UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

DEPARTAMENTO DE ENGENHARIA ELÉTRICA (DEE)

[MOD-PH09] Fotônica para IA e Computação Analógica

Processamento Tensorial Óptico, Malhas MZI (Reck e Clements), SVD e Redes Neurais

Prof. Marcelo Eduardo Vieira Segatto

LabTel — Laboratório de Telecomunicações | IN-Foton





Sumário do Módulo [MOD-PH09]

1. Motivação e o Gargalo de Von Neumann
2. Célula Unitária MZI e Álgebra Unitária
3. Topologias de Malha Interferométrica: Reck vs Clements
4. MVM Arbitrária via Decomposição SVD
5. Arquiteturas WDM e Funções de Ativação
6. Síntese do Módulo e Conexão Smart-PSoC

Limitações dos Aceleradores Eletrônicos (GPU/TPU)

- **Crescimento Exponencial da IA Moderna:**
  - Modelos de linguagem (LLMs) e redes neurais profundas exigem $> 10^{20}$ operações de ponto flutuante (FLOPs) por treino.
  - A operação dominante ($> 85\%$ dos ciclos) é a **Multiplicação Matriz-Vetor (MVM)**: $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$.
- **O Gargalo de Memória e Energia de Von Neumann:**
  - Em arquiteturas digitais convencionais, os pesos são transferidos constantemente da memória DRAM para os núcleos aritméticos.
  - Custo energético do trânsito de dados: ~ 10 a $100\,\text{pJ}$ por acesso, superando em ordens de grandeza o custo da operação lógica em si ($< 0{,}1\,\text{pJ}$).
- **A Promessa da Computação Fotônica Analógica:**
  - A computação ocorre durante a propagação do pulso luminoso: latência sub-nanossegundo (tempo de voo, $\sim 10\,\text{ps/camada}$).
  - A multiplicação e a soma por interferência são passivas ($0\,\text{fJ/MAC}$).

O Interferômetro Mach-Zehnder 2×2 como Operador $SU(2)$

Uma célula MZI balanceada de 2 portas com dois defasadores térmicos ou eletro-ópticos (θ e ϕ) atua como rotação unitária em $SU(2)$:

Matriz de Transferência T_{MZI}

$$T(\theta, \phi) = j e^{j\frac{\theta}{2}} \begin{bmatrix} e^{j\phi} \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \\ e^{j\phi} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \end{bmatrix}$$

Controle Independente:

- θ modula a razão de divisão de amplitude (divisor variável 0 a 100%).
- ϕ ajusta o desfasamento óptico relativo entre as portas de saída.

Conservação de Energia

$$T^\dagger T = I_{2\times 2} \implies \det(T) = e^{j(\phi+\theta)}$$

Propriedades Notáveis:

- Operação estritamente unitária e passiva (sem perdas na hipótese ideal).
- Bloco básico para síntese de qualquer operador unitário arbitrário $U(N)$.

Decomposição de Matrizes Unitárias $U(N)$

Qualquer matriz unitária arbitrária de dimensão $N \times N$ pode ser sintetizada exatamente por uma malha de $N(N-1)/2$ células MZI:

Topologia Triangular (Reck et al., 1994)

- Arranjo triangular cascateado.
- Profundidade óptica assimétrica: varia de 1 a $2N-3$ MZIs.
- **Gargalo:** Desbalanceamento severo de perdas ópticas entre canais e dispersão de fase cumulativa desbalanceada.

Topologia Retangular (Clements et al., 2016)

- Arranjo retangular intercalado e periódico.
- Profundidade uniforme e mínima: exatamente N MZIs para todos os canais.
- Perdas ópticas simétricas e máxima tolerância à variação de processo (padrão em PICs).

Exemplo ($N = 4$): Uma transformação unitária 4×4 requer exatamente $\frac{4\times3}{2} = 6$ células MZI com profundidade óptica de 4 estágios na arquitetura Clements.

Síntese de Matrizes Arbitrárias via Decomposição SVD

Pesos reais/complexos arbitrários $\mathbf{M} \in \mathbb{C}^{M \times N}$ são sintetizados por **SVD**: $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}$.

Mapeamento em Cascata

1. **Malha Unitária $\mathbf{V}^{\dagger}$:** Malha Clements $N \times N$.
2. **Núcleo Diagonal $\mathbf{\Sigma}$:** Atenuadores VOAs ($\sigma_i \leq 1$).
3. **Malha Unitária U:** Malha Clements $M \times M$.

Computação sem Clock

- $\mathbf{x}$ codificado em amplitudes ópticas.
- Produto óptico instantâneo:

$$\mathbf{y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}\mathbf{x}$$

- Latência ultrabaixa de tempo de voo.

Processamento Neuronal WDM e Não-linearidades

Multiplicação WDM com Anéis (MRR)

- Vetor de entrada $\mathbf{x}$ modulado em múltiplos comprimentos de onda $(\lambda_1, \ldots, \lambda_N)$.
- Bancos de MRRs ponderam individualmente cada λ_k por sintonização térmica ou eletro-óptica.
- Fotodetector balanceado realiza soma coerente instantânea pela corrente fotogerada:

$$I_{\text{out}} = \sum_k w_k P_{\text{in}}(\lambda_k)$$

Funções de Ativação Não-Lineares

Redes neurais exigem não-linearidades para representação de funções complexas:

- **Não-linearidade OE-EO:** Fotodiodo → amplificador TIA → modulador não-linear MZM (Sigmoid/ReLU eletro-óptica).
- **Materiais de Mudança de Fase (PCM):** Células de GST ($Ge_2Sb_2Te_5$) com absorção saturável óptica.

Síntese do Módulo e Transição para [MOD-PH10]

Competências Consolidadas em [MOD-PH09]

- Vantagens energéticas e latência na computação óptica analógica.
- Operador unitário MZI e síntese retangular de Clements $N \times N$.
- Mapeamento de matrizes de pesos arbitrários via SVD óptica.
- Convolutores WDM baseados em bancos de ressonadores em anel.
- Alinhamento com os estudos neuromórficos do consórcio Smart-PSoC (Prof. Lorenzo Pavesi, Univ. Trento).

Conexão com o Próximo Módulo

No **[MOD-PH10] Sensores Fotônicos Integrados de Campo Evanescente**, exploraremos a alta sensibilidade refratométrica de guias e cavidades para diagnóstico molecular e biossensoriamento em tempo real.