CHAPTER 2

GIGABIT ETHERNET PROTOCOL

2.1 Objectives of Research

Ethernet protocol has been a robust and dominant technology and it is found to be present on more than 90 percent of all networked devices globally [17] and the percentage has been growing by leaps and bounds. Although the Ethernet protocol speed has evolved from 10Mbps to 400Gbps, the basic Ethernet Frame structure has remained the same with CRC32 error detection technique being employed for handling errors in the communication channel, which involves retransmission of the errored frames. Also, there is a growing demand for data-centric services and multimedia-rich internet applications which has led to the standardization of new Ethernet protocols with improved throughputs namely 10Gbps , 40Gbps, 100Gbps and 400Gbps. This improved throughput has also increased the probability of errors in the channel. Also, Ethernet has been the basis for transmission through fiber optic cables and also advanced wireless technologies like IEEE 802.11 (WLAN) and IEEE 802.16 WiMAX). Hence, we feel there is a need to rethink over the design of the Ethernet frame format, which requires the investigations to be performed with respect to CRC32 error detection capability and possibility of error correction for Ethernet frame which is already available in the new protocols. Therefore the said area was taken up for investigation with following objectives as

1. Development of Simulation model for Error Detection and Correction for Gigabit Ethernet protocol.
2. To implement the Gigabit Ethernet protocol (Ethernet MAC IEEE 802.3z and 802.3ab Standard) on the FPGA (Altera Stratix II GX).
3. To analyse the performance of Gigabit Ethernet protocol for optimum speed.
4. Error Detection in Gigabit Ethernet protocol for Single mode optical fibre media.
5. Error Correction Analysis of Gigabit Ethernet protocol.

2.2 Gigabit Ethernet Protocol Architecture

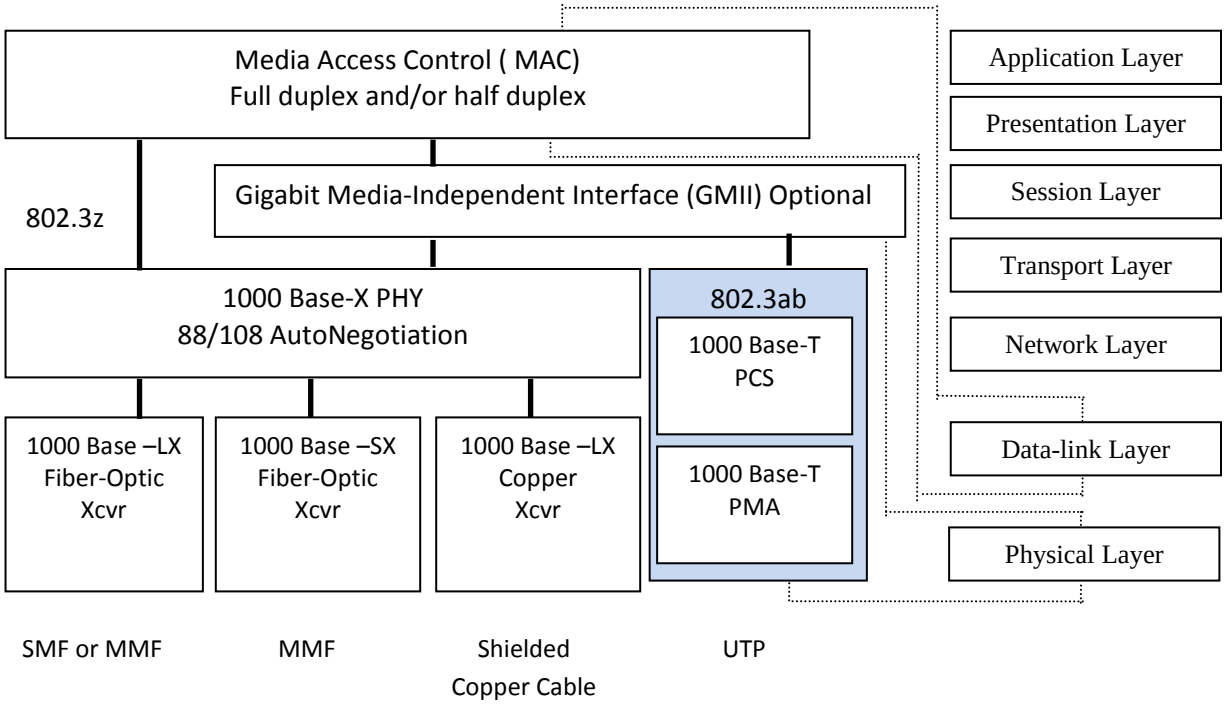

Figure 2.2.1 Gigabit Ethernet layer diagram [89]

Figure 2.2.1 shows the overall protocol architecture for Gigabit Ethernet. Gigabit Ethernet (IEEE802.3z) addresses the two lowest layers of the OSI model Layer 2, the DLL, which describes how data are organized into frames and sent over the network, and Layer 1, the Physical layer, which describes the network medium and signalling specifications [89].

The MAC layer is an enhanced version of the basic 802.3 MAC algorithm. GMII is optional for all the medium options except unshielded twisted-pair (UTP). The GMII consists of independent 8-bit-parallel transmit and receive synchronous data interfaces. It provides a chip-to-chip interface that lets system vendors mix MAC and PHY components from different manufacturers. The 8B/10B encoding scheme is used for optical fiber and shielded copper media, and the pulse amplitude modulation (PAM)-5 is used for UTP.

2.2.1 Media Access Control Layer

The 1000-Mbps specification uses the CSMA/CD (Carrier Sense Multiple Access with Collision Detection) frame format and MAC protocol used in the 10- and 100-Mbps versions of IEEE 802.3.

The basic CSMA/CD scheme has two enhancements for traditional Ethernet hub operation:

Carrier extension

Carrier extension appends a set of special symbols at the end of short MAC frames so that the frame is at least 4096 bit-times in duration compared to the minimum 512 bit-times imposed at 10 and 100 Mbps. Thus the frame length of a transmission becomes longer than the propagation time at 1 Gbps. Figure 1.6.5 shows the Gigabit Ethernet Frame Structure with Carrier Extension.

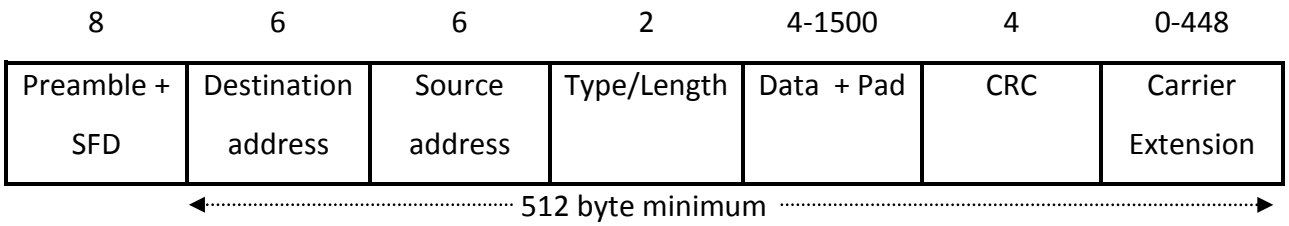

Figure 2.2.2 Gigabit Ethernet Frame Structure with Carrier Extension

Frame bursting

Frame bursting allows multiple short frames to be transmitted successively, up to a limit, without relinquishing control for CSMA/CD between frames. Frame bursting avoids the overhead of carrier extension when a number of small frames are ready to be sent on a single station.

Full Duplex operation

Full duplex operation can be employed by using a LAN switch. Here the CSMA/CD protocol, carrier extension and frame bursting are not used since data transmission and reception at a station can occur simultaneously without interference and with no contention for a shared medium. The Gigabit Ethernet specification expands on the pause protocol used for congestion control defined for 100-Mbps Ethernet by allowing asymmetric flow control. Pause protocol is

implemented by transmission of a short packet known as a pause frame. The frame contains a timer value, which contains a multiple of 512 bit-times and specifies the duration for the transmitter to stop sending frames. If the receiver becomes uncongested before the transmitter's pause timer expires, the receiver may send another Pause frame to the transmitter with a timer value of zero, and the transmitter can resume immediately. The AutoNegotiation protocol is used by a device to indicate that it may send pause frames to its link partner but will not respond to the pause frames.

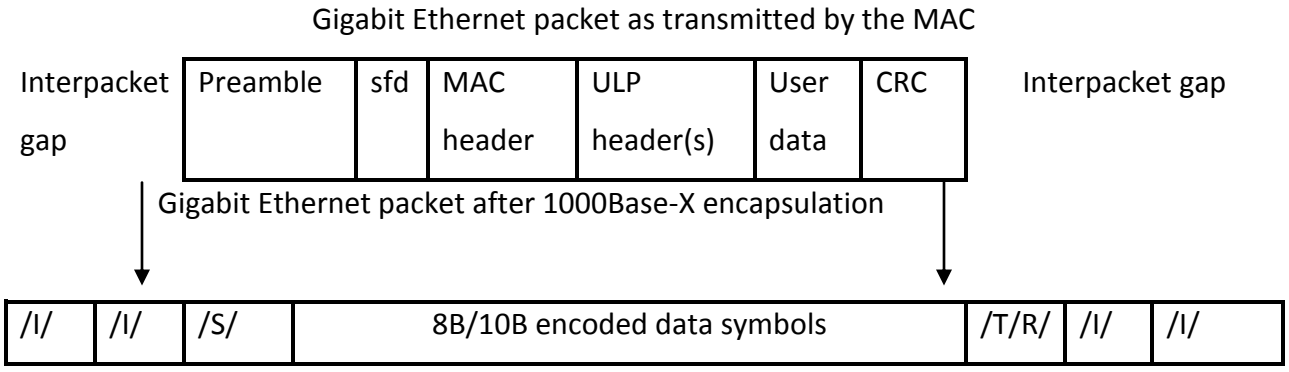

Figure 2.2.3 1000Base-X encapsulation

2.2.2 The Physical Layer

The PCS and PMA together are known as the 1000Base-X PHY. The 1000Base-X PHY performs a link configuration protocol referred to as AutoNegotiation.

Physical Coding Sublayer

The PCS encodes 8-bit data bytes from the GMII into 10-bit code groups, using a block-coding technique known as 8B/10B encoding, as mentioned earlier. The PCS follows a simple set of rules to encapsulate and encode data transmitted by the MAC. As shown in Figure 2.2.3, the /S/ code group is used to indicate the start of a packet. Then, the Preamble, start-frame-delimiter, MAC header, upper layer protocol (ULP) headers, user data and CRC are encoded using the 8B/10B code rules. A pair of special code groups, /T/ and /R/ code groups are used to terminate a packet. Packets containing an odd number of code groups are terminated with a second /R/ code group. The interpacket gap is filled with a two-symbol idle code group represented as /I/,

which is continuously between packets so that the receivers clock recovery circuits at each end of the link maintains phase and frequency lock on the recovered clock and data. Thus the 8B/10B code along with the frame encapsulation rules and a 32-bit CRC, provides Gigabit Ethernet with extraordinary robustness against undetected errors and thus ensures high data integrity.

Physical Medium Attachment

The 10-bit code-groups produced by the PCS are transmitted in a serial fashion by the PMA using a simple non-return-to-zero (NRZ) line coding.

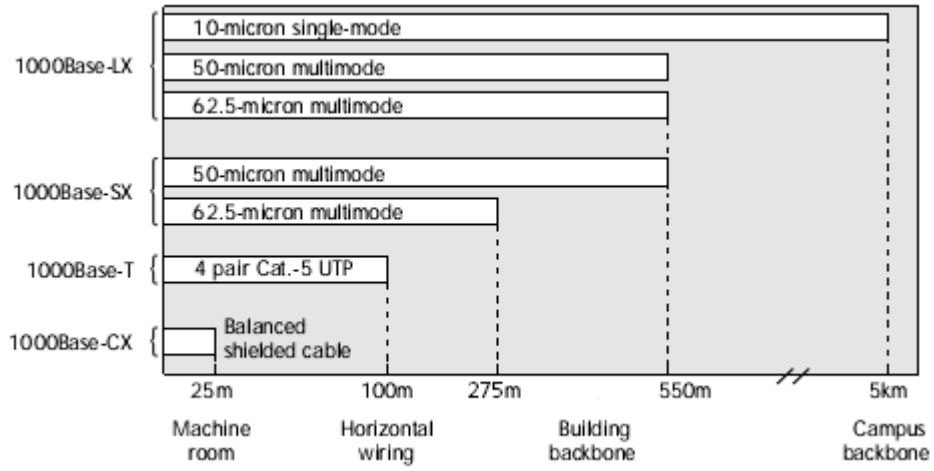

Figure 2.2.4 Gigabit Ethernet application environments and link distances [89]

The 8B/10B code provides excellent DC balance and also has excellent transition density, which is required to perform clock recovery.

AutoNegotiation

AutoNegotiation is used to ensure that the link characteristics are appropriately configured when links are initialized. Defined in Fast Ethernet to automatically select operational speeds between 10 and 100 Mbps, AutoNegotiation was adapted to Gigabit Ethernet mainly to select between duplex mode and the use of link-level flow control. On a 1000Base-X link, a special sequence of 8B/10B codes referred as a /C/ code group is used to exchange configuration information, which transfers bits of configuration information at a time. The AutoNegotiation process takes about 40 milliseconds after the cables are plugged in or the equipment is turned

on. Figure 2.2.4 charts the link distances for the transceiver types supported by Gigabit Ethernet: optical fiber, copper cabling, and UTP.

The 1 Gbps specification for IEEE 802.3 supports the following physical-layer media:

1000Base-LX: uses long-wavelength optic fibre and supports duplex links of upto 550 m of 62.5-µm or 50-µm multimode fiber or upto 5 km of 10-µm single-mode fiber. The wavelengths range from 1270 to 1355 nm. LX is used for longer distance, backbone connections.

1000Base-SX: uses short-wavelength optic fibre and supports duplex links of upto 275 m using 62.5-µm multimode or upto 550 m using 50-µm multimode fiber. The wavelengths range from 770 to 860 nm. SX is cheap and mainly used for short-distance desktop links.

1000Base-CX: uses specialized shielded twisted-pair cable of length less than 25 m. CX is used to connect devices located within a single room or equipment rack, using copper jumpers.

1000Base-T: uses four pairs of Category 5 unshielded twisted-pair copper wires to connect devices over a range of upto 100 m. The data is transmitted on each wire in both directions simultaneously. Data is encoded using 5-level pulse-amplitude modulation (PAM) at a signaling rate of 125 MBaud. AutoNegotiation is done by using patterns of simple link test pulses and facilitates interoperation with existing 10- and 100-Mbps Ethernet standards.

100 Mbps Fast Ethernet is upgraded to 1 Gbps by merging the two technologies, IEEE 802.3 Ethernet and ANSI X3T11 Fibre Channel, together as shown in Figure 2.2.5. Leveraging these two technologies has helped to use the existing high-speed physical interface technology of Fibre Channel and also maintain the IEEE 802.3 Ethernet frame format, backward compatibility for installed media. Full or half-duplex carrier sense multiple access collision detect (CSMA/CD) can be used optionally.

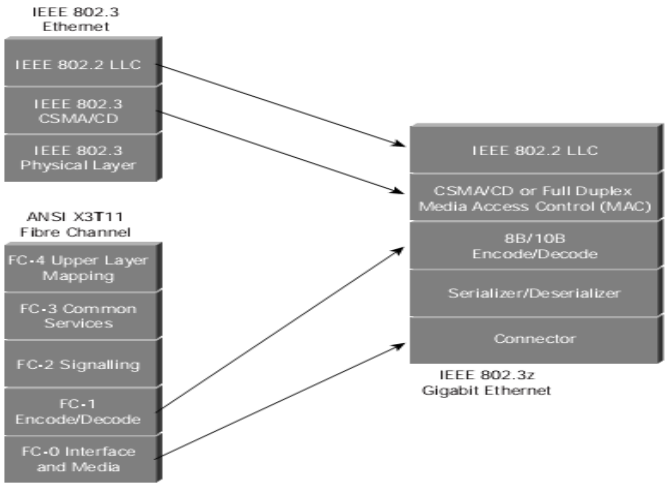

Figure 2.2.5 Gigabit Ethernet Protocol Stack [90]

2.3 Ethernet Media Access layer Implementation

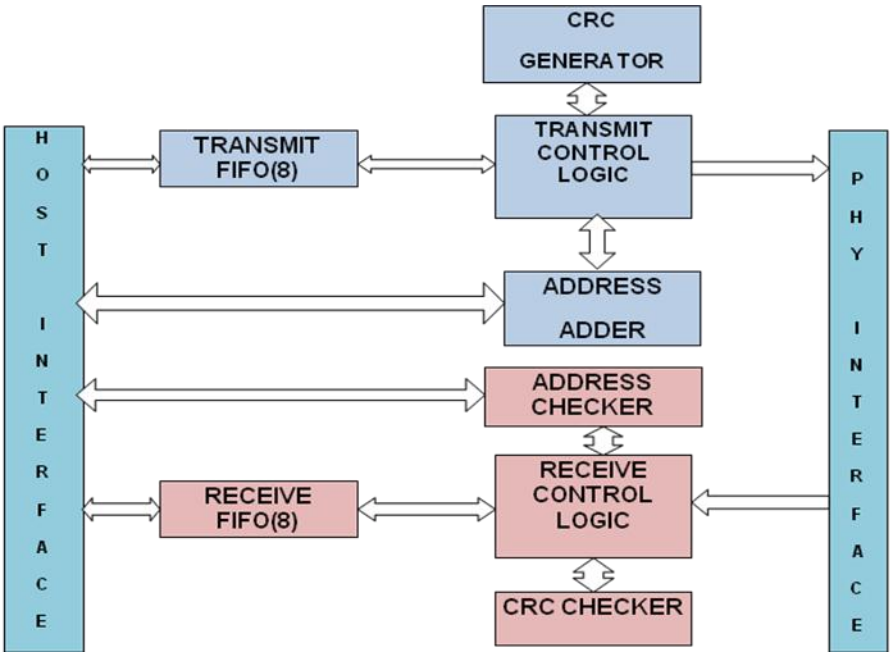

Figure 2.3.1 Proposed MAC Implementation Block Diagram

Figure 2.3.1 gives Block Diagram of the proposed Ethernet MAC architecture.

It comprises mainly of the Transmitter and the Receiver.

- For the transmitter there is a Transmitter FIFO (First In First Out) (8 bit, 1500 bytes) where the data received from host is stored before it is given to the transmitter control module.
- The receiver has a Receiver FIFO (8 bit, 1504 bytes) where it stores the data which it receives from the receiver control module.

From the host interface all the signals are generated which are needed by the MAC module and the respective FIFO buffers.

- The transmitter part contains the Transmitter control module which generates frames.
- The MAC address adder adds the destination address, the source address and the type/length field of the frame which is generated in the transmit control module.
- A CRC generator is included in the transmitter to generate the required CRC bits which is appended as the FCS (Frame Check Sequence).
- The receiver part contains the Receiver control module which decapsulates the frame.
- A CRC checker is used to check for errors in transmission.
- MAC address checker is used for MAC address checking. It also checks for multicast and broadcast packets.

2.4 Methodology of Implementation

To understand the working of the Ethernet MAC Core , we obtained the HDL source code for 10_100_1000 Mbps tri-mode Ethernet MAC IP core of opencores project [91] conforming to IEEE 802.3 specification. The said IP core was simulated and verified using tcl/tk user interface, Modelsim and cygwin platform.

Tcl/tk

Tool Command Language (Tcl) is a scripting language used for rapid prototyping scripted applications, Graphical User Interfaces (GUIs) and testing. Tcl is also used for CGI scripting. The combination of Tcl and the Tk GUI toolkit is referred to as Tcl/Tk.

Modelsim

ModelSim® is a HDL simulator which can be used to simulate electronic circuit behaviour. It can be used with Verilog as well as VHDL.

Cygwin

Cygwin is a Unix-like command-line interface, which can be used with Microsoft Windows. It provides integration of Windows-based applications, data, and other system resources with applications, software tools, and data of the Unix-like environment.The design was compiled using Altera's QUARTUS-II IDE (Integrated Development Environment)software tool.

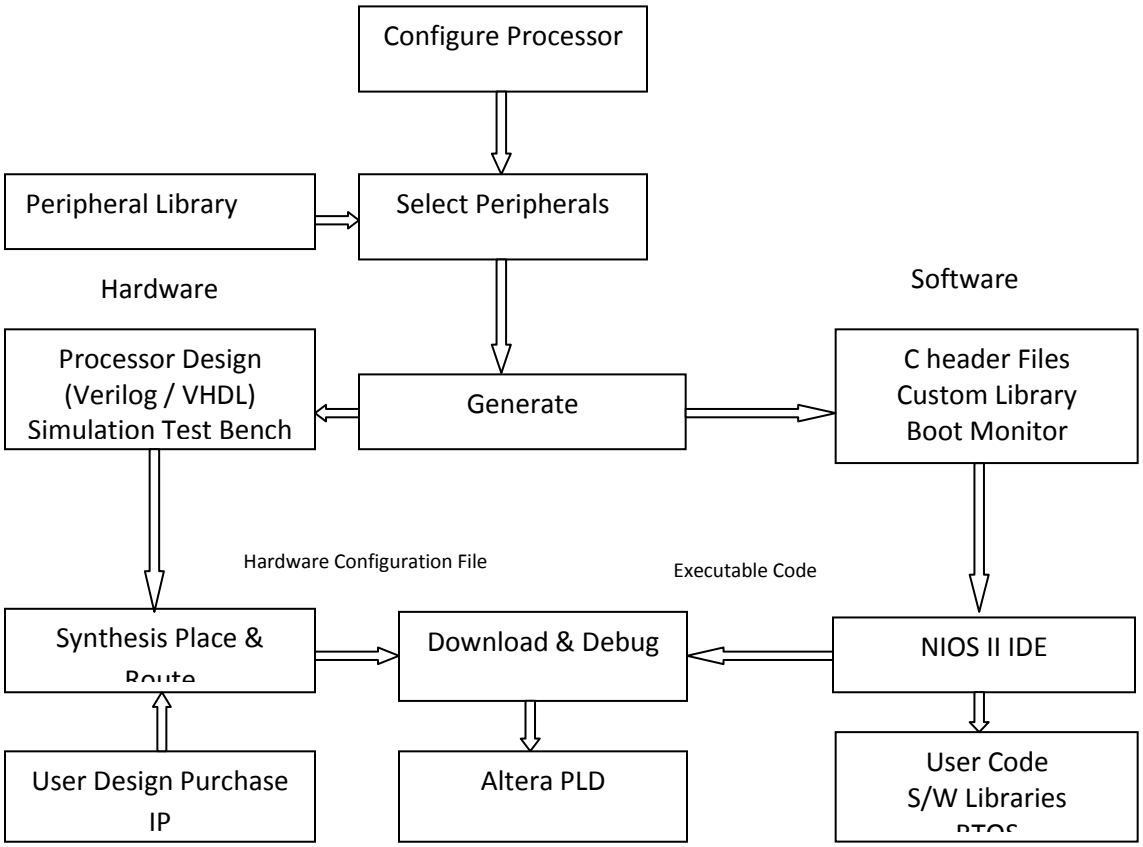

Figure 2.4.1 Various stages associated with QUARTUS-II IDE

Figure 2.4.1 gives the block diagram illustrating the sequence of compilation and the various stages associated with the IDE.

The Gigabit Ethernet protocol design was built using Altera's Triple Speed Ethernet (TSE) IP core. Altera's Quartus II software and System On Programmable Chip (SOPC) builder along with Nios II IDE was used to implement the design on Altera's EP2SGX90FF1508C3 device available on Stratix II GX PCI Express development board. The details of the implementation are described in Chapter 3.

2.5 Evaluation Techniques

The simulation and implementation of network interface architecture is a challenging issue. It requires accurate modeling of the host system architecture, behavior of the processor, memory interface, local peripheral interconnect, and also the design of the interface. All these mechanisms operate asynchronously and involve complex interaction of signals. Also the software such as operating system and device drivers need to be simulated. The Gigabit Ethernet protocol design using Altera's Triple Speed Ethernet IP Core helps to overcome these challenges.

The performance analysis of the implemented Gigabit Ethernet protocol design on Stratix II GX FPGA was carried out by using the test control API, which is provided by Altera Inc., wherein the test of frame transmission and reception can be initiated with different parameters such as speed, transmission mode, frame length, number of frames and also API was used to generate the reports of the test, such as the transmission time, number of frames sent, number of frames received, frames in error and throughput.

A testbed is developed for introducing optical attenuation using single mode optic fibre. Matlab Simulation models are developed for Error Detection and Error Correction Analysis of the Gigabit Ethernet frames. The frames generated by the Gigabit Ethernet protocol design implemented on Stratix II GX FPGA were captured in Wireshark in pcap file format and then interfaced with the Matlab simulation models to obtain the BER curves and the performance was evaluated.