

ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO

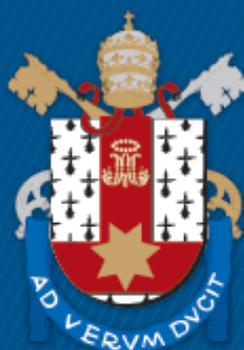
DEIVID ANTUNES TESCH

CSMA COMO MAC PARA DWINOCS

Porto Alegre

2020

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica
do Rio Grande do Sul

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

CSMA COMO MAC PARA DWINOCS

DEIVID ANTUNES TESCH

Tese apresentada como requisito parcial
à obtenção do grau de Doutor em
Ciência da Computação na Pontifícia
Universidade Católica do Rio Grande do
Sul.

Orientador: Prof. Fabiano Passuelo Hessel

**Porto Alegre
2020**

Ficha Catalográfica

T337c Tesch, Deivid Antunes

CSMA como MAC para DWiNoCs / Deivid Antunes Tesch. – 2020.
121 p.
Tese (Doutorado) – Programa de Pós-Graduação em Ciência da
Computação, PUCRS.

Orientador: Prof. Dr. Fabiano Passuelo Hessel.

1. NoC. 2. WiNoC. 3. DWiNoC. 4. CSMA. 5. MAC. I. Hessel, Fabiano
Passuelo. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051

Deivid Antunes Tesch

CSMA como MAC para DWiNOCs

Tese apresentada como requisito parcial para obtenção do grau de Doutor em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação, Escola Politécnica da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovado em 24 de agosto de 2020.

BANCA EXAMINADORA:

Prof. Dr. Jarbas Aryel Nunes da Silveira (PPGETI/UFC)

Prof. Dr. Leonel Pablo Carvalho Tedesco (PPGSPI/UNISC)

Prof. Dr. César Augusto Missio Marcon (PPGEE/PUCRS)

Prof. Dr. Fabiano Passuelo Hessel (PPGCC/PUCRS - Orientador)

DEDICATÓRIA

Dedico este trabalho à minha família.

AGRADECIMENTOS

Gostaria de agradecer a todos que de alguma forma se envolveram com esse trabalho.

Ao meu orientador, professor Fabiano Passuelo Hessel, primeiramente pela oportunidade, mas também pela compreensão, confiança e pelo direcionamento por toda essa caminhada.

Ao colega Everton Berz pelo suporte e pela parceria na escrita de artigos.

Aos colegas da RBS TV e Nelogica pelo apoio e compreensão.

Por fim, aos meus pais, Carlos (*in memoriam*) e Eliane e aos meus irmãos, Douglas e Carla, por todo apoio, dedicação, compreensão e força para enfrentar os desafios.

CSMA COMO MAC PARA DWINOCs

RESUMO

A Rede-em-Chip Sem Fio (*Wireless Network-on-Chip* — WiNoC) surge como uma das mais promissoras tecnologias para enfrentar os problemas de escalabilidade da Rede-em-Chip (*Network-on-Chip* — NoC) convencional. Entretanto, fornecer acesso múltiplo e simultâneo ao meio sem fio compartilhado é um dos principais desafios dessa tecnologia e o uso de múltiplos canais não sobrepostos não é uma solução escalável devido à limitação da largura de banda disponível e da complexidade do transceptor. Para lidar com essa limitação, a WiNoC com Antenas Direcionais (*Directional Wireless Network-on-Chip* — DWiNoC) utiliza outra abordagem que tira proveito das características das antenas direcionais para criar múltiplos canais não sobrepostos. Entretanto, o mecanismo de controle de acesso ao meio utilizado pela DWiNoC é energeticamente ineficiente para baixas cargas de tráfego. Conhecendo a capacidade que o mecanismo de Acesso Múltiplo com Detecção da Portadora (*Carrier Sense Multiple Access* — CSMA) tem de responder bem a essa categoria de tráfego, esse trabalho propõe o uso desse como mecanismo de controle de acesso ao meio para a DWiNoC, visando melhorar a eficiência energética do canal mantendo um nível aceitável de desempenho. O uso desse mecanismo pode reduzir, em média, entre 39 e 42% o consumo de energia de cada canal, considerando uma distribuição uniforme do tráfego de entrada. Isso, sem impactar significativamente a taxa de transferência e a latência para cargas de tráfego de entrada inferiores a 40%.

Palavras-Chave: NoC, WiNoC, DWiNoC, CSMA, MAC.

CSMA AS MAC FOR DWINOCs

ABSTRACT

Wireless Network-on-Chip (WiNoC) emerges as one of the most promising technologies to face the scalability problems of conventional NoCs (Network-on-Chip). However, providing multiple and simultaneous access to the shared wireless medium is one of the main challenges of this technology and the use of multiple non-overlapping channels is not a scalable solution due to the limitation of the available bandwidth and of the transceiver complexity. To address this limitation, Directional Wireless Network-on-Chip (DWiNoC) uses another approach that takes advantage of directional antennas characteristics to create multiple non-overlapping channels. However, the media access control mechanism used by DWiNoC is energy inefficient for low traffic loads. Knowing the capacity that CSMA has to deal well with this category of traffic, this work proposes the use of it as a media access control mechanism for DWiNoCs, with the objective of improving channel energy efficiency while maintaining an acceptable level of performance. The use of this mechanism can reduce, on average, from 39 to 42% the energy consumption of each channel, considering a uniform distribution of the incoming traffic. This, without significantly impacting throughput and latency for inbound traffic loads below 40%.

Keywords: NoC, WiNoC, DWiNoC, CSMA, MAC.

LISTA DE FIGURAS

2.1	Possível organização de uma arquitetura SoC combinando uma NoC 3D e uma NoC fotônica. Fonte: [CPX09].	35
2.2	Categorias de linhas de transmissão. Fonte: [KMTY16]	36
2.3	NoC simétrica. Fonte: [NND10].	37
2.4	Arquitetura da NoC 3D híbrida. Fonte: [CPX09].	38
2.5	Arquitetura da NoC 3D com <i>crossbar</i> em 3D. Fonte: [CPX09].	38
2.6	Intervalos da frequência de operação das diferentes tecnologias de redes em chip sem fio. Fonte: [DGP+12].	40
2.7	Leiaute da antena em formato de zigue-zague. Fonte: [FHO02].	41
2.8	Antena log periódica em chip. Fonte: [SRD14].	42
2.9	Corte transversal da antena PLPA sob o substrato de silício. Fonte: [SRD14].	43
2.10	Perda de retorno da antena PLPA. Fonte: [MGS+17]	43
2.11	Padrão de radiação da antena PLPA ao longo dos planos azimutal e elevação. Fonte: [MGS+17]	44
2.12	Variação do parâmetro $S(2, 1)$ da antena PLPA com a distância. Fonte: [MGS+17]	44
2.13	Variação do parâmetro $S(2, 1)$ com a orientação da antena. Fonte: [MGS+17]	45
2.14	Taxa de bits, taxa de baud e largura de banda do BASK. Fonte: [For12]. . . .	46
2.15	Espectro de potência de um sinal BASK. Fonte: [HM07].	47
2.16	Conceito de funcionamento do BASK. Fonte: [For12].	48
2.17	Modulação FSK. Fonte: [Fre15].	48
2.18	Representação das taxas de bit, baud e largura de banda do BFSK. Fonte: [For12].	49
2.19	Espectro de potência de um sinal BFSK. Fonte: [HM07].	49
2.20	Representação simplificada da implementação de um modulador BFSK coerente. Fonte: [For12].	50
2.21	Representação da modulação PSK. Fonte: [Fre15].	50
2.22	Representação das taxas de bit, baud e largura de banda do BPSK. Fonte: [For12].	51
2.23	Espectro de potência de um sinal BPSK. Fonte: [HM07].	51
2.24	Representação simplificada da implementação de um modulador BPSK coerente. Fonte: [For12].	52
2.25	Receptor de conversão direta. Fonte: [Fre15].	52

2.26	Receptor de conversão direta para FSK e PSK. Fonte: adaptado de [Fre15].	54
2.27	Diagrama de blocos de um transceptor OOK. Fonte: [MGS ⁺ 17].	55
2.28	<i>McWiNoC</i> baseada em uma malha 4 × 4. Fonte: [ZWLK11].	56
2.29	Variação das topologias em função do alcance de transmissão. Fonte: [ZWLK11].	57
2.30	<i>WiNoC</i> 10 × 10. Fonte: [WHB11].	58
2.31	(a) Topologia em malha de uma sub-rede e (b) Topologia <i>small-world</i> interligando os <i>hubs</i> . Fonte: [GCD ⁺ 11].	59
2.32	Espectro de frequência do FDMA. Fonte: [LCL07].	60
2.33	<i>slots</i> do TDMA. Fonte: [LCL07].	61
2.34	Transmissor com espalhamento espectral por sequência direta. Fonte: [Fre15].	62
2.35	Forma de onda dos sinais envolvidos. Fonte: [Fre15].	62
2.36	Regra geral (a) e exemplos (b) de criação de tabelas <i>Walsh</i> . Fonte: [For12].	63
2.37	Modelo espaço-tempo de uma colisão no CSMA. Fonte: [For12].	64
2.38	Tempo vulnerável no CSMA. Fonte: [For12].	65
2.39	Comportamento dos três métodos de persistência. Fonte: [For12].	65
2.40	Diagrama de fluxo dos três métodos de persistência. Fonte: [For12].	66
3.1	Arquitetura do TMU e do <i>flit</i> que atua como ficha. Fonte: [MGY13].	70
3.2	Mecanismo de passagem da ficha implementado sob uma topologia em anel. Fonte: [PCMC15].	71
3.3	Arquitetura comum do P-TDMA e do D-TDMA. Fonte: [MSG16].	72
3.4	Arquitetura da CDMA- <i>WiNoC</i> . Fonte: [VYM ⁺ 14].	74
3.5	Transmissor e receptor da CDMA- <i>WiNoC</i> . Fonte: [VYM ⁺ 14].	74
3.6	Primeiro e segundo níveis de uma <i>WCube</i> . Fonte: [LTP ⁺ 09].	76
3.7	Arquitetura <i>HiWA</i> . Fonte: [RDSZ16].	76
3.8	Visão geral do CSMA <i>p</i> -persistente. Fonte: [CL14].	77
3.9	Diagrama de fluxo do transmissor (à esquerda) e do receptor (à direita) do BRS-MAC. Fonte: [MAT ⁺ 16].	78
3.10	Arquitetura da <i>WiNoC</i> reconfigurável sob demanda. Fonte: [MG15].	79
3.11	Arquitetura da unidade de MAC dinâmica. Fonte: [MG15].	81
4.1	Topologia da <i>DWiNoC</i> . Fonte: [MGS ⁺ 17].	86
4.2	Algoritmo de posicionamento das <i>WIs</i> . Fonte: [MGS ⁺ 17].	89
5.1	Consumo de energia <i>E</i> para transmissão de <i>flits</i> de dados utilizando o mecanismo de passagem da ficha em função da carga de tráfego de entrada <i>G</i> e para diferentes valores do THT. Fonte: o autor.	97

5.2	Consumo de energia E para transmissão dos <i>flits</i> com a ficha do mecanismo de passagem da ficha em função da carga de tráfego de entrada G para diferentes valores do THT. Fonte: o autor.	98
5.3	Razão entre a energia consumida para transmissão de <i>flits</i> com a ficha e entre a energia consumida para transmissão de <i>flits</i> de dados, em função da carga de tráfego de entrada G e para diferentes valores do THT. Fonte: o autor.	99
5.4	Contribuição individual dos <i>flits</i> com a ficha e dos <i>flits</i> de dados no consumo total de energia em relação à carga de tráfego de entrada para diferentes valores de THT. Fonte: o autor.	100
5.5	Consumo de energia sumarizado dos <i>flits</i> com a ficha e dos <i>flits</i> de dados para diferentes valores de THT. Fonte: o autor.	102
5.6	Taxa de transferência do mecanismo de passagem da ficha em relação à carga de tráfego de entrada para diferentes valores do THT. Fonte: o autor.	103
5.7	Taxa de transferência do CSMA não-persistente em relação à carga de tráfego de entrada. Fonte: o autor.	104
5.8	Taxa de transferência do <i>slotted</i> CSMA não-persistente em relação à carga de tráfego de entrada. Fonte: o autor.	105
5.9	Comparação da taxa de transferência entre o mecanismo de passagem da ficha (com diferentes valores do THT), o CSMA não-persistente e o <i>slotted</i> CSMA não-persistente em função da carga de tráfego de entrada. Fonte: o autor.	106
5.10	Latência do mecanismo de passagem da ficha em função da taxa de transferência para diferentes valores do THT. Fonte: o autor.	107
5.11	Latência do CSMA não-persistente em função da taxa de transferência. Fonte: autor.	108
5.12	Latência para o <i>slotted</i> CSMA não persistente em função da taxa de transferência. Fonte: o autor.	109
5.13	Comparação entre a taxa de retransmissão do CSMA não-persistente e do <i>Slotted</i> CSMA não-persistente em função da taxa de transferência. Fonte: o autor.	110
5.14	Comparação da latência entre o mecanismo de passagem da ficha, o CSMA não-persistente e o <i>slotted</i> CSMA não-persistente em função da taxa de transferência. Fonte: o autor.	111

LISTA DE TABELAS

2.1	Dimensões da antena. Fonte: [SRD14].	42
3.1	Características dos mecanismos de controle de acesso ao meio.	82
5.1	Estatísticas do consumo de energia para transmissão de <i>flits</i> com dados. . .	97
5.2	Estatísticas do consumo de energia para transmissão de <i>flits</i> com a ficha. .	98
5.3	Estatísticas dos dados da razão entre o consumo de energia para transmissão dos <i>flits</i> com a ficha e o consumo de energia para transmissão dos <i>flits</i> com dados.	99
5.4	Estatísticas da taxa de transferência do mecanismo de passagem da ficha. .	103
5.5	Estatísticas da taxa de transferência do CSMA não-persistente.	105
5.6	Estatísticas da taxa de transferência do <i>slotted</i> CSMA não-persistente. . . .	106
5.7	Estatísticas dos dados de latência do mecanismo de passagem da ficha. . .	108
5.8	Estatísticas dos dados normalizados da latência do CSMA não-persistente. .	109
5.9	Estatísticas dos dados normalizados da latência do <i>Slotted</i> CSMA não-persistente.	109
5.10	Estatísticas de retransmissões do <i>Slotted</i> CSMA não-persistente e do CSMA não-persistente.	110

LISTA DE SIGLAS

3D – Three-dimensional
ACK – Acknowledgement
ADC – Analog-to-digital Converter
AM – Action Manager
AM – Amplitude Modulation
ARM – Advanced RISC Machines
ASK – Amplitude Shift Keying
BASK – Binary Amplitude Shift Keying
BEB – Binary Exponential Backoff
BER – Bit Error Ratio
BFSK – Binary Frequency Shift Keying
BPSK – Binary Phase-shift Keying
BRS-MAC – Broadcast, Reliability and Sensing MAC
CDMA – Code Division Multiple Access
C-Mesh – Concentrated-mesh
CMOS – Complementary Metal Oxide Semiconductor
CNT – Carbon Nanotubes
CPS – Coplanar Strip
CPW – Coplanar Waveguide
CSMA/CA – Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD – Carrier Sense Multiple Access with Collision Detection
CW – Continuous Wave
DAC – Digital-to-analog Converter
dB – Decibel
DC – Direct Conversion
DSP – Digital Signal Processor
DSSS – Direct-sequence Spread Spectrum
D-TDMA – Dynamic TDMA
FDMA – Frequency Division Multiple Access
FSK – Frequency Shift Keying
GHz – Gigahertz
HiWA – Hierarchical Wireless-based Architecture

IBM – International Business Machines
IC – Integrated Circuit
IF – Intermediate Frequency
I – In-phase
IP – Intellectual Property
LFSR – Linear Feedback Shift Register
LNA – Low-noise Amplifier
LO – Local Oscillator
LPF – Low-pass Filter
MAC – Medium Access Control
McWiNoC – Multi-channel WiNoC
mmWave – Millimeter Wave
MSL – Microstrip Line
NACK – Negative Acknowledgment
NRZ – Non-return-to-zero
OOK – On-off Keying
PE – Processing Element
PE – Processing Element
PLL – Phase-locked Loop
PLPA – Planar Log-periodic Antenna
PSK – Phase-shift Keying
P-TDMA – Proportionate TDMA
Q – Quadrature
QPSK – Quadrature Phase Shift Keying
RF-I – RF-interconnect
RF – Radio Frequency
SIR – Signal-to-interference Ratio
SoC – System-on-Chip
Sub-THz – Sub-Terahertz
TDMA – Time Division Multiple Access
THT – Token Holding Time
THz – Terahertz
TMU – Token Management Unit
TSV – Through-silicon Via

UWB – Ultra-wideband

VC – Virtual Channel

VCO – Voltage-controlled Oscillator

WI – Wireless Interfaces

WiNoC – Wireless Network-on-Chip

WR – Wireless Routers

XOR – Exclusive Or

ZIF – Zero-IF

SUMÁRIO

1	INTRODUÇÃO	27
1.1	OBJETIVOS	30
1.2	CONTRIBUIÇÃO ORIGINAL	31
1.3	ORGANIZAÇÃO	31
2	CONCEITOS BÁSICOS	33
2.1	NOC	33
2.1.1	OUTROS PARADIGMAS DE INTERCONEXÃO EM CHIP	34
2.2	WINOCS	39
2.2.1	ANTENAS	40
2.2.2	TRANSCEPTORES	45
2.2.3	TOPOLOGIA	55
2.3	MECANISMOS MAC	59
2.3.1	MACS DE ACESSO CANALIZADO	60
2.3.2	MACS DE ACESSO ALEATÓRIO	63
2.3.3	MACS DE ACESSO CONTROLADO	66
3	ESTADO-DA-ARTE EM MECANISMOS MAC PARA WINOCS	69
3.1	PASSAGEM DA FICHA	69
3.2	TDMA	70
3.3	CDMA	74
3.4	FDMA	75
3.5	CSMA	76
3.6	DISCUSSÃO	82
4	DESCRIÇÃO DA ARQUITETURA	85
4.1	ARQUITETURA DA DWINOC	85
4.1.1	TOPOLOGIA	85
4.1.2	POSICIONAMENTO DAS WIS	85
4.2	<i>SLOTTED</i> CSMA NÃO-PERSISTENTE	90
5	RESULTADOS	91
5.1	MODELO ANALÍTICO	91

5.1.1	CSMA.....	91
5.1.2	PASSAGEM DA FICHA	93
5.2	ANÁLISE DO CONSUMO DE ENERGIA	96
5.3	ANÁLISE DA TAXA DE TRANSFERÊNCIA.....	102
5.4	ANÁLISE DA LATÊNCIA	106
6	CONCLUSÃO	113
6.1	PUBLICAÇÕES	113
6.2	TRABALHOS FUTUROS	114
	REFERÊNCIAS	115

1. INTRODUÇÃO

A recente escalada tecnológica, impulsionada pela crescente demanda computacional de diversos segmentos de mercado como, por exemplo, o financeiro, o da bioinformática, o da computação gráfica, o da previsão do tempo entre muitos outros, gerou uma corrida para integrar cada vez mais circuitos em um único chip. Isso foi possível graças a redução na tecnologia de fabricação do chip. Por um lado, ela permitiu a redução no consumo de energia e o aumento na frequência de operação desses circuitos, mas, por outro lado, com mais circuitos integrados em um único chip, operando a frequências cada vez maiores, intensificou a demanda do chip por energia e agravou o problema da sua dissipação.

Os sistemas multinúcleo (*multicore*) surgiram como uma solução para o problema de dissipação de energia. Em um sistema multinúcleo vários elementos de processamento (*Processing Element* — PE), são integrados num mesmo *die*¹. Isso permite que diversas tarefas sejam executadas em paralelo, o que geralmente aumenta o desempenho. O PE opera a frequências inferiores, o que, em geral, contribui para a redução no consumo de energia e minimiza os problemas de dissipação de energia. O aumento no grau de paralelismo das aplicações nos últimos anos demandou um crescimento no número de PEs integrados em um único circuito integrado. Isso provocou um aumento na troca de mensagens entre PEs, que tornou mais acirrada a disputa pelo barramento de comunicação compartilhado, agravando o problema de latência na comunicação.

Diversas arquiteturas de barramentos foram propostas para tentar reduzir a latência em sistemas multinúcleo como, por exemplo, a ARM AMBA, a Wishbone e a IBM CoreConnect. No entanto, essas arquiteturas não conseguiram atender a demanda crescente de comunicação, que afetou a latência da comunicação significativamente [PGJ⁺05]. Novos paradigmas de comunicação em chip foram explorados para tentar reduzir essa latência de comunicação. A rede em chip (*Network-on-Chip*, NoC) surgiu como uma promessa para resolver esse problema. Na NoC, os blocos de Propriedade Intelectual (*Intellectual Property* — IP), que podem ser elemento de processamento ou de armazenamento, ficam separados da comunicação. Os blocos IP podem se comunicar entre si através de roteadores conectados à cada bloco IP.

Embora a malha de interconexão tenha evoluído do barramento para a NoC, o meio de interconexão, por fio metálico, permaneceu o mesmo [KMTY16]. Essa categoria de interconexão é uma forma simples e barata de implementar uma interconexão em chip. Nela o sinal é transmitido carregando ou descarregando o fio inteiro. Com a redução da tecnologia CMOS (*Complementary Metal Oxide Semiconductor*), o espaçamento e a espessura desse fio vêm diminuindo continuamente e conseqüentemente aumentado sua

¹ *Die* é um pequeno bloco de material semicondutor onde um circuito integrado é fabricado.

capacitância e resistência [ITR20a] [ITR20b], respectivamente. Isso fez crescer o atraso, visto que ele é inversamente proporcional à resistência e a capacitância [DP98].

Para tentar amenizar esse problema, alguns trabalhos propuseram o uso de fios metálicos de longo alcance [OM06] ou de canais expressos de comunicação entre os nós [KPKJ07]. Outra abordagem propôs o uso de repetidores, que podem ser usados para tentar manter o atraso com um aumento linear com a escalada da tecnologia [BM02]. Embora os repetidores mitiguem o aumento do atraso, eles aumentam significativamente o consumo de energia das interconexões metálicas. Isso limita a escalabilidade no uso de repetidores para mitigar o atraso nas interconexões com fios metálicos e ainda assim o ganho de desempenho dessas abordagens fica limitado ao paradigma de comunicação por fio metálico.

Para as próximas gerações de sistemas multinúcleo, estima-se que até 80% da energia total do chip será consumida pelas interconexões [MKWS04]. Estima-se também que as próximas gerações de sistemas multinúcleo vão demandar a integração de 100 vezes mais núcleos em um único chip em relação aos atuais no estado-da-arte [ITR20b]. A interconexão com fio metálico não está preparada para suprir essa demanda e limitaria o ganho de desempenho provido por essas novas gerações de sistemas multinúcleo.

Para tirar proveito do desempenho desses futuros sistemas multinúcleo, novos paradigmas de interconexão estão sendo explorados, como a fotônica, a por radiofrequência (*Radio Frequency* — RF), a integração em três dimensões (*Three-dimensional* — 3D) e a interconexão sem fio (*wireless*). Todas essas tecnologias têm em comum o benefício de reduzir o atraso e a dissipação de energia, mas, por outro lado, trazem novos desafios de projeto que precisam ser superados.

Dentre esses paradigmas de interconexões emergentes, que ainda enfrentam dificuldades de fabricação e complexidades de projeto, a interconexão sem fio é uma alternativa promissora, pois pode ser implementada utilizando a tecnologia CMOS atual. A interface sem fio (*Wireless Interface* — WI) com transceptores e antenas implementados em chip, permite a comunicação de longa distância, com alta largura de banda e baixa latência. Diversas tecnologias de antenas e transceptores em chip já foram propostas, como a banda ultra larga (*Ultra-wideband* — UWB) [LTP⁺09], nanotubos de carbono (*Carbon Nanotubes* — CNTs) [Han05] e grafeno [SMSG17]. Entretanto, o UWB não consegue atingir longas distâncias em chip, as antenas CNT ou de grafeno ainda estão passando pelo desafio da integração no processo CMOS. Por outro lado, interligações sem fio compatíveis com CMOS utilizando a tecnologia de ondas milimétricas (*Millimeter Wave* — mmWave) vem se mostrando bem promissoras [MGS⁺17].

As antenas utilizadas pela tecnologia mmWave ficam situadas sob a mesma camada de dióxido de silício (SiO_2) e a propagação ocorre através dela e a partir de uma camada inferior de substrato de silício (Si) [ZCS07]. A maioria das arquiteturas de NoC com WI, (*Wireless Network-on-Chip* — WiNoC) existentes, exploram o uso de antenas om-

nidirecionais², como antenas do tipo zigue-zague [FHO02]. Recentemente, outros trabalhos exploraram o uso de antenas direcionais, como antenas planares log periódicas [SRD14], que conseguem concentrar a maior parte da energia irradiada em uma determinada direção. Esse tipo especial de WiNoC com antenas direcionais é denominada como DWiNoC (*Directional Wireless Network-on-Chip* — DWiNoC).

O acesso ao meio sem fio compartilhado é coordenado por um mecanismo de controle de acesso ao meio (*Medium Access Control* — MAC). Uma das abordagens utilizadas por mecanismos MAC no ambiente em chip é a divisão da largura da banda disponível em sub-bandas, seja ela no domínio do tempo, da frequência ou ainda utilizando um esquema de codificação especial. Essas sub-bandas são canais dedicados e permitem a transmissão simultânea a partir do mesmo meio sem fio compartilhado, sem provocar nenhuma interferência. O acesso múltiplo por divisão de frequência (*Frequency Division Multiple Access* — FDMA) [LTP⁺09], o acesso múltiplo por divisão de código (*Code Division Multiple Access* — CDMA) [VYM⁺14] ou ainda o acesso múltiplo por divisão do tempo (*Time Division Multiple Access* — TDMA) são exemplos de mecanismos MAC que utilizam essa abordagem. Entretanto, essas abordagens compartilham de problemas de escalabilidade, seja ela por complexidade do transceptor ou limitação da largura de banda disponível.

A passagem da ficha é outra categoria de mecanismo MAC, ela utiliza o conceito de acesso controlado para compartilhar o meio sem fio. Muitas arquiteturas de WiNoC e DWiNoC adotaram a passagem da ficha como mecanismo MAC. Nela, a ficha circula entre todas as WIs e somente a WI que detêm a ficha tem a permissão de transmitir. Após transmitir, a WI que detêm a ficha a passa para próxima WI. Esse mecanismo está sujeito a uma série de problemas, como, por exemplo, a ficha precisa circular por todas as WIs (mesmo que a WI não tenha nada para ser transmitido) ou ela pode ser perdida, o que ocasionaria um processo de reeleição, aumentando o atraso na circulação da mesma. Além disso, ele possui uma escalabilidade limitada, cada nova WI adicionada à WiNoC aumenta o tempo para a ficha circular por todas as WIs [CL14].

Poucas arquiteturas de WiNoC exploraram o uso do acesso múltiplo por detecção da portadora (*Carrier Sensing Multiple Access* — CSMA) como MAC. Embora o CSMA utilize a divisão do tempo para o uso do canal, o acesso ocorre de uma forma aleatória. Essa abordagem difere da utilizada nos demais mecanismos MAC apresentados até aqui. Nele, a WI precisa verificar se o canal não está em uso por alguma outra WI antes de transmitir. A frequência com que ele verifica o canal é definido pelo modo de persistência do CSMA. Para uma WiNoC não saturada, o CSMA apresenta sua maior eficiência. Por esse motivo, algumas arquiteturas exploram o uso de arquiteturas híbridas usando o CSMA e algum outro mecanismo MAC, como em [MG15] onde é usado em conjunto com a passagem de ficha.

²A antena omnidirecional irradia com a mesma intensidade em todas as direções

Mecanismos MAC eficientes, utilizados em redes de computadores, não são aplicáveis ao contexto em chip devido à sua complexidade, pois exigiriam uma abundância de área e energia, fatores críticos no ambiente em chip. Por esse motivo, o projeto de um mecanismo MAC eficiente e adequado para o ambiente em chip é um dos principais desafios para as WiNoCs [ANACA15].

1.1 Objetivos

O objetivo principal desta tese é propor o uso do CSMA como mecanismo de controle de acesso ao meio para DWiNoC tirando proveito dos canais dedicados, formados a partir das regiões delimitadas pelos lóbulos principais das antenas direcionais, para aumentar a eficiência energética do canal, reduzindo o consumo quando ele está ocioso ou submetido a cargas leves de tráfego.

O mecanismo passagem de ficha, que é o mecanismo MAC — no estado-da-arte — utilizado pela DWiNoC, é menos eficiente para essa categoria de tráfego [MG15]. Além do gasto energético para transmissão da ficha mesmo com o canal ocioso ou com baixa carga de tráfego.

O CSMA consegue lidar melhor com essa categoria de tráfego e com as particularidades do ambiente em chip, como o tempo de propagação significativamente menor em relação ao ambiente fora do chip e a maior facilidade para obter a sincronização entre as WIs, já que todas elas podem operar sob a mesma árvore de *clock*, são atrativas para o uso do CSMA como mecanismo MAC nos diversos canais dedicados da DWiNoC. Essa última particularidade, por exemplo, facilita a exploração do método de persistência *slotted*, onde cada WI, só pode transmitir no início de intervalos de tempo bem definidos e desde que as WIs estejam sincronizadas. Esse método de persistência em conjunto com o tempo de propagação curto ajudam a reduzir significativamente a probabilidade de colisões.

Dessa forma, para alcançar esse objetivo principal, foram traçados os seguintes objetivos secundários:

- Propor o uso do *slotted* CSMA não-persistente como mecanismo MAC interno dos canais dedicados da DWiNoC, adaptando-o às peculiaridades do ambiente em chip.
- Elaborar metodologia para avaliação do desempenho de forma analítica, tanto do mecanismo MAC proposto, quanto do mecanismo de passagem da ficha.
- Comparar o desempenho entre *Slotted* CSMA não-persistente e o mecanismo da passagem da ficha.

1.2 Contribuição Original

A originalidade desse trabalho consiste na proposta inovadora do uso do CSMA como mecanismo MAC para a DWiNoC, tendo foco em melhorar a eficiência energética do canal compartilhado sem impactar significativamente o desempenho. Esse trabalho combina diferentes tecnologias com a finalidade de alcançar o resultado esperado. O uso do CSMA nos canais dedicados de uma DWiNoC representa algo completamente inédito.

1.3 Organização

O presente trabalho está organizado da seguinte forma. Os conceitos básicos de uma WiNoC e dos principais mecanismos de controle de acesso ao meio são apresentados no Capítulo 2. O estado da arte em mecanismos MAC já propostos para WiNoC e DWiNoC é discutido no Capítulo 3. O Capítulo 4 traz os detalhes da arquitetura para o uso do CSMA como mecanismo MAC de uma DWiNoC. O modelo analítico utilizado para avaliar o desempenho e os resultados obtidos são apresentados no Capítulo 5. Por fim, o Capítulo 6 traz a conclusão e sugestões para trabalhos futuros.

2. CONCEITOS BÁSICOS

Neste capítulo serão apresentados os conceitos básicos relacionados à WiNoCs e a mecanismos MAC. Inicialmente, na Seção 2.1 é mostrado o princípio de funcionamento de uma NoC, bem como uma visão geral de outros paradigmas de comunicação em chip. Em seguida, na Seção 2.2, são apresentados os principais conceitos relativos a uma WiNoC, como a antena, a interface sem fio, o modelo de propagação eletromagnética em chip, a topologia e o posicionamento das WIs. Por fim, na Seção 2.3 é apresentado o princípio de funcionamento dos principais mecanismos de controle de acesso ao meio já empregados em WiNoCs.

2.1 NoC

A NoC surgiu como uma promessa para resolver os problemas de latência e escalabilidade dos barramentos. Na NoC, os blocos IP, que podem ser elementos de processamento (*Processing Element* — PE) ou elementos de armazenamento, ficam separados da comunicação. Os blocos IP podem se comunicar entre si através de roteadores conectados à cada bloco IP. Cada roteador é composto de um número específico de portas de entrada e saída conectadas diretamente a outros roteadores através de interconexões em ambos os sentidos (*full-duplex*). O arranjo de como os roteadores estão distribuídos e interconectados é definido pela topologia utilizada.

Os roteadores tomam a decisão para qual de suas portas de saída deve ser encaminhado o pacote recebido na sua entrada. A decisão é baseada na informação contida no cabeçalho do pacote e no algoritmo de roteamento. Esse pode ser determinístico ou não-determinístico. No determinístico, o caminho entre dois nodos específicos sempre será o mesmo. Por outro lado, no não-determinístico, caminhos alternativos podem ser tomados pelo pacote. O caminho pode ser escolhido aleatoriamente ou adaptativamente, considerando o congestionamento em cada roteador. O roteamento adaptativo pode ser parcialmente ou completamente adaptativo, sendo que no primeiro algumas direções não são permitidas no intuito de evitar a dependência cíclica dos recursos (*Deadlock*). No último, o *Deadlock* é evitado utilizando canais virtuais (*Virtual Channels* — VCs).

Os pacotes são quebrados em unidades de controle fluxo (*flow control units* — *flits*). O primeiro *flit* contém o cabeçalho do pacote. A comutação de pacotes define a estratégia de como os dados serão transmitidos da origem até o destino. Buraco de minhoca (*wormhole*) é uma estratégia comum em NoCs, nela o roteador toma a decisão de roteamento e encaminha o pacote assim que o cabeçalho é recebido. Os demais *flits* do pacote seguem o mesmo caminho do *flit* de cabeçalho.

A política de controle de fluxo caracteriza o deslocamento do pacote através da NoC. Tipicamente esse controle é distribuído, onde cada roteador toma sua decisão localmente. O canal de comunicação entre os roteadores pode dispor de um único canal físico ou aplicar o conceito de multiplexação, dividindo-o em diversos VCs, com buffers independentes. O buffer é utilizado para armazenar temporariamente os *flits* antes de serem encaminhados.

2.1.1 Outros Paradigmas de Interconexão em Chip

Nesta subseção é apresentada uma visão geral das principais vantagens e desvantagens das tecnologias emergentes de interconexão em chip, como a fotônica, a por radiofrequência (*RF-interconnect* — RF-I) e a 3D.

2.1.1.1 Interconexão fotônica

A interconexão fotônica utiliza guias de ondas óticas para transmitir o sinal. Para isto, precisa integrar no chip elementos óticos como ressonadores em micro anéis e fotodetectores. A propagação do sinal ocorre na velocidade da luz permitindo uma alta largura de banda e uma baixa latência [DCG⁺15] [PKK⁺09] [JBK⁺09]. Além disso, ela é pouco suscetível a interferências eletromagnéticas e possui capacidade de cobrir longas distâncias. A interconexão fotônica tem sido proposta como uma possível solução para reduzir o impacto da comunicação no consumo global de energia das NoCs. Dado que tira proveito de duas características importantes do meio fotônico. A primeira é que a energia dissipada independe da quantidade de dados que trafegam pela interligação fotônica. A segunda é que a energia dissipada em uma interligação fotônica independe da distância de transmissão [CPX09].

Uma possível organização de um sistema integrado em um único chip (*System-on-Chip* — SoC) 3D com uma NoC fotônica é ilustrada na Figura 2.1. Ele consiste em três categorias de camadas: (a) camada computacional, núcleos de processamento em conjunto com suas memórias locais e interfaces de rede; (b) uma ou mais camadas de armazenamento, que fornecem abundância de memória em chip; (c) camada de comunicação, que hospeda os componentes óticos e dispositivos opto-eletrônicos combinados para fazer a NoC fotônica e fornecer a infraestrutura principal de comunicação para interligar os núcleos entre si ou entre memórias e dispositivos fora do chip [CPX09].

Embora a tecnologia fotônica ofereça vantagens exclusivas para comunicação com alta largura de banda e energeticamente eficiente, as suas limitações com relação à capacidade de computação e armazenamento colocam alguns desafios críticos para o projeto de NoCs fotônicas. Em particular, o *buffering* do *flit* e o processamento dos *control-flits*,

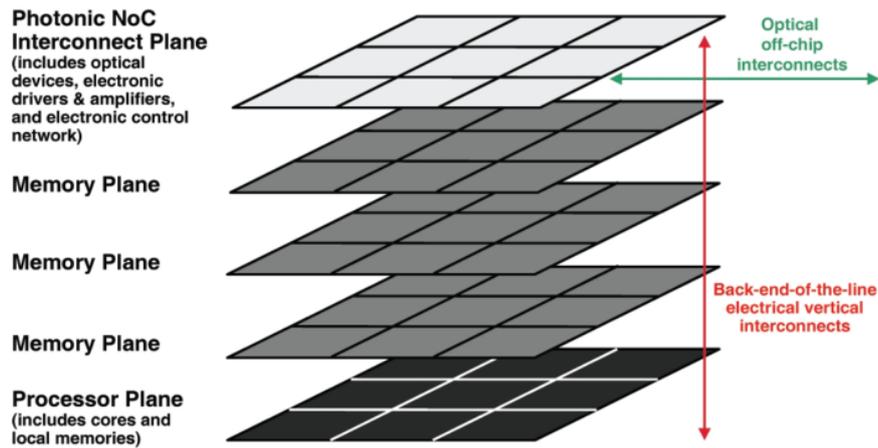


Figura 2.1: Possível organização de uma arquitetura SoC combinando uma NoC 3D e uma NoC fotônica. Fonte: [CPX09].

duas funções importantes de uma NoC com comutação de pacotes que são inviáveis de serem implementadas com dispositivos ópticos, dado que não há equivalentes fotônicos para elementos de armazenamento como, por exemplo, *flip-flops* e registradores [CPX09]. Além disso, na interconexão fotônica, o sinal é atenuado significativamente cada vez passa por bifurcações, isso reduz a capacidade de *fanout*, visto que limita o número de entradas que podem ser conectadas a uma saída [DCG⁺15] [MJK14]. Para aumentar o número de nodos que podem receber um sinal é necessário aumentar a potência do sinal e consequentemente o consumo de energia.

Existe ainda a complexidade no processo de fabricação, os componentes integrados em CMOS ocupam uma área significativa e para transformar o sinal de elétrico para óptico e para rotear o sinal óptico, algumas vezes são necessários componentes não-CMOS [KKD⁺06] [Mil09], inclusive alguns deles, como fontes de laser, precisam ficar fora do chip [HSL⁺03]. As perdas neste acoplamento elétrico-óptico representam à maior parte do consumo de energia [KMTY16]. O roteamento óptico possui serias restrições a curvas acentuadas, onde uma curva muito acentuada causaria uma significativa degradação no sinal e por esse motivo não é permitida. Avanços recentes [MOM⁺14] [Mic11] conseguiram integrar quase todos os componentes ópticos em chip, porém são técnicas imaturas que precisam enfrentar seus próprios desafios [KMTY16]. Além disso, as interligações ópticas são sensíveis a variações térmicas e no processo de fabricação, que podem ocasionar diafonia e perda do sinal [MLC⁺14], devido a incompatibilidades entre transmissor e receptor.

2.1.1.2 Interconexão por RF-I

Na interconexão por linhas de transmissão, ou também conhecida como RF-I, os fios funcionam como guias de onda por onde são propagados os dados na forma de ondas eletromagnéticas. Como mostra a Figura 2.2, existem três tipos principais de linhas

de transmissão em chip: a linha *microstrip* (*Microstrip Line* — MSL), a guia de onda coplanar (*Coplanar Waveguide* — CPW), e a linha diferencial ou faixas coplanares (*Coplanar Strip* — CPS). A MSL é conhecida por sua simplicidade em comparação com a CPS e CPW, enquanto as duas últimas mostram melhor robustez contra a diafonia. Além disso, o CPS é conhecido por sua maior densidade de interconexão em comparação com o CPW [CHX⁺12]. O RF-I precisa de um transceptor para transformar o sinal elétrico em um sinal de radiofrequência e vice-versa. Por usar linhas de transmissão ao invés de antenas para propagação do sinal, o RF-I dissipa e consome menos energia e o espectro de frequência não fica limitado a frequência de ressonância da antena [CCK⁺08].

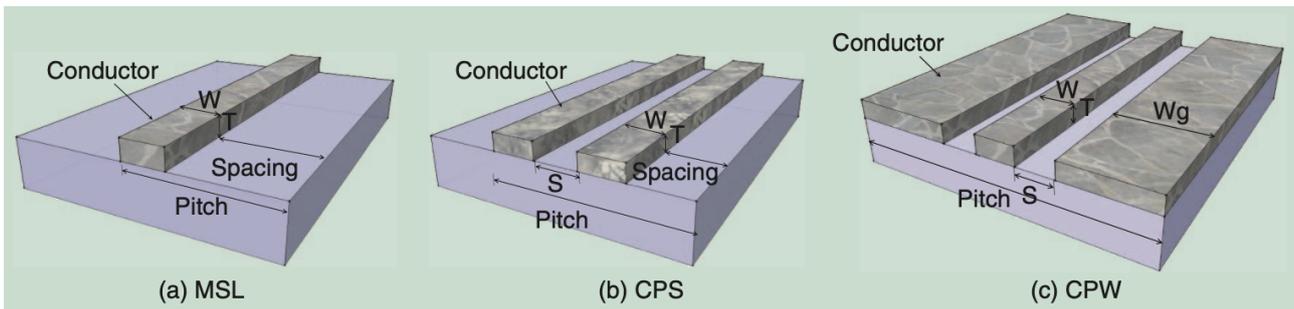


Figura 2.2: Categorias de linhas de transmissão. Fonte: [KMTY16]

Entretanto, o RF-I está sujeito à diafonia entre as linhas de transmissão e entre essas e os circuitos próximos, esse problema é mais evidente em altas frequências ou em linhas de transmissão longas. Reduzir a impedância da linha de transmissão para aumentar sua robustez à diafonia implica em aumentar a dissipação de energia. Além disso, os fios utilizados como linhas de transmissão têm uma maior espessura, o que resulta numa menor resistência, mas, por outro lado, numa grande capacitância. Para evitar o efeito parasita provocado por essa capacitância é necessário utilizar uma camada mais extensa de material dielétrico entre os metais. Isso aumenta a área utilizada bem como a densidade de área no chip utilizada para interconexões. Por fim, ele ainda está sujeito a limitação na frequência de corte intrínseca à tecnologia CMOS e a limitação de pontos de queda [KMTY16].

2.1.1.3 Interconexão 3D

Diversas abordagens já foram propostas para construção de uma NoC 3D, como a simétrica, a híbrida ou ainda a com *crossbar* em 3D ou com os componentes do roteador espalhados por diversas camadas. Na NoC 3D simétrica, onde tanto a comunicação realizada na mesma camada quanto a comunicação entre camadas são realizadas da mesma forma: por saltos passando por todos os roteadores intermediários do caminho [NND10]. Por outro lado, na NoC 3D híbrida, onde ela mistura o conceito de NoC com o de barramento, são criados barramentos verticais compartilhados, controlados por um árbitro central, onde a

comunicação entre camadas pode ser realizada em um único salto através de canais verticais usando TSV¹. Isso reduz o número de saltos na comunicação e, conseqüentemente, a latência e o consumo de energia [PF06].

Pode ser utilizado também um *crossbar* em 3D, onde o *crossbar* de cada roteador de uma camada inferior é interligado verticalmente ao *crossbar* do roteador da camada superior. Essas interligações entre as *crossbar* de cada camada são fisicamente fundidas em uma única *crossbar* 3D. Por fim, na NoC 3D multicamada os componentes do roteador podem ser divididos em subcomponentes menores distribuídos por múltiplas camadas [CPX09].

Essa NoC 3D é chamada simétrica porque, na comunicação, tanto o movimento dentro da mesma camada quanto o movimento entre camadas é realizado por saltos passando por todos os roteadores intermediários do caminho. Por exemplo, o movimento a partir da camada inferior de um chip com 4 camadas para a camada superior requer 3 saltos, como pode ser observado na Figura 2.3 [NND10].

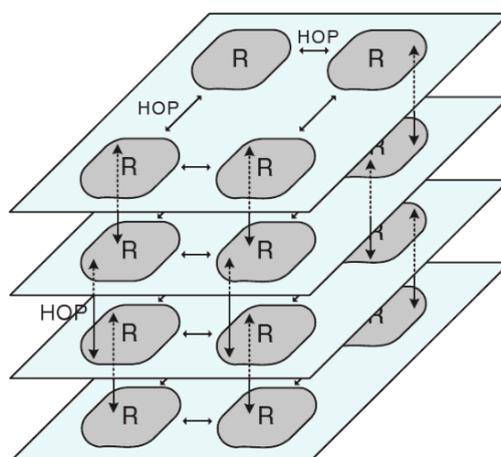


Figura 2.3: NoC simétrica. Fonte: [NND10].

Uma NoC 2D pode ser estendida para uma NoC 3D simétrica realizando uma simples adaptação nos roteadores. Essa adaptação consiste na adição de duas portas físicas em cada roteador, uma para cima e outra para baixo, com *buffers*, árbitros e *crossbar* associados [CPX09].

Como a comunicação entre as camadas é salto a salto, essa entre dois nodos em camadas diferentes demora tanto tempo quanto a comunicação entre dois nodos na mesma camada. Além disso, a adição de duas portas extras exige um *crossbar* maior, que implica em uma significativa sobrecarga de energia e área [CPX09].

A NoC 3D híbrida mistura os conceitos de NoC com outra popular categoria de interligação com acesso compartilhado: o barramento. Nela é criado um barramento vertical através das camadas do chip 3D, Figura 2.4, e os roteadores convencionais precisam ser

¹TSV (*Through-silicon Via*) são interconexões metálicas que atravessam do substrato de silício.

transformados em roteadores híbridos, com uma interconexão entre o roteador e o barramento vertical. O uso do barramento vertical permite que a comunicação entre quaisquer camadas seja realizada em um único salto, diferente da NoC 3D simétrica. Como no barramento vertical o acesso é compartilhado, o acesso ao barramento é controlado por um árbitro central [CPX09].

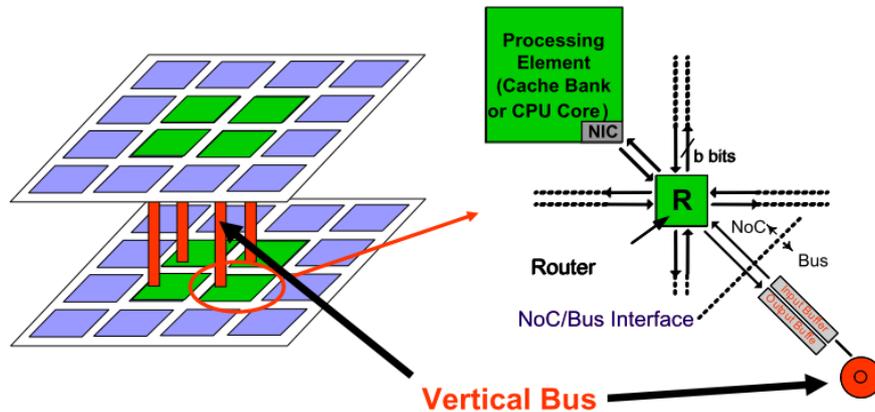


Figura 2.4: Arquitetura da NoC 3D híbrida. Fonte: [CPX09].

Apesar dos benefícios sobre a NoC 3D simétrica, a NoC 3D híbrida também sofre de uma grande desvantagem: ela não permite comunicação simultânea no barramento vertical. Visto que o barramento é um meio compartilhado, ele só pode ser usado por um único *flit* em qualquer dado momento. Isso aumenta drasticamente a contenção e a probabilidade de bloqueio sob uma alta carga na rede [CPX09].

Na NoC 3D com *crossbar* em 3D, o *crossbar* de cada roteador de uma camada inferior é interligado verticalmente ao *crossbar* de cada roteador da camada superior. Essas interligações entre as *crossbar* de cada camada são fisicamente fundidas em uma única *crossbar* 3D. Isso implica na utilização de um *crossbar* 5×5 , dado que não há canais físicos adicionais que possam ser dedicados para a comunicação entre as camadas [CPX09]. A Figura 2.5 ilustra a disposição da *crossbar* 3D.

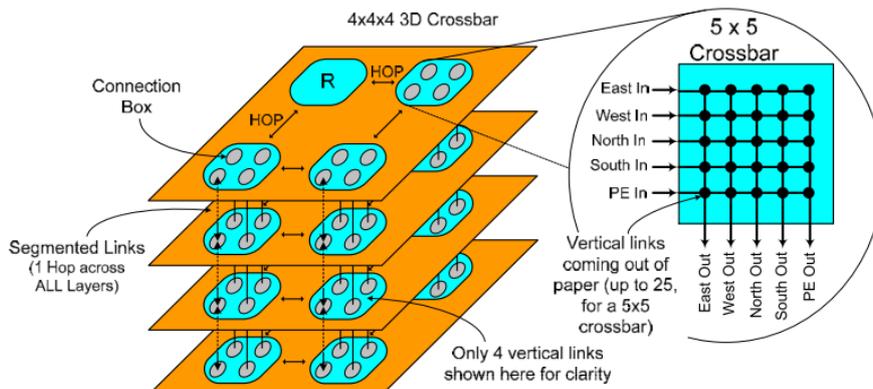


Figura 2.5: Arquitetura da NoC 3D com *crossbar* em 3D. Fonte: [CPX09].

O grande número de interligações verticais em uma *crossbar* 3D, resulta no aumento da diversidade de caminho. Embora o aumento da diversidade pode inicialmente parecer um atributo positivo, ele leva realmente a um aumento drástico na complexidade do árbitro central, que coordena a comunicação entre as camadas no *crossbar* 3D. O árbitro agora precisa decidir entre uma infinidade de possíveis interligações, e requer um número excessivo de sinais de controle para gerenciar todas essas interligações [CPX09].

A NoC 3D multicamada, se baseia na possibilidade de dividir os componentes de cada nodo em subcomponentes menores e distribuí-los por múltiplas camadas. Para isso é necessário estender o roteador convencional para um roteador 3D que abranja múltiplas camadas. Os componentes do roteador podem ser classificados em duas categorias baseadas na capacidade de dividir e distribuir os componentes através de múltiplas camadas: separáveis (*buffers* de entrada e *crossbar*) e não separáveis (lógica de arbitragem e lógica de roteamento) [CPX09].

Essa distribuição de componentes de um nodo por múltiplas camadas acaba gerando uma economia de área. Além disso, já que a maior parte do tráfego de comunicação consiste em *flits* curtos e com padrões frequentes, é possível desligar dinamicamente algumas camadas do roteador multicamada para reduzir o consumo de energia. Entretanto, essa abordagem de projeto ainda é agressiva para a tecnologia atual, podendo se tornar viável com o amadurecimento da tecnologia 3D [CPX09].

Geralmente, as NoCs 3D precisam enfrentar os problemas de dissipação de energia por área de superfície e com a elevação da temperatura de pico como consequência disso [PL06]. Além disso, o processo de fabricação é complexo devido aos problemas de alinhamento entre as camadas [BAC⁺07].

2.2 WiNoCs

Dentre esses paradigmas de interconexões emergentes, que ainda enfrentam dificuldades de fabricação e complexidades de projeto, a interconexão sem fio é uma alternativa promissora, pois pode ser implementada utilizando a tecnologia CMOS atual.

Diversas tecnologias de antenas e transceptores em chip já foram propostas, como a UWB [LTP⁺09] [ZW08], CNTs [Han05] [KRH⁺07], grafeno [SMSG17] e mmWave [Raz09]. As WIs com transceptores e antenas implementados em chip, permitem a comunicação de longa distância, com alta largura de banda e baixa latência.

Essas tecnologias têm como característica diferentes frequências de operação. Os intervalos da frequência de operação de cada uma dessas tecnologias são apresentados na Figura 2.6. O UWB trabalha no intervalo de 1 até uma dezena de GHz. A tecnologia mmWave opera no intervalo de uma dezena de GHz até pouco mais que uma centena de

GHz. O Sub-THz opera a partir disso até uma frequência próxima a um THz. Por fim, a THz opera a partir dessa frequência [DGP⁺12].

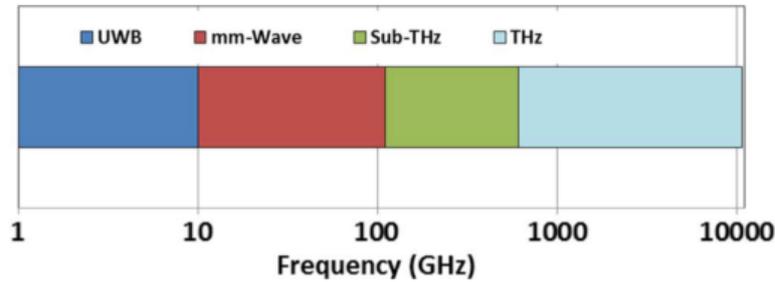


Figura 2.6: Intervalos da frequência de operação das diferentes tecnologias de redes em chip sem fio. Fonte: [DGP⁺12].

Entretanto, algumas dessas tecnologias apresentam algumas desvantagens significativas, como o UWB, que não consegue atingir longas distâncias em chip, e as antenas CNT ou de grafeno, que ainda estão passando pelo desafio da integração no processo CMOS. Por outro lado, interligações sem fio compatíveis com CMOS e operando na escala mmWave vem se mostrando bem promissoras [MGS⁺17].

Por esses motivos, o mecanismo MAC proposto nesta tese tem como referência o mmWave como tecnologia de interconexão sem fio. Nas próximas subseções serão detalhados os conceitos básicos relativos a arquiteturas de WiNoCs que utilizam essa tecnologia de interconexão sem fio.

2.2.1 Antenas

Uma antena ideal para o ambiente em chip é aquela que consegue oferecer o maior ganho ocupando a menor área possível [DCY⁺13]. A maioria das arquiteturas de WiNoCs existentes, exploram o uso de antenas omnidirecionais, como antenas do tipo zigue-zague [FHO02]. Outros trabalhos exploram o uso de antenas direcionais, como antenas planares log periódicas [SRD14] (*Planar Log-periodic Antenna* — PLPA), que conseguem concentrar a maior parte da energia irradiada em uma determinada direção. Essas duas antenas são detalhadas nas subseções 2.2.1.1 e 2.2.1.2, respectivamente.

2.2.1.1 Antena zigue-zague

A antena zigue-zague foi utilizada pela primeira vez em [FHO02] para distribuição do sinal de *clock* usando a tecnologia CMOS de $0,18\ \mu\text{m}$, onde os circuitos foram testados a uma frequência de 15 GHz. O leiaute da antena é mostrado na Figura 2.7. Ela é constituída de um dipolo de 2 mm em formato de zigue-zague. Cada braço da antena possui $10\ \mu\text{m}$ de largura e $80\ \mu\text{m}$ de comprimento, com um ângulo de curvatura de 30° .

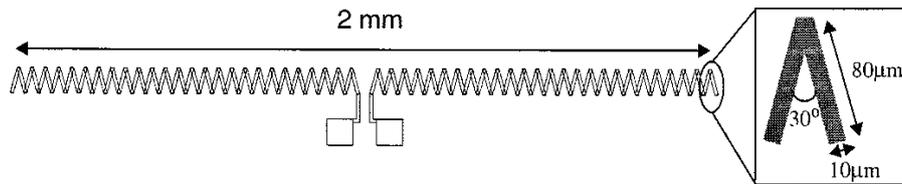


Figura 2.7: Leiaute da antena em formato de zigue-zague. Fonte: [FHO02].

Para ser utilizada em uma WiNoC essa antena foi ajustada, alterando o comprimento do braço para $60 \mu\text{m}$ e reduzindo o diâmetro da antena para ficar compatível com a frequência de operação de 60 GHz [DGC⁺10].

Essa antena apresenta características omnidirecionais, dessa forma, o efeito de rotação (ângulo relativo entre as antenas de transmissão e recepção) tem é desprezível sobre a intensidade do sinal recebido [ZCS07]. O que a torna ideal para transmissões em *broadcast* ou *multicast*, onde o objetivo é atingir todas ou um grupo específico de WIs na WiNoC, respectivamente.

2.2.1.2 Antena planar log periódica

Essa antena consegue operar na faixa de frequência das ondas milimétricas (entre 30 e 100 GHz) e tem como características uma ampla largura de banda e uma alta diretividade². Essa antena adota parâmetros específicos para a frequência da portadora de 60 GHz [SRD14]. O leiaute da antena é mostrado na Figura 2.8 e suas dimensões são apresentadas na Tabela 2.1. O diâmetro da antena coincide com o comprimento de onda no meio dielétrico (λ_g), dado por,

$$\lambda_g = \lambda / \sqrt{\epsilon_r} \quad (2.1)$$

onde, λ é comprimento de onda no espaço livre e ϵ_r é a constante dielétrica do substrato de silício (10 – 20 Ω -cm).

O leiaute da antena é aplicado sob uma camada de metal de 0,055 mm de espessura e está situada sob uma camada de dióxido de silício (SiO_2) de 0,0123 mm. Essa camada serve para fornecer algum isolamento entre a antena e o plano de aterramento. O substrato de silício (Si) possui 0,385 mm de espessura e está situado entre a camada de SiO_2 e o plano de aterramento. A Figura 2.9 apresenta o corte transversal da antena sob o substrato de silício. A viabilidade do uso de antenas neste substrato já foi estudada em [KHO00]. Uma ligação sem fio pode ser estabelecida entre duas antenas que estejam no mesmo substrato.

O gráfico do parâmetro S_{11} em relação à frequência é mostrado na Figura 2.10. Esse parâmetro representa a perda de retorno³ da antena em relação à frequência. A partir

²Diretividade é a capacidade da antena em concentrar a energia irradiada numa determinada direção.

³Perda de retorno representa o quanto da potência do sinal que chega ao terminal de entrada da antena é refletido de volta.

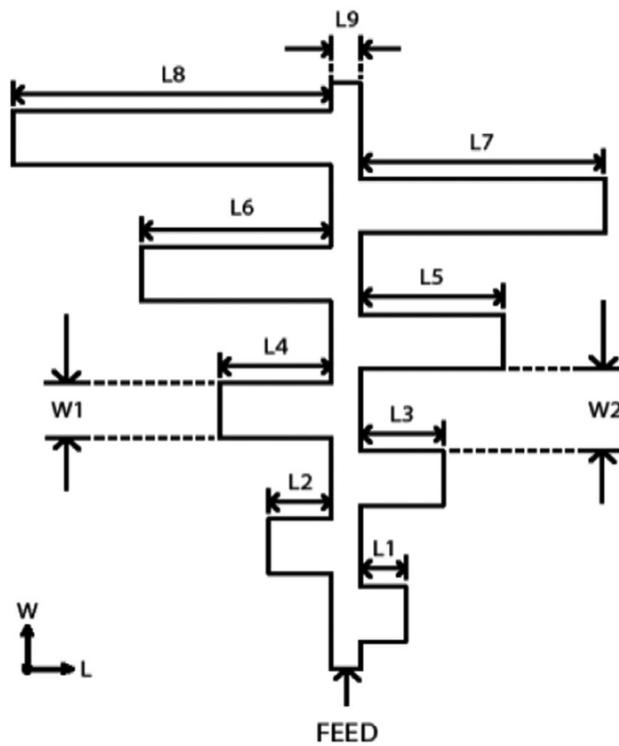


Figura 2.8: Antena log periódica em chip. Fonte: [SRD14].

Tabela 2.1: Dimensões da antena. Fonte: [SRD14].

Dimensão	Comprimento (mm)
L1	0,11000
L2	0,14019
L3	0,17867
L4	0,22772
L5	0,29023
L6	0,36990
L7	0,47144
L8	0,60085
L9	0,05500
W1	0,11000
W2	0,16500
ϕ	1,1825

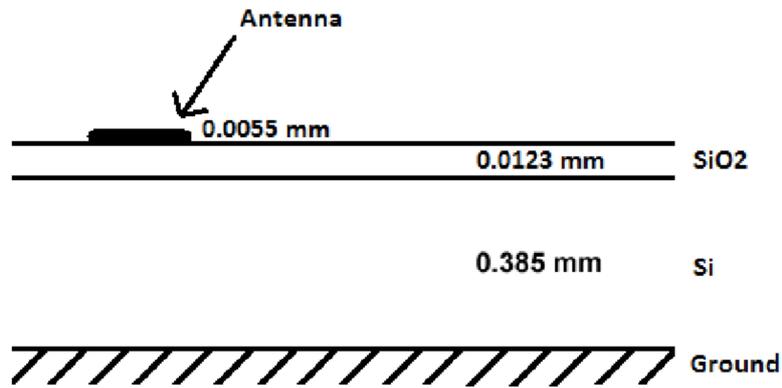


Figura 2.9: Corte transversal da antenna PLPA sob o substrato de silício. Fonte: [SRD14].

do gráfico é possível observar que nas frequências de 44 e 60 GHz apresentam as menores perdas de retorno. Isso indica que a antenna PLPA é uma antenna de banda dupla e consegue operar nestas duas frequências.

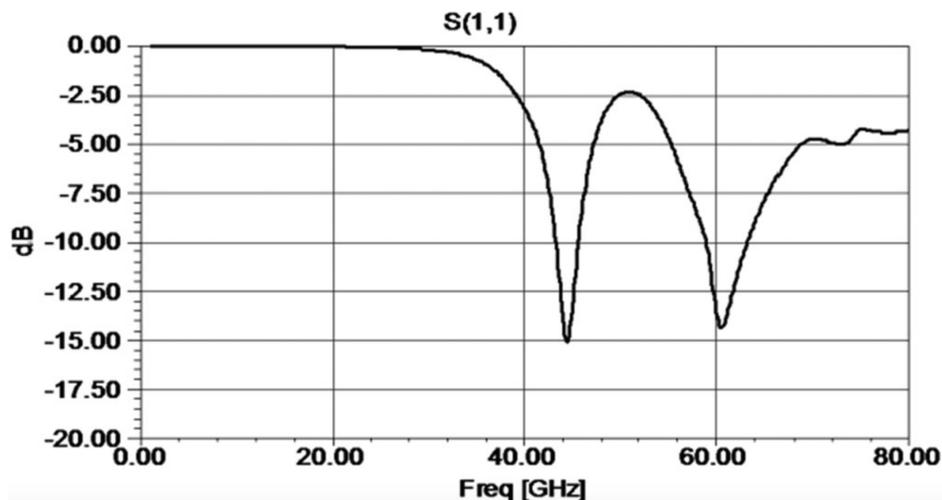


Figura 2.10: Perda de retorno da antenna PLPA. Fonte: [MGS⁺17]

O padrão de radiação nos planos azimutal e de elevação, mostrados na Figura 2.11, não deixam dúvidas quanto a diretividade da antenna. A largura de feixe a meia potência é de 33° no plano azimutal e 30° no plano de elevação.

O ganho de transmissão, G_a , entre duas antenas pode ser calculado a partir dos parâmetros S , conforme a Equação 2.2,

$$G_a = \frac{|S_{21}|^2}{(1 - |S_{11}|^2)(1 - |S_{22}|^2)} \quad (2.2)$$

onde, $S_{11} = S_{22}$ devido à reciprocidade, dado que as antenas de transmissão e recepção são iguais. O parâmetro S_{21} representa a potência transferida da porta 1 para porta 2, considerando uma rede de 2 portas. A porta 1 é a entrada da antenna conectada ao transmissor e a porta 2 é a saída da antenna conectada ao receptor.

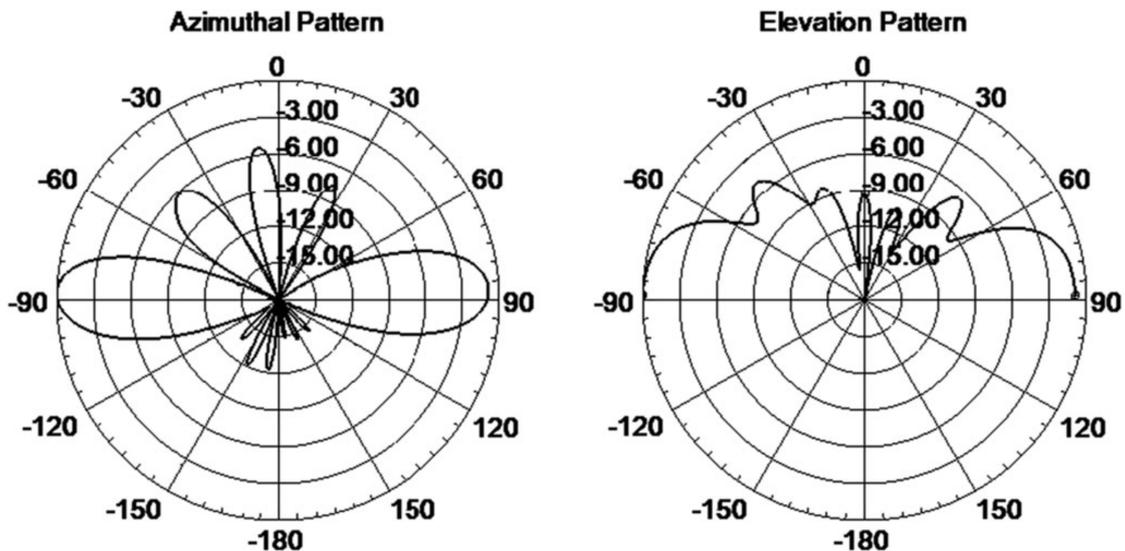


Figura 2.11: Padrão de radiação da antena PLPA ao longo dos planos azimutal e elevação. Fonte: [MGS⁺17]

A variação do parâmetro S_{21} de uma antena PLPA em relação à distância foi analisado em [MGS⁺17], onde a distância entre o transmissor e o receptor variou de 2 até 28 mm, como é mostrado na Figura 2.12. Como pode ser observado, o parâmetro S_{21} varia exponencialmente em relação à distância. Como exemplo, para duas antenas PLPA, situadas no mesmo substrato, separadas por uma distância de 20 mm e alinhadas na direção com maior diretividade uma da outra foi obtido um ganho de $-38,65$ dB. A 60 GHz, esse ganho, se traduz em potência de transmissão requerida na faixa de 0 dBm, considerando a modulação *On-off Keying* (OOK) e exigência de um BER de 10^{-15} [DKKM11]. Essa potência de transmissão pode ser facilmente gerada no ambiente em chip.

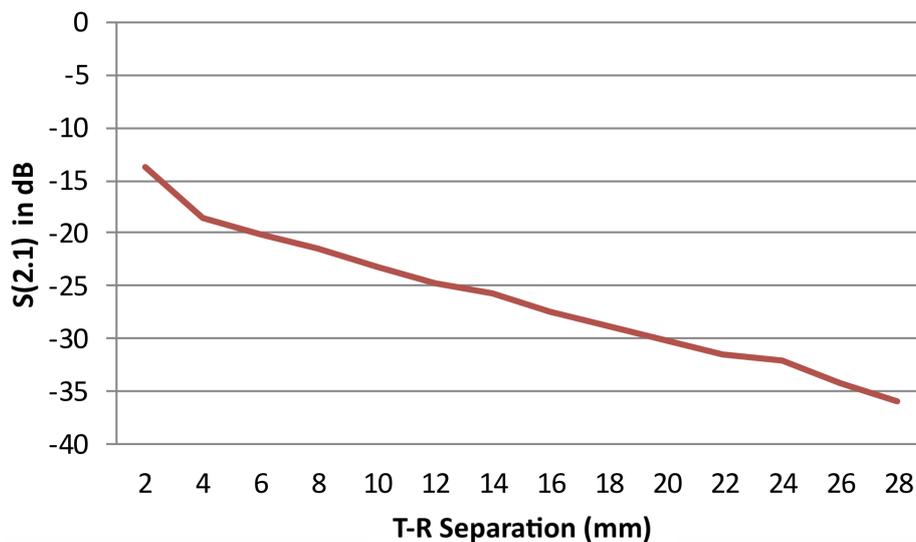


Figura 2.12: Variação do parâmetro $S(2,1)$ da antena PLPA com a distância. Fonte: [MGS⁺17]

A perda no caminho (*path loss*) em dB pode ser calculada a partir da Equação 2.3. Nas simulações em [MGS⁺17] foi obtido o valor de 1,42 para o fator de perda no caminho (Υ), sendo ligeiramente menor ao fator de perda no caminho do vácuo, $\Upsilon = 2$.

$$PL = \Upsilon 10 \log_{10}(d) + NF_{TOTAL} \quad (2.3)$$

onde, NF_{TOTAL} é o ruído de fundo total.

Em [MGS⁺17] foi simulado também a variação na potência do sinal recebido com relação à orientação da antena. Para isso uma das antenas foi girada em seu próprio eixo por 360° e foi plotado na Figura 2.13 a variação do parâmetro S_{21} em conjunto com o padrão de radiação azimutal. A partir dele é possível observar a correlação entre a maior potência recebida com os lóbulos principais do padrão de radiação.

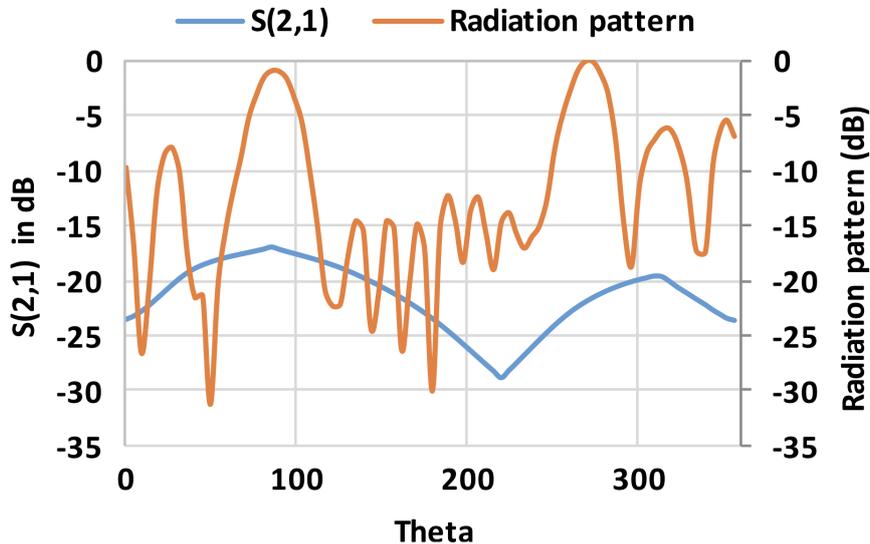


Figura 2.13: Variação do parâmetro $S(2, 1)$ com a orientação da antena. Fonte: [MGS⁺17]

2.2.2 Transceptores

Nessa subseção, serão apresentados alguns detalhes do funcionamento do transceptor⁴. Entretanto, antes é necessário revisar alguns conceitos de modulação e de receptores de conversão direta que serão abordados nas subseções 2.2.2.1 e 2.2.2.2.

2.2.2.1 Conceito de modulação

A modulação em amplitude por ondas quadradas ou pulsos binários retangulares é referida como chaveamento em amplitude (*Amplitude Shift Keying* — ASK). Outra cate-

⁴O transceptor combina os circuitos de transmissão e recepção, compartilhando elementos em comum.

goria de modulação em amplitude pode ser efetuado pelo simples ligar e desligar da onda portadora. Essa categoria de modulação é chamada chaveamento em amplitude binário (*Binary Amplitude Shift Keying* — BASK), mas também é referenciado como chaveamento liga-desliga (*On-off keying* — OOK). Transmissões que utilizam essa categoria de modulação são chamadas transmissões por onda contínua (*Continuous Wave* — CW), e apesar de o fato de apenas a portadora ser transmitida, bandas laterais são geradas por tal sinal liga-desliga. As bandas laterais resultam da frequência ou taxa de repetição dos pulsos mais seus harmônicos [Fre15].

A Figura 2.14 apresenta uma visão conceitual do BASK e do cálculo da sua largura de banda (B), proporcional à taxa de sinal (S), ou *baud rate*⁵. No entanto, há normalmente outro fator envolvido, denominado d , que depende do processo de filtragem. O valor de d pode variar entre 0 e 1 [For12].

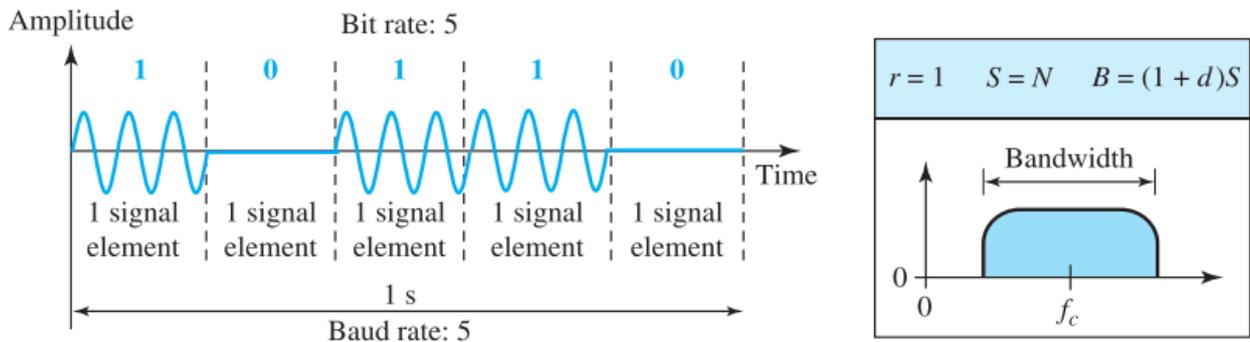


Figura 2.14: Taxa de bits, taxa de baud e largura de banda do BASK. Fonte: [For12].

Para descrição formal do sinal do BASK, como é mostrado em [HM07], é preciso considerar um fluxo de dados binários $b(t)$ definido como

$$b(t) = \begin{cases} \sqrt{E_b} & 1 \\ 0 & 0 \end{cases} \quad (2.4)$$

onde, E_b é a energia por bit.

Multiplicando-se $b(t)$ pela onda portadora sinusoidal com a fase ϕ_c igual a zero, têm-se

$$s(t) = \begin{cases} \sqrt{\frac{E_b}{T_b}} \cos(2\pi f_c t) & 1 \\ 0 & 0 \end{cases} \quad (2.5)$$

onde T_b é a duração do bit, f_c é a frequência central da portadora.

Quando a duração do bit é ocupada pelo símbolo 1, a energia do sinal transmitido é E_b . Quando a duração do bit é ocupada pelo símbolo 0, a energia do sinal transmitido é

⁵Baud rate é o número de elementos (N) do sinal por segundo.

zero. Baseando-se nisso, E_{av} é a energia média do sinal transmitido sendo dada por

$$E_{av} = \frac{E_b}{2} \quad (2.6)$$

A Equação 2.6, no entanto, só é válida se os dois símbolos binários forem equiprováveis, onde o número de 0 e 1 em um fluxo longo de dados binários são iguais.

O espectro do sinal BASK apresenta um componente de linha na frequência central. A largura de banda pode ser medida a partir da largura do lóbulo principal do seu espectro e é igual à $2/T_b$, onde T_b é a duração do bit. Quando a portadora se mantém constante e a duração do bit é reduzida pela metade, a largura do lóbulo principal, que define o envelope do espectro BASK, é dobrada, o que significa que a largura de banda de transmissão do sinal BASK é duplicada. A Figura 2.15 apresenta o espectro de potência de um sinal BASK [HM07].

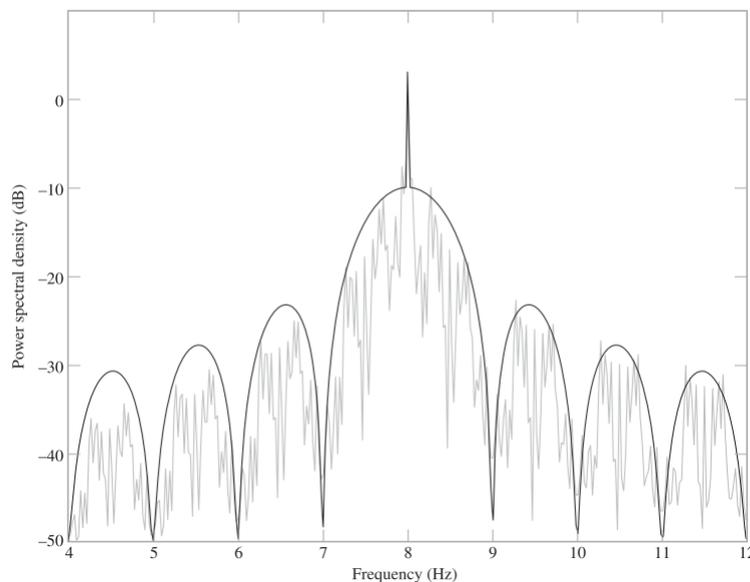


Figura 2.15: Espectro de potência de um sinal BASK. Fonte: [HM07].

O conceito de funcionamento do BASK é mostrado na Figura 2.16. Se os dados digitais forem representados como um sinal digital que não retorna a zero (*Non-return-to-zero* — NRZ) unipolar com o nível alto representado por 1 V e o nível baixo representado por 0 V, a implementação pode ser feita pela multiplicação do sinal digital NRZ pelo da portadora proveniente de um oscilador. Quando o sinal NRZ é 1, a amplitude da frequência da portadora é mantida; quando ele é 0, a amplitude da frequência portadora é zero [For12].

Na modulação por chaveamento em frequência (*Frequency Shift Keying* — FSK), quando o sinal é um fluxo de dados binários, a frequência varia entre a da portadora ou uma superior, para o caso do sinal 0 ou 1, respectivamente. Isso é ilustrado na Figura 2.17 [Fre15].

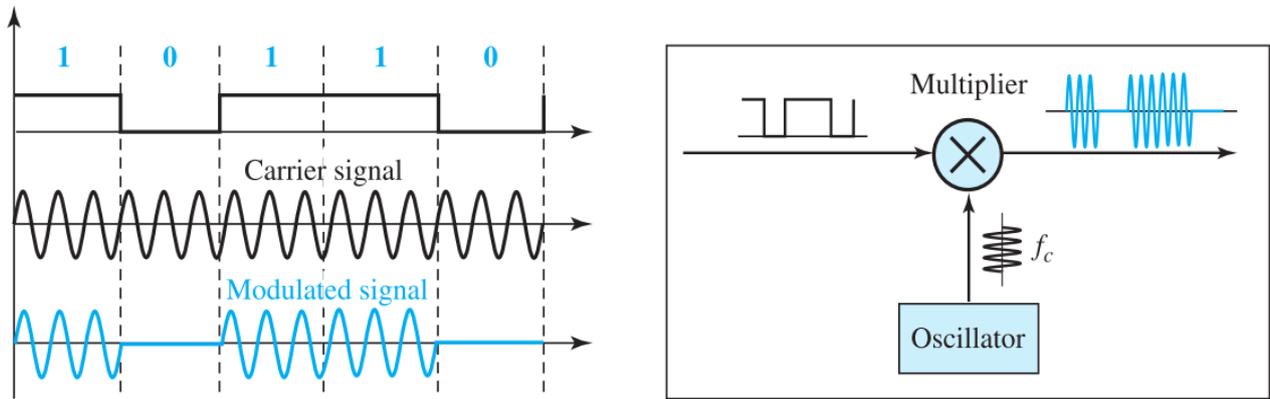


Figura 2.16: Conceito de funcionamento do BASK. Fonte: [For12].

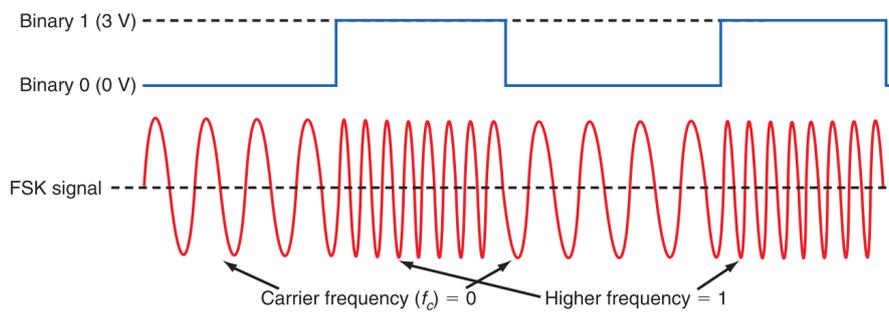


Figura 2.17: Modulação FSK. Fonte: [Fre15].

Essa categoria de modulação também é conhecido como chaveamento em frequência binário, ou *Binary Frequency Shift Keying* (BFSK), os símbolos 0 e 1 são distinguidos um do outro transmitindo uma das duas ondas sinusoidais diferindo em frequência por uma quantidade fixa. Um par típico de ondas sinusoidais é descrito em [HM07] como

$$s_i(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_1 t) & 1 \quad i = 1 \\ \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_2 t) & 0 \quad i = 2 \end{cases} \quad (2.7)$$

onde, E_b é a energia por bit do sinal transmitido.

A Figura 2.18 apresenta as taxas de bit e *baud* e o cálculo da largura de banda da modulação BFSK. Como é mostrado na Figura 2.18, o centro de uma largura de banda é f_1 e o centro da outra é f_2 . Tanto f_1 como f_2 estão separadas do ponto médio entre as duas bandas por uma largura Δf . A diferença entre as duas frequências é $2\Delta f$, onde ele deve ser no mínimo igual a S para garantir que a modulação e a demodulação ocorram corretamente [For12].

A Figura 2.19 apresenta o espectro de potência de um sinal BFSK com uma frequência da portadora $f_c = 8\text{ Hz}$ e duração do bit de 1 s. A partir da Figura 2.19 é possível observar que espectro possui dois componentes de linha nas frequências de $f =$

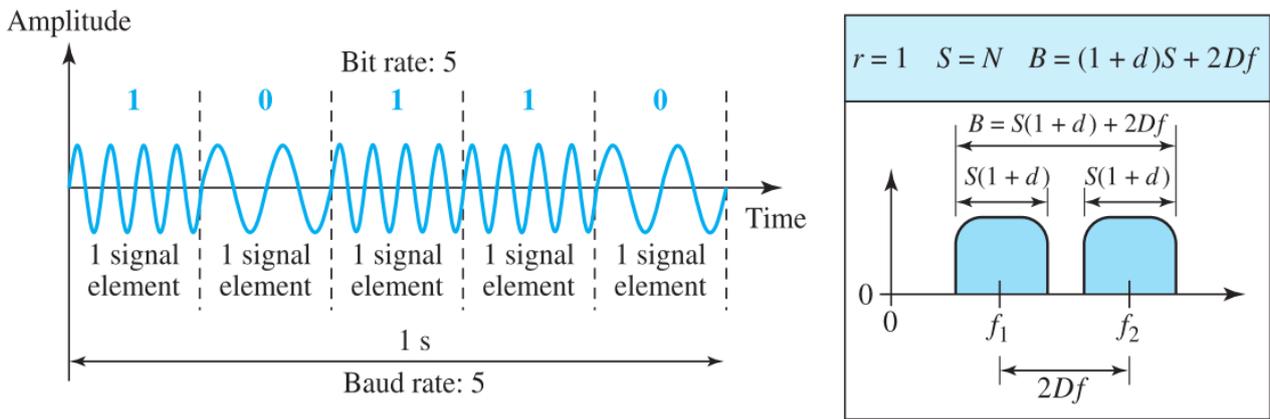


Figura 2.18: Representação das taxas de bit, baud e largura de banda do BFSK. Fonte: [For12].

$f_c \pm 1/(2T_b)$, 7,5 Hz e 8,5 Hz. O lóbulo central ocupa uma banda de largura igual à $3/T_b$ (3 Hz) centrada na frequência da portadora (8 Hz) [HM07].

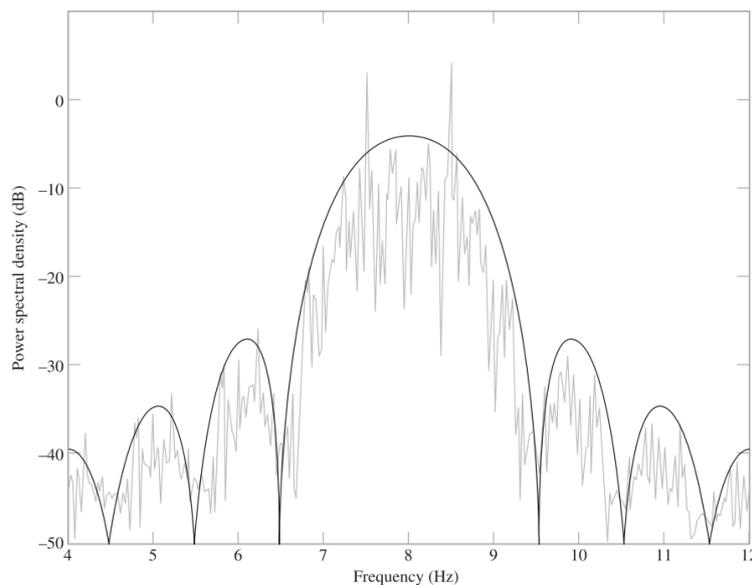


Figura 2.19: Espectro de potência de um sinal BFSK. Fonte: [HM07].

Existem duas possíveis implementações do BFSK: não coerente e coerente. No BFSK não-coerente, pode haver descontinuidade na fase quando um elemento de sinal termina e o próximo começa. No BFSK coerente, a fase se mantém através do limite de dois elementos de sinais. O BFSK não-coerente pode ser implementado tratando o BFSK como duas modulações ASK e usando duas frequências portadoras. O BFSK coerente pode ser implementado usando um oscilador controlado por tensão (*Voltage-controlled Oscillator* — VCO), que muda sua frequência conforme a tensão de entrada. A Figura 2.20 mostra uma versão simplificada do BFSK coerente. A entrada para o oscilador é o sinal NRZ unipolar. Quando a amplitude de NRZ é zero, o oscilador mantém sua frequência regular; quando a amplitude é positiva, a frequência é aumentada [For12].

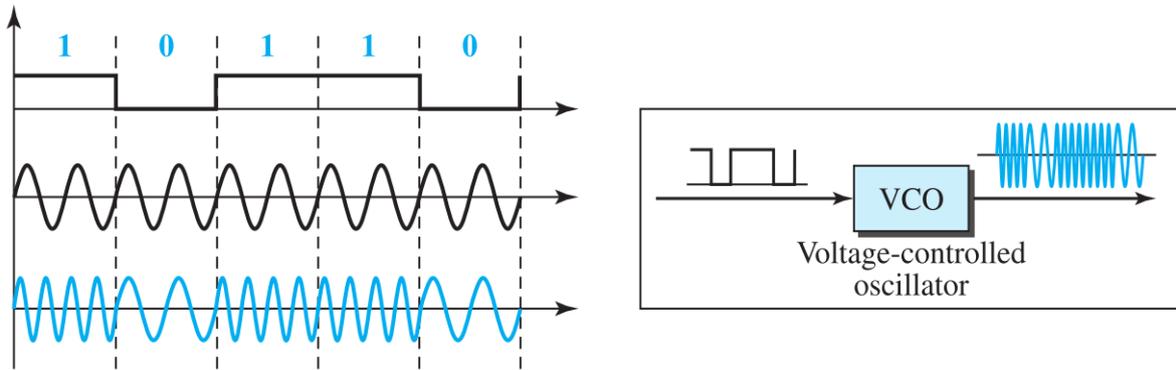


Figura 2.20: Representação simplificada da implementação de um modulador BFSK coerente. Fonte: [For12].

O processo de modulação em fase de uma portadora com dados binários é chamado chaveamento em fase (*Phase-shift Keying* — PSK). O sinal PSK mostrado na Figura 2.21 utiliza um desvio de fase de 180° a partir de uma referência, mas outros valores de desvio de fase podem ser utilizados como, por exemplo, 45° , 90° , 135° ou 225° . Na PSK não ocorre nenhuma variação de frequência. O sinal PSK tem uma frequência constante, mas a fase do mesmo muda em relação ao sinal de referência à medida que varia o sinal binário. Cada vez que o sinal muda de 0 para 1 ou 1 para 0, ocorre uma mudança de fase de 180° [Fre15].

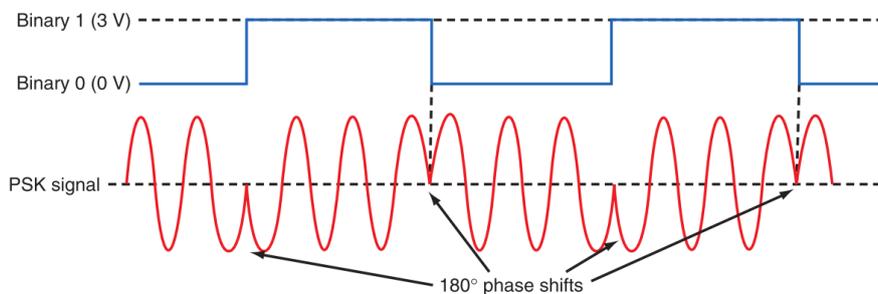


Figura 2.21: Representação da modulação PSK. Fonte: [Fre15].

O chaveamento em fase é conhecido como chaveamento em fase binário (*Binary Phase-shift Keying* — BPSK), o par de sinais $s_1(t)$ e $s_2(t)$ usados para representar os símbolos 1 e 0, respectivamente, são definidos, conforme [HM07], por

$$s_i(t) = \begin{cases} \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t), & 1 \quad (i = 1) \\ \sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t + \pi) = -\sqrt{\frac{2E_b}{T_b}} \cos(2\pi f_c t), & 0 \quad (i = 2) \end{cases} \quad (2.8)$$

onde, $0 \leq t \leq T_b$, com T_b denotando a duração do bit e E_b denotando a energia por bit do sinal transmitido.

O BPSK é como um simples BASK com uma grande vantagem, ele é menos suscetível ao ruído. No ASK, o critério para detecção do bit é a amplitude do sinal, no PSK é a fase. O ruído pode mudar mais facilmente a amplitude do que mudar a fase. O PSK também é superior ao FSK por não precisar dois sinais da portadora. Entretanto, o PSK exige um circuito mais complexo para conseguir distinguir as fases no receptor. A Figura 2.22 apresenta uma visão conceitual e o cálculo da largura de banda utilizada pelo BPSK. A largura de banda é a mesma que o BASK, mas é menor que o BFSK [For12].

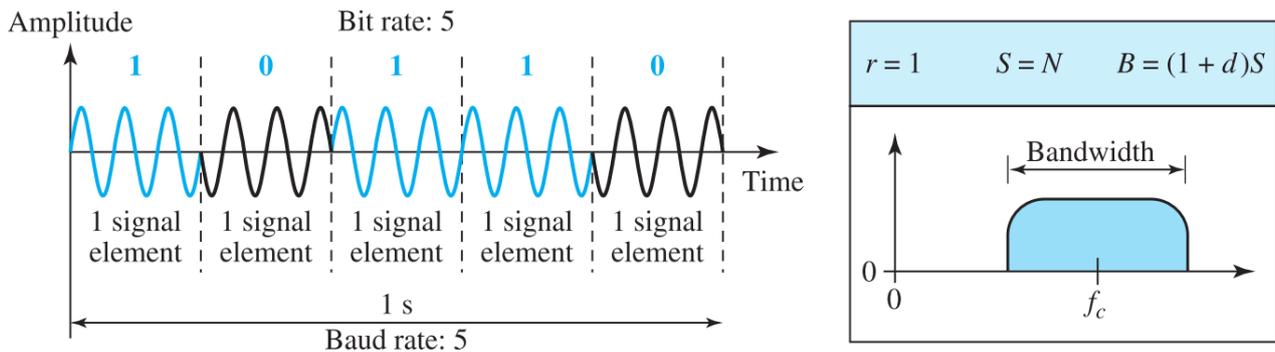


Figura 2.22: Representação das taxas de bit, baud e largura de banda do BPSK. Fonte: [For12].

A Figura 2.23 apresenta o espectro de potência de um sinal BPSK. Como pode ser observado na Figura 2.23 o lóbulo principal ocupa uma largura de banda de $2/T_b$, sendo a mesma observada com o BASK. Diferente do BASK, o sinal BPSK não apresenta a componente da portadora. A supressão da portadora faz com que seja necessário o uso de um receptor coerente para recuperação do fluxo de dados binários [HM07].

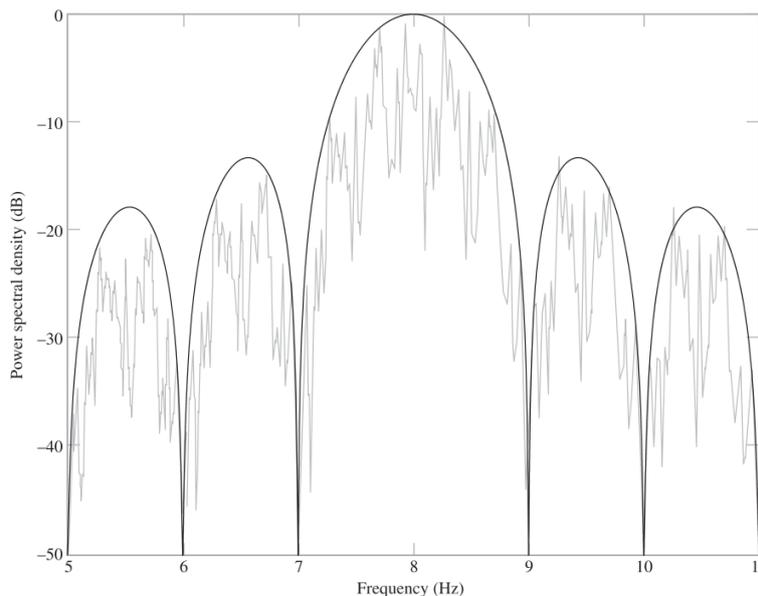


Figura 2.23: Espectro de potência de um sinal BPSK. Fonte: [HM07].

A implementação do BPSK é tão simples quanto a do BASK. O elemento de sinal com fase 180° é o complemento do elemento de sinal com fase 0° . Pode ser usada a

mesma ideia utilizada para o ASK, mas com um sinal polar NRZ em vez de um sinal NRZ unipolar, como mostrado na Figura 2.20. O sinal polar NRZ é multiplicado pela frequência da portadora; o bit 1 (tensão positiva) é representado por uma fase que começa a 0° ; o bit 0 (tensão negativa) é representado por uma fase que começa em 180° [For12].

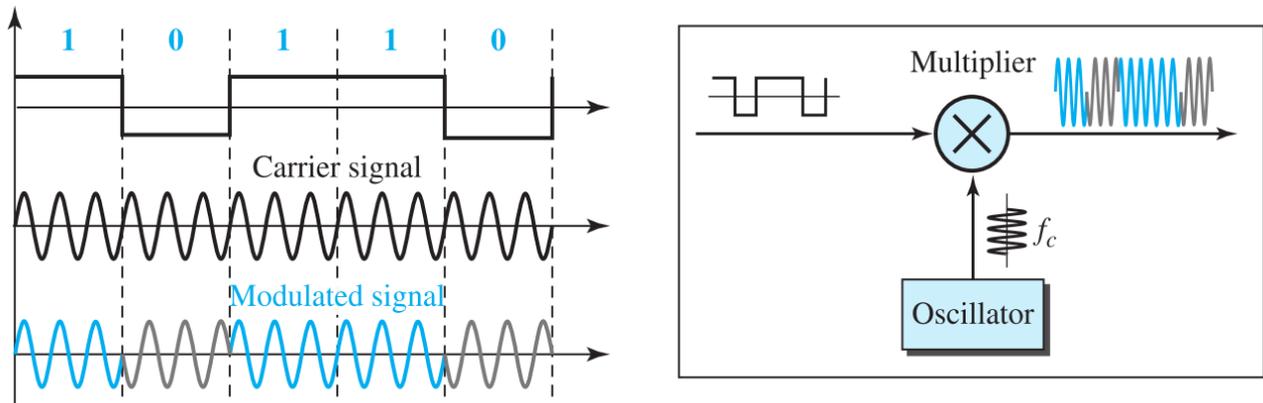


Figura 2.24: Representação simplificada da implementação de um modulador BPSK coerente. Fonte: [For12].

2.2.2.2 Receptor de conversão direta

Os receptores super-heteródinos convertem todos os sinais recebidos em uma frequência mais baixa, conhecida como frequência intermediária (*intermediate frequency* — IF), onde um único conjunto de amplificadores e filtros é usado para fornecer um nível constante de sensibilidade e seletividade. Uma versão especial desses receptores é conhecida como receptor de conversão direta (*Direct Conversion* — DC) ou ainda *Zero-IF* (ZIF). Em vez de traduzir o sinal recebido para uma frequência intermediária, os receptores DC convertem o sinal recebido diretamente para banda base. Em outras palavras, eles realizam a demodulação do sinal como parte da tradução [Fre15].

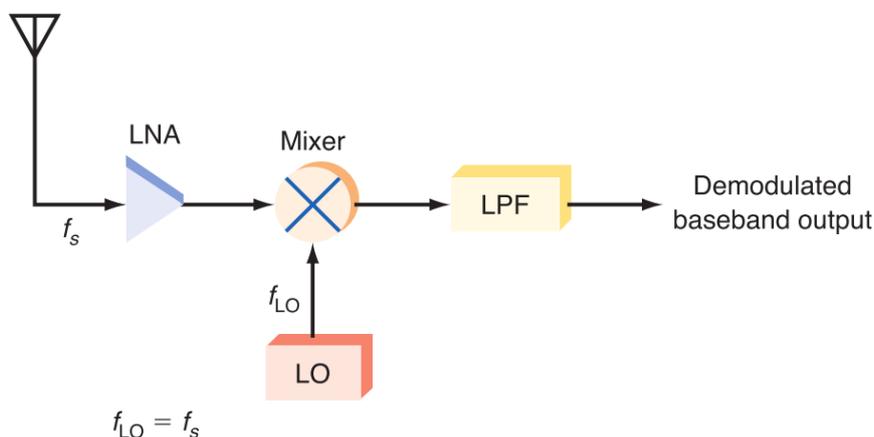


Figura 2.25: Receptor de conversão direta. Fonte: [Fre15].

A Figura 2.25 mostra a arquitetura básica do receptor ZIF. O amplificador de baixo ruído, ou *Low-noise Amplifier* (LNA), aumenta o nível do sinal antes do misturador (*mixer*). A frequência do oscilador local (*Local Oscillator* — LO), f_{LO} , geralmente de um sintetizador de frequência de malha de captura de fase (*Phase-locked Loop* — PLL), é definida para ser a mesma do sinal de entrada f_s ou

$$f_{LO} = f_s \quad (2.9)$$

logo, as frequências de soma e diferença como resultado da mistura são

$$f_{LO} - f_s = 0 \quad f_{LO} + f_s = 2f_{LO} = 2f_s \quad (2.10)$$

A diferença de frequência é zero. Sem modulação, não há saída. Com a modulação em amplitude (*Amplitude Modulation* — AM), as bandas laterais se misturam com a LO para reproduzir o sinal modulador banda base original. Neste caso, o misturador também é o demodulador. A soma é duas vezes a frequência LO, removida pelo filtro de passa-baixa (*Low-pass Filter* — LPF) [Fre15].

No receptor DC, não é necessário um filtro IF separado, já que não ocorre a conversão para uma frequência intermediária. Ele também não necessita de um circuito detector separado porque a demodulação é intrínseca à técnica. Além disso, quando ele é usado em transceptores que usam *half-duplex*, onde o transmissor e o receptor operam na mesma frequência, apenas um sintetizador de frequência PLL é necessário [Fre15]. Por fim, ele não é suscetível aos problemas de imagem que ocorrem nos receptores super-heteródinos [CC09].

Entretanto, o receptor DC apresenta algumas desvantagens, como, por exemplo, é necessário um amplificador de baixo ruído, ou *Low-noise Amplifier* (LNA), senão o sinal LO pode vazar através do misturador para a antena e ser irradiado. Além disso, os circuitos precisam estar perfeitamente balanceados, pois, pode ocorrer um deslocamento indesejável da corrente contínua na saída, que pode alterar os arranjos de polarização em circuitos posteriores, assim como causar a saturação do circuito e impedir a amplificação e outras operações. Finalmente, o receptor Z-IF não reconhece variações de fase ou frequência, como as requeridas pelas modulações FSK e PSK [Fre15].

Para usar essa categoria de receptor com as modulações FSK ou PSK, são necessários dois misturadores em conjunto com um arranjo LO em quadratura, como o mostrado na Figura 2.26. O LNA fornece amplificação. A saída LNA alimenta dois misturadores. O sinal LO alimenta diretamente o misturador superior ($\sin \theta$) e um deslocador de fase de 90° que, alimenta o misturador inferior ($\cos \theta$). Os misturadores fornecem sinais banda base em suas saídas. Os sinais de frequência LO duplicados resultantes da mistura são removidos com os LPFs. Os dois sinais de banda base estão separados em fase por 90° [Fre15].

O sinal superior é geralmente referido como em fase (*In-phase* — I), enquanto o inferior é referido como em quadratura⁶ (*Quadrature* — Q). Os sinais I e Q são então enviados para os conversores analógico-digitais (*Analog-to-digital Converter* — ADC), onde são convertidos em sinais binários. Os sinais binários são então enviados para um processador de sinal digital (*Digital Signal Processor* — DSP). O DSP contém uma sub-rotina pré-armazenada que realiza a demodulação. Esse algoritmo requer dois sinais de quadratura para ter dados suficientes para distinguir as mudanças de fase e de frequência no sinal original resultante da modulação. A saída da sub-rotina de demodulação alimenta um conversor digital-analógico externo (*Digital-to-analog Converter* — DAC) onde o sinal modulador original é reproduzido. Essa arquitetura de conversão direta I-Q é atualmente uma das arquiteturas de receptor mais utilizadas em circuitos integrados (*Integrated Circuit* — IC) de celulares e de redes sem fio [Fre15].

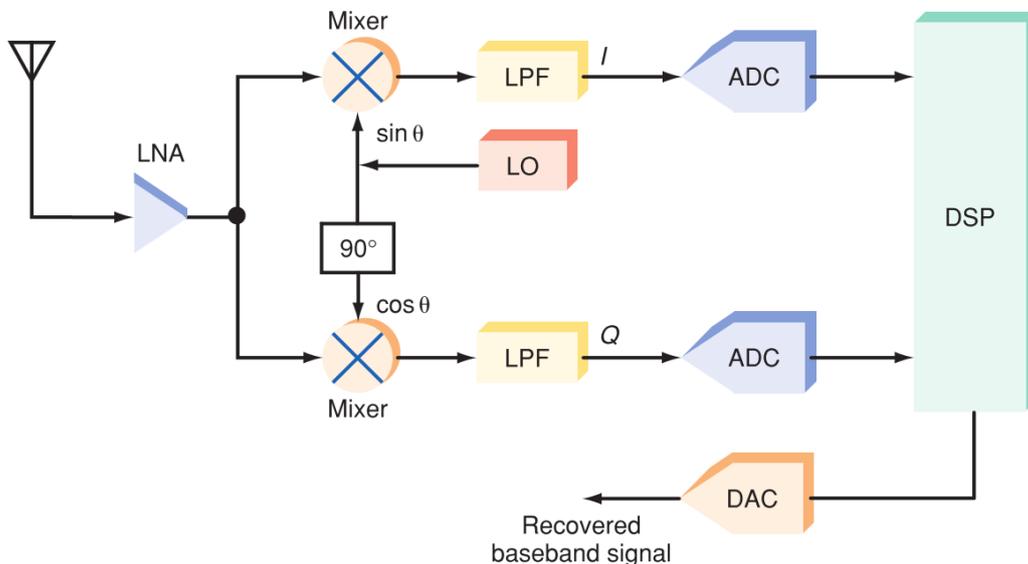


Figura 2.26: Receptor de conversão direta para FSK e PSK. Fonte: adaptado de [Fre15].

2.2.2.3 Arquitetura do transceptor

Nas WIs, as antenas estão interligadas à transceptores. A categoria do transceptor depende da modulação e do mecanismo MAC). As modulações PSK e FSK — que foram apresentadas na Subseção 2.2.2.1 — exigem circuitos complexos que demandam um maior consumo de área e energia [MG15].

A modulação OOK é a mais comum em WiNoCs. Por exigir um circuito simples, ela é facilmente implementável no ambiente em chip e apresenta baixa sobrecarga de área e energia. Como foi explicado na Subseção 2.2.2.1, essa modulação é um tipo especial da modulação ASK. Nela, o sinal varia entre presença ou ausência da portadora conforme o

⁶Quadratura significa uma diferença de fase de 90°.

nível lógico do sinal em banda base for alto ou baixo, respectivamente. Sendo desta forma um circuito simples de ser implementado, além de não exigir muita energia.

A Figura 2.27 apresenta um típico transceptor para WiNoCs. Tanto o transmissor quanto o receptor compartilham o mesmo VCO. No lado do receptor, é utilizado um LNA, um misturador de conversão para baixa frequência e um amplificador banda base. Por outro lado, no transmissor é usado um misturador de conversão para alta frequência e um amplificador de potência.

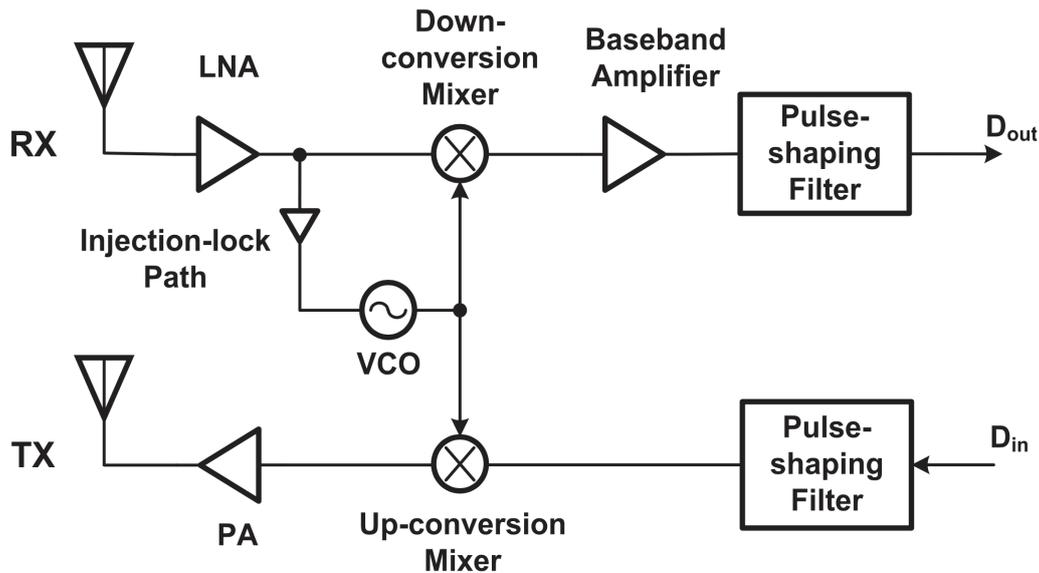


Figura 2.27: Diagrama de blocos de um transceptor OOK. Fonte: [MGS⁺17].

2.2.3 Topologia

A topologia das WiNoCs difere completamente das topologias convencionais, podendo ser uma topologia puramente sem fio ou híbrida, incluindo interligações com e sem fio [WJ14].

2.2.3.1 Topologia puramente sem fio

Em um projeto de NoC puramente sem fio, todas as interligações são substituídas por interligações sem fio [WJ14]. Entretanto, devido às limitações de canais atuais e futuras da tecnologia de comunicação em chip sem fio, não é possível suportar que todas as interligações sem fio funcionem simultaneamente. Isso compromete a escalabilidade dessa topologia [ZW08].

Em [ZWLK11] é proposto uma WiNoC multicanal (*Multi-channel WiNoC* — McWiNoC), a Figura 2.28 ilustra uma McWiNoC em uma malha 2D 4 × 4. A McWiNoC é constituída de um nodo sem fio associado a um processador. Cada nodo sem fio possui uma

antena e um transceptor. Além disso, cada nodo sem fio consegue realizar roteamento, arbitração de múltiplos canais e *buffering*.

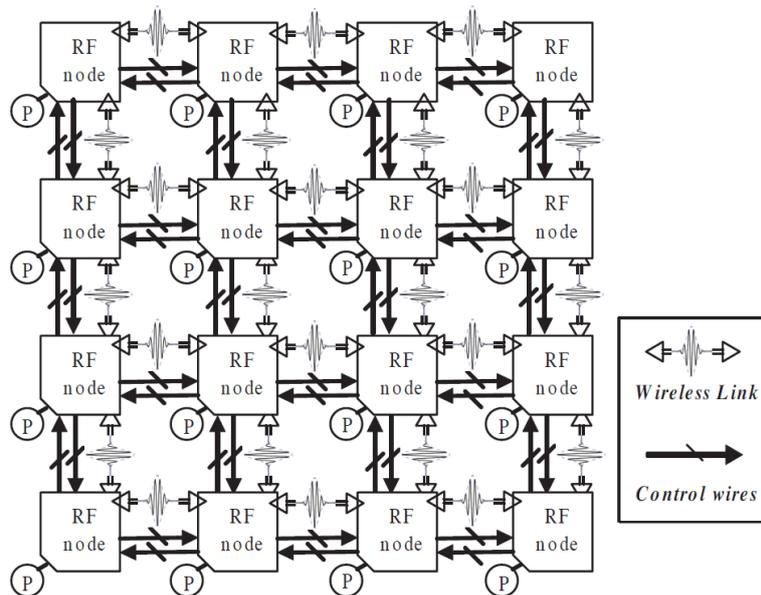


Figura 2.28: McWiNoC baseada em uma malha 4×4 . Fonte: [ZWLK11].

Em contraste a uma NoC regular, a McWiNoC apresenta uma infraestrutura sem fio muito flexível para construir qualquer topologia. É possível gerar diferentes topologias alterando o alcance da transmissão sem fio. Se o alcance da transmissão T é definido com a mesma distância que a distância entre dois nodos sem fio L , uma arquitetura em malha é formada, Figura 2.29a. Quando o T é aumentado de L para $2L$ ou $5L$, são geradas outras topologias, como é ilustrado nas figuras 2.29b e 2.29c, respectivamente. Com o aumento de T , o comprimento médio da interligação para a entrega de dados diminui, o que conseqüentemente reduz o atraso na comunicação fim-a-fim. Entretanto, em simultâneo, o aumento de T pode induzir a uma maior arbitragem de canal que pode aumentar a latência. Além disso, uma área maior do chip precisa ser reservada para nodos sem fio mais potentes [ZWLK11].

2.2.3.2 Topologia híbrida

Em vez de utilizar apenas interligações sem fios para transferir dados entre diferentes nodos, o projeto híbrido utiliza interligações com e sem fios. Ele explora a ideia de que a interligação com fio simples e confiável de curta distância é boa para o tráfego local dentro de um pequeno grupo de nodos, ao passo que a interligação sem fios pode melhorar significativamente o desempenho e a eficiência energética para a comunicação de longa distância. Devido à sua baixa complexidade de projeto e conveniência para ser estendido a partir das arquiteturas NoC existentes, o projeto híbrido atualmente domina nas pesquisas WiNoC [WJ14].

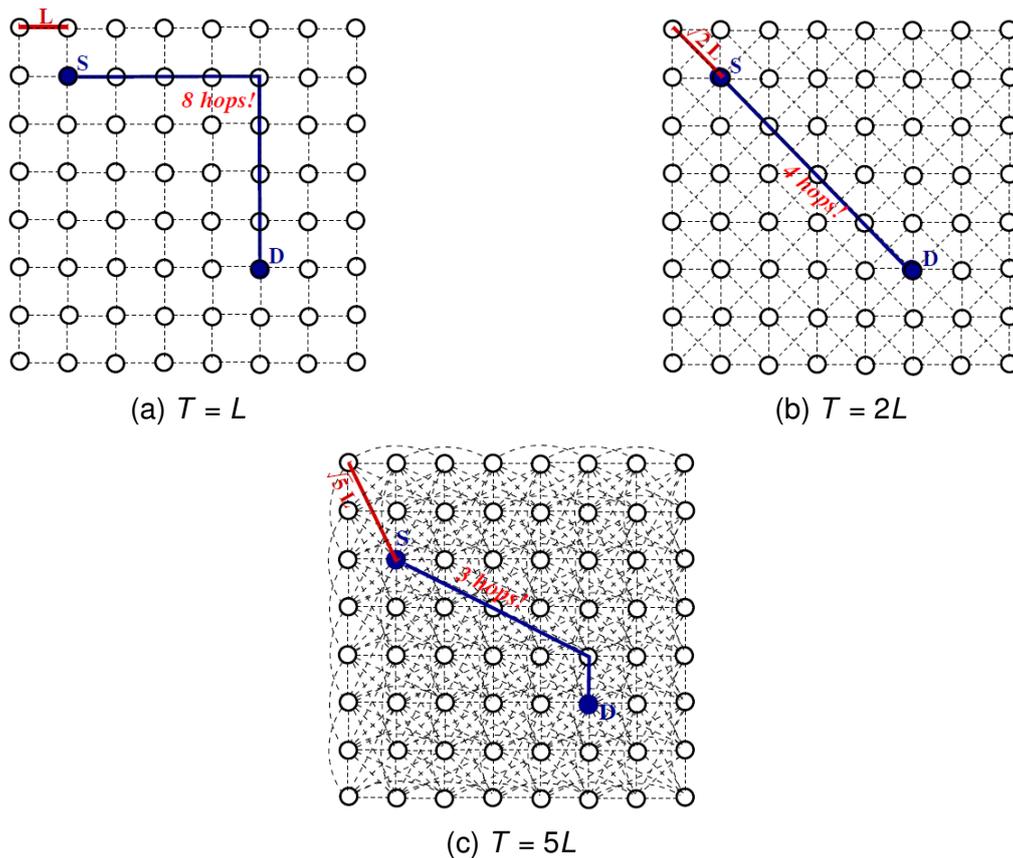


Figura 2.29: Variação das topologias em função do alcance de transmissão. Fonte: [ZWLK11].

Em [WHB11] é apresentada uma arquitetura baseada em uma NoC em malha 2D convencional, porém substituindo alguns dos roteadores convencionais por roteadores sem fio (*Wireless Routers* — WR), sendo que estes possuem interfaces com e sem fio. Dessa forma, os WRs conseguem transferir pacotes a partir de interligações com ou sem fio. Na Figura 2.30 é ilustrada uma WiNoC 10×10 dividida em 4 sub-redes retangulares, onde os WRs estão situados no centro de cada sub-rede. As linhas contínuas e as linhas pontilhadas representam as interligações com fio e sem fio, respectivamente. Cada par transmissor e receptor sem fio utilizam uma frequência da portadora independente para acomodar os dados dos diferentes canais.

A separação de sub-redes e endereços locais permite uma rápida decisão de roteamento e facilita o projeto de sistemas hierárquicos escaláveis, isso também pode simplificar o projeto do roteador e reduzir a complexidade do hardware [WHB11].

Redes de pequeno mundo (*small-world*) tem uma distância média entre qualquer par de nodos, muito pequena. O comprimento médio do percurso mais curto dos grafos pequeno mundo é delimitado por um polinômio $\log(N)$, onde N é o número de nodos, o que a torna particularmente interessante para comunicação eficiente com o mínimo de recursos. Esse recurso dos grafos pequeno mundo torna particularmente atraente para construção de redes em chip sem fio escaláveis [GCD⁺11].

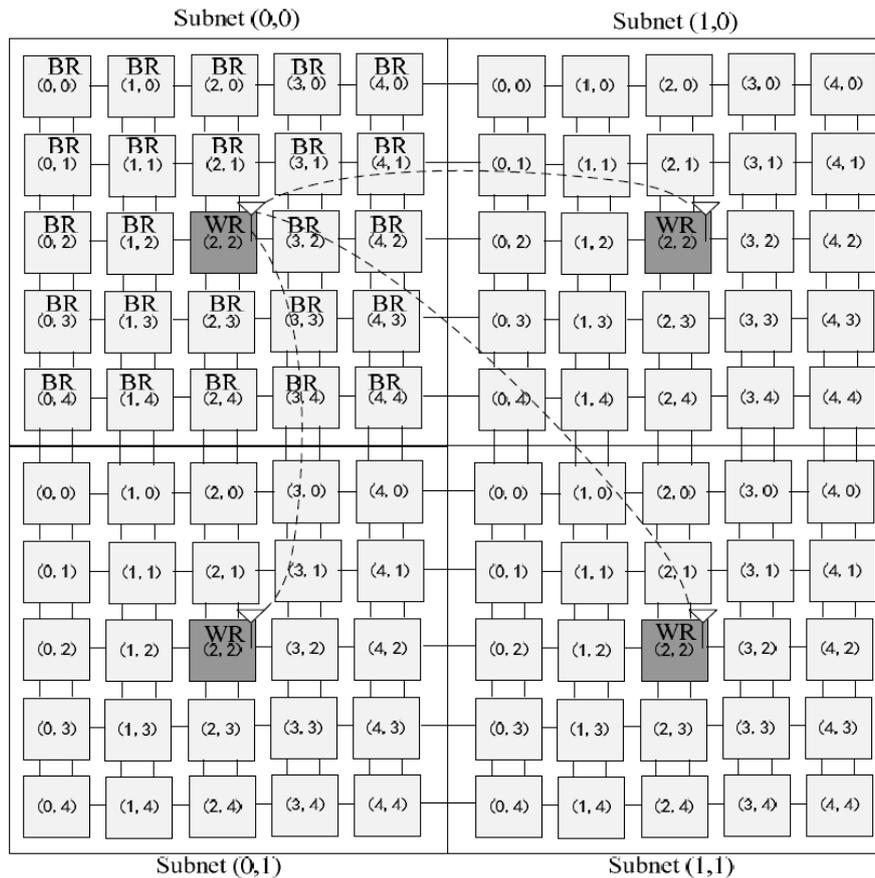


Figura 2.30: *WiNoC* 10 × 10. Fonte: [WHB11].

Em [GCD⁺11] foi utilizada a abordagem pequeno mundo para construir uma NoC altamente eficiente com base em ambas as interligações com e sem fios. Inicialmente todo o sistema foi dividido em vários pequenos aglomerados de núcleos vizinhos, chamados sub-redes. A Figura 2.31a mostra uma sub-rede com topologia em malha. Essa sub-rede tem comutadores da NoC e interligações como em uma NoC comum baseada em malha. Os núcleos são conectados a um concentrador localizado no centro através de interligações diretas e os concentradores de todas as sub-redes são conectados em uma rede de segundo nível formando uma rede hierárquica.

Esse nível superior da hierarquia é projetado para ter características de grafos pequeno mundo. Devido a um número limitado de possíveis interligações sem fios, certos vizinhos estão ligados por fio, formando um anel bidirecional e algumas interligações sem fios são distribuídas entre os núcleos separados por distâncias relativamente longas. Reduzir os múltiplos saltos na comunicação com fio de longa distância é essencial, de modo a obter o máximo benefício da rede em chip sem fio para sistemas com múltiplos núcleos [GCD⁺11].

A Figura 2.31b mostra uma possível topologia de interconexão com oito concentradores e três interligações sem fios. O tamanho e o número de sub-redes são escolhidos de tal modo que nem as sub-redes, nem o nível superior da hierarquia tornem-se demasiada-

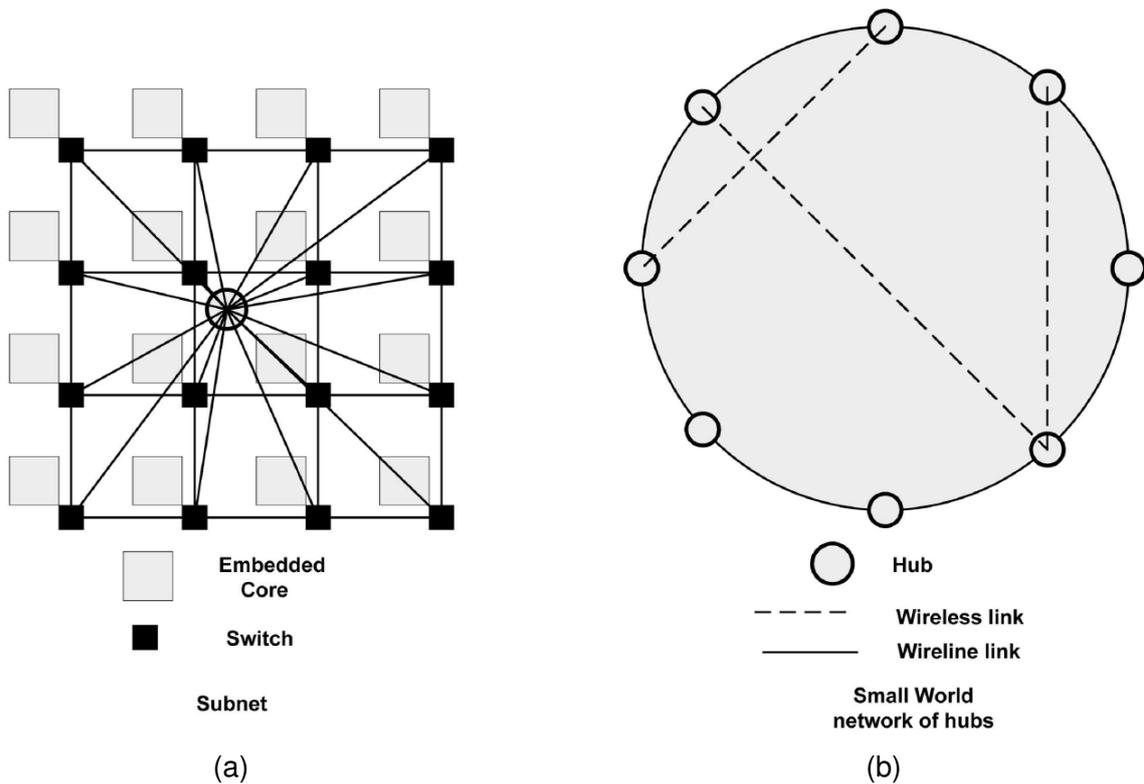


Figura 2.31: (a) Topologia em malha de uma sub-rede e (b) Topologia *small-world* interligando os *hubs*. Fonte: [GCD⁺11].

mente grandes. Isso porque, caso um nível se torne muito grande o concentrador daquela sub-rede será um gargalo, devido à limitação de banda do concentrador [GCD⁺11].

2.3 Mecanismos MAC

Quando interfaces sem fio compartilham o mesmo meio é necessário um mecanismo de acesso-múltiplo para coordenar o acesso ao mesmo. O problema de controlar o acesso ao meio é semelhante às regras de falar em uma discussão. Os procedimentos garantem o direito de falar e garantem que duas pessoas não falem em simultâneo, não se interrompam, não monopolizem a discussão, e assim por diante. Muitos protocolos foram concebidos para lidar com o acesso ao meio compartilhado. Todos esses protocolos pertencem a subcamada MAC situada acima do nível físico [For12]. O acesso múltiplo se refere a forma de como os nodos estão alocados no espectro de frequência atribuído. Os métodos de acesso são as maneiras pelas quais muitos nodos compartilham uma quantidade limitada de espectro [Fre15].

Esses mecanismos podem ser classificados em três grupos: de acesso canalizado, de acesso aleatório e de acesso controlado. Nas próximas seções são detalhados os

mecanismos pertencentes a cada um desses grupos relevantes para o entendimento desta tese.

2.3.1 MACs de Acesso Canalizado

Os mecanismos de acesso canalizado têm como característica o compartilhamento da largura de banda disponível nos domínios da frequência, do tempo ou por códigos especiais. As subseções 2.3.1.1, 2.3.1.2 e 2.3.1.3 detalham esses mecanismos.

2.3.1.1 FDMA

No FDMA, a largura de banda disponível é dividida em sub-bandas de frequência. A cada nodo é atribuído a uma dessas sub-bandas para enviar seus dados e essa fica reservada para essa interface sem fio o tempo todo. Cada nodo também usa um filtro passa-faixa⁷ para limitar a recepção do espectro somente relativo à sua sub-banda [For12].

Como cada nodo ocupa a faixa de frequência atribuída ao mesmo. Para um nodo k , a mensagem m_k é modulada digitalmente para a faixa de frequência atribuída f_k . Os sinais dos nodos que estão transmitindo aparecerão na antena de todos os nodos que estão recebendo. No receptor, o sinal que aparece na antena é a sobreposição de todos os sinais transmitidos de todos os nodos ativos. No entanto, como todos os nodos ativos são atribuídos com bandas de frequência diferentes, os sinais transmitidos dos usuários não se sobrepõem no domínio da frequência [LCL07].

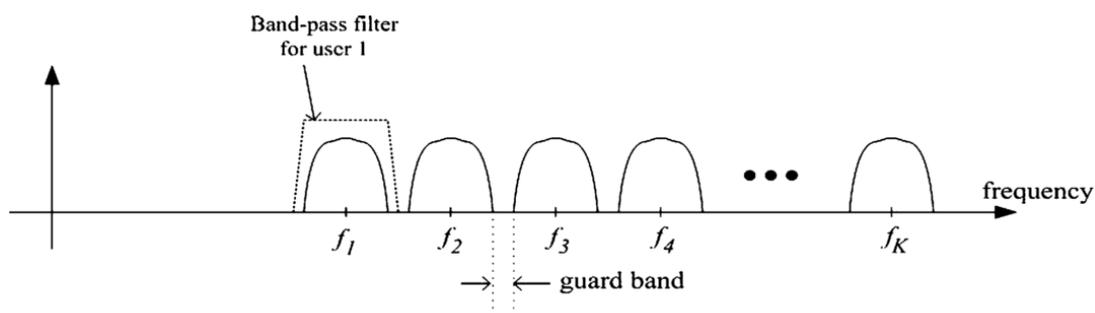


Figura 2.32: Espectro de frequência do FDMA. Fonte: [LCL07].

A Figura 2.32 ilustra o espectro do sinal recebido na antena, que alimenta então os filtros passa-faixa K com frequências centrais f_1, f_2, \dots, f_k . Por exemplo, o filtro de passa-faixa para o primeiro nodo só pode deixar passar o sinal em torno da frequência central f_1 com a largura de banda adequada e rejeitar os outros. Portanto, o sinal de saída do primeiro filtro passa-faixa contém apenas a forma de onda transmitida do primeiro nodo

⁷Filtro passa-faixa é um circuito que permite a passagem de frequências de uma determinada faixa e atenua frequências fora dessa faixa.

sem interferência dos outros. Devido ao efeito não ideal do filtro passa-faixa, bandas de proteção são utilizadas para minimizar a interferência de um canal no outro [LCL07].

2.3.1.2 TDMA

No TDMA, os nodos compartilham a largura de banda do canal no domínio do tempo. Cada nodo recebe um subintervalo de tempo durante o qual pode enviar seus dados. Como é mostrado na Figura 2.33, o nodo 1 pode transmitir dados apenas nos intervalos de tempo atribuídos a ele. Da mesma forma, o nodo 2 também só pode transmitir dados nos intervalos de tempo atribuídos a ele e assim sucessivamente. Assim como ocorre no FDMA, como cada nodo opera em um canal diferente, um nodo não interfere na transmissão do outro [LCL07].

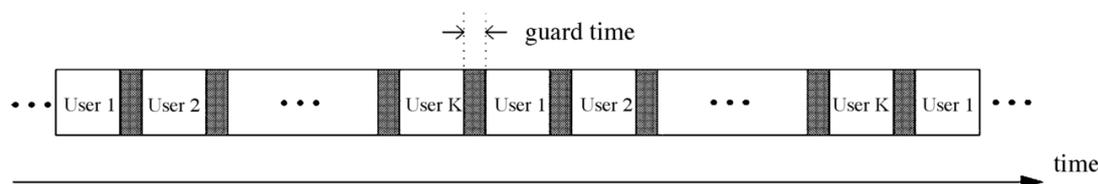


Figura 2.33: *slots* do TDMA. Fonte: [LCL07].

O principal problema no TDMA consiste em conseguir obter a sincronização entre os diferentes nodos. Cada nodo precisa saber o início e a localização do seu intervalo de tempo e para isso precisam estar sincronizados. Isso pode ser difícil devido a atrasos de propagação introduzidos no sistema, principalmente se os nodos estiverem espalhados por uma grande área. Para compensar os atrasos, tempos de proteção podem ser inseridos entre os subintervalos de tempo. A sincronização é normalmente realizada por alguns bits de sincronização (normalmente referidos como bits de preâmbulo) no início de cada intervalo [For12].

2.3.1.3 CDMA

O CDMA utiliza o espalhamento espectral⁸ por sequência direta (*Direct-sequence Spread Spectrum* — DSSS). Nessa categoria de espalhamento espectral, um código binário pseudoaleatório a uma taxa superior é misturado com os dados binários seriais. A Figura 2.34 ilustra um transmissor com DSSS, onde os dados binários seriais são aplicados a uma porta Ou-Exclusivo (*Exclusive Or* — XOR) em conjunto com o código pseudoaleatório [Fre15].

Cada bit do código pseudoaleatório é chamado *chip* e a sua taxa é chamada taxa de *chipping*. O sinal de saída da porta XOR é então aplicado a um modulador PSK, típica-

⁸O espalhamento espectral é uma técnica de modulação e multiplexação que distribui o sinal e suas bandas laterais em uma largura de banda muito maior.

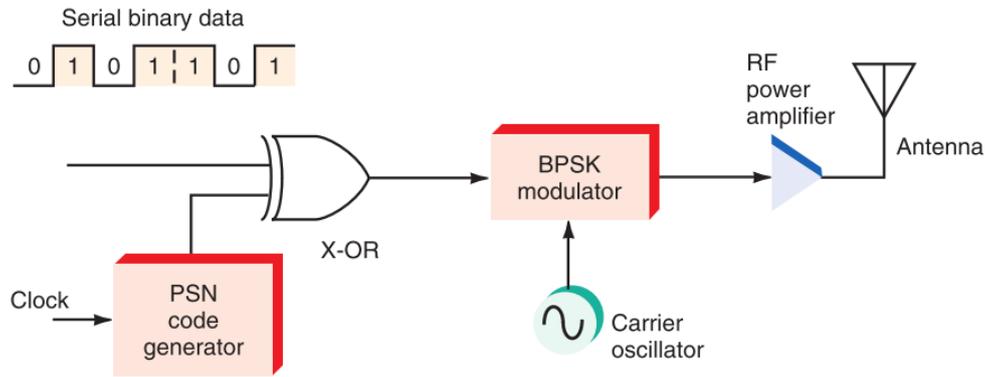


Figura 2.34: Transmissor com espalhamento espectral por sequência direta. Fonte: [Fre15].

mente BPSK, mas também pode ser utilizado com QPSK⁹ ou outras formas do PSK. A fase da onda portadora é comutada entre 0° e 180° conforme a saída da porta XOR for 0 ou 1. A Figura 2.35 mostra a forma de onda dos sinais envolvidos [Fre15].

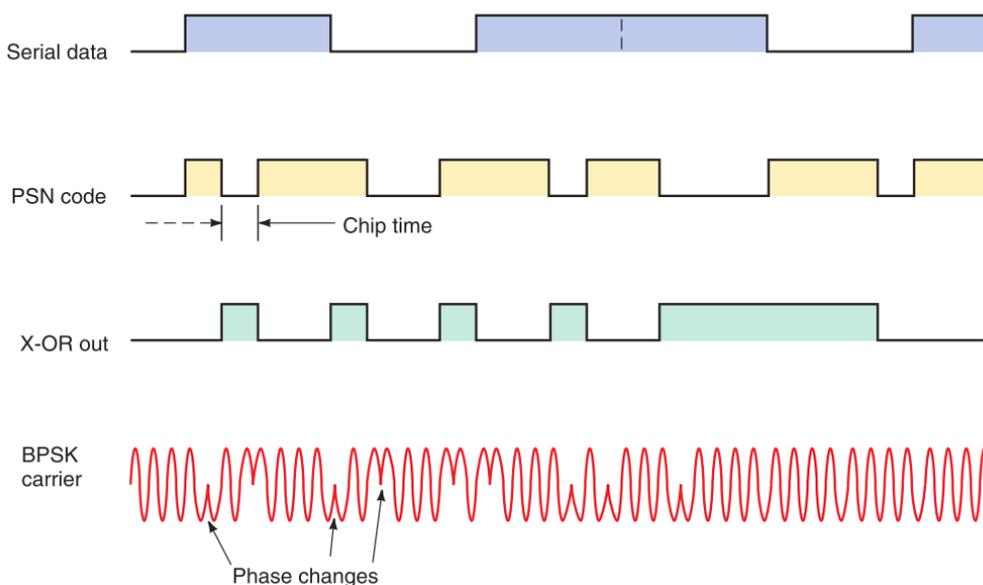


Figura 2.35: Forma de onda dos sinais envolvidos. Fonte: [Fre15].

Para gerar as seqüências de código *chip*, é utilizada a tabela de *Walsh*, que é uma tabela bidimensional com um número igual de linhas e colunas. A regra geral para criação da Tabela é mostrada na Figura 2.36a [For12].

Na Tabela de *Walsh*, cada linha é uma seqüência de *chips*. Para uma seqüência de um *chip*, W_1 , temos uma linha e uma coluna. Podem ser escolhidos os valores -1 ou +1 para o *chip* dessa tabela (No exemplo da Figura 2.36a, foi escolhido o +1). Ainda conforme a regra geral, ao conhecer a tabela para N seqüências W_N , podem ser criadas tabelas para $2N$ seqüências W_{N2} [For12].

⁹QPSK (*Quadrature Phase Shift Keying*) é uma modulação derivada do PSK, que utiliza os parâmetros de fase e quadratura para modular o sinal.

$$\begin{aligned}
 W_1 &= \begin{bmatrix} +1 \end{bmatrix} & W_{2N} &= \begin{bmatrix} W_N & W_N \\ W_N & \overline{W_N} \end{bmatrix} \\
 & & & \text{(a)} \\
 W_2 &= \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} & W_4 &= \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \\
 & & & \text{(b)}
 \end{aligned}$$

Figura 2.36: Regra geral (a) e exemplos (b) de criação de tabelas *Walsh*. Fonte: [For12].

A Figura 2.36b mostra como podem ser criadas as tabelas W_2 e W_4 a partir do W_1 . Após selecionar a tabela W_1 , a tabela W_2 pode ser criada de quatro tabelas W_1 , com a última sendo o complemento da tabela W_1 . Depois que a tabela W_2 é gerada, a tabela W_4 pode ser criada de quatro tabelas W_2 , com a última sendo o complemento da tabela W_2 . Após a tabela W_N ser gerada, cada nodo recebe um *chip* correspondente a uma linha [For12].

2.3.2 MACs de Acesso Aleatório

Em métodos de acesso aleatório ou de contenção, nenhum nodo é superior ao outro e a nenhum nodo é atribuído o controle sobre o outro. Em cada instância, um nodo que possui dados para enviar usa um procedimento definido pelo protocolo para decidir sobre se deve ou não enviar seus dados. Essa decisão depende do estado do meio (ocioso ou ocupado). Cada nodo pode transmitir quando deseja, se seguir o procedimento predefinido, incluindo o teste do estado do meio [For12].

Dois recursos dão a esse método o nome. Primeiro, o nodo pode transmitir a qualquer tempo. A transmissão é aleatória entre os nodos. É por isso que esses métodos são chamados acesso aleatório. Em segundo lugar, não há uma ordem pré-estabelecida da ordem de quem envia primeiro. Os nodos competem entre si para acessar o meio. É por isso que esses métodos também são chamados métodos de contenção [For12].

Em um método de acesso aleatório, cada nodo tem o direito ao meio sem ser controlado por qualquer outro nodo. No entanto, se mais de uma estação tentar enviar, ocorre um conflito de acesso, colisão, e os dados serão destruídos ou modificados. Para evitar conflitos de acesso ou para resolvê-lo quando acontece, cada nodo segue um procedimento descrito pelo respectivo protocolo [For12].

2.3.2.1 CSMA

O CSMA requer que cada nodo primeiro escute o meio (ou verifique o estado do meio) antes de transmitir. O CSMA baseia-se no princípio “ouça antes de falar” (“*listen before talk*”). Ao seguir esse princípio ele pode reduzir a possibilidade de colisão, mas não a eliminar. A possibilidade de colisão ainda existe devido ao atraso de propagação, como é ilustrado no modelo espacial e temporal de uma rede CSMA da Figura 2.37, onde os nodos estão conectados a um canal compartilhado [For12].

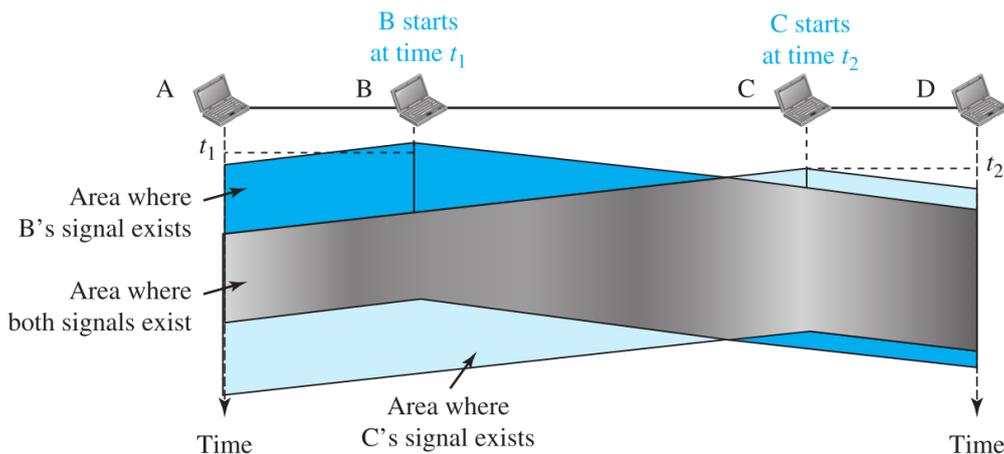


Figura 2.37: Modelo espaço-tempo de uma colisão no CSMA. Fonte: [For12].

Quando uma estação envia dados, ainda leva um tempo para o primeiro bit chegar aos outros nodos e para cada nodo detectá-lo. Em um dado momento algum nodo poderia sentir o meio e encontrá-lo ocioso, apenas porque o primeiro bit enviado por outro nodo ainda não foi recebido. Como pode ser observado na Figura 2.37, no instante t_1 , o nó B detecta o meio e o encontra ocioso, então ele envia os dados. No tempo t_2 ($t_2 > t_1$), o nó C detecta o meio e o encontra ocioso porque, agora, os primeiros bits do nó B ainda não chegaram nele. O nó C também envia os dados. Os dois sinais colidem e ambos os dados são destruídos [For12].

O tempo de vulnerabilidade do CSMA é igual ao tempo de propagação T_p . Esse é o tempo necessário para que um sinal se propague de uma extremidade do meio para a outra. Quando um nó envia dados e qualquer outro nó envia dados nesse mesmo instante, uma colisão ocorre. Entretanto, se o primeiro bit chegar ao final da extremidade do meio, todos os nodos já terão escutado o bit e se absterão de enviar. A Figura 2.38 mostra o pior caso, onde o nó mais à esquerda, A, envia dados no tempo t_1 , que atinge o nó mais à direita, D, no tempo $t_1 + T_p$. A área em azul mostra a área vulnerável no tempo [For12].

O CSMA pode ter três comportamentos diferentes caso encontre o canal ocupado, dependendo do método de persistência adotado. O método 1-persistente, é simples e direto. Nesse método, após o nó encontrar o meio ocioso, ele envia os bits imediatamente,

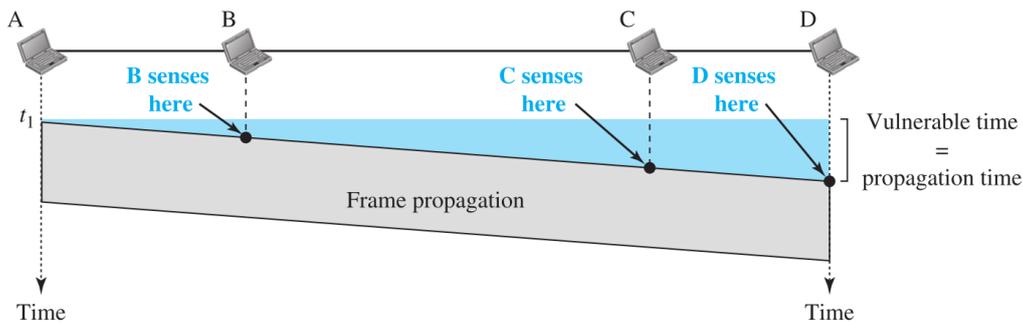


Figura 2.38: Tempo vulnerável no CSMA. Fonte: [For12].

com probabilidade 1, como é mostrado na Figura 2.39a e no diagrama de fluxo da Figura 2.40a. Esse método tem uma maior probabilidade de colisão, porque dois ou mais nós podem encontrar o meio ocioso e enviar os seus bits imediatamente. Por outro lado, no método não persistente, Figura 2.39b, o nó que tem dados a serem enviados escuta o meio, se estiver ocioso, ele envia os seus bits imediatamente. Caso contrário, se ele estiver ocupado, ele espera um intervalo de tempo aleatório e então escuta o meio novamente. Esse fluxo é mostrado na Figura 2.40b. A abordagem não-persistente reduz a probabilidade de colisão porque é improvável que dois ou mais nós aguardem a mesma quantidade de tempo e tentem enviar simultaneamente novamente. No entanto, esse método reduz a eficiência da rede porque o meio permanece inativo enquanto poderia haver nós enviando dados. [For12].

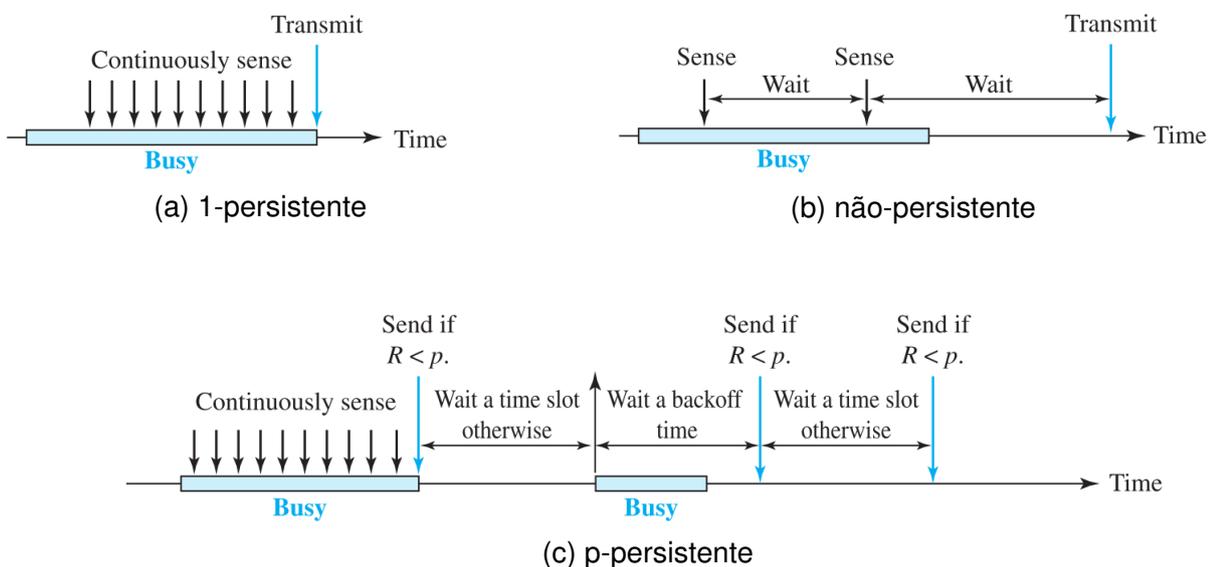


Figura 2.39: Comportamento dos três métodos de persistência. Fonte: [For12].

Por fim, o método p-persistente é usado se o canal tem intervalos de tempo com duração igual ou maior que o tempo máximo de propagação. Essa abordagem combina as vantagens das outras duas estratégias. Ela reduz a probabilidade de colisão e aumenta a eficiência. Como é mostrado na Figura 2.40c, nesse método, após o nó encontrar o meio ocioso, ele pode enviar os dados com probabilidade p , ou não enviar os dados com

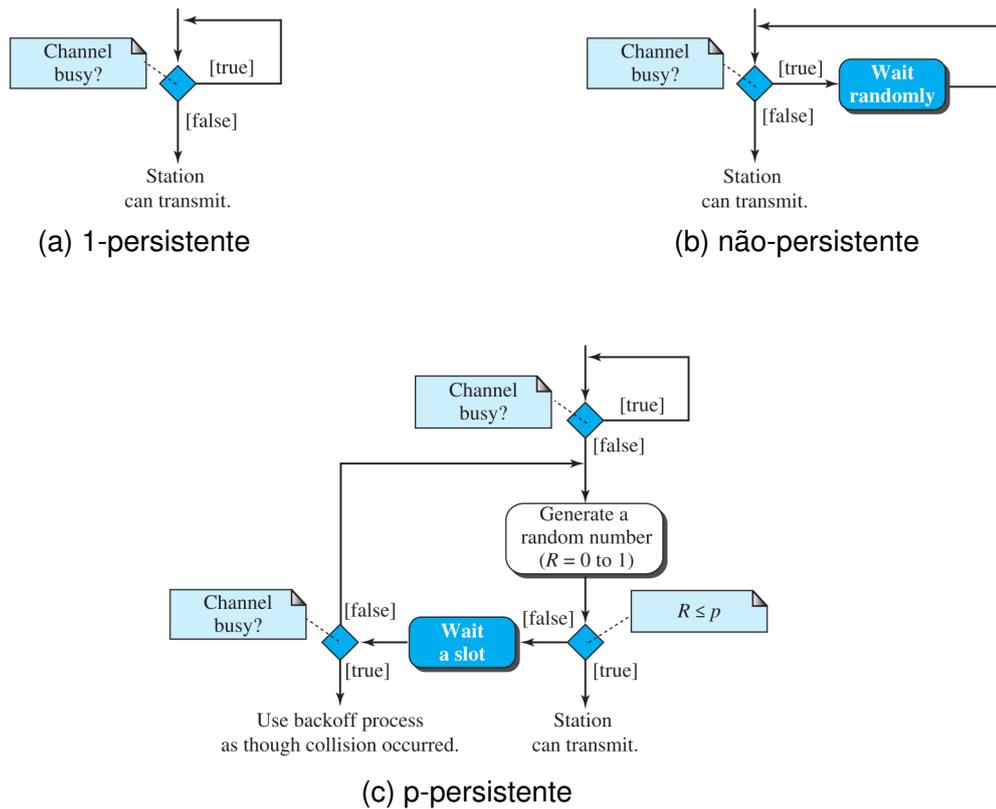


Figura 2.40: Diagrama de fluxo dos três métodos de persistência. Fonte: [For12].

probabilidade $q = 1 - p$. Nesse caso, o nó espera o início do próximo intervalo de tempo e escuta o meio novamente. Se o canal estiver ocioso ele envia com probabilidade p novamente. Caso contrário, se o canal estiver ocupado, ele atua como se tivesse ocorrido uma colisão e aguarda por um tempo aleatório de recuo. Esse tempo cresce exponencialmente com o número de tentativas de enviar os dados sem êxito ou com o número de colisões [For12].

2.3.3 MACs de Acesso Controlado

No mecanismo de acesso de controlado, os nós precisam de uma autorização para poder acessar o meio. Dessa forma, um nó só pode enviar seus dados se estiver com a posse dessa autorização. Nesse método os nós precisam consultar um ao outro para determinar o nó que tem o direito de transmitir [For12].

2.3.3.1 Acesso múltiplo por passagem da ficha

Nesse método, um pacote que contém uma ficha circula através dos nós. A posse da ficha dá ao nó o direito de acessar o canal e enviar seus dados. Quando um nó tem alguns dados para enviar, aguarda até receber a ficha do nó antecessor. Em

seguida, ele se mantém com a ficha e envia seus dados. Quando o nodo não tem mais dados para enviar, libera a ficha, passando-a para o próximo nodo. O nodo não pode enviar dados até receber a ficha novamente na próxima rodada. Nesse processo, quando um nodo recebe a ficha e não tem dados para enviar, ele apenas passa a ficha para o próximo nodo [For12].

Os nodos precisam ser limitados no tempo em que podem permanecer com a ficha. A ficha também deve ser monitorada para garantir que ela não tenha sido perdida ou destruída. Por exemplo, se um nodo que está segurando a ficha falhar, a ficha desaparecerá da rede. Por isso o gerenciamento da ficha é necessário para esse método de acesso. Outra função do gerenciamento da ficha é atribuir prioridade aos nodos e as categorias de dados que estão sendo transmitidos. Por fim, o gerenciamento da ficha é necessário para que os nodos de baixa prioridade liberem a ficha para os nodos de alta prioridade [For12].

3. ESTADO-DA-ARTE EM MECANISMOS MAC PARA WINOCS

A arquitetura considerada nessa tese, DWiNoC, utiliza a passagem da ficha como mecanismo MAC nas regiões formadas pela união dos lóbulos principais das antenas direcionais do transmissor e receptor [MGS⁺17]. Uma ampla pesquisa bibliográfica por publicações que envolvessem mecanismos MAC específicos para DWiNoCs foi realizada. Entretanto, não foram encontrados mais trabalhos com MACs específicos para DWiNoCs. Desta forma, optou-se por realizar uma pesquisa mais ampla incluindo publicações que envolvessem mecanismos MAC utilizados também em WiNoCs. Os mecanismos MAC já propostos para WiNoCs podem ser classificados em três grupos: com acesso controlado (Passagem da ficha), canalizados (CDMA, FDMA e TDMA) e com acesso aleatório (CSMA). Nas próximas seções são apresentados os mecanismos MAC no estado-da-arte relativos a cada um desses grupos. Ao final, é realizada uma discussão comparando esses mecanismos MAC com o mecanismo proposto nesse trabalho.

3.1 Passagem da Ficha

O mecanismo de controle de acesso ao meio por passagem da ficha já foi utilizado por diversas arquiteturas de WiNoCs [CDG⁺12] [DKKM11] [PCMC15] [MLW⁺14] [MKW⁺14] [DCY⁺13] [WKM⁺14] [WMK⁺14] [DKM⁺15] [DGC⁺10]. Nesse mecanismo, que teve seu funcionamento detalhado na Subseção 2.3.3.1, um *flit* especial denominado ficha circula através das WIs. A WI que detêm a posse da ficha ganha a permissão de acessar o meio sem fio compartilhado. O restante das WIs precisa aguardar as próximas rodadas de circulação da ficha, que ocorre logo após a liberação da mesma pela WI que no momento detêm a sua posse.

Poucos trabalhos propuseram otimizações em cima do algoritmo básico de funcionamento desse mecanismo. Em [MGY13] é proposto um aprimoramento desse mecanismo para garantir maior robustez aos problemas inerentes ao mecanismo da passagem da ficha, como a duplicação ou a perda da ficha. Para isso, foram adicionadas informações adicionais à ficha, o endereço do da última WI e endereço da próxima WI. Cada WI é equipada com uma unidade de gerenciamento da ficha (*Token Management Unit* — TMU), que possui registradores de estado, que mantêm as informações pertencentes à ficha. A Figura 3.1 apresenta a arquitetura do TMU bem como a estrutura do *flit* que atua como ficha.

O controle do TMU é realizado pelo gerente de ação (*Action Manager* — AM). O AM atualiza o registrador do *ID_current* após receber a ficha com o *nextWI* e define o registrador *HasToken* como 1 se o *nextWI* for igual ao *ID* do TMU local. A WI só pode transmitir se *ID_Current* for igual ao *ID* da WI e *HasToken* estiver definido para 1. Ao validar

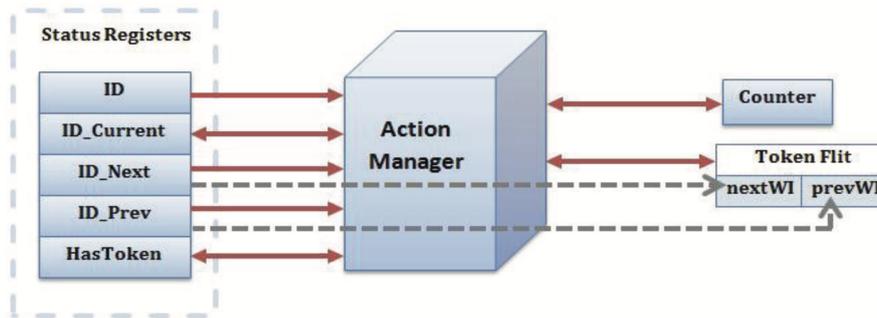


Figura 3.1: Arquitetura do TMU e do *flit* que atua como ficha. Fonte: [MGY13].

o $ID_Current$ em conjunto com $HasToken$ o TMU garante que a ficha não será duplicada. Caso a ficha seja perdida o TMU a regenerará na próxima WI. O AM inicia um contador quando ele recebe uma ficha onde o $nextWI$ é igual ao ID_Prev . Se a ficha não for liberada no tempo T_{max} a ficha é regenerada com o correto $nextWI$ e $prevWI$. Os resultados mostraram que o TMU conseguiu mitigar a redução na taxa de transferência e o aumento do consumo de energia decorrente dos problemas de perda da ficha ou de ficha duplicada que impactavam significativamente o desempenho do canal compartilhado.

Em [PCMC15] é proposto outro aprimoramento para o mecanismo de passagem da ficha, onde o tempo em que cada WI permanece com a ficha é ajustado dinamicamente e fica abaixo de um determinado limiar. Com isso, se comparado com mesmo mecanismo sem o aprimoramento proposto, a latência e o consumo foram reduzidos em 30% e 25%, respectivamente. A Figura 3.2 apresenta esse mecanismo baseado na passagem da ficha implementado a partir de uma topologia em anel, pelo qual a ficha circula. Outra abordagem é aliar o mecanismo da passagem da ficha com algum outro mecanismo MAC. Na Seção 3.5 é apresentado um trabalho que propôs esse tipo mecanismo híbrido entre a passagem da ficha e o CSMA.

Por ser amplamente utilizado em WiNoCs, o mecanismo de passagem da ficha é geralmente utilizado como referência para fins de comparação de desempenho com outros mecanismos de controle de acesso ao meio.

3.2 TDMA

Em [MSG16] são propostos dois mecanismos MAC baseados no TDMA, um proporcional (*Proportionate TDMA* — P-TDMA) e o outro dinâmico (*Dynamic TDMA* — D-TDMA), capazes de ajustar dinamicamente os intervalos do tempo de transmissão de cada WI. Tanto o P-TDMA quanto o D-TDMA exigem um mecanismo de predição da largura de banda. A arquitetura desse mecanismo, comum tanto no P-TDMA quanto no D-TDMA é mostrado na Figura 3.3.

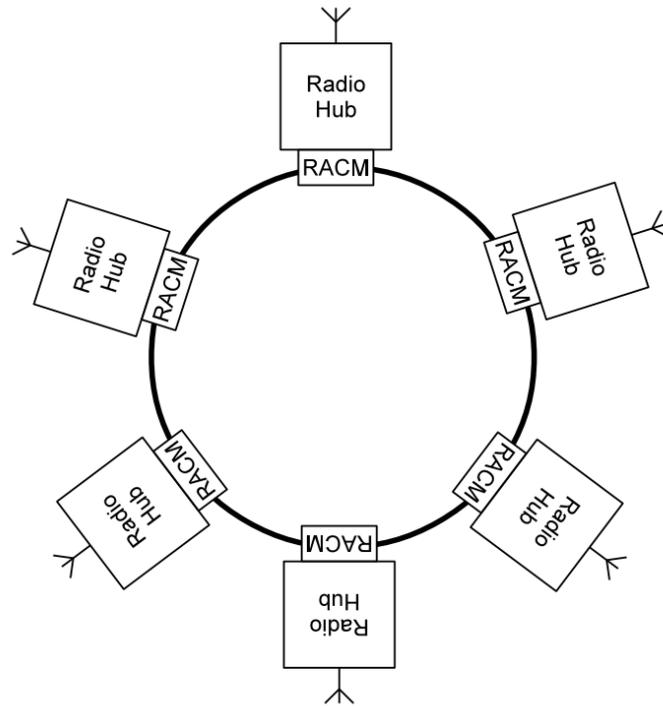


Figura 3.2: Mecanismo de passagem da ficha implementado sob uma topologia em anel. Fonte: [PCMC15].

O ajuste dinâmico ocorre com base em uma estimativa da demanda prevista. Esses mecanismos atuam como um complemento do mecanismo de passagem da ficha ajustando o tempo de retenção da ficha. Para isso, o *flit* com a ficha carrega também os campos *TokenID*, *PrevWI*, *NextWI* e *Demand*. A demanda de largura de banda prevista é dada por

$$\hat{B}^{TP_{j+1}} = \frac{BD^{TP_j} + \overline{BD}^{TP}}{2} \quad (3.1)$$

onde, BD^{TP_j} é a demanda atual de largura de banda de uma WI, medida como o número total de *flits* que chegam no período de posse da ficha j e \overline{BD}^{TP} é a demanda média de largura de banda prevista por uma WI no período de posse da ficha 0 até $j - 1$. A média móvel dos últimos períodos de posse da ficha captura a demanda no estado estacionário das WIs. A demanda do último período captura a variação mais recente da demanda. Assim, a demanda de largura de banda prevista captura tanto as demandas de longo prazo quanto as instantâneas de uma WI. Este valor de demanda prevista é usado para ajustar os intervalos de tempo para o próximo período de posse da ficha.

No P-TDMA, o período de posse da ficha é adaptado dinamicamente para cada WI com base na demanda proporcional de largura de banda da WI em comparação a outras WIs. O número de intervalos de tempo no período de posse da ficha de uma WI é alocado dinamicamente no início de cada período de posse da ficha para lidar com a variação na demanda por largura de banda das WIs. No entanto, essa alocação de intervalos de tempo

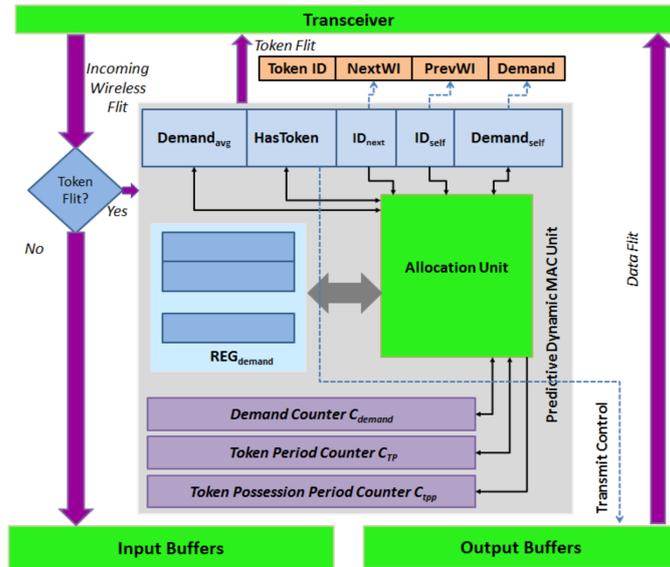


Figura 3.3: Arquitetura comum do P-TDMA e do D-TDMA. Fonte: [MSG16].

é restrita de tal forma que o período de posse da ficha permanece constante entre as alocações. Os intervalos atribuídos de um WI, i no período de posse da ficha, $j + 1$, é dado por,

$$S_i^{TP_{j+1}} = \frac{\hat{B}_i^{TP_j}}{\sum_{i=1}^N \hat{B}_i^{TP_j}} \times S_{TP} \quad (3.2)$$

onde, S_{TP} é o número de intervalos de tempo para os *flits* de dados no período de posse da ficha e N é o número de WIs.

Na unidade preditiva do MAC dinâmico de cada WI são utilizados cinco registradores: ID_{self} , ID_{next} , $HasToken$, $Demand_{avg}$ e $Demand_{self}$, para armazenar seu próprio ID, o ID da próxima WI na circulação da ficha, a posse da ficha, a demanda média de largura de banda e a demanda de largura de banda prevista para a WI, respectivamente. Um vetor de registradores, REG_{demand} é usado para armazenar a demanda prevista de largura de banda de outras WIs. O contador de demanda, C_{demand} é usado para contar a utilização durante o período de posse da ficha. O contador do período de posse ficha, o C_{TP} , efetua a contagem regressiva do valor constante do período da ficha até zero. Quando o contador do período de posse da ficha, C_{TP} , expira, o C_{Tpp} é carregado com o número de intervalos de tempo para o próximo período de posse da ficha. Em seguida, o valor de C_{demand} e $Demand_{avg}$ são usados para calcular a demanda prevista de largura de banda. Esta previsão é armazenada no $Demand_{self}$. Então, $Demand_{avg}$ é atualizado usando $Demand_{avg}$ e o $Demand_{self}$. Depois disso, o C_{demand} é zerado para capturar a demanda de largura de banda do próximo período de posse da ficha.

No entanto, para determinar o número de intervalos de tempo de forma distribuída, o valor do registro, $Demand_{self}$ de cada WI é compartilhado com outras WIs. Para conseguir isso, foi adicionado um campo $Demand$ no *flit* com a ficha, preenchido com o $Demand_{self}$ para compartilhar a demanda de largura de banda prevista da WI na passagem da ficha.

Quando este *flit* com a ficha é transmitido por uma WI e recebido por outras WIs, o valor no campo *Demand* é usado para atualizar o vetor de registradores, REG_{demand} em todas as WIs. Mesmo que um WI não tenha pacotes de dados para transmitir, ele recebe o *flit* com a ficha da WI anterior, atualiza a ficha com sua demanda e passa para a próxima WI.

No D-TDMA, o período de posse da ficha para cada WI é dinamicamente adaptado com base na demanda prevista de largura de banda da WI. Entretanto, diferente do esquema P-TDMA, o número de intervalos de tempo no período de posse da ficha é igual à demanda prevista de largura de banda da WI. Portanto, o período de posse da ficha para o D-TDMA muda sempre que os períodos de posse da ficha são calculados. Os intervalos de tempo alocados no período de posse da ficha de uma WI, i em um período, $j + 1$, é dado por

$$s_i^{TP_{j+1}} = \min(\hat{B}_i^{TP_j}, M) \quad (3.3)$$

onde, $\hat{B}_i^{TP_j}$ é a demanda de largura de banda prevista para a WI i no período de posse da ficha j e M é o número máximo de intervalos de tempo que podem ser alocados para qualquer WI. Esse número máximo de intervalos de tempo garante que nenhuma WI tenha que esperar muitos intervalos de tempo para ter acesso ao meio sem fio. Desta forma, o número de intervalos de tempo (incluindo os *flits* de dados e o *flit* com a ficha) no novo período de posse da ficha TP_{j+1} calculado no final do período de posse da ficha atual TP_j é dado por

$$TP_{j+1} = \sum_{i=1}^N s_i^{TP_{j+1}} + N \quad (3.4)$$

onde, $s_i^{TP_{j+1}}$ é o número de intervalos de tempo alocados no período de posse da ficha para a WI i no período de posse da ficha $j + 1$ e N é o número total de WIs.

Portanto, o número máximo de intervalos de tempo para a contagem regressiva no contador do período de posse da ficha, o C_{TP} , pode ser definido como $N(M + 1)$, controlando assim o seu tamanho. Durante a operação, o valor do novo período de posse da ficha é calculado a partir da Equação 3.4 é usado para redefinir o contador, ao contrário do que acontece no P-TDMA. A unidade preditiva do MAC dinâmico para o D-TDMA contém os mesmos registradores e contadores que o do P-TDMA. A funcionalidade do D-TDMA é igual à do P-TDMA, exceto a lógica de alocação que segue a Equação 3.3. Esse mecanismo MAC dinâmico baseado em TDMA com alocação preditiva de intervalos de tempo superou o mecanismo MAC baseado na passagem da ficha tanto para padrões de tráfego sintéticos como para aplicações específicas.

3.3 CDMA

O uso do CDMA como mecanismo de controle de acesso ao meio em uma WiNoC foi proposto em [VYM⁺14]. Nesta arquitetura, o CDMA é utilizado em uma WiNoC com a topologia *small-world*. A Figura 3.4 apresenta a arquitetura da CDMA-WiNoC.

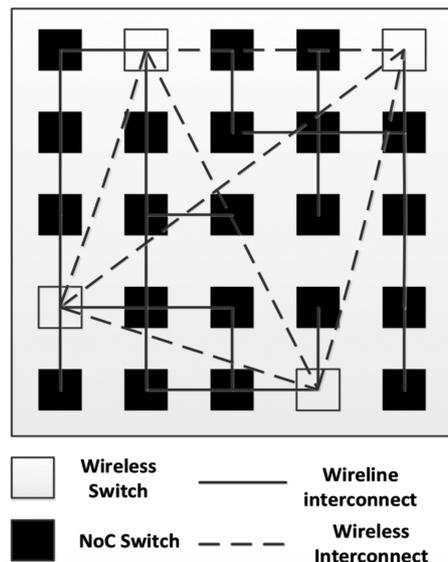


Figura 3.4: Arquitetura da CDMA-WiNoC. Fonte: [VYM⁺14].

Os transceptores utilizados pelo CDMA operam na mesma frequência (57,5 GHz) e conseguem suportar uma taxa de transmissão máxima de 6 Gbps. Embora estejam na mesma frequência, eles podem ser usados simultaneamente sem qualquer controle ou arbitragem centralizada. Isto graças a multiplexação usada no CDMA. Nela, cada transmissor codifica seus bits usando uma palavra-chave (Códigos *Walsh*) única constituída de vários *chips* de código. Cada código é ortogonal aos outros códigos, de modo que a correlação cruzada entre diferentes palavras-chave é zero. Isso elimina a interferência entre transmissões de diferentes WIs usando diferentes palavras-chave.

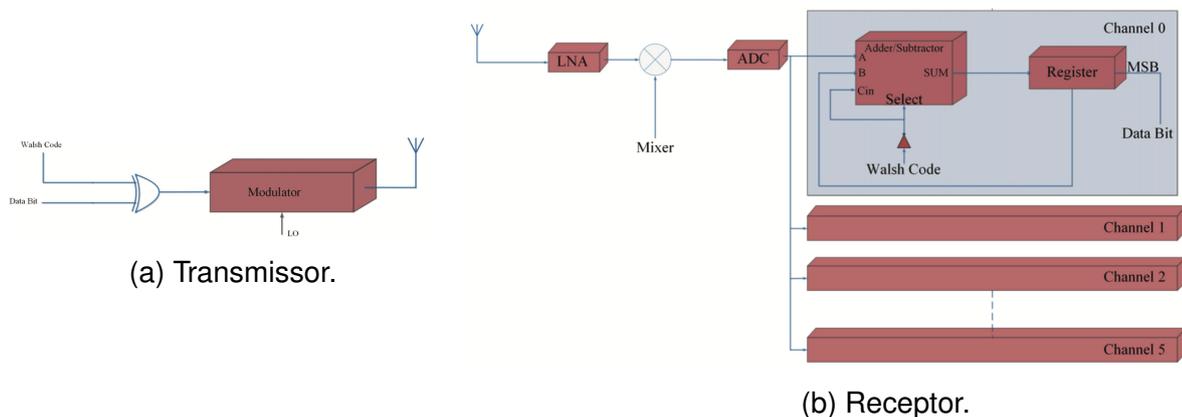


Figura 3.5: Transmissor e receptor da CDMA-WiNoC. Fonte: [VYM⁺14].

Para permitir a criação de múltiplos canais de código ortogonais é adotado o espalhamento espectral por sequência direta, onde o sinal, originalmente modulado por fase, passa por uma segunda etapa de modulação sendo modulado também por código. A codificação pode ser realizada por um simples XOR entre os bits de dados e a palavra-chave. O resultado é então modulado e misturado com a portadora usando o BPSK. A Figura 3.5a mostra o transmissor CDMA.

No lado do receptor é adicionado um decodificador CDMA. Para decodificar digitalmente o sinal CDMA recebido, são necessários códigos *Walsh* ortogonais e balanceados. Os códigos ortogonais garantem que, no caso ideal, quando todos os transmissores estão sincronizados de modo que eles enviam seus bits em simultâneo, a correlação entre diferentes canais de código é zero e os bits transmitidos em outros canais não afetem o bit recebido. O correlator é implementado digitalmente por um acumulador, que adiciona ou subtrai o sinal recebido, dependendo se esse *chip* de código particular do código *Walsh* é alto ou baixo. Os códigos balanceados têm um número igual de *chips* altos e baixos. Conseqüentemente, o sinal do resultado pode ser usado para determinar o bit transmitido. A Figura 3.5b mostra o receptor CDMA. Nos resultados o CDMA conseguiu melhorar significativamente o desempenho e, em simultâneo, reduzir a dissipação de energia, quando comparado com outras arquiteturas de WiNoC utilizando transceptores similares.

3.4 FDMA

Em [LTP+09] é proposto o uso do FDMA em uma arquitetura WCube de multinível. A Figura 3.6b apresenta o segundo nível de uma arquitetura WCube, onde cada $WCube_n$ possui um roteador híbrido com uma WI, o transceptor desta interface, possui um transmissor e $2n$ receptores, onde n é o número de WCube existentes na WiNoC. Desta forma, uma transmissão é recebida $2n$ receptores. Para isso, cada um deles e sua respectiva antena precisa ser ajustado para um diferente canal de frequência. As WIs operam na frequência de subterahertz (1 a 500 GHz).

A Figura 3.6a ilustra o primeiro nível da arquitetura WCube, onde cada roteador híbrido atende a m roteadores e estes, atendem a k PEs, onde $k \geq 1$ e $k \leq 4$. A arquitetura WCube em conjunto com FDMA como mecanismo MAC, conseguiram reduzir entre 20 e 40% a latência média, quando comparado a uma arquitetura de malha concentrada (*Concentrated-mesh* — C-Mesh), 16×16 .

Outra arquitetura que também usa o FDMA, é denominada HiWA (*Hierarchical Wireless-based Architecture*), proposta em [RDSZ16]. As WIs desta arquitetura também operam na faixa de frequência de subterahertz (1 a 500 GHz), onde é possível obter até 16 canais. A Figura 3.7 apresenta a arquitetura HiWA, onde os pontos em verde representam os roteadores híbridos.

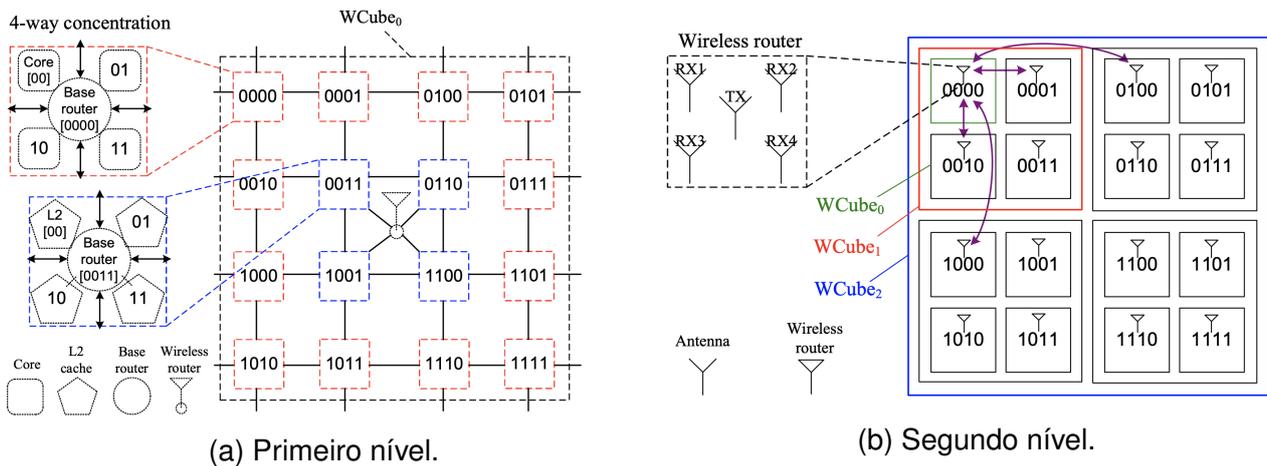


Figura 3.6: Primeiro e segundo níveis de uma WCube. Fonte: [LTP⁺09].

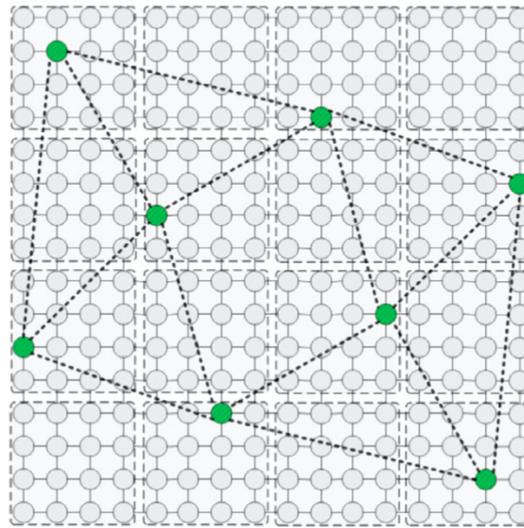


Figura 3.7: Arquitetura HiWA. Fonte: [RDSZ16].

Para minimizar a potência de transmissão entre as WIs, reduzindo a distância entre as WIs, o posicionamento dos roteadores híbridos foi realizado por meio da heurística da otimização de colônia de formigas (*Ant Colony Optimization* — ACO). A ACO é uma abordagem baseada na população para resolver problemas de otimização combinatória inspirada no comportamento de forrageamento das formigas e sua capacidade inerente de encontrar o caminho mais curto de uma fonte de alimento para seu ninho. Esta arquitetura apresentou uma redução 16% na latência e de 14% no consumo de energia em relação a uma NoC convencional.

3.5 CSMA

Em [CL14] é proposto o uso do CSMA p-persistente em uma WiNoC. Como foi explicado na Subseção 2.3.2.1, no CSMA p-persistente o nodo que tem um pacote para ser

enviado precisa escutar o canal. Se o canal estiver ocioso, o nodo sem fio pode transmitir o pacote com probabilidade p ou não transmitir com probabilidade $1 - p$, onde $0 < p \leq 1$.

Se o canal está ocupado o nodo sem fio continua escutando o canal até ele ficar ocioso. Para garantir que a detecção da portadora funcione corretamente, o tempo entre cada vez que o nodo sem fio escuta o canal foi definido como o tempo máximo de propagação τ em segundos. Os períodos de acesso ao canal são categorizados em dois tipos: ocioso I (*idle*) e ocupado B (*busy*). Isso é mostrado na Figura 3.8.

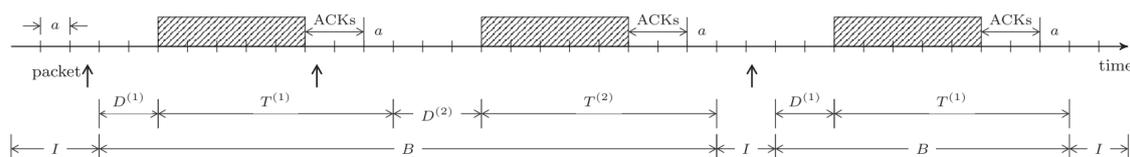


Figura 3.8: Visão geral do CSMA p-persistente. Fonte: [CL14].

Para avaliação do desempenho foram simuladas em alto nível duas arquiteturas de WiNoC, uma com 16 WIs e outra com 64 WIs. O desempenho do mecanismo CSMA p-persistente foi comparado com o mecanismo de passagem da ficha. Nos resultados o CSMA p-persistente atingiu mais rapidamente a taxa de transferência máxima em relação à taxa de injeção de pacotes e a partir deste ponto permaneceu constante, enquanto o mecanismo de passagem da ficha precisou de uma maior taxa de injeção de pacotes para atingir a sua taxa de transferência máxima. A partir deste ponto de saturação, permaneceu constante.

A taxa de transferência máxima apresentada pelo CSMA p-persistente foi significativamente maior que o da passagem da ficha. Quanto a latência, no CSMA p-persistente a latência cresce lentamente com o aumento da carga na rede até estabilizar. Na passagem da ficha o comportamento é semelhante ao do CSMA até o dado nível de carga na rede e a partir deste ponto ele cresce rapidamente.

Em [MAT+16] é proposto o BRS-MAC (*Broadcast, Reliability and Sensing MAC*) como sendo um protocolo intermediário entre o acesso múltiplo com detecção da portadora com prevenção de colisão (*Carrier Sense Multiple Access with Collision Avoidance* — CSMA/CA) e o acesso múltiplo com detecção da portadora com detecção de colisão (*Carrier Sense Multiple Access with Collision Detection* — CSMA/CD) e realiza a detecção da colisão baseada no preâmbulo da mensagem.

O diagrama de fluxo do BRS-MAC é mostrado na Figura 3.9. Neste protocolo, quando uma WI tem dados para enviar, precisa aguardar até que o canal esteja ocioso. O transmissor então envia o preâmbulo e aguarda. As WIs que receberam corretamente esses bits iniciais permanecem em silêncio até o fim da transmissão, enquanto as WIs que detectaram uma colisão começam a transmitir uma confirmação negativa (*Negative Acknowledgment*, NACK). O restante das WIs escuta os sinais de NACK e não podem

iniciar uma nova transmissão de dados. As WIs onde as mensagens colidiram, cancelam suas transmissões e recuam (*backoff*). Na ausência de um sinal NACK, a transmissão do preâmbulo é considerada bem-sucedida e a WI transmite o restante dos dados.

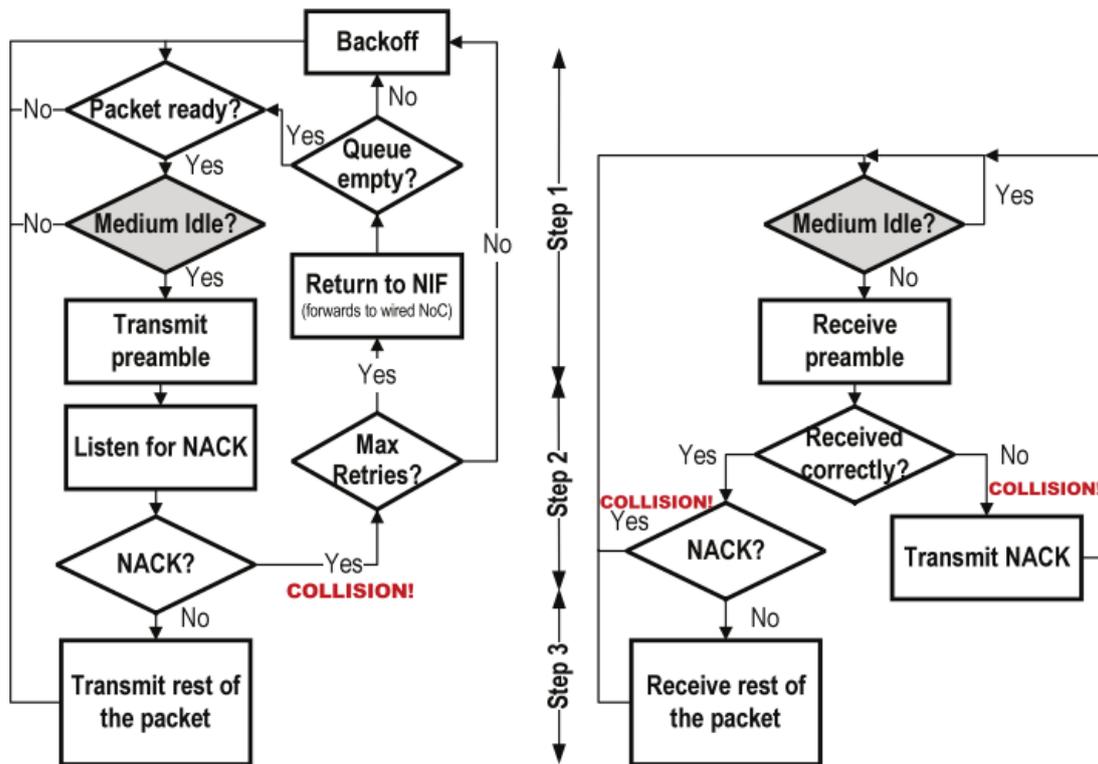


Figura 3.9: Diagrama de fluxo do transmissor (à esquerda) e do receptor (à direita) do BRS-MAC. Fonte: [MAT⁺16].

A detecção de colisão explora a singularidade do cenário em-chip, com seu meio de propagação invariante e conhecido no tempo, e pode empregar três diferentes formas para detectar a colisão: (i) uma verificação de redundância realizada no preâmbulo; (ii) uma comparação entre a amplitude ou fase recebida e o valor esperado dependendo do endereço do nodo de origem, também indicado no preâmbulo; ou (iii) o uso de técnicas de correlação para avaliar a integridade de assinaturas únicas colocadas no preâmbulo. Os NACKs são gerados utilizando o esquema de comunicação coletiva delineado em [OPZ11], onde os NACKs são modelados como a presença ou ausência do sinal. Desta forma, não ocorre colisão, mas sim uma agregação do sinal NACK proveniente de alguns nodos. A presença do sinal é interpretada como um NACK pelo transmissor original, que faz com que ele execute o procedimento de recuo (*backoff*) e tente novamente após aguardar um tempo aleatório. Como estratégia de retransmissão, esse mecanismo utiliza o recuo binário exponencial (*Binary Exponential Backoff* — BEB), onde o tempo de recuo r depende do número de tentativas att como,

$$r = r_0(2^{att} - 1) \quad (3.5)$$

onde r_0 é tempo mínimo de recuo, igual ao tempo médio de transmissão.

Esse protocolo mostrou em simulações utilizando um simulador baseado no Omnet++, um pico de taxa de transferência 27% maior do que os protocolos CSMA convencionais, e uma latência entre uma e duas ordens de grandeza inferiores às NoCs convencionais com carga de tráfego moderada.

Em [MG15] é proposto um mecanismo de MAC dinâmico que usa CSMA para baixas cargas de tráfego e passagem da ficha para altas cargas de tráfego. Esse protocolo MAC permite que as WIs acessem o meio sem fio formando interligações sob demanda e criando uma topologia reconfigurável. A arquitetura dessa WiNoC reconfigurável é mostrada na Figura 3.10.

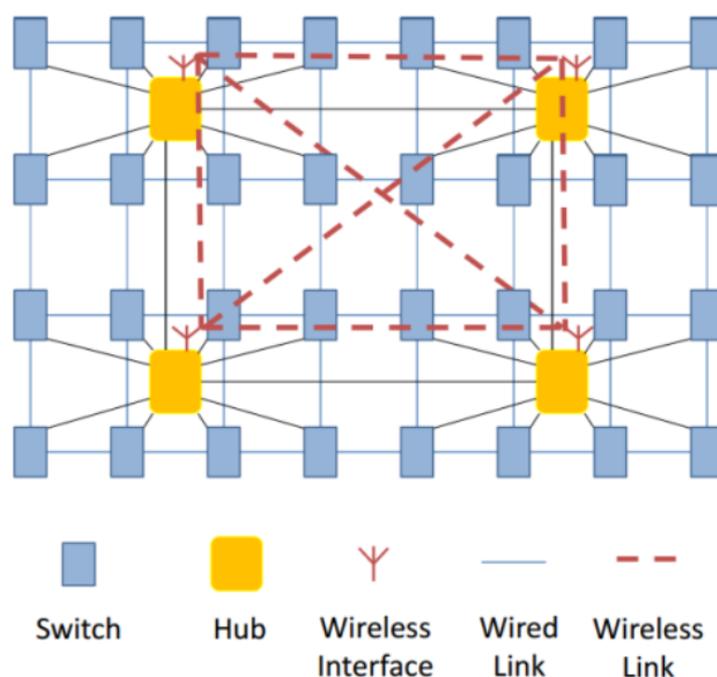


Figura 3.10: Arquitetura da WiNoC reconfigurável sob demanda. Fonte: [MG15].

Para habilitar este mecanismo, cada WI é equipada com uma unidade MAC dinâmica. O MAC dinâmico funciona como um MAC de passagem da ficha quando a utilização sem fio é maior que um limite predefinido. No modo de passagem da ficha o acesso do meio sem fio é concedido ao possuidor da ficha, que circula entre as WIs.

Uma WI pode transmitir dados apenas se possuir a ficha. A circulação da ficha garante o acesso justo ao meio sem fio sendo mantida usando o endereço da WI que supostamente deveria receber a ficha seguida do endereço da WI que a liberou. O endereço da próxima WI é necessário para distinguir a WI pretendida de outras WIs, pois as antenas em-chip adotadas não são direcionais e todas as WIs recebem a ficha. Além disso, o endereço da WI que liberou a ficha é necessário em todas as WIs para atualizar a métrica de utilização na matriz de cada WI sendo utilizada para monitorar a utilização das WIs.

A ficha é liberada para a próxima WI quando a WI atual não possui mais dados para transmitir ou transmitiu um pacote inteiro. Depois que a ficha é recebida com sucesso

na próxima WI, ela verifica os *buffers* e se tiver *flits* para enviar, ela faz um *broadcast* de um *transmit flit* com o endereço da WI de destino e o número de *flits* que deseja enviar para essa WI. O número de *flits* indica a duração da transmissão para as outras WIs. Assim, ao receber o *transmit flit*, todas as WIs, menos a WI de destino, entram em estado de suspensão e esperam até o fim da transmissão para sair desse estado. Consequentemente, o consumo de energia no modo de passagem da ficha é reduzido apenas ao consumo de envio e recebimento.

O MAC dinâmico opera no modo CSMA quando a utilização da rede sem fio é baixa. Neste modo as WIs sentem o canal para verificar se existe alguma transmissão em curso. Quando o canal está livre e a WI tem dados para enviar, a WI envia um pequeno *query flit*, apenas com o endereço da WI destino ao qual deseja enviar os dados. Assim que o *query flit* chega na WI desejada, a WI copia o conteúdo da *query flit* e faz um *broadcast* de um *reply flit* contendo este endereço. Assim que a WI de origem recebe de volta esse *reply flit*, ele cria um *transmit flit* com o número de *flits* que serão enviados para a WI de destino e faz um *broadcast* desse *transmit flit*. Tanto o *query flit* como o *transmit flit* são recebidos por todas as WIs, mas apenas a WI desejada manterá o seu receptor ativo durante o período mencionado *transmit flit*. As outras WIs entram em estado de suspensão e após o término desse período elas saem desse estado. Isso ajuda a manter a eficiência energética.

A colisão é detectada se ocorre uma falha em receber o *reply flit* na WI do transmissor de origem dentro de um tempo específico após o *query flit* ter sido enviado. No caso de colisão, as WI que colidiram recuam e deixam de transmitir no canal sem fio por um tempo B_T dado por,

$$B_T = r \times n_B \times t_{flit} \quad (3.6)$$

onde, n_B é o número de *flits* que uma WI precisa esperar antes de tentar acessar o meio sem fio novamente, t_{flit} é o tempo para transmitir um *flit* e r é um inteiro aleatório gerado por um Registrador de Deslocamento com Realimentação Linear (*Linear Feedback Shift Register* — LFSR) dentro de um intervalo de 0 a R . Aqui R é o número máximo de colisões que qualquer WI pode ter. Este intervalo é escolhido de modo que seja suficientemente grande para garantir que o mecanismo MAC mude de CSMA para passagem da ficha antes de incrementar muito o número de colisões. Esse recuo, reduz a probabilidade de ocorrer uma nova colisão.

Quando uma WI entra no estado de recuo ela só pode receber *flits*. No período de recuo, se uma WI receber um *query flit* e um *transmit flit* então ela entra em um estado de suspensão e após o término do período descrito pelo *transmit flit* ela sai desse estado.

A unidade MAC dinâmica permite a mudança entre o modo de passagem da ficha e o modo CSMA. Esta mudança é baseada no monitoramento da métrica de utilização das WIs. A utilização de uma WI é medida como o número de *flits* nos *buffers* da WI. Por isso a utilização da WI é uma indicação da carga de tráfego instantânea na WI. Se esta carga

for superior a um limiar superior, o esquema MAC muda de CSMA para passagem da ficha. Quando a utilização é inferior a um limiar inferior, o MAC muda de passagem da ficha para o modo CSMA.

Os dois limiares separados são usados para evitar a troca frequente entre os dois modos sendo determinados experimentalmente. A arquitetura da unidade MAC dinâmica é mostrada na Figura 3.11. A unidade MAC dinâmica contém um contador de *back-off* ($B_{counter}$) e um contador de *wake-up* ($W_{counter}$). Existem cinco registradores de estado. O registrador de estado CM_s é usado para armazenar o modo operacional atual do MAC dinâmico, T é usado como um indicador de posse da ficha quando CM_s está configurado para modo de passagem da ficha, ID_{self} e ID_{next} são registradores utilizados para armazenar o ID da própria WI e da próxima WI na ordem de circulação da ficha por *Round-robin*. Um registrador $UTILIZATION_{self}$ armazena a utilização da própria WI. Uma série de registradores $UTILIZATION_ARRAY$ são usados para armazenar a utilização das outras WIs. A unidade de comutação (SU) é a unidade de controle que decide alternar entre os diferentes modos operacionais com base na utilização das WIs.

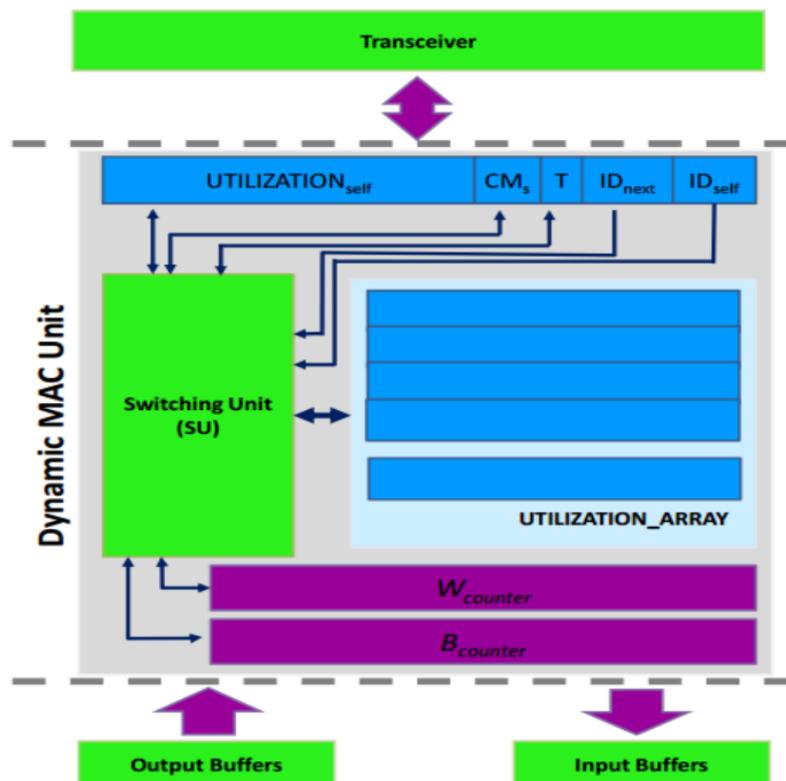


Figura 3.11: Arquitetura da unidade de MAC dinâmica. Fonte: [MG15].

As WIs dessa arquitetura utilizam a antena no formato zigue-zague, que fornece uma largura de banda de 3 dB a 16 GHz, com uma frequência central ao redor de 60 GHz com um alcance de 20 mm. O transceptor utilizado pela WI utiliza a modulação OOK não-

coerente. O *wormhole* é utilizado para comutação e o algoritmo XY para o roteamento de pacotes.

Esse mecanismo de MAC dinâmico apresentou uma redução de 1,5 vezes no número de saltos necessários para o pacote chegar ao seu destino em comparação a uma NoC em malha convencional. Também foi observada uma redução de 8% no consumo de energia e um aumento de 3,4% na largura de banda de pico em comparação com uma arquitetura com conexões sem fio dedicadas.

3.6 Discussão

A Tabela 3.1 relaciona algumas das características dos mecanismos de controle de acesso ao meio para WiNoCs detalhados nesse capítulo. Nessa tabela são relacionados, o mecanismo de controle de acesso ao meio, a complexidade do transceptor, a categoria de antena, a escalabilidade, a latência, a taxa de transferência e o consumo de energia desses mecanismos. Essa tabela também relaciona as mesmas características do mecanismo proposto por este trabalho.

Tabela 3.1: Características dos mecanismos de controle de acesso ao meio.

	MAC	Transceptor	Antena	Escalável	Latência	Taxa de transferência	Consumo de energia
[VYM ⁺ 14]	CDMA	Complexo	Omnidirecional	Não	Baixa	Alta	Alto
[CDG ⁺ 12] [DKKM ⁺ 11] [MLW ⁺ 14] [MKW ⁺ 14] [DCY ⁺ 13] [WKM ⁺ 14] [WMK ⁺ 14] [DKM ⁺ 15] [DGC ⁺ 10]	Passagem da ficha	Simplex	Omnidirecional	Não	Alta	Moderada	Moderado
[PCMC ⁺ 15]	Passagem da ficha (Aprimorada)	Simplex	Omnidirecional	Não	Moderada	Moderada	Alto
[CL ⁺ 14]	CSMA p-persistente	Simplex	Omnidirecional	Sim	Baixa	Baixa	N/D
[MAT ⁺ 16]	CSMA (baseado)	Complexo	Omnidirecional	Não	Baixa	Baixa	Moderado
[LTP ⁺ 09]	FDMA	Complexo	Omnidirecional	Não	Baixa	Alta	Alto
[RS ⁺ DT14]	FDMA	Complexo	Omnidirecional	Não	Baixa	Alta	Alto
[MG ⁺ 15]	CSMA e passagem da ficha	Simplex	Omnidirecional	Sim	Baixa	Baixa	Moderado
[MSG ⁺ 16]	TDMA	Simplex	Omnidirecional	Não	Alta	Moderada	Moderado
Este trabalho	Slotted CSMA não-persistente	Simplex	Direcional	Sim	Moderada	Moderada	Baixo

A maior parte das publicações referentes a mecanismos MAC para WiNoCs encontrados na bibliografia utilizam o acesso controlado por passagem da ficha para oferecer

acesso múltiplo. Entretanto, esse mecanismo não escala com o crescimento do número de WIs [DGP⁺12]. Ele também não se demonstra energeticamente eficiente quando está sujeito a uma baixa carga de tráfego. Mesmo sem demanda de tráfego nas WIs, a ficha continua tendo que circular entre elas. Entretanto, com o canal ocioso, a ficha circula em intervalos menores, visto que não ocorre o atraso relativo à transmissão de um ou mais pacotes entre a transmissão da ficha. Por consequência, isso gera uma sobrecarga no consumo de energia quando a WiNoC está sujeita a essa categoria de tráfego [MG15].

Outras publicações oferecem acesso múltiplo e simultâneo ao meio sem fio através de mecanismos de canalização, como o TDMA, FDMA e CDMA. Esses mecanismos oferecem uma alta taxa de transferência, mas tornam-se impraticáveis em redes de alta densidade, dado que cada canal adicional aumenta a complexidade do transceptor ou reduz o desempenho geral. Além disso, esses mecanismos são inerentemente rígidos e podem não funcionar bem com tráfego altamente variável [MAT⁺16][DGP⁺12].

Poucos trabalhos exploraram o uso do acesso múltiplo através de mecanismos de acesso aleatório, que podem fornecer baixa latência quando a carga é moderada-baixa e se adapta às mudanças no tráfego [MAT⁺16]. Entretanto, o CSMA sofre uma degradação no desempenho para cargas de tráfego mais altas, como foi observado em [MG15], que contorna esse problema utilizando um mecanismo híbrido que utiliza CSMA, quando a carga na rede é baixa, e passagem da ficha, quando a carga na rede é alta. De qualquer forma, este mecanismo híbrido ainda estaria sujeito aos problemas de escalabilidade presentes no mecanismo de passagem da ficha.

A utilização de antenas direcionais, se aproveitando das regiões formadas da união dos lóbulos principais das antenas do transmissor-receptor nas DWiNoC trazem, de certa forma, o benefício dos mecanismos MAC canalizados. Visto que permitem a criação de canais independentes a partir dessas regiões, onde consegue fornecer acesso múltiplo, internamente em cada canal, e simultâneo, considerando as diversas regiões que podem ser formadas sob a área da DWiNoC. Entretanto, como o mecanismo MAC utilizado internamente nesses canais é o de passagem da ficha, logo ele sofre com a degradação na eficiência energética quando o canal está submetido a uma baixa carga de tráfego. Considerando a alta variabilidade no tráfego da WiNoC esse problema pode influenciar de forma significativa a eficiência energética da DWiNoC.

Como pode ser observado na Tabela 3.1, o uso de um transceptor complexo geralmente reflete em um consumo de energia alto. Os mecanismos baseados na canalização são os únicos que apresentam taxa de transferência e latência baixas, entretanto a custo da escalabilidade e de um consumo de energia alto. Boa parte dos mecanismos de passagem da ficha apresentam consumo de energia alto e baixa latência a um custo de energia moderado. Todos os trabalhos exploraram apenas o uso de antenas omnidirecionais. O mecanismo proposto nessa tese estabelece um meio-termo entre desempenho e o consumo

de energia através das oportunidades oferecidas pelas antenas direcionais, aliado ao uso do *slotted* CSMA não-persistente.

4. DESCRIÇÃO DA ARQUITETURA

Nesse capítulo é apresentado os detalhes da arquitetura para o uso do CSMA como MAC em uma DWiNoC em substituição ao mecanismo de passagem da ficha, visando reduzir o consumo de energia quando a DWiNoC está submetida a baixas cargas de tráfego. Para facilitar o entendimento, na Seção 4.1 é apresentada a arquitetura da DWiNoC e na Seção 4.2 são apresentados alguns detalhes do uso do *Slotted* CSMA não-persistente no ambiente em-chip.

4.1 Arquitetura da DWiNoC

Nessa seção são abordados os conceitos específicos de uma DWiNoC, que apresentam diferença em relação a uma WiNoC e não foram abordados no Capítulo 2.

4.1.1 Topologia

A topologia da DWiNoC segue os princípios de uma rede pequeno-mundo [OM06], caracterizada por muitas ligações de curta distância e algumas ligações de longa distância, que permitem que nodos distantes sejam alcançados com menos saltos. Nessa topologia, mostrada na Figura 4.1 as ligações com fios são estabelecidas seguindo a distribuição da inversa da função de potência [MGS⁺17].

$$P(i, j) = \frac{d_{ij}^{-\alpha} f_{ij}}{\sum_{\forall i} \sum_{\forall j} d_{ij}^{-\alpha} f_{ij}} \quad (4.1)$$

onde, $P(i, j)$ é a probabilidade do estabelecimento de uma ligação entre o roteador i e j . D_{ij} é a distância de Manhattan entre o roteador i e j . f_{ij} é a frequência de comunicação entre os roteadores i e j .

4.1.2 Posicionamento das WIs

O posicionamento das WIs é uma etapa fundamental para que a DWiNoC consiga reduzir a latência e o consumo de energia através de ligações sem fio, múltiplas e concorrentes, mas livre de interferências. Para isso, é necessário determinar para quais pares de roteadores seria relevante a inserção de WIs para uma dada topologia e padrão de tráfego,

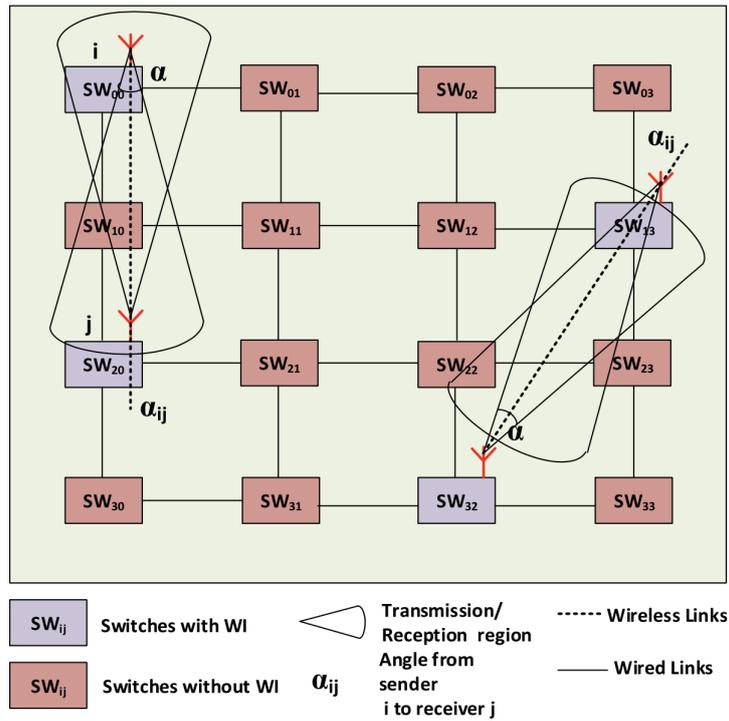


Figura 4.1: Topologia da DWiNoC. Fonte: [MGS⁺17].

sem desrespeitar as restrições de posicionamento das WIs. O posicionamento das WIs na DWiNoC pode ser definido como um problema de otimização [MGS⁺17], definido por

$$\min_W \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N h_{ij} f_{ij} \quad h_{ij} = K d_{ij-WI} + (1-K) d_{ij} \quad (4.2)$$

onde, W é uma matriz que contém os pares de WI. N é o número total de nodos em uma topologia mesh. f_{ij} é a frequência de comunicação entre os nodos i e j . h_{ij} é o número de saltos entre os nodos i e j . K é uma variável lógica que indica a presença “1” ou a ausência “0” de uma ligação sem fio ao longo do caminho.

A primeira restrição que precisa ser respeitada diz respeito ao número máximo de WIs na DWiNoC. Cada WI adicionada implica em um aumento na área e no consumo de energia. Esta restrição estabelece um limite tolerável para sobrecarga de área e energia [MGS⁺17] e pode ser definida como

$$2\text{Rows}(W) \leq N_{WI} \quad (4.3)$$

onde, N_{WI} é o número máximo de WIs permitidas onde a sobrecarga de área e energia fique dentro de um limite tolerável.

Como as antenas usadas no DWiNoC são direcionais e estão alinhadas, conforme os lóbulos principais do padrão de radiação de seus pares. A topologia DWiNoC torna possível múltiplas comunicações sem fio simultâneas na mesma frequência, permitindo

que cada WI seja capaz de se comunicar apenas com uma WI. Isso é imposto restringindo cada WI a apenas um par enquanto se otimiza o posicionamento das WIs [MGS⁺17]. Essa restrição é representada por

$$\text{unique}(W) = W \quad (4.4)$$

Como as ligações sem fios são mais eficientes quando utilizadas na comunicação de longo alcance, a terceira restrição estabelece uma distância mínima entre um par de WIs, a partir da qual se obteria uma redução no consumo de energia em relação a uma ligação com fio [MGS⁺17]. Esta restrição pode ser representada por

$$\left| W_X * \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right| + \left| W_Y * \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right| \geq D_{th} * \mathbf{1} \quad (4.5)$$

onde, W_X e W_Y são matrizes representando as coordenadas XY dos roteadores em uma rede em malha. $\mathbf{1}$ é um vetor de uns com o mesmo tamanho do número de linhas de W . D_{th} é o limiar da distância mínima da distância de Manhattan entre os pares de nodos. Essa distância mínima geralmente varia entre 7 e 9 mm [DGP⁺12] [GD15].

Para garantir que a comunicação concorrente das ligações sem fio não interfira uma na outra e a comunicação possa ocorrer de forma confiável é necessário respeitar algumas restrições de posicionamento das WIs para evitar interferências [MGS⁺17]. Estas restrições podem ser representadas por

$$\begin{aligned} \mathbf{A}(W^v) &= \mathbf{0} \\ W^v &= W * \begin{bmatrix} N & 1 & N+1 & 0 \\ 1 & N & 0 & N \end{bmatrix} - N * \mathbf{1}_{rows(W) \times 4} \end{aligned} \quad (4.6)$$

onde, \mathbf{A} é uma matriz $N \times N$ que representa as restrições para evitar a interferência entre quaisquer duas interligações sem fio. W^v converte os valores da matriz W para posição apropriada.

4.1.2.1 Restrições para evitar interferência

As restrições no posicionamento para evitar interferência são baseadas nas propriedades direcionais e nas características de transmissão da antena PLPA, como ganho, largura de feixe, taxa de erro em bits, perdas no caminho e da razão sinal-interferência (*Signal-to-interference Ratio* — SIR). Estas informações são obtidas a partir do padrão de radiação da antena PLPA, da variação da perda no caminho em função da distância, da potência recebida e da potência interferente recebida de diferentes direções [MGS⁺17].

Cada roteador híbrido, que possui uma WI, pode ser representado como um ponto $N_1(x_1, y_1)$ no plano XY e cada ligação sem fio pode ser representada como um segmento de

reta conectando dois pontos $\overline{N_1 N_2}$. Para um dado padrão de radiação de uma antena PLPA, temos θ_{ML} e θ_{SL} como a largura dos lóbulos principal e lateral, respectivamente. Assumindo que a potência de transmissão de cada WI emprega apenas a potência suficiente para alcançar a outra antena em que está emparelhada. A partir das larguras dos lóbulos e sabendo que a potência cai significativamente com a distância, quatro regiões triangulares podem ser formadas das duas antenas de uma ligação sem fio [MGS⁺17]. Estas regiões são ilustradas na Figura 4.1.

A interferência em qualquer transmissão sem fio na DWiNoC pode ser evitada se

1. Dois segmentos de ligação sem fio, que não se sobreponham, não se interseccionarem em nenhum ponto.
2. A antena de uma WI, que não pertença à ligação sem fio do segmento $\overline{N_1 N_2}$, mas que esteja nas regiões triangulares, precisa estar posicionada na mesma direção que as antenas deste segmento. Além disso, todas as WIs cobertas pelas regiões triangulares precisam utilizar o mecanismo de controle de acesso ao meio proposto neste trabalho.
3. Fora a condição citada no item anterior, nenhuma WI que não pertença à ligação sem fio de um determinado segmento pode estar nas regiões triangulares.

4.1.2.2 Algoritmo de otimização

O diagrama de fluxo do algoritmo de otimização proposto em [MGS⁺17] é mostrado na Figura 4.2. Ele inicialmente verifica qual é a média de saltos de uma dada NoC com topologia em malha. A partir disto, ele considera cada roteador como candidato a se tornar um roteador híbrido, recebendo uma WI, e os adiciona a uma lista de roteadores candidatos a ter uma WI. Então, a cada iteração, ele forma pares com os roteadores desta lista e os adiciona a NoC, avaliando o impacto no número médio de saltos. Ao fim da avaliação, ele retorna o par de WIs que foi mais significativo na redução do número de saltos e então verifica se ele obedece à todas as restrições. Após passar por esta verificação o par de WIs é então adicionado à lista de solução, caso contrário ele é adicionado a lista de pares inelegíveis e removido de futuras iterações. Este processo segue se repetindo até:

1. Atingir o número de máximo de WIs estabelecido pela Equação 4.3.
2. A Solução da iteração atual retornar a mesma solução da iteração anterior.
3. Não existir mais pares disponíveis na lista de pares de WIs elegíveis.

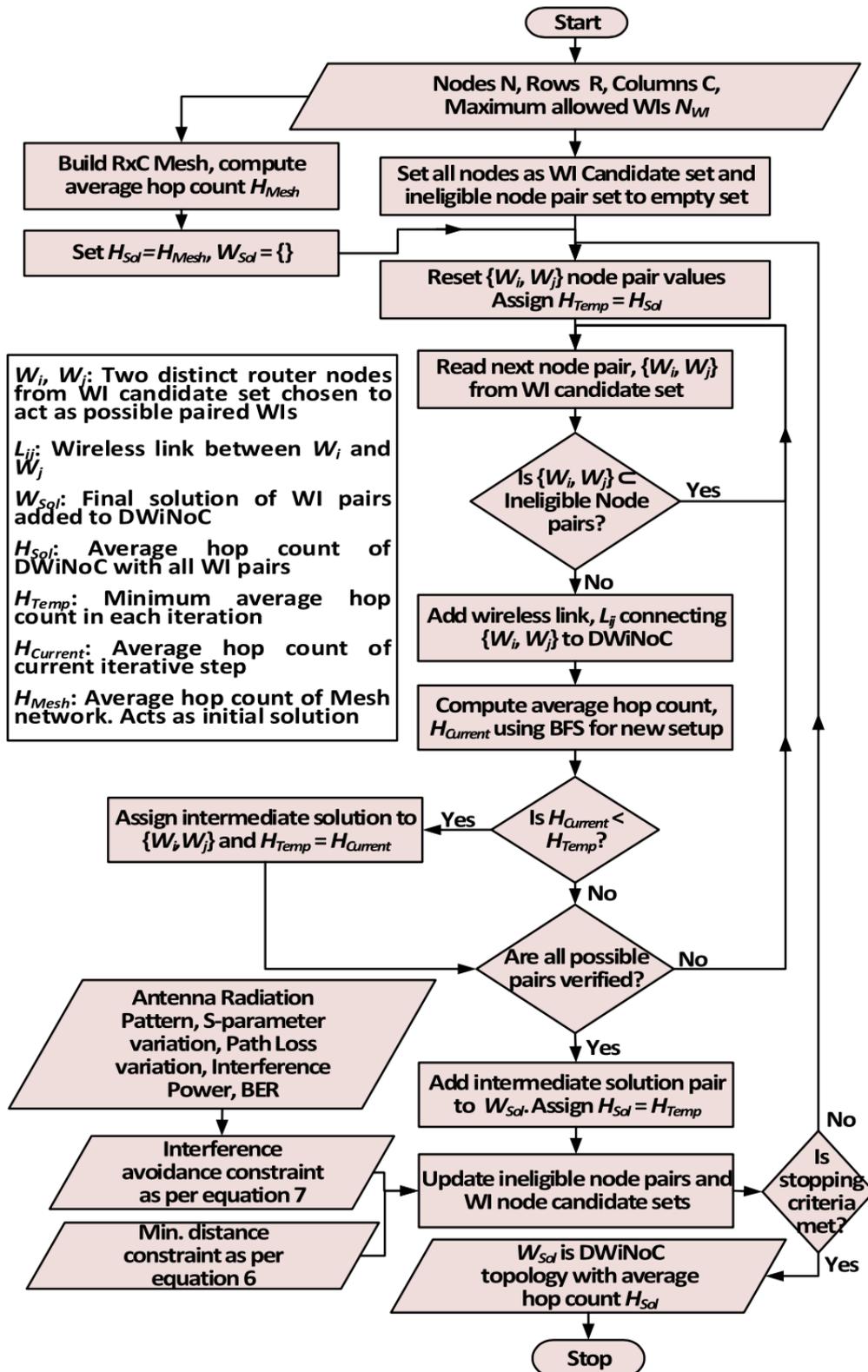


Figura 4.2: Algoritmo de posicionamento das WIs. Fonte: [MGS⁺17].

4.2 ***Slotted CSMA não-persistente***

Primeiramente é necessário justificar o motivo pelo qual foi escolhido utilizar a modo não-persistente do CSMA ao invés dos outros modos de persistência. O CSMA não-persistente fica situado no meio-termo entre desempenho e complexidade de implementação em chip. Os detalhes do funcionamento do CSMA, bem como o detalhamento dos seus métodos de persistência foram apresentados na Subseção 2.3.2.1.

No contexto de uma DWiNoC, seu mecanismo de funcionamento é composto basicamente de três partes: a detecção da portadora, realizada indiretamente pelo próprio circuito do receptor; o gerador de números pseudoaleatórios, que pode ser implementado facilmente utilizando LFSRs; um contador decrescente e um detector de borda de clock.

5. RESULTADOS

Nesse capítulo são apresentados os modelos utilizados para avaliação do desempenho, bem como os resultados da avaliação do desempenho do CSMA como mecanismo MAC para DWiNoCs. Os modelos analíticos, tanto do CSMA quanto do mecanismo de passagem da ficha, utilizado para comparação dos resultados, são apresentados na Seção 5.1. A análise nos resultados do consumo de energia é apresentada na Seção 5.2. Por fim, a análise dos resultados da taxa de transferência e da latência são apresentados nas seções 5.3 e 5.4, respectivamente.

5.1 Modelo Analítico

Inicialmente, na Seção 5.1.1 são apresentados os modelos analíticos utilizados no CSMA não-persistente e no *slotted* CSMA não-persistente e como, a partir deles, são inferidos a taxa de transferência e a latência. Da mesma forma, para realizar a comparação do desempenho e avaliar o impacto da circulação da ficha no consumo de energia, o modelo analítico do mecanismo de passagem da ficha — mecanismo MAC no estado-da-arte para DWiNoCs — é apresentado na Seção 5.1.2. Todos os modelos que serão apresentados presumem o modelo apenas do tráfego no canal. O modelo para o todo o tráfego da WiNoC foge ao escopo desse trabalho.

5.1.1 CSMA

O modelo de tráfego utilizado para análise do CSMA é o proposto em [KT75], que presume um número infinito de WIs, que coletiva e independentemente formam uma distribuição de Poisson, com uma média agregada de geração de pacotes de λ/s .

Essa é uma aproximação para uma população grande, porém finita onde cada WI gera pacotes com pouca frequência e cada pacote pode ser transmitido com sucesso dentro de um intervalo de tempo muito inferior ao tempo médio entre pacotes sucessivos gerados por uma determinada WI. Assume-se que cada WI na população infinita tem no máximo um pacote, e esse pode requerer ser transmitido a qualquer tempo.

Cada pacote tem tamanho constante, exigindo T segundos para ser transmitido. $S = \lambda T$ é o número médio de pacotes gerados por tempo de transmissão, ou seja, é a taxa de entrada normalizada em relação a T . Em condições de estado estável, S também pode ser referido como a taxa de transferência do canal ou utilização do canal. Devido ao problema de interferência inerente à natureza aleatória desse mecanismo, a taxa de trans-

ferência alcançável será sempre inferior a 1. A taxa de transferência máxima alcançável para um modo de acesso chama-se capacidade do canal.

Esse modelo considera que se dentro de algum prazo especificado, após a transmissão de um pacote, se uma WI não receber um reconhecimento (*Acknowledgement* — ACK), ela sabe que ocorreu uma colisão. Cada WI atrasa a transmissão de um pacote previamente colidido por algum tempo aleatório, cuja média é \bar{X} , onde um número é escolhido uniformemente entre 0 e $X_{max} = 2\bar{X}$. O tráfego oferecido ao canal a partir de um grupo de WIs consiste não só em novos pacotes, mas também em pacotes previamente colididos, isto aumenta a taxa média de tráfego oferecido G onde $G \geq S$. A capacidade de canal é encontrada pela razão S/G e denota a capacidade de uma transmissão bem-sucedida e G/S é o número médio de vezes que um pacote tem de ser retransmitido (reescalonado) até o sucesso.

O modelo ainda presume que o atraso médio de retransmissão é grande em comparação com T e o tempo entre as chegadas do processo pontual¹, definido pelos tempos de início de todos os pacotes mais as retransmissões, são independentes e exponencialmente distribuídos. Quando uma WI fica sabendo que sua transmissão não foi bem-sucedida, ele reescala a transmissão do pacote de acordo com um atraso de retransmissão distribuído aleatoriamente. Neste novo momento, o transmissor detecta o canal e repete o algoritmo ditado pelo protocolo.

Uma WI pode, a qualquer momento, estar transmitindo ou recebendo (mas não ambos simultaneamente). Entretanto, o atraso incorrido para mudar de um modo para o outro é insignificante. Além disso, o tempo necessário para detectar a portadora é insignificante. Todos os pacotes são de comprimento constante sendo transmitidos por um suposto canal ideal (sem ruído). O modelo assume que uma sobreposição de qualquer fração de dois pacotes resulta em interferência destrutiva e ambos os pacotes devem ser retransmitidos e que o atraso de propagação é idêntico para todos os pares origem-destino.

Apenas retomando o funcionamento do algoritmo do CSMA não-persistente (mostrado na Subseção 2.3.2.1), uma WI pronta para transmitir um pacote detecta o canal e opera da seguinte forma: se o canal estiver ocioso, ela transmite; se ele estiver ocupado, então a WI reescala a retransmissão do pacote para algum momento posterior conforme a distribuição do atraso de retransmissão. Neste novo momento, a WI detecta o canal novamente e repete o algoritmo descrito. O modo *slotted* do CSMA não-persistente também é considerado pelo modelo. Nela, cada *slot* tem um tempo igual ao atraso de propagação τ . Todas as WIs são sincronizadas sendo forçadas a iniciar a transmissão somente no início de um *slot*.

¹Processos pontuais são definidos como um conjunto de pontos irregularmente distribuídos gerados por um mecanismo estocástico.

5.1.1.1 Modelo para taxa de transferência

Segundo [KT75], a equação básica para a taxa de transferência S do CSMA não-persistente pode ser escrita como

$$S = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}} \quad (5.1)$$

onde, $a = \tau/T$ é a razão entre o atraso de propagação e o tempo de transmissão do pacote e G é a taxa de tráfego oferecida ao canal.

Da mesma forma, para o modo *slotted* do CSMA não-persistente temos

$$S = \frac{aGe^{-aG}}{(1 - e^{-aG}) + a} \quad (5.2)$$

Tanto o CSMA não-persistente quanto para o *slotted* CSMA não persistente é válida o seguinte

$$\lim_{a \rightarrow 0} = \frac{G}{(1 + G)} \quad (5.3)$$

Isso mostra que quando $a = 0$, a taxa de transferência de 1 pode teoricamente ser obtida para um tráfego infinito oferecido ao canal.

5.1.1.2 Modelo para latência

Nesse modelo, a latência D esperada do pacote é definido como o tempo médio desde quando um pacote é gerado até que seja recebido com sucesso. Esse modelo presume que os pacotes de reconhecimento (ACKs) são sempre recebidos corretamente com probabilidade um e o tempo de processamento necessário para executar a verificação da soma (*checksum*) e para gerar o pacote de reconhecimento é insignificante [KT75].

Considerando todas as chegadas de pacotes de maneira uniforme, podemos escrever a latência como

$$D = (G/S - 1)(1 + 2a + \alpha + \delta) + 1 + a \quad (5.4)$$

onde, G/S é dado pela equação da taxa de transferência do CSMA não persistente, Equação 5.1. $\alpha = \tau/T$ e $\delta = \bar{X}/T$ são o tempo de propagação e o atraso médio de retransmissão normalizados em relação ao tempo de transmissão de um pacote, respectivamente.

5.1.2 Passagem da Ficha

Assim como no modelo para o CSMA, o modelo para o mecanismo de passagem da ficha presume, que para cada WI, os processos de chegada de mensagens são proces-

dos Poisson com taxa média de chegada de pacotes λ e o tamanho do pacote de dados \bar{X} . Todas as filas são infinitas em tamanho. Quando uma WI tem a ficha, ela transmite todas as mensagens na fila (serviço exaustivo) [HO86].

5.1.2.1 Modelo para taxa de transferência

A taxa de transferência para o mecanismo de passagem da ficha, conforme é mostrado em [HO86], pode ser obtida a partir de

$$S = \frac{M\lambda\bar{X}}{R} \quad (5.5)$$

onde M é o número de WIs que compartilham o canal, λ é a taxa média de chegada de pacotes, \bar{X} é o tamanho do pacote de dados e R é a taxa de transferência do canal em bits/s.

Entretanto, a taxa máxima de transferência é afetada pelo tempo que cada WI permanece com a posse ficha (*Token Holding Time* — THT), essa relação é expressa pela Equação 5.6 [HO86].

$$S_{max} = \frac{T_h}{T_h + T_t + \frac{\tau}{3}} \quad (5.6)$$

onde T_h é o THT, T_t é o tempo de transmissão do pacote com a ficha e τ é o tempo de propagação fim-a-fim.

5.1.2.2 Modelo para latência

Conforme [HO86], a latência média do mecanismo de passagem da ficha é dada por

$$T = \frac{\bar{X}}{R} + \frac{\tau}{3} + \frac{MX_t \left(1 - \frac{S}{M}\right)}{2R(1-S)} + \frac{\tau(M-S)}{6(1-S)} + \frac{S\bar{X}}{2R(1-S)} \quad (5.7)$$

onde, \bar{X} é o tamanho do pacote de dados, R é a taxa de transferência do canal em bits/s, M é o número de WIs que compartilham o canal, τ é o tempo de propagação fim-a-fim, X_t é o tamanho do pacote com a ficha e S é a taxa de transferência.

5.1.2.3 Modelo para o consumo de energia

Podemos obter o consumo de energia para os *flits* com dados multiplicando a taxa de transferência S pela taxa de transferência do canal R e pela energia necessária para

transmitir 1 bit e_{bit} , como é mostrado na Equação 5.8.

$$E = SRe_{bit} \quad (5.8)$$

O consumo de energia dos *flits* com a ficha pode ser obtido a partir da soma do consumo de energia quando o canal está ocupado E_{busy} , nesse caso o THT é respeitado e a WI permanece com a ficha durante o período do THT, e da soma do consumo de energia quando o canal está ocioso E_{idle} , nessa caso o THT não é respeitado e a ficha circula livremente entre as WIs.

Quando o canal está ocupado, o consumo de energia dos *flits* com a ficha pode ser obtido a partir do número de *flits* com a ficha enviados em 1 s multiplicado pelo tamanho do pacote do *flit* com a ficha X_t e pela energia consumida por bit e_{bit} , conforme é descrito pela equação 5.9.

$$E_{busy} = \left(\frac{1}{T_h + T_t + \frac{\tau}{3}} \right) * X_t * e_{bit} \quad (5.9)$$

onde T_h é o THT, T_t é o tempo de transmissão do pacote com a ficha e τ é o tempo de propagação fim-a-fim.

Para determinar o consumo de energia dos *flits* com a ficha quando o canal está ocioso primeiramente temos que determinar por quanto tempo o canal fica ocioso. Isso pode ser obtido a partir da Equação 5.10

$$T_{idle} = \frac{(R - (S * R) - (1 - S_{max}) * R)}{R} \quad (5.10)$$

onde R é a taxa de transferência do canal, S é a taxa de transferência e S_{max} é taxa de transferência máxima limitada pelo THT.

A partir do tempo que o canal fica ocioso T_{idle} é possível determinar o número de *flits* com a ficha enviados e a partir desse, multiplicado pelo tamanho do pacote do *flit* com a ficha X_t e pela energia consumida por bit e_{bit} é possível determinar o consumo de energia dos *flits* com a ficha quando o canal está ocioso, conforme é descrito pela equação 5.11.

$$E_{idle} = \left(\frac{T_{idle}}{T_{WI} + T_t + \frac{\tau}{3}} \right) * X_t * e_{bit} \quad (5.11)$$

onde T_{WI} é o tempo que WI demora, após receber o *flit* com a ficha, para iniciar a transmissão desse *flit* para a próxima WI, quando essa não tem dados para transmitir.

Por fim, o consumo de energia total dos *flits* com a ficha corresponde a soma da energia consumida por esses *flits* quando o canal está ocioso e quando o canal está

ocupado, como é descrito na Equação 5.12.

$$E = E_{busy} + E_{idle} \quad (5.12)$$

5.2 Análise do Consumo de Energia

Como o foco principal do mecanismo proposto nesse trabalho é a redução do consumo de energia, essa métrica será a primeira a ser analisada. Como foi mencionado na Subseção 2.3.3.1, no mecanismo de passagem da ficha, mesmo que as WIs não estejam transmitindo dados, o *flit* com a ficha continua circulando livremente entre as WIs e consequentemente consumindo energia. O objetivo aqui é analisar o impacto energético disso, no canal, bem como mostrar em que momentos essa característica é mais acentuada.

Como referência para o consumo de energia por bit foi utilizado o valor 1,8320 pJ/bit, extraído de [MGS⁺17], como a energia necessária para transmitir 1 bit entre duas WIs à 10 mm de distância uma da outra. Foram escolhidos os valores de 20, 40 e 80 ns para o THT, atendendo ao requisito de que o THT seja suficiente grande para que uma WI consiga enviar um pacote por completo, considerando o tempo de propagação, o tempo de transmissão da ficha e o tempo para transmissão dos dados.

O tamanho de pacote foi definido em 8 flits, onde cada flit tem o tamanho de 32 bits. Para o tempo que a WI demora para, após receber o *flit* com a ficha, iniciar a transmissão desse flit para a próxima WI, quando essa não tem dados para transmitir (T_{WI}) foi utilizado como referência o tempo de salto (*hop time*) definido em [MGS⁺17] como 3 ciclos de *clock*, considerando um *clock* de 2,5 GHz. A energia é normalizada a partir do consumo máximo de energia do canal, que corresponde a taxa de transferência máxima do canal (16 Gbit/s) multiplicado pelo consumo de energia por bit.

A Figura 5.1 apresenta o gráfico do consumo de energia E apenas para transmissão dos flits com dados em função da carga de tráfego de entrada oferecida ao canal G para os diferentes valores do THT. A partir dela podemos observar que o consumo de energia na transmissão de *flits* com dados segue uma tendência de crescimento linear coerente com o crescimento da carga de tráfego de entrada oferecida ao canal.

Os diferentes valores do THT não influenciaram a energia consumida para transmissão dos flits com dados em função de G , mas sim no máximo da energia consumida para cada um dos valores do THT. Essa variação está estritamente relacionada ao impacto do tempo do THT na taxa de transferência máxima do canal, como será mostrado na Seção 5.3.

O consumo de energia mínimo e máximo para os valores do THT de 20, 40 e 80 ns foi de 0,2, 0,2 e 0,2 e 0,87, 0,92 e 0,97, respectivamente. A média de consumo de energia foi de 0,445, 0,47 e 0,495 com um desvio padrão de 0,2669, 0,2814, 0,2958 para os

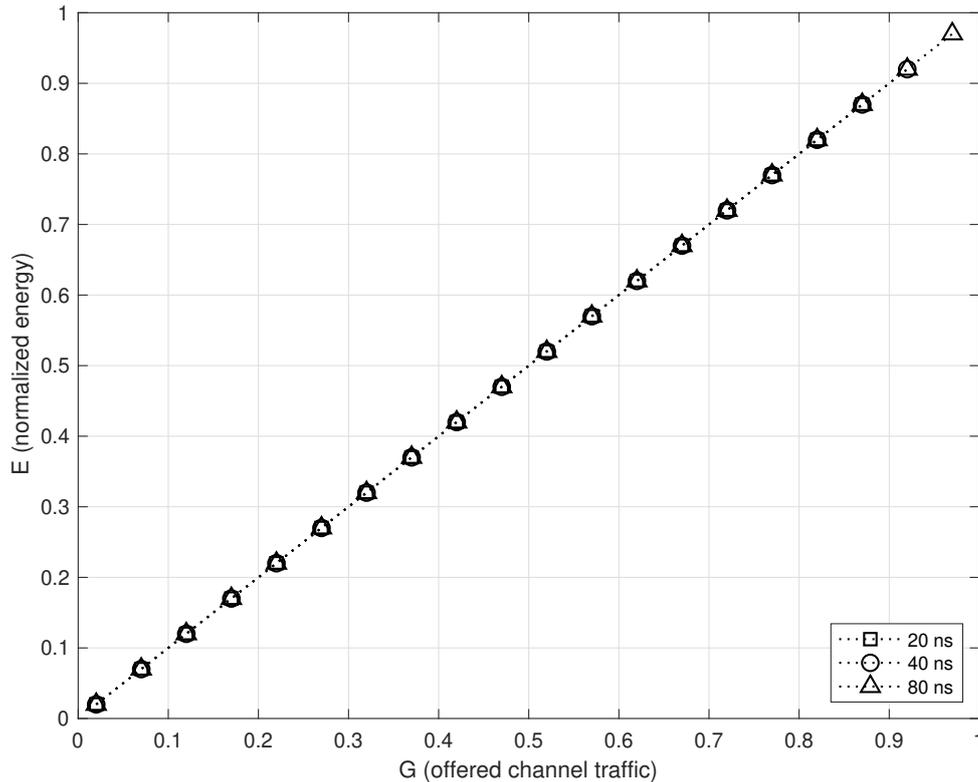


Figura 5.1: Consumo de energia E para transmissão de *flits* de dados utilizando o mecanismo de passagem da ficha em função da carga de tráfego de entrada G e para diferentes valores do THT. Fonte: o autor.

valores do THT de 20, 40 e 80 ns, respectivamente. Essas estatísticas estão consolidadas na Tabela 5.1.

Tabela 5.1: Estatísticas do consumo de energia para transmissão de *flits* com dados.

THT (ns)	<i>min</i>	<i>max</i>	μ	σ
20	0,02	0,8700	0,4450	0,2669
40	0,02	0,9200	0,4700	0,2814
80	0,02	0,9700	0,4950	0,2958

Para continuar a análise, agora é necessário avaliar o consumo de energia E apenas na transmissão do *flit* com a ficha. Isso é mostrado na Figura 5.2, onde essa informação é apresentada em relação à carga de tráfego de entrada oferecido ao canal G e ao tempo do THT. Esse gráfico apresenta uma curva linear com uma tendência de decrescimento com o aumento de G . Isso ocorre devido à redução no tempo que o canal fica ocioso conforme a carga de tráfego de entrada vai aumentando, que faz com que menos fichas sejam enviadas.

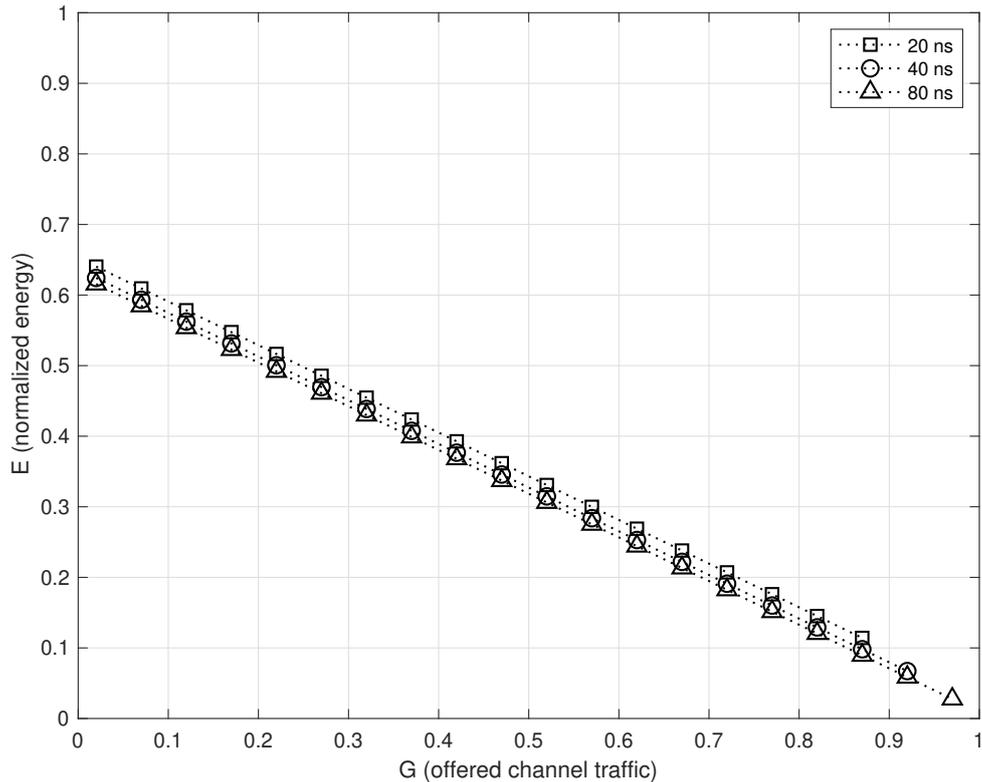


Figura 5.2: Consumo de energia E para transmissão dos *flits* com a ficha do mecanismo de passagem da ficha em função da carga de tráfego de entrada G para diferentes valores do THT. Fonte: o autor.

A Tabela 5.2 mostra as estatísticas do consumo de energia para transmissão dos *flits* com a ficha. O consumo de energia mínimo e máximo para os valores do THT de 20, 40 e 80 ns foi de 0,1142, 0,0671 e 0,0276 e 0,6403, 0,6243 e 0,6157, respectivamente. A média de consumo de energia foi de 0,3772, 0,3457 e 0,3216 com um desvio padrão de 0,1652, 0,1742, 0,1831 para os valores do THT de 20, 40 e 80 ns, respectivamente. O THT provocou uma redução média no consumo de energia de 8,35% do THT de 20 ns para o THT de 40 ns. Entre o THT de 20 ns e o de 80 ns essa redução foi de 14,74%. Por fim, a redução entre o THT de 40 ns e o de 80 ns foi de 6,97%.

Tabela 5.2: Estatísticas do consumo de energia para transmissão de *flits* com a ficha.

THT (ns)	<i>min</i>	<i>max</i>	μ	σ
20	0,1142	0,6403	0,3772	0,1652
40	0,0671	0,6243	0,3457	0,1742
80	0,0276	0,6157	0,3216	0,1831

Para ficar mais evidente a sobrecarga no uso de energia pelo *flits* com a ficha, para cada nível de tráfego de entrada oferecido ao canal, foi calculada a razão entre a

energia consumida para transmitir *flits* com a ficha e a energia consumida para transmitir *flits* com dados. Essas informações são mostradas no gráfico da Figura 5.3. Como pode ser observado nesse gráfico, para cargas de tráfego de entrada abaixo de 40%, a maior parte da energia é consumida para transmissão dos *flits* com a ficha.

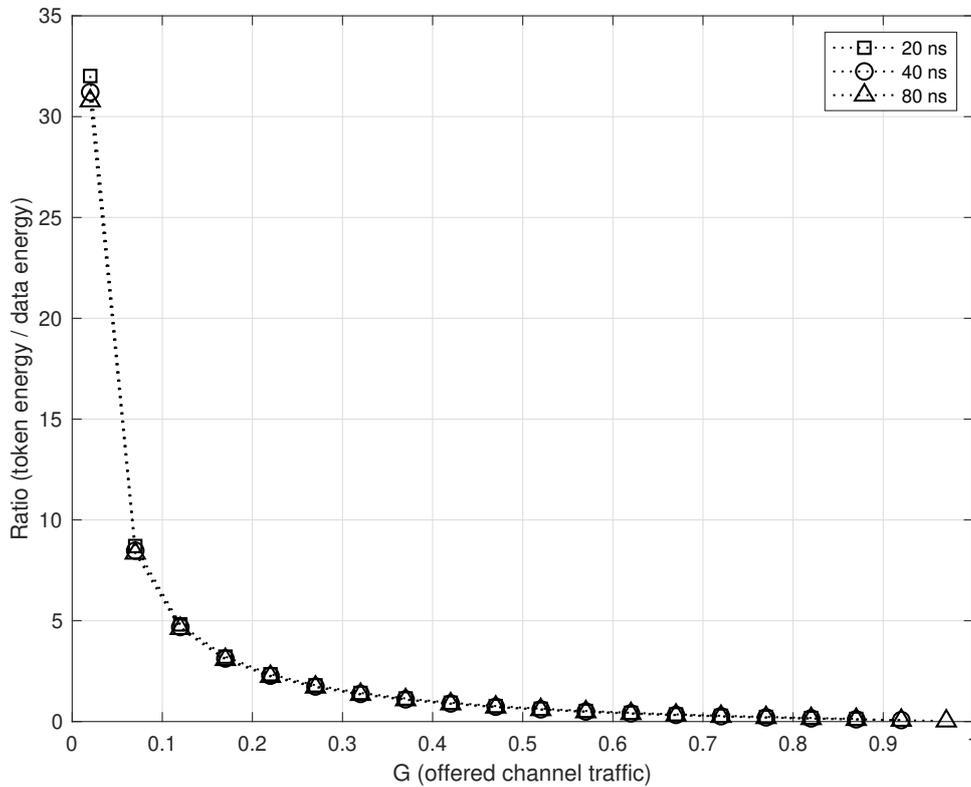


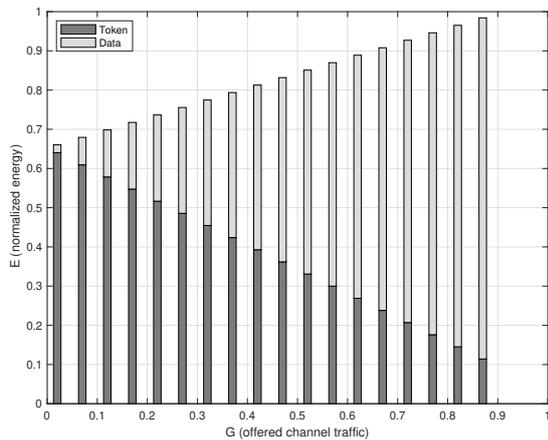
Figura 5.3: Razão entre a energia consumida para transmissão de *flits* com a ficha e entre a energia consumida para transmissão de *flits* de dados, em função da carga de tráfego de entrada G e para diferentes valores do THT. Fonte: o autor.

Os valores mínimo e máximo da razão entre o consumo de energia para transmissão dos *flits* com a ficha e dos *flits* com dados para os valores do THT de 20, 40 e 80 ns foi de 0,1312, 0,0730, 0,0284 e 32,02, 31,21, 30,78, respectivamente. A média dessa razão foi de 3,331, 3,067 e 2,868 com um desvio padrão de 7,473, 7,1210 e 6,87 para os valores do THT de 20, 40 e 80 ns, respectivamente. Essas estatísticas estão consolidados na Tabela 5.3.

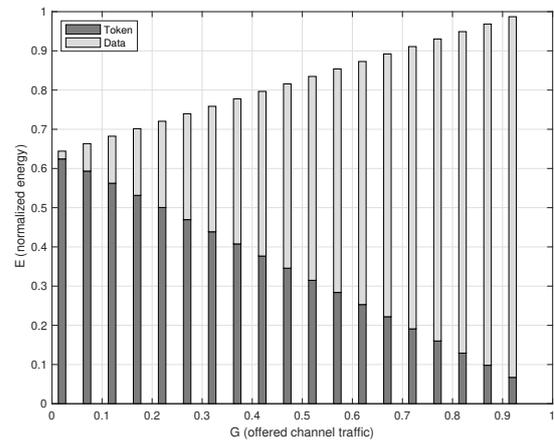
Tabela 5.3: Estatísticas dos dados da razão entre o consumo de energia para transmissão dos *flits* com a ficha e o consumo de energia para transmissão dos *flits* com dados.

THT (ns)	<i>min</i>	<i>max</i>	μ	σ
20	0,1312	32,0200	3,3310	7,4730
40	0,0730	31,2100	3,0670	7,1210
80	0,0284	30,7800	2,8680	6,8700

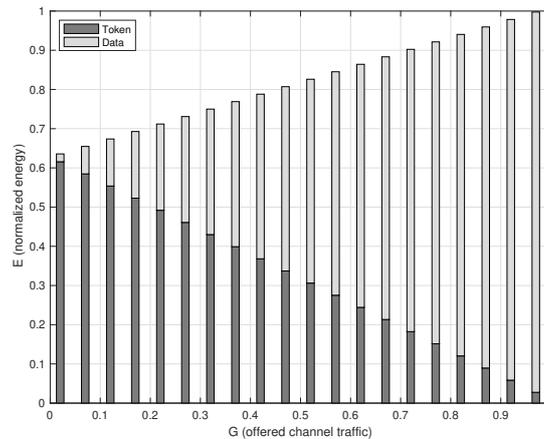
O consumo de energia E combinado de *flits* com a ficha e de *flits* com dados em relação à carga de tráfego de entrada G é mostrado na Figura 5.4. Nela são mostrados três gráficos, cada um para um valor diferente do THT. No primeiro (Figura 5.4a) o valor do THT é de 20 ns, no segundo (Figura 5.4b) o valor do THT é de 40 ns. Por fim, no último (Figura 5.4c) o valor do THT é de 80 ns.



(a) THT=20 ns.



(b) THT=40 ns.



(c) THT=80 ns.

Figura 5.4: Contribuição individual dos *flits* com a ficha e dos *flits* de dados no consumo total de energia em relação à carga de tráfego de entrada para diferentes valores de THT. Fonte: o autor.

Algumas características comuns em todos os gráficos da Figura 5.4 podem ser destacadas. Primeiro, a soma do consumo total para cada valor da carga de entrada nunca é maior que 1 e nem menor que 0, o que é coerente com o esperado, dado que os valores do consumo de energia estão normalizados em relação ao consumo de energia máximo do canal. Os gráficos seguem o mesmo comportamento apresentado nos gráficos das Figuras 5.2 e 5.1 para a energia consumida pelos *flits* com a ficha e para os *flits* com dados, respectivamente.

Assim como foi mostrado no gráfico da Figura 5.3, em todos os gráficos da Figura 5.4 fica evidente o maior consumo de energia para transmissão dos *flits* com a ficha nas cargas de tráfego de entrada inferiores a 40%. Para cargas superiores a essa, o consumo de energia para transmissão dos *flits* com dados é superior.

O consumo de energia dos *flits* com a ficha segue uma tendência de decréscimo linear com o aumento da carga do tráfego de entrada. Por outro lado, o consumo de energia dos *flits* com dados segue uma tendência de crescimento linear com o aumento da carga do tráfego de entrada. O gráfico da Figura 5.4a não apresenta os dados relativos à carga de entrada de 0,92 e 0,97, devido à redução na taxa de transferência máxima imposta pelo valor do THT, conforme a Equação 5.6. O mesmo ocorre no gráfico da Figura 5.4b, para a carga de entrada de 0,97.

O gráfico da Figura 5.5 sumariza os dados da Figura 5.4, a partir dele é possível observar a proporção dos valores do consumo de energia dos *flits* com ficha e dos *flits* com dados em relação ao consumo de energia total para cada um dos valores do THT.

A Figura 5.5a apresenta essa proporção para o THT de 20 ns, nela o consumo de energia dos *flits* com ficha corresponde a 46%, enquanto o dos *flits* com dados corresponde a 54%. Para o THT de 40 ns, Figura 5.5b, o consumo de energia dos *flits* com a ficha representa 42% da energia total, conseqüentemente o dos *flits* com dados representam 58%. Isso representa uma redução de 4% no consumo dos *flits* com a ficha com o THT de 40 ns em relação ao consumo apresentado pelo THT de 20 ns.

Por fim, a Figura 5.5c apresenta a proporção do consumo de energia para o THT de 80 ns. Nesse caso, o consumo de energia dos *flits* com a ficha corresponde a 39%, enquanto o dos *flits* com dados corresponde a 61%. Isso representa uma redução 7% em relação ao THT de 20 ns e 3% em relação ao THT de 40 ns.

A partir dos dados mostrados nesta seção é possível inferir que o consumo de energia para transmissão dos *flits* com a ficha nas cargas de tráfego de entrada abaixo de 40% corresponde a maior parte do consumo de energia. Como o tráfego de aplicações reais em NoCs possui uma natureza não uniforme [MG15], com intervalos onde os dados são transmitidos em rajadas e outros intervalos em que não são transmitidos dados e o canal fica ocioso. Esses intervalos de ociosidade do canal representam um problema para o mecanismo de passagem da ficha, visto que é neles que ocorrem a maior parte do consumo de energia para transmissão dos *flits* com a ficha. Nesses intervalos, a ficha fica circulando livremente entre as WIs até que alguma delas tenha dados para transmitir. Isso deixa claro o problema do consumo de energia do mecanismo de passagem da ficha.

Como essa tese propõe o uso do *slotted* CSMA não-persistente como mecanismo de controle de acesso ao meio para DWiNoCs em substituição ao mecanismo de passagem da ficha visando reduzir o consumo de energia sem impactar significativamente a taxa de transferência e a latência, nas próximas seções serão analisados os resultados da com-

paração entre o mecanismo de passagem da ficha e o CSMA (*slotted* não-persistente e não-persistente) com relação à taxa de transferência e a latência.

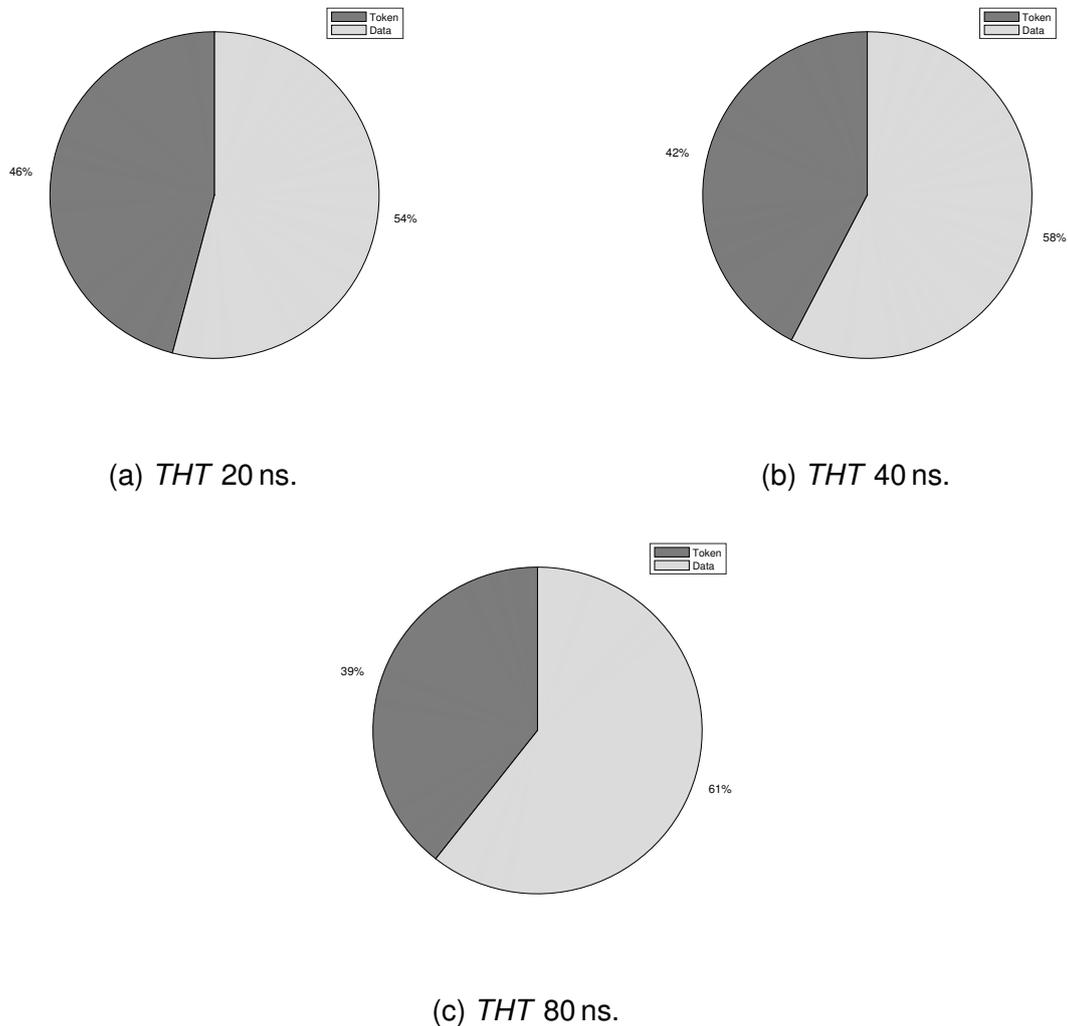


Figura 5.5: Consumo de energia sumarizado dos *flits* com a ficha e dos *flits* de dados para diferentes valores de THT. Fonte: o autor.

5.3 Análise da Taxa de Transferência

Na última Seção foi analisado o quanto de energia pode ser economizada com o uso do CSMA como mecanismo MAC, nessa seção o objetivo é analisar o impacto do CSMA na taxa de transferência.

Primeiramente será analisada a taxa de transferência do mecanismo de passagem da ficha. A Figura 5.6 apresenta o gráfico da taxa de transferência S em relação à carga de tráfego de entrada G para três diferentes valores do THT, 20, 40 e 80 ns. Nele podemos observar uma tendência de crescimento linear da taxa de transferência com o aumento da

carga de tráfego de entrada. Os diferentes valores do THT não provocaram uma variação no valor da taxa de transferência para os diferentes níveis de carga de tráfego de entrada. Também é possível observar no gráfico o efeito do THT, conforme a Equação 5.6, sobre a taxa de transferência máxima.

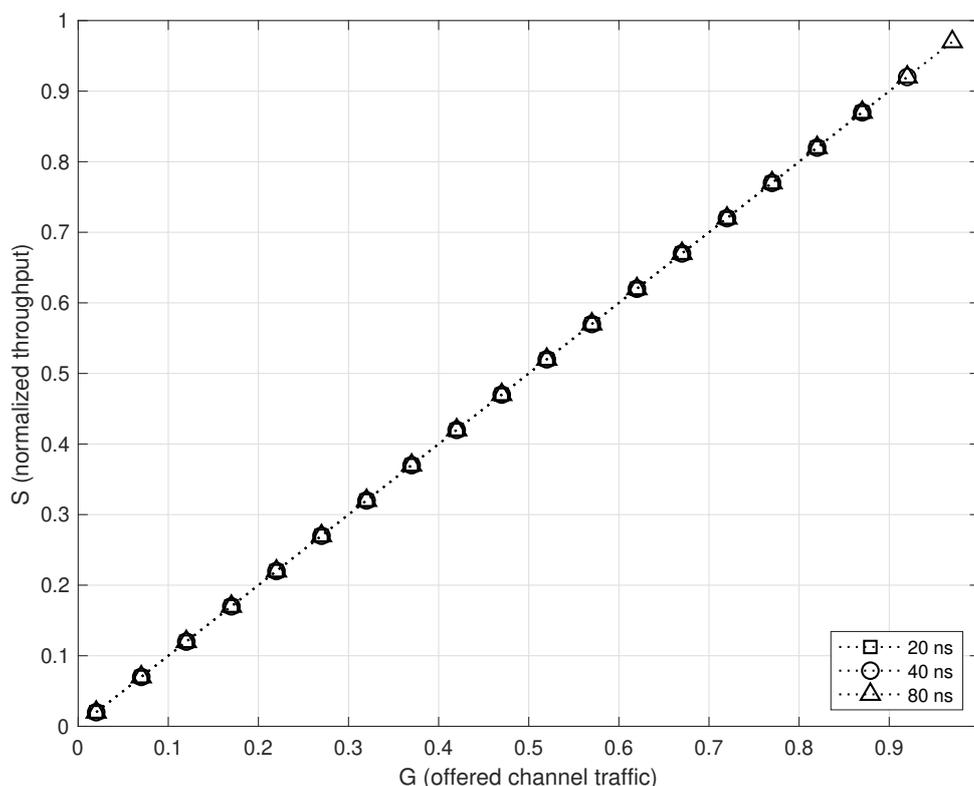


Figura 5.6: Taxa de transferência do mecanismo de passagem da ficha em relação à carga de tráfego de entrada para diferentes valores do THT. Fonte: o autor.

A taxa de transferência mínima foi igual para todos os valores do THT, 0,02. Por outro lado, o valor máximo da taxa de transferência variou para cada valor de THT. Para os valores de THT de 20, 40 e 80 ns as taxas de transferência máximas foram de 0,87, 0,92 e 0,97, respectivamente. A taxa de transferência média para os valores do THT de 20, 40 e 80 ns foram de 0,445, 0,470 e 0,495, respectivamente. O desvio padrão foi de 0,2669, 0,2814 e 0,2958 para os valores de THT de 20, 40 e 80 ns, respectivamente. Essas estatísticas do gráfico da Figura 5.6 estão consolidados na Tabela 5.4.

Tabela 5.4: Estatísticas da taxa de transferência do mecanismo de passagem da ficha.

THT (ns)	<i>min</i>	<i>max</i>	μ	σ
20	0,02	0,87	0,445	0,2669
40	0,02	0,92	0,470	0,2814
80	0,02	0,97	0,495	0,2958

Para o CSMA não-persistente, a taxa de transferência S em relação à carga de tráfego de entrada G é apresentada no gráfico da Figura 5.7. A taxa de transferência está normalizada em relação à capacidade máxima do canal. Podemos observar uma tendência de crescimento praticamente linear com o aumento da carga de tráfego de entrada.

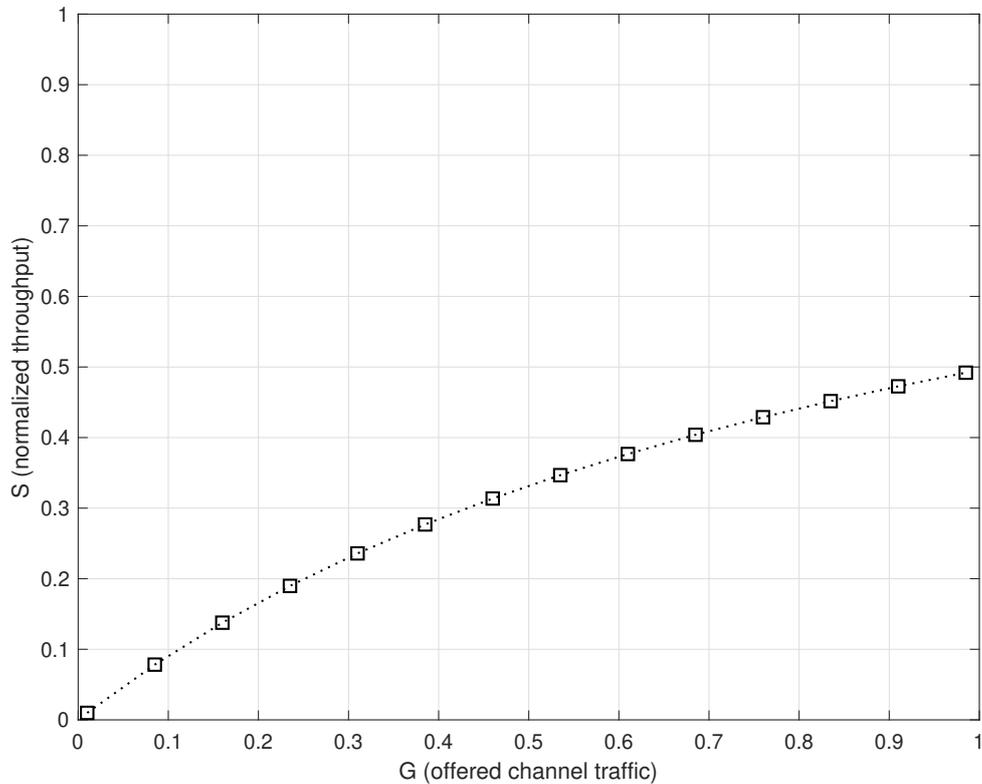


Figura 5.7: Taxa de transferência do CSMA não-persistente em relação à carga de tráfego de entrada. Fonte: o autor.

Cabe ressaltar aqui que o CSMA por sua natureza aleatória suporta uma carga de tráfego de entrada superior à capacidade do canal sem de fato ultrapassar a capacidade máxima do canal. Como o algoritmo do CSMA é executado para cada novo pacote que chega na WI, pode ocorrer de um pacote chegar na WI e ela não conseguir transmiti-lo, por encontrar o canal ocupado, e num instante seguinte, outro pacote ser transmitido, na mesma WI, por encontrar o canal livre.

As estatísticas do gráfico da Figura 5.7 são apresentadas na Tabela 5.5. Como pode ser observado nessa tabela, a taxa de transferência mínima foi de 0,0099. Por outro lado, a taxa de transferência máxima foi de 0,4919. A taxa de transferência média ficou em 0,301 com um desvio padrão de 0,1523.

Na taxa de transferência do *slotted* CSMA não-persistente, Figura 5.8, foi observado um comportamento semelhante ao observado no CSMA não-persistente, mostrado na Figura 5.7. Entretanto, é possível observar um pequeno aumento no valor máximo da taxa de transferência bem como na média em relação ao CSMA não-persistente. O valor

Tabela 5.5: Estatísticas da taxa de transferência do CSMA não-persistente.

min	max	μ	σ
0,0099	0,4919	0,301	0,1523

máximo foi de 0,4941, enquanto a média ficou em 0,3019 com um desvio padrão de 0,153. O valor mínimo foi igual ao obtido no CSMA não-persistente, 0,0099. Essas estatísticas estão consolidadas da Tabela 5.6.

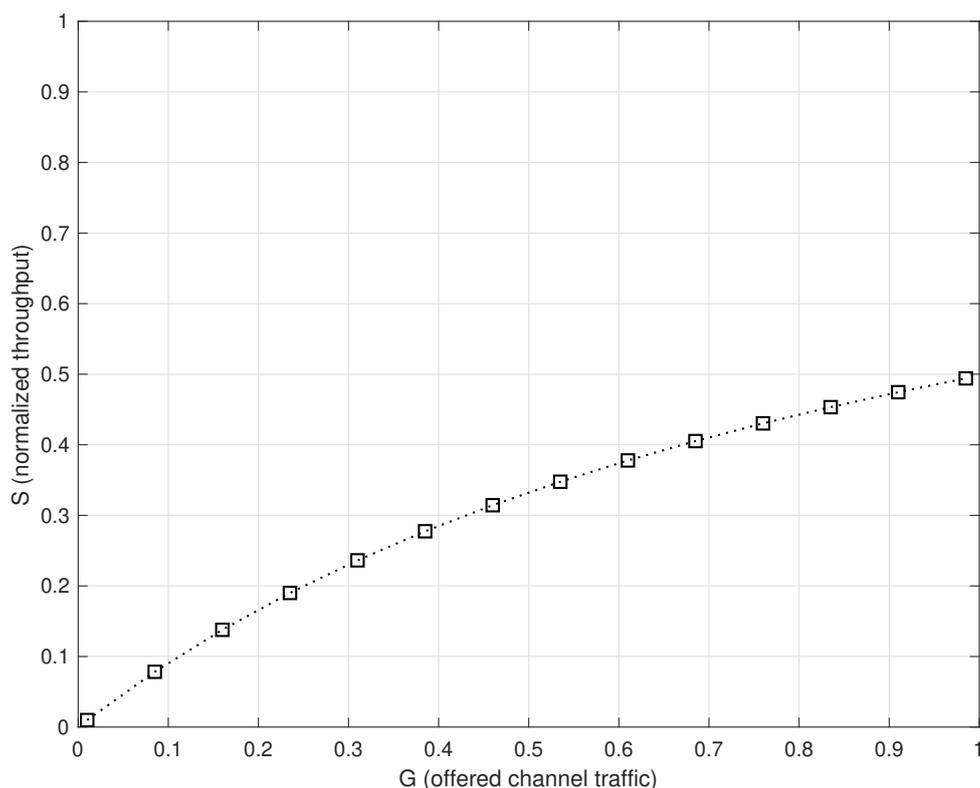


Figura 5.8: Taxa de transferência do *slotted* CSMA não-persistente em relação à carga de tráfego de entrada. Fonte: o autor.

Um dos benefícios do *slotted* CSMA não-persistente, que seria o aumento da taxa de transferência em função da redução no número de colisões e conseqüentemente no número de retransmissões, teve pouco efeito no gráfico da Figura 5.8, devido ao pequeno número de retransmissões quando o canal está submetido a essa carga de tráfego de entrada. O número de retransmissões em função do aumento da taxa de transferência pode ser observado no gráfico da Figura 5.13.

Por fim, a Figura 5.9 apresenta a comparação da taxa de transferência S em função da carga de tráfego de entrada G entre o mecanismo de passagem da ficha, o CSMA não-persistente e o *slotted* CSMA não-persistente. Como pode ser observado nesse gráfico, o mecanismo de passagem da ficha apresenta uma maior taxa de transferência para

Tabela 5.6: Estatísticas da taxa de transferência do *slotted* CSMA não-persistente.

min	max	μ	σ
0,0099	0,4941	0,3019	0,153

todos os níveis de carga de tráfego de entrada em comparação ao CSMA não-persistente e ao *slotted* CSMA não-persistente. Essa diferença cresce com o aumento da carga de tráfego de entrada. Em média, o mecanismo de passagem da ficha, com o THT de 20, 40 e 80 ns, teve um desempenho de 47,84%, 56,14% e 64,45% superior ao do CSMA não-persistente e 47,39%, 55,68% e 63,96% superior ao *Slotted* CSMA não-persistente, respectivamente.

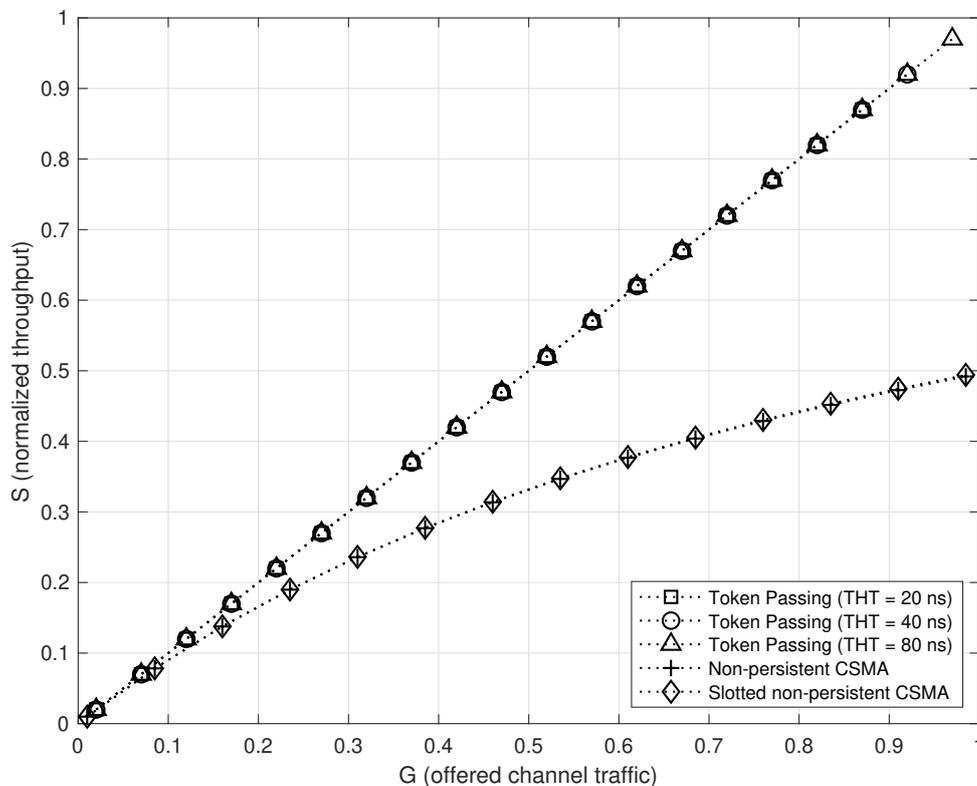


Figura 5.9: Comparação da taxa de transferência entre o mecanismo de passagem da ficha (com diferentes valores do THT), o CSMA não-persistente e o *slotted* CSMA não-persistente em função da carga de tráfego de entrada. Fonte: o autor.

5.4 Análise da Latência

A Figura 5.10 apresenta o gráfico da latência (em nanosegundos) em função da taxa de transferência (normalizada em relação à taxa de transferência máxima do canal —

16 Gbit/s). Cabe ressaltar aqui que o modelo analítico utilizado para cálculo da latência do mecanismo de passagem da ficha considera uma disciplina de serviço exaustiva, onde a WI que tem a posse ficha transmite todos os pacotes quem têm na fila de espera até esvaziá-la completamente, desconsiderando o tempo limite que deveria ficar com a ficha (THT). O cálculo da latência considerando a disciplina de serviço limitada, que deveria ser a empregada quando o THT é utilizado [HO86], exigiria a aplicação de um modelo iterativo que fugiria ao escopo desse trabalho.

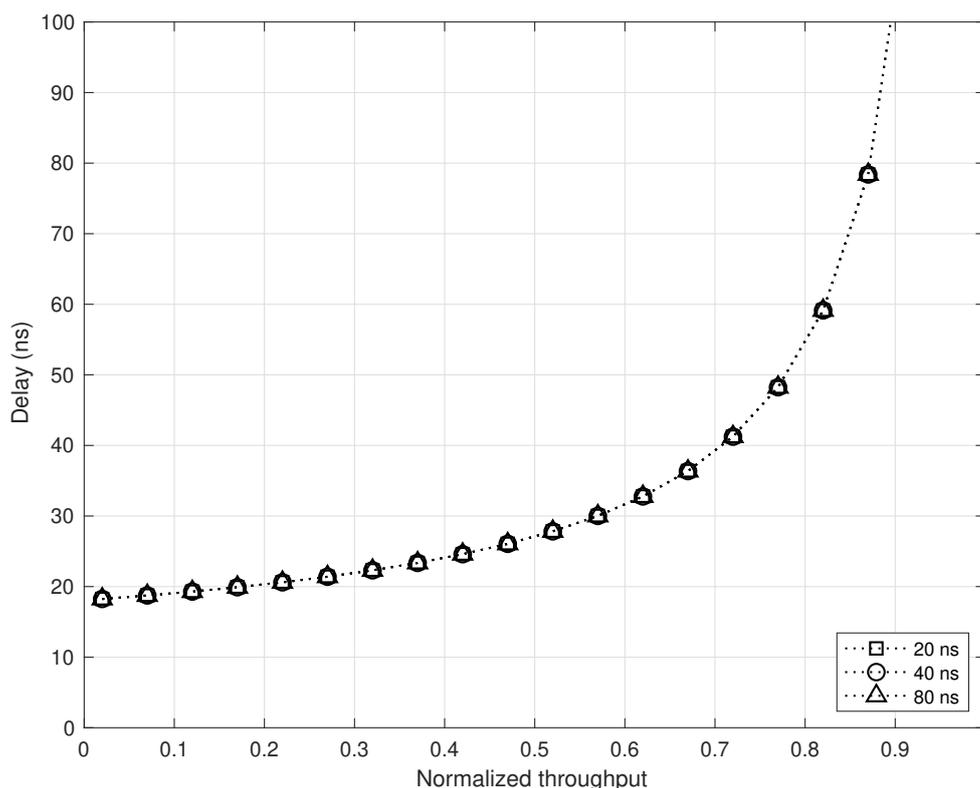


Figura 5.10: Latência do mecanismo de passagem da ficha em função da taxa de transferência para diferentes valores do THT. Fonte: o autor.

Como pode ser observado no gráfico da Figura 5.10 a latência aumenta linearmente até certo ponto, ponto de saturação, e depois segue aumentando em escala exponencial. A latência mínima foi igual para todos os valores do THT, 18,25. Por outro lado, o valor máximo da latência variou para cada valor de THT. Para os valores de THT de 20, 40 e 80 ns a latência máxima foi de 78,4, 121,7 e 309,6, respectivamente. A latência média para os valores do THT de 20, 40 e 80 ns foram de 31,58, 36,33 e 49,99 com um desvio padrão de 16,18, 25,98 e 66,12, respectivamente. Essas estatísticas do gráfico da Figura 5.10 estão consolidados na Tabela 5.7.

A latência do CSMA não-persistente em função da taxa de transferência normalizada é mostrada no gráfico da Figura 5.11. Assim como no gráfico da latência do mecanismo de passagem da ficha (Figura 5.10), no gráfico do CSMA não-persistente a latência

Tabela 5.7: Estatísticas dos dados de latência do mecanismo de passagem da ficha.

<i>THT</i> (ns)	<i>min</i>	<i>max</i>	μ	σ
20	18,25	78,4	31,58	16,18
40	18,25	121,7	36,33	25,98
80	18,25	309,6	49,99	66,12

também aumentou linearmente até o ponto de saturação e a partir desse ponto cresceu exponencialmente.

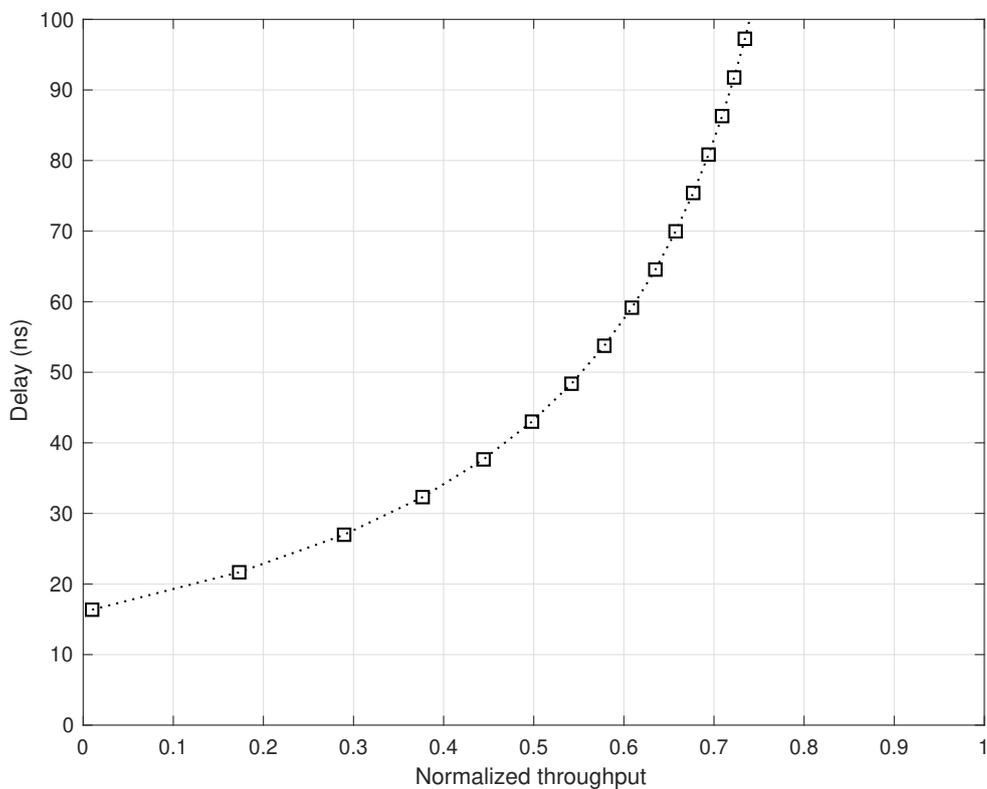


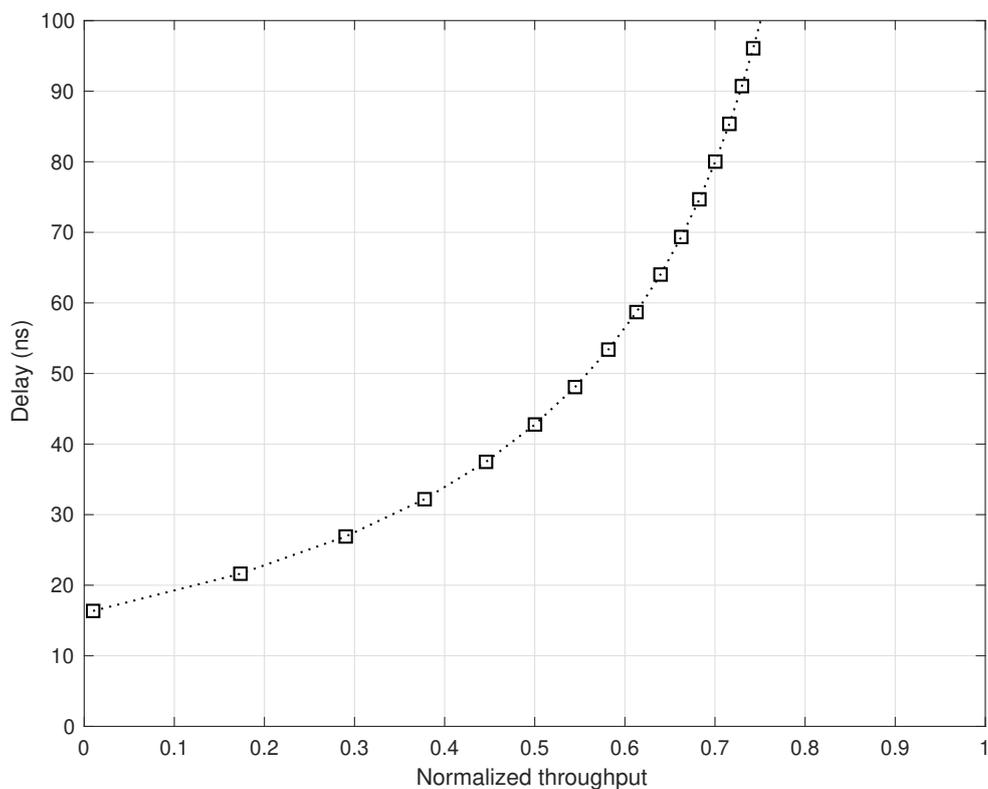
Figura 5.11: Latência do CSMA não-persistente em função da taxa de transferência. Fonte: autor.

As estatísticas do gráfico da Figura 5.11 são apresentadas na Tabela 5.8. Como pode ser observado nessa tabela, a latência mínima foi de 16,36 ns. Por outro lado, a latência máxima foi de 119,3 ns. A latência média ficou em 67,47 ns com um desvio padrão de 32,05 ns.

O gráfico da Figura 5.12 apresenta a latência (em nanossegundos) do *Slotted CSMA* não-persistente em função da taxa de transferência normalizada. De forma análoga ao CSMA não-persistente e ao mecanismo de passagem de ficha a latência seguiu aumentando linearmente até o ponto de saturação e após esse ponto continuou crescendo, porém em escala exponencial.

Tabela 5.8: Estatísticas dos dados normalizados da latência do CSMA não-persistente.

min	max	μ	σ
16,36	119,3	67,47	32,05

Figura 5.12: Latência para o *slotted* CSMA não persistente em função da taxa de transferência. Fonte: o autor.

A Tabela 5.9 apresenta as estatísticas do gráfico da Figura 5.12. Como pode ser observado nessa tabela, a latência mínima foi de 16,36 ns, enquanto a máxima foi de 117,6 ns. A média da latência ficou em 66,79 ns com um desvio padrão de 31,52 ns.

Tabela 5.9: Estatísticas dos dados normalizados da latência do *Slotted* CSMA não-persistente.

min	max	μ	σ
16,36	117,6	66,79	31,52

O gráfico da Figura 5.13 apresenta as taxas de retransmissão tanto do CSMA não-persistente quanto do *Slotted* CSMA não-persistente. Como pode ser observado, as curvas desse gráfico apresentam um comportamento de semelhante de crescimento linear até um certo ponto e depois segue aumentando exponencialmente. O *Slotted* CSMA não-

persistente apresentou a menor taxa de retransmissão. Em média, ele teve uma taxa de retransmissão 17,88% inferior ao CSMA não-persistente.

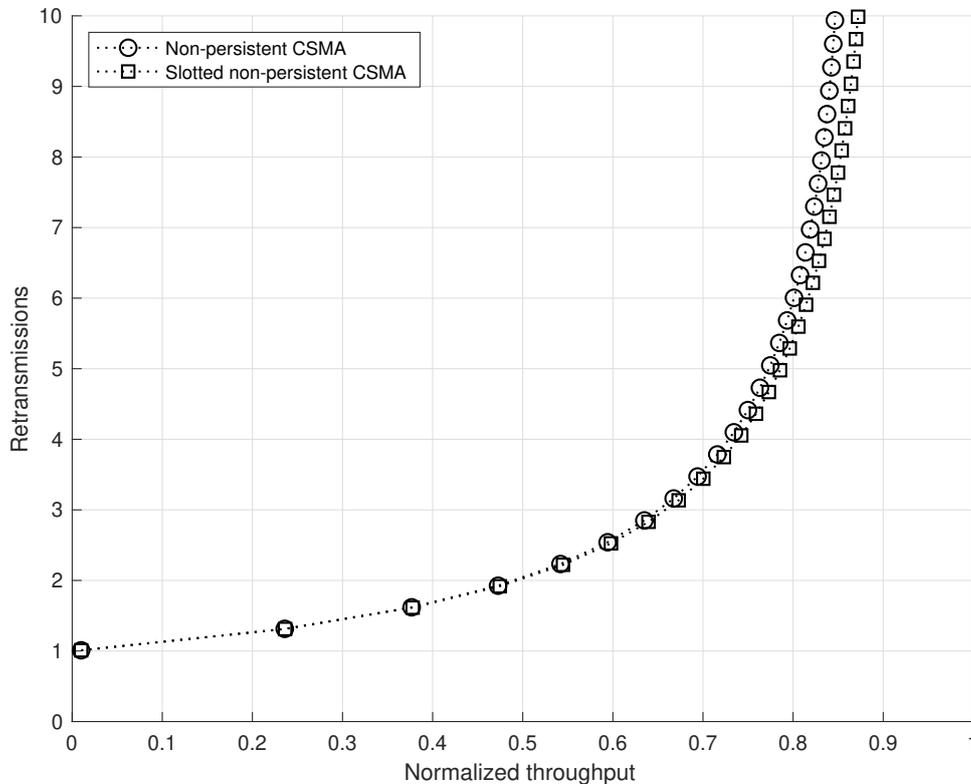


Figura 5.13: Comparação entre a taxa de retransmissão do CSMA não-persistente e do *Slotted* CSMA não-persistente em função da taxa de transferência. Fonte: o autor.

A Tabela 5.10 apresenta as estatísticas do gráfico da Figura 5.13. Como pode ser observado nessa tabela, para o CSMA não-persistente o número mínimo de retransmissões foi de 1,01, enquanto o máximo foi de 182,1. A média de retransmissões ficou em 76,3 com um desvio padrão de 52,24. Para o *Slotted* CSMA não-persistente o número mínimo de retransmissões também ficou em 1,01, entretanto o máximo foi de 137,4. A média de retransmissões ficou em 62,65 com um desvio padrão de 39,45.

Tabela 5.10: Estatísticas de retransmissões do *Slotted* CSMA não-persistente e do CSMA não-persistente.

		<i>min</i>	<i>max</i>	μ	σ
Taxa de retransmissão	<i>Slotted</i> CSMA não-persistente	1,01	137,4	62,65	39,45
	CSMA não-persistente	1,01	182,1	76,3	52,24

Por fim, a comparação da latência do mecanismo de passagem da ficha com o CSMA não-persistente e com o *Slotted* CSMA não-persistente é mostrado na Figura 5.14. Como é possível observar nesse gráfico, tanto o CSMA não-persistente quanto o *Slotted*

CSMA não-persistente apresentaram uma menor latência para as taxas de transferência inferiores a 0,1. A partir desse ponto o mecanismo de passagem da ficha apresentou uma menor latência. Em média, o CSMA não-persistente e o *Slotted* CSMA não-persistente tiveram uma latência superior ao mecanismo de passagem da ficha com o THT de 20, 40 e 80 ns em 113,65%, 85,71% e 34,96% e 111,49%, 83,84% e 33,60%, respectivamente.

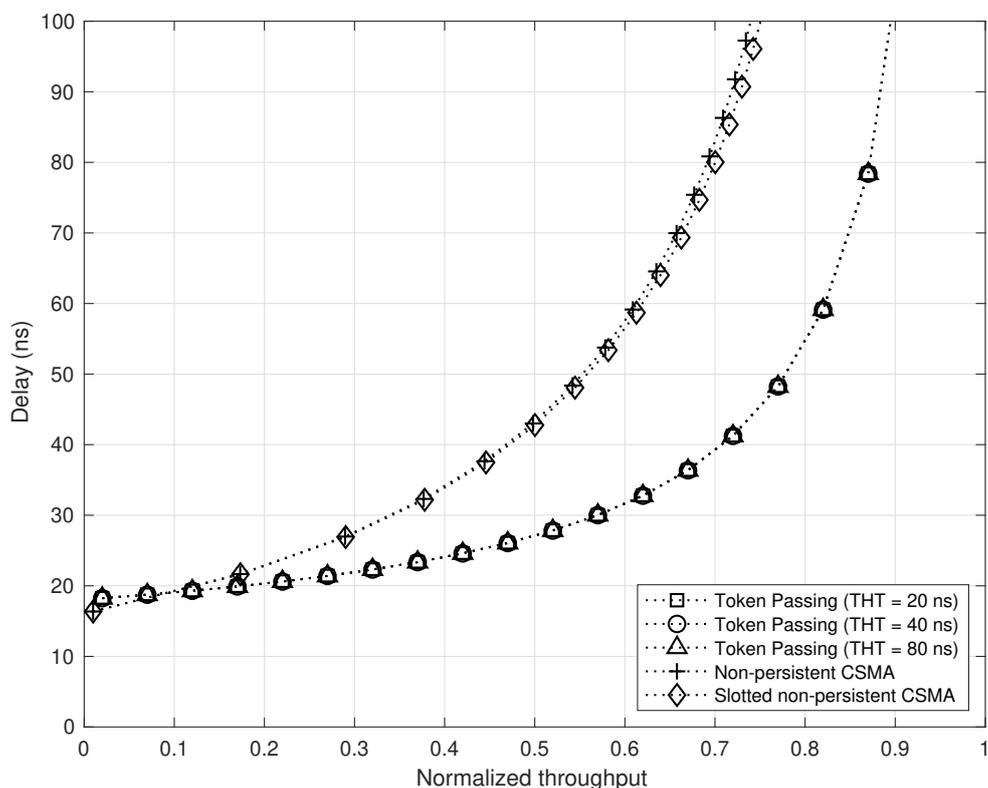


Figura 5.14: Comparação da latência entre o mecanismo de passagem da ficha, o CSMA não-persistente e o *slotted* CSMA não-persistente em função da taxa de transferência. Fonte: o autor.

6. CONCLUSÃO

Essa tese propôs o uso do CSMA como mecanismo de controle de acesso ao meio para DWiNoCs, tirando proveito dos canais não sobrepostos criados por essa arquitetura de WiNoC com antenas direcionais e do comportamento do CSMA para determinadas categorias de tráfego, para aumentar a eficiência energética do canal, quando ele está submetido a baixas cargas de tráfego.

O mecanismo proposto foi analisado a partir de modelos analíticos para taxa de transferência, latência e consumo de energia. Os resultados mostraram que tanto o CSMA não-persistente, quanto o *slotted* CSMA não-persistente são significativamente mais eficientes em termos do consumo de energia para cargas leves (abaixo de 40%) de tráfego de entrada do que o mecanismo de passagem da ficha. O uso desse mecanismo pode reduzir, em média, entre 39 e 42% o consumo de energia de cada canal, considerando uma distribuição uniforme do tráfego de entrada.

O desempenho do canal não foi afetado de forma significativa pelo uso do CSMA como MAC para cargas de tráfego de entrada inferiores a 40%. Com relação à taxa de transferência, o uso do CSMA impactou de forma pouco significativa para cargas de tráfego de entrada inferiores a 40% a partir desse ponto a taxa de transferência é impactada significativamente.

O CSMA apresentou menor latência em relação ao mecanismo de passagem da ficha para taxas de transferência inferiores a 10%. Para taxas de transferência entre 10% e 40% a latência do CSMA foi impactada de forma pouco significativa. Tanto o mecanismo de passagem da ficha quanto o CSMA apresentam crescimento exponencial da latência a partir do ponto de saturação. Entretanto, o ponto de saturação do CSMA ocorre mais cedo.

6.1 Publicações

Os seguintes artigos científicos foram publicados:

- E. L. Berz, D. A. Tesch and F. P. Hessel, “RFID indoor localization based on support vector regression and k-means”. In: Proceedings of the 24th International Symposium on Industrial Electronics, 2015, pp. 1418-1423
- D. A. Tesch, E. L. Berz and F. P. Hessel, “RFID indoor localization based on Doppler effect”. In: Proceedings of the 16th International Symposium on Quality Electronic Design, 2015, pp. 556-560

- E. L. Berz, D. A. Tesch and F. P. Hessel, “Machine-learning-based system for multi-sensor 3D localisation of stationary objects”. IET Cyber-Physical Systems: Theory & Applications, vol. 3, no. 2, Jun 2018, pp. 81-88.
- E. L. Berz, D. A. Tesch and F. P. Hessel, “A hybrid RFID and CV system for item-level localization of stationary objects”. In: Proceedings of the 18th International Symposium on Quality Electronic Design, 2017, pp. 331-336

6.2 Trabalhos Futuros

Durante o andamento deste trabalho foram identificadas algumas possibilidades de temas que podem ser abordados em trabalhos futuros, como:

- Visto que para o modo p-persistente do CSMA, existe um p ótimo que consegue se adaptar bem a qualquer carga de tráfego. Um trabalho futuro poderia explorar esse aspecto e propor um mecanismo MAC dinâmico que ajusta a probabilidade p conforme a carga na WI.
- Outra possibilidade seria explorar a característica *dual-band* das antenas direcionais PLPA, usando essa segunda largura de banda para transmissão de ACKs.
- Uma possível abordagem para reduzir o consumo de energia, porém pelo receptor, seria colocar o receptor das WIs que não estejam envolvidas na transmissão corrente para dormir.
- Substituir os transmissores convencionais por transmissores adaptativos, que ajustam a potência de transmissão conforme a distância entre transmissor e receptor, pode permitir a criação de subcanais dedicados para melhorar ainda mais a eficiência do canal.
- Ao utilizar os transceptores mencionados no último item, as WIs estariam sujeitas ao problema do nodo oculto. Um trabalho futuro poderia propor um mecanismo para detecção da portadora de forma distribuída, formando agrupamentos (*clusters*) de detecção da portadora para tentar mitigar o problema do nodo oculto.
- Por fim, existe uma carência muito significativa de simuladores de WiNoCs que consigam simular canais compartilhados e sejam mais flexíveis para testes de novos mecanismos de acesso ao meio. Um novo simulador poderia ser proposto cobrindo essas lacunas.

REFERÊNCIAS BIBLIOGRÁFICAS

- [ANACA15] Abadal, S.; Nemirovsky, M.; Alarcón, E.; Cabellos-Aparicio, A. “Networking challenges and prospective impact of broadcast-oriented wireless networks-on-chip”. In: Proceedings of the 9th International Symposium on Networks-on-Chip, 2015, pp. 12:1–12:8.
- [BAC+07] Bernstein, K.; Andry, P.; Cann, J.; Emma, P.; Greenberg, D.; Haensch, W.; Ignatowski, M.; Koester, S.; Magerlein, J.; Puri, R.; Young, A. “Interconnects in the third dimension: design challenges for 3D ICs”. In: Proceedings of the 44th Design Automation Conference, 2007, pp. 562–567.
- [BM02] Banerjee, K.; Mehrotra, A. “A power-optimal repeater insertion methodology for global interconnects in nanometer designs”, *IEEE Transactions on Electron Devices*, vol. 49–11, nov. 2002, pp. 2001–2007.
- [CC09] Carlson, A. B.; Crilly, P. B. “Communication systems: an introduction to signals and noise in electrical communication”. McGraw-Hill Education, 2009, 944p.
- [CCK+08] Chang, M. F.; Cong, J.; Kaplan, A.; Naik, M.; Reinman, G.; Socher, E.; Tam, S. W. “CMP network-on-chip overlaid with multi-band RF-interconnect”. In: Proceedings of the 14th International Symposium on High Performance Computer Architecture, 2008, pp. 191–202.
- [CDG+12] Chang, K.; Deb, S.; Ganguly, A.; Yu, X.; Sah, S. P.; Pande, P. P.; Belzer, B.; Heo, D. “Performance evaluation and design trade-offs for wireless network-on-chip architectures”, *ACM Journal on Emerging Technologies in Computing Systems*, vol. 8–3, ago. 2012, pp. 23:1–23:25.
- [CHX+12] Carpenter, A.; Hu, J.; Xu, J.; Huang, M.; Wu, H.; Liu, P. “Using transmission lines for global on-chip communication”, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2–2, jun. 2012, pp. 183–193.
- [CL14] Chen, J.; Lai, Y.-H. “A study of CSMA-based and token-based wireless interconnects network-on-chip”. In: Proceedings of the International Conference on Communication Problem-solving, 2014, pp. 205–209.
- [CPX09] Carloni, L. P.; Pande, P.; Xie, Y. “Networks-on-chip in emerging interconnect paradigms: advantages and challenges”. In: Proceedings of the 3rd International Symposium on Networks-on-Chip, 2009, pp. 93–102.
- [DCG+15] Dong, P.; Chen, Y.-K.; Gu, T.; Buhl, L. L.; Neilson, D. T.; Sinsky, J. H. “Reconfigurable 100 Gb/s silicon photonic network-on-chip”, *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7–1, jan. 2015, pp. A37–A43.

- [DCY⁺13] Deb, S.; Chang, K.; Yu, X.; Sah, S. P.; Cosic, M.; Ganguly, A.; Pande, P. P.; Belzer, B.; Heo, D. "Design of an energy-efficient CMOS-compatible NoC architecture with millimeter-wave wireless interconnects", *IEEE Transactions on Computers*, vol. 62–12, dez. 2013, pp. 2382–2396.
- [DGC⁺10] Deb, S.; Ganguly, A.; Chang, K.; Pande, P.; Beizer, B.; Heo, D. "Enhancing performance of network-on-chip architectures with millimeter-wave wireless interconnects". In: Proceedings of the 21st International Conference on Application-specific Systems, Architectures and Processors, 2010, pp. 73–80.
- [DGP⁺12] Deb, S.; Ganguly, A.; Pande, P. P.; Belzer, B.; Heo, D. "Wireless NoC as interconnection backbone for multicore chips: promises and challenges", *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2–2, jun. 2012, pp. 228–239.
- [DKKM11] DiTomaso, D.; Kodi, A.; Kaya, S.; Matolak, D. "iWISE: inter-router wireless scalable express channels for network-on-chips (NoCs) architecture". In: Proceedings of the 19th Annual Symposium on High-Performance Interconnects, 2011, pp. 11–18.
- [DKM⁺15] DiTomaso, D.; Kodi, A.; Matolak, D.; Kaya, S.; Laha, S.; Rayess, W. "A-WiNoC: adaptive wireless network-on-chip architecture for chip multiprocessors", *IEEE Transactions on Parallel and Distributed Systems*, vol. 26–12, dez. 2015, pp. 3289–3302.
- [DP98] Dally, W. J.; Poulton, J. W. "Digital systems engineering". Cambridge University Press, 1998, 663p.
- [FHO02] Floyd, B. A.; Hung, C.-M.; O, K. K. "Intra-chip wireless interconnect for clock distribution implemented with integrated antennas, receivers, and transmitters", *IEEE Journal of Solid-State Circuits*, vol. 37–5, mai. 2002, pp. 543–552.
- [For12] Forouzan, B. A. "Data communications and networking". McGraw-Hill Education, 2012, 1264p.
- [Fre15] Frenzel, L. "Principles of electronic communication systems". McGraw-Hill, 2015, 944p.
- [GCD⁺11] Ganguly, A.; Chang, K.; Deb, S.; Pande, P. P.; Belzer, B.; Teuscher, C. "Scalable hybrid wireless network-on-chip architectures for multicore systems", *IEEE Transactions on Computers*, vol. 60–10, out. 2011, pp. 1485–1502.
- [GD15] Gade, S. H.; Deb, S. "Achievable performance enhancements with mm-wave wireless interconnects in NoC". In: Proceedings of the 9th International Symposium on Networks-on-Chip, 2015, pp. 1–2.

- [Han05] Hanson, G. "Fundamental transmitting properties of carbon nanotube antennas", *IEEE Transactions on Antennas and Propagation*, vol. 53–11, nov. 2005, pp. 3426–3435.
- [HM07] Haykin, S.; Moher, M. "Introduction to analog and digital communications". John Wiley & Sons, Inc., 2007, 515p.
- [HO86] Hammond, J. L.; O'Reilly, P. P. "Performance analysis of local computer networks". Addison-Wesley Longman Publishing Co., Inc., 1986, 411p.
- [HSL+03] Huang, D.; Sze, T.; Landin, A.; Lytel, R.; Davidson, H. "Optical interconnects: out of the box forever?", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 9–2, mar. 2003, pp. 614–623.
- [ITR20a] ITRS. "International Technology Roadmap for Semiconductors 2009 Edition". Recuperado de: <http://www.itrs2.net/itrs-reports.html>, jul. 2020.
- [ITR20b] ITRS. "International Technology Roadmap for Semiconductors 2012 Edition". Recuperado de: <http://www.itrs2.net/itrs-reports.html>, jul. 2020.
- [JBK+09] Joshi, A.; Batten, C.; Kwon, Y.-J.; Beamer, S.; Shamim, I.; Asanovic, K.; Stojanovic, V. "Silicon-photonics networks for global on-chip communication". In: Proceedings of the 3rd International Symposium on Networks-on-Chip, 2009, pp. 124–133.
- [KHO00] Kihong Kim; Hyun Yoon; O, K. "On-chip wireless interconnection with integrated antennas". In: Proceedings of the International Electron Devices Meeting, 2000, pp. 485–488.
- [KKD+06] Kirman, N.; Kirman, M.; Dokania, R. K.; Martinez, J. F.; Apsel, A. B.; Watkins, M. A.; Albonesi, D. H. "Leveraging optical technology in future bus-based chip multiprocessors". In: Proceedings of the 39th Annual International Symposium on Microarchitecture, 2006, pp. 492–503.
- [KMTY16] Karkar, A.; Mak, T.; Tong, K. F.; Yakovlev, A. "A survey of emerging interconnects for on-chip efficient multicast and broadcast in many-core", *IEEE Circuits and Systems Magazine*, vol. 16–1, 2016, pp. 58–72.
- [KPKJ07] Kumar, A.; Peh, L.; Kundu, P.; Jha, N. K. "Express virtual channels: towards the ideal interconnection fabric". In: Proceedings of the 34th Annual International Symposium on Computer Architecture, 2007, pp. 150–161.
- [KRH+07] Kempa, K.; Rybczynski, J.; Huang, Z.; Gregorczyk, K.; Vidan, A.; Kimball, B.; Carlson, J.; Benham, G.; Wang, Y.; Herczynski, A.; Ren, Z. F. "Carbon

nanotubes as optical antennae”, *Advanced Materials*, vol. 19–3, 2007, pp. 421–426.

- [KT75] Kleinrock, L.; Tobagi, F. “Packet switching in radio channels: part I - carrier sense multiple-access modes and their throughput-delay characteristics”, *IEEE Transactions on Communications*, vol. 23–12, dez. 1975, pp. 1400–1416.
- [LCL07] Lee, R. C. T.; Chiu, M.-C.; Lin, J.-S. “Communications engineering: essentials for computer scientists and electrical engineers”. John Wiley & Sons (Asia) Pte Ltd, 2007, 272p.
- [LTP⁺09] Lee, S.-B.; Tam, S.-W.; Pefkianakis, I.; Lu, S.; Chang, M. F.; Guo, C.; Reinman, G.; Peng, C.; Naik, M.; Zhang, L.; Cong, J. “A scalable micro wireless interconnect structure for CMPs”. In: Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, 2009, pp. 217–228.
- [MAT⁺16] Mestres, A.; Abadal, S.; Torrellas, J.; Alarcón, E.; Cabellos-Aparicio, A. “A MAC protocol for reliable broadcast communications in wireless network-on-chip”. In: Proceedings of the 9th International Workshop on Network on Chip Architectures, 2016, pp. 21–26.
- [MG15] Mansoor, N.; Ganguly, A. “Reconfigurable wireless network-on-chip with a dynamic medium access mechanism”. In: Proceedings of the 9th International Symposium on Networks-on-Chip, 2015, pp. 13:1–13:8.
- [MGS⁺17] Mondal, H. K.; Gade, S. H.; Shamim, M. S.; Deb, S.; Ganguly, A. “Interference-aware wireless network-on-chip architecture using directional antennas”, *IEEE Transactions on Multi-Scale Computing Systems*, vol. 3–3, jul. 2017, pp. 193–205.
- [MGY13] Mansoor, N.; Ganguly, A.; Yuvaraj, M. P. “An energy-efficient and robust millimeter-wave wireless network-on-chip architecture”. In: Proceedings of the International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems, 2013, pp. 19–24.
- [Mic11] Mickelson, A. R. “Silicon photonics for on-chip interconnections”. In: Proceedings of the Custom Integrated Circuits Conference, 2011, pp. 1–8.
- [Mil09] Miller, D. A. B. “Device requirements for optical interconnects to silicon chips”, *Proceedings of the IEEE*, vol. 97–7, jul. 2009, pp. 1166–1185.
- [MJK14] Morris, R.; Jolley, E.; Kodi, A. K. “Extending the performance and energy-efficiency of shared memory multicores with nanophotonic technology”, *IEEE*

Transactions on Parallel and Distributed Systems, vol. 25–1, jan. 2014, pp. 83–92.

- [MKW⁺14] Murray, J.; Kim, R.; Wettin, P.; Pande, P. P.; Shirazi, B. “Performance evaluation of congestion-aware routing with DVFS on a millimeter-wave small-world wireless NoC”, *ACM Journal on Emerging Technologies in Computing Systems*, vol. 11–2, nov. 2014, pp. 17:1–17:22.
- [MKWS04] Magen, N.; Kolodny, A.; Weiser, U.; Shamir, N. “Interconnect-power dissipation in a microprocessor”. In: *Proceedings of the International Workshop on System Level Interconnect Prediction*, 2004, pp. 7–13.
- [MLC⁺14] Mohamed, M.; Li, Z.; Chen, X.; Shang, L.; Mickelson, A. R. “Reliability-aware design flow for silicon photonics on-chip interconnect”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22–8, ago. 2014, pp. 1763–1776.
- [MLW⁺14] Murray, J.; Lu, T.; Wettin, P.; Pande, P. P.; Shirazi, B. “Dual-level DVFS-enabled millimeter-wave wireless NoC architectures”, *ACM Journal on Emerging Technologies in Computing Systems*, vol. 10–4, jun. 2014, pp. 27:1–27:27.
- [MOM⁺14] Meade, R.; Orcutt, J. S.; Mehta, K.; Tehar-Zahav, O.; Miller, D.; Georgas, M.; Moss, B.; Sun, C.; Chen, Y.-H.; Shainline, J.; Wade, M.; Bafrafi, R.; Sternberg, Z.; Machavariani, G.; Sandhu, G.; Popović, M.; Ram, R.; Stojanović, V. “Integration of silicon photonics in bulk CMOS”. In: *Proceedings of the Symposium on VLSI Technology*, 2014, pp. 1–2.
- [MSG16] Mansoor, N.; Shamim, M. S.; Ganguly, A. “A demand-aware predictive dynamic bandwidth allocation mechanism for wireless network-on-chip”. In: *Proceedings of the 18th System Level Interconnect Prediction Workshop*, 2016, pp. 8:1–8:8.
- [NND10] Nicopoulos, C.; Narayanan, V.; Das, C. R. “Network-on-chip architectures: a holistic design exploration”. Springer, 2010, 223p.
- [OM06] Ogras, U. Y.; Marculescu, R. ““It’s a small world after all”: NoC performance optimization via long-range link insertion”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 14–7, jul. 2006, pp. 693–706.
- [OPZ11] Oh, J.; Prvulovic, M.; Zajic, A. “TLSync: support for multiple fast barriers using on-chip transmission lines”. In: *Proceedings of the 38th Annual International Symposium on Computer Architecture*, 2011, pp. 105–116.
- [PCMC15] Palesi, M.; Collotta, M.; Mineo, A.; Catania, V. “An efficient radio access control mechanism for wireless network-on-chip architectures”, *Journal of Low Power Electronics and Applications*, vol. 5–2, mar. 2015, pp. 38–56.

- [PF06] Pavlidis, V. F.; Friedman, E. G. “3-D topologies for networks-on-chip”. In: Proceedings of the International SOC Conference, 2006, pp. 285–288.
- [PGJ⁺05] Pande, P.; Grecu, C.; Jones, M.; Ivanov, A.; Saleh, R. “Performance evaluation and design trade-offs for network-on-chip interconnect architectures”, *IEEE Transactions on Computers*, vol. 54–8, ago. 2005, pp. 1025–1040.
- [PKK⁺09] Pan, Y.; Kumar, P.; Kim, J.; Memik, G.; Zhang, Y.; Choudhary, A. “Firefly: illuminating future network-on-chip with nanophotonics”. In: Proceedings of the 36th Annual International Symposium on Computer Architecture, 2009, pp. 429–440.
- [PL06] Puttaswamy, K.; Loh, G. H. “Thermal analysis of a 3D die-stacked high-performance microprocessor”. In: Proceedings of the 16th Great Lakes Symposium on VLSI, 2006, pp. 19–24.
- [Raz09] Razavi, B. “Design of millimeter-wave CMOS radios: a tutorial”, *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56–1, jan. 2009, pp. 4–16.
- [RDSZ16] Rezaei, A.; Daneshtalab, M.; Safaei, F.; Zhao, D. “Hierarchical approach for hybrid wireless network-on-chip in many-core era”, *Computers & Electrical Engineering*, vol. 51, abr. 2016, pp. 225–234.
- [RSdT14] Rezaei, A.; Safaei, F.; Daneshtalab, M.; Tenhunen, H. “HiWA: a hierarchical wireless network-on-chip architecture”. In: Proceedings of the International Conference on High Performance Computing Simulation, 2014, pp. 499–505.
- [SMsG17] Saxena, S.; Manur, D. S.; Shamim, M. S.; Ganguly, A. “A folded wireless network-on-chip using graphene based THz-band antennas”. In: Proceedings of the 4th International Conference on Nanoscale Computing and Communication, 2017, pp. 1–6.
- [SRD14] Samaiyar, A.; Ram, S. S.; Deb, S. “Millimeter-wave planar log periodic antenna for on-chip wireless interconnects”. In: Proceedings of the 8th European Conference on Antennas and Propagation, 2014, pp. 1007–1009.
- [VYM⁺14] Vijayakumaran, V.; Yuvaraj, M. P.; Mansoor, N.; Nerurkar, N.; Ganguly, A.; Kwasinski, A. “CDMA enabled wireless network-on-chip”, *ACM Journal on Emerging Technologies in Computing Systems*, vol. 10–4, jun. 2014, pp. 28:1–28:20.
- [WHB11] Wang, C.; Hu, W. H.; Bagherzadeh, N. “A wireless network-on-chip design for multicore platforms”. In: Proceedings of the 19th International Euromicro

Conference on Parallel, Distributed and Network-Based Processing, 2011, pp. 409–416.

- [WJ14] Wang, S.; Jin, T. “Wireless network-on-chip: a survey”, *The Journal of Engineering*, vol. 2014–3, mar. 2014, pp. 98–104.
- [WKM⁺14] Wettin, P.; Kim, R.; Murray, J.; Yu, X.; Pande, P. P.; Ganguly, A.; Heoamlan, D. “Design space exploration for wireless NoCs incorporating irregular network routing”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33–11, nov. 2014, pp. 1732–1745.
- [WMK⁺14] Wettin, P.; Murray, J.; Kim, R.; Yu, X.; Pande, P. P.; Heo, D. “Performance evaluation of wireless NoCs in presence of irregular network routing strategies”. In: Proceedings of the Design, Automation Test in Europe Conference Exhibition, 2014, pp. 1–6.
- [ZCS07] Zhang, Y. P.; Chen, Z. M.; Sun, M. “Propagation mechanisms of radio waves over intra-chip channels with integrated antennas: frequency-domain measurements and time-domain analysis”, *IEEE Transactions on Antennas and Propagation*, vol. 55–10, out. 2007, pp. 2900–2906.
- [ZW08] Zhao, D.; Wang, Y. “SD-MAC: design and synthesis of a hardware-efficient collision-free QoS-aware MAC protocol for wireless network-on-chip”, *IEEE Transactions on Computers*, vol. 57–9, set. 2008, pp. 1230–1245.
- [ZWLK11] Zhao, D.; Wang, Y.; Li, J.; Kikkawa, T. “Design of multi-channel wireless NoC to improve on-chip communication capacity”. In: Proceedings of the 5th International Symposium on Networks-on-Chip, 2011, pp. 177–184.



Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br