高精度な行列演算に向けた光集積回路アーキテクチャの提案 An Integrated Photonic Architecture for High-Accuracy Matrix Multiplication

^o唐 睿⁻, 岡野 誠², トープラサートポン カシディット⁻, 高木 信一⁻¹, 竹中 充⁻¹ (¹東大院工, ²産総研)

[°]Rui Tang¹, Makoto Okano², Kasidit Toprasertpong¹, Shinichi Takagi¹, Mitsuru Takenaka¹ (¹The University of Tokyo, ²AIST)

E-mail: ruitang@mosfet.t.u-tokyo.ac.jp

1. 背景

光集積回路(PIC)を用いると高速かつ低消費電力 な行列演算が実行可能であることから、深層学習 への応用が活発に研究されている。従来の PIC 構 造として、マッハ・ツェンダー干渉計(MZI)に基 づいたユニバーサル光回路及びマイクロリング 共振器(MRR)に基づいた光クロスバーアレイな どが提案・実証された[1,2]。しかし、いずれの構 造も、回路規模が大きくなると計算エラーが増加 するため、高精度な行列演算を行うことが困難で ある[3]。本報告では、高精度な行列演算に向けた 新しい PIC 構造を提案する。

2. 提案構造

我々が提案する構造を図1に示す。単一波長の光 を入力させ、入力側の変調器アレイで各導波路内 の光パワーをベクトルの各要素に比例して変調 する。次に、二層目の導波路を経由して行列用変 調器アレイで再び変調し、ベクトルと行列の各要 素の掛け算を行う。二層導波路構造を使うことで 導波路交差がなくなり、各経路の損失差やクロス トークを低減できる。最後に、二回変調された光 パワーを多ポート光受光器(PD)で一括受信し、出 力電流を読み取ることで足し算が実行され、行列 演算の結果が得られる。ここで、相変化材料など の可変光吸収体を変調器として使うことも可能 である。図2に多ポート Ge PD の設計及び計算 上の特性を示す。100 ポート程度の PD サイズで は、行列演算にとって十分な数 GHz の応答速度 と100 nA以下の暗電流は容易に実現可能である。 3. ハードウェアエラー

本構造の行列変調領域で生じたエラーとユニバ ーサル光回路で生じたエラーを比較した。このエ ラーは式(1)に示す行列のフロベニウスノルムよ り計算できる。

$$\varepsilon = \frac{\|\mathbf{M} - \mathbf{W}\|}{\|\mathbf{M}\|} \tag{1}$$

ここで、Mは目標の行列、Wは実際に得た行列で ある。MZI 変調器を用いた場合、位相シフトの量 子化によるエラー(ϵ_0)と光スプリッタ分岐比の誤 差によるエラー(ϵ_s)が主なエラーになる。図3に 異なる規模のデバイスにおける ϵ_0 と ϵ_s の計算結 果を示す。各点はランダムなM(2500個)を実行 した場合の平均エラーであり、エラーバンドはエ ラーの最大値と最小値の範囲を表す。ここで、位 相量子化の精度は12bitを想定した。また、各光 スプリッタにおいて、分岐比が理想値との偏差は 正規分布 $\mathcal{N}(0, \sigma^2)$ に従うことを想定した。図3よ り、ユニバーサル光回路ではエラーが MZIの数 に比例して増えていくことに対して、本構造では 平均エラーが MZI の数に依存しないことが分か る。これは、各 MZI で生じたエラーが、ユニバ ーサル光回路では後ろのステージにも影響する 一方、本構造ではエラーが他の素子に影響しない ことに起因している。

4. 結論

二層導波路構造と多ポート PD を用いて、高精度 な行列演算に向けた新規な PIC 構造を提案した。 回路規模が大きくなってもハードウェアエラー が増えないことを明らかにした。



Fig. 1. Proposed structure for matrix-vector multiplication.



Fig. 2. (a) Conceived design for the multiport PD. (b) Estimated 3-dB bandwidth and dark current.



Fig. 3. Hardware errors of this proposal and the universal unitary multiport interferometer (UUMI).

謝辞 本研究の一部は JST、CREST、JPMJCR2004 の支援を受けて実施した。

参考文献

- [1] W. R. Clements et al., Optica 3, 1460-1465 (2016).
- [2] S. Ohno et al., arXiv:2106.04351 (2021).
- [3] S. Bandyopadhyay et al., Optica 8, 1247-1255 (2021).