Vol.2019-ARC-237 No.28
2019/7/26

光アプロキシメートコンピューティングの実現に向けた電力性能解析

川上 哲志[1,a)]　谷本 輝夫[2]　北 翔太[3,4]　新家 昭彦[3,4]　小野 貴継[1]　納富 雅也[3,4]　井上 弘士[1]

概要：ナノフォトニックデバイスは光を情報媒体として扱う性質上，極めて突出した利点（極低レイテンシ，低損失、高い耐雑音性，および，広い帯域幅）を有する反面，集積度の低さや電/光・光/電変換（電力・時間）コストの高さといった問題がある．このような特異なデバイスを活用し効率的なコンピュータシステムを構築するには，デバイスの利点の最大化と欠点の隠蔽をシステムレベルで達成する必要があり，デバイス・アーキテクチャ・ソフトウェアを協調設計が極めて重要となる．本稿では，効率的な光コンピュータの実現に向けて，アプロキシメートコンピューティング応用に着目し光アナログ行列ベクトル乗算器のスループット，消費電力，計算精度の性能モデルを示す．性能モデルベースの評価結果は，光コンピュータの電力効率限界のみならず，どのパラメータが電力効率に大きく寄与しているかも示唆する．さらに，光コンピュータのハードウェアレベルのノイズが，アプリケーションレベルの最終出力精度に与える影響をシミュレーションで評価する．これらの結果と性能モデルは，光コンピュータの電力効率と精度の間の複雑なトレードオフを明確化し，システム階層を跨いだ協調設計の指針となる．

1. はじめに

アプロキシメートコンピューティングは，計算機システムの電力効率，すなわち 1 ワット当たりの性能を最大化するための手法として注目を集めている．ターゲットアプリケーションの QoS（Quality of Service）を満たす制約下で計算エラーを受け入れることは，冗長な操作の排除を可能にし，その結果，著しい性能改善と電力削減をもたらす．さらに，この計算パラダイムと Memristor [16] や RERAM [4] といった新奇デバイスに基づく超効率的なアナログ演算機を組み合わせることで，ニューラルネットワーク処理の高速化が提案されている．基本概念は，QoS 基準内で計算精度のわずかな低下を許容することによってノイズ効果によるハードウェア不安定性を補償することである．このような新奇デバイスを活用したアプロキシメートアナログコンピューティングは，従来の CMOS デジタル計算機の限界を突破する大きな可能性を秘めている．

1 九州大学 大学院システム情報科学研究院
Faculty of Information Science and Electrical Engineering, Kyushu University
2 九州大学 情報基盤研究開発センター
Research Institute for Information Technology, Kyushu University
3 NTT ナノフォトニクスセンタ
NTT Nanophotonics Center
4 NTT 物性科学基礎研究所
NTT Basic Research Laboratories
a) satoshi.kawakami(at)cpc.ait.kyushu-u.ac.jp

ナノフォトニックデバイスの活用，すなわち，光アプロキシメートコンピューティング（OAC: Optical Approximate Computing）は，計算機の電力効率を劇的に改善するための有望なアプローチの一つである．近年のナノフォトニクス技術の発展により，フォトニック結晶に代表される光の伝搬を制御可能な物質の大規模集積化，小型化，低消費エネルギー化が可能になりつつある [10]．ナノフォトニック・デバイスは，かつては困難であった光の波長程度の極微小領域への光の閉じ込めを可能にすることが報告されており [21]，今後も CMOS トランジスタと同様の割合で発展することが予想されている [10]．また，近年小型・高速・超低電力な光検出器も提案されており，光/電気変換の効率も劇的に改善されている [11]．これらの革新技術により，高性能・低電力な光コンピュータが実現可能になりつつある．

本研究では, CMOS コンピュータを凌駕する光コンピュータの実現を目指す．特に，本稿ではニューラルネットワークに代表されるように多くの情報処理の中核演算となるベクトル行列積（VMM: Vector-by-Matrix Multiplier）に焦点を当てる．光 VMM のデバイスレベルの実装は活発に研究されてきた [17], [20] が，OAC の実現可能性を探るためにデバイス，マイクロアーキテクチャ，およびアプリケーション特性間の相互作用を考慮した論文はない．主な課題は，ハードウェアレベルの性能予測とアプリケーションレベルの QoS 分析とを橋渡しすることであり，本研究の主な

© 2019 Information Processing Society of Japan

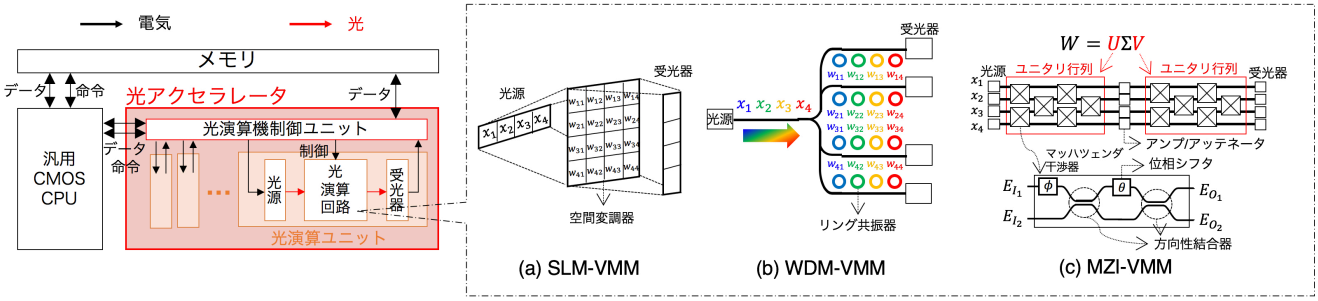


図 1: 光コンピュータ概要と光ベクトル-行列積演算回路（VMM）

貢献点は以下の通りである．

- ノイズの影響を考慮したパフォーマンスモデルを提案し，それにより VMM ベースの OAC のスループット，電力，および精度を分析することを可能にする．提案モデルは我々の分析の基礎である．本稿では，VMM のみをターゲットにしているが，他の光アナログ計算ユニットにも簡単に拡張可能である．
- 光 VMM の理論ピーク電力効率と精度のトレードオフを分析することで，デバイスパラメータ（変調器の透過率）とアーキテクチャパラメータ（VMM の規模）の協調設計が重要であることを明らかにした．さらに，最適な協調設計により光 VMM は，理論ピークで 256TOPS/W が達成できることも確認した．
- OAC アプリケーションとして，ニューラルネットワーク，Sobel フィルタ，高速フーリエ変換（FFT）をターゲットとし，アプリケーションレベルでの QoS 評価を示す．その結果，アプリケーションレベルでの精度劣化がほとんど無しに電力効率を 247[TOPS/W] から 1480[TOPS/W] まで向上できることが分かった．

2. 光コンピュータ

光アプロキシメートコンピューティングの実現に向けて，本研究で想定する光コンピュータを図 1 に示す．光演算機での汎用処理やメモリ回路の実現の困難性を鑑み，ディスクリートアクセラレータとして従来の CMOS コンピュータシステムに組み込むことを前提とする．光アクセラレータは，光演算機制御ユニットといくつかの光演算ユニットで構成される．光演算機制御ユニットは，ホスト CPU またはメインメモリから読み出される命令およびデータを光演算ユニットへ提供し管理する．光演算ユニットは，光源（LS: Light Source），光子を運ぶための導波路，光演算回路を構成する変調器，光波を検出するための受光器（PD: Photo Detector）で構成される．本研究で着目する光演算ユニットは光アナログ VMM とする．ベクトル行列積は，あらゆる種類の情報処理，特にニューラルネットワークに代表されるアプロキシメートコンピューティングアプリケーションの基本演算である．本セクションでは，光アナログ VMM を実現する 3 つの回路を説明する．

2.1 SLM-VMM

1 つ目の構成は空間光変調器（SLM: Spatial Light Modulator）を活用した VMM である [2]．図 1(a) に示すように，VMM の入力ベクトルは 1 組の LS によって表され，VMM の行列は SLM によって表され，VMM の出力ベクトルは 1 組の PD によって表される．各 LS からの自由空間伝播光は，マトリックス内の単一の列を照らすようにレンズによって垂直に分光され，次にマトリックス内のそれぞれの行がレンズによって検出器アレイ内の単一の検出器に集光される．SLM-VMM は自由空間伝搬光を利用するため，光演算回路のサイズは大きくなる．しかしながら，近年プレーナ（平面）集積が可能な SLM-VMM も提案されている [6]．平面集積は，自由空間光学システムを電気 VLSI 技術との統合を可能にする．入力値は光の強度レベルを用いて符号化され，SLM のセルを通過することでその強度は任意の割合で弱められる．これは，入力ベクトルの要素と行列要素の積を意味する．したがって，SLM-VMM では，入力値 x は $x \leq 0$ に制限され，行列要素は $0 \leq w \leq 1$ に制限される．SLM は，液晶ディスプレイ（LCD: Liquid Crystal Display）や微小電気機械システム（MEMS: Micro-Electro-Mechanical systems）技術に基づく光学機械マイクロミラーデバイスによって実現できる．

2.2 WDM-VMM

2 つ目の構成は，波長多重（WDM: Wavelength Division Multiplexing）を活用した VMM である [20]．図 1(b) に示すように，VMM の入力ベクトルは LS からの波長多重光で表され，リング共振器が各々 VMM の行列要素を表し，PD によって検出される光強度が VMM の出力ベクトルを表す．入力ベクトルの要素は，波長多重化された固有の波長キャリアを割り当てられる．波長多重信号は，カプラによって均等な光強度で分配され，各行列要素値を示すリング共振器で変調される．リング共振器は導波路の隣に配置されて特定の波長の光のみを変調するため，ある入力ベクトルの要素とある行列要素の積を実現できる．リング共振

© 2019 Information Processing Society of Japan

器を通過した波長多重光は，波長に依存せず総信号強度がPD によって検出されるため，これが和演算と同等となる．入力値は固有の波長を有する光の強度レベルを用いて符号化され，光信号は対応するリング共振器によって任意の強度へ弱められる．したがって，WDM-VMM でも入力値 x は $x \leq 0$ に制限され，行列要素は $0 \leq w \leq 1$ に制限される．リング共振器による特定の波長信号の損失率は，リングに電荷を注入するか，その温度を変えることで調整できる．

2.3 MZI-VMM

３つ目の構成は，マッハツェンダ干渉器（MZI: Mach-Zehnder Interferometer）を活用した VMM である [17]．図 1(c) に示すように，MZI は２つの方向性結合器（DC: Directional Coupler）と２つの位相シフタ（PS: Phase Shifter）で構成される．DC は２つの入力信号を各々均等な光強度で２つの出力ポートへ分配し，PS は信号光波の任意の移相変化が可能である．PS の移相量（ϕ,θ）は，キャリア注入もしくは温度制御によって調節できる．MZI の伝達関数は次のように表すことができる．

$$\begin{pmatrix} E_{O_1} \\ E_{O_2} \end{pmatrix} = e^{i(\frac{\theta+\pi}{2})} \begin{pmatrix} e^{i\phi}\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \\ e^{i\phi}\cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \end{pmatrix} \begin{pmatrix} E_{I_1} \\ E_{I_2} \end{pmatrix} \tag{1}$$

ここで，E_{O1}, E_{02} と E_{I1}, E_{I2} は，各々出力信号と入力信号の電界振幅である．MZI は，（ϕ,θ）を調整することにより各入力ポートの光信号強度を２つの出力ポートに任意の比率で分配することができる．Reck や Clements は，この MZI を活用し任意の $N \times N$ ユニタリ変換を実現できることを示した [5], [14]．Clements の実装は Reck のものと比較して，より高い忠実度，優れたノイズ耐性，そして小さなサイズでの実現が可能である．これらの違いは，MZI の配置によるものである．Reck の回路では三角形型に，Clements の回路では長方形型に MZI を配置（図 1）する．詳細な動作原理と数学的証明は文献 [5], [14] を参照されたい．図 1(c) に示すとおり，VMM は 2 つの MZI ベースのユニタリ変換回路と 1 組のアンプまたはアッテネータで構成される．これは，任意の行列を特異値分解することで２つのユニタリ行列（S,V）と対角行列（Σ）に分解することができることに対応する．すなわち，対角行列（Σ）はアンプまたはアッテネータで実現される．

3. 光コンピュータの電力性能モデル

本節では，光 VMM の性能モデル，すなわちスループット，消費電力（電力効率），演算精度モデルについて述べる．性能モデリングのターゲットシステムは，図 1 に示した光演算ユニットとする．これは，LS，導波路，VMM，および PD で構成される光信号で閉じた系である．まず，VMM の電力効率モデルを 3.1 節で説明し，演算精度モデルについては 3.2 節で述べる．

3.1 電力効率モデル

電力効率のモデルについて，スループットと消費電力を個別に説明する．なお，著者らは MZI-VMM の性能モデルについて文献 [24] で提案しており，本節ではその内容を簡単にまとめる．ここでは，スループットを 1 秒あたりの積和演算の数として定義する．VMM は 2.1 節-2.3 節に示すようにさまざまな構成の下で動作するが，すべての VMM のスループットは次のように表すことができる．

$$Throughput = N^2 f_{VMM} \tag{2}$$
$$f_{VMM} = min(f_{LS}, f_{PD}, f_{mod}, 1/L)$$

ここで，N は VMM サイズ（$N \times N$ 行列），f_{VMM} は VMM の周波数を表す．VMM 回路の周波数 f_{VMM} は，光源の周波数 f_{LS}，受光器の周波数 f_{PD}，変調器の周波数 f_{mod}，もしくは VMM 回路のレイテンシの逆数 $1/L$ の何れかによって律速される．変調器の周波数 f_{mod} は，SLM-VMM 中の SLM 特性，WDM-VMM 中のリング共振器特性，および MZI-VMM 中の PS 特性に依存する．VMM 回路のレイテンシ L は，回路を構成するデイバイス長や VMM 構成に依存する．詳細は，文献 [24] を参照されたい．なお，本モデルは L のみが VMM の構成に依存するため，レイテンシモデルの変更により他の光アナログ演算回路用の性能モデルに容易に拡張可能である．

エネルギー消費は，基本的に系のエンドポイント（PD）と変調器で発生する．LS から放射された光信号は PD に到達するが，計算のために変調器によって失われる．そして，PD に到達した光信号は電流値に変換される．計算データに関係なく，PD での光信号のエネルギーが 100%失われる（実際には変換効率 < 30%程度）と仮定すると，VMM の消費電力は次のように表される．

$$Power = NP_{LS} \tag{3}$$

ここで，N は VMM サイズ（$N \times N$ 行列），P_{LS} は LS からの出力光の強度である．この性能モデルは，エネルギーを追加するアクティブデバイスが存在しない LS から PD までの閉じた系で使用できる．MZI-VMM がアッテネータで構成されている場合は，2 節のすべての VMM をアクティブデバイスなしで実現できる．

式 (2) および (3) に基づき，光 VMM の電力効率は以下のように導出される．

$$Power\ efficiency = \frac{Nf_{VMM}}{P_{LS}} \tag{4}$$

3.2 演算精度モデル

本節では，光 VMM の演算精度モデルについて説明する．精度は，出力信号が目的の量子化範囲内に収まる確率として定義する．精度が無視できれば電力効率は際限なく向上する事ができるため，ノイズは電力効率を制限する非常に

© 2019 Information Processing Society of Japan

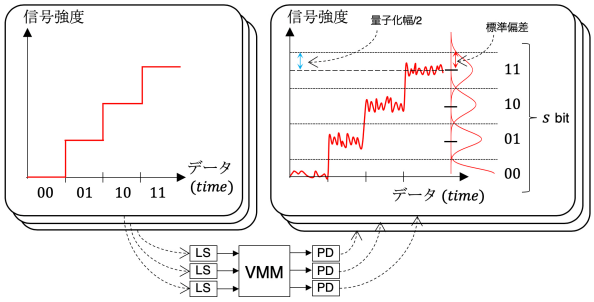


図 2: 光信号の s ビット量子化

重要な要素となる．なお，著者らは VMM ノイズモデルについては，文献 [25] で提案しており，その詳細については本稿では省略する．

本システムの入力信号と出力信号の分解能は s ビットとし，最大光信号で PD に入射する光子の平均数を N_{photon} とする．s ビットに相当するアナログ信号を表すためには，PD に到着する光子数を幅 $\Delta = N_{photon}/(2^s - 1)$ 毎に離散数値化する．つまり，光演算ユニットの演算精度を定義するためには，PD での光子数が所望の量子化幅に収まる確率を考えれば良い．PD に到達する光子数は，ノイズやロスの影響で量子化幅を超え（下回り）誤った値として認識される可能性がある．したがって，演算精度を向上させるには LS の光強度を増加させ，N_{photon} を増加させる必要がある．

LS から照射された光信号は，図 2 に示すように何らかのノイズによる変動を伴って PD に到達する． 1 光子が PD に到達し，η（量子効率）個の電子が励起されると，PD の最大平均電流は $I_{max} = N_{photon}\eta q$（$q$: 電気素量）となる．したがって， PD の電流信号分布の標準偏差の α 倍までが量子化範囲内に収まることを保証するためには，次の式を満たす必要がある．

$$\frac{1}{2}\frac{I_{max}}{(2^s - 1)} \geq \alpha\sigma \tag{5}$$

ここで，左側は量子化範囲/2（図 2 の青い双方向矢印），σ はノイズによる電流分布の標準偏差（赤い双方向矢印）であり，α は演算精度係数として定義する．σ は，信号光波のショットノイズ，PD の暗電流によるショットノイズ，PD の熱雑音，変調器の制御信号ノイズに依存し，その詳細は文献 [25] を参照されたい．光 VMM 電流分布が正規分布であり，式 (5) の右左辺が等しいとすると，VMM の演算精度は図 3 に示す通り α で規定できる．つまり，式 (2) を満たす場合，α に応じて図 3 に示す VMM の演算精度を最低限保証することができる．

PD における最大平均電流（I_{max}）は次式で表される．

$$I_{max} = N_{photon}\eta q = \frac{P_{LS}T_{VMM}X}{hc/\lambda}\eta q \tag{6}$$

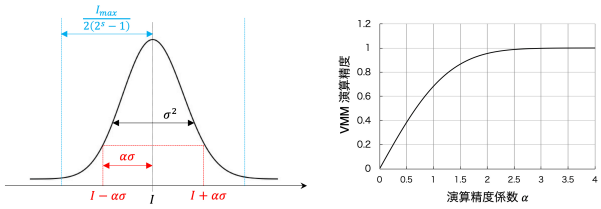


図 3: VMM 演算精度と演算精度係数 (α) の関係

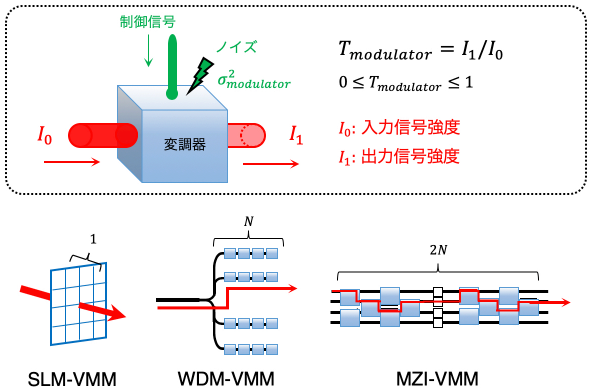


図 4: VMM における変調器の制御信号ノイズと透過率

$$T_{VMM} = \begin{cases} T_{mod} & \text{(SLM-VMM)} \\ T_{mod}^{N} & \text{(WDM-VMM)} \\ T_{mod}^{2N} & \text{(MZI-VMM)} \end{cases} \tag{7}$$

ここで，P_{LS} は LS からの光信号の最大強度，T_{VMM} は VMM 透過率，X は電流増倍係数（アバランシェフォトダイオードのみ），h はプランク定数，c は光速，λ は光信号の波長である．$P_{LS}T_{VMM}X$ は PD に到達した信号の最大電力であり，hc/λ は光子 1 個のエネルギーであるため，これらの商は単位時間あたりの光子数 N_{photon} を表す．T_{VMM} は，VMM 構成に依存するため，式 (7) に示すように，VMM にて通過する変調器の段数に依存する．図 4 は，$N \times N$ の VMM についての各最長パスを示す．T_{mod} は入出力強度比なので，N 段変調器を通過したときの透過率は T^N となる．

4. モデルに基づく電力性能解析

本設では，前節で示した性能モデルに基づいて電力効率と精度を分析する．まず，MZI-VMM に着目しハードウェアの観点から評価を実施する．4.1 節で，MZI-VMM の潜在的な電力効率を評価し，また，どのデバイスパラメータが性能に大きく寄与するのかを解析する．これにより，デバイスとアーキテクチャの最適化を目指す．次に，アプロキシメートコンピューティング向けのアプリケーションを対象とし，アプリレベルのエラーを評価する．4.2 節では，アプリレベルのエラーと演算精度の関係を明らかにし，OAC の電力性能について解析することでデバイス～アプ

© 2019 Information Processing Society of Japan

表 1: 固定パラメータ

q [C]	電気素量	$1.60 * 10^{-19}$	-
h [Js]	プランク定数	$6.63 * 10^{-34}$	-
c [m/s]	光速	$3.00 * 10^{8}$	-
k [J/K]	光の波長	$1.38 * 10^{-23}$	-
T [K]	絶対温度	300	-
- [-]	屈折率	3.47	[10]
λ [μm]	光の波長	1.55	[10]
η [-]	量子効率	0.7	[12]
R_L [Ω]	PD 負荷抵抗	50	[12]
I_d [nA]	PD 暗電流	10	[12]
M [-]	電流増倍係数	1	-
x [-]	過剰雑音指数	1	-
σ^2_{mod} [-]	PS の制御信号ゆらぎ	10^{-15}	-
α [-]	演算精度係数	3	-
s [bit]	ビット数	8	-

リケーションの最適化を目指す．

4.1 VMM のハードウェアレベル性能評価

4.1.1 評価概要

前節にて，電力効率と演算精度モデルを構築したが，これらを個別に評価してもこの 2 つの指標のトレードオフ関係は明らかとならない．そこで，2 つのモデルを統合的に扱い光 VMM のハードウェアの潜在的な電力効率を見積もる．特定の精度を保証する条件下で電力効率の最大値を分析するために，式 (5) に式 (6) を代入し下記のように整理する．

$$\frac{N f_{VMM}}{P_{LS}} \leq \frac{N q \eta T_{VMM} X f_{VMM}}{2\alpha\sigma(2^s - 1)hc/\lambda} \tag{8}$$

ここで，左辺は式 (4) に示すとおり VMM の電力効率を表す．つまり，右辺が特定の演算精度を保証する際の電力効率の上限となる．

本評価では，現在のナノフォトニクス技術（Current パラメータ）と将来技術（Future パラメータ）との比較により，どのパラメータが電力効率に大きく寄与しているかを明らかにする．なお，同目的の評価は文献 [25] でも実施しているが，本稿では図 2 で示す入出力信号のデコード・エンコードの整合性を考慮したことで，モデル式 (6) ならびに評価結果が異なることに注意されたい．式 (8) に含まれるすべてのパラメータを表 1 と表 2 に示す．表 1 は，固定パラメータである．VMM の光信号の波長は，ファイバー通信でもよく使われる 1.55μm である．受光器は，PIN-PD（すなわち，$M = 1, x = 1$ を意味する）を想定し，環境温度は 300[K] とする．受光器の詳細スペックは [12] を参照されたい．PS の制御信号ノイズ σ^2_{mod} は信号発生器（MP1800A [23]）とスペクトラムアナライザ（MS2692A [24]）を用いて実験的に計測した．最後に，本入出力信号は 8bit であり，

表 2: 可変パラメータ

Symbol	Description	Current	Future	Ref.
L_{mod} [ps]	MZI のレイテンシ	1	0.01	[10]
f_{mod} [GHz]	PS の周波数	12.5	100	[1]
NP_{LS} [dBm]	全 LS の最大電力	−30 to 2	−40 to 20	fig. 5
f_{PD} [GHz]	PD の周波数	40	100	[12]
N [-]	$N \times N$ VMM サイズ	6	1,024	[18]
T_{mod} [-]	MZI の透過率	0.9	0.99	[22]

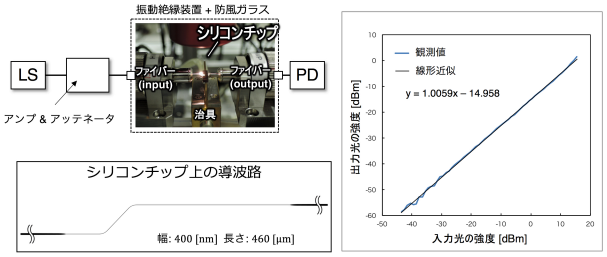


図 5: 導波路の線形特性範囲

$\alpha = 3$（演算精度は，99/3%）であると仮定する．

表 2 は，将来のナノフォトニック技術で改善されると想定される可変パラメータである．現在のパラメータ値は各々の文献を参照されたい．将来のパラメータ値は，すべてについて正確な証拠を示すわけではないが，非現実的な値は設定していない．具体的には，MZI の素子長は光波長とほぼ同じサイズとなった場合を想定し [21]，文献 [23] では f_{PD} と f_{mod} が 100GHz を超える可能性が示唆されており，$N \times N$ に関しては 4092 デバイスでの実装が報告されており [18]，T_{mod} は文献 [19] で 0.97 を確認している．仮に，将来パラメータ値が達成不可能であっても，電力効率向上との相関を分析するのに十分である．LS から照射される光信号の総電力は，導波路の 2 光子吸収と呼ばれる現象によって制限される．2 光子吸収は導波路の線形特性を奪い，これは導波路の機能が失われることを意味します．2 光子吸収が発生しないダイナミックレンジは，図 5 に示すように測定により設定する．なお，本測定においては，LS，PD，アンプ，および光強度を調整するためのアッテネータは，導波路が実装されているチップの外側に配置される．そのため，ファイバとチップ間の接続損失が大きい（約 15dBm）が，線形特性のレンジは十分に確認できる．

4.1.2 電力効率分析

本節では，VMM ハードウェアレベルの電力効率評価結果について説明する．表 3 は，表 2 のパラメータを変えたときの理論ピーク電力効率を示す．将来のパラメータで評価する場合は，∘ が各パラメータのセルにマークされており，無印の場合は現行パラメータである．No. 1 は，すべてのパラメータに対して現行パラメータ値を使用しており，正規化のためのベースラインである．No. 2〜7 では電力効率はそれほど改善されないため，各パラメータを個別に改善しても電力効率の改善にはそれほど寄与し無いことが

© 2019 Information Processing Society of Japan

表 3: MZI-VMM の理論ピーク電力効率

Configuration No.		1†	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Future パラメータか？	L_{mod}		○						○			○		○	○	○	○
	f_{mod}			○					○				○	○	○	○	○
	NP_{LS}				○					○					○	○	○
	f_{PD}					○				○					○	○	○
	N						○				○	○	○	○	○		○
	T_{mod}							○			○	○	○	○		○	○
電力効率 [TOPS/W]		8.64	8.64	15.5	8.64	8.64	8.64	26.3	15.5	8.64	87.3	88.8	94.4	161	24.5	74.7	256
正規化電力効率 † Baseline		1	1	1.8	1	1	1	3	1.8	1	10.1	10.3	10.9	18.7	2.8	8.6	29.7

分かる．No. 8〜10 にて VMM サイズと透過率を同時に向上させると，電力効率が大幅に向上することが分かる．また，No.16 と No.14 または No.15 との比較では，N または T_{mod} のみが現行パラメータ値である場合に電力効率が大幅に制限されていることが分かる．これは，N と T_{mod} が電力効率改善のキーパラメータであることを意味する．また，N と T_{mod} が将来パラメータである際には，L_{mod} と f_{mod} の改善が VMM の電力効率に効果的であることを示している（No. 10〜13）．

VMM サイズ N と透過率 T_{mod} が重要なパラメータである理由を明確にするために，図 6 に N と T_{mod} に対する電力効率の相関を示す．y 軸は電力効率，x 軸は VMM サイズである．そして凡例はそれぞれの透過率の値を示す．T_{mod} が向上するにつれて，電力性能のピーク（赤い菱形で表示）は右にシフトする．基本的に N の増加に応じて，スループットは N^2 で増加し，電力は N で増加する．これは N に比例して電力効率が改善されることを意味する．ただし，光信号強度は通過する変調器の段数に応じて指数関数的に減衰するため，演算精度の低下が予想される．演算精度をある一定以上に保つためには，LS の光信号強度を上げる必要がある．したがって，N が大きい地域では，精度を担保するために多くの電力が使用されるため電力効率が低下する．

実用的な光アナログ VMM を実現するためには，変調器の透過率を優先的に向上させ，適切な VMM サイズを選択することが最も重要である．ナノフォトニクスでは，微細化に多くの労力が注がれているが，それだけに焦点を当てると光コンピュータの優位性を失う危険性がある．前述したように，透過率 T_{mod} を改善し N が適切に設計された後であれば，微細化によるレイテンシ L_{mod} の減少が効果的となる．つまり，VMM の電力効率改善という観点において，パラメータ改善の優先順位が分かる．

これらの考察に基づき適切なデバイスの改善やアーキテクチャ設計ができれば，光コンピュータが CMOS コンピュータを凌駕する可能性がある．最新の CMOS ニューラルネットワークアクセラレータの電力効率は，TPU@TDP では 1.23[TOPS/W] [7]，ISAAC では 381[GOPS/W] [16]，DaDianNao では 286[GOPS/W] [3] である．表 3 の現行パラメータ（No. 1）における電力効率 8.64[TOPS/W] でも，これらの CMOS アクセラレータより優れていることが分かる．ただし，CMOS アクセラレータは，メモリ，I/O などの周辺機器を含めた消費電力による電力効率の見積もりであり，また，ISAAC と DaDianNao は 16 ビット積和演算を実行するため，これは公平な比較では無いことに注意されたい．しかしながら，光 VMM でも，十分に高い電力効率を期待できることは明らかである．

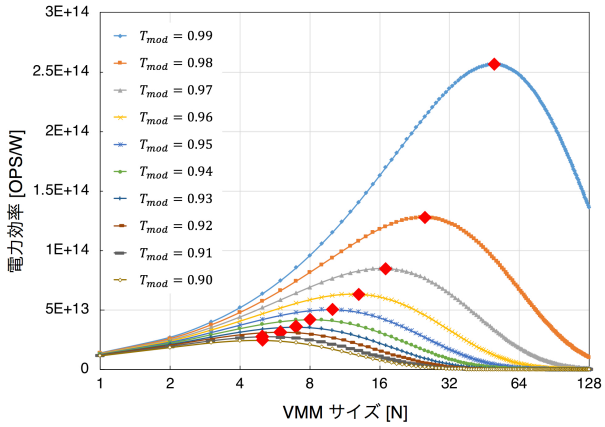


図 6: N と T_{mod} 電力効率の相関

4.2 アプリケーションレベル性能評価

4.2.1 評価概要

本節では，アプロキシメートコンピューティング向けのいくつかのベンチマークを使用して，光 VMM の演算結果の不正確さに対してアプリケーションレベルの精度がどれだけ許容できるかを評価する．表 4 は，アプロキシメートコンピューティング研究で一般的に使われ，かつ，その主要演算がベクトル行列積であるベンチマークアプリケーションの一覧である．ベンチマークのベクトル行列積の出力の各要素は 8 ビット量子化されており，演算精度係数 α に基づいて演算結果に誤差が生じる．アプリケーションの最終出力のエラーは，表 4 に示す各アプリケーション固有のエラー指標に基づいて算出する．つまり，各ベンチマークのベクトル行列積においてノイズなし出力とノイズあり出力を比較しアプリケーション精度を算出する．たとえば，エッジ検出を行うソーベルフィルタでは，ノイズがある場合とない場合とで異なる出力（エッジか否か）となるピクセ

表 4: ベンチマークアプリケーション

	アプリケーション	データセット	詳細	エラー指標
ニューラルネットワーク	MLP: 分類問題	MNIST [7]	ネットワークアーキテクチャ 784→50→10	分類ミス率
		Fashion-MNIST [7]	784→50→10	分類ミス率
		CIFAR-10 [8]	3072→500→10	分類ミス率
	RNN: 64bit-シリアル加算器	64bit 乱数	2→4→1 64bit全て一致: ヒット その他: ミス	ミス率
ソーベルフィルター: エッジ検出		256×256-pixel 画像	カーネルサイズ= 3 出力値 ≥ 127: エッジ その他: エッジでない	ミス率
FFT-1D: 信号処理		256 乱数	サンプリング数= 256 サンプリング周波数= 100	平均絶対誤差

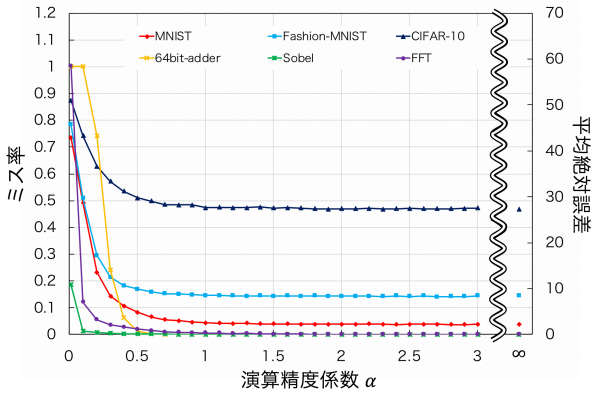


図 7: アプリケーションレベルの精度と演算精度係数 α

ルの割合で「ミス率」を計算する．ニューラルネットワークでは，入力データセットとトレーニングデータセットは異なる．64 ビットシリアル加算器の場合は，トレーニングと推論で異なるランダム入力セットを使用する [8], [9].

4.2.2 アプリケーションレベルの精度評価

演算精度係数 α に対するアプリケーションレベルの精度の相関を図 7 に示す．FFT-1D の精度（平均絶対誤差）は右軸に示され，他のアプリケーションの誤差（ミス率）は左軸に表す．$\alpha = \infty$ は，VMM 演算出力に全く誤差が無いことを意味する．アプリケーションレベルの精度は，$\alpha = \infty$ から $\alpha = 0.5$ までの精度とそれほど変わらないことが確認できる．つまり，これらのアプリケーションでは，OAC で $\alpha = 0.5$ までの誤差を許容できることを意味する．

次に，VMM の電力効率と演算精度 α のトレードオフ関係を図 8 に示す．黒い線は，表 3 の No. 16 に示す通り全て将来パラメータ値である電力効率曲線を示している．なお，表 3 の結果は，$\alpha = 3$（演算精度 99.7 %）上の線に対応する．つまり，前節で述べたとおりデバイスパラメータ（特に T_{mod}）を改善し，適切な VMM サイズ N を選択することで，図 8 に示す通り電力効率を上に移動（改善）することができる．さらに，図 7 により，OAC 向けアプリケーションは $\alpha = 0.5$ までの誤差を許容できることが判明した．よって，図 8 により $\alpha = 3$ から $\alpha = 0.5$ に変更したとしても，アプリレベルでの精度劣化はほぼ無く電力効率を 247TOPS/W から 1480TOPS/W に向上させることができる．演算精度係数 α に関して，長距離光ファイバ通信では

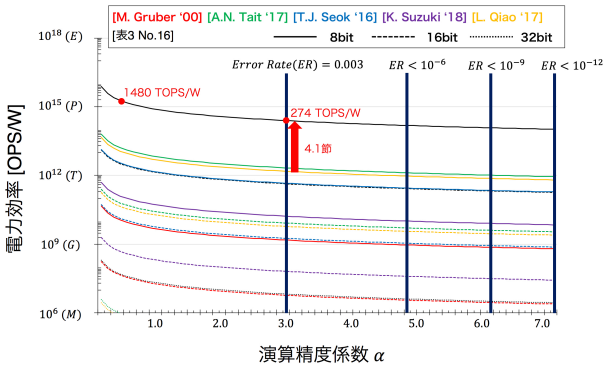


図 8: VMM の電力効率と演算精度係数 α の相関

10^{-12}〜10^{-9} 以下（$6.1 < \alpha < 7.1$），電話回線では 10^{-9} 以下（$5.8 < \alpha$）のエラーレートを要求される．しかしながら，OAC では非常に高い演算誤差耐性があることと，ならびに，誤差許容により極めて大きな電力効率改善の可能性があることが分かる．

また，図 8 の凡例には，表 5 に示す VMM のプロトタイプのパラメータに基づいた電力効率曲線を示す．これらのプロトタイプ実装の中には，行列積演算が目的ではなくネットワークスイッチのものもあり，実現可能性に重点が置かれているため低パフォーマンスのデバイスを使用し，結果として VMM としての性能が制限されているものもある．したがって，これらの VMM を比較することは厳密には不公平である．ただし，本性能モデルによってあらゆるの種類の VMM の電力効率と精度を簡単に推定できることを意味する．つまり，提案の性能モデルは，VMM の性能，電力，精度の複雑なトレードオフ解析を可能とし光コンピュータの協調設計の指針となる．

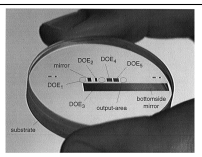
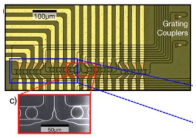
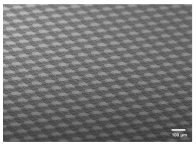
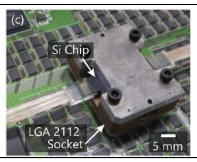
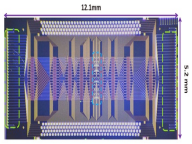
5. おわりに

本論文では，電力効率と計算精度のトレードオフを明らかにするために，光アナログ VMM のスループット，消費電力，計算精度の性能モデルを提案した．このモデルにより，光 VMM の電力効率限界だけでなく，デバイステクノロジおよびアーキテクチャ設計の方向性も明らかとなった．また，アプリケーションレベルでのノイズの影響を評価し，アプロキシメートコンピューティングにおける演算の不正確さに対する高い許容度を明確にした．この研究は，新しい光デバイスを使ったアナログ計算の将来のデバイス-アーキテクチャ-ソフトウェア協調設計のために非常に価値があると信じている．今後は，モデルから得た洞察を使用して最適な光 VMM を実際に実装する予定である．また，メモリインタフェースを含む光アクセラレータを設計し，光コンピュータの実現を目指す．

謝辞　本研究の一部は，科学技術振興機構の戦略的創造研究推進事業「新たな光機能や光物性の発見・利活用を基軸とする次世代フォトニクスの基盤技術（Grant Number: JPMJCR15N4）」の助成により行われた．

© 2019 Information Processing Society of Japan

表 5: VMM プロトタイプ一覧

参考文献	M. Gruber '00 [6]	A.N. Tait '17 [20]	T.J. Seok '16 [15]	K. Suzuki '18 [19]	L. Qiao '17 [13]
VMM の変調器	† MEMS (SLM)	リング共振器 (WDM)	MEMS クロスバー	位相シフタ (MZI)	
				TO 効果	キャリア注入
Overview					
スイッチングタイム ($1/f_{mod}$)	12 [μs]	1 [ns]	1 [μs]	1 [μs]	1 [ns]
VMM サイズ ($N \times N$)	10×10	4×4	64×64	32×32	32×32
損失 ($1 - T_{VMM}$)	10 [dB]	4 [dB]	3.7 [dB]	6.4 [dB]	14 [dB]
チップサイズ	10×10 [mm^2]	625×16 [μm^2]	8.6×8.6 [mm^2]	25×11 [mm^2]	12.1×5.2 [mm^2]

† 変調器は実装されていないが，文献中で MEMS SLM の利用が提案されている．

参考文献

[1] Akiyama, S., Baba, T., Imai, M., Akagawa, T., Takahashi, M., Hirayama, N., Takahashi, H., Noguchi, Y., Okayama, H., Horikawa, T. et al.: 12.5-Gb/s operation with 0.29-V· cm V π L using silicon Mach-Zehnder modulator based-on forward-biased pin diode, *Optics express*, Vol. 20, No. 3, pp. 2911–2923 (2012).

[2] Cartwright, S.: New optical matrix–vector multiplier, *Applied optics*, Vol. 23, No. 11, pp. 1683–1684 (1984).

[3] Chen, Y., Luo, T., Liu, S., Zhang, S., He, L., Wang, J., Li, L., Chen, T., Xu, Z., Sun, N. et al.: Dadiannao: A machine-learning supercomputer, *Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture*, IEEE Computer Society, pp. 609–622 (2014).

[4] Chi, P., Li, S., Qi, Z., Gu, P., Xu, C., Zhang, T., Zhao, J., Liu, Y., Wang, Y. and Xie, Y.: PRIME: A Novel Processing-In-Memory Architecture for Neural Network Computation in ReRAM-based Main Memory, *Proceedings of ISCA*, Vol. 43 (2016).

[5] Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. and Walmsley, I. A.: An Optimal Design for Universal Multiport Interferometers, *arXiv preprint arXiv:1603.08788* (2016).

[6] Gruber, M., Jahns, J. and Sinzinger, S.: Planar-integrated optical vector-matrix multiplier, *Applied optics*, Vol. 39, No. 29, pp. 5367–5373 (2000).

[7] Jouppi, N. P. e. a.: In-Datacenter Performance Analysis of a Tensor Processing Unit, *Proceedings of the 44th Annual ACM/IEEE International Symposium on Computer Architecture*, pp. 1–12 (online), DOI: 10.1145/3079856.3080246 (2017).

[8] Krizhevsky, A. and Hinton, G.: Learning multiple layers of features from tiny images, Technical report, Computer Science Department, University of Toronto (2009).

[9] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324 (1998).

[10] Notomi, M., Nozaki, K., Shinya, A., Matsuo, S. and Kuramochi, E.: Toward fJ/bit optical communication in a chip, *Optics Communications*, Vol. 314, pp. 3–17 (2014).

[11] Nozaki, K., Matsuo, S., Fujii, T., Takeda, K., Shinya, A., Kuramochi, E. and Notomi, M.: Forward-biased nanophotonic detector for ultralow-energy dissipation receiver, *APL Photonics*, Vol. 3, No. 4, p. 046101 (2018).

[12] Optilab LCC: A 40 GHz Linear InGaAs PIN Photodetector PD-40, http://www.optilab.com/images/datasheets/PD-40.pdf.

[13] Qiao, L., Tang, W. and Chu, T.: 32× 32 silicon electro-optic switch with built-in monitors and balanced-status units, *Scientific Reports*, Vol. 7, p. 42306 (2017).

[14] Reck, M., Zeilinger, A., Bernstein, H. J. and Bertani, P.: Experimental realization of any discrete unitary operator, *Phys. Rev. Lett.*, Vol. 73, pp. 58–61 (1994).

[15] Seok, T. J., Quack, N., Han, S., Muller, R. S. and Wu, M. C.: Large-scale broadband digital silicon photonic switches with vertical adiabatic couplers, *Optica*, Vol. 3, No. 1, pp. 64–70 (2016).

[16] Shafiee, A., Nag, A., Muralimanohar, N., Balasubramonian, R., Strachan, J. P., Hu, M., Williams, R. S. and Srikumar, V.: ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars, *Proc. ISCA* (2016).

[17] Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D. et al.: Deep learning with coherent nanophotonic circuits, *Nature Photonics* (2017).

[18] Sun, J., Timurdogan, E., Yaacobi, A., Hosseini, E. S. and Watts, M. R.: Large-scale nanophotonic phased array, *Nature*, Vol. 493, No. 7431, p. 195 (2013).

[19] Suzuki, K., Konoike, R., Hasegawa, J., Suda, S., Matsuura, H., Ikeda, K., Namiki, S. and Kawashima, H.: Low insertion loss and power efficient 32× 32 silicon photonics switch with extremely-high-Δ PLC connector, *Proceedings of Optical Fiber Communication Conference*, Optical Society of America, pp. Th4B–5 (2018).

[20] Tait, A. N., de Lima, T. F., Zhou, E., Wu, A. X., Nahmias, M. A., Shastri, B. J. and Prucnal, P. R.: Neuromorphic photonic networks using silicon photonic weight banks, *Scientific Reports*, Vol. 7, No. 1, p. 7430 (online), DOI: 10.1038/s41598-017-07754-z (2017).

[21] Takahashi, Y., Inui, Y., Chihara, M., Asano, T., Terawaki, R. and Noda, S.: A micrometre-scale Raman silicon laser with a microwatt threshold, *Nature*, Vol. 498, No. 7455, pp. 470–474 (2013).

[22] Tanizawa, K., Suzuki, K., Toyama, M., Ohtsuka, M., Yokoyama, N., Matsumaro, K., Seki, M., Koshino, K., Sugaya, T., Suda, S. et al.: Ultra-compact 32× 32 strictly-non-blocking Si-wire optical switch with fan-out LGA interposer, *Optics express*, Vol. 23, No. 13, pp. 17599–17606 (2015).

[23] Vivien, L., Polzer, A., Marris-Morini, D., Osmond, J., Hartmann, J. M., Crozat, P., Cassan, E., Kopp, C., Zimmermann, H. and Fédéli, J. M.: Zero-bias 40Gbit/s germanium waveguide photodetector on silicon, *Optics express*, Vol. 20, No. 2, pp. 1096–1101 (2012).

[24] 川上哲志，　磯部聖，浅井里奈，小野貴継，本田宏明，井上弘士，納富雅也：ナノフォトニック・ニューラルアクセラレーション構想，情報処理学会研究報告，Vol. 2016-ARC-223 (2016).

[25] 川上哲志，浅井里奈，小野貴継，本田宏明，井上弘士，　北翔太，納富雅也：ナノフォトニックコンピューティングの性能限界，情報処理学会研究報告，Vol. 2017-ARC227 (2017).

© 2019 Information Processing Society of Japan