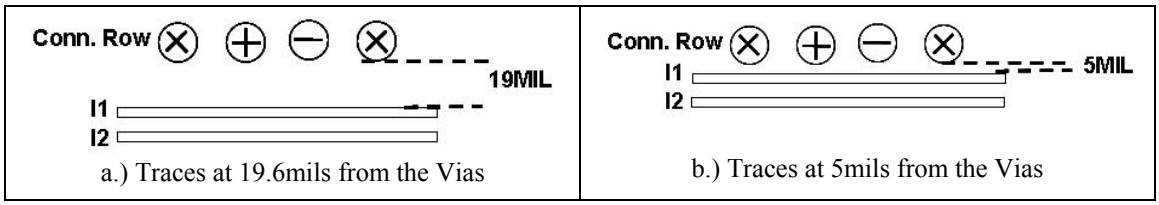
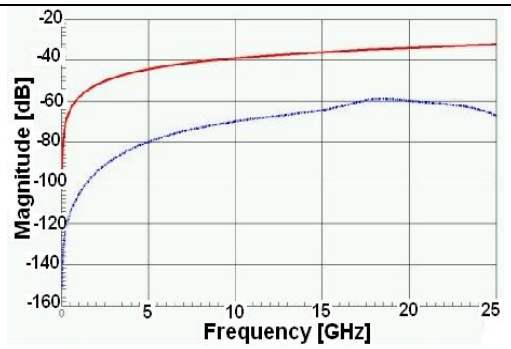
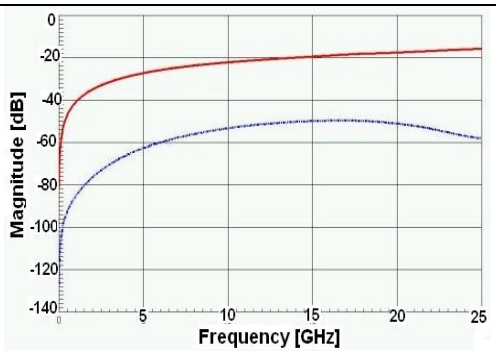


Figure 5.19: Differential traces placement relative to the connector pins

Both these extreme routing cases have been analyzed with HFSS and the Figures 5.20 a.) and b.) present the crosstalk in terms of insertion loss (S12) from the aggressors' vias to both, the nearest and furthest trace.



a.) Traces at 19.6mils from the Via (~ -40dB)



b.) Traces at 5mils from the Via (~ -20dB)

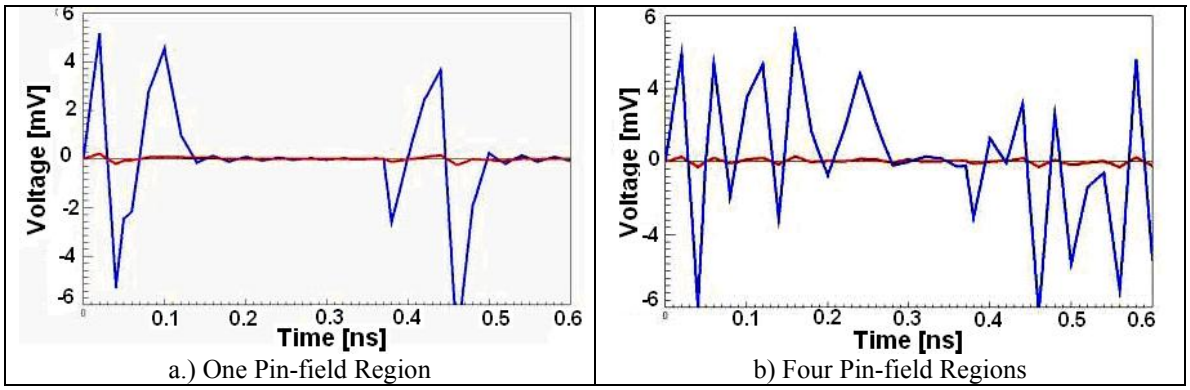
Figure 5.20: Crosstalk versus Frequency

Figure 5.20 a.) shows the results for the extreme far position of the traces to the via. At the frequency of interest for our signals there is only about -40dB of crosstalk even for the closest trace, so the differential noise is extremely low for this case. In Figure 5.20 b.) the results are shown for the extreme close spacing between the traces and the vias. The overall amount of crosstalk in this situation has risen by an order of magnitude from -40dB to -20dB, while the amount of difference between the crosstalk induced in the two traces remained the same as in the case a.). This observation constitutes a proof that the simulator was 'seeing' the correct effects.

The level of -20dB insertion loss (crosstalk effect) introduced in the closest trace when the differential pair is situated at a minimum distance of 5mil from the aggressors vias is considered acceptable. However, it remained the possibility that even such a small effect might accumulate with successive passes through several connector pin-field sections. To study this possibility, we exported the frequency-domain data (S-parameters) provided by HFSS, in the time-domain. A time-domain analysis performed by using a circuit simulator 'SPICE-like' allowed us to concatenate several connector pin-field models and such way to get a better estimate of the induced crosstalk magnitude over an entire switching plane.

We first performed the time-domain analysis by using Maxwell PSPICE. This is a basic SPICE simulator, incorporated in HFSS, which uses 'ideal' models for the representation of non-linear devices, such as transmitter and receiver. This is of course not the real case, but for the beginning was enough to give a first indication regarding the level of induced crosstalk.

An ideal signal pulse with the rise time of 50ps and the amplitude of 0.5V has been used for the excitation of the aggressors' vias in the connector pin-field. The Figures 5.21 show the amplitude of noise picked-up by the differential traces when they are passing through one a.), respectively up to four connector pin-field regions b.) on the backplane.



a.) One Pin-field Region b) Four Pin-field Regions

Figure 5.21: Time Domain Simulation in Maxwell PSpice

The maximum amplitude of pickup noise through one connector region a.) appeared to be about 1% of the aggressor value and was synchronous with the rising and falling edges of the aggressor pulse. The additions from the three other connector zones b.) does not greatly affect the amplitude (which only increases to approximately 1.2%), rather it just adds noise of the same magnitude but out of phase. The pickup on the furthest stripline is also shown in Figure 5.21 and it is practically zero. Hence all the accumulated noise will show up as differential crosstalk at the receiver, as concerned.

By simulating the concatenation of four pin-field section models in Maxwell PSPICE, we established an upper bound of any crosstalk effect that might occur within the primary switching plane on the backplane. However, for obtaining a more realistic estimate of the crosstalk effect accumulated within a full-path transmission channel, we had to analyse the behaviour of a realistic backplane environment and to use a more performant time-domain simulator.

5.4 Simulating a Realistic Backplane Environment

5.4.1 The Realistic Backplane Environment

Minimum components required for simulating a backplane based system are two semiconductors (the transmitter and the receiver), situated on different daughter-cards, two connectors and the PCB traces to link all together. Such a transmission channel over the backplane is shown below.

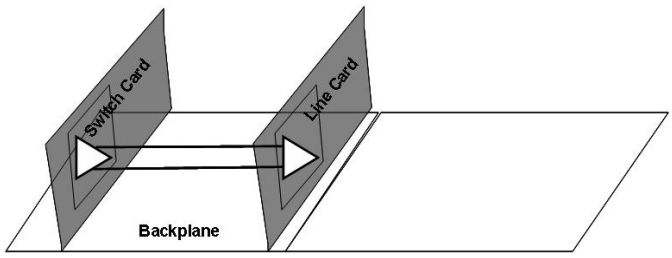


Figure 5.22: An Ideal Path in a Backplane Environment

A more realistic backplane system application requires including along the above transmission path the escape vias and pads corresponding to the connector launch structure at the daughter-card interface, and the models of the PCB traces passing through the connector pin-field regions on the backplane. Figure 5.23 shows one such transmission path.

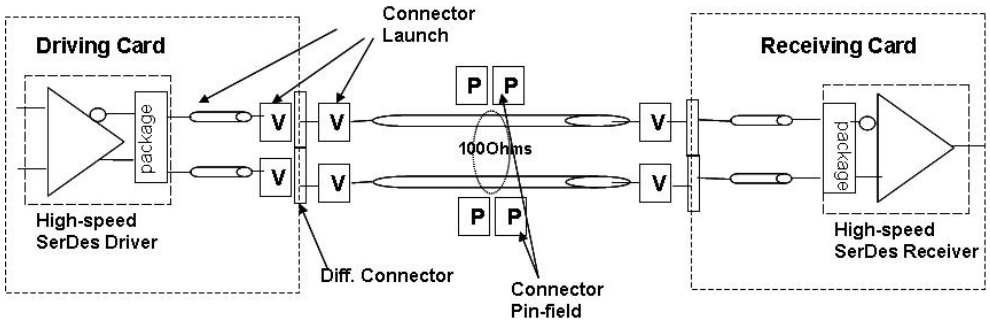


Figure 5.23: A Realistic Full-Path Backplane Environment

The full signal path then consists of the transmitter circuit on a daughter-board driving a short trace segment between it and the connector (and its launch structure) which leads it into the backplane. Then, on the backplane there are the traverses through one or more successive connector pin-field's sections, and finally the signal is routed off the backplane via the receiver daughter-card connector and passes through a short trace before arriving to the receiving chip. The full-path transmission channel in Figure 5.23 includes several types of models with different representations such as: behaviour models for the active devices, multi-coupled RLCG for the connectors, a W-element model extracted by HSPICE bi-dimensional field solver for the PCB traces and the full-wave model from HFSS for the connector launch and the pin-field's structures. We need to use a SPICE-like simulator whose engine would be able to treat and handle a time-domain analysis of such a complicated channel. Due to a reliable 2D field-solver and to the behaviour models of the active devices supplied by the manufacturer in HSPICE format, we choose the HSPICE simulator for doing this work. In addition, the simulation environment in HSPICE has been previously established and calibrated in chapter 4.

5.4.2 Full-path Transmission Channel on the Backplane

The signal transmitted across the backplane can pick up noise in the launch structure of the connector at the backplane-card interface, and on the backplane from the adjacent vias of the connector pins. To study the overall amount of the induced crosstalk effect, we chose a representative transmission channel shown in Figure 5.24 where we included the models of the connector launch and the connector pin-fields, previously obtained from the full-wave simulation in HFSS.

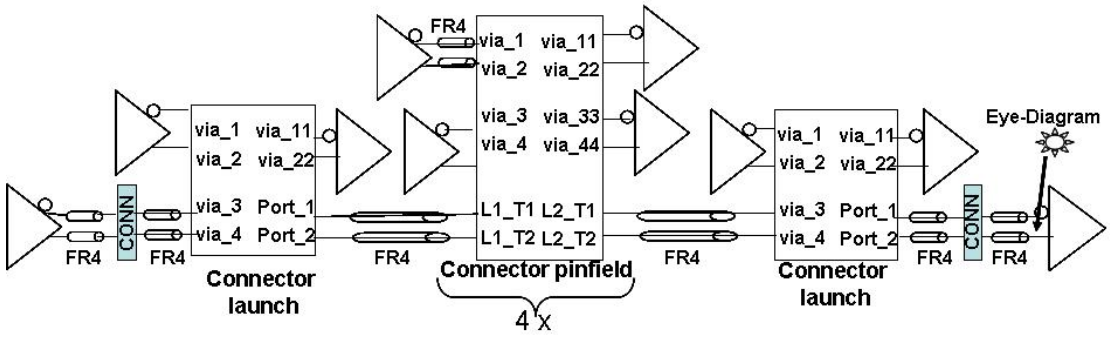


Figure 5.24: Transmission Channel for Studying the Crosstalk Effect

Starting at the transmitter, the signal traverses some inches of PCB trace on the transmit daughter-card before arriving at the connector that leads it to the backplane. The launch structure corresponding to this connector has been described using HFSS as a four port device. Two ports handle the ingress and egress of the signal passing through the traces and the two other ports carry signals that pass through adjacent vias. These vias are stimulated by other transmitters just as would be the case in a fully populated connector. The signal then traverses the pins of the connector corresponding to the Primary Switching Plane on the backplane. This region is represented by four adjacent connector pin-field sections. The model for a connector pin-field section is a six port device, which contains the ingress and egress ports for the signal traces in addition to four other ports representing stimulated via's on either side of the main signal traces. Finally, the signal is routed to the receivers chip through the daughter-board connector, its corresponding launch structure and other inches of trace on that board.

The transmission channel illustrated in Figure 5.24 is complicated and involves a large number of active devices used for stimulating all the differential pairs of connector pins (vias) along the path. The current practice for characterizing the transmission channel is to define patterns (see paragraph 3.3.3.5) as an equivalent for real data transmission. The active devices used in the backplane simulations are represented by Marvell Alaska 10Gbps transceiver. For analyzing the backplane environment within the transmission channel above, the aggressors connector pins (of the launch structure and connector pin-field geometry) were stimulated with 3.125Gbps PRBS pattern provided by the testing facilities incorporated inside the Alaska transmitters.

5.4.3 Study of the Crosstalk Effect

The time-domain simulations realized and presented in this paragraph are intended to study the induced crosstalk effect in the backplane environment of the 10Gbps Ethernet switch. The crosstalk analysis was therefore performed on the transmission channel illustrated in Figure 5.24 for different situations and scenarios. The induced crosstalk was measured at the receiver point of the differential pair using inactive probes.

5.4.3.1 Influence of the Routing Topology

The analysis was performed by considering the two extreme cases for the trace routing topology described in paragraph 5.3.4.2 (very close proximity, respectively very far away to the aggressing vias). The Figure 5.25 shows the differential crosstalk amount which has been introduced in the differential pair of traces for these two routing topologies.

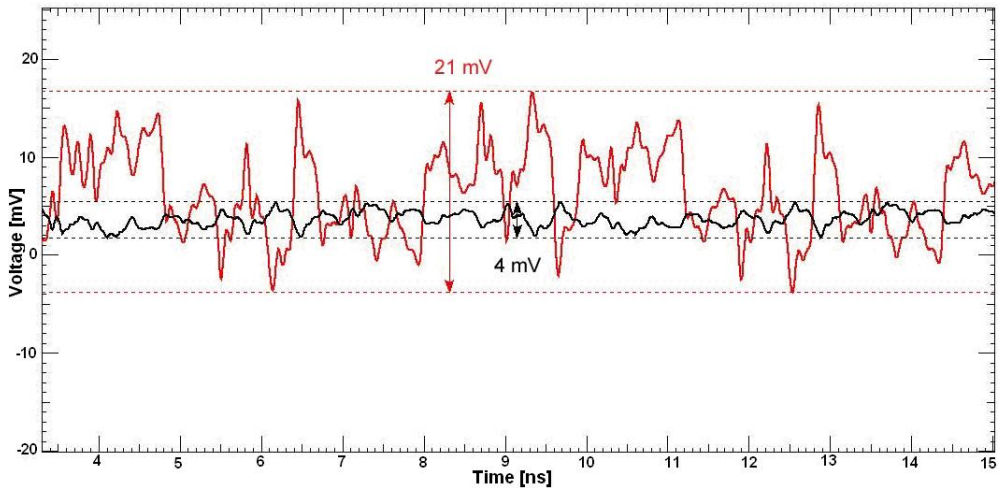


Figure 5.25: Differential Crosstalk after Traversing Four Backplane Connector Pin-fields

The level of crosstalk appeared to be less than 0.5% when the traces were located as far as 19.6mil from the aggressing vias and it increased to about 2% for the close routing topology case (5mil).

In the paragraph 5.1.2, the design rules for the 10Gbps backplane were setting the differential pair of traces at a distance of 7.2mil from the vias with the reserve of changing it to a more relaxed distance if the simulation would show that this creates any problems. We have seen first (in the paragraph 5.3.4.2) that the crosstalk in frequency-domain (in terms of insertion-loss) introduced in the closest trace is low enough not to constitute a concern. In this paragraph however, a better estimate of the crosstalk is given in the time-domain for an entire channel across the 10Gbps backplane and its level is much lower than 4%, defined as a maximum amount of acceptable crosstalk as defined in the XAUI interface specifications. In these conditions we can now conclude that the distance trace-to-via of 7.2mil within the constrained routing channel which is stipulated in the design rules specifications does not involve any crosstalk issue and it will be kept as it is.

5.4.3.2 Influence of the Aggressing Signals

On the same channel presented in Figure 5.24, we further studied an extreme situation which would set-up an upper limit for the amount of differential crosstalk possible to be induced from the connector pins in the adjacent traces. We considered the two different trace routing topologies as before, but this time we used an ‘ideal’ PRBS pattern running

at 10Gbps with a very fast rise time (5ps) to stimulate the aggressing vias. Figure 5.26 shows the amount of crosstalk induced in the traces for these situations.

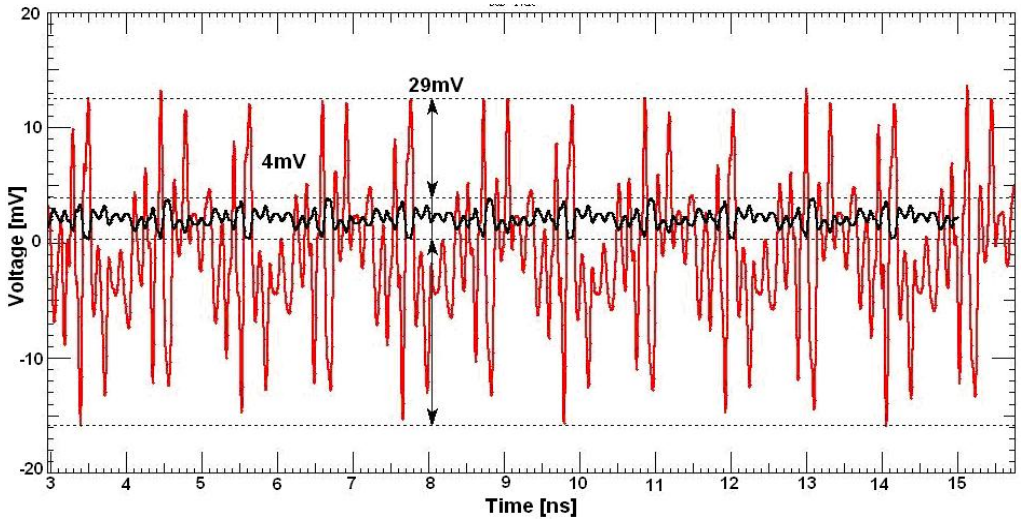


Figure 5.26: Differential Crosstalk after Traversing Four Backplane Connector Pin-fields

The amount of crosstalk induced in the far-away routed traces appeared not to be higher than in the previous case (Figure 5.25). For the case of close routed traces, the crosstalk increased however from less than 2% to almost 3%. Even though, this was not really such a significant increase as we suspected, but it showed that the simulator was seeing a certain increase of the crosstalk effect determined by the routing topology and the increased frequency of the aggressors signals.

5.4.3.3 Conclusions on the Study of the Crosstalk Effects

Our studies confirmed that there was certainly a measurable effect causing the injection of differential crosstalk from adjacent aggressing vias inside the PCB traces passing through the connector pins. We performed the analysis for two cases which constitute the worst possible scenarios in terms of routing topology and signal aggressing bit rate. Even though, for the most constrained channel (very close routing topology and 10Gbps aggressing signals data rates) the maximum amount of induced noise was less than 3%. The design rules we provided suggested a more 'relaxed' routing distance for the traces passing between the connector pins than the worst-case simulated scenario. It appeared than unlikely that the amount of crosstalk introduced in the proposed design for the 10Gbps backplane would exceed the acceptable level imposed by the XAUI interface specifications.

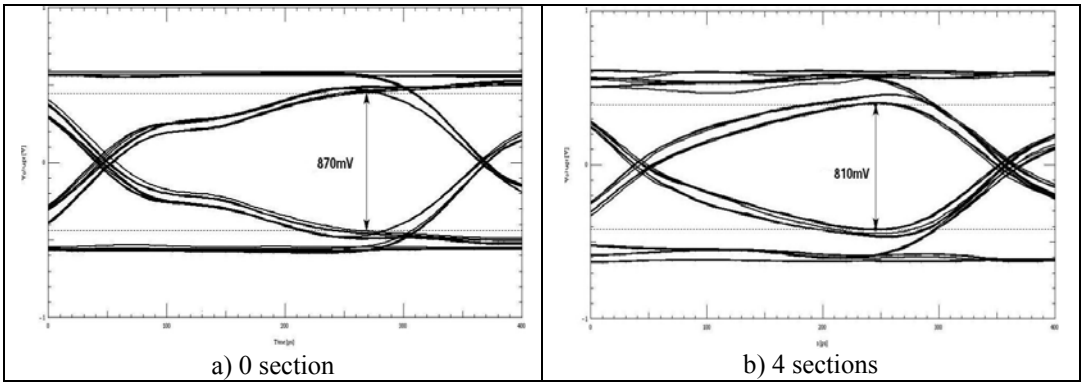
5.4.4 Simulating the 10 Gbps Transmission Channel

For the purposes of analysing the overall performance of the full-path transmission channel corresponding to the 10Gbps Ethernet switching backplane, we simulated the transmission channel in Figure 5.24 with different parameters. The simulated waveform

at the receiver is plotted as an eye-diagram, whose shape and dimensions will give information about the quality of the signal transmission across the channel.

5.4.4.1 Influence of the Connector's Pin-fields Area

The following analysis was performed for a full-path channel of 14 inches of PCB differential trace which passes through two connectors. Two situations were considered over the backplane: no connector pin-field (Figure 5.27 a.), respectively up to four sections of connector pin-fields (b). When included, the model of the pin-field was corresponding to the close routed trace topology in the connector. The effect at the receiver is shown in Figure 5.27 for both cases.



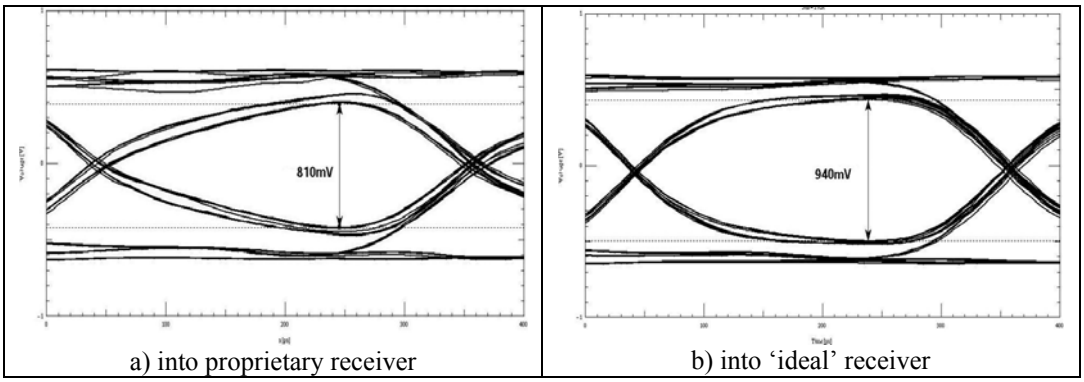
a) 0 section b) 4 sections

Figure 5.27: Influence of several connector pin-field sections on the backplane

The main observable effect of the pin-fields presence is that the added capacitance has reduced the high frequency components of the received signal giving it a reduced rise and fall time and thus reducing the eye opening from 870mV without pin-fields down to 810mV for the case when four extra pin-field models were added on the backplane. This corresponded to a loss of 6.8%.

5.4.4.2 Influence of the Marvell Receiver and its Package

A percentage of the loss we previously observed in the eye opening was due to the limiting effect of the driver and receiver's packages of the Alaska 88X2040 device. To demonstrate this affirmation, we used the same channel as in Figure 5.24, but we replaces the Marvell's proprietary receiver with an 'ideal' one which consists only of a resistive termination and a voltage probe. In Figure 5.28 we plotted the eye when using as receiver the Marvell chip model a.), respectively the eye obtained by using an ideal receiver model b.).



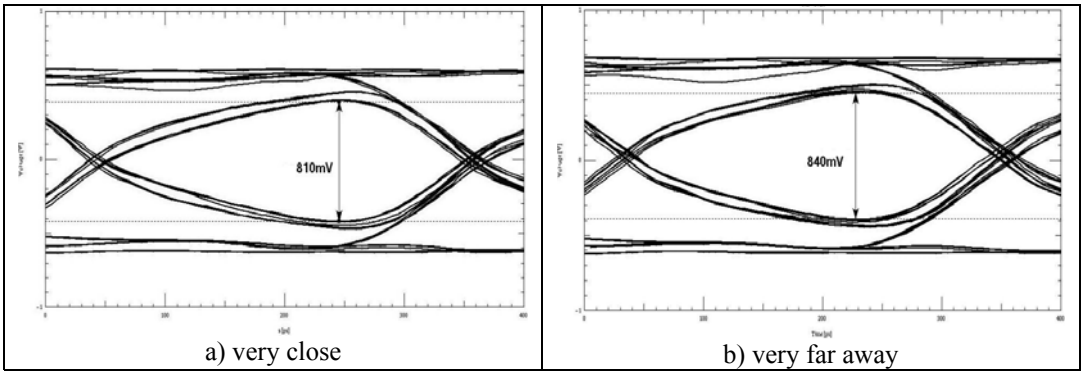
a) into proprietary receiver b) into 'ideal' receiver

Figure 5.28: Signal Reception

The difference between the two receiver models was basically only in the parasitic capacitance and inductance that derives from the mechanics of the chip packaging, fixed by the silicon fabrication process. The analysis showed that the packaging effects account for a loss of nearly 14% in the available eye opening for the present simulated channel.

5.4.4.3 Influence of the Routing Topology

The crosstalk effect induced by the routing topology within a channel comprising two connectors and four pin-fields along with 14 inches of differential trace, was presented in Figure 5.25. We simulated the same channel and analysed the effect encountered on the eye-diagram when signal traces were routed very close, respectively very far from the aggressing vias. The results are shown in Figure 5.29 a.) and b.).



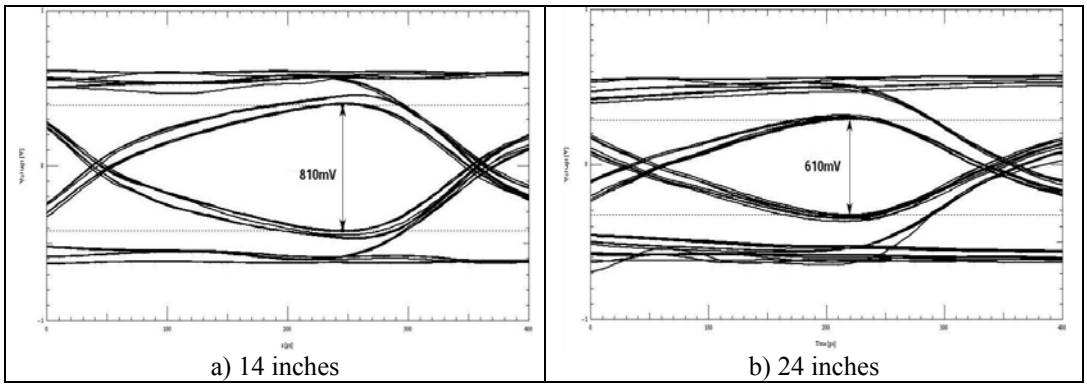
a) very close b) very far away

Figure 5.29: Eye Opening for the Proprietary Aggressors and Victim (3.125Gbps)

The shape of the eye resulted almost the same for the two cases simulated above. However, a difference of about 30mV was measured between the two eye openings. This corresponded to an eye closing of approximatively 3.5%.

5.4.4.4 Influence of the PCB Trace Lengths

For the purpose of studying the traces lengths influence on the overall performance of the transmission channel, we considered therefore that the daughter-cards were separated by two different lengths of traces passing across the backplane. The PCB trace configuration and dimensions as well as the FR4 dielectric characteristics are as discussed in the paragraph 5.1.2. The results for 14 inches and 24 inches trace lengths are shown in Figure 5.30.



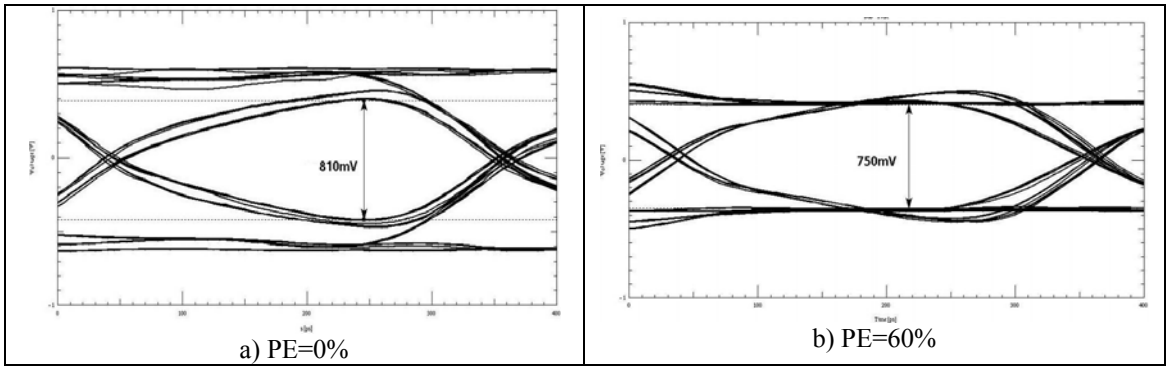
a) 14 inches b) 24 inches

Figure 5.30: Eye Opening

At 3.125Gbps, the most detrimental effects are due to the high-frequency losses in FR4 dielectric materials, which are directly increasing with the length of the transmission medium. By doubling the length of the PCB trace across the backplane, we could indeed notice an important decrease of 200mV in the eye opening, corresponding to a loss of about 25%. However, this loss was encountered for 24 inches of trace already which sets up the maximum distance to the furthest daughter-card in the chassis.

5.4.4.5 The Influence of Pre-Emphasis

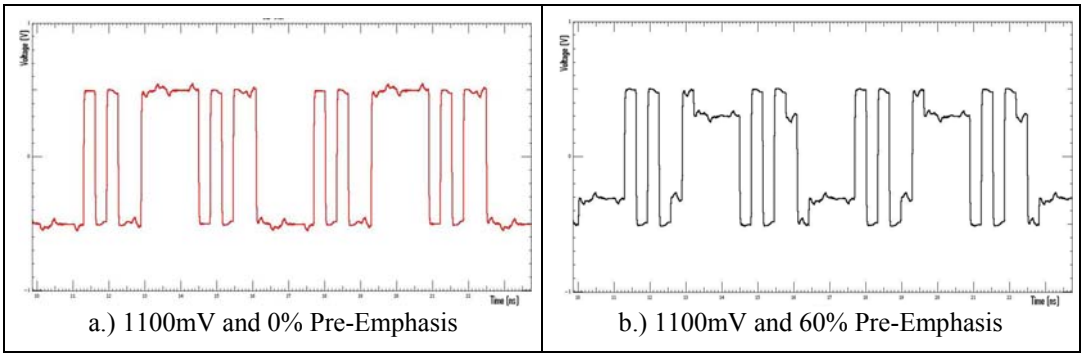
We did further analysis for studying the effects of pre-emphasis on the overall performances of the signal transmission over the 10Gbps backplane. It is well known that pre-emphasis is used for improving the quality of the transmission, and therefore we compared the eye-diagrams at the receiver end in the channel in Figure 5.24 for 0, respectively 60% of pre-emphasis. The results are plotted in Figure 5.31.



a) PE=0% b) PE=60%

Figure 5.31: Eye Opening

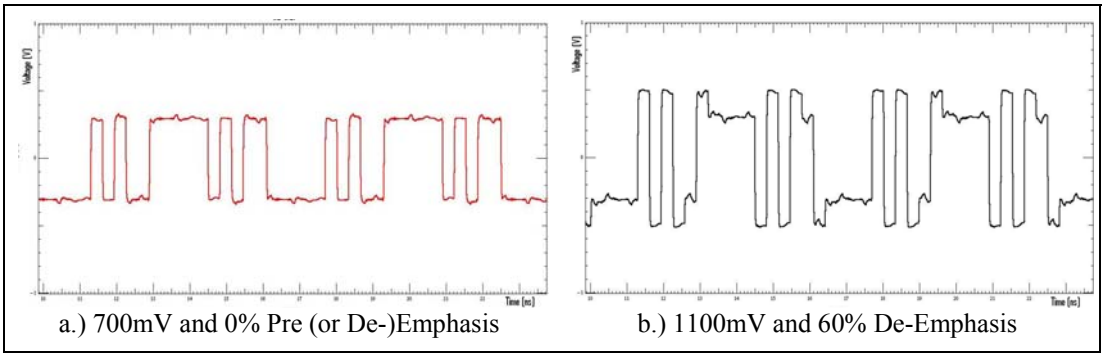
The first thing we remarked was that the increase of the pre-emphasis is associated with a decrease of the eye measured at the receiver. This is the complete opposite to the expected effect and needed a fuller understanding. We consider first the voltage levels output from the Marvel transmitters with and without pre-emphasis as shown in Figure 5.32.



a.) 1100mV and 0% Pre-Emphasis b.) 1100mV and 60% Pre-Emphasis

Figure 5.32: Waveform at the transmitter

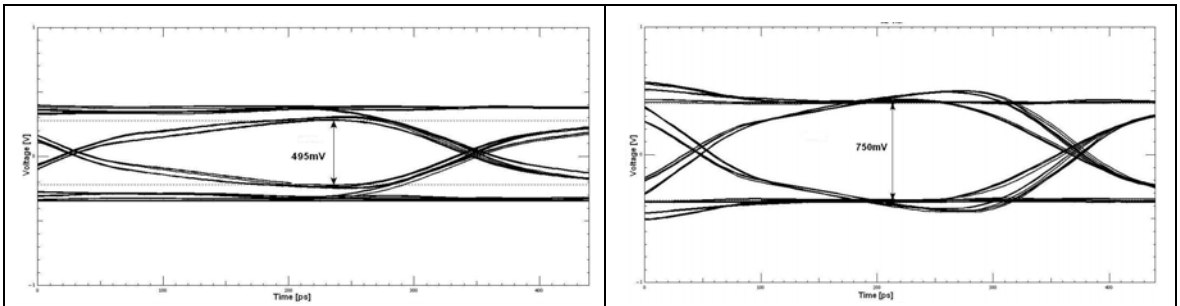
It is clear that adding pre-emphasis to the Marvel model maintains the maximum amplitude of the signal at the pre-set voltage, in this case 1100mV, but reduces the amplitude of the bits subsequent to the transition bit. We concluded that what it was defined as pre-emphasis in the Marvell Alaska 88X2040 datasheet was in practice, a de-emphasis effect. This means that to properly compare the value of an eye diagram with or without pre (actually more correctly being said de) -emphasis we have to change two parameters: the signal amplitude and what is now understood to be the de-emphasis. An example is shown below in Figure 5.33. In Figure 5.33 a.) we observe a non-emphasised signal having an amplitude of 700mV. In Figure 5.33 b.) we defined a signal of 1100mV and a de-emphasis level of 60%. 60% of 1100 is 660mv which is less than 6% of 700mv.



a.) 700mV and 0% Pre (or De-)Emphasis b.) 1100mV and 60% De-Emphasis

Figure 5.33: De-emphasis Effect on the Transmitted Signal

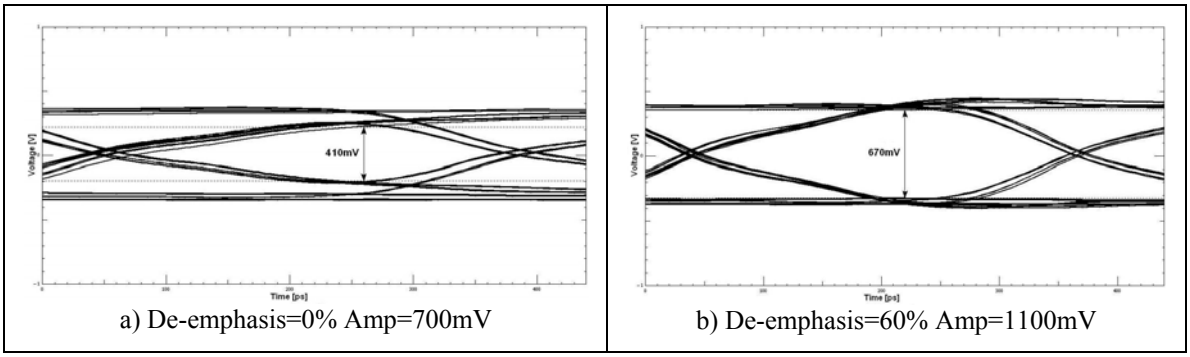
Only having a few discrete voltages and levels of de-emphasis to choose from in the settings of Alaska 88X2040 device, means that we cannot reach a combination having a perfect match, so comparisons have to be made bearing this in mind. In this case illustrated in Figure 5.34, adding 60% of de-emphasis made an improvement of over 50% in the eye-opening of the received signal across 14 inches of PCB trace.



a) PE=0% Amp=700mV b) PE=60% Amp=1100mV

Figure 5.34: Eye Openings for 14in. of FR4 trace

We now consider the same signal options across the longest distance on the backplane. Figure 5.35 plots the correct signal comparison, when 0%, respectively 60% of de-emphasis was added on the signal transmitted across 24 inches of PCB trace.



a) De-emphasis=0% Amp=700mV b) De-emphasis=60% Amp=1100mV

Figure 5.35: Eye Openings for 24in. of FR4 trace

The improvement achieved in this case corresponds to 63.4%.

5.4.4.6 Conclusions on 10Gbps backplane simulation

A full-path transmission channel was analysed from different perspectives, such as the influence of the routing topology of the traces on the backplane, of different number of the connector pin-fields area, of the trace lengths and the pre/de-emphasis. On one side, we analysed the influence of the routing topology by varying to the extreme limits the PCB trace placement within the connector routing channel and by increasing to a factor of three the bit rate of the aggressing signals. On the other side, we considered the influence introduced by a different number of connector pin-field sections along with varying the PCB trace length and the de-emphasis level. Resulting eyes had higher or smaller openings, depending on the level of induced crosstalk or on the high-frequency losses encountered.

The XAUI specifications, which are themselves highly conservative, call for a stated eye opening of 400mV. Even for the most constrained case we analysed with no de-emphasis effect added on the transmitted signal, the remaining eye-opening exceed that value for over 50% of the bit period. We could then conclude that the eye-opening in our backplane based system which has a more ‘relaxed’ routing than the simulated case,

would have more than enough margin to sustain error free operation at the design speed of 3.125 Gbps.

5.5 Limits to Performance

5.5.1 Introduction

The previous study on the performances of the 10Gbps channel confirmed the possibility of building the 10Gbps Ethernet switching backplane with the current technology for the silicon and PCB manufacturing. It was reasonable then to ask where the limits are for switching Ethernet and how will the next generation evolve. Before the falling market in the telecom industry at the beginning of 2002 the view was that, perhaps, the next step would be the usual factor of 10 up to 100Gbps, whereas other voices called for a factor four to keep track with Sonet at 40Gbps. In the light of past developments 40Gbps looks like the more credible of the two options, so the question should be asked, how feasible is this? The implicit assumption in 40Gbps switching was that the current switch architectures and the chip packaging remained the same while the switching capacity was quadrupling by relying on Moore's law. Moore's law doesn't work for PCB traces however and therefore we further consider this issue in terms of the limitations caused by signal integrity and packaging [ESTA-R2].

5.5.2 Current Channel Performance

Starting from the base line established in paragraph 5.4 and using the same Marvell Alaska's driver/receiver models as before, we simulated a simple channel circuit driving the current switch architecture. The simulation took into account the traverse of a transmission channel as described and illustrated in the Figure 5.24. The simulations have been performed in the presence of noise induced in the realistic backplane environment from adjacent vias in the connectors.

We ran two sets of successive simulations with PRBS patterns at increasing bit rates and we observed the evolution of the eye opening. For the first set of simulations we were using the Marvell's proprietary driver/receiver models while the second set involved the use of 'ideal' drivers and receivers. We ran both those simulations up to the point where the eye had been reduced by a factor two, 10 Gigabit Ethernet standard. The obtained results are shown in Figure 5.36.

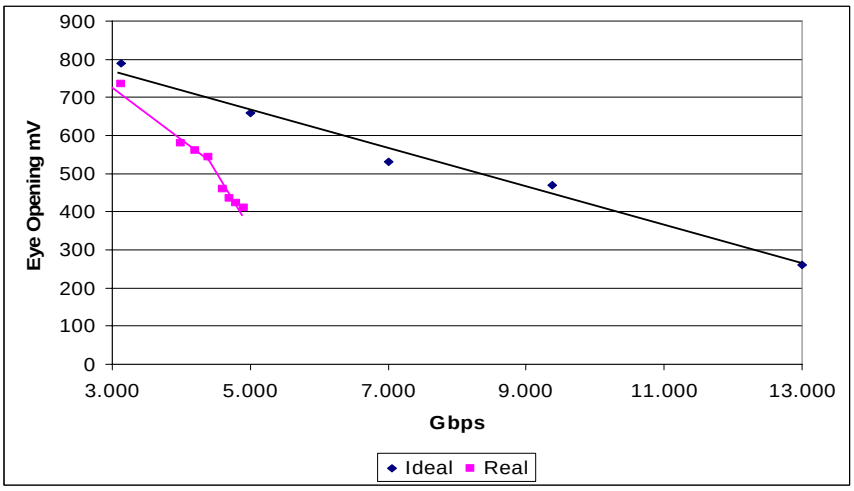


Figure 5.36: Reducing Eye Opening with Frequency

From this study it resulted that the Marvell's commercial model for the driver/receiver would be reaching its practical limit at about 5Gbps, whereas the channel limit itself was over 10Gbps. An eye opening of minimum 400mV could be achieved for the current channel up to 10Gbps, while the channel itself has an upper bound of 15Gbps for this length of trace. At 15Gbps the eye was completely closed.

Where are the limitations? Even with low-cost FR4 dielectric the commercial driver and receiver was far from exploiting the available channel bandwidth. The Marvell's driver performs very well and the use of pre/de-emphasis goes a long way to reducing the effect of FR4 losses on the longer PCB traces. However, the slew rate of the transmitted signal is bound by the package parasitics in terms of internal inductance between the chips power lines and the driver pins. For the output current on the drivers package pin to change rapidly, the internal inductance must be as small as possible. The receiver chip is also limiting and this due mostly to the input capacitance of the receiver circuitry. The limitations appear to be due to the geometry of the chip package.

Our conclusion following the present study was that further gains in the 10Gbps transmission can be achieved by using better chip packages available. However, the edge in performance that they offer is not required for achieving the current bit rate and allowing a successful transmission over the 10Gbps backplane of the Ethernet switch.

5.5.3 Future Channel Performance

In obtaining the progress to a factor of four in transmitting signals over copper, there are currently two approaches due to the usual competitive claim between the two interested groups. One approach is on improving the chips technology and the packaging, while a second approach addresses more performant encoding schemes for the signaling. In addition to these, the PCB technologies have a great impact on the transmission final performance.

5.5.3.1 Silicon

On the one hand the elusive factor four can be achieved by further refining the silicon process, paying close attention to chip packaging and, a judicious choice of transmitted signal swing and receiver sensitivity. Better performance of the silicon can be achieved also by improving the packaging, by making the die and the corresponding package parasitics smaller. In addition, further gains can be obtained by reducing the signal level swing between the 'zero' and 'one' state, which would mean that the rise time from one state to the next is faster and so higher bit rates can be supported. This also means however that there is less margin against noise and therefore careful engineering of the transmission channel is once again required to minimize any possible undesirable signal integrity effects.

5.5.3.2 Multi-level Signaling

The second approach in achieving higher rates is to use multi-level signaling solutions such as Pulse Amplitude Modulation (PAM). PAM-4 for example, divides the dynamic range into four levels such that the resulting vertical eye is only one-third that of a normal NRZ signal. With four levels two bits per bit period can be then transmitted. The resulting signaling rate is half the bit rate. This means that a 10Gbps transmission will be seen on the line as 5Gbps from the point of view of driver, receiver and PCB loss considerations. This is extremely important as lowering the signaling rate and the high-frequency components, more likely it is that the interconnect will meet the signal integrity requirements.

It is true that the PAM encoded signal occupies a lower part of the bandwidth spectrum and hence has less loss to deal with from any given transmission channel. However this has been achieved at the cost of reducing the eye height since the voltage on the driver chip remains the same and the signal is now divided into three eyes instead of the one for the classical NRZ encoding. This is not particularly a problem for a cable driver where crosstalk is very limited, however in the context of a backplane it can become an issue because of possible higher levels of the induced crosstalk noise.

5.5.3.3 PCB Technologies

The issue of backplane loss can be greatly improved on as well. There has been very little impetus to do this so far on a wide scale because the driver technology, in particular here the pre/de-emphasis, removed the need. Scaling up speed by a factor of four however brings the issue back into focus.

FR4 is one of the lossier materials available to build PCBs and is employed simply because it is cheap. There are much more performant materials available as can be seen in the following Figure 5.37 (reproduced from [REF-loss]).

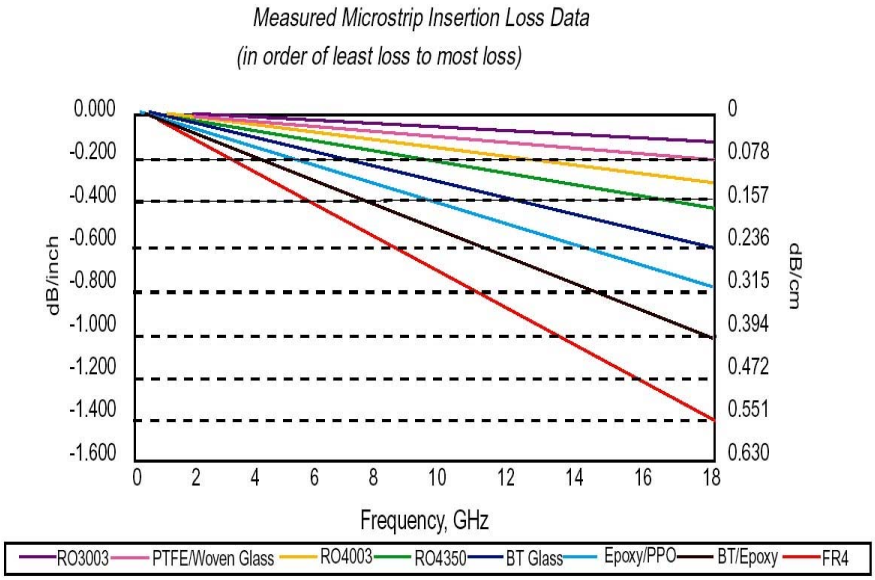


Figure 5.37: Relative loss performance of available PCB dielectrics

Clearly high performance can be achieved by using lower loss materials. Almost an order of magnitude improvement in loss can be obtained, but the issue is one of cost. The following Figure 5.38 gives relative prices for different dielectrics [REF-diel].

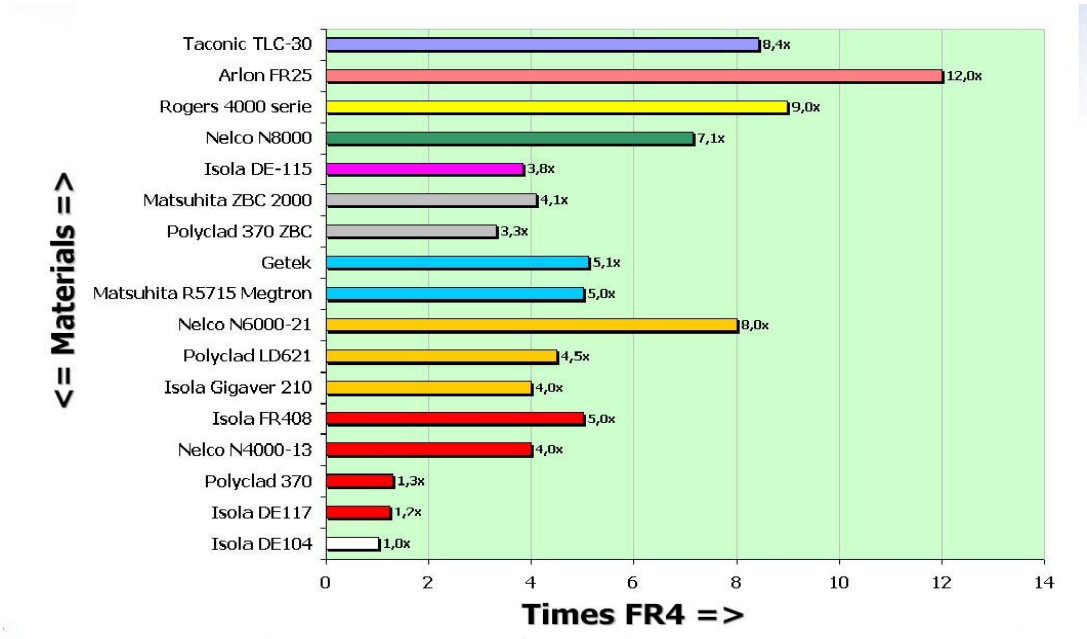


Figure 5.38: Relative dielectric costs

Unsurprisingly the factor ten in performance comes with a factor 10 increase in cost. It is not as bad as it seems however because current practice is to employ the high performance laminates only around the signal layers, while several power distribution

layers are still constructed in traditional FR4. The increase in cost for the higher performance is more likely to be therefore a factor four or five rather than ten.

5.5.4 Performance Conclusions

At the time of the present study, by the end of 2002, the current line driving technologies for copper were operating up to 6Gbps point using two state differential NRZ encoded signaling. With improvements in chip design and more importantly chip packaging, there seems to be no reason not to expect a further factor four in performance. By changing the encoding scheme to a multi-level signaling scheme a factor of two has already been achieved. Crosstalk noise is still an issue but the material technologies to limit it are well known and available at a price.

We did not see any impediment regarding the fundamental electric limits for achieving higher speeds, rather than the cost. From this perspective, there is no technical reason why the interconnect architecture of current 10Gbps switches that use four channels of 3.125Gbps cannot simply be scaled up by a factor four to provide the bandwidth needed for a 40Gbps switch.

5.6 Summary

In this chapter we have presented a proposal for the design and the PCB stack-up for the backplane of the 10Gbps Ethernet switch. This proposal has been originally worked out by taking into consideration the routing constraints on the backplane within the connector pin-field areas corresponding to the primary or secondary switching plane. In order to validate it however, we have established a methodology, hereby presented, which allowed us through extensive modeling and simulation, to perform a pre-layout validation study of different possible approaches in achieving 10Gbps transmission across the Ethernet switching backplane.

We have first full-wave modelled the connector pin-fields in HFSS field-solver, after having a-priori calibrated the modeling environment. To study the amount of induced crosstalk and the overall performance of the transmission channel across the backplane afterwards, we had to resort to circuit/system simulation in HSPICE. We have analysed in HSPICE ones of the worse possible design scenarios, in order to setup a maximum limit for the resulted effects.

Through all the extensive analysis we performed and presented in the current chapter, we established that the influence of the critical regions of the connector pin-field regions on the backplane does not constitute a limiting factor for the 10Gbps transmission over the backplane and that our backplane proposal was well within current capabilities.

6 Advanced Telecom Computing Architecture

6.1 AdvancedTCA Compatibility Study

6.1.1 AdvancedTCA

The implementation of the 10Gbps Ethernet switch fabric was based on a passive backplane into which multiple line cards and two redundant switch fabric cards were plugged. Through modeling and simulation study, presented in chapter 5, we defined and confirmed the design rules to follow for the 10Gbps backplane. At this point it was the responsibility of the manufacturer to build the backplane board according to the proposed design solution. However, not long before finishing the design and starting the production of this backplane, the PCI Industrial Computer Manufacturers Group (PICMG) released its new standard, for the Advanced Telecom Computing Architecture (AdvancedTCA or ATCA) ([ATCA], PICMG-02). This standard addressed the needs of the telecom market which had shown little or no interest in PICMG's previous standard for CompactPCI. AdvancedTCA was targeting the modular approach, enabling the use of the same AdvancedTCA chassis/backplane with different modules, for multiple types of telecommunication equipment [OLT-UPB]. It addressed the following requirements:

- Scalable Capacity of up to 2.5Tbps (per chassis);
- Modularity and configurability to enable multiple modules with various interfaces, different technologies and storage media to be mixed and matched for diverse applications in the same platform;
- Redundancy;
- Advanced power distribution and cooling concepts.

The base specifications for the new AdvancedTCA family were included in PICMG 3.0 and adopted in January 2003. AdvancedTCA specifications incorporate the latest trends in high-speed interconnect technologies, next generation processors and improved reliability, by providing specifications for electromechanical issues (rack and chassis form factors), interconnect topology and electrical characteristics (backplane connectivity, power sources, cooling systems, management interfaces) for modular system chassis. The AdvancedTCA standard defines the passive backplane, power sources, the chassis and also provides a set of necessary guidelines to assist with the development of the boards which will be accommodated within the chassis slots. In addition, the standard recognized that there was a strong customer proprietary interest and can support different link technologies such as, for example, PCI Express, Infiniband, or 10Gbps Ethernet.

The key feature of this standard is its very high bandwidth application and its big advantages are the rapid development time, scalability and the ability to mix different types of modules and technologies in one single chassis. Due to all these benefits, the AdvancedTCA standard was rapidly adopted by several telecommunication equipment manufacturers as the backplane solution of choice for their modular chassis systems.

Although we had satisfied ourselves with extensive simulation that our own backplane design met the requirements with a sufficient margin of error, there are definite commercial risks in developing a proprietary solution in cases where a standard exists. We had therefore to assure ourselves that the new standard covered all our own concerns and was indeed appropriate for the evolving 10Gbps Ethernet switch design. If so the next step would be to adapt the system design to take advantage of the emerging AdvancedTCA architecture.

6.1.2 Comparison Study

6.1.2.1 The Study

We studied the nominal specification of the AdvancedTCA standard and compared them with the guidelines we had already established.

One first advantage was given by the definition in the standard of the same dual-star topology as chosen for the 10Gbps Ethernet switch. The standard specifications allowed for the use of either FR4 or Nelco dielectric materials for building the backplane board stack-up. However, it was highly recommended to use FR4 and it was left up to the manufacturers to produce the required 100Ω differential impedance for the PCB traces. The stack-up proposed by the standard for FR4 with a relative permittivity of εr = 4.0 is shown in Figure 6.1.

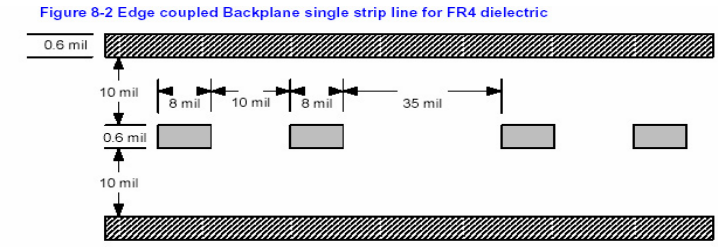


Figure 6.1: AdvancedTCA Recommended Stripline Dimensions

To compare the impedance target with our own estimations, we used the same stack-up calculator (SpecctraQuest) and we obtained:

- AdvancedTCA specification: (FR4 with εr =4.0, lt=0.018)
  - Z_{diff} = 97.3Ω and Z_0 = 53.3Ω
- Our specifications (FR4 with εr =4.3, lt=0.022)
  - Z_{diff} = 100.6Ω and Z_0 = 51.6Ω

In practice either of these two proposals still meets the specification which states that the differential impedance of the backplane and board serial links shall be 100Ω ± 10%.

The signal connector of choice for the AdvancedTCA is the Tyco ZPACK HM-Zd or Erni ERmet ZD [ERNI-03], which was also our choice. One area of concern here was the new anti-pad specification, which was more generously dimensioned (63mil) than the connector specs originally called for (50 mil diameter). The direct effect of this was on

reducing the routing channel from 48mil in our design to 34mil in theirs. By increasing the anti-pad clearance, there was a corresponding reduction of via to ground or power plane capacitance and therefore less signal loss across the via. Figure 6.2 shows the resulting difference in stripline configuration between the AdvancedTCA specification (A) and the less constrained model (B) that we had developed.

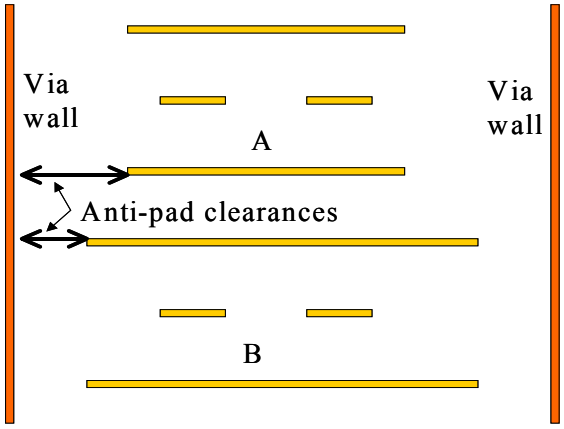


Figure 6.2: Comparison of stripline geometries

The drawing is to scale and shows that the reduction in via-to-plane proximity has come at the cost of reduced ground plane coverage for the return current. This will affect the impedance of the traces as they pass through the connector pin-field. It must be assumed that the AdvancedTCA simulations took this into account and that the trade-off was acceptable.

The manufacturer (ASIS) which build the 10Gbps backplane following the specs of AdvancedTCA standard, cross-checked the simulations performed by the PICMG 3.0 team by their own simulation and their measurements of the final backplane board [ASIS]. Their results, proven by simulations and real data measurements, were showing that even the worse case trace met all the PICMG 3.0 requirements. The discontinuity provided by the ZD connector was minimal, the impedance match was within limits and the eye-diagram at the receiver was open.

6.1.2.2 Connectivity

The main drawback of the AdvancedTCA solution compared with our design requirements is that there are only eight channels between the Node boards and switching fabric Hubs. This is only sufficient for four links, but we need actually six links in each direction (as specified in the paragraph 5.1.1) to deliver the required bandwidth. In this way we would achieve a raw and effective throughput on the AdvancedTCA backplane of respectively 18.75Gbps and 15Gbps.

The fact that the AdvancedTCA standard does not include this extra requirement for the backplane is not such a drawback as might first appear, because the architectural description of the same standard provides a user definable zone 3 which is intended for rear I/O access (see section 6.2.1). Nothing prevents using this area to mount a copy of the main data fabric design for the required extra two channels in each direction with a

board built to exactly the same standards. This does not detract from the advantages of using standard parts while still meeting the bandwidth requirements.

6.1.2.3 Conclusions on the Compatibility Study

The differences between our model and that adopted by the AdvancedTCA group were so small as to make building a proprietary backplane a poor economic choice. There are also clear advantages in following a standard, notably the availability of components at lower cost from third party vendors. This includes mechanics, power supplies, cooling, and the backplane itself. Then there is the potential availability and interchangeability of more complex units such as the management processors. The risks of being an early adopter of a standard are also clear. If the projected market fails to materialise then the cost benefits and easy availability of components will also disappear. The benefits outweigh the risks and the decision within the ESTA project was taken to adopt the standard.

6.2 AdvancedTCA Backplane Architecture

6.2.1 AdvancedTCA Architecture

The AdvancedTCA backplane can contain up to the maximum of 16 slots. The 10Gbps backplane provided by ASIS to us however, has only 14 slots and therefore the explanations will be made for this particular case.

According to the AdvancedTCA standard specifications a slot may have different physical slot and logical slot number designations. The logical slot numbers are used to determine channel mapping between slots, while the physical slot defines the real position within the chassis. A mapping of physical slot designations to logical slot designations for every slot in the AdvancedTCA backplane is shown in Table 6.1.

Logical	1	2	3	4	5	6	7	8	9	10	11	12
Physical	15	13	11	9	7	5	3	1	2	4	6	8

Table 6.1: Logical to Physical slot mapping for the AdvancedTCA backplane

Every slot of the AdvancedTCA backplane is populated with connectors, which are disposed over three zones. Figure 6.3 taken from the standard specifications illustrates these three connector zones on the AdvancedTCA backplane and on the matching board to be plugged in.

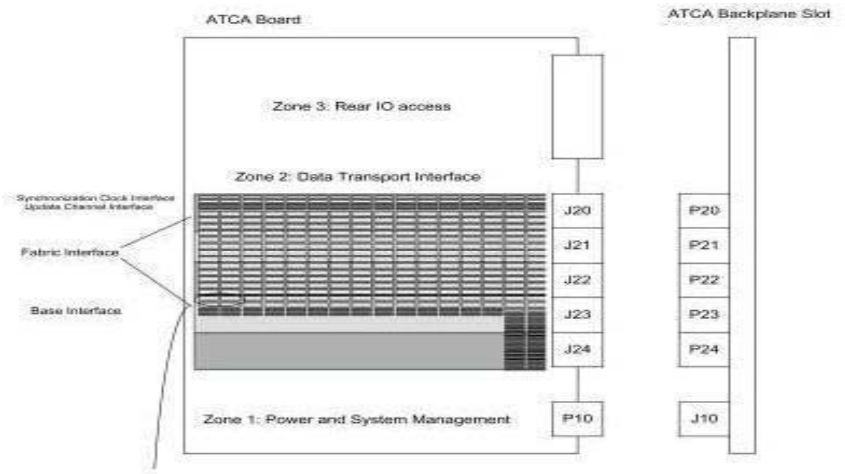


Figure 6.3: AdvancedTCA Backplane and Board Interface

Zone 1 is for power and system management. Power is delivered through a dual redundant path fed at 48V through connectors [CON-01] in Zone 1. A complete AdvancedTCA chassis may consume up to 3.2kW. Zone 3 is for rear I/O access. Zone 2 is the primary data transport interface which contains five high-speed connectors for four separate interfaces: base interface, fabric interface, update channel interface, and synchronization clock interface. All of the high-speed signals of interest for testing the backplane are accessed through the fabric interface connectors (Erni ERmet ZD or Tyco HM-Zd) in Zone 2.

6.2.2 Fabric Interface

Within the specifications of the standard, of special interest for us was the interconnect architecture achieved by using point-to-point high-speed serial links. Backplanes were available with a choice of two topologies, either dual star fabrics or an all-to-all full.

We were interested in the dual-star fabric topology across the backplane, where two dedicated slots of the AdvancedTCA backplane are used for the switching fabric cards (defined as Hub1 and Hub2) in logical slots1 and 2 and the other twelve logical slots in the AdvancedTCA chassis (slots3 to 14) are dedicated to the Line Cards (defined as Node boards). The redundant connection for the dual-star architecture between each Node and both Hub boards is achieved via two Fabric Channels. Figure 6.4 illustrates by arrows the available Fabric Channels of the dual-star fabric interface for an AdvancedTCA chassis implementing twelve Node boards and two redundant Hubs.

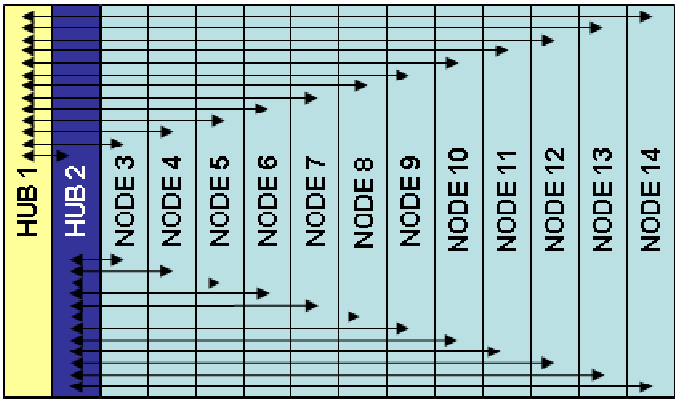


Figure 6.4: Fabric interface Dual Star topology

Each Hub board will have twelve Fabric Channels for the connection to all the Node boards plus one channel to connect Hub boards together. As a consequence, on the whole AdvancedTCA backplane above, there is a total of 25 Fabric Channels which assure the dual-star communication in the chassis.

The standard data fabric supplies 8 pairs per fabric channel and as described in 6.1.2.2 we provide another four pairs using a supplementary data fabric backplane mounted in Zone 3. Thus there are a total of 300 differential pairs (25 Fabric Channels x 12 differential pairs/channel) routed across the switch backplane, each operating at 3.125Gbps.

The PICMG® 3.0 specifications define the physical implementation of each Fabric Channel in terms of slot and pin assignment on backplane connectors. In the Hub boards, the dual-star Fabric Interface uses therefore a total of four connectors (P20 to P23) to cover the standard fabric connectivity towards all the Node board's slots in the AdvancedTCA backplane, and uses as well an extra connector (P19) to enable the connectivity of the two extra links required by the interface.

6.2.3 Base Interface

All PICMG 3.0 compliant backplanes support a Base Interface in a dual-star topology. Each link in the dual-star Base Interface defines a Base Channel, starting in one of the Hubs and having the end point in any other Node slot of the backplane. The transport supported by each Base Channel can arrive to a maximum of 1Gbps (the protocol supported can be chosen between 10/100/1000 Base-T).

The Base Channel is recommended to be used for the control of the system accommodated inside the AdvancedTCA chassis. Let us consider for the simplicity of the explanation that only a single point of control, respectively that only the primary star of the Base Interface starting from Hub1 is used. On the entire AdvancedTCA backplane, there are then used 13 Base Channels to achieve the connection between Hub1 and all the other existent slots.

Similarly to the dual-star Fabric Interface, the PICMG specifications define the physical implementation of each Base Channel in terms of slots and pin assignment on the connector. When mapped into the high-speed connectors, a Base Channel occupies an

entire row. On the AdvancedTCA backplane, the Base Interface uses two connectors (P23 and P24) on the Hub1 to support connections to all slots in the AdvancedTCA chassis. The Base Channel starting in Hub1 is received at a different point in the connector in Hub2 than in a Node board, due to an asymmetry in the system caused by the presence of the Shelf Management Controls (SHMC). Therefore, on the AdvancedTCA backplane an extra Base Channel is needed to be used for the connection between Hub1 and Hub2. Thus a total of 14 Base Channels are involved for the primary star interconnection in the AdvancedTCA backplane.

Each Base Channel itself is based on four differential pairs: two pairs for transmitting and the other two used to receive the signal. The fourteen Base Channels in the AdvancedTCA backplane, correspondingly occupy 64 (14 x 4) differential pairs.

6.3 Summary

In this chapter we described the AdvancedTCA architecture, defined in the corresponding standard issued by the PCI Industrial Computer Manufacturers Group (PICMG) in the beginning of 2003. The standard was quickly adopted by some sectors of the industry and it became more evident for us in the middle of 2003 to follow that trend instead of building our own proprietary 10Gbps backplane. We had first to convince ourselves that the new standard was appropriate for the backplane of the 10Gbps Ethernet switch and therefore the AdvancedTCA specifications have been extensively analysed in a study presented in this chapter. The conclusions of this analysis confirmed the compatibility of the AdvancedTCA backplane with our own design and therefore its suitability for the 10Gbps Ethernet switch. The final decision was taken to adopt the standard.

7 AdvancedTCA Backplane Tester

The ESTA research project required not just that a working design be developed on the basis of a theoretical model but that the model itself is put to a practical test. A methodology needs therefore to be developed that will serve the purpose of validating the design results and establish practical performance metrics for a manufactured backplane. In addition we need to demonstrate what the achieved margins of correct performance are, and in so doing provide an accurate estimate for achievable performance in the future. Finally we needed a method that would fully test the manufactured backplane to ensure that no backplane manufacturing defects existed that might compromise the performance of the 10Gbps Ethernet Switch.

7.1 Context and Motivation

Leading edge companies, of whom a few were early adopters of the AdvancedTCA standard, are driven by the need to bring their products to market as early as possible.

The AdvancedTCA backplane design was subject to extensive and professional simulation, testing and demonstration before meeting the specifications included in the standard. From our experience in simulating our initial design we understood however that even the most sophisticated tools available can only simulate, at best, a small number of PCB tracks simultaneously and would not be able to give a figure for the backplane performance as a whole. In practice however the actual construction of the backplane can be itself subject to PCB manufacturing control problems, which in turn could have a negative effect on signal integrity and system performance.

Most of the companies which adopted AdvancedTCA standard for their designs since the beginning have to bring their products as early as possible to the market. One consequence of this is that early product releases will often consist of prototype silicon, mounted on prototype PCBs, running high speed signals across a new standardized AdvancedTCA backplane solution. It exists therefore a significant possibility that errors could affect the functionality of these prototype systems. The errors can be due to the logic or to the PCB quality control during the PCB manufacturing process.

For the 10 Gigabit Ethernet switch, as for any other commercially produced equipment, a low quality AdvancedTCA backplane combined with marginal silicon performance could cause low level errors in the final system. Tracking down the cause of these errors and fixing them in the final product, can sometimes take as long as the original design cycle. It would greatly simply the commissioning process if a tool were available that could validate and characterize the backplane sufficiently to demonstrate that it could not be the source of any observed low level errors.

At first, we considered specific test systems existing on the market for backplane validation. However, it turned out that these systems are essentially only checking the connectivity, working only at very low frequencies or DC, and don’t have the capability of detecting high-speed signal integrity problems. More performant testing systems are of

course available, in the form of high speed programmable data streams. These are designed to test that the PHY levels of an interconnect meet the standards BER under the most exacting of timing conditions. They have a very high level of functionality, and a correspondingly high cost per channel. It was therefore decided within ESTA to develop our own, cost effective, tester which would serve not just the requirements for prototype testing, but also as a means to test any AdvancedTCA production backplane samples. This tester system is intended to be not just a simple continuity tester, but a real high-speed, multi-channel, data traffic generator that is capable of simultaneously checking every connection on a dual-star topology for a fully populated AdvancedTCA backplane.

It is named therefore later on throughout the thesis as the AdvancedTCA Backplane Tester. Some of the high-level requirements we have for the AdvancedTCA Backplane Tester system are:

- the possibility to provide a relatively inexpensive AdvancedTCA backplane validation tool that goes beyond connectivity testers;
- the possibility to demonstrate that any given AdvancedTCA backplane assembly can sustain full bandwidth error free communication across all links of the dual-star topology simultaneously;
- the possibility to easily identify fault conditions to assist in the manufacturing process.

In the following paragraphs, detailed information related to the design and the practical architectural hardware aspects of the AdvancedTCA Backplane Tester will be described.

7.2 AdvancedTCA Backplane Tester Hardware Design

7.2.1 Traffic Generation on Fabric Channels

7.2.1.1 Fabric Interface Backplane Routing

The Fabric Interface was used in the AdvancedTCA Backplane Tester to distribute the high-speed data in the system. The 25 Fabric Channels, 12 differential pairs each, occupy a total of 300 differential pairs on the AdvancedTCA backplane. The routing assignment in terms of logical slot in the AdvancedTCA chassis and pin assignment in the high-speed connector for the Fabric Interface is reported in Table 7.1.

			Logical slot #													
Channel #	**Connector**	**Row**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Fabric Channel 13	P20 P19	9-10 2	14-1	14-2	Not implemented on backplane											
Fabric Channel 12	P21 P19	1-2 3	13-1	13-2												
Fabric Channel 11	P21 P19	3-4 4	12-1	12-2												
Fabric Channel 10	P21 P19	5-6 5	11-1	11-2												
Fabric Channel 9	P21 P19	7-8 6	10-1	10-2												
Fabric Channel 8	P21 P19	9-10 7	9-1	9-2												
Fabric Channel 7	P22 P19	1-2 8	8-1	8-2												
Fabric Channel 6	P22 P19	3-4 9	7-1	7-2												
Fabric Channel 5	P22 P19	5-6 10	6-1	6-2												
Fabric Channel 4	P22 P20	7-8 5	5-1	5-2												
Fabric Channel 3	P22 P20	9-10 6	4-1	4-2												
Fabric Channel 2	P23 P20	1-2 7	3-1	3-2	2-3	2-4	2-5	2-6	2-7	2-8	2-9	2-10	2-11	2-12	2-13	2-14
Fabric Channel 1	P23 P20	3-4 8	2-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14

Table 7.1: Fabric Interface and Base Interface routing assignment

7.2.1.2 SerDes for Traffic Generation

The AdvancedTCA Backplane Tester we designed consists of 14 boards, which are plugged into the backplane and are able to generate 10Gbps traffic that is distributed over the dual-star topology of the Fabric Interface. To drive each differential pair of every Fabric Channel over the backplane with controllable traffic, PICMG specifications recommend the use of 'Serdes' interfaces (see paragraph 3.3.3.1). The basic idea was to use a set of high-speed SerDes devices mounted on the Hubs and the Nodes boards of the Backplane Tester, in order to fully-load with traffic any differential pair within the dual-star Fabric Interface.

There were several SerDes devices available on the market, but we chose the Alaska 88X2040 transceiver from Marvell. This device was chosen for the design of the 10Gbps Ethernet switch and its performances were already studied in chapter3. The Alaska 88X2040 has a built-in BER circuitry which permits to generate and check traffic in terms of 10Gbps Ethernet patterns without any external devices. The 88X2040 functional block diagram is reported in Figure 7.1.

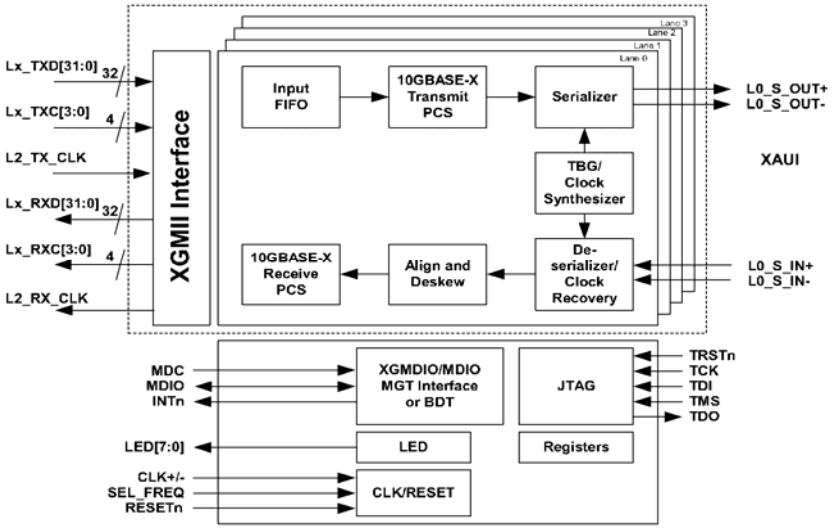


Figure 7.1: 88X2040 transceiver functional block diagram

The traffic is internally generated by the self test circuitry in each of the four lanes (Lane_0 to Lane_3) of the device and it is serialized, encoded and sent out via a set of four differential output pins of its XAUI interface. The generated data is then transmitted over the differential pairs on the backplane and upon reception at the other end of each pair, the received data is de-serialized, aligned, de-skewed, decoded and finally checked. The BER feature of the Alaska device permits to test independently each differential pair connection by monitoring the errors appearing on each pair involved in the backplane test. For that, at the Serdes receiver the incoming data stream is checked for errors and one 32-bit error counter is used per Lane to maintain the BER statistics.

7.2.1.3 SerDes Implementation in the Tester

Each Alaska 88X2040 SerDes device embeds four differential drivers and four differential receivers operating at 3.125Gbps. Three of them are required to saturate the two Fabric Channels, respectively 12 differential pairs corresponding to each Node board connection. In the full AdvancedTCA Backplane Tester system, 76 Marvell Alaska devices are involved to fully occupy all 300 differential pairs in the Fabric Interface with 20 devices for each Hub board and 3 devices for each Node board. Figure 7.2 shows the implementation of Alaska devices in the AdvancedTCA Backplane Tester system.

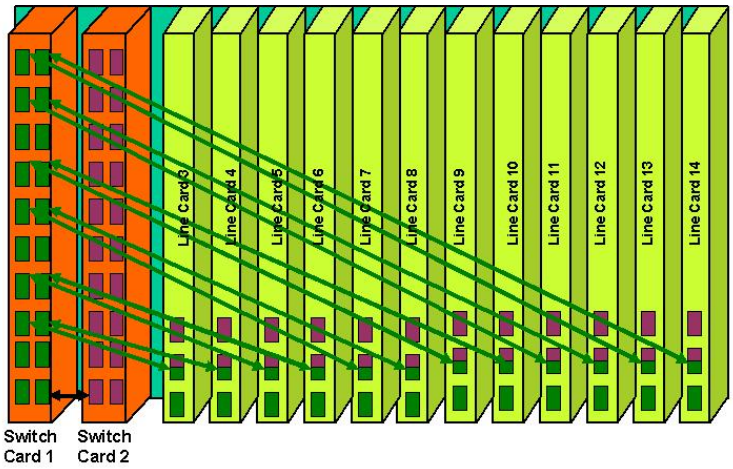


Figure 7.2: Marvell Serdes as traffic generators in the active AdvancedTCA backplane tester system

Only the star connection starting from Hub1 is shown in the figure for clarity, but a similar dual-star corresponds to the redundant Hub2 board. Only one and a half Serdes devices on each Node board is involved in the communication with the primary Hub1. The connectivity between the SerDes devices in the Backplane Tester system is shown in Figure 7.3 which indicates the mapping of individual Serdes devices onto the interconnects between Hub1 and 2 to each Node card.

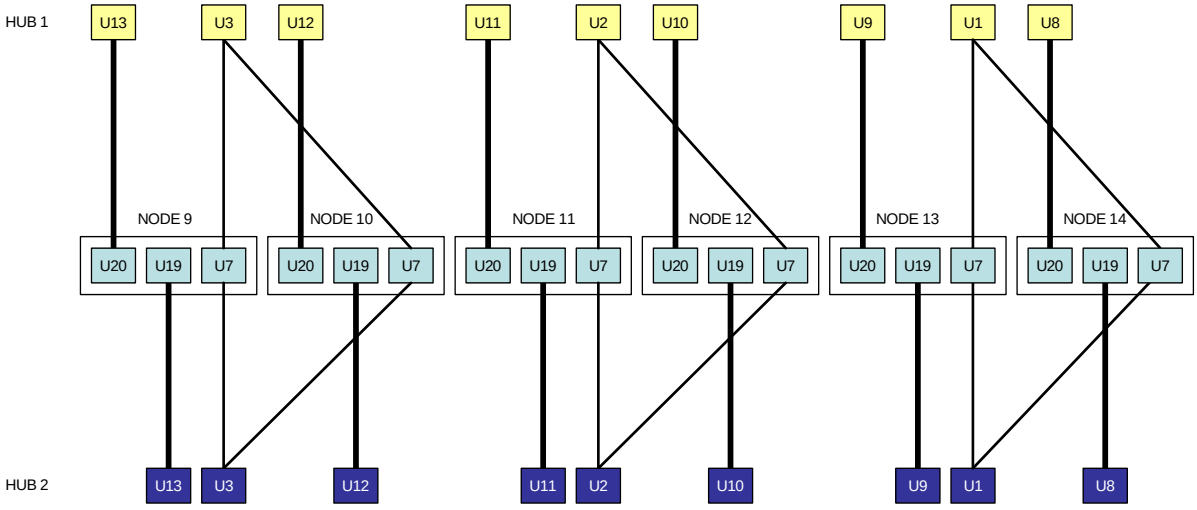


Figure 7.3: 88X2040 transceivers implementation in the Backplane Tester system

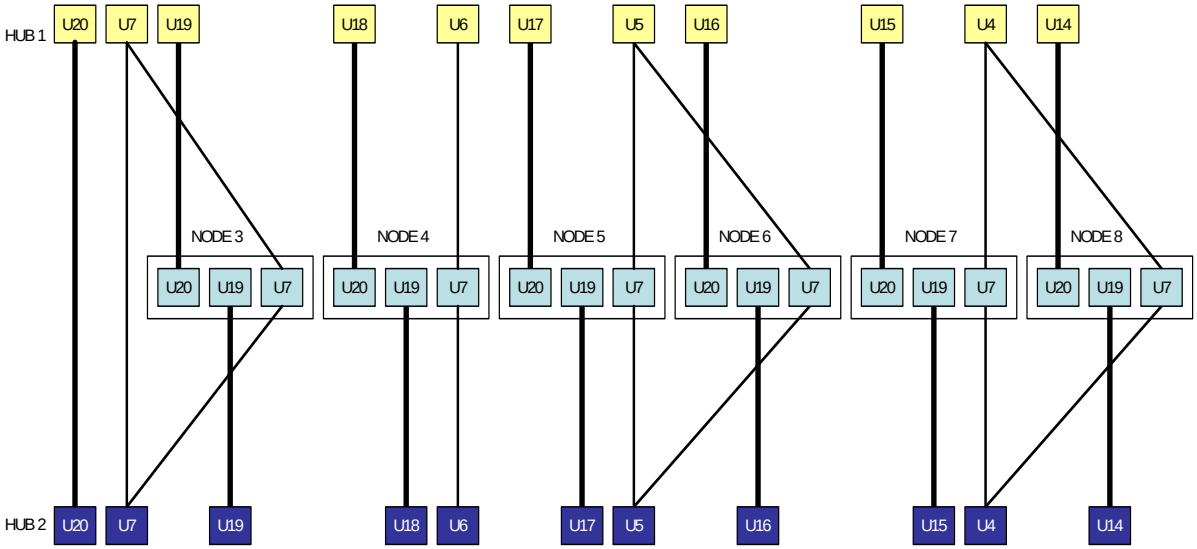


Figure 7.3: 88X2040 transceivers implementation in the Backplane Tester system

The Serdes components are illustrated by their physical addresses. Thick lines represent a Fabric Channel link composed of four differential pairs per direction. Thin lines represent the extension of extra two differential pairs per direction, required by the architectural design of the 10Gbps Ethernet switch.

7.2.2 AdvancedTCA Backplane Tester Control

7.2.2.1 MDIO Interface

The 10 Gigabit standard defines the Management Data Input/Output (MDIO) interface [IEEE-02] between a Station Management (STA) and a managed physical layer device (PHY) entity, in order to allow the access to the internal registers of the PHY device. The chips handling 10Gbps functions, in particular the SerDes, have such a standardized MDIO control path.

MDIO is defined as a serial interface that supports two types of frames - write and read - as shown in Figure 7.4.

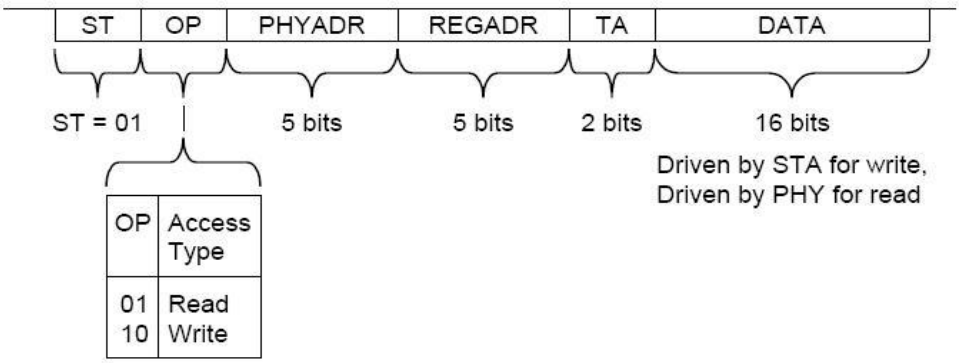


Figure 7.4: MDIO write and read frame as defined in 10 Gigabit standard

Within these, the write frame is driven by the STA master to the PHY device, synchronously with the management data clock (MDC), while the read frame is driven by the device itself.

From a hardware point of view, the MDIO control bus is a slow (0 to 10MHz) two wire bussed connection. A maximum of 32 devices (limitation due to 5 bits of addressing available in the 'PHYADR' field of the MDIO frame) can be located on a single MDIO chain where each device address is individually set by hardwired jumpers on the device motherboard. Every device on the MDIO chain receives four signals: RESET, MDC, MDIO (data) bidirectional line and a signal, READ, which indicates the direction of the data on the line. A typical MDIO sequence for the read instruction is shown in the Figure 7.5.

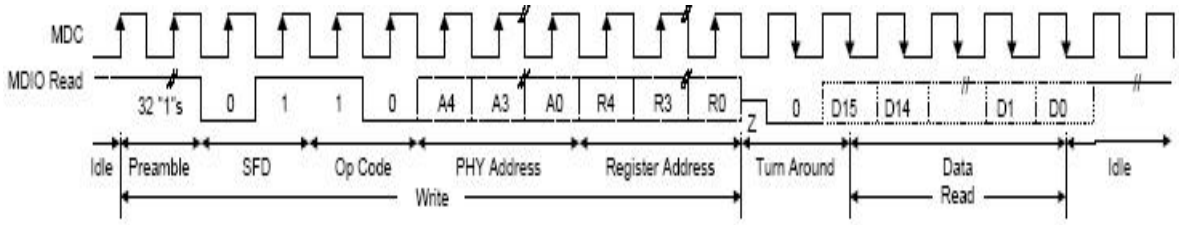


Figure 7.5: Typical MDIO sequence

In the example above the MDIO line is first driven by the STA master which issues a read frame containing the device address and the register address. On the MDIO chain, only the device which recognizes its address responds to the read command by driving the MDIO bus with the requested data.

All the Marvell SerDes devices in the AdvancedTCA Backplane Tester are programmed and controlled via the MDIO interface. In Figure 7.6 is illustrated the transmission path between two Marvell Alaska 88X2040 devices across the AdvancedTCA backplane. For simplicity reasons, we consider only a single lane in the device.

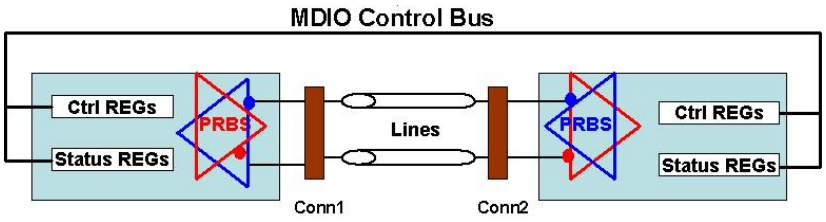


Figure 7.6: Single test channel between two SerDes devices

The typical configuration phase for every Marvel Alaska includes device initialization, the pre/de-emphasis settings for every SerDes lane and the setup of the device for driving PRBS or Jitter patterns. All the operations required for the configuration of the device are realized by writing the Alaska's internal control registers (Ctrl REGs) with the appropriate values for achieving a certain task. Monitoring of status at any time is achieved by reading the values contained by the device status registers (Status REGs).

7.2.2.2 Point of Control

To simplify board design, we designed in a single point of control for the whole test system, implemented in Hub1. The Base Interface Channels issued from Hub1 are then used to distribute control signals to all Node boards and to Hub2. This limits the cost and complexity of the control plane.

The MDIO interface has no critical timing constraints and the STA could be any processor or microcontroller. We chose a microcontroller that offered a maximum of embedded functionality. The Lantronix DSTni-LX-001 [LTR-01] offered many programmable I/O pins as well as inbuilt field bus ports, Ethernet ports and RS232 for booting simple processes. The Lantronix is a 16-bit microprocessor, Intel 80186 compatible, driving a set of industrial field bus ports plus 32 independent I/O lines. The microcontroller also has 256KB of on-chip SRAM memory. We chose to develop the software on a reference platform. In addition to the controller the reference platform carries 1MB of on-board RAM as well as parallel and serial flash memories. An Ethernet 10/100 Mbps controller is integrated and allows connection to an external PHY device. The system block of the Lantronix DSTni-LX-001 is illustrated in Figure 7.7.

The Lantronix controller supports a real-time operating system (RTOS), an embedded TCP/IP stack and it is designed to use standard Intel/AMD x86 development tools.

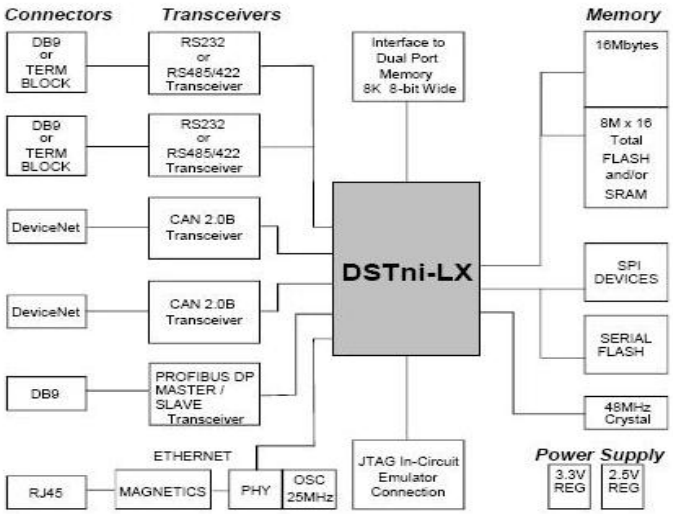


Figure 7.7: DSTNI-1 System Block Diagram

7.2.2.3 Backplane

The primary star of the dual-star Base Interface was used to fan-out the control data. The 14 Base Channels, 4 differential pairs each, occupy a total of 64 differential pairs on the AdvancedTCA backplane. The routing assignment in terms of logical slot position and pin assignments in the high-speed connectors for the Base Interface is reported in Table 7.2.

			Logical slot #													
Channel #	**Connector**	**Row**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
	P20	8														
Base Channel 1	P23	5	Not used		1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-12	1-13	1-14
Base Channel 2	P23	6	2-1	1-2	Not used											
Base Channel 3	P23	7	3-1	Not used	Not implemented on backplane											
Base Channel 4	P23	8	4-1													
Base Channel 5	P23	9	5-1													
Base Channel 5	P23	10	6-1													
Base Channel 7	P24	1	7-1													
Base Channel 8	P24	2	8-1													
Base Channel 9	P24	3	9-1													
Base Channel 10	P24	4	10-1													
Base Channel 11	P24	5	11-1													
Base Channel 12	P24	6	12-1													
Base Channel 13	P24	7	13-1													
Base Channel 14	P24	8	14-1													

Table 7.2: Fabric Interface and Base Interface routing assignment

7.2.2.4 Control Signals

Each differential pair on the backplane has controlled-impedance PCB traces designed for carrying 10/100/1000Mbps Ethernet signaling so control signal integrity is not a problem. However, the corresponding signals at the microcontroller level need to be buffered to drive differentially the Base Interface Channels. Since the MDIO lines are bidirectional, the direction the data line buffers are driving is specified by a READIO signal provided by the microcontroller. In addition to the three control lines (MDC, MDIO and RESET), this signal (READIO) is then sent on the fourth signal pair. Moreover, some fan-out is required at the controller level because there are not enough lines to give every Marvell SerDes an individual control pair.

In the AdvancedTCA standard specifications, each Base Channel occupies a row with four differential pairs in the ZD connector. It was easy then for us to assign each MDIO signal to one differential pair as presented in Table 7.3.

Diff. Pair	HUB1 Connector Pair	HUB2/NODE Connector Pair
MDIO	a-b	c-d
READIO	c-d	a-b
MDC	e-f	g-h
RESET	g-h	e-f

Table 7.3: Differential Control Pair Assignments

There was however further complication in that a fully-populated Backplane Tester was controlled by only one STA (Lantronix) which needs to manage all 76 SerDes devices. The MDIO frame only supports five bits of addressing for any physical device port which gives a maximum of 32 addressable devices possible to be accessed per individual MDIO bus. Since there are twenty MDIO ports on each Hub and three MDIO ports on each Node board, the control point needs to manage more than one single MDIO chain. We therefore partitioned the design so that four MDIO busses are used: two of the busses for controlling each Hub and other two for the rest of the twelve Node boards.

- Chain A: Node boards located in slots 3:8 with a total of 18 SerDes devices;
- Chain B: Node boards located in slots 9:14 with a total of 18 SerDes devices;
- Chain C: Hub 2 with 20 SerDes devices;
- Chain D: Hub 1 with 20 SerDes devices.

The Figure 7.8 show the use of the I/O pins in the Lantronix controller in order to achieve the distribution of the required control signals within all four MDIO chains.

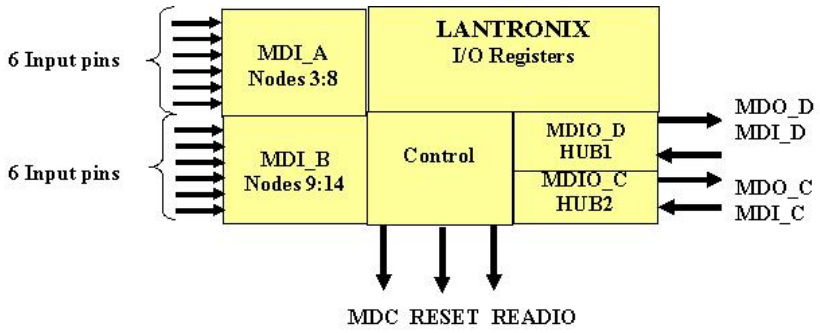


Figure 7.8: I/O pins in the Lantronix used to generate the control signals

MDC, RESET and READIO can be common signals for all the 4 MDIO chains, if they are buffered at the Lantronix level. To handle these signals, three I/O pins were defined as outputs in the Lantronix. However, the MDIO signals have to be individually generated and distributed across the backplane, according to their corresponding chain and their directions (MDIO write or read, respectively output or input).

The Lantronix ports can be defined as input or output but not bi-directional. Therefore, the MDIO line in our design is split into Input (MDI) and Output (MDO) at the level of the controller. By splitting the MDI and MDO, we can bus together the outputs according to which chain they are on and drive them differentially across the backplane.

The transceivers that interface between the single ended Lantronix I/O pins and the differential ports have totem-pole outputs and therefore cannot be used as a wired-OR. Therefore one Lantronix pin can be used to drive all the MDIO lines on each of the four MDO lines but each slots MDI has to be received individually by one Lantronix pin per input.

7.2.2.5 Control Architecture

The control architecture has been defined so that a single PCB can be used either for Nodes or for Hubs design, by populating the boards in a different way. The control architecture is reported in Figure 7.9 and to illustrate the different functionalities, different colour nuances are used. All the logic shown in a light (yellow) colour is only mounted on Hub1 and it is absent on the Hub2 and the Node cards. The components in dark blue are present on the Hub2 board and the ones highlighted in light blue are mounted on the Node boards.

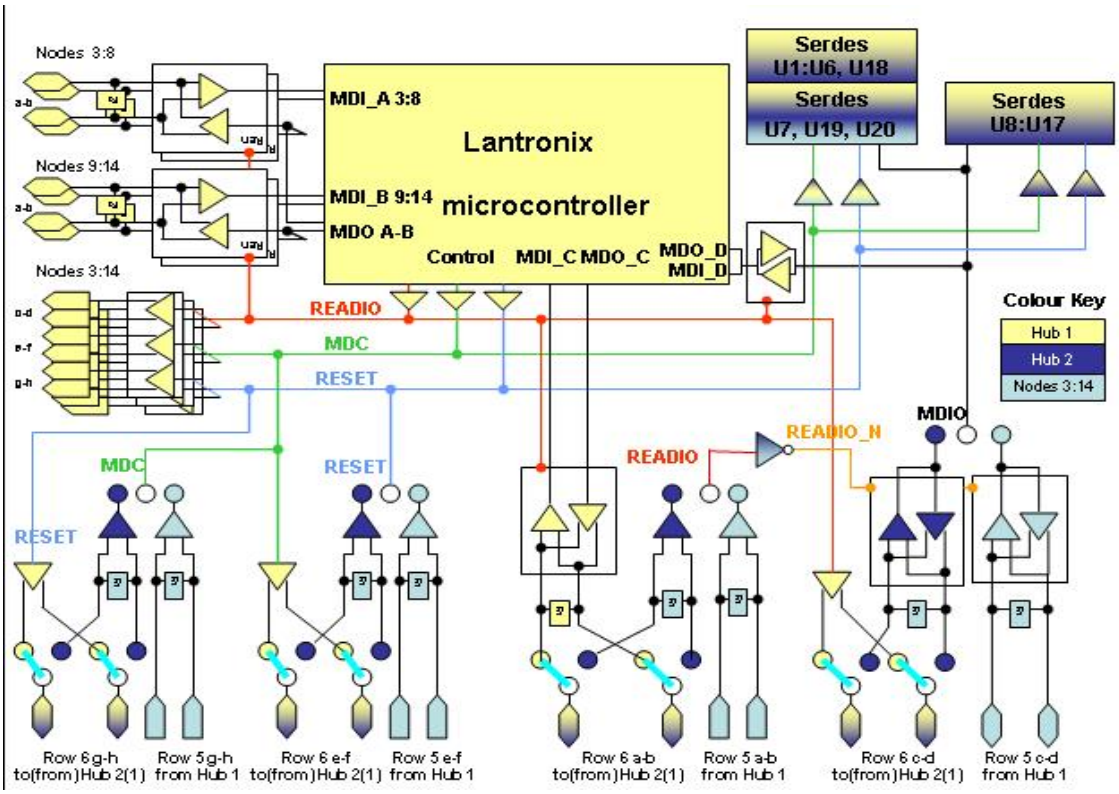


Figure 7.9: Control architecture of the AdvancedTCA Backplane Tester

The Hub boards are fully populated with twenty 88X2040 SerDes (U1 to U20) whereas the Node boards (U7, U19 and U20) are partially populated with only three of them. The four control signals (MDIO, MDC, READIO and RESET) are distributed in the

AdvancedTCA backplane from the Lantronix microcontroller on Hub1, over 14 Base Channels.

There are two levels of jumpers which are used to define whether the board is a Hub or a Node and thus whether to drive or to receive the various control and data lines. The array of jumpers in the lower part of the figure is used to differentiate the functionality of the Hub1 to all the other boards (Hub2 and Nodes). Due to the reported asymmetry in the Base Channels (caused by the presence of the SHMC), there must be some jumpers to select the function of a board as being either a Hub2 or a Node card, respectively the appropriate settings for rows 5 and 6 in the backplane connector. To allow the choice between a Hub2 or a Node at the receiving point of the control signals, there is a second level of jumpers placed in the upper level of the figure.

The twenty 88X2040 SerDes on each Hub board are split into two groups presenting 10 loads each to the buffered control signals RESET and MDC. The three control signals (READIO, MDC and RESET) are buffered at the level of the microcontroller chip since there are 13 line drivers to serve (12 Node boards and Hub2) as the internal loads. MDC and RESET lines were buffered however to ensure that the edges are rapid and not likely to create false triggering inside the 88X2040 transceiver logic.

The MDIO data line is buffered at the level of SerDes devices since the microcontroller pin does not have much drive capacity and there are twenty loads to be served. Double buffering is not required for the MDIO data line at the controller because the line has a long time to settle in between clocks. Serial termination is included at the level of each individual single to differential buffer to avoid possible signal overshoot. For MDIO access to the Hub1 Serdes, there is no requirement for differential lines and the bi-directional problem is resolved with a buffer under control of the Lantronix.

Let us just take the RESET signal as an example to explain the above control architecture. When a board is placed in the Hub1 position, the RESET signal is generated and fanned out to all other slots in the AdvancedTCA chassis. The fan-out to Hub2 uses pair Row 6 g-h in connector P23. When the board is placed in the Hub2 position, Row 6 g-h is used to receive the clock MDC, so there is a first level of jumpers to cover this double usage of pair Row 6 g-h. Now, when the board is placed in any of the Node slots, MDC is received from Row 5 g-h. There is therefore a second layer of jumpers to distinguish between double sources of this control signal. Note that when the board is being used in the Hub1 position, there are no jumpers installed at all on this second layer since all the control signals to the Serdes are generated locally on the board. When being used as Hub2 or Node card, the microcontroller and MDIO buffers are not mounted so there is no driving conflict on the control lines.

7.2.2.6 Microcontroller Implementation

The control of the AdvancedTCA Backplane Tester is achieved from an externally remote PC over two serial interfaces as shown in the Figure 7.10: a RS-232 serial link and an Ethernet interface. The control of the Backplane Tester requires the distribution of the software between the remote PC and the Lantronix.

The serial port (RS-232) is used to upload software executables to the microcontroller and the Ethernet interface is used for communication between the applications operating on the microcontroller and on the remote PC.

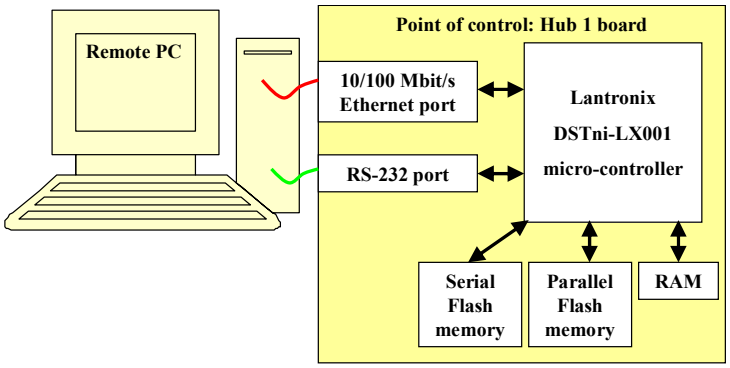


Figure 7.10: Microcontroller implementation

7.3 AdvancedTCA Backplane Tester Boards

The Backplane Tester boards have the dimensions 322.5mm x 280mm and they are built with a 14-layers stack-up, as reported in Table 7.4.

Layer	Name	Copper thickness (mils)	Dielectric thickness (mils)
1 (coated microstrip)	**Sig-1**	2.1[(1)]	
Pre-preg			7.5
2 (ref. plane)	**P1V5/GND**	1.4	
Core			8
3 (ref. plane)	**P1V8/P2V5/GND**	1.4	
Pre-preg			7.5
4 (ref. plane)	**P3V3/GND**	1.4	
Core			8
5 (ref. plane)	**GND**	1.4	
Pre-preg			8.5
6 (stripline)	**Sig-2**	0.7	
Core			8
7 (ref. plane)	**GND**	1.4	
Pre-preg			8.5
8 (stripline)	**Sig-3**	0.7	
Core			8
9 (ref. plane)	**GND**	1.4	
Pre-preg			8.5
10 (stripline)	**Sig-4**	0.7	
Core			8
11 (ref. plane)	**GND**	1.4	
Pre-preg			8.5
12 (stripline)	**Sig-5**	0.7	
Core			8
13 (ref. plane)	**GND**	1.4	
Pre-preg			7.5
14 (coated microstrip)	**Sig-6**	2.1[(1)]	
Total thickness (including plating) +/- 10%		**3.1 mm**	

Table 7.4: Backplane tester PCB stack-up

[(1)] 1.5 OZ Cu Finish.

The AdvancedTCA Backplane Tester boards corresponding to this stack-up have been built. The Figure 7.11 shows one of the signal layers for one board populated for Hub1 purpose. The board is shown with the backplane connectors on the upper edge and the front panel connectors on the lower edge. Each of the Serdes can be identified as a small square footprint from which a pair of tracks makes its way to the backplane connectors. It can be seen that the board represents a real practical case where the distances from Serdes to connector are both varied, and in the case of the most distant ones, also quite long.

The key design challenge in the PCB design of Hub1 was the routing strategy for the 156 differential pairs running at 3.125Gbps (13 Fabric Channels x 12 differential pairs) related to Fabric Interface Channels. Among the 6 signal layers reported in Table 7.4, four of them are dedicated for routing those signals.

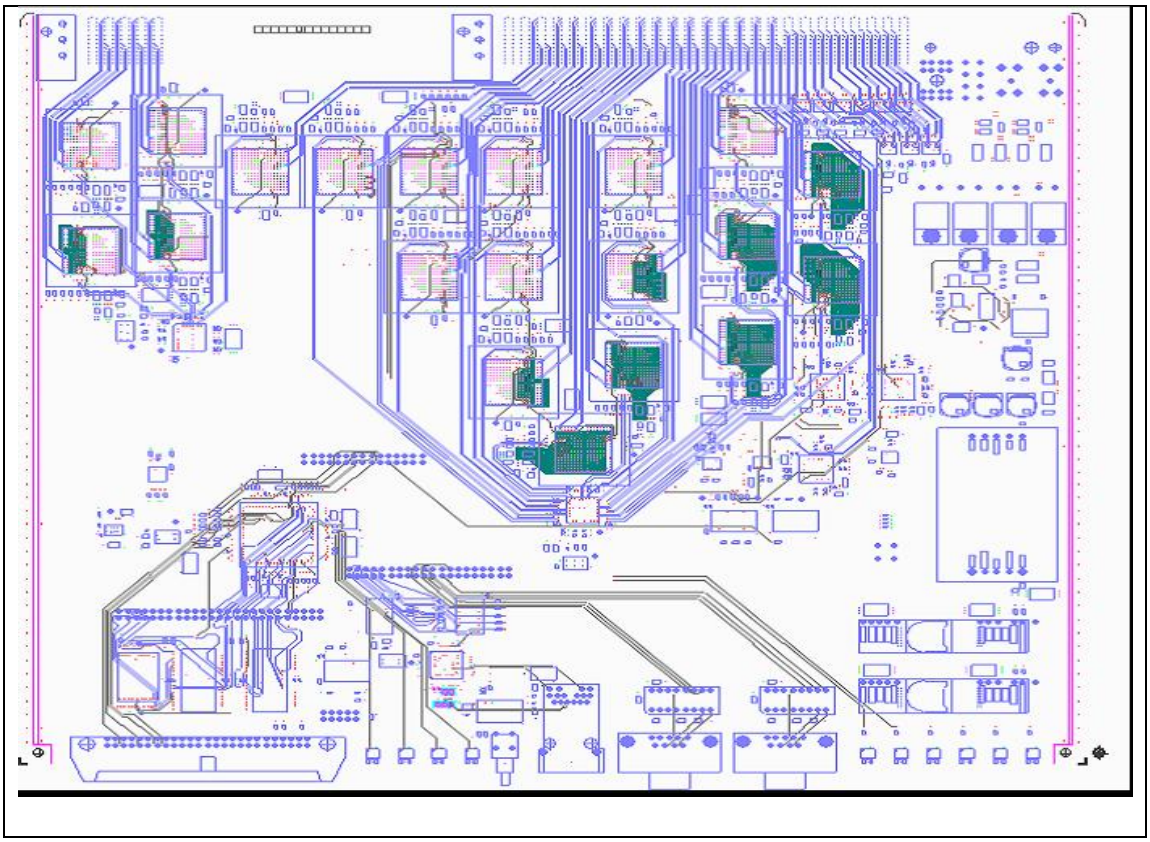


Figure 7.11: Routing on layer Sig-2

Under some of the Serdes can be seen blocked in areas. These correspond to small 'islands' of power set into a lower layer. There are not enough layers to provide a power plane for every voltage, so islands of one voltage are constructed within a plane of a different voltage (or ground).

In the right hand side, 20% of the board is dedicated to the generation of the various voltages required all derived from the incoming 48V DC. The Lantronix controller and its ancillary logic is in the lower left corner of the board.

7.4 AdvancedTCA Backplane Tester Control Software Design

7.4.1 Software Architecture Model

The most commonly used paradigm in constructing distributed applications is the client-server model [OLT-R3]. In this scheme, the client application request services from the server process, which normally listens at a well known address (port) for service requests. When a connection is requested by a client to the server's address, this one services it, by performing whatever appropriate actions the client requests. The second task performed by the server is to supply information regarding the status of its host. The client and

server require a well known set of conventions, known as a ‘protocol’, which must be implemented at both ends of a connection.

This client-server model suits quite well our Backplane Tester application and we decided therefore to develop the control software based on it. From a remote PC, which is connected to the Backplane Tester system, the user can completely define the test to be performed over the AdvancedTCA backplane. The controller task is to set-up the traffic across the backplane according to the user-defined test, and to maintain the statistics to the remote PC. It comes then naturally the distribution of the client tasks for the remote PC and the server task to the controller on Hub1. The connection between the client and the server, through which the client requests and the server statistics are sent, is made via TCP/IP over a 100Mbps Ethernet link.

In the following sections we will consider some of the problems in developing client and server applications and look more closely how the user interacts with the tester system for defining and performing a test of the AdvancedTCA backplane.

7.4.2 Software Architecture Design Issues

Rather than invest effort in developing our own graphical user interface (GUI) we took the approach of using available desktop packages as far as possible before committing to custom software. The user interface for the Backplane Tester is therefore implemented on an Excel spreadsheet (see paragraph 7.4.3), with which the user interacts directly on the remote PC, by defining the test(s) to be performed over the backplane.

A first design issue was related to choosing the place to perform the conversion of the Excel data into a valid request for the server. There were two options considered. One was to download the Excel table onto the controller using a FTP connection and to convert the data at the server level. The second possibility was to convert the Excel table into a client request on the remote PC and then to sent this request to the server. We finally decided on this second possibility to keep the server application as simple as possible for debugging purposes, and to have a direct access to the conversion result on the remote PC.

In order to have a simplified server application, we decided to restrict its functionality to the management of the MDIO bus, being able this way to execute individual ‘write’ or ‘read’ MDIO instructions. The client application has to convert the Excel user specifications and to provide the corresponding MDIO instructions to the client, according to a well-defined communication protocol. The Excel conversion could therefore just be directly into a long succession of the resulting MDIO commands.

However there is another requirement which is to be able to manually issue simple commands for the purposes of hardware debugging. For this to be effective it should be done at a level higher than the MDIO instruction since it can take tens of such instructions to set up just one link. We therefore have created an intermediate level between the abstraction of the spreadsheet and the fine level detail of the individual MDIO instructions. To serve this requirement, a set of macros that fulfil certain basic functions such as resetting, initialising etc. have been defined. For debugging purpose a

command line interface is provided where commands can be issued in terms of the macro and link identifier. In that way, the user has full control over the hardware at a level appropriate for the task.

This intermediate stage then defines the remaining software functions to be implemented. To run automated tests we have to first extract from the spreadsheet a set of macro calls that correspond to the tasks called for on each link under test. These macros are grouped into three text files called configuration files. This extraction is beyond the scope of pure Excel functions but could be done by programming or scripting. C programming in this application would be unnecessarily heavy and difficult to manage. We chose therefore to do it in Visual Basic scripting which has the advantage of being simple and free of any intermediate compilation steps. This is a pre-requisite when the user will be rapidly changing test conditions in order to track down the source of any observed errors.

Whether the macros have been created from the spreadsheet input or from the command line interface, the next stage is to expand them into the series of MDIO instructions to be executed by the client.

A final issue was related to the communication protocol between the client and the server. The client requests, now in format of write/read MDIO instructions, could be sent to the server instruction-by-instruction or all grouped together in one single transaction at the macro level. At issue here is the speed of the transaction, the amount of buffering required in the client, how to handle errors resulting from any one given MDIO instruction and how to associate acknowledgements and responses with the requests that generated them. The choice was made to partition requests at the macro level and send one packet request to the server for each macro, and in reply to this, only one acknowledgement packet per macro will be received back to the client.

Having as a start these architectural decisions, the conversion of the user interface (paragraph 7.4.3), the client (paragraph 7.4.4) and the server (paragraph 7.4.5) applications have been created and their description is given in the following paragraphs.

7.4.3 User Interface

The user interface is implemented on the Excel spreadsheet located on the remote PC. The Excel spreadsheet contains three different sheets. Only one of them is directly accessible to the user for defining the test to be performed on the AdvancedTCA backplane and this one is shown in Figure 7.12.

		STANDARD DATA TRANSPORT								EXTENDED DATA TRANSPORT				PARAMETER SETTINGS			
		Node[3:14] to HUB1				HUB1 to Node[3:14]				Node[3:14] to HUB1		HUB1 to Node[3:14]		Pre-Emphasis	PRBS	Jitter	TEST interval [sec]
Node	HUB	P23 Row4 [a:b]	P23 Row4 [e:f]	P23 Row3 [a:b]	P23 Row3 [e:f]	P23 Row4 [c:d]	P23 Row4 [g:h]	P23 Row3 [c:d]	P23 Row3 [g:h]	P20 Row8 [a:b]	P20 Row8 [e:f]	P20 Row8 [c:d]	P20 Row8 [g:h]				
3	1	1	1	1	1	1	1	1	1	1	0	1	1	0xA402	2^23-1	-1	400
4	1	1	1	1	1	1	1	1	1	1	1	1	1				
5	1	1	1	1	1	1	1	1	1	1	1	1	1				
6	1	1	1	1	1	1	1	1	1	1	1	1	1				
7	1	1	1	1	1	1	1	1	1	1	1	1	1				
8	1	1	1	1	1	1	1	1	1	1	1	1	1				
9	1	1	1	1	1	1	1	1	1	1	1	1	1				
10	1	1	1	1	1	1	1	1	1	1	1	1	1				
11	1	1	1	1	1	1	1	1	1	1	1	1	1				
12	1	1	1	1	1	1	1	1	1	1	1	1	1				
13	1	1	1	1	1	1	1	1	1	1	1	1	1				
14	1	1	1	1	1	1	1	1	1	1	1	1	1				

		HUB1 to HUB2				HUB2 to HUB1				HUB1 to HUB2		HUB2 to HUB1	
		P23 Row4 [c:d]	P23 Row4 [g:h]	P23 Row3 [c:d]	P23 Row3 [g:h]	P23 Row4 [a:b]	P23 Row4 [e:f]	P23 Row3 [a:b]	P23 Row3 [e:f]	P20 Row8 [c:d]	P20 Row8 [g:h]	P20 Row8 [a:b]	P20 Row8 [e:f]
HUB1	HUB2	1	1	1	1	1	1	1	1	1	1	1	1

		Node[3:14] to HUB2				HUB2 to Node[3:14]				Node[3:14] to HUB2		HUB2 to Node[3:14]	
Node	HUB	P23 Row2 [a:b]	P23 Row2 [e:f]	P23 Row1 [a:b]	P23 Row1 [e:f]	P23 Row2 [c:d]	P23 Row2 [g:h]	P23 Row1 [c:d]	P23 Row1 [g:h]	P20 Row7 [a:b]	P20 Row7 [e:f]	P20 Row7 [c:d]	P20 Row7 [g:h]
3	2	1	1	1	1	1	1	1	1	1	1	1	1
4	2	1	1	1	1	1	1	1	1	1	1	1	1
5	2	1	1	1	1	1	1	1	1	1	1	1	1
6	2	1	1	1	1	1	1	1	1	1	1	1	1
7	2	1	1	1	1	1	1	1	1	1	1	1	1
8	2	1	1	1	1	1	1	1	1	1	1	1	1
9	2	1	1	1	1	1	1	1	1	1	1	1	1
10	2	1	1	1	1	1	1	1	1	1	1	1	1
11	2	1	1	1	1	1	1	1	1	1	1	1	1
12	2	1	1	1	1	1	1	1	1	1	1	1	1
13	2	1	1	1	1	1	1	1	1	1	1	1	1
14	2	1	1	1	1	1	1	1	1	1	1	1	1

Possible PRBS values	Possible Jitter values
2^23-1	48A.1
2^31-2	48A.2
2^7-1	48A.3
	48A.4
	48A.5

Figure 7.12: View of the user interface

The other two Excel spreadsheets are completely transparent for the user and they define the connectivity on the entire AdvancedTCA backplane in terms of MDIO buses and Serdes location in terms of physical addresses.

For defining a test over the AdvancedTCA backplane however, the user has to modify only the pre-defined area of the first spreadsheet (Figure 7.12). This contains four areas, where the user can define the active differential pairs in the standard or extended data transport interface on the backplane, the levels of signal amplitude and pre/de-emphasis, the traffic type and the test duration. For example, in the current test all differential pairs on the AdvancedTCA backplane are defined as active (‘1’ in all the corresponding fields) with one exception - a pair between Node3 and Hub1 is disabled (‘0’ in the extended data transport).

Once the test configuration is defined in this Excel interface, the user has to run an existent Visual Basic script which automatically generates three configuration files (<setup.txt>, <reset_counters.txt> and <read.txt>). The other option would be to build up directly the configuration files by introducing directly the macros from the command user interface. These files will be further used as inputs by the client application. This process is schematically shown in Figure 7.13.

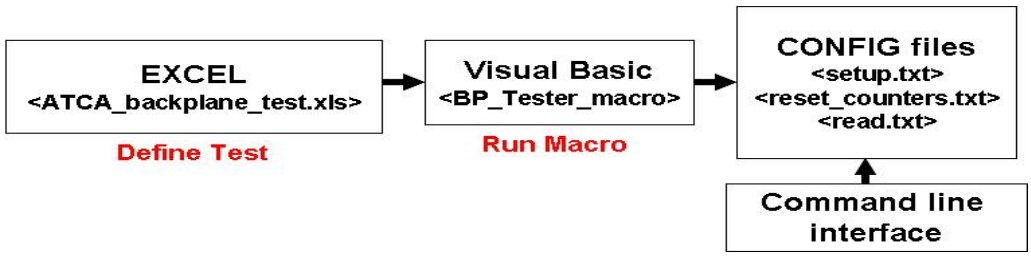


Figure 7.13: User interface overview

7.4.4 Client Application

The role of the client is to first take the user specifications for the test, to transform and send them to the server, according to the defined client-server communication protocol. The second task is to receive the statistics back from the server (based on the same communication protocol) and to report these statistics in real-time to the user. For achieving the first task, the client application takes as input parameters the configuration text files (<setup.txt>, <reset_counters.txt> and <read.txt>), which have been created based on the user specifications.

Each of these three configuration files contains a series of macros (groups of MDIO instructions for achieving a certain task). File <setup.txt>, for example, is used to configure each individual lane for transmitting and/or receiving traffic. The setup of a communication link implies an initial global reset of the system, the initialization of the Serdes devices involved in the communication, the setup of the signal voltage level and pre/de-emphasis, the disabling of the links not involved in the test defined by the user and the setting up of a certain traffic type to be generated. Files <read.txt> and <reset_counters.txt> are used periodically to read the SerDes devices status, their error counters and to reset them.

The client application expands each macro into an equivalent sequence of MDIO commands and sends this sequence to the server. All the macros in the <setup.txt> and <reset_counter.txt> configuration files are transformed into an equivalent sequence of 'writemdio' commands, while the macros in the <read.txt> file are converted into several 'readmdio' instructions. Once the macro has been expanded, the next step is to encapsulate the resulting instructions into a packet, and send it as a request to the server application, via the TCP/IP connection. In the communication protocol we defined between the client and the server, a packet request has the following format:

MACRO_Name\\MDIOinstr[1]\\MDIOinstr[2]\\…\\MDIOinstr[n]

Table 7.5: Format of the client request to the server

MDIOinstr[i] can be either a write or a read MDIO instruction.

After the client request has been sent, it is the server application's task to execute it command by command and to return back an acknowledge (ACK) packet, containing the execution results of each command. The ACK package format as defined by the client-server communication protocol is the following:

MACRO_Name\\ACK[1]:…:ACK[n]\\MDIOinstr[1]\\…\\MDIOinstr[n]

Table 7.6: ACK received from the server

ACK[i] contains the result of the MDIOinstr[i] execution, which can be the indication of a successful operation in the case of a write command, or the value read on the MDIO bus, in the case of a read instruction. On the remote PC, the client displays in real-time on the terminal console (and stores in corresponding log files) the processed received information from server, indicating the status of the communication links and the number of erroneous bits registered during signal transmission over the AdvancedTCA backplane. A brief overview of the client-server inter-operability is given in Figure 7.14.

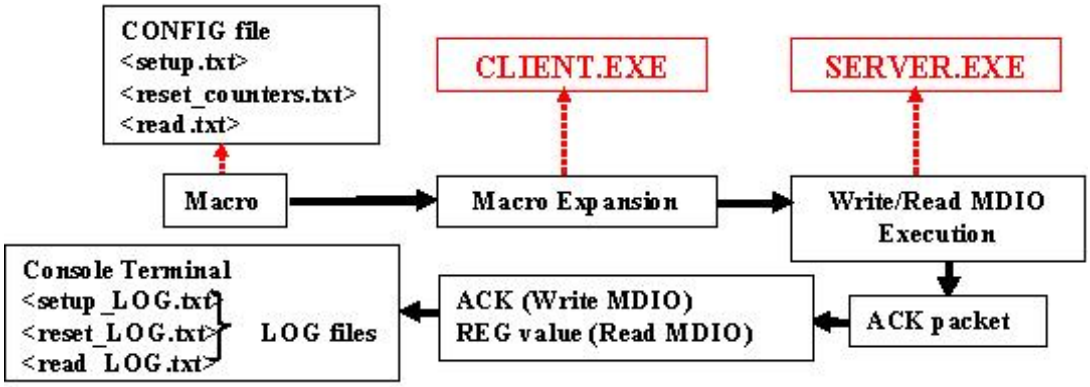


Figure 7.14: Client-server inter-operability overview

The test performed on the AdvancedTCA backplane can run for a user defined time or it can be stopped anytime by pressing <Ctrl^C> buttons. The results of a test can be read anytime by accessing the automatically generated LOG files.

7.4.5 Server Application

The server runs on the Lantronix controller. The application is downloaded via the RS-232 serial port from the remote PC and it is stored into the RAM memory, the serial flash or the parallel flash on the Hub1 board. The Lantronix processor controls the traffic on the backplane and monitors the state registers of the 88X2040 devices, through four MDIO control chains. The control, by the mean of write and/or read access to the internal registers, is emulated through the I/O pins of the Lantronix, by implementing at the server level the corresponding MDIO commands. These commands permit a low level access to the MDIO bus, by implementing the write/read protocol synchronously to a locally generated clock. The format of the corresponding MDIO commands sent by the client and received by the server can be seen in Table 7.7.

write_mdio:MDIO_BUS:PHY_ADR:REG_ADR:REG_DATA
read_mdio:MDIO_BUS:PHY_ADR:REG_ADR

Table 7.7: Format of the WRITE MDIO and READ MDIO commands

MDIO_BUS can have any value between A, B, C or D. The values for the device physical address (PHY_ADR), the register address (REG_ADR), respectively the register

data (REG_DATA) are given in hexadecimal. PHY_ADR can have any value between 0 and 19 for the Hubs (MDIO_BUS C or D) and values between 0 and 2 for the Node boards (MDIO_BUS A or B).

The client request at the server is received as MDIO commands, encapsulated and formatted according to the client-server communication protocol into a packet as presented in Table 7.5. According to the client-server protocol, the server unpacks the client request into low-level MDIO commands, which then it solves and gets an ACK for each of them individually.

7.5 Summary

The multiple requirements for the validation of the backplane design model, the measurement of the performance and the estimation of the available performance margins can be met by the backplane tester as described in this chapter. The basic element employed is a state of the art high-speed Serdes device which has the possibility to vary its speed, amplitude and level of pre/de-emphasis. By deploying many such devices, simultaneously across the whole switching fabric, the system wide performance and limits can be accurately determined. This chapter describes the architecture and design details for deploying and controlling such a large and complex array of traffic generators. Also it describes the architecture of the software that operates the control plane and the user interface for setting up and running tests over the AdvancedTCA backplane.

8 Backplane Tester System Integration

This chapter covers the commissioning of the AdvancedTCA Backplane Tester once the production phase had completed. Once the system was fully working we then describe the exploitation phase and report on the measurement results obtained [ESTA-R4].

8.1 Control Software Validation

The Lantronix DSTni-LX evaluation board (Figure 8.1) has been used for preparing and validating the control software of the Backplane Tester in parallel with the boards development and manufacturing. The evaluation platform was equivalent to the Hub1 board since the Hubs controller section was designed as a copy of the reference platform including the controller itself, adjacent memory (RAM, serial flash and parallel flash) and the Ethernet interface.

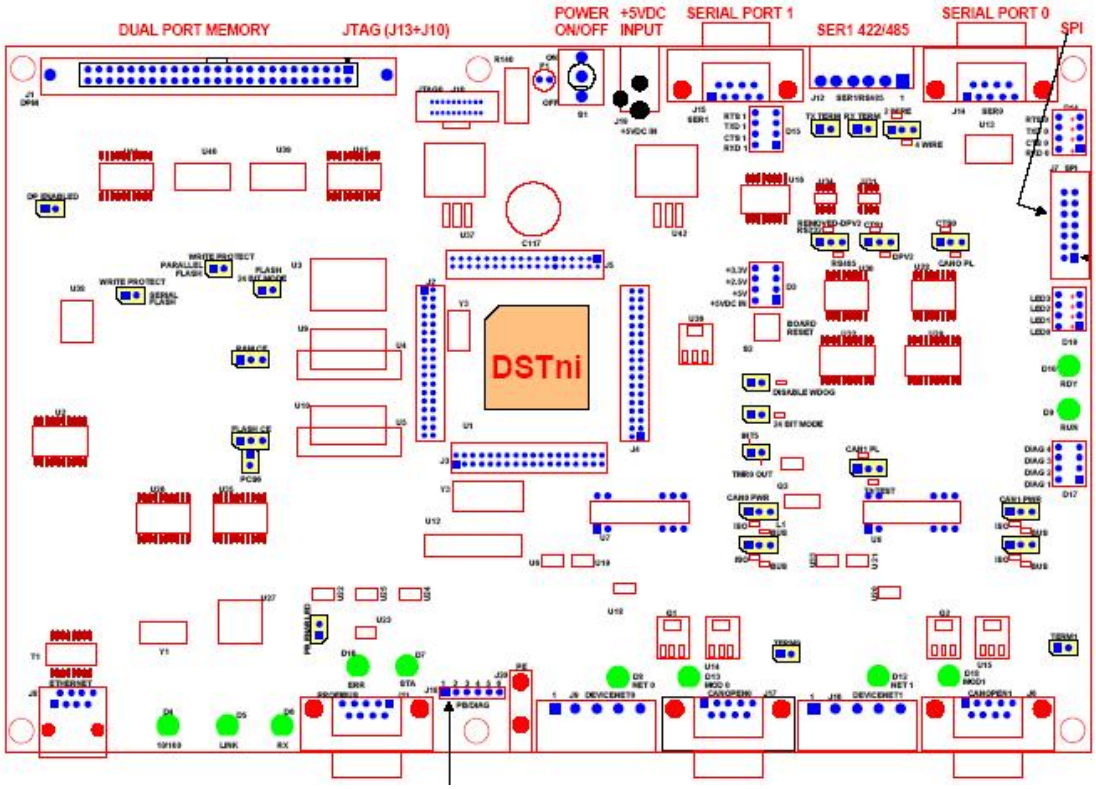


Figure 8.1: DSTni-LX evaluation platform

As described in the previous chapter, the control software of the Backplane Tester system is based on a client-server distributed application: the client located on the remote PC and the server application running on the Lantronix microcontroller. The client sends write and read MDIO instructions, which are executed on the server and acknowledged. To verify that the MDIO emulation was functioning correctly we used an oscilloscope to

monitor the MDIO bit sequences and timing relative to the MDC clock for each of the command sequences that we used.

After performing several measurements for different MDIO instructions, the conclusion was that the MDIO commands were properly generated by the client and correctly executed by the Lantronix controller. However, the behaviour of the Backplane Tester as a traffic generator cannot be predicted by the validation tests as no Alaska devices were present on the evaluation platform. The complete and the final functionality tests of the control software application can only be done with all the Backplane Tester boards inserted in the AdvancedTCA chassis.

8.2 Hub1 Board Testing

8.2.1 Communication with the Micro-Controller

Hub1 was the first board received from the manufacturer. It has the most important functionality, being the single point of control of the system containing all the required logic.

Following a well-defined test plan, we first performed standard power-off and power-on hardware tests of Hub1, which were successfully passed. Specific software applications were created for test purposes of the external components on the Hub1 (the serial port, the external RAM memory, the serial flash and the parallel flash). A successful upload via the serial port, storage and execution of the software from RAM, the serial or the parallel flash indicated a proper functionality of all the internal and external devices associated with the controller. This process established that we had complete access to the micro-controller which we could then use to test the remaining of the boards functions.

8.2.2 Testing the Client-Server Communication

The next step was to load the server into the micro-controller. This emulates the MDIO interface to the Alaska Serdes devices, by performing simple write and read MDIO instructions. The user interacts directly with the client software, running on the remote PC, either via a terminal line to which the MDIO commands are directly introduced, or by executing a script text file containing a set of MDIO commands.

Tests of the client-server application and of the server functionality were performed by writing and reading different registers of the Serdes devices. For each instance the expected values were written and read from the registers and the correct Alaska status was indicated for each device.

8.2.3 Traffic Generation

Once the client-server application was up and running, we could execute the required MDIO instructions to configure the Alaska devices for the traffic generation. We tested first the transmission and reception of the PRBS and Jitter traffic locally on the Hub1 board plugged into the AdvancedTCA backplane. We had for this two possibilities: to generate traffic internally in the SerDes device (deep loop-back) or externally on the same board and the same chip (external loop-back).

When the deep loop-back was activated in the Alaska 88X2040, the transmitter data in each lane is internally generated, passed through the entire digital circuit, retimed and then looped-back to the receiver PCS. Nineteen devices on Hub1, were configured to send and receive the same type of traffic (PRBS $2^{23}-1$), while the twentieth Serdes was programmed to receive a different type of pattern (PRBS $2^{7}-1$) than its sender was transmitting (PRBS $2^{23}-1$). For all twenty Alaska devices on Hub1, the status and the counter registers were checked. Traffic generation/checker worked without errors when the transmitter and the receiver were configured for the same type of traffic, while errors were reported, as expected, when the transmitter and the receiver were configured for different traffic types.

For test purposes, during the hardware design, two lanes (Lane2 and Lane3) of an Alaska device on Hub1 were externally looped back. The traffic generated on these two lanes, goes outside on the board and comes back to the lanes inputs of the same device. The other two lanes of the same device (Lane0 and Lane1) were connected with a Serdes located on a different card (in slot4) that had not been initialised. By generating traffic on this particular Alaska Serdes and by reading the lanes status, we have seen synchronization and no errors for the externally looped-back lanes, while the status of the counters for Lane0 and Lane1 were showing a lot of errors.

The deep and external loop-back tests performed locally on Hub1 in the AdvancedTCA chassis were therefore showing the expected functionality of traffic transmission between Alaska devices.

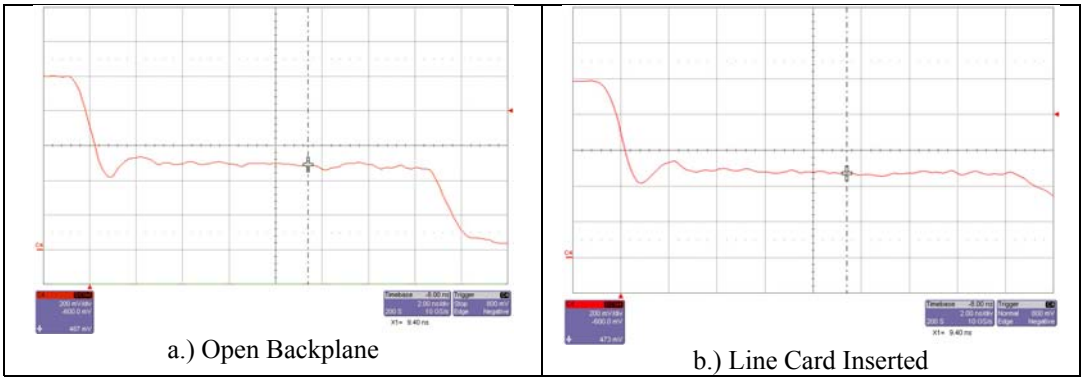
8.3 Backplane Tester Exploitation

8.3.1 Commissioning Problems

Fourteen boards have been manufactured for the Backplane Tester system, as specified in the system design specifications: Hub1, Hub2 and Nodes3 to 14. First problem we discovered when plugging into the chassis was a wrong position of the P19 high-speed connector on any of the boards. As this connector was used for achieving the extended connectivity of extra two differential pairs/directions required by the 10Gbps Ethernet switch architecture (see paragraph 6.1.2.2), it appeared we were not able to test 100 pairs (25Fabric Channels x 2 diff. pairs) within the total of 300 differential pairs on the backplane. That left us with the possibility of testing only the standard interface including

200 differential pairs. This design error can easily be rectified with a respin of the layout. It does not affect the main purpose of the tester since the extended backplane is separate from the standard one which is the primary goal of this validation exercise.

The first validation tests involving traffic generation and transmission across the backplane have been done with Hub1 and Node3 boards plugged into the AdvancedTCA backplane. It soon became obvious that traffic transmission between two 88X2040 transceivers located on different boards showed systematic errors. As we had previously tested the traffic generation in the Alaska 88X2040 and concluded that the traffic generation worked as advertised, the first suspicion was the AdvancedTCA backplane itself. We speculated that this backplane was out of specifications and that signal integrity issues were forcing errors into the system. To seek supporting evidence of this theory and therefore we performed some Time-Domain Reflectometry (TDR) tests. A 1ns rise time TTL signal was injected through a 50Ω coaxial cable to a connector plugged into the backplane. An oscilloscope was set up to measure the signal as it entered the coaxial cable and trigger on the passing negative edge. The resulting signal is shown in Figure 8.2. The display is set so that the center line appears at the end of the coaxial cable. To the left of that line are reflections from the cable, to the right reflections from the backplane. Figure 8.2 a.) shows the TDR measurement of the reflection seen at the Hub1 when driving an empty slot and (b.) the reflection seen at the starting point when a line card inserted at the end.



a.) Open Backplane b.) Line Card Inserted

Figure 8.2: TDR Measurements of the AdvancedTCA Backplane

A 1ns rise time is not sufficiently fast to perform an analysis of true high -peed effects but it is certainly sufficient to determine impedance mismatches, especially on relatively long PCB traces such as the ones being measured. The lack of DC level shift in the signal after exiting the cable and entering the backplane shows that the backplane trace is within a very small margin the same as the nominal 50Ω cable impedance. Equally the extra trace length in Figure 8.2 b.) shows that the trace on the target line card is also closely matched to the nominal impedance. There was no visible perturbation due to the connector or any other possible mismatch on the backplane. In conclusion, if either the backplane or the receive tracks on the line card were not 50Ω then it was by such a small percentage in difference that we could not detect it. After performing these TDR measurements, the possibility of a backplane mismatch had to be excluded.

We still continued performing some tests locally on Hub1 between two Alaska devices by using external loop-back on the on-board high-speed connectors. The twenty Alaska Serdes mounted on Hub1 are divided into two groups of 10 devices, each of them with its

own clock oscillator. We loop-back and test the communication between two Alaska devices on Hub1 for three different situations presented in Table 8.1.

	Valid 8B/10B characters: D21.5, K28.7 or K28.5	**PRBS 2^7-1, 2^23-1 or 2^31-1**
Two independent oscillators, 15cm between the two transceivers (short wires at backplane connector)	No synchronization	No synchronization
Same oscillator, 15cm between the two transceivers (short wires at backplane connector)	Synchronization no error reported	Synchronization, no error reported
Same oscillator 1m between the two transceivers (twisted-pair cable at backplane connector)	Synchronization no error reported	No synchronization

Table 8.1: Transmission Results between two 88X2040 Alaska devices

After performing several tests, the conclusion was that the correct transmission/reception of the traffic was depending on the clocking scheme, the type of traffic and the length of the physical medium. These results were reported to Marvell to understand why the transmission cannot be done between two asynchronous transceivers. Part of the answer was provided: Marvell confirmed that PRBS patterns could not be correctly transmitted between two asynchronous 88X2040 devices because their clock data recovery circuitry simply does not work with those patterns. This limitation was not reported in the 88X2040 datasheet. Equally Marvell had a problem with the special Jitter D21.5, K28.7 and K28.5 characters. When these are transmitted between two Serdes at the XGMII level any drift between the sender and receiver clocks is accommodated by the insertion or deletion of padding characters. There is no logic to do this when operating in standalone test pattern mode without using the XGMII, so the Jitter patterns require synchronous clock operation as well. In conclusion, the communication between two Alaska devices could only run synchronously since in test mode there was no higher level protocol to add or subtract the pad characters allowing for asynchronously running clocks. This was a major and fundamental issue, in the absence of any documentation to the contrary, the assumption had been made that the Serdes in test mode would be as performant as they were in normal active mode.

To address this problem, the solution was to add and integrate into the design the distribution of the same global reference clock towards all the fourteen boards of the Backplane Tester. Since the backplane tester was originally designed to operate in an asynchronous mode, major modifications are required to make it operate in synchronous mode. This was the integration of the clock fan-out board into the AdvancedTCA Backplane Tester system.

8.3.2 Clock Fan-Out Board

A clock fan-out board (Figure 8.3) was been designed and built to distribute the same reference clock for all the boards of the Backplane Tester system.

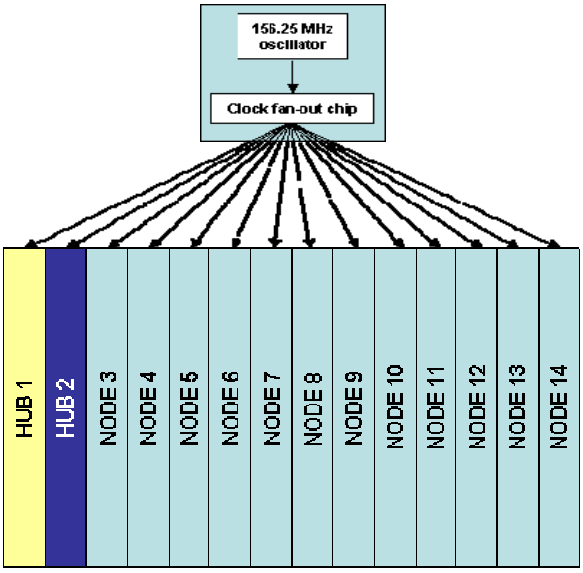


Figure 8.3: Architecture for Synchronous Mode

Instead of being clocked by independent clock oscillators, two 88X2040 devices located on two separate boards belong now to the same clock domain. The fan-out board itself (Figure 8.4) consists of a simple standalone printed circuit platform carrying one clock fan-out chip and a set of 50Ω Lemo sockets. Special attention has been paid to equalizing the lengths of the fan-out tracks, as can be seen from the visible serpentine traces, so that there is no need to compensate for inequalities with different length cables from the fan-out board to the respective node boards. Although the clock signals are 100Ω differential on board, they can be transported over two separate, but equal length, 50Ω single ended cables to the Backplane Tester target board without compromising signal integrity. Figure 8.5 shows the modification done on the individual Backplane Tester boards in order to receive the same external reference clock. Coaxial Lemo sockets are epoxy glued to the board and onboard co-axial cable was brought from them, to the two internal clock fan-out drivers.

Figure 8.4: Global Clock Fanout Circuit

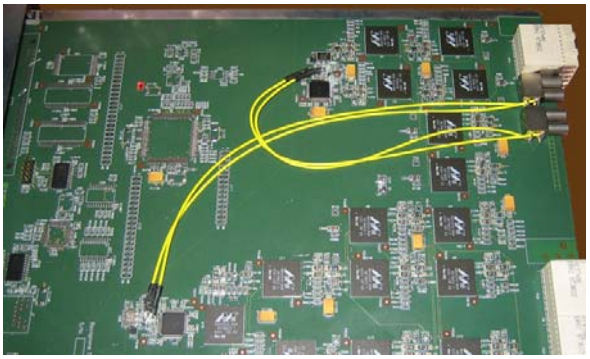


Figure 8.5: Clock Modifications to Hub1 Board

8.3.3 Final Successful Integration

Figure 8.6 shows the front of an AdvancedTCA fully populated chassis. The fan-out clock board was mounted in the open rack above the AdvancedTCA Backplane Tester and the cables were brought from the back of the chassis. Figure 8.7 shows how the clock cables were brought in the back of the AdvancedTCA chassis from the fan-out clock board to all the individual Backplane Tester boards.

Figure 8.6: Fully occupied Backplane Tester Chassis

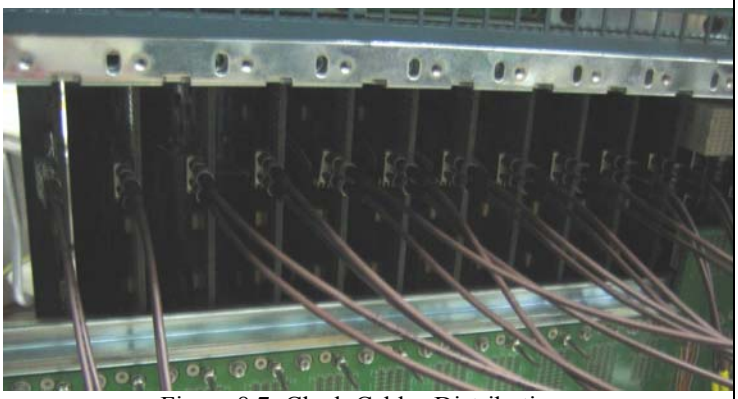


Figure 8.7: Clock Cables Distribution

The integration of the clock fan-out board and the required modifications of the Backplane Tester boards were done 'in extremis' to cope with the unexpected behaviour of the Alaska 88X2040 devices. However, they can easily be incorporated into the basic tester boards design with a re-spin of the layout. The fan-out circuit can be placed on the redundant Hub2 board and the Base Interface for that board used to distribute the clocks. There would be a slight asymmetry because the control plane from Hub1 to Hub2 was already used for the MDIO control, but in this particular case we could use the update channel instead.

Once the synchronous system has been setup, we quickly established that the global clock distribution works as well as we had hoped. Measurements made of the resulting clock distribution chain showed that concatenating two fan-out chips, the central one plus the on-board one, did not significantly deteriorate the received clock quality and the whole system worked as expected. The Serdes were insensitive to small phase differences due to variations in supply cable lengths. The essential thing was that all have the same clock frequency. With the clock fan-out board in place and the new system setup, we were then able to progress to the exploitation of the AdvancedTCA Backplane Tester.

8.3.4 BER Measurements

Once the AdvancedTCA Backplane Tester system was setup and working, we performed a series of BER tests by sending traffic for several hours under different conditions on all 200 differential pairs available in the dual-star Fabric Interface.

One of the main reasons for choosing the Marvell Alaska as a traffic generator and checker on the AdvancedTCA backplane was its programmability features. We could program the Alaska Serdes to output Jitter or PRBS patterns with a wide range of values corresponding to different amplitudes, pre/de-emphasis levels and frequencies of operation. We tested the transmission across the AdvancedTCA backplane for different values of those parameters in order to establish an envelope of functionality for the working condition of the AdvancedTCA backplane. BER tests have been performed for four different patterns (PRBS $2^{23}-1$) and Jitter – low frequency - 48A.2 and – mixed frequency - 48A.3) at two different frequencies of operation (1.25Gbps and 3.125Gbps) and for each set of tests were considered three different values for the amplitude (700mV, 900mV and 1100mV) and de-emphasis (0%, 60% and 300%). It is useful here to recall the discussion of pre and de-emphasis in section 5.4.4.5.

The results of those tests are shown in the Table 8.2, where each cell of a matrices represents one BER test.

	1.25 Gbps					**3.125 Gbps**			
48A.2	**Amplitude [mV]** **De-emphasis [%]**	**700**	**900**	**1100**		**Amplitude [mV]** **De-emphasis [%]**	**700**	**900**	**1100**
	0	PASS	PASS	PASS		0	PASS	PASS	PASS
	60	PASS	PASS	PASS		60	PASS	PASS	PASS
	300	PASS	PASS	PASS		300	PASS	PASS	PASS
48A.3	**Amplitude De-emphasis**	**700**	**900**	**1100**		**Amplitude De-emphasis**	**700**	**900**	**1100**
	0	PASS	PASS	PASS		0	PASS	PASS	PASS
	60	PASS	PASS	PASS		60	PASS	PASS	PASS
	300	PASS	PASS	PASS		300	FAIL	FAIL	FAIL
PRBS 2^23-1	**Amplitude De-emphasis**	**700**	**900**	**1100**		**Amplitude De-emphasis**	**700**	**900**	**1100**
	0	PASS	PASS	PASS		0	PASS	PASS	PASS
	60	PASS	PASS	PASS		60	PASS	PASS	PASS
	300	PASS	PASS	PASS		300	FAIL	FAIL	FAIL

Table 8.2: The Results of the BER Tests

Consider first, the top row in each of the six matrices in Table 8.2. There was no pre/de-emphasis added to the original transmitted signal and by moving from left to the right, the programmed amplitude is increased from 700mV to 900mV and then 1100mV. It could be seen that for 0% of de-emphasis all tests performed were successfully passed.

It is interesting now to observe what happens by increasing the degree of de-emphasis to 60 and 300%, as shown in the second and third rows of the matrices. In the case of PRBS and 48A.3 Jitter pattern at a bit rate of 3.125Gbps, the bit errors are counted for a maximum level of 300% de-emphasis. These errors are due to a combination of excessive amount of de-emphasis and attenuation in the signal spectrum of the high-frequency components contained in those patterns. No errors were detected for 48A.2, which is the low frequency pattern and hence the less constraint.

Error free operation has been observed for lower levels of pre/de-emphasis. This was a proof that the loss due to the AdvancedTCA backplane is limited and so no de-emphasis effect is required to present a sufficiently open eye. Only in the 'extreme' conditions of very high-level of de-emphasis, the BER tests for PRBS and 48A.3 started showing any transmission errors on this particular backplane.

Since we don't really need pre/de-emphasis to work with this quality of backplane we could instead use it to artificially reduce the eye-diagram to the point where small anomalies in backplane construction will cause the eye to close sufficiently to create bit errors. We could then use the AdvancedTCA backplane tester for exploring backplane behaviour under sub-optimal conditions and thus establish an envelope of functionality for this backplane. This way, we obtained a quantitative and a qualitative estimate for the margin of security on the AdvancedTCA backplane under test.

8.3.5 Signal Integrity Evaluation

8.3.5.1 High-Speed Measurements

We performed measurements along the backplane to highlight high-speed signal integrity aspects as a cross-check of the previous software BER tests and to establish a degree of credibility in the simulations performed for the AdvancedTCA backplane.

For performing accurate high-speed measurements, it is very important to have adequate equipment. The following measurements have been performed with a Lecroy Serial Data Analyzer (SDA 5000A) [LEC-01] with an analog bandwidth of 5GHz, used together with a 6GHz differential probe (WaveLink Differential Probe D600A-AT) [LEC-02].

8.3.5.2 Test Conditions

The Marvell Alaska Serdes functionality has already been tested and validated in chapter 4 (see paragraph 4.3). The circuit allows 10Gbps transmission and reception to be shared over four differential pairs, each of them running at 3.125Gbps. Each 3.125Gbps signal is generated in the Alaska device, fanned-out and transmitted from Hub1 across the

backplane. We have done the measurements on a high-speed connector which can be inserted in any of the slots along the AdvancedTCA backplane. The ending signal pair on this connector was terminated with 100Ω resistive termination, intended to remove possible reflections at the receiver.

For test purposes we used the 48A.3 (K28.5) Jitter mixed frequency test pattern, commonly specified for measurements in Ethernet system operating between 1Gbps and 3.125Gbps. K28.5 is a special character in 8B/10B-coding table which contains a repetitive sequence of five consecutive 'ones', five consecutive 'zeros', but it also includes an isolated 'one' ('zero-one-zero') and an isolated 'zero' ('one-zero-one'), intended to stress the low-pass filter capability of the transmission channel. The measured signal waveforms corresponding to this pattern are shown in Figure 8.8.

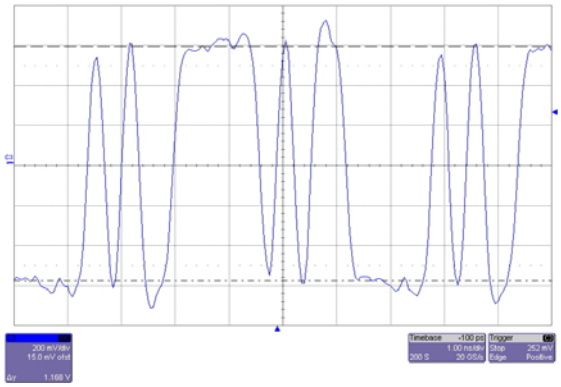
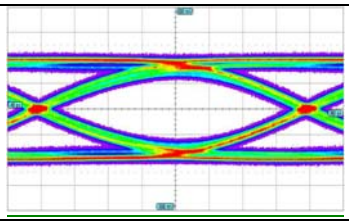
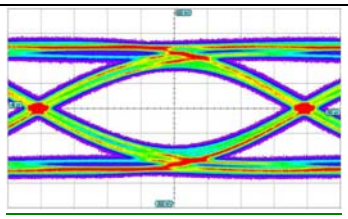
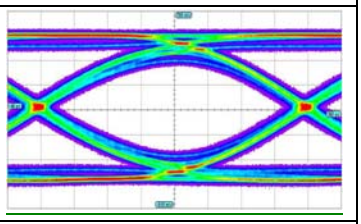


Figure 8.8: Simulated vs. measured waveforms

The Lecroy measurement equipment used for these measurements could adequately cover only up to the third harmonic in the 3.125Gbps signal under test. As the rise and the fall times of the signal depends mostly on the third and the fifth harmonic, some degradation of the signal edges was expected. We could indeed notice some rise/falling time degradations due to the scope and the probe limitations. For instance, the measured waveform cannot resolve a plateau at the middle of the bit period, where for a 'zero-one-zero' transition, the signal just reaches the design amplitude before falling again. Despite the limitation in the equipment, we could establish a degree of confidence in the fact that the programmable bit pattern, signal amplitude and de-emphasis were correctly obtained.

8.3.5.3 Effect of Pre/De-Emphasis

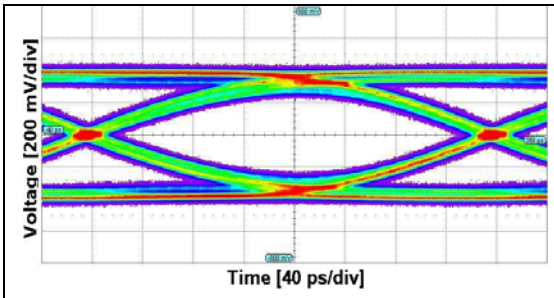
We captured eye-diagram representations and used the scope in statistical mode to measure their corresponding eye-openings. First we performed measurements on the longest tracks on the AdvancedTCA chassis, having the probe attached to the high-speed connector which is plugged into slot14. The transmitted signal had 0% of pre/de- emphasis and different levels of amplitude. The eye-diagram at the receiver is shown in Figure 8.9.

Amp [mV]	700mV	900mV	1100mV
PE=0 [%]			
Figure 8.9: AdvancedTCA Backplane Measurements (Amplitude benefits)			
	a.) Eye-opening=510 mV	b.) Eye-opening=574 mV	b.) Eye-opening=720 mV

By increasing the amplitude at the transmitter, the vertical eye-opening at the receiver was also increasing. In all the three cases above, a more than sufficient eye-opening is present at the receiver, allowing for any Alaska device to reconstruct the received signal without any errors. This first set of measurements confirms the previous BER tests which showed zero errors for the 48A.3 pattern. Obtaining such a good signal quality with such little loss when no pre/de-emphasis is added to the transmitted signal, is a sign of good quality of the AdvancedTCA backplane.

This measurement also confirms that the simulations reported in Chapter 5 were predicting the correct order of magnitude effects. If we compare Figure 5.29 a.) which was for an 1100mV signal received after 14" of FR4 we have a simulated eye opening of 810mV compared with the measured one here of 720mV. Bearing in mind that the measured result is reduced by the bandwidth constraints of the oscilloscope this is a good comparison.

The next test is to obtain measurements for the de-emphasis to compare with the simulated effects reported in Chapter 5. We reproduced the conditions simulated and shown in Figure 5.33. In order to highlight the improvements at the receiver achieved through this effect, we had to compare two signals with the same nominal amplitude, such as a 700mV non-emphasised signal with a 1100mV signal de-emphasised by 60%. Figure 8.10 shows the eye-diagram measurements at the receiver made in slot14 for these two transmitted signals. In the simulation the nominal amplitude was 700mV and 660mV with eye openings of 495 and 750mV respectively, a gain of 51.5 %. In the measured case we have nominal amplitudes of 690 and 720mV with openings of 510 and 600mV respectively, a gain of only 17%.



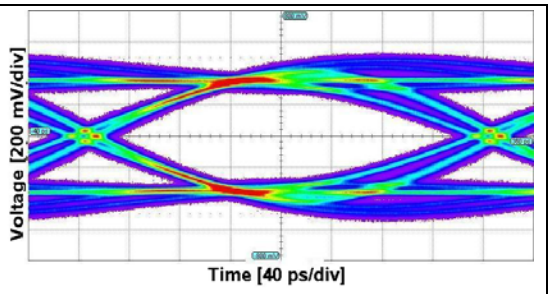


Figure 8.10: AdvancedTCA Backplane Measurements (Pre/De-emphasis benefits)

a.) Transmitted amplitude=700mV De-emphasis=0% Eye-opening=510 mV Eye-amplitude=690 mV	b.) Transmitted amplitude=1100mV De-emphasis=60% Eye-opening=600 mV Eye-amplitude=720 mV

Trying to understand where the differences between our previous simulations in chapter 5 and the measurements performed for the AdvancedTCA backplane come, we have to consider the dielectric characteristics which were used in both cases. In the simulation the dielectric FR4 had a dielectric constant $\varepsilon r = 4.3$ and a loss tangent lt=0.022, while the AdvancedTCA standard specifies for the same FR4 $\varepsilon r = 4$ and lt=0.018. In order to quantify the impact of those differences, we consider the results presented by [Johnson-03] in Figure 8.11, which gives an estimation of the loss percentage per inch function of the dielectric loss tangent.

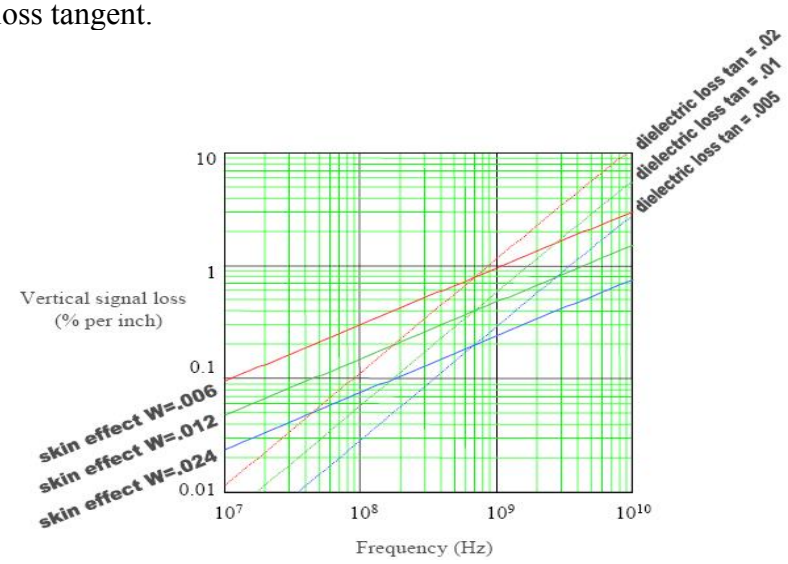


Figure 8.11: Signal loss function of dielectric loss tangent (reproduced from [Johnson-03])

Considering the two values of loss tangent (simulation versus measurement) in the portion of the figure between 1GHz and 10GHz, we note that for small variations in dielectric loss there can be as much as 2% or more signal loss per inch. Given that our tests were run over 14 inches of FR4 this could give rise to as much as a 30% of difference in received amplitude which is what we observed (by comparing Figure 5.33 with Figure 8.10). We therefore will use the new range of loss tangent and permittivity specified for the AdvancedTCA backplane for further simulation comparisons which are reported in paragraph 8.3.6.

For the Figure 8.10, the only truly comparable values are the launch amplitudes and the non de-emphasised eye opening. This would seem to indicate that the FR4 used in practice does not have the same amount of loss as that used in the simulation model. Indeed looking more closely at Figure 8.10 b.) it shows that there is too much de-emphasis for optimal results with this track because there is significant overshoot of the signal above and below nominal signal swing. We further tried combinations of different nominal voltage and de-emphasis levels at the transmitter, and measure the received signal on the high-speed connector in slot 14. The best results were obtained for 900mV signal amplitude with 33% de-emphasis. The eye-diagrams corresponding to those signals measured at the receiver are shown in Figure 8.12.

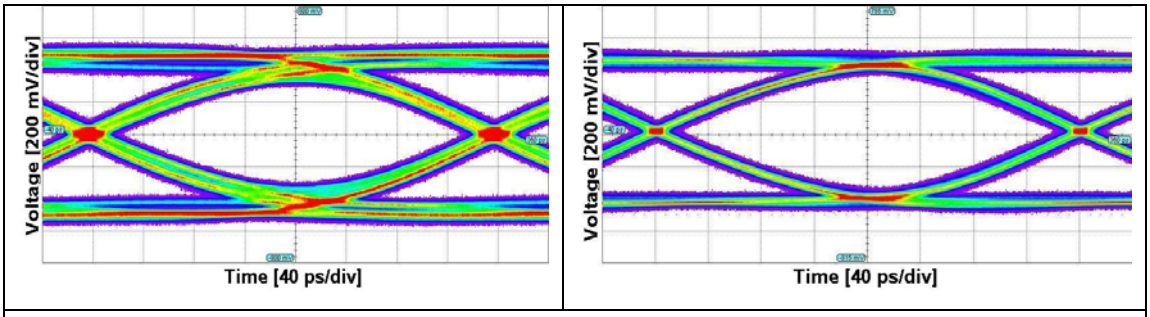


Figure 8.12: AdvancedTCA Backplane Measurements (Pre/De-emphasis benefits)

a.) Transmitted amplitude=900mV De-emphasis=0% Eye-opening=575 mV Eye-amplitude=850 mV	b.) Transmitted amplitude=1100mV De-emphasis=33% Eye-opening=740 mV Eye-amplitude=820 mV

In this case, an improvement of 28.7% was achieved by using de-emphasized signal with an eye opening of 740mV and no overshoots resulting in a clean and balanced eye-diagram. This shows the extent to which the actual effects of de-emphasis are hard to predict with simulation since the real values for loss in any given material can only be estimated ahead of production. However these measurements also demonstrate that even in a case where the loss is so low that pre/de-emphasis is not strictly necessary, it is still possible to achieve substantial improvements in the eye opening and hence a greater degree of security for error free operation by employing an appropriate value of de-emphasis. Finding that value is however an empirical process for each and every track.

Finally we used the effect of extreme de-emphasis to degrade the signal to the point where errors would start to occur. In the limit we can require up to 300% of de-emphasis which for a first bit swing of maximum 1100mV yields a nominal bit swing of only 366mV for the subsequent bits in the pattern. For comparison it is shown in Figure 8.13 the measured signal transmitted with the amplitude of 1100mV, without de-emphasis (a.), respectively with a maximum of 300% added de-emphasis (b.).

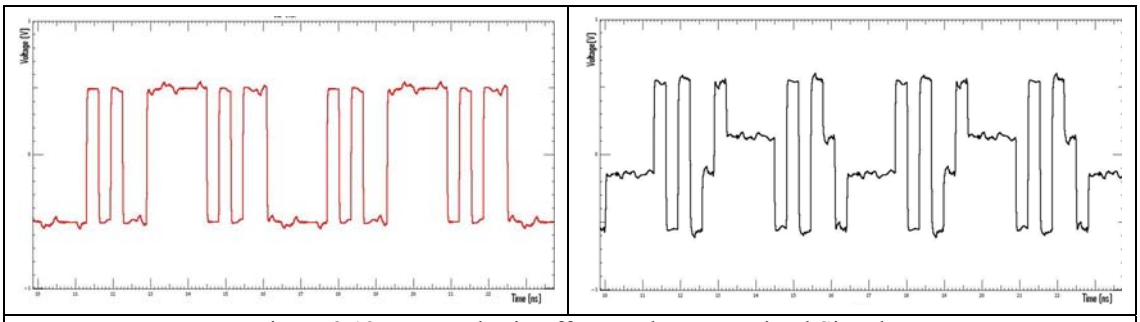


Figure 8.13: De-emphasis Effect on the Transmitted Signal

a.) 1100mV and 0% De-Emphasis | b.) 1100mV and 300% De-Emphasis

The de-emphasis of 300% has the effect of reducing the transmitted signal to much less than the nominal 400mV required by the XAUI standard. The eye at the receiver for this de-emphasized signal is shown in Figure 8.14.

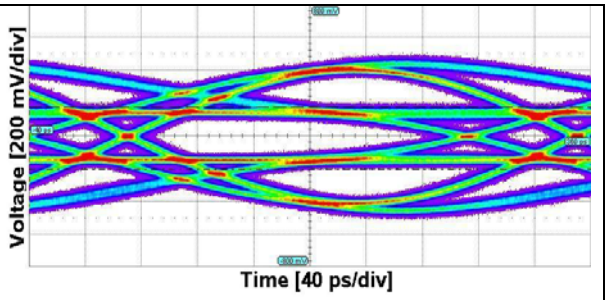


Figure 8.14: AdvancedTCA Backplane Measurement (De-emphasis=300%)

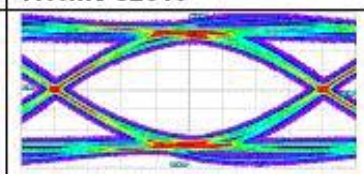
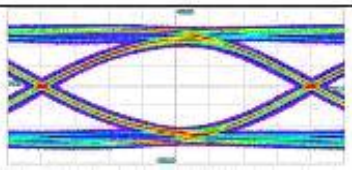
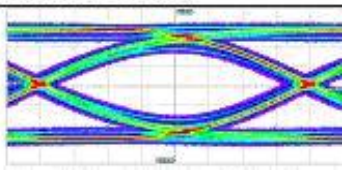
Transmitted amplitude=1100mV
De-emphasis=300%
Eye-opening=220 mV
Eye-opening=300 mV

The eye-opening has been artificially reduced to only 220mV together with an obvious signal overshoot clearly seen in the eye-diagram. Only under these extreme conditions of de-emphasis did the system start to generate errors as detected by the BER tests and reported in table 8.2.

In this section we have presented a study of the pre/de-emphasis effects. We have measured the real results and compared them with the simulated results from our earlier modeling. The comparison is favorable, apart from the obvious differences caused by the lower loss in practice than had been used in modeling. We observed a considerable improvement that can be made by involving a certain small amount of de-emphasis in the signal transmission. Error free operation has been observed for eye-openings corresponding to lower levels of de-emphasis. By making use of the de-emphasis effect, we could explore backplane behaviour under sub-optimal conditions. Observing errors on the backplane only in extreme conditions, demonstrates not only that the backplane is well defined and manufactured but that there is enough headroom to achieve error free operation and that it clearly exists the possibility of achieving the much higher rates that are described in the AdvancedTCA roadmap.

8.3.5.4 Effect of Track Length

All the eye-diagram measurements presented previously have been performed for the longest distance on the AdvancedTCA backplane, on the high-speed connector inserted in slot 14. In this paragraph, we will be looking in somewhat more detail at differences in transmitted quality as the differential traces along the backplane get longer by moving from adjacent, to increasingly distant, slots. Figure 8.15 shows the received eye measured after passing through 3 different distances of AdvancedTCA traces: Slot3 (shortest distance), Slot7 (intermediate distance) and Slot14 (longest distance). For all these measurements, the transmitted signal has the maximum amplitude (1100mV) and 0% of de-emphasis. The respective eye openings at the receiver are given below.

Slot 3	Slot 7	Slot 14
PE 0 		
Figure 8.15: Eye-Diagrams measured over three different distances on the AdvancedTCA backplane		
Eye opening = 992mV	Eye opening = 749mV	Eye opening = 716mV

We observe a successive decrease in the eye-opening with the increase of the distance, and hence the loss, along the backplane. However, in all cases there is more than sufficient eye opening for any Serdes to reconstruct the signal with zero BER.

8.3.5.5 Crosstalk

Tests until now had concentrated on the signal integrity of one differential pair while all the other pairs were running on the backplane concurrently. In the current paragraph, the signal quality was measured for a chosen differential pair when running alone (Figure 8.16 b.) or in the presence of all available pairs on the backplane running simultaneously (Figure 8.16 a.).

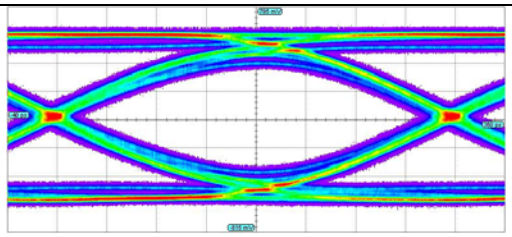
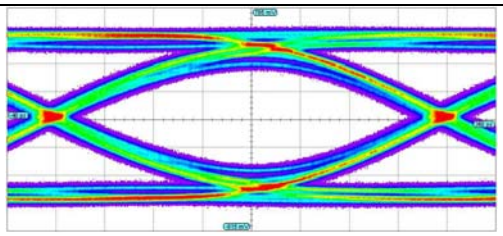
PE	a.) All pairs running	b.) One pair running
0 [%]		

Figure 8.16: Effect of crosstalk on the AdvancedTCA backplane

As can be seen there was no easily visible difference between the two conditions. Without de-emphasis we observe an eye opening of 772 mV for the single link running that only reduces by less than 4% to 740 mV when all links are running. We added

de-emphasis to reduce the eye to see if the effect would be more visible but in practice the already small difference disappears altogether.

8.3.6 Simulations versus Measurements

We have already drawn attention to the deviation of some results from the simulated predictions (see paragraph 8.3.5.3). This was due to the differences in the characteristics of the dielectric medium and the dimensions of the tracks defined in the AdvancedTCA specifications. We return to the model we developed before adopting the new standard, incorporate the changes and then compare the predictions with reality.

8.3.6.1 Effect of Pre/De-Emphasis

The transmission channel used for simulation is similar to that in Figure 5.24 and it contains 2 connectors and four connector pin-field regions whose model was developed by us. The difference now consists in a new W-element model used for the PCB trace. Using the new model we simulate one track and compare the receiver input signal with measurements taken from the backplane as shown in Figure 8.17.

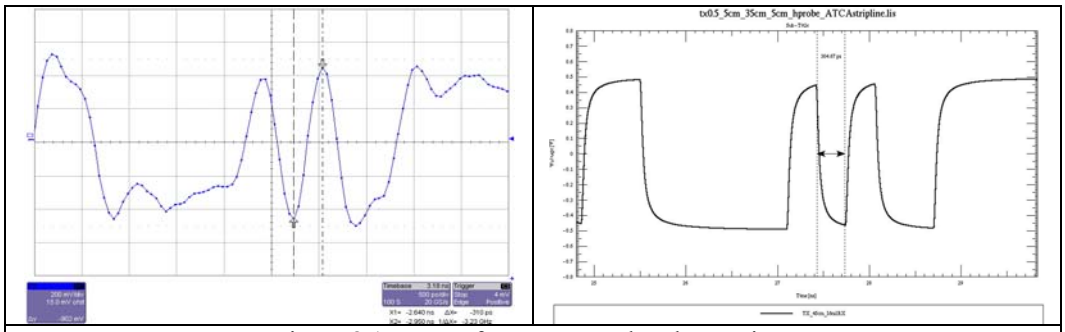


Figure 8.17: Waveform measured at the receiver

a) Measured	b) Simulated Response

We can see that the waveforms corresponding to the transitions ‘zero-one’ or ‘one-zero’ have absolutely identical, somewhat artificial shapes in the simulated signal, while in the measured signal the swings have slightly different amplitudes. The eye-diagrams corresponding to the signal above are shown in Figure 8.18.

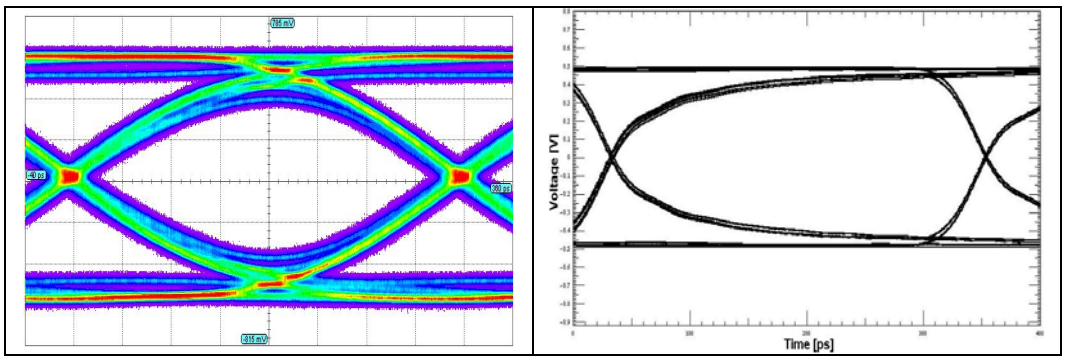


Figure 8.18: Measurement vs. Simulation – eye-diagram

As we can see, the simulated and the measured eye does not look at all the same. The simulation itself contains more high-frequency components than the actual measurement which is bandwidth limited by the oscilloscope used for the measurements. We decided therefore to pass the simulated signal through a low pass filter with the cut-off frequency corresponding to 5GHz bandwidth of the oscilloscope we were using for this acquisition. We used a low-pass finite impulse response filter having the sampling frequency corresponding to the simulation data step, the pass frequency of 3.125GHz and the stop frequency of 5GHz.

Figure 8.19 shows the comparison between the same measured signal as in Figure 8.17 and the simulated waveform obtained at the output of the low-pass filter.

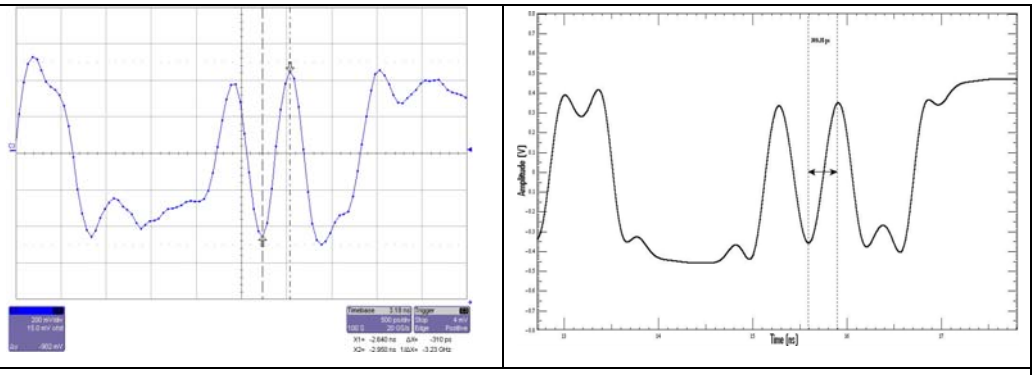


Figure 8.19: Waveform measured at the receiver

a) Measured	b) Simulated Response after filtering

By filtering out the high-frequency components present in the simulation, the result now looks much more comparable with the measurements. Using this technique we can now take the simulation results performed for transmitted signals with and without de-emphasis, pass them through the filter and plot their eye-diagrams for comparison with the measurements shown in Figure 8.12. Figure 8.20 shows the processed simulations for receiver eye diagrams with 900mV signal amplitude and, respectively 0% and 33% de-emphasis.

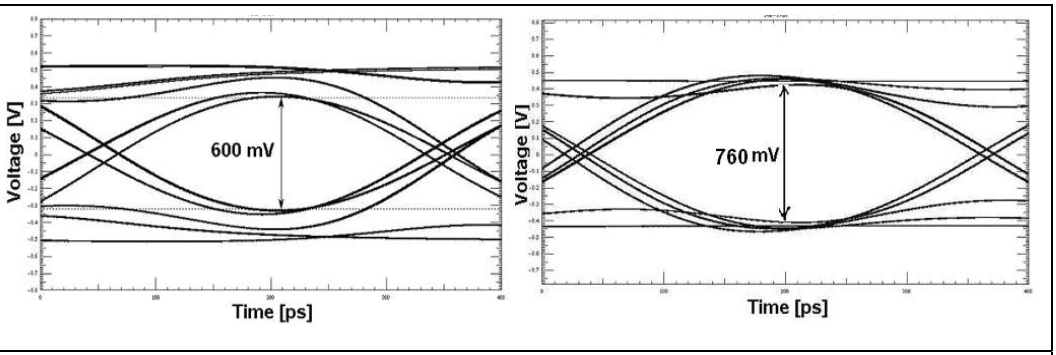


Figure 8.20: Simulated eye-diagram

a.) 0% De-emphasis Eye Opening = 600 mV	b.) 33% De-Emphasis Eye Opening = 760 mV

These figures can fairly be compared with the measured results given in Figure 8.12 which gave eye openings of 575mV and 740mV respectively. The measurements for the eye-opening were showing slightly reduced amplitudes due to more loss and signal

integrity interactions existing in the real environment. However, the values obtained through simulation are well within the expected range of performance.

8.3.6.2 Effect of Track Length

We also performed simulations for three different lengths of PCB trace passing across the transmission channel in Figure 5.24. The values were 10cm, 25cm and respectively 50cm and we consider them as being the correct ones corresponding to slot3, slot7 and slot14 in the AdvancedTCA backplane. The eye-diagrams, plotted after being passed through the low-pass filter, are shown in Figure 8.21 and the eye openings compared with the measurements reported in Figure 8.15.

Slot 3: 10cm	**Slot 7: 25cm**	**Slot 14: 50cm**

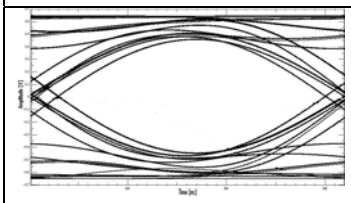
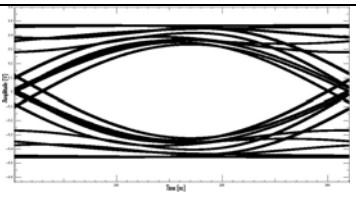
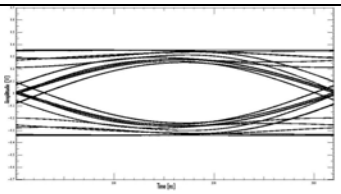

		

Figure 8.21: Eye-Diagrams measured over three different distances on the AdvancedTCA backplane		
Simulated eye = 1020mV Measured eye = 992mV	Simulated eye = 775mV Measured eye = 749mV	Simulated eye = 725mV Measured eye = 716mV

As in the previous case, the simulations show slightly more optimistic results than has been measured, but still within 5%, which is well within the bounds of experimental error.

8.3.6.3 Conclusions on the Simulation-Measurement Correlation

We found a proper methodology to achieve a proper comparison between the measurements and the simulations by passing the simulated waveform through a low-pass filter with the frequency cut-off of 5GHz.

The simulations show slightly more optimistic values that it has been measured in practice. In despite of these small differences however, we could conclude that the values of the eye openings are still within 5%. These differences appear most probably because the simulation takes into account the transmission channel over the backplane and only the closest aggressing neighbours, while in practice the realistic environment on the backplane contains hundreds of differential pairs of traces running close together. We have measured that effect which appears on the backplane (see paragraph 8.3.5.5). It is very small, less than 4% but it exists and can be a cause of the reported difference noted between the simulations and the measurements.

8.3.7 Effect of Adding Further Loss

The low level of loss noted in the AdvancedTCA backplane was unexpected, mainly because it is built with standard FR4 dielectric material which usually does not show good loss characteristics. We therefore carried out a small experiment, a test of available 'headroom', to see just how far it would be possible to drive the switching fabric before loss set in and what the application of pre/de-emphasis could do to solve it.

We constructed one point-to-point connection using dual co-axial cables to propagate the differential signal 97cm away from the slot 14 of the backplane and bring them to be terminated with 50Ω each in two of channel inputs of the oscilloscope. The cable used is commonly employed at CERN for high frequency interconnects. It is of type 96-IEC-50-2-1 and has the following characteristics:

Characteristic Impedance	50 ± 2	[Ω]
Rated Capacitance	101	[pF/m]
Propagation Rate	0.66	
Rated Delay	5.03	[ns/m]
Maximum Attenuation	10 100 200 9.5 29 42	[MHz] [dB/100m]
DC Resistance of Inner Conductor	126	[m Ω /m]
DC Resistance of Screen	40	[m Ω /m]

Table 8.3: Characteristics of the cable type 96-IEC-50-2-1

In a second experiment, we constructed another connection using dual Gore co-axial cables 140cm long and terminated in 100Ω and measured the signal at the termination with the high frequency probe. Details of this cable are not available since Gore is fairly secretive about their proprietary technology but an indication can be found at [Gore].

The resulting eye diagrams for varying levels of de-emphasis are shown in Figure 8.22. Note that due to the auto-ranging of the eye-diagrams, the vertical scale in this measurement is not constant. Column a.) in Figure 8.22 show the eye-diagram as previously measured at slot14 on the backplane. The eye is well open but deteriorates as the de-emphasis is increased. Column b.) shows the eye-diagram of the signal which from slot14 has been further propagated 97cm over the coaxial cable. With zero de-emphasis the eye is still open enough to use with nearly 200mV opening and with 60% of de-emphasis this is doubled to over 400mV. Increasing the de-emphasis further however only has a negative effect. Column c.) shows results for 140cm of cable at the end of which the eye is completely closed. With 60% de-emphasis it has been opened sufficiently to be useful again with 86mV of opening. This also deteriorates with further de-emphasis.

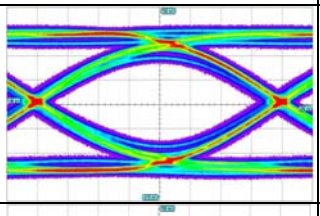
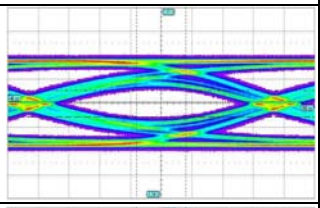
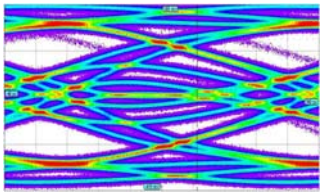
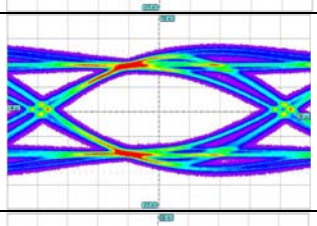
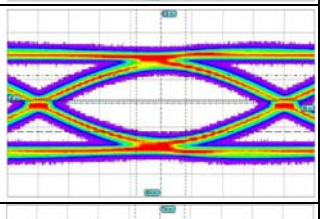
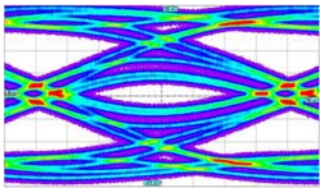
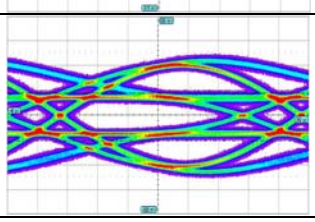
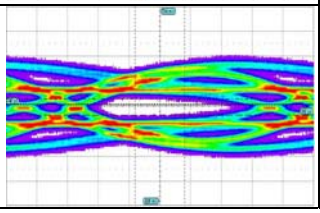
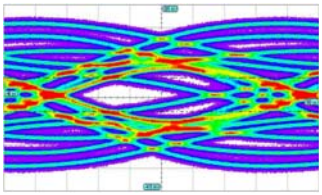
PE DE [%]	a.) Slot 13, no extra cable	b.) Slot 13 + 97 cms cable	c.) Slot 13 +140 cms cable
0			
60			
300			

Figure 8.22: Eye diagrams with and without extra cable losses

In this paragraph we have shown the improvement that can be achieved by using de-emphasis to compensate for high-frequency loss introduced by the cable environment. The point here was not in testing the cable quality but in highlighting the effect of the de-emphasis for the transmission on the backplane. The conclusion was that with the combination of the Marvel transmitter and the well engineered backplane, the Hub to Node communication not only works over the longest distance in the AdvancedTCA chassis, but that it has sufficient headroom in reserve to continue over more than a meter of external cable and still yield a recoverable signal at 3.125Gbps.

8.4 Summary

The present chapter includes the testing, the validation and the exploitation of the AdvancedTCA Backplane Tester boards.

Due to some unexpected problems and limitations of the Marvell device when used in test mode, we had to transform the initial asynchronous tester design into a synchronous system, by distributing to all the boards in the chassis a single global clock. To achieve that, we have designed, built and integrated a fan-out clock board into the Backplane Tester system. We were able to measure that the backplane met its performance targets without error. We were then able to make measurements of the achieved levels of signal integrity on the backplane itself and compare these with our simulated predictions. We have established that our simulation technique provided predictions accurate to within 5% for most measurements.

We made use of the programmability of the Alaska Serdes (voltage and de-emphasis) in order to compensate for losses of the transmission medium. We have quantified in this way the de-emphasis circuitry performance of the chosen Serdes technology and use this technique to explore backplane behaviour under sub-optimal conditions. We established an envelope of functionality for this backplane, by exaggerating the programmable settings up to point where artificially low thresholds of operation occur in the system. We could conclude that the AdvancedTCA Backplane Tester completely fulfils its requirement as an instrument for quality control which can establish an estimate of the headroom available on any particular backplane.

9 Conclusions and the Author's Contributions

9.1 Conclusions

This thesis reports on research and development carried out within the EU funded ESTA project [ESTA] whose goal was to develop the necessary architecture and modules with which a scalable Ethernet switch could be built. The goal of the project was also to examine the limits of interconnects' performance with a view to addressing future generations of networking equipment. To achieve this, a demonstrator switch with eighteen slots, capable of handling up to 32 10 Gigabit links, was to be built and used to quantify to what extent the predicted performance goals had been met.

The performance of the backplane of the switch was identified as a potential weak point of the system. Handling high-speed data operations inside silicon chips is a well understood and manageable problem. However once signals are being transmitted from one module to another over complicated interconnects, through multiple connectors and in close proximity to many other high-speed signals, then they become subject to amplitude and frequency loss as well as receiving external noise and crosstalk. This is a known problem even for previous generation technologies and solving it demands special attention to all aspects of design. This thesis focuses on the research done, and the methods employed, to develop a backplane design [OLT-R2] that would not only meet the required specifications but also permit exploitation at higher speeds [ESTA-R2]. To prove that the practical implementation of the design both met and exceeded requirements, a test facility was developed and measurements taken from it were used to validate the predicted results. This tester was also used to ensure that there were no defects in the production of the prototype backplanes which in turn allowed for the error- free operation of the demonstrator switch.

The interconnects of a switch design are a key element since they must not only deliver the required connectivity and bandwidth but do so with a performance that does not permit the introduction of systematic errors. At the time the study was undertaken, there had been proofs of concept and vendor demonstrations of prototype products that could successfully drive up to eight differential pairs of signals over different dielectric media, from one transmitter card to one receiver card. There had been no systematic study however of how a backplane supporting up to eighteen daughter-cards and containing several hundred asynchronously communicating links, all running concurrently, would perform.

The first part of the present thesis proposes and validates a solution for such an eighteen slots backplane design which is supposed to accommodate the 10 Gigabit Ethernet switch prototype built within ESTA. The main goal was to study the performances of the high-speed interconnects existent on the switching backplane and to deliver an adequate methodology to validate the proposed backplane interconnect design after employing an extensive signal integrity analysis.

In addition to the high-speed signal integrity aspects, the copper interconnects are limited by the technology (of connectors and transceivers), by the backplane materials and by the

PCB manufacturing process itself. We chose standard FR4 on the grounds of cost performance and since other studies suggested that it was capable of supporting the signal speeds required. It is however subject to large variations in mechanical and dielectric properties and these variations were examined in detail and taken into account in the design. Connectors are another critical component and several connector specifications proposed by different manufacturers have been carefully analysed from the electrical and mechanical point of view. The most appropriate connector solution was then chosen for our backplane design. On the silicon market, most of the components were state of the art pre-standard 10Gbps components. However the collapse of the telecom market in 2002 left us with a restrictive set of choices. We chose the 10 Gigabit Ethernet transceiver from Marvell (Alaska 88X2040). We carried out a careful analysis to determine its limits of operation and not only rely on the publicity material provided by the company itself [ESTA-R1]. The silicon itself performed very well and we even established that the design will hold without modification for a factor four increase in bandwidth provided that chip packaging can be developed to match the required speeds.

Achieving well-controlled impedance values for the interconnects on the backplane and ensuring good signal integrity practice for them is essential for assuring reliable transmission at high-speeds. Hundreds of differential pairs carrying 3.125Gbps signals were routed on the backplane in order to make possible the connectivity between the sixteen line-cards and the two redundant switch-cards required by the design. The backplane interconnect design was not a trivial issue since it involved making continual trade-offs between what was 'desirable' and what was 'possible', especially when interconnects had to be routed within a limited amount of space between the conector footprint in the region of switch fabric cards. We identified this connector pin-field area as being the worst case source of signal interference in the design and studied the question in depth.

We have carried out detailed simulations of the transmission properties of complex structures to within the limits of the existing tools [OLT-R1]. We first developed and calibrated a model for the connector pin-field using the 3D full-wave electromagnetic field solver Ansoft HFSS. In order to quantify the influence of the connector pin-field regions on the overall transmission on the backplane, we performed further analysis in an HSPICE system simulation environment. Worst case scenarios were considered from the point of view of the routing topology, the aggressor signals, the number of connector pin-field sections and the length of the PCB interconnect. In the first part of the thesis, we developed and presented a methodology, which allowed us, through full-wave modeling of the connector pin-field and simulation of a very realistic 10Gbps backplane environment, to perform a pre-layout validation study of the proposed 10Gbps Ethernet backplane design. This way, we have validated the set of proposed interconnect design criteria and established that they would allow for the manufacture of a backplane with a high level of reliability within current technologies.

When the first release of a new standard for telecom applications (Advanced Telecom Computing Architecture) became available, we compared its description with the design criteria we have derived for the 10 Gigabit Ethernet switch design and found them essentially equivalent. The use of the AdvancedTCA backplane was therefore adopted by the ESTA project [OLT-UPB].

With the simulation and design process completed, attention was then turned to the issues of validating the theoretical results and measuring the achieved performance. We also had to estimate the extent to which achievable performance exceeded requirements and finally to ensure that the prototype backplane had been built to specification. All of these goals can be achieved by the deployment of high performance Serdes operating in test mode, for every link, simultaneously.

The second part of the thesis includes therefore the development and presentation of the design for the AdvancedTCA Backplane Tester system ([OLT-PGNET05], [OLT-RTC05]). We have designed, built and deployed this system in order to test up to 300 concurrent data streams, all running at up to 3.125Gbps over the AdvancedTCA backplane. This design is considered as unique on the market at the time it was conceived. The tester uses the same driver and receiver technology, and at the same speed, as would be employed by the switch fabric in the 10 Gigabit switch design (Marvell Alaska 88X2040). Each link is monitored for errors, and the signals could be modified under command, to establish the limits within which error free operation could be established and sustained on the AdvancedTCA backplane.

During the first tests, we encountered unexpected problems which were due to limitations of the Marvell chip when used to generate 10Gbps patterns in test mode. This problem was not mentioned in the device datasheet and the solution involved the design of an extra board to fan-out a global clock and to transform the original asynchronous design into a synchronous system. Once the system was working correctly it was employed to test the prototype backplane at full speed, and full load, and report on any transmission errors encountered. Extensive measurements of the achieved signal integrity were undertaken and the results evaluated.

The measurements showed very little loss in the dielectric material and little evidence of any crosstalk caused by the concurrent operation of multiple links on the AdvancedTCA backplane. To obtain the quality estimation for any particular AdvancedTCA backplane we used the pre/de-emphasis circuitry of the Marvell Alaska technology to introduce artificially low thresholds of operation so that any significant PCB manufacturing anomalies would result in detectable error conditions. We demonstrated that the AdvancedTCA backplane could perform well within the requirements of the standard and was capable of delivering the specified cross-sectional bandwidth with an impressive margin.

The original simulations had taken a worst case value of permittivity for FR4 to ensure that the design would meet specifications irrespective of the FR4 quality. This made it difficult to compare simulations with measured results so the simulations were repeated using the boundary limits as defined in the Advanced TCA standard. The measurements taken were then compared with the updated predictions. These showed close correlations to the practical measurements. We could establish this way that the methodology proposed in the first part of the thesis was valid and could deliver a backplane interconnect design that sustains error free operation for today's available technologies.

This completed the design cycle which started from a theoretical analysis of the problem set, simulation of the worst case scenario's and then the successful testing, measurement and evaluation of the completed design

9.2 The Author's Contributions

The first part of the thesis has made the following contributions:

- Analysis and selection of the appropriate components involved in the transmission across the backplane of the 10 Gigabit Ethernet switch, including the connectors, I/O Serdes transceivers and PCB trace configurations:
    - Validation of the chosen components and derivation of their limit of operation.
- Proposal for the backplane interconnect design, including the stack-up and the design rules of the PCB traces on the backplane.
- Establishing a methodology – through modeling and simulation – and validating the proposed design for the backplane in the pre-layout phase:
    - Identification of a critical region on the backplane inside the connector pin-fields and establishing a valid model for this region;
    - Simulation of the realistic transmission environment on the 10 Gigabit Ethernet backplane for the worst-case scenarios.

The second part of the thesis ahs made the following contributions:

- Definition, design and construction of a 10 Gigabit Ethernet traffic generator to quantify the performance of the prototype AdvancedTCA backplane.
- Transformation of the final system – by building a new fan-out clock board – in order to overcome the problems which arise due to the undocumented behaviour of the Serdes transceiver.
- Exploitation of the AdvancedTCA backplane tester and development of an automatic methodology (Bit-Error Rate tests) to determine if any given sample of backplane operates according to its design specifications.
- Evaluation of the signal integrity aspects encountered by the high-speed transmission on the backplane, by performing measurements:
    - Providing an envelope of functionality for the backplane by employing pre/de-emphasis circuitry.
- Obtaining a proper correlation between simulations based on the methodology developed in the first part and the practical measurements realized for the AdvancedTCA backplane:
    - Validation of the methodology developed in the first part of the thesis and complete the design cycle undertaken for the 10Gbps Ethernet switching backplane interconnects.

9.3 Future Work

The commissioning of the backplane tester exhibited one layout error and the need for global clocking. The original design needs to be modified accordingly so that a re-spin becomes possible in case that a company chooses to adopt the tester for quality control purposes.

The technique of using a single, simple Ethernet based controller to monitor logical functions in every slot of the AdvancedTCA chassis has attracted interest from physicists replacing legacy hardware at CERN. The relevant parts of the design should be extracted and re-compiled as a 'component' that can be integrated into a third party design.

Appendix. Modeling with HFSS

The High Frequency Structure Simulator (HFSS) is a fullwave electromagnetic field solver which uses the Finite Element Method for the numerical computation of Maxwell's equations inside a physical 3D geometry. The typical representations obtained from this computation are the electromagnetic interactions in terms of field configurations at any point in the volume space and the Scattering-parameters representation for the passive geometry under study.

Inputs to HFSS

A typical electromagnetic field problem solved by HFSS is described by defining the 3D geometry under study, the electromagnetic properties of the materials forming it and the boundary conditions.

Regarding the geometry, as a general rule of thumb, its size and complexity should be kept to aspect rations (length/thickness) of less than 1000. In contrary, the computation might fail or might take extremely long time. For each computation, boundary conditions should be included to specify the electromagnetic fields behaviour on various surfaces of the geometry. Boundary specifications are required as initial conditions and for obtaining a unique solution in the computation of Maxwell's equations. In the same category with boundaries, there are to be defined 2D surfaces (referred to as ports) through which excitation signals (in terms of power waves) are applied to the geometry. A port can be thought of as an 'aperture' face upon which the field distribution and orientation is known for steady-state finite element solution. In HFSS, the final computed solution is highly dependent on port's proper size and location and therefore special attention is required for the ports setup [AN-HFSS-01].

In the case when the passive geometry to study in HFSS is in a differential configuration, such as the differential transmission lines, differential PCB vias or connector pins, so called 'driven-terminal wave ports' are recommended to be used [AN-HFSS-03]. The wave ports can handle multi-modes and support coupled excitations as required by the differential transmission. For example, if there are multiple traces sharing one wave port they take line coupling into account. The driven terminal means that a terminal line is defined between the ground plane and each conductor to specify the voltage ($\int \vec{E} \cdot \mathrm{d}\vec{s}$). In addition to each terminal line, a calibration line is defined to specify the E-field polarization. This information about the terminals of each port, enable the port solver in HFSS to make linear combinations of the propagating modes existing inside the structure and place the desired signals on the desired terminals. Wave ports are always matched termination to a transmission line. So, if a stripline was defined with dimensions and dielectrics that make 53 Ω, then HFSS will respect that and report 53 Ω impedance. However, at the end of the solution process, the single-ended S-parameters solution given

by HFSS should be renormalized to 50 Ω. Renormalization means attaching a 50 Ω source or termination at the location of the port and it will affect mostly the magnitude of the S-parameters. To obtain the differential S-parameters, we have to compute them within the Post-Processing field of HFSS from the single-ended solution given by the solver, and in the same time to renormalize their value to 100 Ω. Only after the renormalization the solution obtained from HFSS can be properly compared with measurements for the same geometry or it can be exported to a circuit simulator.

If due to a complicated geometry, positioning of a wave-port is not possible, then another type of ports – lumped ports – can be used. This port can be placed inside the geometry structure and works with less geometry information than the wave port. A lumped port needs specifying what the impedance is supposed to be. Based on that, the port solver will adjust the magnitudes of *E* and *H* in its solution to obtain the desired impedance. Even though lumped ports do not handle coupling, it is possible to use them for differential signals if after the computation in HFSS, the results are exported to a circuit simulator where the terminals are excited differentially.

Solution Process in HFSS

The solution process in HFSS has two main stages. First, the finite element methods divide the geometry volume into a large number of smaller regions, referred as the finite element mesh. The initial mesh generated by HFSS is coarse, but during an iterative process is refined and optimized until the desired accuracy is achieved in the field computation. The second stage includes the solution process, during which HFSS solver numerically computes Maxwell's equations in any mesh element and reduces the full electromagnetic behaviour of the whole geometry to a set of high frequency circuit parameters (S-Parameters). For calculating these parameters, HFSS solution process follows the steps presented in the next flowchart.

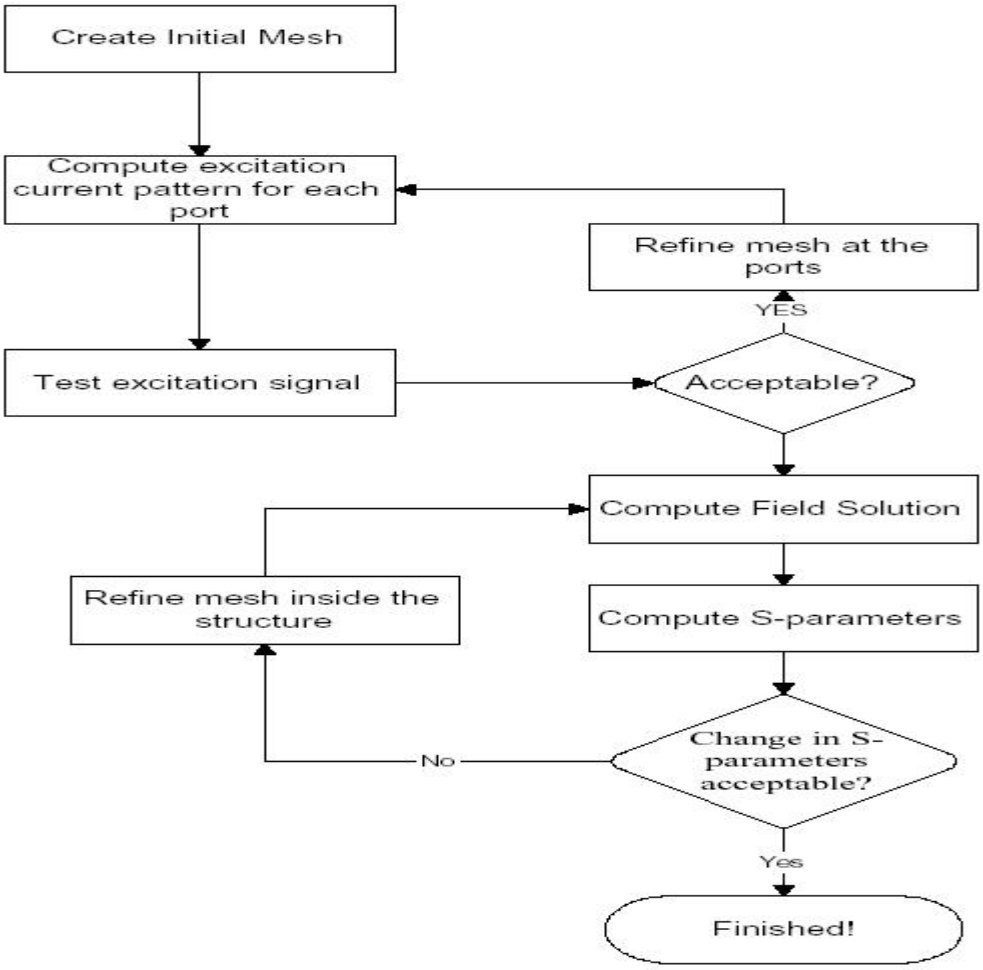


Figure 1: Flowchart of the solution process in **HFSS**

(1) Compute excitation current pattern for each port

The computation in this step corresponds to a two-dimensional problem for calculating the electromagnetic field distributions at the port surface. The goal is to calculate the electromagnetic propagation modes that can exist inside a transmission line with the same cross-sectional characteristics as the port boundary. The process is iteratively applied, for mesh refinement and the computation of the electric and magnetic configuration on the port surface.

(2) Compute Field Solution

The computation corresponds to a 3D problem, resulting in a solution for the field distribution inside the whole structure. The goal is to calculate the full electromagnetic field inside the whole geometry.

(3) Compute S-Parameters

The goal is to compute the modal S-Parameters from the amount of reflection and transmission that occurs in the structure. Once the electromagnetic field distributions

have been obtained in the previous step (2), the power through a port can be easily calculated as:

$$P = \mathrm{Re} \iint \vec{E} \times \vec{H}^{*} dS$$

This step is also an iterative one and the computation continues until the changes encountered in S-Parameters values are sufficiently small compared to some error tolerances. If the corresponding HFSS error criteria are not meet, the mesh will be further refined. For the refined mesh the field solution (2) and the S-Parameters (3) will be computed again. The computation will stop when the error criteria are met.

The results of the final computation in HFSS are in terms of electromagnetic field distributions, modal Scattering-Parameters and characteristic impedances. In order to later use these results, a renormalization to single 50 Ω, respectively differential 100 Ω is required for the reasons explained in 'Inputs to HFSS' section.

References

Author's Publications

[OLT-RTC05] Oltean, A., Martin, B., *ATCA: Its performance and Application For Real Time Systems*, IEEE Real Time Conference 2005, Stockholm, June 4-10 2005, on CD Conference Proceedings.

[OLT-PGNET05] Oltean, A., Martin, B., *AdvancedTCA Backplane Tester, 6th Postgraduate Symposium on The Convergence of Telecommunications*, Networking & Broadcasting,, Liverpool John Moores University, UK, June 27-28 2005, Conference Proceedings pp.67-72.

[OLT-UPB] Oltean, A., Buzuloiu, B., *Backplane Characterization for the next generation telecommunication systems architecture*, Politehnica University Bucuresti, Scientific Bulletin, November 2005, submitted.

[OLT-R3] Oltean, A., *Advanced TCA Backplane Tester*, PhD report, Politehnica University Bucharest, CERN Geneva, 2004.

[OLT-R2] Oltean, A., *Architectural Design Study for a 10Gbps Ethernet Switch*, PhD report, Politehnica University Bucharest, Romania, 2004.

[OLT-R1] Oltean, A., *Electromagnetic Modeling and Circuit Simulation for High Speed PCB Design*, PhD report, Politehnica University Bucharest, Romania, 2004.

[OLT-02] Oltean, A., *Signal Integrity Analysis for the Backplane Design of a Gigabit Ethernet Switch*, diploma thesis, CERN and Transilvania University Brasov, Romania, 2002.

ESTA Publications

[ESTA] *Ethernet Switching at Ten Gigabit and Above*, IST-2001-33182; see http://www.ist-esta.org

[ESTA-R1] *Report on Component I/O Characteristics –ESTA deliverable for "High Speed System Engineering workpackage*, (main contributors: Oltean, A., Martin, B.), CERN, February 2003.

[ESTA-R2] *Report on the Impact of Backplane and Transmission Technologies on Next Generation (40Gbps) Ethernet Switching-ESTA deliverable for "High Speed System Engineering workpackage*, (main contributors: Oltean, A., Martin, B.), CERN, August 2003; see http://www.ist-esta.org/E_gatekeeper.cfm?FileID=83

[ESTA-R3] *Report on Design Rules and Packaging Study, ESTA deliverable for "High Speed System Engineering" workpackage*, (main contributors: Oltean, A., Martin, B.), CERN, August 2003; see http://www.ist-esta.org/E_gatekeeper.cfm?FileID=82

[ESTA-R4] *Report and Demonstration of a Backplane Tester–ESTA deliverable for "The Switching Fabric" workpackage*, (main contributors: Oltean, A., Lestra, A., Martin, B.), CERN, December 2004.

References

[AGI-00] *Agilent High-Frequency Structure Simulator*, User's Guide, September, 2000.

[ALT-03] *Using Pre-Emphasis and Equalization with Stratix GX*, Altera White Paper, September 2003, WP-SGXEQUAL; see *http://www.altera.com/literature/wp/wp_pre-emphasis.pdf*

[AME-97] Amey, D. I., Horowitz, S. J., *Tests Characterize High-Frequency Material Properties*, Microwaves and RF, vol. 36, August 1997 , p.68-75.

[AMP-01] *Z-PACK HM-Zd PWB Footprint Optimization for Routing – AMP*, Interconnect Application Note, Rev. A, January 2001.

[AN-01] *Creating High-Performance Cost-Effective PCBs – Ansoft*, November 2001, Ansoft Presentation at the Ride the Wave Workshop; see http://www.ansoft.com/thewave/New_High_performance_cost_effective_PCBs.pdf

[AN-Q3D-1] *Ansoft Q3D – Tutorial*; see http://wwwce.web.cern.ch/wwwce/ae/Maxwell/documentation.html

[AN-Q3D-2] *Maxwell Online Help System*, 1996-2001, Ansoft Corporation.

[AN-HFSS] *Ansoft HFSS Guide*; see http://computing.ee.ethz.ch/sepp/hfss-8.5-ds/hfss/filemenu.pdf

[AN-HFSS-01] *Port Tutorial Series: Wave vs. Lumped Port Selection*, Ansoft Corp., Online Technical Support Training Material; see http://www.ansoft.com/ots/training.cfm

[AN-HFSS-03] *Modeling Components in and on Printed-Circuit Board with HFSS*, Ansoft Corp., White Paper, 2003.

[ARO-02] Aronson, L., *XFP - The Ultimate 10Gb/s Solution Finisar*, October 17, 2002; see http://www.kogent.co.jp/product/finisar/pdf/xfp_article_rev_e_200212.pdf

[ASIS] *SI Final Report – AdvancedTCA*, ASIS Ltd., 2003.

[ATCA] *Advanced Telecommunication Computing Architecture (ATCA) – PICMG 3.0*, December 2002, Version 0.91; see http://www.picmg.org/newinitiative.stm or http://www.intel.com/technology/atca/spec.htm

[BEY-02] Beyene W. et al., *Interconnect Design and Modeling of a 3.2 Gb/s/pair Bi-Directional Memory System*, DesignCon, 2002; see http://ieeexplore.ieee.org/application/enterprise/entfilecabinet.jsp?ResultStart=0

[BOL-02] Bollesen, M., *Physical-Layer Solutions for Switch and Router Backplane Interconnect*, TechOnLine Publication, Febr. 2002; see http://www.techonline.com/community/ed_resource/feature_article/20146?print

[BRI-90] *Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges*, IEEE 802.1d, IEEE Press, 1990.

[BRO-03] Brown, P., *10-Gbit/s Switch Fabric Interface Shootout*, CommsDesign, January 2003; see http://www.commsdesign.com/design_corner/showArticle.jhtml?articleID=16500276

[CAM-x] Campbell, N., *Comparison of Simulations of a simple BGA package using HFSS and Spicelink*, Presentation in HFSS Users Workshop, February 2002.

[CLI-03] Clink, J., *Maximizing 10Gbps Transmission Path Length in Copper Backplanes with and without Transceiver Technology*, High-Performance System Design Conference, in Proc. DesignCon 2003, pp. 27-50.

[CLI-04] Clink, J., *Crosstalk in 10Gbps Serial Channels: Optimizing the forgotten component*, Chris Flaherty, DesignCon, 2004; see http://www.littoninterconnect.com/main/Crosstalk%20in%20Serial%2010Gbps%20NRZ%20channels%20Feb%2004.pdf

[COO-79] Coombs, C. F., *Printed Circuits Handbook*, McGraw Hill ed. 1979, ISBN 0071350160.

[DAL-01] Dalmia, K., *High Performance Backplane Design Using the Marvell Alaska X Quad 3.125 Gbps Serdes*, Marvell White Paper, November 2001; see http://i.cmpnet.com/chipcenter/networking/images/prod/prod268.pdf

[DAL-02] Dalmia, K., *10 Gbit Rates Shake Up Backplane Design Marvell Semiconductor*, Special Report Gigabit Ethernet and Beyond, RTC Group, Special Report Gigabit Ethernet and Beyond, October 2002; see http://www.rtcmagazine.com/pdfs/2002/10/rtc10-specialreport.pdf

[DER-x] Dermer, G., Svensson, Ch., *Time domain modeling of lossy interconnects*, IEEE Transactions on Advanced Packaging, May, 2001, Vol. 24, No. 02, pp. 191-196.

[DEU-98] A. Deutsch, *Electrical Characteristics of Interconnections for High-Performance Systems*, Proc. IEEE 86, No. 2, 315–357 (February 1998).

[DEU-01] A. Deutsch, G. V. Kopcsay, P. W. Coteus, C. W. Surovic, P. E. Dahlen, D. L. Heckmann, and D.-W. Duan, *Frequency-Dependent Losses on High-Performance Interconnections*, IEEE Trans. Electromagn. Compatibility 43, No. 4, 446–465 (November 2001).

[ELC-02] Elco, R., *SPICE models with frequency dependent conductor and dielectric losses*, High Frequency Electronics, Summit Technical Media, 2002; see http://www.interactive-products.com/publications/HFE1102_Elco-high-freq-design.pdf

[ERNI-03] *AdvancedTCA Connectors according to PICMG*, ERNI Catalogue, February 2003; see http://www.erni.com/DB/PDF/ERmetZD/AdvancedTCAConFlyer-e.pdf

[ETH-85] *Local Area Networks: Carrier Sense Multiple Access with*

Collision Detection (CSMA/CD) - (ETHERNET), IEEE 802.3, IEEE Press, 1985.

[FAR-99] Farjad-Rad, R., Chinh-Kong, K. J., Horowitz, M. A., *A 0.4-um CMOS 10-Gb/s 4-PAM Pre-Emphasis Serial Link Transmitter*, IEEE Journal of Solid-State Circuits, Vol. 34, No. 5, May 1999.

[FAST-95] *Local and Metropolitan Area Networks-Supplement - Media Access Control (MAC) Parameters, Physical Layer, Medium Attachment Units and Repeater for 100Mb/s Operation, Type 100BASE-T (Clauses 21-30)*, IEEE 802.3u, IEEE press, 1995.

[FIB-93] *Fiber Optic Active and Passive Star-Based Segments, Type 10BASE-F (Sections 15-18), Supplement to 802.3*, IEEE 802.3j, IEEE Press, 1993

[FRA-99] Frazier, H., Johnson, H., *Gigabit Ethernet: From 100 to 1,000 Mbps, IEEE Internet Computing*, January-February 1999; see http://computer.org/internet

[GEE-01] De Geest, Sercu, S., Nadolny, J., *Electrical modeling of high speed interconnection links*, Proc. Of 51st Electronics Component and Technologies Conference, Orlando Florida, May 2001, pp. 187-198.

[GIG-98] *Media Access Control Parameters, Physical Layers, Repeater and Management Parameters for 1,000 Mb/s Operation, Supplement to Information Technology - Local and Metropolitan Area Networks - Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*, IEEE 802.3z, IEEE Press, 1998.

[GEE-03] De Geest, J., Nadolny, J., Sercu, S., *How to make optimal use of signal conditioning in 40 Gb/s copper interconnects*, DesignCon 2003; see http://www.fciconnect.com//pdffiles/tlib/DC03_PaperSignalConditioning_FINAL.pdf

[HALL-00] Hall, S., Hall, G., W., McCall, J., A., *High-Speed Digital System Design – A Handbook of Interconnect Theory and Design Practices*, John Wiley & Sons, Inc., 2000, ISBN 0-471-36090-2.

[HSPICE-04] *Avant Star-HSPICE Manual,* Avant! Corporation, Release 2000.4

[HUG-05] Hughes, J., E., Patel, P., S., Zapata, I., R., et al., *BladeCenter midplane and media interface card, IBM Journal of Research and Development*, Volume 49, No. 6, 2005; see http://www.research.ibm.com/journal/rd/496/patelref.html

[IEEE-02] *Standard for Information Technology – 10 Gigabit Ethernet*, IEEE 802.3ae, IEEE Press, 2002.

[IEEE-10G_BP] *Challenges in Designing 10 GB/S Backplanes*, IEEE 802.3ae Task Force, July 2000, San Diego California.

[INT-x] *10 Gigabit Ethernet Technology Overview*, Intel White Paper, IntelPRO Network Connections; see http://www.intel.com/network/connectivity/resources/doc_library/white_papers/pro10gbe_lr_sa_wp.pdf

[Johnson-93] Johnson, H., W., Graham, M., *High-Speed Digital Design – A Handbook of Black Magic*, Prentice Hall, Inc., 1993, ISBN 0-13-395724-1.

[Johnson-03] Johnson, H., W., Graham, M., *High-Speed Signal Propagation – Advanced Black Magic*, Prentice Hall, Inc., 2003, ISBN 0-13-084408-X.

[KOL-03] Kollipara, R., et al., *Design, Modeling and Characterization of High Speed Backplane Interconnects*, High-Performance System Design Conference, DesignCon 2003; see *http://www.rambus.com/news/technical_docs/designcon2003_paper.pdf*.

[KOL-04] Kollipara, R., Chia, B., et al., *Modeling and verification of backplane press-fit PTH vias*, DesignCon 2004; see http://www.rambus.com/downloads/ravi_designcon2004.pdf

[KOL-x] Kollipara, R., Patel, G., *Speed demands accurate models*, EE Times, Jan. 27, 2003; see http://www.eetimes.com/story/OEG20030124S0021

[KUZ-02] Kuznetzov, D., *W Element—Multiconductor Lossy Frequency Dependent transmission line*, Avanti Start-HSPICE 97.2 Release Notes, Avanti Corp., Milpitas, CA.

[LAE-01] Laermans, E., De Geest, J., De Zutter, D., etc, *Modeling complex via holes*, IEEE Trans. On Advanced Packaging, Vol. 24, No. 3, August 2001

[LAE-02] Laermans, E., De Geest, J., De Zutter, D., etc, *A Signal Integrity Design Tool for Printed Circuit Board Systems*, May 2002, Volume 25, No. 2, pp. 206- 214; see http://intecweb.intec.ugent.be/em/presentations_papers/signal_intregrity_design_tool.pdf

[LAQ-02] Laquai, B., Luehman, K., *Jitter testing for Gigabit serial communication transceivers*, IEEE Design&Test of Computers, January-February 2002, pp. 66-74.

[LES-02] Lestra, A., Soude, T., *Signal Integrity Challenges of a Low-Cost 3.125Gbit/s SerDes-Based Backplane Interconnect*, SAME 2002, Sophia Antipolis, October, 2002; see http://www.same-conference.org/same_2002/papers/session_4/session4_P2.doc

[LIP-02] Lipman, B., Luehman, K., *Models make the difference in high-speed pc-board design*, IEEE Design&Test of Computers, January-February 2002, pp. 66-74.

[LI-x] Li, M., Tao, Y., Wang, S., et al., *Studies on FIR Filter Pre-Emphasis for High-Speed Backplane Data Transmission*, Altera; see http://www.altera.com/literature/cp/gspx/fir-pre-filter.pdf

[LTR-01] *Lantronix DSTni-LX*, Datasheet Rev. 10/01, 2001; see http://www.lantronix.com/pdf/ds_dstni-lx.pdf

[MAD-x] Madduri, V., *High Speed Backplane in Communications systems*, National Semiconductor Corporation, Communications Design Conference, March 2004; see http://www.national.com/appinfo/lvds/files/cdc_presentation.pdf

[MAR-03] *88X2040 Datasheet – Integrated Single Chip Quad 3.125/3.1875 Gbps Transceiver*, Rev. E, September 15, 2003.

[MAX-02] *Equalizing Gigabit-per-Second Signals on Copper Media*, Application Note HFAN-06.0.1, MAXIM High-Frequency/Fiber Communications Group, March 2002; see http://pdfserv.maxim-ic.com/en/an/7hfan6011.pdf

[MAY-03] Mayrand, J., DiMinico, C., *EDA Development for 10 Gigabit Ethernet measurement based HSPICE model library*, High-performance System

Design Conference, DesignCon 2003; see http://www.tek.com/products/oscilloscopes/sampling/interconnect_analysis/customer_papers/10gbit_designcon03.pdf

[MCM-03] McMorrow S., et al., *Attention to details: Preparing for the transition from 2.5 to 10Gbps*, High-performance System Design Conference, DesignCon 2003, Santa Clara Ca.

[MIT-04] Mitchell, J., *A High-Channel-Density, Ultra-High Bandwidth Reference Backplane Designed and Manufactured for 10Gb/s NRZ Serial Signaling*, High-Performance System Design Conference, DesignCon 2004.

[MOE-03] Moedinger R., et al., *New backplane connector technology for 10 Gbps applications*, in Proc. DesignCon East 2003, pp.123-130.

[NAD-x] Nadolny, J., Forman S., *Optimizing Connector Selection for Gigabit Signal Speeds*; see http://www.ecnmag.com/article/CA45245.html

[NEL-02] Nelson, R., *Serdes devices challenge ATE, Test & Measurement World*, 8/1/2002; see http://www.read-electronics.com/tmworld

[OSI-94] *Information technology - Open Systems Interconnection - Basic Reference Model: The Basic Model*, ISO/IEC 7498-1:1994.

[PLA-02] Plant, D., Munguia, M., *The Path to 10 Gb/s: Connector and Backplane Capabilities for High-Speed Applications*, Agilent Technologies White Paper, June 2002, www.agilent.com/semiconductors

[RAD-99] Ramin F., et al., *A 0.4-um CMOS 10-Gb/s 4-PAM Pre-Emphasis Serial Link Transmitter*, IEEE Journal of Solid-State Circuits, vol. 34, no. 5, May 1999, pp. 580-585.

[POZ-98] Pozar D., *Microwave engineering*, Second edition, John Wiley & Sons, New York, 1998.

[RED-04] Redd, J., Lyon C., *Spectral content of NRZ test patterns*, Maxim Integrated Circuits, Design Feature, September 2, 2004; see http://www.edn.com/article/CA447007.html

[ROO-02] *RO4003 High frequency material insertion loss comparison with other material types*, Rogers Corp., Microwave Materials Division, http://www.rogers-corp.com/mwu

[SAY-x] Sayre E. P., Baxter M., *Comparisons of the Digital Gigabit Signal Bandwidth of Common PCB Designs and Materials*, Application Note for North East Systems Associates, Inc.; see www.nesa.com

[SER-02] Sercu S., et al., *An Efficient modeling method for via holes in 10 Gb/s serial copper links*, High-Performance System Design Conference, TecForum HP-TF4, DesignCon 2002.

[SHA-03] Shafer, M., Das S., Patel, G., *Connector and chip vendors unite to produce a high-performance 10Gb/s NRZ-capable serial backplane*, Proc. DesignCon 2003 pp. 253–271.

[SHO-02] Shott, T., Gautam P., *Overcoming Design Challenges to Reach One-Terabit Data/Rate for a Star Configuration Backplane Using FR-4*, DesignCon 2002; see http://www.teradyne.com/prods/tcs/resource_center/white_papers/DesignCon2002.PDF

[SIV-97] Sayre, E., Baxter, M., *Specification and Characterization of a MultiGigaHertz Differential Backplane Connector*, High-Performance System Design Conference, Proc. Design SuperCon 1997, pp. 187-194.

[SMO-03] Smolyansky, D., *Lossy Line Simulation and Analysis*, Printed Circuit Design Magazine, 2003; see www.pcdmag.com

[STA-01] Staggs, E., Parkes, J., *Evaluating IC packaging, RF integrated circuits*, www.rfdesign.com, February 2001

[SWI-99] Swirbel, T., Naujoks, A., Walkins, M., *Electrical design and simulation of high density printed circuit boards*, IEEE Trans. On Advanced Packaging, Vol. 22, No. 3, pp. 416-423, August 1999.

[SZI-03] Szilagyi, S., Blakey, S., Showell, J., *XFP Poses Challenges to Networking Designs*, 5 June 2003, www.commsdesign.com

[TAI-04] Taich, D., Bar-Niv, A.m D'Ambrosia, J., *Selecting optimal pre-emphasis value for 10G 4-Lanes Ethernet*, IEEE802.3 Plenary Meeting, November 2004; see http://www.ieee802.org/3/ap/public/nov04/taich_01_1104.pdf

[TAI-x] Taich D., Roth, Z., *Proposal for Definition of Preemphasis*, IEEE802.3 Plenary Meeting, November 2004; see http://www.ieee802.org/3/10GBCX4/public/dec02/MysticomCX4_Dec0602;6_taich.pdf

[TAN-96] Tanenbaum, A. S., *Computer Networks (Third Edition)*, Prentice-Hall International, 1996.

[UTP-90] *System Considerations for Multisegment 10 M/S Baseband Networks (Section 13) and Twisted-Pair Medium Attachment Unit and Baseband Med Spec, Type 10BASE-T (Section 14)*, IEEE 802.3i, Supplement to 802.3 - IEEE Press, 1990.

[VIC-04] Vicich, B., McMorrow, S., et al., *Advances in Design, Modeling, Simulation and Measurement Validation of High Performance Board-to-Board 5 to 10 Gbps Interconnects*, HSDDS book, pp. 627-680, April 2004

[VOG-x] Vogel M., Subramaniam, S., Cole B., Method for Optimizing a 10 Gb/s PCB Signal Launch, DesignCon 2004; see http://www.xilinx.com/systemio/high_speed/MartinVogel.pdf

[WAL-00] Walling, L., Polstyanki, S., *Solving A High-Frequency Signal-Integrity Problem*, Microwave & RF, October 2000; see http://www.ansoft.com/news/articles/solving_high_frequency_si_problem.pdf

[WAL-x] Wallenhorst, U., *Practical Considerations In Implementing Multi-Gigabit Backplane Interconnection Systems*, HARTING Electronics GmbH; see http://www.intellectuk.org/sectors/components_and_manufacturing/HDI_Dr%20Ulrich%20Wallenhorst_Part2.pdf

[WAT-02] Waters, G., *10 Gigabit Ethernet and the XAUI Interface – New testing challenges for the 71612C 12.5 Gb-s error performance analyzer*, Application Note, Agilent Technologies, January 2002; see http://www.educatorscorner.com/media/LW_71612c_10Gb_Ethernet.pdf

[WED-04] Wedge, S., *Accurate Multi-Gigabit link simulation with HSPICE,* Xcell Journal, Summer 2004, pp. 56-59

[WIL-02] Williams, L., Rouselle, S., Boots, B., *Circuit board design for 10Gbit XFP optical modules*, EDN, 29 May 2002, pp. 63-70.

[WIL-x] Williams, L., *Designing High-Speed Interconnects for High-Bandwidth FPGAs*, Xcell Journal, Summer 2004, pp. 20-23

[WOO-02] Woodruff, B., Roselle K., *Why 5G is Critical for Intrasystem Links*, Communications Design Conference, September 2002, San Jose, Ca., WS216.

[XFP-MSA] *Information Specification INF-8077i, Appendix D - Optimum Via Design, 10 Gigabit Small Form Factor Pluggable Module*, Revision 4.0, SFF Committee, April 2004; see http://www.xfpmsa.org/XFP_SFF_INF_8077i_Rev4_0.pdf

[Young-01] Young, B., *Digital Signal Integrity - Modeling and Simulation with Interconnects and Packages*, Prentice Hall, Inc., 2001, ISBN 0-13-028904-3.

[ZER-04] Zerbe, J., Werner, C., Kollipara, R., et al., *A flexible serial link for 5-10Gb/s in realistic backplane environments*, DesignCon2004; see http://www.rambus.com/news/technical_docs/jaredzerbe_designcon2004.pdf

[ZHA-02] Zhang, J., Wong Zhi, *White Paper on Transmit Pre-Emphasis and Receive Equalization*, Mindspeed Technologies, October 31, 2002; see http://www.analogzone.com/io_shoot_mindspeed.pdf

Web Links

[CAD_SQ] http://www.cadence.com/products/si_pk_bd/

[CON-1] www.connectpositronic.com

[Gore] http://preview.gore.com/MungoBlobs/eyeopener_2mm_vdhmhsd_plug_datasheet,7.pdf

[HM-ZD] http://catalog.tycoelectronics.com/TE/GeneralInfo/1309266_0102.pdf

[LEC-01] http://www.lecroy.com/tm/products/Analyzers/home.asp?mseries=10

[LEC-02] http://www.lecroy.com/tm/products/Probes/Differential/WaveLink/default.asp

[PICMG-02] http://www.picmg.org

[REF-loss] http://www.rogerscorporation.com/mwu/pdf/ro297.pdf

[REF-diel] http://www.fciconnect.com/pdffiles/highspeed/dc_PCB_Design.pdf

[REF_pnc] http://www.pnconline.com/Process.html