

# 単純再帰型ニューラルネットワーク向け ナノフォトニックアクセラレータの設計

佐藤 英人<sup>1,a)</sup> 川上 哲志<sup>1,b)</sup> 岡 慶太郎<sup>1,c)</sup> 谷本 輝夫<sup>1,d)</sup> 小野 貴継<sup>1,e)</sup> 井上 弘士<sup>1,f)</sup>

**概要：**本研究では CMOS アクセラレータを凌駕する高性能・低消費電力な再帰型ニューラルネットワーク向け光アクセラレータの実現を目的とし、光演算回路の設計・検証を行う。その際に、光回路による非線形関数である活性化関数の実現、再帰経路と順方向経路の位相差に起因して発生する演算誤差、再帰経路が存在することによるノイズの蓄積という課題がある。そこで、コヒーレントレーバと光電/電光変換器を用いた ReLU 活性化関数実現法、光再帰回路における再帰光の位相補正法、ならびに、光ベクトル-行列積演算器を搭載した単純再帰型ニューラルネットワーク向け光演算回路を提案する。その機能検証のために、光回路シミュレータを用いて提案回路を設計し、それを用いた提案回路の演算精度評価及び実アプリケーションの推定精度評価を行う。また、提案回路を設計し、光回路シミュレータを用いた機能検証、演算精度評価、ならびに、実アプリケーションの推定精度評価を行う。さらに、再帰処理に起因するノイズの蓄積が演算精度に与える影響を評価し、提案回路のノイズ耐性を解析する。

## 1. はじめに

近年、機械学習に基づく情報処理手段としてニューラルネットワークが普及しており、その高性能化・低消費電力化を実現するための様々な CMOS アクセラレータが提案されている [3] [7]。しかしながら、CMOS 技術を支えてきた半導体の微細化は近い将来に限界を迎えると言われており、今後のさらなる高性能化・低消費電力化は困難になると考えられる [2]。

この問題を解決する手段の一つとして、ナノフォトニクス技術を用いた光アクセラレータの実現が注目されている。これまでに順伝播型ニューラルネットワークを対象とするナノフォトニックニューラルネットワークアクセラレータ (NNA: Nanophotonic Neural network Accelerator) が提案され、CMOS アクセラレータ [3] に対する約 10 倍の電力効率が示されている [5] [9]。

一方、言語処理や音声処理、株式市場予測といった時系列処理型アプリケーションの急速な普及にともない、再帰型ニューラルネットワーク (RNN: Recurrent Neural Network) の重要性が増している。しかしながら、NNA に関する既

存研究ではその殆どが順伝搬型を対象としており、直接的に RNN 型アプリケーションを実行することができない。RNN ではその性質上、「記憶」機能が必要となる。そのため、順伝搬型 NNA を基本として RNN 処理を実現する場合には、処理結果を電氣的メモリ素子に記憶し、その内容を光デバイスの入力へとフィードバックする必要がある。そのため、光アナログ信号から電気デジタル信号への変換 (ならびにその逆変換) がフィードバック毎に発生し、これが大きなオーバーヘッドの原因となる。この問題を解決し、RNN の効率的な加速実行を実現するためには、光/電気変換ならびにアナログ/デジタル変換を伴わない光アナログ再帰処理を実現しなければならない。

そこで、本研究では高性能/低消費電力な RNN 向けフォトニックアクセラレータの実現を目指し、その中核となる光演算回路を提案・設計・検証する。本光演算回路の特徴は、光演算回路に再帰経路を実現することで「記憶」機能を実現し、情報担体の変換オーバーヘッドを削減することにある。本稿では、光再帰に基づく演算を可能にすべく、光遅延時間差に起因する演算誤差の補正機構、ならびに、非線形演算のための光 ReLU 演算器を提案する。また、これらの主要構成要素にマッハツェンダ干渉器型ベクトル-行列積演算器 (MZI-VMM: Mach-Zehnder Interferometer - Vector by Matrix Multiplier) を組み合わせた RNN 向け光演算回路を設計する。さらに、光回路シミュレータによる評価によって、提案回路の動作検証ならびにノイズ耐性評

<sup>1</sup> 九州大学

<sup>a)</sup> eito.sato@cpc.ait.kyushu-u.ac.jp

<sup>b)</sup> satoshi.kawakami@cpc.ait.kyushu-u.ac.jp

<sup>c)</sup> keitaro.oka@cpc.ait.kyushu-u.ac.jp

<sup>d)</sup> tteruo@kyudai.jp

<sup>e)</sup> takatsugu.ono@cpc.ait.kyushu-u.ac.jp

<sup>f)</sup> inoue@ait.kyushu-u.ac.jp

価を実施する。

本稿の構成は以下の通りである。第2章で再帰型ニューラルネットワークについて説明し、第3章でナノフォトニックデバイスの基本的動作原理を紹介する。第4章では、提案回路の概要、ならびに、その主要構成要素の設計を示す。第5章で、提案回路の動作検証・ノイズ耐性評価結果を示し、最後に第6章で本研究をまとめる。

## 2. 再帰型ニューラルネットワーク

### 2.1 ニューラルネットワークの概要

ニューラルネットワークはニューロンとシナプスを基本構成要素とする生体の神経回路網を模した数学モデルである [8]。ニューロンとは、受け取った信号を非線形処理し出力を行う演算素子である。このとき、非線形処理に用いられる関数を活性化関数という。また、シナプスとはニューロン間で信号を伝達する素子であり、信号を伝達するニューロンとの結合強度により伝達する情報の重要度が決定する。この結合強度のことを重みという。ニューラルネットワークは一般に層と呼ばれる複数ニューロンが集合した構造を内部に持つ。層には入力層、中間層、出力層と呼ばれる3つの種類が存在する。それぞれの層は以下の機能を持つ。

- 入力層：ネットワークへの入力を受け取り、それを出力する。
- 中間層：前の層から受け取った情報を非線形処理し、次の層へ渡す。
- 出力層：ネットワークからの出力を行う。

ニューラルネットワークには学習と推定という二つの動作がある。学習とは、データをもとに問題を解くための最適な重みとバイアスを決定する処理を指す。ここで、バイアスとは入力値を一定の範囲に偏らせるために用いられる仕組みである。推論とは、学習で得た重みとバイアスの値をもとに問題の解を推測する処理を指す。

### 2.2 再帰型ニューラルネットワーク

ニューラルネットワークの中でネットワーク内に再帰経路を持つものを再帰型ニューラルネットワーク (RNN: Recurrent Neural Network) という。RNN は、図1で示すように再帰経路により中間層が時系列的に接続されており、再帰経路を信号が伝達する回数 (以降、再帰回数と呼ぶ) が増加するに従って時系列的に深いネットワークが構成される。

次に、本研究でアクセラレーション対象としている RNN の推論動作について説明する。図2は RNN の推論動作時の処理フローを示している。時刻  $t$  における RNN の中間層の伝搬式を式 (1) に、出力層の伝搬式を式 (2) に示す。式中の文字の意味を表1に示す。

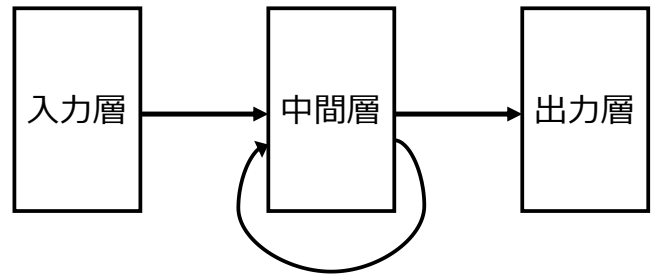


図1 再帰型ニューラルネットワークの概要

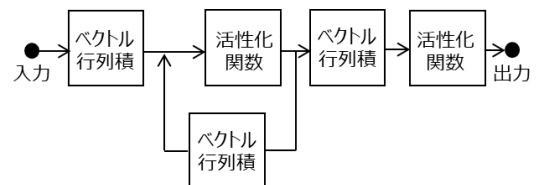


図2 再帰型ニューラルネットワークの処理フロー

$$u_l(t) = \sum_{i=1}^L w_{li}^{in} x_i(t) + \sum_{l'=1}^M w_{ll'}^{rec} z_{l'}(t-1) + b_{rec} \quad (1)$$

$$z_l(t) = f(u_l(t))$$

$$v_k(t) = \sum_{l=1}^M w_{kl}^{out} z_l(t) + b_{out} \quad (2)$$

$$y_k(t) = g(v_k(t))$$

ある時刻に入力層から入力された信号は、中間層へ伝搬される。その際に、シナプスの重みと信号が掛け合わされる。これは、ベクトル-行列積演算で表わすことができる。中間層のニューロンは、受け取った信号を活性化関数で非線形処理し出力する。このとき、出力は出力層に向かう信号と再帰経路を通り中間層へ再帰する信号が存在する。出力層に向かった信号は、ベクトル-行列積演算が行われた後、出力層のニューロンによって非線形処理されネットワーク全体の出力として出力される。再帰した信号は、次の時刻に入力層に入力された信号と足し合わされ中間層に入力される。

### 2.3 単純再帰型ニューラルネットワーク

単純再帰型ニューラルネットワーク (SRNN: Simple Recurrent Neural Network) の概要を図3に示す。SRNN は、RNN の動作に必要な最低限の要素のみで構成される最も単純な RNN である。必要最低限の要素とは、入力層ニューロン、中間層ニューロン、出力層ニューロン、バイアス項用ニューロン、入力層-中間層シナプス、中間層-中間層シナプス、中間層-出力層シナプスである。

## 3. ナノフォトニック回路

### 3.1 基本デバイス

本節では NNA に用いられる基本素子となるナノフォト

表1 式(1)及び式(2)中の文字の意味

$x_i$	入力層の $i$ 番目ニューロンの出力値
$u_l$	中間層の $l$ 番目ニューロンの入力値
$z_l$	中間層の $l$ 番目ニューロンの出力値
$v_k$	出力層の $k$ 番目ニューロンの入力値
$y_k$	出力層の $k$ 番目ニューロンの出力値
$w_{li}^{in}$	入力層の $i$ 番目-中間層の $l$ 番目のニューロン間の重み
$w_{ll'}^{rec}$	中間層の $l'$ 番目-中間層の $l$ 番目のニューロン間の重み
$w_{kl}^{out}$	中間層の $l$ 番目-出力層の $k$ 番目のニューロン間の重み
$b_{rec}$	入力層から中間層のバイアス値
$b_{out}$	中間層から出力層のバイアス値
$f$	中間層のニューロンの活性化関数
$g$	出力層のニューロンの活性化関数
$L$	入力層のニューロン数
$M$	中間層のニューロン数

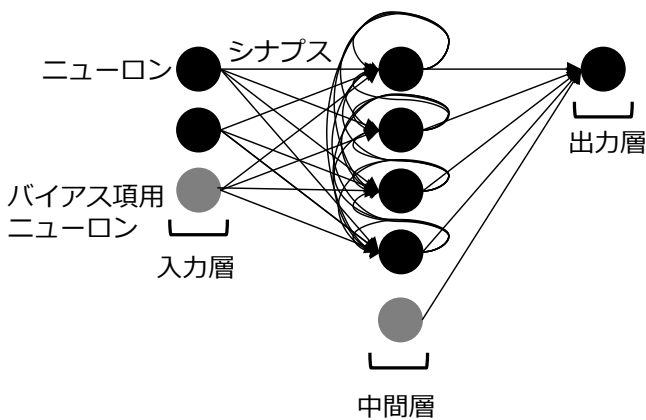


図3 単純再帰型ニューラルネットワーク

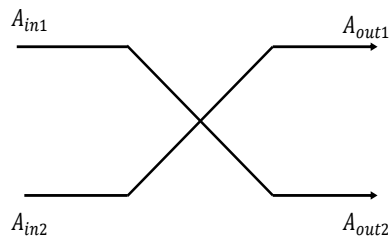


図4 方向性結合器

ニックデバイスについて述べる。ここで、光波は  $A$  を電界振幅、 $\omega$  を角速度、 $t$  を時間、 $\theta$  を初期位相として下記の式で表す。

$$E = Ae^{j\omega t + \theta} \quad (3)$$

ナノフォトニック回路において、データは光の振幅にエンコードされる。したがって、各光子素子の特性は、光信号の変化を数理的に表す伝達行列で説明できる。

### 3.1.1 方向性結合器

図4に方向性結合器の概要図を示す。方向性結合器は光信号を分割または結合する2入力2出力のナノフォトニックデバイスである。各々の入力ポートから照射された光は、一定の比率で2つの出力ポートから出力される。このとき

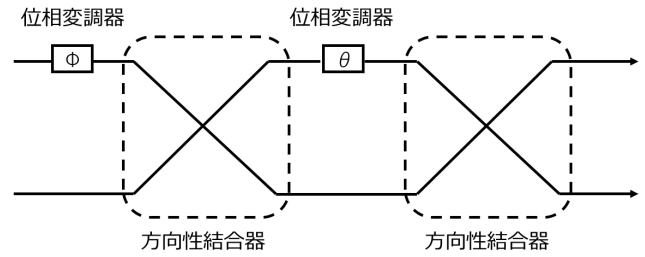


図5 マッハツェンダ干渉器

各入力ポートに対し上下逆方向の出力ポートに移動するエネルギーの割合を透過率という。また、導波路型の方向性結合器は光の導波路間の移動時に位相が  $\frac{\pi}{2}$  シフトする。以上の特性を式(4)の伝達行列で示す。

$$\begin{pmatrix} A_{out1} \\ A_{out2} \end{pmatrix} = \alpha \begin{pmatrix} \sqrt{1-c} & j\sqrt{c} \\ j\sqrt{c} & \sqrt{1-c} \end{pmatrix} \begin{pmatrix} A_{in1} \\ A_{in2} \end{pmatrix} \quad (4)$$

$A_{in1}$ ,  $A_{in2}$  はそれぞれ2つの導波路の入力電界振幅,  $A_{out1}$ ,  $A_{out2}$  は2つの導波路の出力電界振幅,  $\alpha$  は信号の電界振幅の損失係数,  $c$  は透過率を表す。

### 3.1.2 位相変調器

位相変調器は熱やキャリア注入により導波路の屈折率を変化させることで、光信号の位相を変化させるナノフォトニックデバイスである。入出力関係は、式(5)で表す。

$$A_{out} = A_{in} e^{j\Delta\phi} \quad (5)$$

$A_{in}$ ,  $A_{out}$  はそれぞれ位相変調器の入力光と出力光,  $\Delta\phi$  は位相変調度を表す。

### 3.1.3 マッハツェンダ干渉器

図5にマッハツェンダ干渉器 (MZI : Mach-Zehnder Interferometer) の概要を示す。MZIは入力された光信号を任意の比率で分配し出力するナノフォトニックデバイスである。導波路型のMZIは図5で示すように、2つの方向性結合器と2つの位相変調器によって構成される。位相変調度を適切に設定することで出力光の分配比率を決定できる。以上の特性を式(6)の伝達行列で示す。

$$\begin{pmatrix} A_{out1} \\ A_{out2} \end{pmatrix} = e^{j(\frac{\theta+\phi}{2})} \begin{pmatrix} e^{j\phi} \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \\ e^{j\phi} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \end{pmatrix} \begin{pmatrix} A_{in1} \\ A_{in2} \end{pmatrix} \quad (6)$$

ここで、 $A_{in1}$ ,  $A_{in2}$  はそれぞれの導波路への入力光の電界振幅,  $A_{out1}$ ,  $A_{out2}$  はそれぞれの導波路からの出力電界振幅,  $\theta$ ,  $\phi$  は2つの位相変調器の位相変調度を表す。

### 3.1.4 コヒーレントレシーバ

コヒーレントレシーバは光信号の位相を検知するナノフォトニックデバイスである [4]。図6に概要図を示す。コヒーレントレシーバは位相の検知対象である信号光と位相の基準となる参照光を入力とし、信号光の位相情報を持つ電気信号を出力する。以下に動作原理を示す。信号光  $E_{sig}$  と参照光  $E_{LO}$  を以下のように定義する。

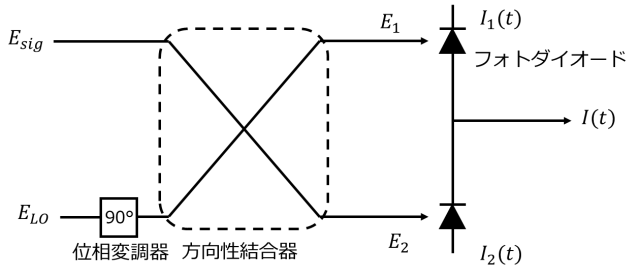


図6 コヒーレントレシーバ

$$\begin{aligned} E_{sig} &= A_{sig} e^{j\omega_{sig}t} \\ E_{LO} &= A_{LO} e^{j\omega_{LO}t} \end{aligned} \quad (7)$$

このとき  $A_{sig}$  と  $A_{LO}$  はそれぞれ信号光と参照光の電界振幅、 $\omega_{sig}$  と  $\omega_{LO}$  はそれぞれ信号光と参照光の角周波数である。まず、信号光または参照光の位相を位相変調器及び方向性結合器を用いて  $\pi$  シフトする。結合器を通過後の上部ポートと下部ポートの出力は下のように表せる。

$$\begin{aligned} E_1 &= \frac{1}{\sqrt{2}}(E_{sig} + E_{LO}) \\ E_2 &= \frac{1}{\sqrt{2}}(E_{sig} - E_{LO}) \end{aligned} \quad (8)$$

その後、フォトダイオードで光-電変換を行う。それぞれの出力電流は以下の通りである。

$$\begin{aligned} I_1 &= \frac{R}{2}(|A_{sig}|^2 + |A_{LO}|^2 + 2A_{sig}A_{LO} \cos(\omega_{sig} - \omega_{LO}t + \theta_{sig} - \theta_{LO})) \\ I_2 &= \frac{R}{2}(|A_{sig}|^2 + |A_{LO}|^2 - 2A_{sig}A_{LO} \cos(\omega_{sig} - \omega_{LO}t + \theta_{sig} - \theta_{LO})) \end{aligned} \quad (9)$$

このとき、 $I_1$ 、 $I_2$  は方向性結合器の各出力ポートへの出力電流、 $\theta_{sig}$ 、 $\theta_{LO}$  はそれぞれ信号光と参照光の位相、 $R$  はダイオードの応答性を指す。これにより、コヒーレントレシーバの出力電流は以下のよう表せる。

$$\begin{aligned} I &= I_1 - I_2 \\ &= 2RA_{sig}A_{LO} \cos(\omega_{sig} - \omega_{LO}t + \theta_{sig} - \theta_{LO}) \end{aligned} \quad (10)$$

### 3.2 MZI-VMM

MZI-VMM は任意のサイズの  $N \times M$  行列  $A$  を表現することができるナノフォトニックデバイスである。例として  $4 \times 4$  MZI-VMM の概要図を図7に示す。MZI-VMM は2つのユニタリ変換回路と1組の光増幅器または光減衰器で構成される。つまり、MZI-VMM の構成は行列  $A$  を特異値分解を用いて

$$A = U\Sigma V \quad (11)$$

と変換することと等しい。ここで、 $U$ 、 $V$  をユニタリ行列、 $\Sigma$  を対角要素に降順の特異値を持つ対角行列とする。また、ユニタリ変換回路は MZI を図7のように四角形状に配置することで実現できる [1]。

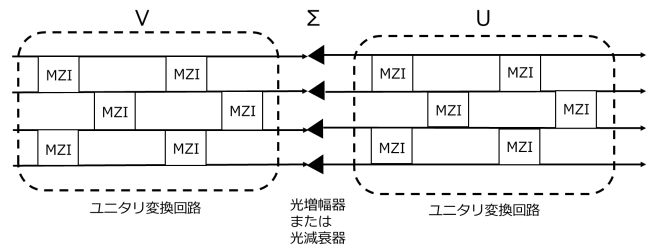


図7 MZI-VMM

表2 SRNN と提案回路の機能対応

SRNN	提案回路
入力層	光源～入力側 VMM の入力
入力層-中間層シナプス	入力側 VMM
中間層	入力側 VMM の出力～活性化関数出力
中間層-中間層シナプス	再帰側 VMM
中間層-出力層シナプス	出力側 VMM
出力層	出力側 VMM の出力～受光器

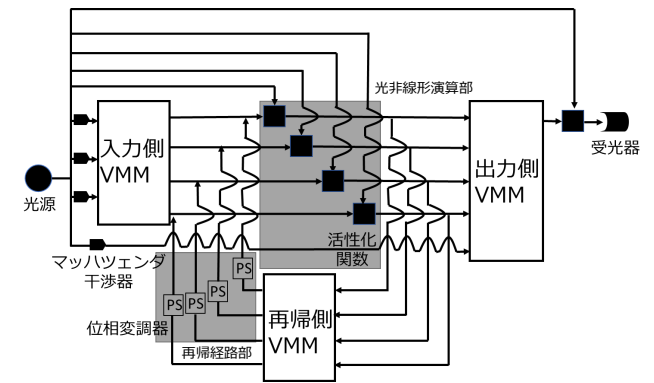


図8 SRNN 向け光演算回路

## 4. 提案回路

### 4.1 提案回路概要

本研究では図3で示した SRNN を想定し、回路設計を行う。その構成要素は、ニューロン、順方向シナプス、中間層から中間層への再帰シナプスである。

また、入力層と中間層にはバイアス項用のニューロンが存在し、層間の結合方式はバイアス項用のニューロンを除き全結合である。以下に SRNN の推論動作時の伝搬式を示す。

図8に図3を基にした提案回路の概要を示す。また、表2に図3と図8の対応を示す。シナプスにおける重みの計算は行列積演算を行う MZI-VMM で実現し、中間層・出力層で行われる非線形処理は ReLU 関数を実現する回路を光デバイスで構築した。これらを接続した提案回路の処理フローを図2に示す。

本研究では、実数空間における動作を想定する。ここで、実数空間における動作とは式(3)で

$$\omega t = 2\pi n \quad (12)$$

が成り立つことを指す。このとき、 $n$ は整数である。また、演算対象は光の電界振幅である。提案回路の光源の光強度は一定で、各VMMに入射する光の電界振幅の調整はVMM入射前にMZIを通過させることで実装している。提案回路は主要構成要素として、ベクトル-行列積演算用のMZI-VMM、光非線形演算部、再帰経路部を持つ。

## 4.2 光非線形演算部

既存研究において活性化関数は電気回路で実現されている[5]。しかしながら、提案回路を低消費電力で動作させるためには光回路で活性化関数を実現することが必要不可欠である。そこで、本研究では複数のナノフォトニックデバイスを組み合わせることで上限付きReLU関数を実現した。図9にその概要を示す。また、以下に想定するReLU関数の式を示す。

$$f(x) = \begin{cases} 0 & (x < 0) \\ x & (0 \leq x \leq 256) \\ 256 & (256 < x) \end{cases} \quad (13)$$

提案回路における活性化関数部は、コヒーレントレシーバ、電-光変換を行うポンプレーザー、光自乗器で構成される。ポンプレーザーは以下のような入出力関数をもつレーザーである。

$$f_{pump}(x_{pump}) = \begin{cases} 0 & (x_{pump} < i_{th}) \\ sx_{pump} & (i_{th} \leq x_{pump} \leq i_{max}) \\ s i_{max} & (i_{max} < x_{pump}) \end{cases} \quad (14)$$

ここで、 $f_{pump}(x_{pump})$ をポンプレーザーの出力光、 $x_{pump}$ をポンプレーザーの入力電流、 $i_{th}$ をポンプレーザーが動作するしきい値電流、 $i_{max}$ をポンプレーザーの最大入力電流、 $s$ をポンプレーザーの電-光変換効率とする。本研究では実数空間の動作を想定しているため、コヒーレントレシーバの出力電流の正負により入力光信号の電界振幅の正負を判定できる。ゆえに、ポンプレーザーのしきい値電流を0[A]に設定することで負電界の切り捨てを実現できる。正電界の場合は、ポンプレーザーの電-光変換効率を適切に設定し、最後に光自乗器を用いることで活性化関数部の入力電界振幅と出力電界振幅を一致させることができる。

次に、ポンプレーザーの電-光変換効率の決定法を述べる。ここで、式(13)を提案回路で実現するためには、式(13)の定義域をナノフォトニックデバイスの入出力特性により変換する必要があることに注意する。その変換率は、式(13)で入力最大ときに式(14)で示したポンプレーザーの最大入力電流が生じるように決定する。このとき、中間層部分の活性化関数部は、入力前と出力後にそれぞれ再帰光の合流と分離が生じるため当該箇所電界振幅が $\frac{1}{\sqrt{2}}$ 倍される点に注意する。これらを踏まえると変換率は、提案回路が実数空間の動作を想定していることを考慮すると式(10)、式(14)より、以下のように表わすことができる。

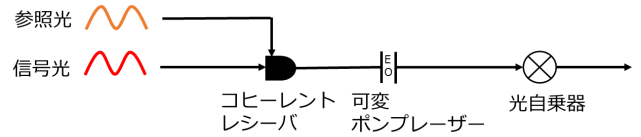


図9 光非線形演算部

$$c = \frac{i_{max}}{256 \sqrt{2} RA_{LO}} \quad (15)$$

ここで $c$ は定義域の変換率を示す。よって、提案回路における活性化関数部の入出力関数は以下の通りである。

$$A_{out} = \begin{cases} 0 & (A_{sig} < 0) \\ RA_{sig} A_{LO} s & (0 \leq A_{sig} \leq \frac{i_{max}}{\sqrt{2} RA_{LO}}) \\ \frac{i_{max}}{\sqrt{2} RA_{LO}} & (\frac{i_{max}}{\sqrt{2} RA_{LO}} < A_{sig}) \end{cases} \quad (16)$$

ここで、 $A_{out}$ は出力電界振幅、 $s$ はポンプレーザーの電-光変換効率である。この点と式(16)より $s$ の値は、

$$s_{mid} = \frac{2}{RA_{LO}} \quad (17)$$

$$s_{out} = \frac{1}{RA_{LO}}$$

と表せる。ここで、 $s_{mid}$ と $s_{out}$ はそれぞれ中間層側活性化関数部のポンプレーザーの電-光変換効率と出力層側活性化関数部のポンプレーザーの電-光変換効率を表わす。また、第4.1節で述べた通り光源の光強度は一定より、 $A_{LO}$ は一定であるのでそれぞれの電-光変換効率の値は一意に定まる。

本研究で想定するReLU関数は、式(13)で示す通り出力上限を設けている。これは、再帰回数増加に伴う回路の出力電界振幅の増大化を防ぐためである。出力上限はポンプレーザーの入力電流があるしきい値を超えると出力電界振幅が一定になるよう設定し実装している。

## 4.3 再帰経路部

提案回路において光の再帰時には入力光と再帰光の干渉が発生する。その際に、干渉地点までの経路長が異なるため二つの光に位相差が発生する。ゆえに、干渉後の光の電界振幅が意図しない値になるという問題がある。そこで、干渉地点の直前に位相変調器を設置し、適切な位相変調を行うことで二つの光の位相差を解消する。

位相変調の決定法を述べる。中間層側活性化関数部から位相変調器までの遅延時間を $t$ 、光源から干渉地点までの遅延時間を $t'$ 、光源の周波数を $f$ とすると位相変調 $\phi$ は以下のように表せる。

$$\phi = 2\pi f(t - t') \quad (18)$$

## 5. 回路動作検証と精度評価

### 5.1 概要

本章では提案回路の機能検証を目的とする回路シミュ

レーションによる評価について述べる。まず、ノイズを考慮しない場合の演算精度と実アプリケーションを用いたRNNの推論精度を評価する。演算精度の評価にはランダムな重みと入力を用いる。また、実アプリケーションを用いた評価では8bit Serial Adderを対象とする。この評価により、提案回路がRNNの推論を実行できることを確認する。

次に、提案回路の演算精度に対するノイズの影響を評価する。提案回路はアナログ再帰回路を採用しているため、ノイズが蓄積することが予想される。そのため、再帰回数の増加に伴い演算精度が低下することが考えられる。そこで、提案回路の入力電界振幅にノイズを付加し、再帰回数ごとの出力電界振幅と理論値の絶対誤差を測定する。

## 5.2 考慮するノイズのモデル

本評価で付加するノイズは正規分布  $\sim N(\mu, \sigma^2)$  に従うものとする。また、ノイズは入力側VMMの入力電界振幅に対して付加する。これは、入力側VMMの入力電界振幅は式(1)において  $x_i$  にあたり、入力側VMMの入力電界振幅にノイズを付加することで、提案回路で行う演算全体にノイズの影響を与えることができるためである。また、ノイズの分散は光波の主なノイズであるショットノイズを想定した小さな値から、位相変調器やフォトダイオードに生じる暗電流や熱雑音により発生するノイズが付加されたことを想定したときの大きな値までを考慮し網羅的に調査する。

## 5.3 評価環境

本研究における評価はOptisystem 17.1 [6] 上でシミュレーションによって行う。表3に評価時の各光デバイスのパラメータを示す。シミュレーション上ではポンプレーザーの最大入力電流が1[A]であり、参照光強度が10[mW]である。よって、シミュレーション上の入力電界振幅は式(16)を  $\frac{5\sqrt{2}}{256}$  倍に変換した値となる。ゆえに、評価環境における提案回路の活性化関数部の入出力関数は以下のように表せる。

$$A_{out} = \begin{cases} 0 & (A_{sig} < 0) \\ RA_{sig}A_{LOS} & (0 \leq A_{sig} \leq 5\sqrt{2}) \\ 5\sqrt{2} & (5\sqrt{2} < A_{sig}) \end{cases} \quad (19)$$

## 5.4 評価項目

### (1) 演算精度検証

ノイズは存在しないと仮定する。評価時の再帰回数は7回を想定する。提案回路が、式(1)、式(2)の演算を行う機能を持つことを確認するため、演算精度検証を行う。簡単化のため、重み行列の要素  $w$  は  $-10 \leq w \leq 10$  の整数値からランダムに決定する。同様の理由で、入力電界振幅  $A_{in}$  は  $-256 \leq A_{in} \leq 256$  の整数値からラン

表3 光デバイスのパラメータ一覧

$LMZI$ [ps]	MZI一つ当たりの遅延時間	1
$t_m$ [ps]	入力時間間隔	8
$P_{LO}$ [mW]	参照光強度	10
$f_{LS}$ [THz]	レーザーの周波数	193.1
$\lambda_{LS}$ [MHz]	レーザーのバンド幅	0.01
$I_{pump}$ [A]	ポンプレーザーの最大入力電流	1
$R$ [W/A]	フォトダイオードの応答性	1
$S_{mid}$ [A/W]	中間層側ポンプレーザーの電-光変換効率	20
$S_{out}$ [A/W]	出力層側ポンプレーザーの電-光変換効率	10

ダムに選びそれを  $10^{-3}$  した値とする。このとき、重みと入力電界振幅の整数値は一樣分布に従う乱数により決定する。設定した重みと入力値で回路を動作させた後、式(15)によって変換した定義域を元の定義域に再変換する。本評価環境では、再変換時に測定値は  $\frac{256}{5\sqrt{2}}$  倍される。評価は、再変換した測定値と理論値を比較する。この試行を100回行う。

### (2) 推定精度検証

ノイズは存在しないと仮定する。提案回路がRNNの推定動作を高精度で実行する機能を持つことを確かめるため、実アプリケーションを用いた推定精度検証を行う。用いるアプリケーションは8bit Serial Adderである。これは、2つの8ビットの2進数の和をとるアプリケーションであり、最下位ビットから順に計算を行う。最終的に正しい8ビットの和を求めるためには、計算途中で一つ前の桁から繰り上がりが発生するかしないかという情報を取得する必要がある。これをRNNの動作に当てはめると、一つ前の時刻に入力された情報の演算結果を次の時刻に入力された情報に足し合わせるという動作に相当する。よって、このアプリケーションを実行したときの推定精度によって提案回路がRNNの推定動作を高精度で実行できるか確認することができる。本評価において、和をとる2つの2進数は符号なし2進数を想定する。2つの2進数の値は0~127からランダムに選んだ値を2進数に変換して決定する。このとき選ぶ値は、一樣分布に従う乱数により決定する。選んだ値を提案回路に入力する際には、最下位ビットから1ビットずつ入力される。その際に、入力を8ビットに拡張し、「1」を「11111111」、「0」を「00000000」とし、さらにその値を10進数に変換する。つまり、回路の入力電界振幅は和をとる2進数のあるビットが「1」のとき「255」、「0」のとき「0」となる。ここで、RNNの推定は学習時に得た重みを使用して行われる。本評価においては、学習済みの重みのデータは従来の汎用コンピュータを用いて取得する。各ビットの和の推定値は、出力電界振幅の領域の中央値をしきい値として、しきい値より大きければ「1」、小さければ「0」と判定する。最終的に、各

ビットの推論値を結合して8ビットの2進数の和の推定値とする。そして、理論値と推定値を比較し評価する。この試行を100回行う。

### (3) 演算精度へのノイズの影響の検証

評価時の再帰回数は20回を想定する。ノイズによる影響を明確に示すために出力理論値を固定する。その値は、出力値域の中央値の $\frac{5\sqrt{2}}{2}$ とする。これは、活性化関数の上限と下限によってノイズの値が制限され、正確な影響が計測できなくなることを防ぐためである。また、重みによってもノイズの値が増減することが考えられるため、この点についても考慮する必要がある。よって、これを踏まえて重みを表わすVMMの行列要素を決定する。入力側VMMと出力側VMMの行列要素は、VMMの出力ベクトルの要素値と入力ベクトルの要素値が一致するように決定し、再帰側VMMの行列要素は、VMMの出力ベクトルの要素値が入力ベクトルの要素値の $\frac{1}{2}$ 倍になるように設定する。重みが上記の条件で決定し、出力理論値は $\frac{5\sqrt{2}}{2}$ で固定するため、入力電界振幅は1回目は $\frac{5\sqrt{2}}{2}$ 、2回目以降は $\frac{5\sqrt{2}}{4}$ とする。入力側VMMに光信号が入射する直前に分散が異なるノイズを付加する。想定する分散値は $10^{-15}$ 、 $10^{-12}$ 、 $10^{-9}$ 、 $10^{-6}$ 、 $10^{-3}$ である。分散ごとに再帰回数0回から20回までの提案回路の出力電界振幅を測定し、出力理論値との絶対誤差を算出する。この試行を分散ごとに100回行い、平均絶対誤差を計算する。以下に平均絶対誤差の式を示す。

$$m = \frac{1}{n} \sum_{k=1}^n |f_i - y_i| \quad (20)$$

ここで、 $n$ は試行回数、 $f_i$ は真値、 $y_i$ は測定値を表す。最後に、変換した定義域を元の値に再変換し、演算精度に対するノイズの影響を評価する。

## 5.5 評価結果

### (1) 演算精度検証

図10は、第5.4節で決定した入力値と重みに基づいた演算結果の理論値に対する回路動作によって得た演算結果のグラフである。出力電界振幅の理論値と測定値は非常に近似した。このことから、提案回路はSRNNの伝達時の演算を高精度で実行する機能があることがわかった。

### (2) 推定精度検証

図11は、第5.4節で決定した2進数の和の理論値に対する回路の出力結果に基づく2進数の和の推定値のグラフである。評価の結果、2進数の和の推定値と理論値が一致した。ゆえに、提案回路は十分高精度なRNNの推論動作を行えることが確認できた。

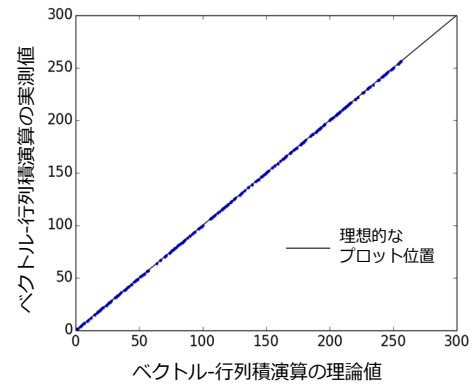


図10 提案回路による光行列積演算の出力値

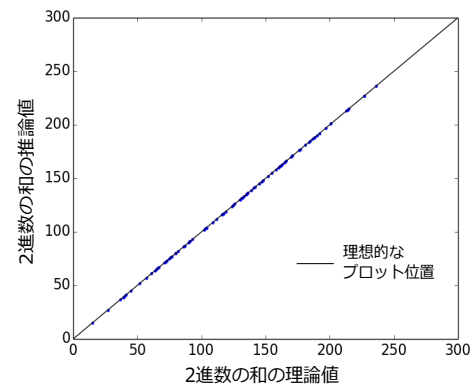


図11 実アプリケーションを用いたときの推定値

### (3) 演算精度へのノイズの影響評価

図12は、回路動作時の再帰回数に対する提案回路の出力理論値と出力値の平均絶対誤差のグラフである。どの分散においても再帰回数が増加すると演算精度が低下している。これにより、ノイズの蓄積が発生していることが分かる。最も分散が小さい $10^{-15}$ では再帰回数13回目で演算精度が劣化している。また、最も分散が大きい $10^{-3}$ では他の分散の場合と比べて再帰回数0回の時点で理論値との誤差が大きい。演算精度の低下をアプリケーション側で隠ぺいできるならば、回路における精度の低下は問題にならない。本研究で対象としているニューラルネットワークはある程度の演算誤差が許容される性質があるため、演算精度の低下を隠ぺいできる可能性がある。具体的な演算精度の低下による影響の調査は今後の課題である。一方、回路に高精度の演算が求められる場合は回路内にノイズを低減する機構を追加するなどの対策が考えられる。

## 6. おわりに

本稿では、単純再帰型ニューラルネットワークを対象とした光演算器を設計し、その演算精度のノイズ耐性を明らかにするために、付加したノイズの大きさと再帰回数ごとの演算精度の関係解析を行った。光演算器の設計では、活性化関数として非線形関数であるReLU関数を実現し、遅

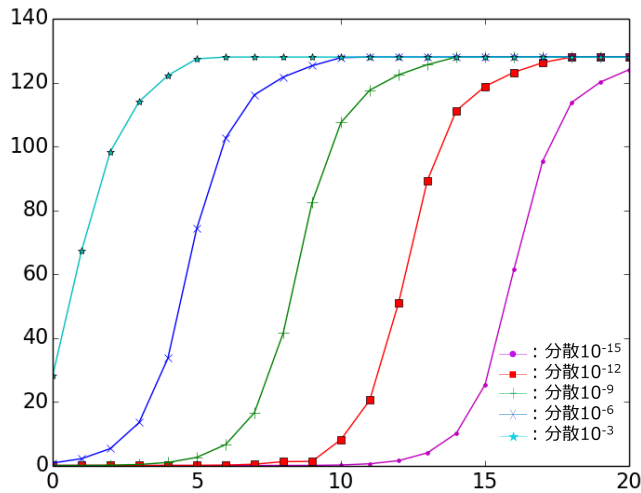


図12 ノイズの分散と再帰回数ごとの演算精度の関係

ニック・ニューラルネットワークアクセラレータ向け統合評価環境, 電子情報通信学会論文誌 A, Vol. 102, No. 6, pp. 182–193 (2019).

延時間差に由来する再帰時の光の位相差を解消した。ノイズ耐性の評価では、ノイズの分散のオーダーが  $10^{-15}$  程度でも再帰 13 回で高精度演算が破綻することが判明した。今後は、ノイズを考慮した実アプリケーションによる推定精度検証を行うことで、演算精度の劣化が推定精度に与える影響を調査し、提案回路の有用性を検討する予定である。

**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基盤技術 (Grant Number: JPMJCR15N4)」ならびに、JSPS 科研費 JP20K19771 の助成により行われた。

## 参考文献

- [1] Clements, W. R., Humphreys, P. C., Metcalf, B. J., Kolthammer, W. S. and Walmsley, I. A.: Optimal design for universal multiport interferometers, *Optica*, Vol. 3, No. 12, pp. 1460–1465 (2016).
- [2] Haron, N. Z. and Hamdioui, S.: Why is CMOS scaling coming to an END?, *2008 3rd International Design and Test Workshop*, pp. 98–103 (2008).
- [3] Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., Bates, S., Bhatia, S., Boden, N., Borchers, A. et al.: In-datacenter performance analysis of a tensor processing unit, *Proceedings of the 44th annual international symposium on computer architecture*, pp. 1–12 (2017).
- [4] Kikuchi, K.: Coherent Optical Communications: Historical Perspectives and Future Directions (2010).
- [5] Kitayama, K., Notomi, M., Naruse, M., Inoue, K., Kawakami, S. and Uchida, A.: Novel frontier of photonics for data processing—Photonic accelerator (2019).
- [6] Optiwave: OptiSystem, (online), available from (<http://optiwave.jp/home/optisystem/>).
- [7] Shafee, 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, *ACM SIGARCH Computer Architecture News*, Vol. 44, No. 3, pp. 14–26 (2016).
- [8] 稲場文男, 一岡芳樹ほか: 光コンピューティングの事典, 朝倉書店 (2006).
- [9] 川上哲志, 小野貴継, 納富雅也, 井上弘士: ナノフォト