PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA

MÓDULOS DE HARDWARE PARA TRANSMISSÃO E RECEPÇÃO DE QUADROS ETHERNET A 40 Gbps

Porto Alegre, 26 de Junho de 2019.

Autor: Tanauan Cardoso da Cunha
Pontifícia Universidade Católica do Rio Grande do Sul
Curso de Engenharia Elétrica
Av. Ipiranga, 6681 - Prédio 30 - CEP: 90619-900 - Porto Alegre - RS - Brasil
E-mail: tanauan.cunha@acad.pucrs.br

Orientador: Prof. Dr. Fernando G. Moraes
Pontifícia Universidade Católica do Rio Grande do Sul
Av. Ipiranga, 6681 - Prédio 32 - Sala 717 - CEP: 90619-900 - Porto Alegre - RS - Brasil
E-mail: fernando.moraes@pucrs.br

RESUMO

A crescente demanda por dispositivos conectados à Internet requer redes de dados com taxas de transferência cada vez maiores. Dispositivos reconfiguráveis do tipo FPGA podem ser utilizados para o desenvolvimento de circuitos digitais para redes de alta velocidade. Já existem soluções disponíveis na forma de *IP cores* (módulos de hardware pré-caracterizados) para transmissão e recepção de dados em taxas de até 10 Gbps. Porém para taxas superiores inexistem *IPs* disponíveis publicamente, dada a complexidade inerente ao tratamento de, por exemplo, dados sendo transmitidos com taxas de 40 Gbps. No contexto de um projeto em parceria com a indústria, a empresa TERACOM, este trabalho compreende o estudo, domínio e compreensão das camadas de rede 1 e 2 relacionados ao padrão IEEE 802.3ba visando o desenvolvimento de módulos de recepção e transmissão de dados a taxa de 40 Gbps. O estudo permite definir a arquitetura de hardware para transmissão e recepção de quadros Ethernet nesta taxa, possibilitando a implementação dos módulos de hardaware em linguagem de descrição de hardware. As interfaces externas dos módulos compreendem de um lado a interface com o circuito óptico (interface PMD - *Physical Medium Dependent*) e do outro lado a interface XLGMII. A interface PMD é responsável por conectar o FPGA com os tranceptores ópticos (SFP+) e a interface XLGMII corresponde aos quadros Ethernet na taxa de 40 Gbps. O desafio do projeto está na frequência de operação suportada pelo FPGA. Por exemplo, trabalhando-se na frequência de 312 MHZ, frequência limite para os dispositivos empregados, a largura dos barramentos de dados deve ser de 128 bits, requerendo implementações paralelas de hardware. Ao final deste TCC, é disponibilizado aos projetistas de hardware uma solução que compreenda os módulos básicos de tratamento de fluxos de dados 40 Gbps, como o PCS (*Physical Coding Sublayer*) e o MAC (*Medium Access Control*). No presente trabalho, os fluxos de dados 40 Gbps serão tratados como 4 fluxos 10 Gbps, conforme definido na norma IEEE 802.3ae.

Palavras-chave: redes de alta velocidade, módulos de hardware, Ethernet 40 Gbps, FPGAs.

SUMÁRIO

1 INTRODUÇÃO

A crescente demanda por dispositivos conectados à Internet induz o desenvolvimento das atuais redes de comunicação a operarem em taxas de transferência cada vez maiores, por exemplo, 40 ou 100 Gbps. Estas redes são utilizadas para transmitir e receber um elevado volume de tráfego de dados. Estatísticas apontam que em 2021, 80% do tráfego da Internet será de vídeo [CIS16]. Para isso, alguns protocolos são definidos com o objetivo de padronizar as operações necessárias para controlar este tráfego. Servidores, switches, e roteadores são alguns exemplos de equipamentos que utilizam protocolos de redes de alto desempenho para atender diversos tipos de aplicações.

Um exemplo de protocolo é o IEEE 802.3ba [IEE10], que define o conjunto de regras a serem adotadas por equipamentos de rede que operam em velocidades de 40 ou 100 Gbps. Além de atender às especificações da norma, desenvolver uma boa arquitetura de hardware é essencial para prover confiabilidade ao produto a ser desenvolvido e garantir a interoperabilidade entre equipamentos.

Dispositivos FPGA (*Field Programmable Gate Array*) podem ser utilizados para o desenvolvimento de circuitos digitais para suporte a redes de alta velocidade. FPGAS proveem flexibilidade de projeto, pois pode-se rapidamente avaliar diferentes opções de projeto, além de ser possível atualizar o hardware caso erros sejam encontrados no projeto. Além da flexibilidade provida pelo emprego de FPGAs, o *time-to-market* é menor que soluções baseadas em circuitos integrados específicos (ASICs).

Existem no mercado módulos de hardwares pré-caracterizados (*IP cores*), descritos em linguagem de descrição de hardware como VHDL ou Verilog, para suporte a taxas de transmissão de 40 ou 100 Gbps. Entretanto, estes módulos são proprietários, e não possuem distribuição gratuita (*open source*).

Logo, há a necessidade de se desenvolver módulos de hardware para redes de alta velocidade. Assim, este trabalho descreve o desenvolvimento realizado de um *IP core open source – Eth40SR*, capaz de transmitir e receber quadros Ethernet operando a uma taxa de 40 Gbps, o qual implementa o conjunto de regras definidas na norma IEEE802.3ba [IEE10]. O *Eth40SR* utiliza 4 fluxos paralelos de dados, cada um operando a 10 Gbps. O *Eth40SR* possibilita a integração com transceptores ópticos para realizar a interface física através de fibras ópticas. O funcionamento do *Eth40SR* foi validado através de simulações que contemplam desde a criação até o recebimento dos quadros Ethernet, conforme as orientações previstas na norma.

Este projeto foi desenvolvido em parceria com a TERACOM Telemática Ltda., empresa brasileira atuante no ramo de telecomunicações e financiadora da pesquisa desenvolvida neste trabalho, e o GAPH (Grupo de Apoio ao Projeto de Hardware) grupo de pesquisa pertencente à Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul.

1.1 Motivação

A principal motivação para o desenvolvimento deste trabalho é o atual crescimento de sistemas interligados por redes de comunicações. A demanda por aplicações que exigem elevado de tráfego de dados se faz presente no cenário atual. Sendo assim, continuamente são desenvolvidos protocolos com o objetivo de padronizar as regras necessárias para o funcionamento dos equipamentos responsáveis por estabelecer estas conexões de rede.

Dada a carência por sistemas *open source* capazes de operar a uma taxa de 40 Gbps, o qual implementa as abstrações necessárias para o tráfego de dados, este trabalho de conclusão de curso visa cobrir esta lacuna. Desenvolvido em HDL (*Hardware Description Language*), o conjunto de módulos de hardware atende a norma vigente distribuída pela IEEE para redes deste tipo.

Disponibilizado de forma open *source*, estes módulos trazem uma flexibilidade aos projetistas, possibilitando desenvolver aplicações para equipamentos de rede em 40Gbps a serem implementados em FPGA, agregando um bom custo-benefício.

1.2 Objetivos

Os objetivos deste trabalho são:

- Estudo, domínio e compreensão das camadas de rede física e de enlace, descritas no modelo OSI, relacionados ao padrão IEEE 802.3;
- Definição de uma arquitetura de hardware capaz de transmitir e receber dados operando a taxas de 40 Gbps;
- Implementação, integração e validação dos módulos de hardware.

A compreensão das camadas de rede física e enlace é imprescindível para execução deste trabalho. O protocolo define as decisões que devem ser tomadas para o correto funcionamento do sistema.

A Figura 1 ilustra a visão geral da arquitetura adotada. A esquerda estão os transceptores ópticos (*transceiver SFP+*) [SFF09] que realizam a interface física externa. Estes dispositivos são conectados a fibras ópticas e correspondem ao enlace físico de comunicação. Conectados aos

módulos transceptores encontram-se os módulos PCS (*Physical Coding Sublayer*), abstração da camada de rede física. Este conjunto de módulos desempenha a codificação dos quadros Ethernet, com finalidade de alimentar os transceptores ópticos. No sentido inverso, os módulos PCS decodificam os dados provenientes dos transceptores, assim obtendo o quadro Ethernet. Os módulos de transmissão e recepção implementam as funcionalidades básicas descritas na subcamada de rede MAC (*Medium Access Control*), isto é, codificar e decodificar a informação útil transmitida no quadro Ethernet. Após implementados os módulos de transmissão e recepção, estes são integrados a uma determinada aplicação de rede, apresentado na Figura 1.

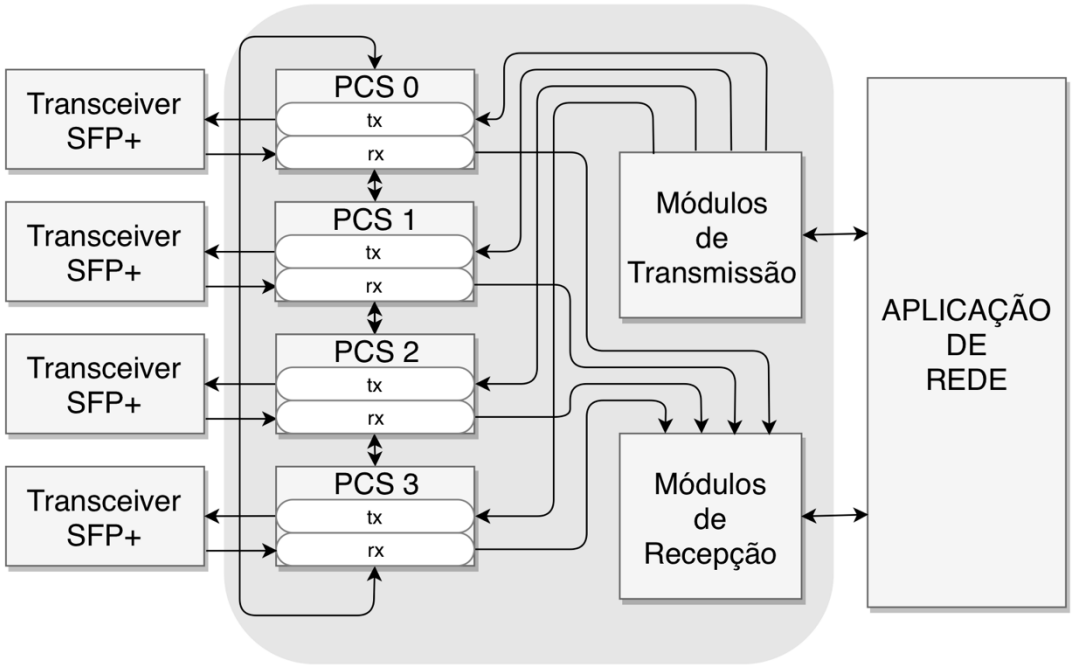

Figura 1 – Visão geral da arquitetura proposta (fonte: Autor).

1.3 Delimitações do Trabalho

A aplicação de rede e a lógica de configuração dos transceptores ópticos estão fora do escopo deste TCC. O foco do TCC está na integração dos PCSs e no desenvolvimento dos módulos de transmissão e recepção.

1.4 Estrutura do Documento

Este documento está organizado como segue. O Capítulo 2 apresenta o referencial teórico necessário para a compreensão deste trabalho. O Capítulo 3 descreve a metodologia empregada para a implementação do hardware proposto. O Capítulo 4 apresenta a principal contribuição deste TCC, que é o desenvolvimento dos módulos de transmissão e recepção 40 Gbps. O Capítulo 5 apresenta os resultados, e o Capítulo 6 as conclusões deste trabalho.

2 REFERENCIAL TEÓRICO

Este Capítulo apresenta o protocolo de comunicação Ethernet 802.3ba, também denominado padrão 40GBASE. Esse padrão foi homologado pela IEEE para dispositivos que atuam com velocidade de 40 e 100 Gbps. Nele são definidas as interfaces que atuam nas camadas física e de enlace do modelo de referência OSI (*Open Systems Interconnection model*) [MIC14], e que tem como principais funções a recepção e transmissão de dados através de um canal e a formação de um enlace entre os dois nodos conectados, respectivamente [COM13].

Devido à existência de várias formas de implementação do padrão 40GBASE, são definidas famílias de padrões que variam conforme o tipo de sinal e o meio físico utilizado [SPU14]. A família em estudo pelo Autor é a 40GBASE-R, a qual especifica a camada física para interfaces que utilizam meio de transmissão óptico. A Figura 2 ilustra o modelo de implementação da família 40GBASE e suas respectivas camadas, conforme o modelo de referência OSI.

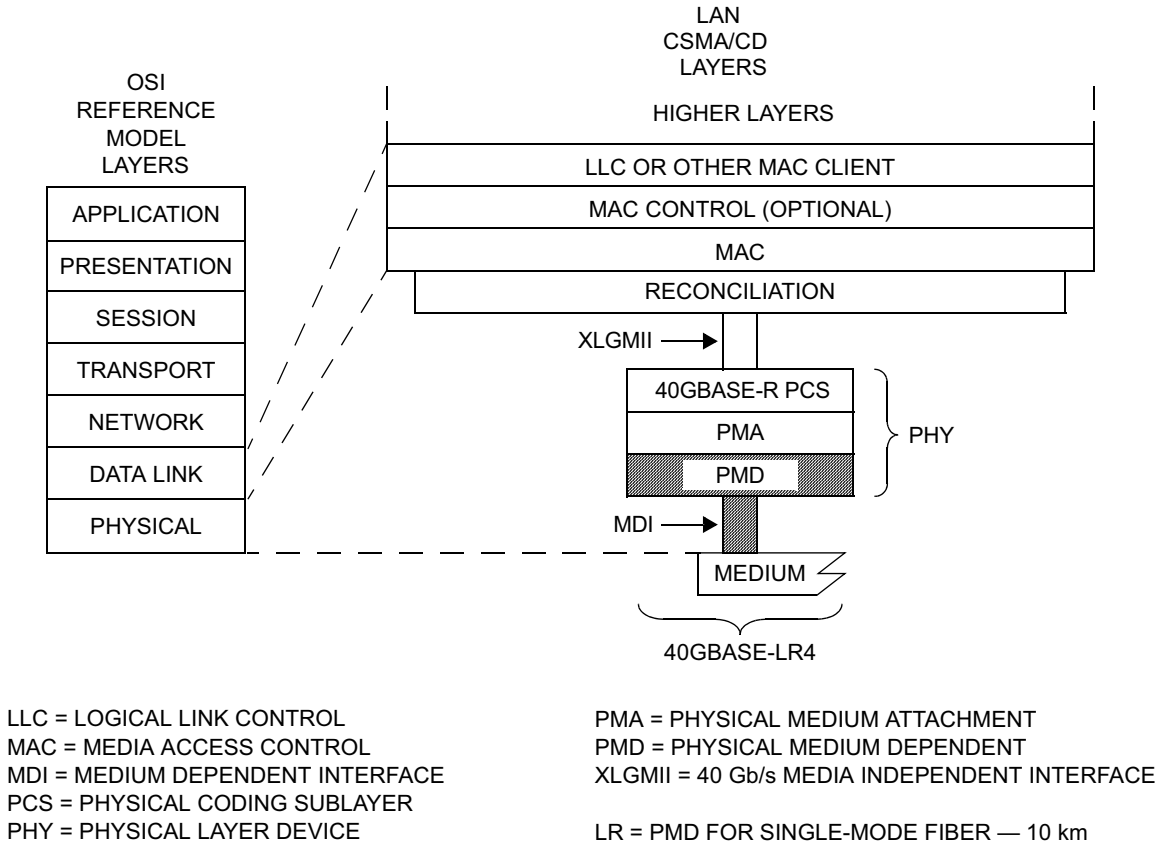

Figura 2 – Família de padrões 40GBASE [IEE10]

A família 40GBASE-R define as interfaces necessárias para a implementação da camada física do modelo OSI. Ela é baseada no modelo de codificação 64b/66b e definida através de três subcamadas e uma interface para comunicação com a camada de enlace de dados, essas são

denominadas: PMD (*Physical Medium Dependent*), PMA (*Physical Medium Attachment*), PCS (*Physical Coding Sublayer*), e interface XLGMII (40 *Gigabit Media Independent Interface*) [IEE10].

Dentro dessa família também existe uma classificação dos dispositivos ópticos quanto ao tipo de comprimento de onda emitido pelo transmissor através da fibra óptica. No projeto desenvolvido, foram utilizados dispositivos ópticos da classe 10GBASE-LR (*long reach)*. Esses dispositivos são especificados para interfaces de longo alcance através de fibra multimodo (unidirecional) e com comprimento de onda de 1310 nm. Esses dispositivos são amplamente utilizados em redes LAN de até 10km de distância [SPU14].

2.1 Physical Medium Dependent (PMD)

A camada PMD é definida como a interface de acesso do sistema ao meio físico. O padrão 40GBASE-R é responsável pela tradução dos sinais ópticos para sinais elétricos no receptor, e a tradução de sinais elétricos para sinais ópticos no transmissor, permitindo tráfego de dados seriais nos dois sentidos [IEE10]. Existem diferentes modelos de dispositivos que realizam a função da camada PMD. O dispositivo utilizado no desenvolvimento deste projeto é o SFP+ (*Enhanced Small Form-factor Pluggable)*.

2.2 Physical Medium Attachment (PMA)

A camada PMA realiza a de-serialização dos dados provenientes do módulo PMD e a transferência dos mesmos em formato de pacotes para a camada PCS. O processo inverso acontece no sentido de transmissão, onde ocorrem a serialização dos dados recebidos da camada PCS e a transferência serial dos mesmos para a camada PMD. Essa camada também é responsável pela recuperação do *clock* do transmissor, no qual os dados recebidos são originados. Para isso, é preciso que durante a inicialização do dispositivo esse *clock* seja monitorado e recuperado. Essa função é necessária para o ajuste de fase entre os dois dispositivos, que apesar de operarem na mesma frequência, enfrentam sempre alguma diferença de fase, mesmo que pequena, e que necessita ser tratada, evitando possíveis erros de sincronismo [IEE10].

2.3 Physical Coding Sublayer (PCS)

A camada PCS [ATH05] é responsável pelo processo de codificação dos pacotes de dados provenientes da camada de enlace. O objetivo desse processo é codificar os pacotes de forma a facilitar a recepção e alinhamento dos mesmos no dispositivo ao qual a interface está conectada. A implementação da camada PCS é feita de forma bidirecional. No sentido de transmissão, pacotes de dados em formato XGMII são codificados, embaralhados e transmitidos. No sentido de recepção,

ocorrem processos de alinhamento dos pacotes recebidos, desembaralhamento e decodificação para o formato XLGMII [IEE10].

2.4 Medium Access Control (MAC)

O controle de acesso ao meio (MAC - *Medium Access Control*) é a primeira subcamada pertencente à camada de enlace (*data link* – Figura 2) e a mais próxima da camada física (camada 1), referente ao modelo de referência OSI [IEE10]. As funções dessa camada são:

- Recebimento dos dados provenientes da camada física;
- Delimitação e sincronismo dos pacotes de dados;
- Encapsulamento e desencapsulamento: remoção dos delimitadores de início e fim de pacote e preâmbulo;
- Controle de erro a partir da utilização de FCM (*Frame Check Sequence*);
- Descarte de pacotes com erro;
- Inserir Interpacket Gap: o espaçamento mínimo necessário entre pacotes.

A Figura 3 ilustra o os campos do quadro Ethernet. Este quadro possui um número fixo de campos, padronizados pelo padrão IEEE802.3ba [IEE10].

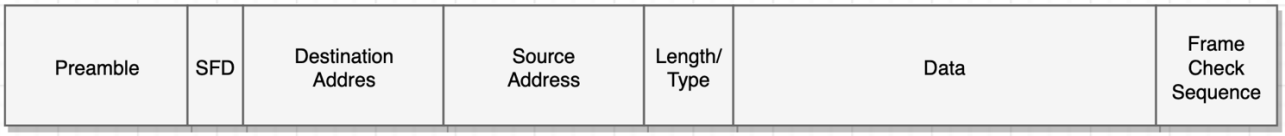

Figura 3 – Modelo geral de um quadro ethernet (fonte: Autor).

Os campos do quadro Ethernet correspondem a:

- ***Preamble*:** possui 7 bytes, onde cada byte possui valor igual a 0x55. Tem a finalidade de indicar o início de uma nova transmissão;
- ***Start Frame Delimiter* (SFD):** é representado através de 1 byte e tem valor igual 0xD5. Serve como delimitador de um novo quadro Ethernet;
- ***Destination Address***: campo com 6 bytes onde é inserido o endereço MAC destino, isto é, o endereço físico da estação receptora a qual se destina o quadro Ethernet;
- ***Source Address*:** campo com 6 bytes preenchidos com o endereço físico da estação transmissora, MAC fonte.

- ***Lenght/Type*:** este campo é preenchido com 2 bytes e indica a quantidade de bytes de informação útil (*payload*) que será transmitido pelo quadro Ethernet;
- ***Data*:** este campo contém a informação útil a qual se deseja transmitir. Possui tamanho que varia de 46 a 1500 bytes;
- ***Frame Check Sequence* (FCS):** campo destinado para verificação do quadro Ethernet. Com tamanho de 4 bytes, armazena o valor de CRC (*Cyclic Reduddancy Check*), o que possibilita analisar a integridade do quadro Ethernet recebido.

A subcamada MAC permite realizar a comunicação entre a estação transmissora e receptora abstraindo o meio físico no qual será transportado o quadro Ethernet. O uso desta abordagem permite a conexão com diversas formas de implementação da camada física de rede, visando o meio que se deseja transmitir.

2.5 Interface XLGMII

A interface XLGMII (4*0 Gigabit Medium Independent Interface*) tem por objetivo padronizar os pacotes de dados provenientes de diferentes formas de implementações de camadas físicas do protocolo Ethernet, adequando-as à um padrão único que possa ser interpretado nas camadas superiores. O tamanho do pacote enviado e recebido por essa camada pode variar conforme a implementação. No padrão Ethernet, a interface é descrita utilizando barramentos de 32 bits de dados e 4 bits de controle. A interface utiliza um bit de controle para cada byte de dados, indicando que tipo de informação o byte contém. Quando o sinal de controle é ‘0’, dados normais são transmitidos, já o sinal de controle ‘1’ indica que caracteres especiais estão sendo transmitidos. Dentro desses estão inclusos os sinais de início e fim de pacote, canal de transmissão ocioso (*Idle*) e erro na transmissão [FRA00][GUS08].

3 METODOLOGIA

Para o processo de desenvolvimento deste trabalho, foi necessário realizar inicialmente o estudo das normas que padronizam o funcionamento de redes de comunicações que operam em alta velocidade, sendo elas IEEE802.3AE [IEE02] a qual normaliza o funcionamento para um fluxo de dados de 10 Gbps, assim como a IEEE802.3ba [IEE10], esta para redes que operam em 40 e 100 Gbps. Além das normas, foi utilizado como referência para o desenvolvimento deste TCC, um trabalho de fim de curso desenvolvido na PUCRS o qual refere-se à comunicação de alto desempenho em FPGA [THU15] visando redes de 10 Gbps. Através deste trabalho, e um artigo publicado [JUR17] foi possível analisar e escolher uma abordagem para a implementação que seja capaz de operar em 40 Gbps.

Utilizando os módulos disponíveis no projeto anterior desenvolvido com a empresa TERACOM, que operam em 10Gbps, foi criado um ambiente de simulação descrito em linguagem SystemC, o qual instancia estes módulos a fim de analisar os procedimentos a serem adotados para que o sistema opere conforme o padrão IEEE802.3AE.

Aplicando as especificações para 40Gbps, foi implementado um conjunto de módulos de hardware, descritos em VHDL, os quais implementam os métodos de codificação e decodificação dos dados que se deseja transmitir e receber.

Foi desenvolvido um ambiente de simulação que permite alterar os parâmetros relativos aos quadros Ethernet, por exemplo, o tamanho e o espaçamento dos quadros. Desta forma, foram comparados dados originalmente transmitidos com dados recebidos, a fim de se verificar a corretude do sistema desenvolvido. Compreendidas as particularidades requisitadas para o correto funcionamento em 40 Gbps, sob o ponto de vista de transmissão, foram criados módulos de circuitos que são alimentados por uma interface genérica. Para a recepção, o procedimento foi análogo.

4 ARQUITETURA ETH40SR

Este Capítulo descreve a plataforma de hardware, os procedimentos adotados, bem como as soluções desenvolvidas. Apresentam as interfaces de comunicação de cada módulo de forma a explicar o fluxo de dados entre os módulos. Este Capítulo é divido em 3 seções: visão geral da arquitetura, processo de transmissão e processo de recepção.

4.1 Visão Geral da Arquitetura Eth40SR

O processo de transmitir e receber dados acontece por meio de uma transmissão *full duplex,* possibilitando que uma estação inicie uma transmissão a qualquer momento sem que haja colisão de dados. A conexão entre a subcamada MAC e a camada física de rede é dada através de um barramento MII (*Media Independent Interface)*. A codificação MII utiliza um barramento de dados paralelo a um barramento de controle, sendo assim possível realizar a codificação do quadro Ethernet visando o meio físico no qual se deseja realizar a transmissão.

No sentido de realizar a codificação para o meio físico, o MAC é conectado com o PCS (Seção 2.3). O PCS é encarregado por recodificar os dados presentes no barramento MII, unificando os barramentos de dados e controle em um único barramento. O novo barramento necessita ser codificado e embaralhado antes de sua transmissão no meio físico. A Figura 4 exemplifica este processo.

O desenvolvimento deste TCC utiliza como arquitetura legada, um trabalho desenvolvido na PUCRS [THU15], o qual realiza os procedimentos de geração e recepção de quadros Ethernet, à uma taxa de 10 Gbps, com foco na implementação em FPGA. O Autor faz uso de uma implementação *open source* de um *core IP* MAC desenvolvido para trabalhar a uma taxa de 10 Gbps [OPE08]. Para realizar a implementação da camada física de rede, o autor fez uso do *core IP* da empresa Xilinx, adequado para operar em taxas 10 Gbps e disponibilizado de forma *open source* [XIL16].

Para que seja alcançada a taxa de 40 Gbps, a norma possibilita a implementação em 4 fluxos paralelos de dados, cada qual operando a uma taxa de 10 Gbps. Sendo assim, torna-se possível a utilização da arquitetura disponível para 10 Gbps. Outro fator que é necessário considerar é a frequência de operação para o projeto. Tendo em vista a prototipação do circuito em dispositivo FPGA [DIG15] e a complexidade do sistema, não é possível utilizar uma frequência de operação muito elevada, por exemplo, usar uma frequência 4 vezes maior que a utilizada no projeto 10G. Esta limitação aponta para construção de uma arquitetura que trate os fluxos de 10Gbps de forma paralela, mantendo a frequência de operação.

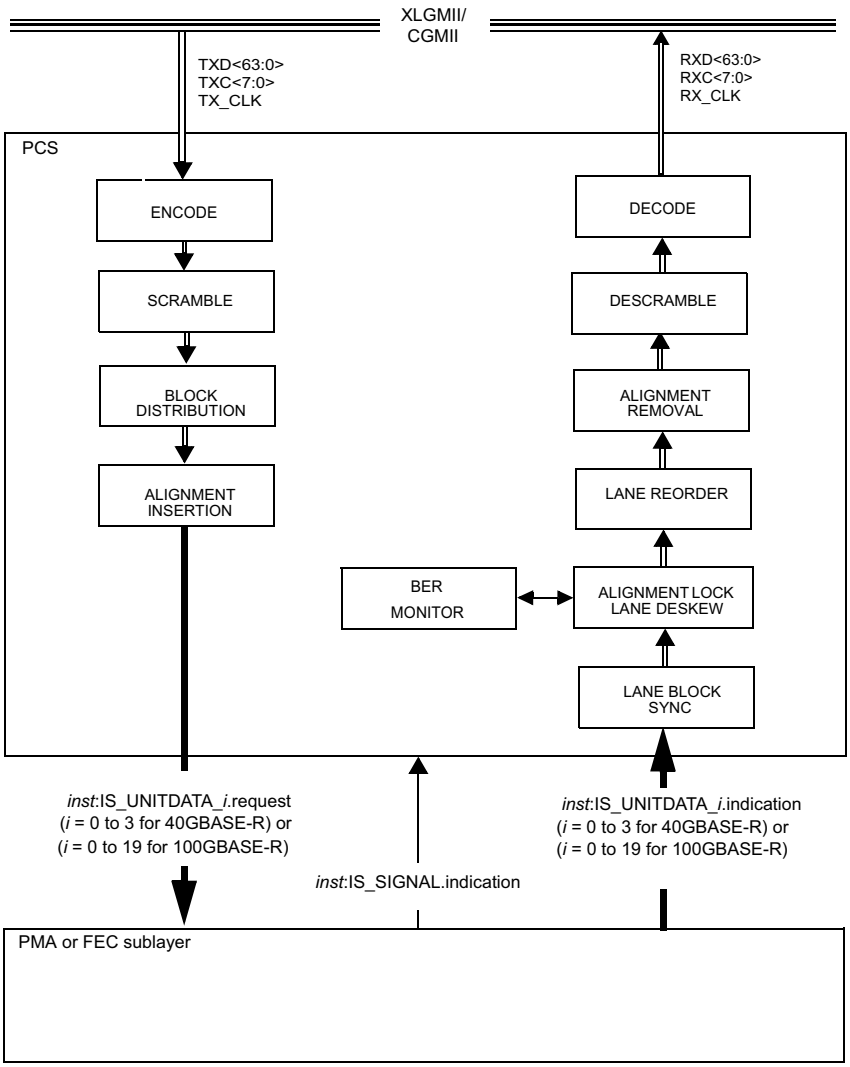

Figura 4 – Interconexão da camada PCS aos barramentos MII e à camada PMA [IEE10].

O presente trabalho contribui com a construção da arquitetura *Eth40SR*, capaz de interpretar informações provenientes de uma aplicação de rede, encapsulando-as em quadros Ethernet. Para obter a taxa de 40 Gbps, o módulo MAC desenvolvido distribui o quadro Ethernet em quatro barramentos MII de 64 bits.

Sendo assim, é possível adaptar a implementação da camada PCS, disponibilizada pela Xilinx. O uso desta implementação requer uma interligação entre as máquinas de estados que realizam a codificação do quadro ethernet XLGMII (*40 Gigabit Media Independent Interface*), visto que a implementação dos PCS legados é orientada à um único fluxo de dados com taxa de 10 Gbps. Obtendo a correta codificação dos PCS, é possível realizar técnicas de paralelização de hardware de modo a utilizar a arquitetura legada das subcamadas físicas de rede, PMA e PMD, no sentido de realizar a conexão com os transceptores ópticos.

A Figura 5 apresenta na parte superior os principais módulos relacionados à transmissão de dados, a na parte inferior os principais módulos relacionados à recepção de dados.

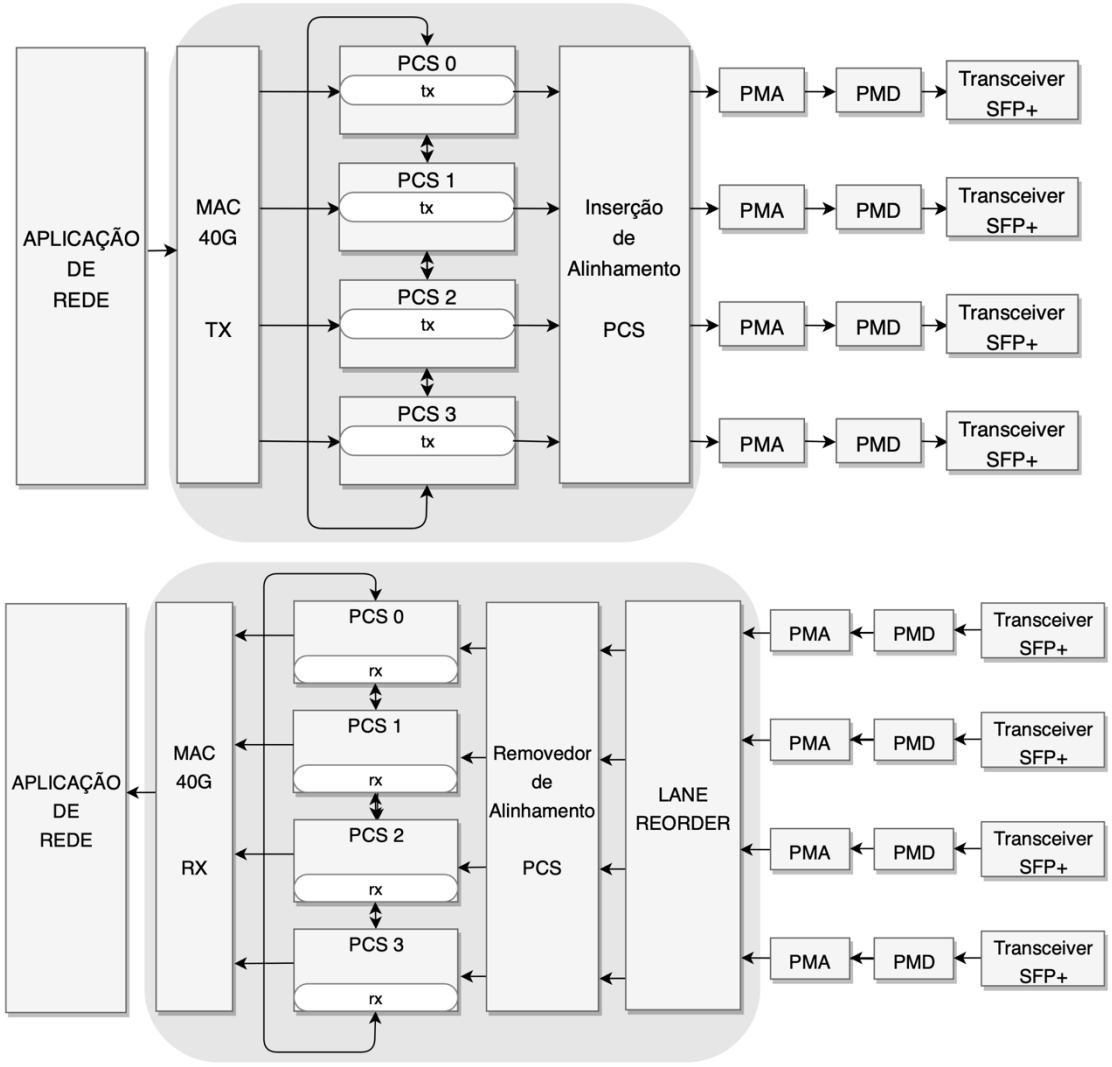

Figura 5 – Visão dos principais módulos da arquitetura *Eth40SR*. A parte superior é denominada *Eth40_Send* e a parte inferior *Eth40_Receive* (fonte: Autor).

4.2 Processo de Transmissão de Dados

Esta Seção busca apresentar os módulos de hardware desenvolvidos relacionados à transmissão de dados, da interface de entrada do módulo MAC até os sinais conectados à camada PMA, os quais inserem os marcadores de alinhamento em cada fluxo de dados provindos dos PCS, conforme ilustrado na Figura 5 . Esta Seção compreende 3 subseções: obtenção do quadro XLGMII, codificação PCS e inserção de alinhamento.

4.2.1 Obtenção do quadro XLGMII

A Figura 6 apresenta as interfaces do modulo MAC 40G. A entrada deste MAC é um conjunto de sinais de controle e um barramento de dados de 256 bits, proveniente da aplicação de

rede. Os sinais de saída correspondem a 4 pares de barramento de dados e controle, com largura de 64 bits e 8 bits, respectivamente. O barramento de dados recebe os bytes presentes no quadro Ethernet, desde o primeiro byte do endereço MAC destino até o último byte de informação útil (*payload*) que se deseja transmitir (Figura 3).

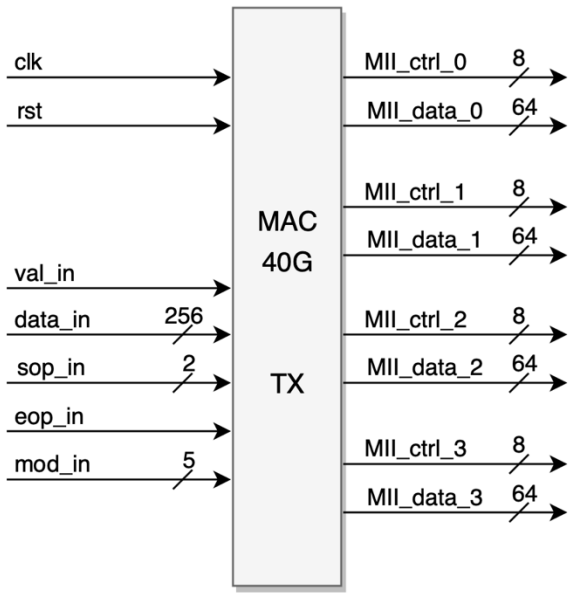

Figura 6 – Interface MAC 40G (fonte: Autor).

Para ser gerado um novo quadro MII, a aplicação de rede deve ativar o sinal **val_in**, e mantê-lo em nível lógico alto até ser indicado o último byte válido do pacote o qual se deseja gerar o quadro MII. Além disto, a aplicação de rede deve preencher o barramento de dados a partir do byte 0 ou a partir do byte 16, indicado pelo sinal de controle **sop_in** (2 bits), o qual sinaliza se o primeiro byte válido do quadro Ethernet está no byte 0 ou no byte 16 do barramento de dados, através do valor lógico "10" ou "11", respectivamente. Este sinal deve ser ativado com o valor "10" ou "11" apenas no início de uma nova transmissão. Ao longo do processo de geração do quadro MII, este sinal deve permanecer com o valor "00". O barramento de dados, **data_in** (largura de 256 bits) recebe 32 bytes por ciclo de relógio, com a orientação dos bytes da direita para esquerda (organização *little endian*), conforme ilustrado na Figura 7. Após sinalizado o início de um novo quadro, a aplicação de rede deve continuar enviando dados para o barramento **data_in**, até o último byte de dados referente ao pacote.

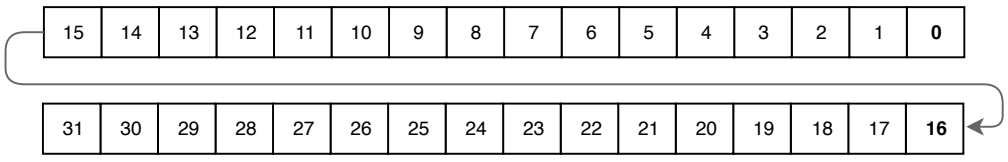

Figura 7 – Representação dos bytes do barramento **data_in** (fonte: Autor).

Para sinalizar o final de quadro, utiliza-se o sinal de controle **eop_in** (ativo baixo) e o barramento **mod_in** (5 bits), o qual indica em qual byte de **data_in** o quadro terminou**.** O **mod_in** sinaliza para o MAC 40G o último byte válido do pacote que se deseja gerar o quadro MII, tendo em vista que os quadros Ethernet possuem tamanhos variados entre 64 e 1500 bytes.

Com foco no aproveitamento da estrutura básica dos PCS, utilizada no projeto 10G, os barramentos XLGMII devem operar na mesma frequência de operação do projeto legado. Para realizar o encapsulamento em quadro Ethernet, o MAC_TX (parte relativa no MAC responsável pela transmissão) possui internamente quatro módulos de circuitos interconectados, exemplificados na Figura 8

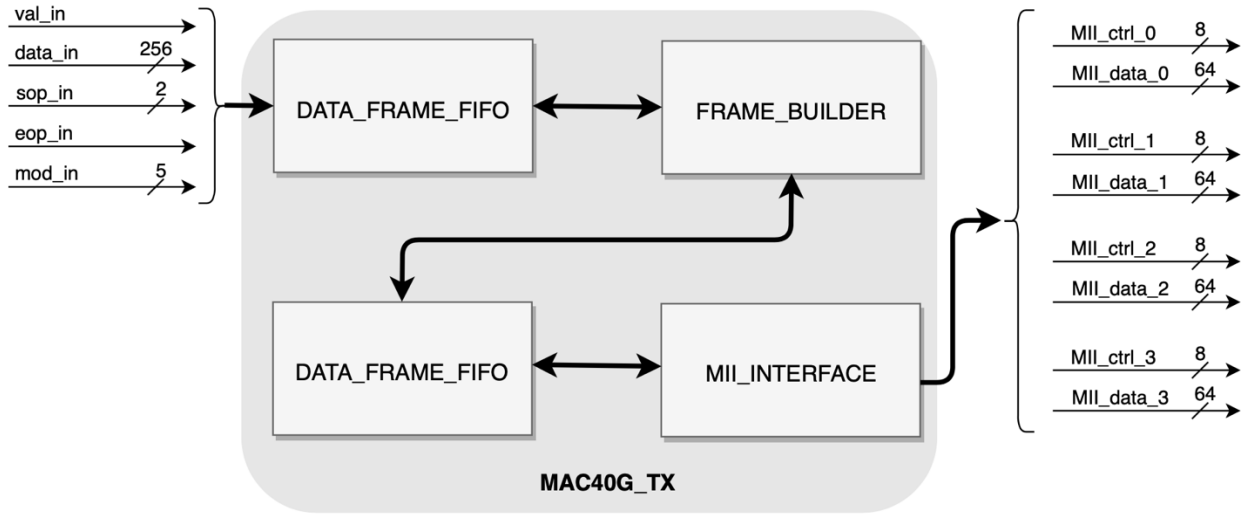

Figura 8 – Visão geral do MAC_TX (fonte: Autor).

O primeiro módulo do MAC_TX é o DATA_FRAME_FIFO (Figura 9). Este módulo é responsável por armazenar os barramentos de controle e dados em um agrupamento de 4 instanciações de FIFO_SYNC_MACRO (FIFO_*xx* na Figura). As FIFO_SYNC_MACROs são geradas pelas ferramentas de Xilinx, na forma de macros, utilizado Block RAMs (módulos de memória do FPGA). É necessária a utilização de 4 instanciações da macro, pois a largura máxima por linha da FIFO é de 72 bits. O DATA_FRAME_FIFO armazena os valores dos barramentos de dados e de controle. Isto possibilita a gravação e leitura de todos os sinais de dados e controle à cada ciclo de relógio. Observar que cada FIFO armazena 64 bits do barramento **data_in**. Por decisão de projeto, o **sop_in** é armazenado na FIFO_L1, e os demais sinais de controle na FIFO_H1.

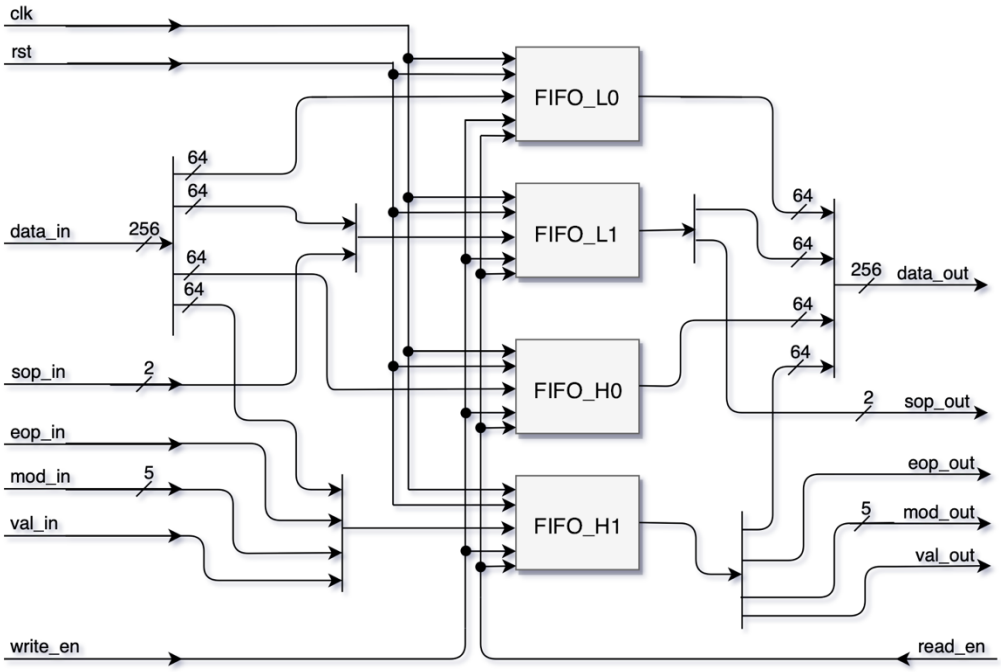

Figura 9 – Módulo DATA_FRAME_FIFO (fonte: Autor).

O DATA_FRAME_FIFO é conectado ao módulo FRAME_BUILDER – Figura 10. Este módulo é responsável por encapsular o conjunto de dados armazenados nas filas, agregando a cada início de pacote os 7 bytes de preâmbulo mais 1 byte de SFD. O FRAME_BUILDER também calcula o valor CRC (*Cyclic Redundancy Code*) desde o primeiro byte do MAC destino até o último byte de *payload* do pacote, agregando o resultado gerado, com tamanho de 4 bytes, ao final do último byte de *payload*. Por meio do cálculo do CRC, é possível fazer a verificação de integridade do quadro Ethernet no seu recebimento. Desta forma é determinado se houve erro na transmissão do quadro Ethernet.

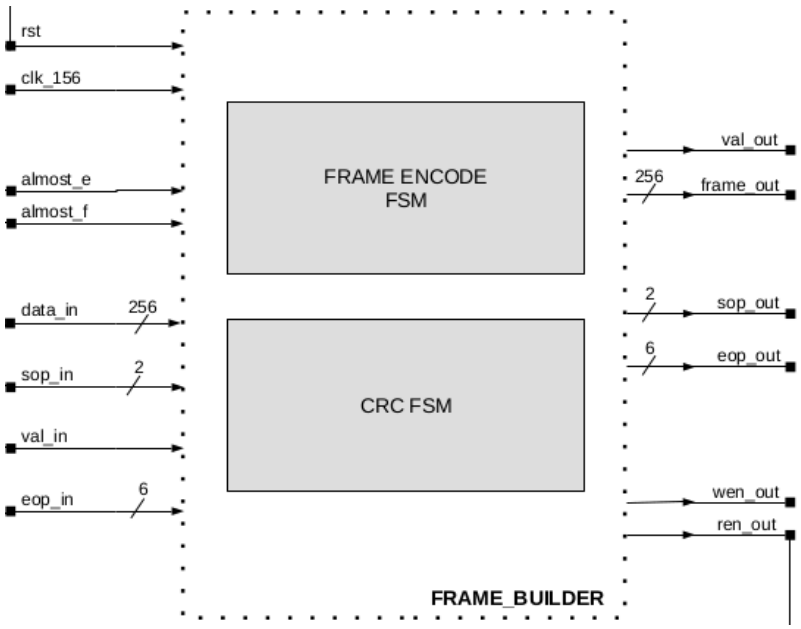

Figura 10 – Interface de comunicação do circuito FRAME_BUILDER (fonte: Autor).

A cada novo quadro gerado são recalculados os valores de início e término do quadro Ethernet (Figura 3). Tendo em vista a necessidade de gerar a codificação XLGMII, todos os sinais de saída do FRAME_BUILDER são então conectados ao módulo MII_INTERFACE - Figura 11. Este módulo consiste em conjunto de FIFOs, possuindo a mesma interface de comunicação de entrada e saída, porém este módulo armazena o quadro Ethernet completo, desde o preâmbulo até o 4º byte de CRC, com as mesmas sinalizações de início e término do quadro, referentes aos barramentos de controle, agora recalculado.

O módulo MII_INTERFACE conecta-se a quatro pares de barramentos de dados e controle, com 64 e 8 bits, respectivamente, definidos através da codificação XLGMII. Adicionalmente, o MII_INTERFACE insere o Inter Packt Gap (IPG) definido na norma [IEE10]. O IPG é um intervalo entre o término de um quadro Ethernet e o início de um novo quadro. A implementação do XLGMII requer um IPG mínimo de 12 bytes, significando que entre dois quadros Ethernet válidos, é necessário um intervalo de 12 bytes em estado IDLE.

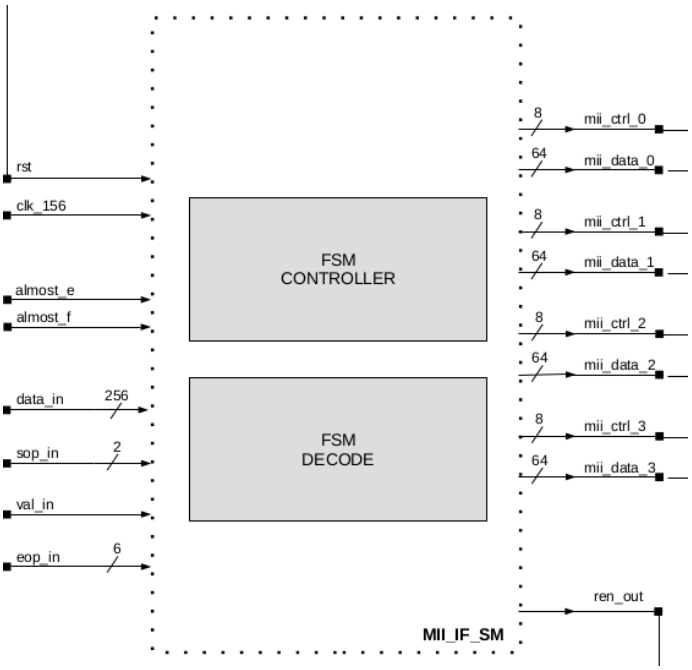

Figura 11 – Sinais de entrada e saída do módulo MII_INTERFACE (fonte: Autor).

4.2.2 Codificação PCS

Após ser gerado o quadro Ethernet, apresentado na Seção anterior, é necessário conectar os barramentos MII com a camada física de rede. Esta camada visa recodificar cada fluxo do barramento MII, gerando uma nova codificação para ser conectada com a camada PMA. O resultado do processo realizado pelo PCS é um novo barramento com 66 bits. Esta recodificação utiliza cada par dos barramentos de dados e de controle providos do protocolo XLGMII gerando este novo barramento.

Como mencionado, a empresa Xilinx disponibiliza um *IP core* com a implementação do processo a ser realizado pelo PCS para operar na taxa de 10GBASE-R. Esta implementação foi projetada para operar com apenas um único fluxo de dados. Os dados provenientes do barramento MII são recodificados através do módulo interno ENCODE, com o propósito de originar um novo barramento de 66 bits, mesclando os barramentos de controle e dados. O módulo ENCODE gera dois barramentos, um de 2 bits e o outro de 64 bits, chamado de *tx_header_out* e *tx_data_out*, respectivamente. Estes barramentos são concatenados e armazenado em uma FIFO com dois domínios de relógio, visando o sincronismo com o padrão estabelecido, no caso 40GBASE-R. Os sinais de saída do módulo alimentam o circuito de embaralhamento, chamado SCRAMBLER.

Para utilizar o *IP core* disponível foi necessário alterar os módulos internos do *PCS_CORE* da Xilinx, implementados em Verilog, adicionando às suas interfaces sinais de entrada e saída aos módulos ENCODER e SCRAMBLER. Esta alteração foi necessária pois o sistema previa que todas as informações necessárias para realizar a codificação em 66 bits estavam em um único fluxo de dados. Devido ao fato de utilizar 4 fluxos de forma paralela, as máquinas de estados internas precisaram ser conectadas e modificadas, de forma a habilitar algumas transições estados as quais não eram necessárias no projeto original (10GBASE-R).

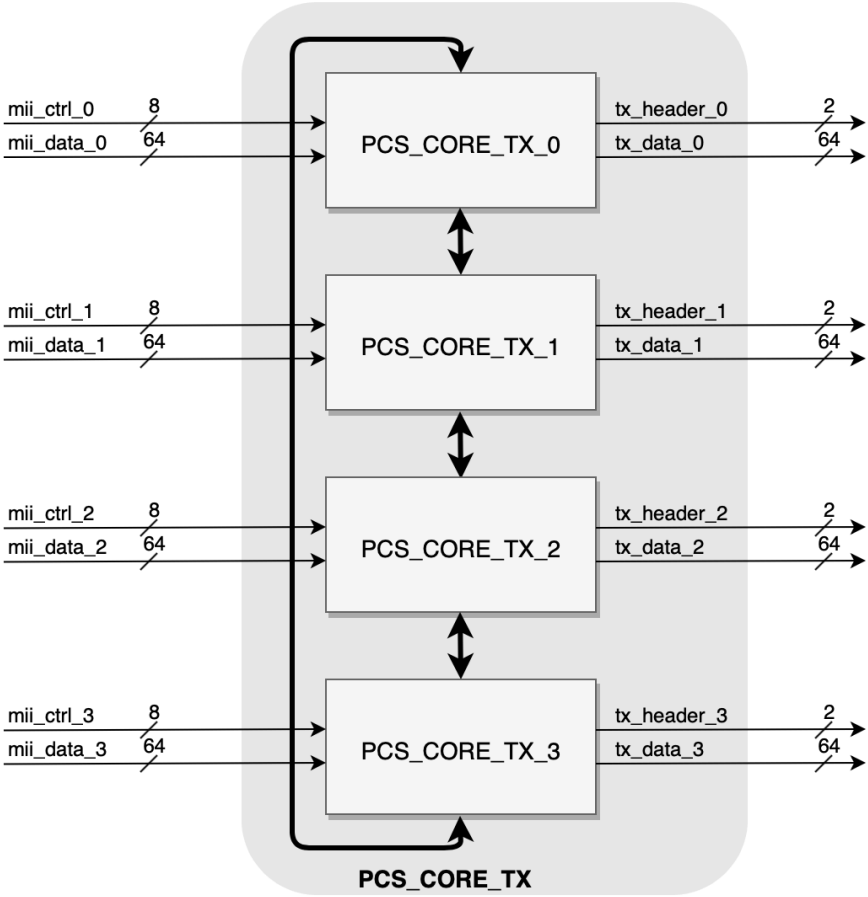

Figura 12 – Conexões adicionados entre os PCS_CORE (fonte: Autor).

Com estas alterações foi possível gerar a codificação PCS conforme a definição da norma, de modo que cada fluxo de dados gere os blocos de 66 bits codificados e embaralhados. Todo este processo é realizado de forma paralela. A Figura 12 ilustra a conexão entre os módulos PCS_CORE. As vantagens de realizar esta abordagem deve-se ao fato de reutilizar uma estrutura de máquinas de codificação da camada física, que possui relevante complexidade, adaptando-as a desempenhar os requisitos impostos pelo padrão 40GBASE-R. Além do mais, permite que o circuito opere com as mesmas frequências do padrão 10GBASE-R facilitando a prototipação do circuito em FPGA.

Conforme ilustrado na Figura 13, o padrão 40GBASE-R possui uma estrutura chamada PCS_DISTRIBUITION. Esta estrutura visa distribuir os fluxos gerados através de um único PCS_CORE, conforme ilustra a Figura 13, em *n* fluxos (*lanes*). Todavia, a implementação realizada na plataforma *Eth40SR* já trata os PCS_CORE em 4 fluxos de dados paralelos, portanto não necessitando desta estrutura.

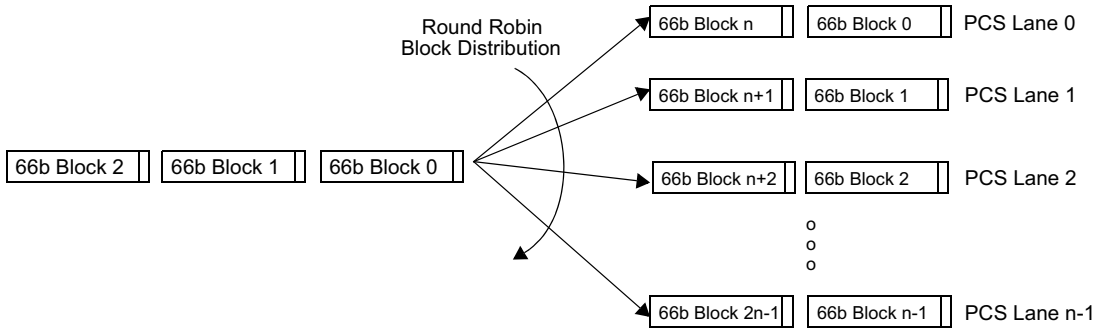

Figura 13 – PCS Block distribution [IEE10].

4.2.3 Inserção de marcadores de alinhamento.

A última etapa realizada pelo PCS objetivando realizar uma transmissão no padrão 40GBASE-R, é a inserção de marcadores de alinhamento. Esta inserção é necessária visando identificar cada um dos quatro fluxos paralelos de dados gerados, de forma com que o sistema de recepção possa identificar o correto sequenciamento entre os fluxos de dados embaralhados.

Se eventualmente for conectado fisicamente de forma incorreta os fluxos de 66 bits no sistema de recepção, não é possível desembaralhar os blocos transmitidos. A fim de resolver este problema, os marcadores são injetados de forma periódica, a cada 16.383 blocos de 66 bits gerados em cada fluxo dos PCSs, conforme mostrado na Figura 14.

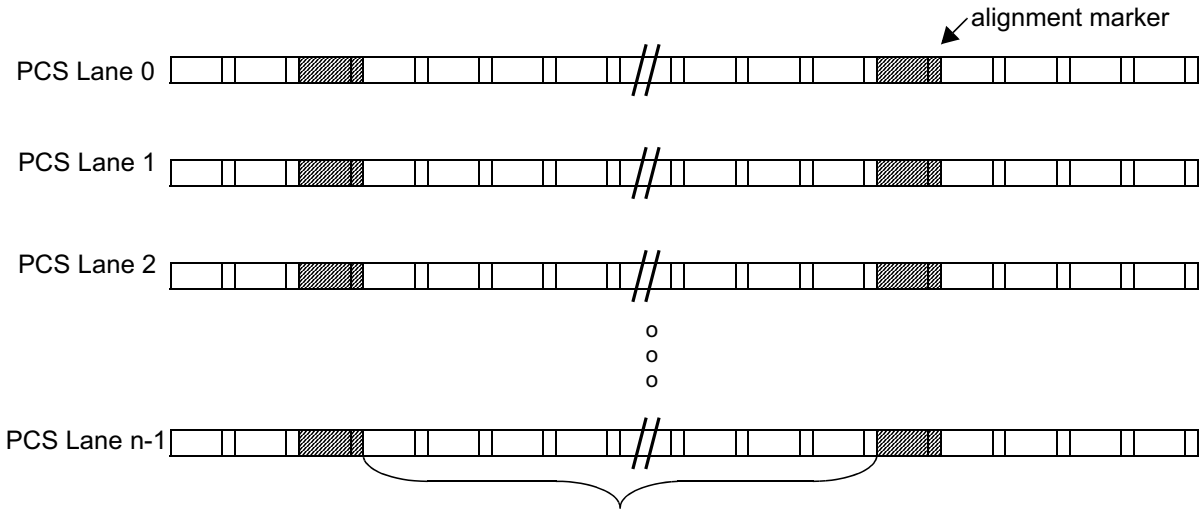

Figura 14 – Blocos entre marcadores de alinhamento [IEE10].

Os marcadores de alinhamento possuem uma formatação padronizada, ilustrada na Figura 15. O barramento de dados possui capacidade para oito bytes, sendo os três primeiros bytes que determinam a sequência lógica de cada fluxo de dados embaralhado. A Tabela 1 exibe os valores dos marcadores de alinhamento injetados em cada fluxo, referentes aos campos M0, M1, M2, M4, M5 e M6. O quarto byte, chamado de BIP3, executa a operação *xor* entre os bits dos blocos de 66 bits gerados pelo SCRAMBLER, com a finalidade de gerar um resultado de paridade. Os bytes M4, M5, M6 e BIP7 são campos onde é realizado a inversão bit a bit dos campos M0, M1, M2 e BIP3, respectivamente. O valor do BIP é uma métrica simples de paridade, porém capaz de permitir ao módulo de recepção identificar erros durante a transmissão. A Tabela 2 identifica os bits atribuídos ao BIP para a realização da operação *xor* entre os blocos de 66 bits gerados pelo circuito SCRAMBLER.

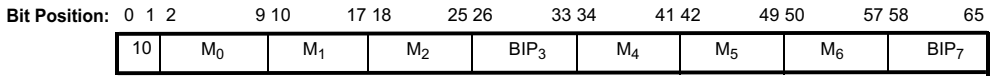

Figura 15 – Formato do marcador de alinhamento [IEE10].

Tabela 1 - Codificação dos marcadores de alinhamento [IEE10].

PCSlane number	Encoding[a] $\{M_0, M_1, M_2, BIP_3, M_4, M_5, M_6, BIP_7\}$
0	0x90, 0x76, 0x47, BIP_3, 0x6F, 0x89, 0xB8, BIP_7
1	0xF0, 0xC4, 0xE6, BIP_3, 0x0F, 0x3B, 0x19, BIP_7
2	0xC5, 0x65, 0x9B, BIP_3, 0x3A, 0x9A, 0x64, BIP_7
3	0xA2, 0x79, 0x3D, BIP_3, 0x5D, 0x86, 0xC2, BIP_7

Tabela 2 - Atribuição dos bits referentes ao byte que recebe o valor BIP [IEE10].

BIP_3 bit number	**Assigned 66-bit word bits**
0	2, 10, 18, 26, 34, 42, 50, 58
1	3, 11, 19, 27, 35, 43, 51, 59
2	4, 12, 20, 28, 36, 44, 52, 60
3	0, 5, 13, 21, 29, 37, 45, 53, 61
4	1, 6, 14, 22, 30, 38, 46, 54, 62
5	7, 15, 23, 31, 39, 47, 55, 63
6	8, 16, 24, 32, 40, 48, 56, 64
7	9, 17, 25, 33, 41, 49, 57, 65

Para que seja injetado a marcação de alinhamento nos fluxos de dados oriundos dos embaralhadores presentes nos módulos PCS_CORE, é necessário compensar esta inserção através da retirada periódica de bytes de *Inter Packet Gap* (IPG), presente no quadro XLGMII [IEE10]. A não realização de compensação com IPG acarreta no enchimento das filas internas dos PCS, devido aos marcadores de alinhamento não serem embaralhados. Desta forma, a cada alinhamento inserido é necessário realizar uma pausa no embaralhamento gerando uma pausa na leitura da FIFO que alimenta o SCRAMBLER.

Para cumprir a restrição mencionadas acima, o circuito desenvolvido denominado PCS_ALGNMENT, insere a marcação de alinhamento bem como o cálculo do BIP. Além disto interpreta a codificação gerada pelo circuito ENCODER, em cada fluxo dos PCS, de modo a identificar e retirar o IPG gerado nos barramentos de 66 bits. Sendo assim, o circuito desenvolvido faz a compensação da inserção de alinhamento com a retirada dos blocos de IPG presentes nos fluxos entre os marcadores de alinhamento.

Após a inserção dos marcadores de alinhamento, a arquitetura de transmissão esta pronta para ser conectada aos módulos que realizam a conversão do sinal digital para o meio de transmissão físico. Estes por sua vez, são estruturas que se conectam aos transceptores ópticos, a fim de transportar a informação através de fibras ópticas.

4.3 Processo de recepção

Esta Seção apresenta os módulos implementados presentes na arquitetura *Eth40SR* para a recepção dos dados. É explicada a sequência de procedimentos realizados para que a plataforma seja

capaz de receber os barramentos provindos da subcamada física de rede PMA, conforme ilustrado na Figura 16.

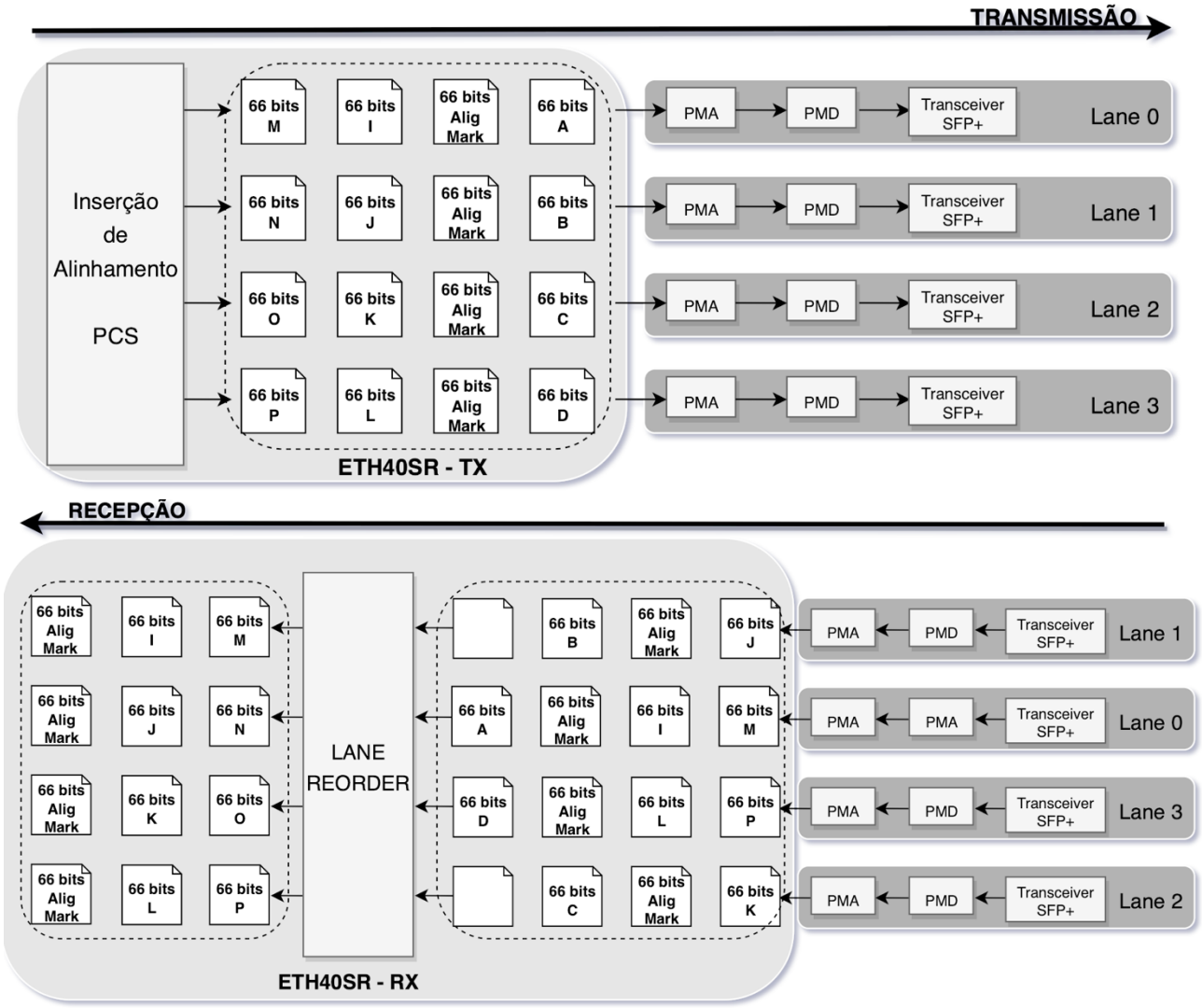

Figura 16 – Funcionalidades do circuito LANE_REORDER (fonte: Autor).

A arquitetura de recebimento deve cumprir com os requisitos impostas pelo padrão de rede 40GBASE-R, de modo que os circuitos internos sejam capazes de sincronizar os quatro fluxos de dados provenientes da subcamada PMA.

Este processo compreende desde o sincronismo dos fluxos de dados recebidos da camada PMA, passando pela retirada de atraso temporal (*skew*) entre os fluxos de dados, além de sincronizar os fluxos recebidos e rearranjá-los de forma a disponibilizar os blocos de dados de 66 bits para os módulos PCS_CORE de recepção. Estes por sua vez devem decodificar os quadros XLGMII e disponibilizá-los ao MAC40G_RX. O MAC40G_RX deve interpretar os barramentos MII recebidos através dos módulos PCS, de forma a realizar a verificação de integridade dos quadros e disponibilizar as informações transmitidas à aplicação de rede.

Para obter melhor compreensão dos procedimentos implementados e suas funcionalidades, esta seção é subdividida em sincronismo de recebimento, remoção dos marcadores de alinhamento, decodificação dos quadros XLGMII e MAC40G_RX.

4.3.1 Sincronismo de recebimento.

O padrão 40GBASE-R define as funcionalidades que a camada física de rede necessita atender para que haja comunicação entre duas estações físicas distantes, através de fibras ópticas. As regras definidas no padrão 40GBASE-R fundamentam o desenvolvimento do sistema de recepção implementado na arquitetura *Eth40SR*.

A camada PMA disponibiliza em quatro fluxos paralelos os blocos de 66 bits os quais foram gerados através dos PCS_CORE de transmissão. O objetivo dos módulos implementados é estabelecer um sistema de comunicação através de fibras ópticas, como meio físico. A utilização desta abordagem requer as seguintes ações: (*i*) correta identificação de cada fluxo de dados conectados aos transceptores ópticos; (*ii*) remoção de atraso temporal, inerente a variação física do comprimento das fibras ópticas; (*iii*) disponibilização dos 4 fluxos de dados de forma sincronizada aos PCS_CORE, para que estes decodifiquem os barramentos MII.

A fim de disponibilizar os blocos de dados para realizar a remoção dos marcadores de alinhamento, a arquitetura *Eth40SR* possui um bloco de circuito denominado LANE_REORDER. Este circuito disponibiliza em suas saídas os quatro fluxos de dados, de modo que os blocos estejam reordenados logicamente e sem atraso entre os fluxos. Esta reordenação possibilita a cada ciclo de relógio possuir todas as informações necessárias aos PCS_CORE.

A Figura 17 ilustra as funcionalidades presentes no módulo LANE_REORDER. O circuito LANE_REORDER foi implementado com quatro estágios de *pipeline.* O primeiro estágio procura pelas marcações de alinhamento, as quais são inseridas periodicamente em cada *lane* (Seção 4.2.3). Após a chegada dos marcadores, é possível identificar a sequência lógica de cada *lane,* a qual sinaliza para o último estágio para reordenar as *lanes* e disponibilizar nas saídas do módulo os fluxos ordenados.

O segundo estágio de pipeline realiza o cálculo do *Bit Interleved Parity* (BIP) e o compara com os valores gerados na transmissão, segundo o método apresentado na Seção 4.2.3. Obtido este valor, é possível verificar se houve erro de recebimento dos blocos de 66 bits gerados pelos PCS_CORE, em cada fluxo.

Uma funcionalidade essencial do circuito LANE_REORDER é realizar a compensação de atraso entre os fluxos recebidos, este atraso é chamado de *skew*. Este efeito é gerado devido as fibras

ópticas possuírem comprimentos diferentes, na escala de micrômetros. Isto ocasiona um atraso de recebimentos dos dados, providos da PMA, relativo os fluxos paralelos de recebimento. O padrão 40GBASE-R define um atraso mínimo e máximo quanto ao *skew,* exibido na Tabela 3.

Tabela 3 – Tolerância de *skew* [802.3_ba].

PCS	Maximum Skew	Maximum Skew Variation
40GBASE-R	180 ns (~1856 bits)	4ns (~41 bits)
100GBASE-R	180 ns (~928 bits)	4ns (~21 bits)

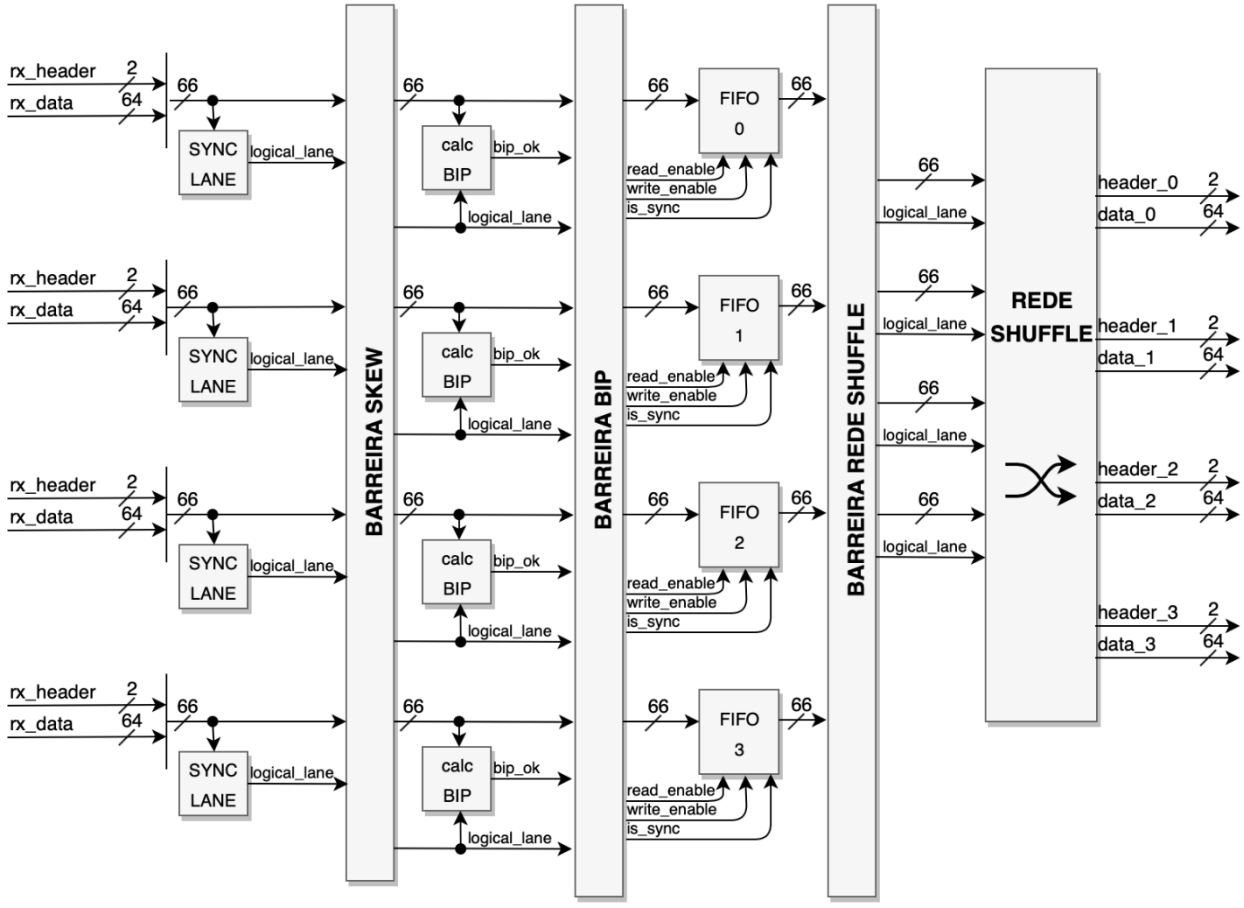

Figura 17 – Arquitetura do circuito LANE_REORDER (fonte: Autor).

Atendendo às especificações impostas pelo padrão, o terceiro estágio de pipeline implementa uma estrutura do tipo fila a qual armazena os blocos de dados a partir do primeiro marcador de alinhamento recebido, para cada fluxo, habilitando a leitura somente quando todos os fluxos já estiverem sincronizados.

O último estágio do LANE_REORDER consiste em uma rede s*huffle* para reordenar os fluxos, e disponibilizar em suas saídas, os blocos alinhados e ordenados, sem a presença de *skew.* Foi escolhido chaveamento através de rede *shuffle* ao invés de um *crossbar* visando a otimização em área e número de conexões, objetivando a implementação em FPGA. O controle dos switches da rede é

baseado no valor do *logical_lane* (*número correto da lane após sua identificação*) de cada fluxo (obtido no primeiro estágio de pipeline).

Neste ponto os dados estão ordenados e prontos para serem entregues aos próximos módulos, a fim de se realizar a decodificação dos quatro fluxos paralelos obtendo os barramentos XLGMII.

4.3.2 Remoção dos marcadores de alinhamento

Para disponibilizar os quatro fluxos de dados ordenados aos PCS_CORE de recepção é necessário realizar uma compensação referente os marcadores de alinhamento inseridos pelo sistema de transmissão. Devido os marcadores não serem embaralhados, a arquitetura de recebimento tem de retirar a marcação de alinhamento inserida antes dos módulos PCS_CORE.

A não compensação da retirada dos marcadores, inferindo codificação IDLE, reflete em esvaziamento das filas internas dos PCS, visto que estes foram projetados para executar a codificação/decodificação de apenas um único fluxo de dados.

O circuito PCS_ALGNMENT_REMOVAL recebe os quatro fluxos de dados ordenados do circuito LANE_REORDER, o qual gera a sinalização quanto aos marcadores de alinhamento. Para que estes barramentos não sejam desembaralhados e escritos nas filas internas dos PCS_CORE de recepção, os fluxos de dados provindos do LANE_REORDER são registrados por dois ciclos de relógio, e direcionado à saída, do circuito PCS_ALIGNMENTE_REMOVAL os valores mais atrasados. Quando os marcadores de alinhamento forem sinalizados, o circuito chaveia para a saída os fluxos de dados menos atrasados, fazendo com que os marcadores não cheguem aos PCS_CORE de recepção.

Para que o circuito volte a consumir os blocos de dados do registrador mais atrasado, o módulo PCS_ALGNMENTE_REMOVAL recebe os fluxos decodificados dos DESCRAMBLER internos dos PCS_CORE_RX, de forma a monitorar a codificação em *IDLE* (IPG) dos quatro fluxos. No ciclo de relógio em que for identificado um bloco de dados em codificação IDLE, é injetado diretamente nas filas internas (não passa pelo módulo DESCRAMBLER), codificação IDLE nos 4 fluxos, de forma que está inserção não altera a decodificação das máquinas que interpretam os quadros MII.

Esta ação permite entregar os dados embaralhados através do registrados mais atrasados, de forma com que a chegada dos próximos marcadores de alinhamento este processo seja repetido, não ocasionando esvaziamento das filas internas além de não alterar o processo de decodificação dos quadros MII.

4.3.3 Decodificação dos quadros XLGMII

Esta etapa objetiva a decodificação dos quadros XLGMII através de 4 blocos de 66 bits embaralhados e gerados através dos PCS_CORE_TX. Os blocos de dados provindos dos PCS_ALGNMENT_REMOVAL direcionam aos PCS_CORE_RX os quatro fluxos de dados em blocos de 66 bits, alinhados e sem a presença dos marcadores de alinhamento, a fim de os blocos serem desembaralhado e após decodificados em quatro barramentos MII.

Análogo ao processo de transmissão descrito na Seção 4.2.2, o procedimento de obtenção dos quadros XLGMII utiliza o mesmo PCS_CORE_IP da Xilinx, legado do projeto 10GBASE. Assim como no processo de transmissão, adaptações foram necessárias visando o funcionamento de forma paralela, de modo que o sistema implementado realiza quatro instanciações dos PCS_CORE. As mudanças realizadas foram feitas nos circuitos DESCRAMBLER e DECODE, circuitos descritos em Verilog, presentes nos PCS_CORE_RX, conforme lustrado na Figura 18.

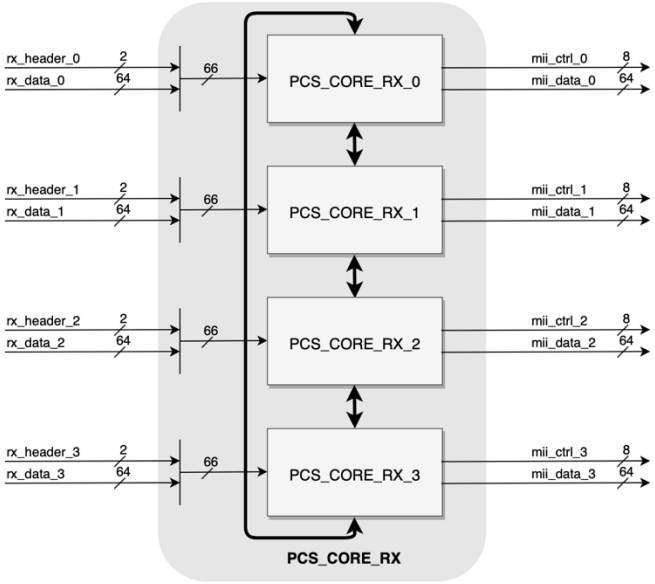

Figura 18 – Interligação dos circuitos PCS_CORE_RX (fonte: Autor).

Sob a ótica de funcionamento em 10GBASE-R, ou seja, um único fluxo de dados, o circuito DESCRAMBLER necessita dos dados desembaralhados do ciclo de relógio anterior para que possa continuar a desembaralhar os blocos de 66 bits. Todavia a implementação em 40GBASE-R interliga os PCS_CORE instanciados, de modo que para desembaralhar os dados do PCS_CORE_1 se faz necessário o bloco desembaralhado do PCS_CORE_0 e assim sucessivamente.

Após os dados serem desembaralhados, estes são armazenados em FIFO_DUAL_CLOCK, estrutura do tipo fila, descrito na Seção 4.2.2, elas realizam a troca de domínio de relógio, para que o

circuito DECODE seja capaz de extrair os barramentos MII decodificados a partir dos blocos de dados de 66 bits desembaralhados.

O circuito DECODE recebe em suas entradas os blocos de dados desembaralhados. Para o circuito ser capaz de decodificar os barramentos MII, foi interligado as máquinas de estados internas do circuito DECODE com as máquinas de estados dos outros PCS_CORE instanciados, sendo possível a decodificação sincronizada no mesmo ciclo de relógio, disponibilizando nas saídas do circuito de recepção dos PCS_CORE_RX, os quatro fluxos MII com os barramentos de dados e controle, com 64 e 8 bits respectivamente. A Figura 19 exibe a interligação dos circuitos internos dos PCS_CORE_RX.

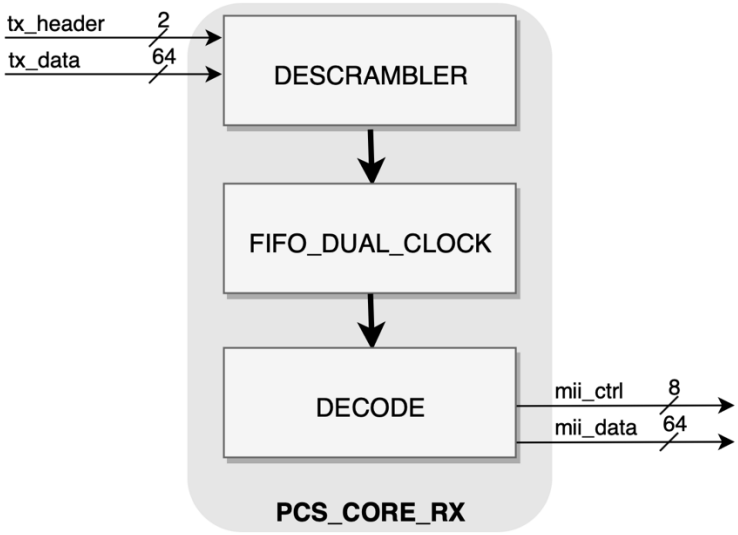

Figura 19 – Circuitos internos dos PCS_CORE_RX (fonte: Autor).

4.3.4 MAC40G-RX

A partir dos barramentos MII decodificados pela camada física de rede, a plataforma *Eth40SR* desemcapsula os dados trafegados por meio dos quadros MII, de forma a disponibilizar as informações originalmente transmitidas pela aplicação de rede, conferindo a integridade de cada quadro transmitido.

Para executar estes procedimentos, o projeto legado 10GBASE utiliza um *Core_IP* habilitado para operar a uma taxa de 10Gbps, disponibilizado de forma *open source* pela OpenCores [OPE08]. Visando uma arquitetura apta a funcionar no padrão 40GBASE, a adequação do *Core_IP* disponível é inviável, dada a complexidade das adequações necessárias.

Para que os dados transmitidos a taxa de 40Gbps possam ser desencapsulados dos quadros MII, foi construído um conjunto de circuitos que implementam as funcionalidades básicas de um

MAC. O circuito retira as constantes de preâmbulo e SFD e calcula o valor de CRC do quadro recepcionado, comparando-o com o valor de CRC transmitido no próprio quadro MII, a fim de verificar a integridade de cada quadro recebido.

O propósito deste módulo de circuito é disponibilizar uma interface de saída a qual sinaliza o início e término de pacotes recebidos, bem como a sinalização de integridade dos pacotes trafegados. Este processo deve ser feito sem causar pausa ou interrupção no fluxo de recepção dos barramentos MII provindos dos PCS_CORE_RX. Ilustrado na Figura 20 está a interface de comunicação de entradas e saídas do circuito MAC40G_RX.

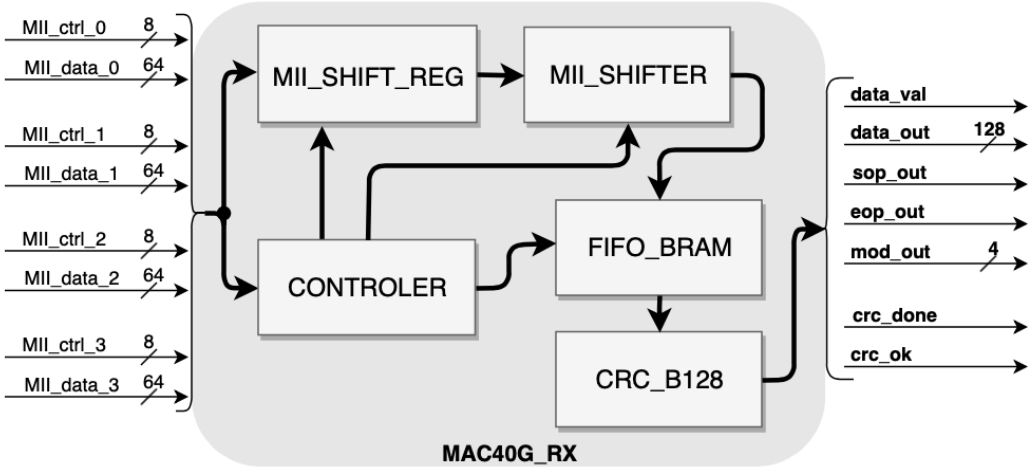

Figura 20 – Interface de comunicação do circuito MAC40G_RX (fonte: Autor).

O circuito MAC40G_RX é composto por um conjunto de cinco módulos de circuitos internos. Estes são responsáveis pela identificação e sinalização sobre as informações presentes nos barramentos de controle e dados dos quatro fluxos MII recebidos.

O módulo de circuito CONTROLER interpreta as sinalizações codificadas nos barramentos MII, de modo que os demais circuitos internos do MAC40G_RX utilizam estes sinais gerados, visando a organização dos barramentos de dados e sincronismo entre os módulos internos. Foi escolhida esta abordagem tendo em vista a necessidade de interpretar continuamente os fluxos MII recebidos. O circuito CONTROLER possui decodificadores que interpretam início e término dos quadros MII recebidos. Estas sinalizações são direcionadas aos demais módulos internos, através de uma sequência de registradores de maneira a se obter sincronismo entre os circuitos.

Através das sinalizações provindas do circuito CONTROLER os demais módulos internos MII_SHIFT_REG e MII_SHIFTER reordenam os barramentos de dados. São nesses módulos que efetivamente estão localizados os dados trafegados, desde o primeiro byte de preâmbulo até o último byte do quadro MII, sinalizando o término do quadro.

O circuito MII_SHIFT_REG, recebe em suas entradas os quatro barramentos de dados do MII com 64 bits cada. O circuito registra o conteúdo dos barramentos de dados por dois ciclos de relógio e direciona para suas saídas dois barramentos com 256 bits cada**: out_0** e **out_1**. Estes registradores são utilizados para atender as situações em que há o mínimo intervalo permitido entre dois quadros MII válidos, isto é, casos de menor IPG permitido no padrão 40GBASE. O padrão estabelece que o menor intervalo na codificação XLGMII é de 12 bytes [IEE10]. Esta informação indica que um novo quadro MII pode ser iniciado no mesmo ciclo de relógio em que está terminando o recebimento do quadro MII anterior.

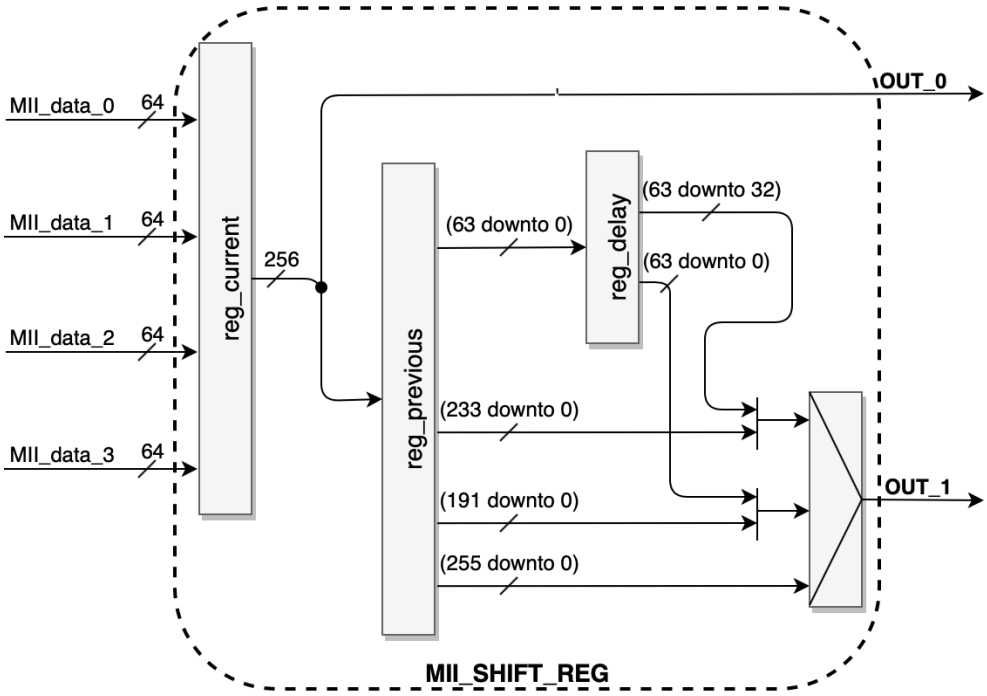

Figura 21 – Arquitetura do circuito MII_SHIFT_REG (fonte: Autor).

As saídas do circuito MII_SHIFT_REG são entradas para o módulo MII_SHIFTER. O circuito MII_SHIFTER reordena os dois barramentos de entrada para apenas um barramento de saída de 256 bits chamado **data_out**. O módulo permite buscar trechos do quadro transmitido nos registradores **out_0** e **out_1** para preencher **data_out** corretamente, eliminando as constantes de preâmbulo e *SFD* presentes nos quadros MII. Este circuito baseia-se em deslocamentos mínimos de 32 bits (preâmbulo + *SFD*). O barramento de saída do módulo MII_SHIFTER alimenta um conjunto de estruturas do tipo fila (*FIFO*) com os quadros MII ordenados ciclo a ciclo no barramento **data_out**. A Figura 22 ilustra a do circuito MII_SHIFTER.

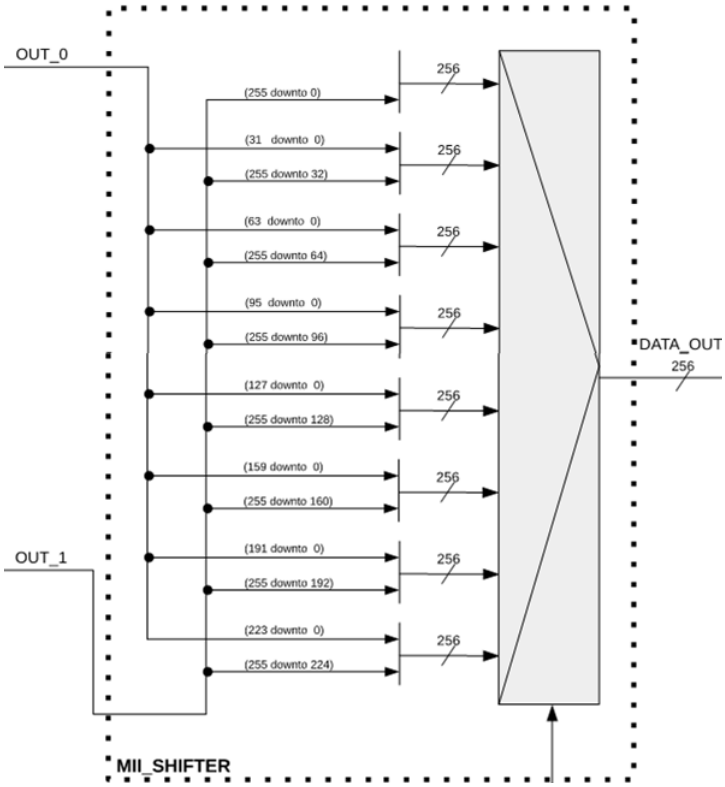

Figura 22 – Arquitetura do circuito MII_SHIFTER (fonte: Autor).

Com o objetivo de se obter os dados pertencentes aos quadros MII de forma sequencial e organizada para que seja calculado o valor CRC correspondente a cada quadro MII, o módulo FIFO_BRAM instancia quatro macros FIFO_DUAL_PORT, disponíveis pela empresa Xilinx. Estas estruturas do tipo fila são implementadas em Block RAMs.

O circuito FIFO_BRAM armazena os bytes do MAC destino até o último byte de CRC válido do mesmo quadro MII. Os dados são armazenados juntamente com as devidas sinalizações de início e término de pacote assim como a sinalização de validade do quadro transmitido. A leitura das FIFOs acontece de forma alternada, isto é, as informações armazenadas nas filas são direcionadas a saída do módulo FIFO_BRAM por um barramento de 128 bits. Para que a taxa de 40Gbps seja mantida e o barramento de dados possua 128 bits, as leituras dos blocos de dados devem acontecer ao dobro da frequência utilizada pelos barramentos de 256 bits, ou seja, a leitura deve ser realizada no dobro da frequência de decodificação dos PCS_CORE_RX.

A partir dos quadros MII organizados e enfileirados sequencialmente pelo circuito FIFO_BRAM, este é conectado com o último módulo de circuito interno da estrutura MAC40G_RX, o módulo CRC_B128. Este módulo gera o valor de CRC dos quadros MII recebidos e compara o valor calculado na recepção com o valor CRC transmitido no próprio quadro MII.

O sistema calcula o CRC sem que haja necessidade de gerar pausa na recepção dos dados, pois a retirada das constantes preâmbulo e SFD, somados com o mínimo intervalo permitido (IPG mínimo) entre dois quadros MII válidos na codificação XLGMII, totaliza no mínimo 20 bytes livres entre um quadro MII e outro. Assim, é garantido que o início do próximo quadro MII recebido, que é a próxima informação válida para o cálculo de CRC, está disponível no próximo ciclo.

Sendo assim, o circuito CRC_B128 possui todas as sinalizações necessárias para efetuar o cálculo de CRC do quadro MII recebido. O módulo garante a integridade do quadro recebido através do valor do CRC injetado na etapa de transmissão. O módulo também disponibiliza em suas saídas os valores originalmente transmitidos pela aplicação de rede, e juntamente com o último byte válido sinaliza se o quadro recebido está íntegro ou não.

5 RESULTADOS

Este Capítulo apresenta a simulação dos módulos de hardware desenvolvidos na plataforma *Eth40SR*, em VHDL. Os circuitos da *Eth40SR* foram simulados através da ferramenta Modelsim, sendo possível a visualização das formas de onda das interfaces de comunicação de entrada e saída. O objetivo é ilustrar em grandes blocos, as interfaces de entrada e saída descritas nas Seções de Transmissão 4.2 e Recepção 4.3 do Eth40SR. Para melhor compreensão da simulação executada, as formas de onda estão subdivididas em processo de transmissão e recepção.

5.1 Simulação do processo de transmissão

A Figura 23 apresenta as formas de ondas dos principais sinais de entrada e saída dos blocos de circuitos ilustrados na Figura 5. Os seguintes eventos são apresentados em numeração vermelha como segue:

1. Este bloco demonstra os sinais gerados pela aplicação de rede, onde um pacote de dados deve ser encapsulado no padrão XLGMII. Neste caso os estímulos são gerados através da linguagem *SystemC*. Este conjunto de sinais são conectados à interface de entrada MAC40G_TX.
2. Sinais de saída do módulo MAC40G_TX, onde estão os barramentos MII codificados no padrão XLGMII. Estes foram gerados através dos sinais conectados pela aplicação de rede.
3. Visualização dos blocos de 66 bits (header e data) embaralhados, gerados pelos módulos PCS_CORE_TX. Estes blocos alimentam o circuito PCS_ALIGNMENT o qual injeta os marcadores de alinhamento.
4. Este bloco representa a inserção dos marcadores de alinhamento gerados através do circuito PCS_ALGNMENT, possibilitando a identificação de cada *logical lane*. As formas de ondas representadas na cor magenta correspondem aos blocos de 66 bits conectados à camada PMA, para que esta direcione o conjunto de dados gerados aos *tranceivers ópticos.*

A barra vertical em 1625 ns corresponde ao primeiro momento de inserção dos marcadores de alinhamento. As constantes presentes nos sinais *lane_X_dataout* corresponde aos padrões de alinhamento apresentados na Tabela 1.

Os sinais *lane_X_header_out* (bloco com numeração 4) com valor 10 correspondem a aos dados sendo transmitidos. Assim a latência interna de transmissão do módulo de transmissão é de aproximadamente 1.000 ns.

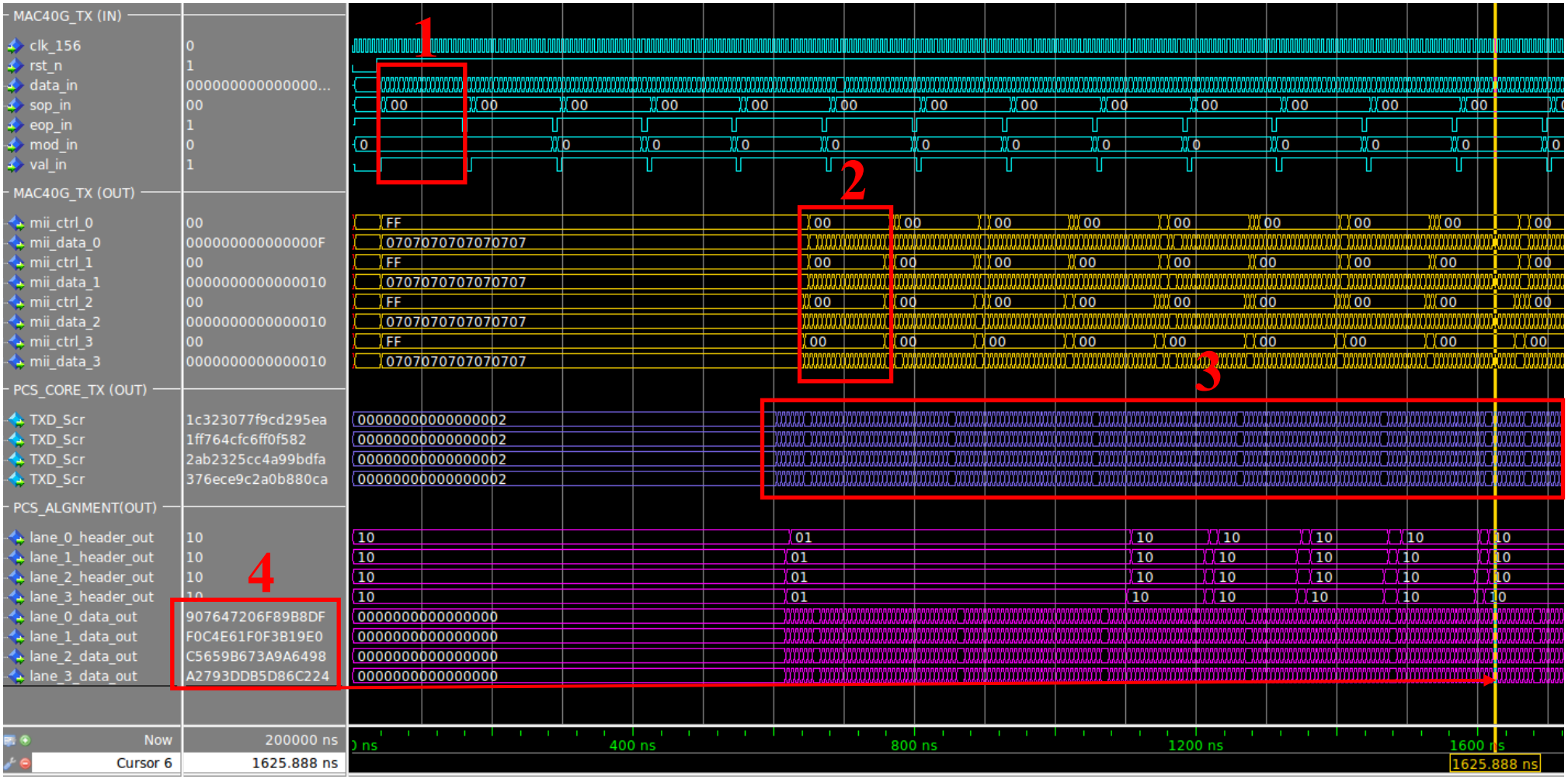

Figura 23– Simulação do processo de transmissão (fonte: Autor).

5.2 Simulação do processo de recepção

Esta Seção apresenta a simulação referente a etapa de recepção detalhada na seção 4.3 deste documento. A ilustração das formas de onda é dividida nas Figura 24 e Figura 25. A explicação destas é dada pela sequência de eventos na etapa de recepção, numeradas na cor vermelho, análogas a ilustração do processo de transmissão.

1. Aqui são representados os sinais *header* e *data* recebidos através das interfaces PMA, sob o ponto de vista de recepção. No ambiente de simulação criado, estes são conectados aos sinais de saída do módulo PCS_ALIGNMENT, presente na etapa de transmissão (ou seja, transmissão e recepção em *loop-back*). Os dados recebidos nesta interface de entrada são passíveis de atraso temporal, bem como conexão incorreta entre os blocos (*header* e *data*) de cada fluxo.
2. A partir deste ponto, os blocos são entregues pela interface de saída do módulo LANE_REORDER, ordenados e alinhados
3. O módulo LANE_REORDER identifica e sinaliza os marcadores de alinhamento de forma a disponibilizar a sinalização gerada à interface de entrada do circuito PCS_ALGNMENT_REMOVAL.
4. Este evento refere-se à interface de saída do circuito PCS_ALIGNMENT_REMOVAL. Os blocos (*header* e *data*) são direcionados em quatro fluxos à interface de entrada do circuito PCS_CORE_RX. Neste ponto, são extraídos os marcadores de alinhamento e compensados através da inserção de IPG.
5. O quinto evento refere-se à decodificação através dos PCS_CORE_RX. Os sinais destacados em amarelo representam barramentos *data* e *control*, relacionados ao padrão XLGMII.
6. Neste ponto está representada a interface de saída do circuito CRC_B128, presente no conjunto de módulos de circuito MAC40G_RX. Os sinais destacados na cor ciano referem-se aos barramentos de dados de 128 bits e suas sinalizações de início e término de quadro bem como a sinalização de CRC a qual identifica a integridade dos quadros MII recebidos. A partir deste ponto estão disponíveis os dados gerados originalmente, prontos para serem consumidos pela aplicação de rede na recepção.

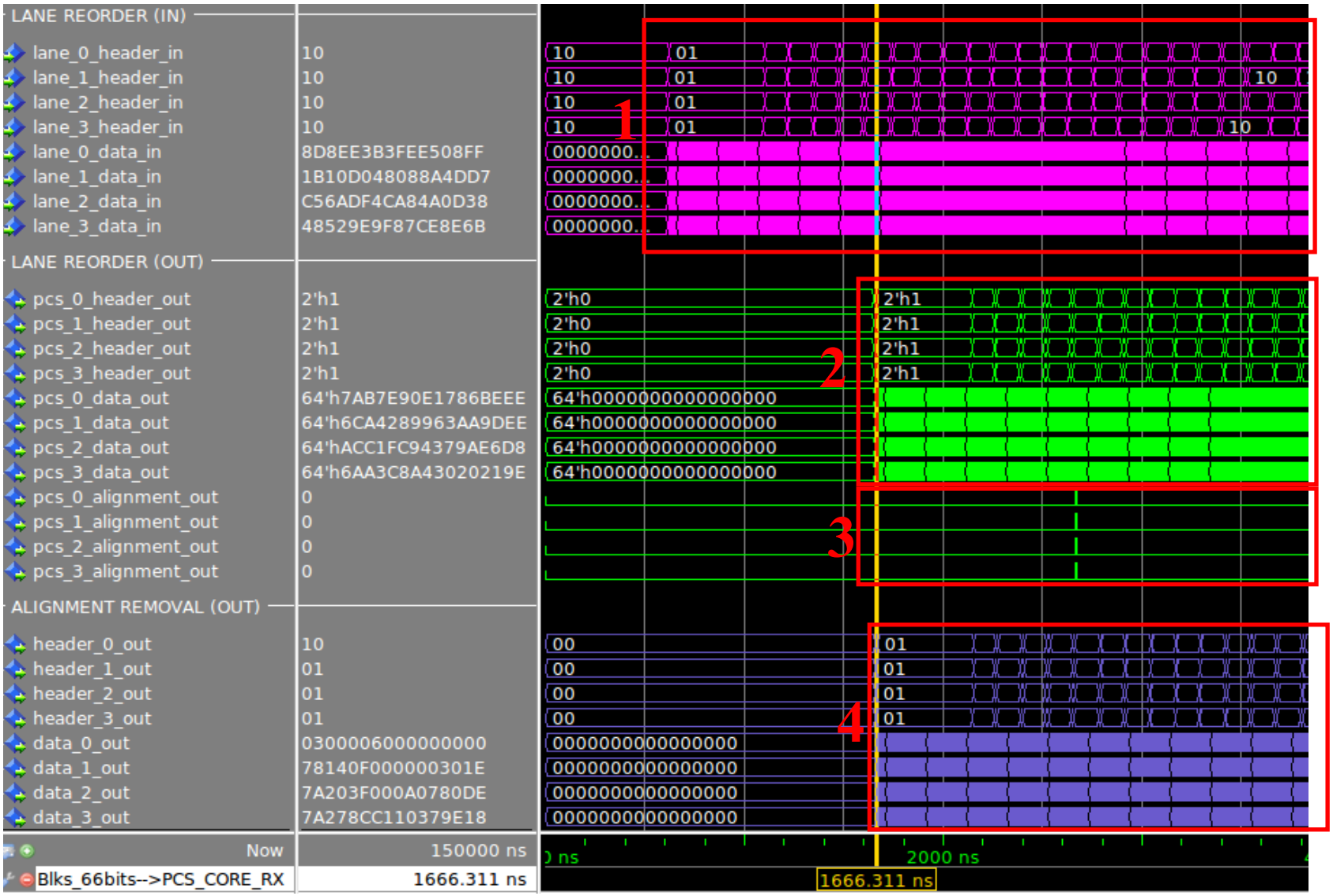

Figura 24 – Simulação do processo de recepção - parte 1 (fonte: Autor).

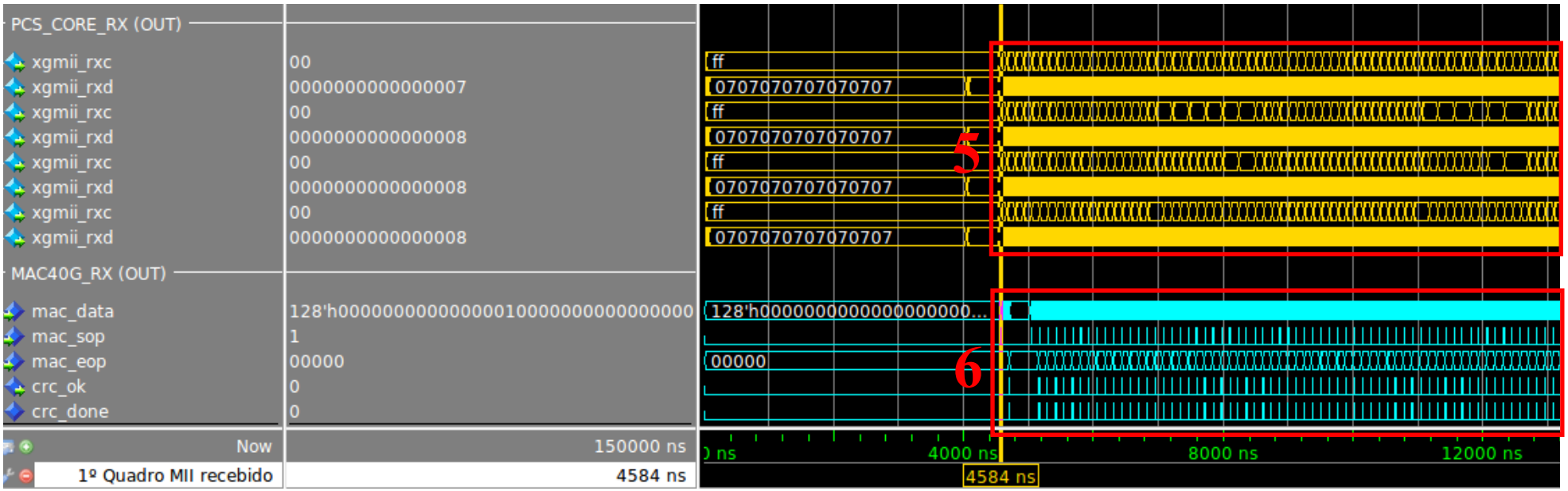

Figura 25 – Simulação do processo de recepção - parte 2 (fonte: Autor).

6 CONCLUSÃO

Este trabalho apresentou uma arquitetura apta a operar em uma taxa de 40 Gbps, a qual utilizou um trabalho de referência que implementa um único fluxo de dados em uma taxa de 10 Gpbs. A arquitetura desenvolvida é composta por quatro fluxos paralelos interconectados de modo que a taxa alvo seja obtida. A abordagem foi definida de maneira a manter a frequência original dos PCSs, 156.25 MHz.

No nível da camada PCS desenvolveu-se uma série de interconexões entre os módulos PCS, para que estes possibilitem o tráfego de dados a taxa de 40 Gbps. Esta tarefa demandou alto esforço de projeto, pois foi necessário compreender um *IP core* de terceiros, de forma a modificá-lo para atender o sincronismo entre os mesmos. Após esta etapa, percebeu-se que não era possível utilizar o MAC 10 Gbps, dado que para utilizá-lo seria necessária a utilização de uma frequência não suportada pelo FPGA, pois o mesmo não suporta frequências acima de 350 MHz. Assim, foi desenvolvido um conjunto de módulos descritos em linguagem de descrição de hardware, capazes de interpretar informações de uma aplicação de rede genérica e encapsular tais informações no padrão de codificação XLGMII, resultando em um MAC 40 Gbps.

O desenvolvimento deste trabalho foi construído baseado no padrão disponibilizado pelo IEEE para redes de 40 e 100 Gbps. Os módulos de hardware desenvolvidos podem servir de base para desenvolvimento de novas aplicações que demandem alta velocidade de comunicação.

Em relação ao curso, o principal estímulo para o desenvolvimento deste TCC foi a oportunidade de agregar a um estudante de Engenharia Elétrica a capacidade de compreender e gerar soluções para questões relacionadas a sistemas de telecomunicações de alto desempenho. Além do mais, possibilitou ter uma vivência prática com ferramentas atuais para resolução de problemas deste porte. O trabalho possibilitou a vivência em assuntos de grande relevância para a indústria, além de ofertar uma solução que faz uso de recursos disponíveis para trabalhar com interfaces de rede de alta velocidade.

O conjunto de módulos de hardwares desenvolvidos neste trabalho pode servir como base para trabalhos futuros. Duas linhas de trabalhos podem ser elencadas. A primeira é a efetiva prototipação dos módulos desenvolvidos em FPGA, pois os mesmos foram apenas validados por simulação. A segunda direção de trabalho futuro consiste na integração com aplicações de rede, como um testador que segue a norma RFC2544.

7 REFERÊNCIAS BIBLIOGRÁFICAS

[ATH05] Athavale, A.; Christensen, C. “High-Speed Serial I/O Made Simple”. Disponível em: http://xilinx.eetrend.com/files-eetrend-xilinx/forum/201407/7408-13218-xilinx_serial_io_101_gao_su_chuan_xing_ru_men_.pdf , Abril, 2005, 210p.

[CIS16] Cisco, Inc. “Cisco Visual Networking Index Predicts Global Annual IP Traffic to Exceed Three Zettabytes by 2021”. Disponível em: https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1853168, Disponível em: Junho 2017.

[COM13] Comer, D. “Internetworking with TCP/IP Volume One”, Pearson, sixth edition, Maio, 2013.

[DIG15] Digilent Inc. “Net FPGA Sume”. Disponível em: http://digilentinc.com/sume, Setembro, 2015.

[FRA00] Frazier, H. “IEEE P802.3ae - 10 Gigabit Ethernet Task Force: XGMII Update”, Disponível em: http://www.ieee802.org/3/ae/public/jul00/frazier_1_0700.pdf, Julho, 2000.

[GUS08] Gustlin, M. “XL/CGMII and RS Proposal”, Disponível em: http://www.ieee802.org/3/ba/public/may08/gustlin_02_0508.pdf, Maio, 2008.

[IEE02] The Institute of Electrical and Electronic Engineers Inc. (IEEE) “IEEE Std 802.3ae”, Disponível em: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1040118, Agosto, 2002.

[IEE10] The Institute of Electrical and Electronic Engineers Inc. (IEEE) “IEEE Std 802.3ba”, Disponível em: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5501740, Junho 2010.

[JUR17] Juracy, L.; Lazzarotto, F.; Pigatto, D.; Calazans, N.; Moraes, F. “XGT4: An Industrial Grade, Open Source Tester for Multi-Gigabit Networks”. In: ICECS, 2017, 4p.

[MIC14] Microsoft Support "The OSI Model's Seven Layers Defined and Functions Explained", Disponível em: https://support.microsoft.com/en-us/kb/103884, Março, 2014.

[OPE08] Open Cores “Ethernet 10GE MAC”, Disponível em: www.opencores.org/project,xge_mac, Maio, 2008.

[SFF09] SFF Committee. “SFF-8431 SFP+ 10 Gb/s and Low Speed Electrical Interface”. Disponível em: ftp://ftp.seagate.com/sff/SFF-8431.PDF, Julho 2009.

[SPU14] Spurgeon, C., Zimmerman, J. “Ethernet: The Definitive Guide”, O’Reilly, second edition, Abril, 2014.

[THU15] Thum, E. V. " Comunicação de Alto Desempenho em Dispositivos FPGAs Através de Interfaces Ópticas ", Trabalho de Conclusão de Curso, Engenharia de Computação, PUCRS, 2015, 70p.

[XIL16] Xilinx Inc. “10G Ethernet PCS/PMA v6.0”. Disponível em: https://www.xilinx.com/support/documentation/ip_documentation/ten_gig_eth_pcs_pma/v6_0/pg068-ten-gig-eth-pcs-pma.pdf, Disponível em: Outubro, 2016.