Evaluation of Vector-Matrix Multiplier using optical devices

Koji Sato

Kyushu-University

Agenda

- Background
- Purpose
- Introduction of optical VMM
- Evaluation
  - Area
  - latency
- Plan
- Summary

Agenda

- Background
- Purpose
- Introduction of optical VMM
- Evaluation
  - Area
  - latency
- Plan
- Summary

Background

- Need for processing more data with low latency, and low power consumption
  - ✓Cyber physical system

- Physical limitation of CMOS device
  - ✓Ohmic loss
  - ✓Leakage current

Optical device

- Advantage of optical device
  - ✓Low latency

- Improvement of optical device
  - ✓Miniaturization
  - ✓Low power consumption
  - ✓Large scale integrated

Compute with optical device!

Problem

- Vector-Matrix Multiplier(VMM)
  - ✓ Vector-Matrix operation is used in many applications
    e.g.) neural network, image processing

- VMM can be realized with optical device

⇩

No comparison with optical device and CMOS

Purpose

- Purpose of my work
  - ➢Evaluation of VMM consisting of optical devices
    - ✓Latency
    - ✓Area
    - ✓Power consumption
    - ✓Accuracy
  - ➢Contrasting optical VMM with other VMMs
    - ASIC(Application Specific Integrated circuit)
    - GPU

Agenda

- Background
- Purpose
- Introduction of optical VMM
- Evaluation
  - Area
  - latency
- Plan
- Summary

Introduction of optical VMM

- This is design of analog vector-matrix multiplier
- Singular Value Decomposition(SVD)

  M × N matrix (A) can be decomposed as :

$$A = U\Sigma V$$

  U : M × M unitary matrix

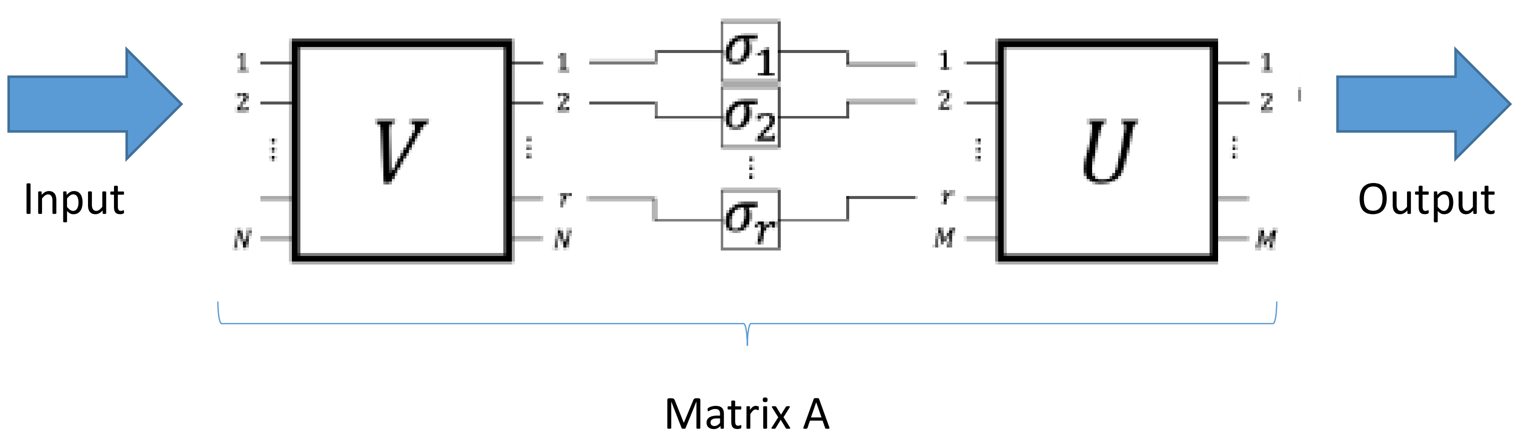
  V : N × N unitary matrix

  Σ : M × N diagonal matrix with non-negative real number

- Any matrix can be decomposed into two unitary matrix and a diagonal matrix

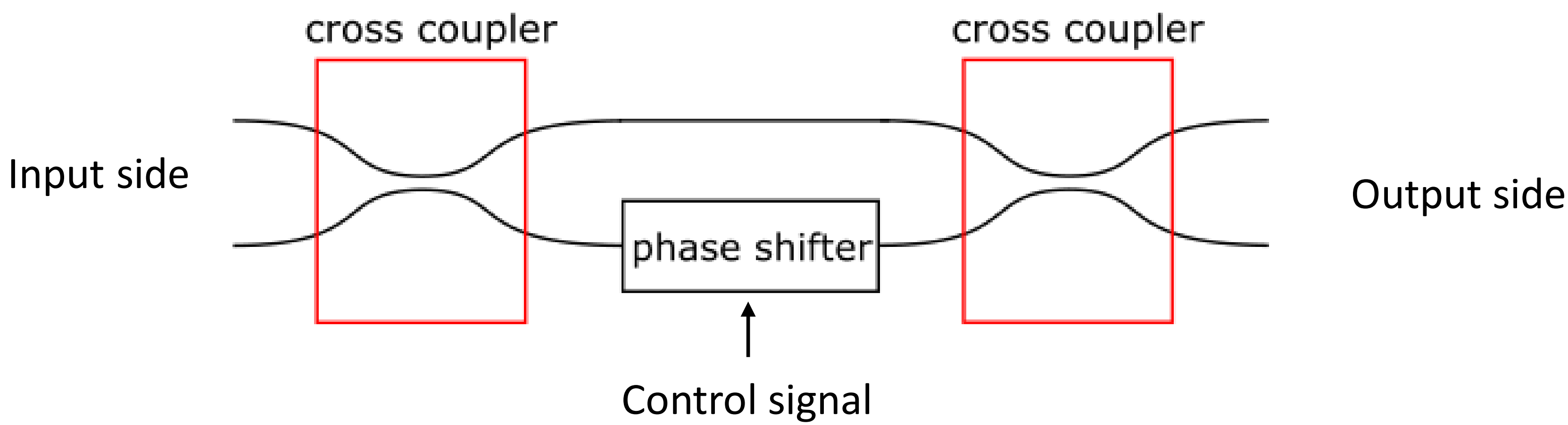
Introduction of optical VMM

- Diagonal matrix Σ can be implemented with amplifier
  - ✓ $\sigma_1, \sigma_2, \ldots, \sigma_r$ are amplifiers (r : rank of A)
- Unitary matrix U and V can be implemented with Mach-Zender interferometer(MZI) and phase shifter



Introduction of MZI

- MZI(Mach-Zenhder Interferometer)

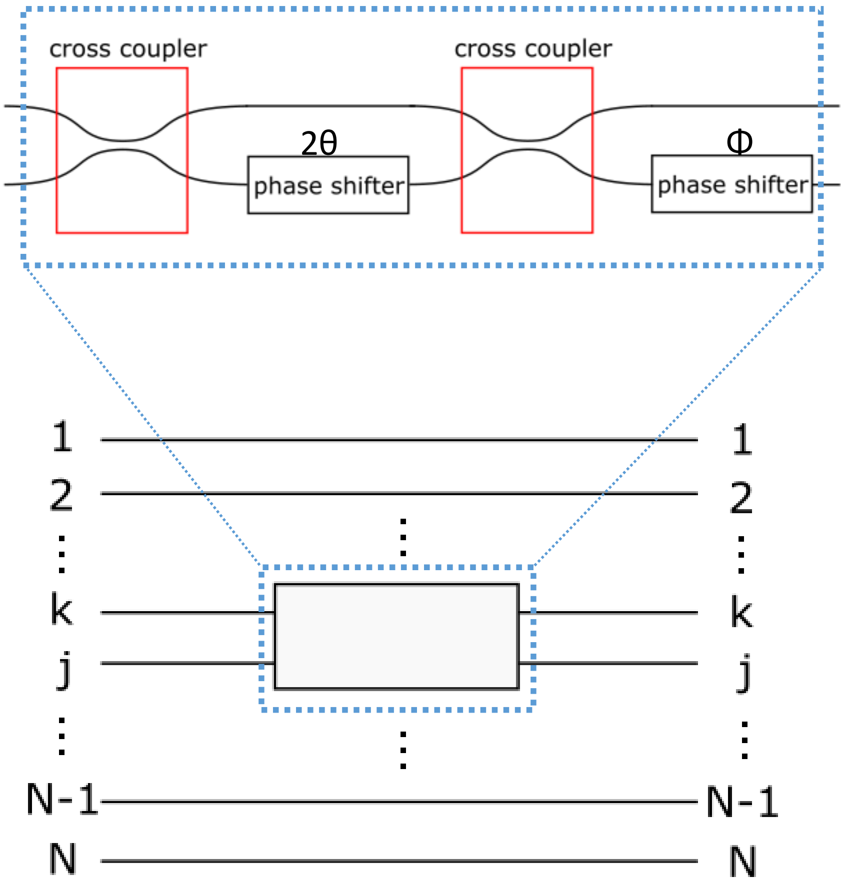


Cross coupler : coupling two of input lightwave

Phase shifter : shift lightwave’s phase depending on control signal

Introduction of MZI VMM

- MZI unitary matrix transformation



This component corresponds to
a following matrix transformation:

$$T(\theta,\phi) = \begin{pmatrix} e^{i\phi}\sin\theta & e^{i\phi}\cos\theta \\ \cos\theta & -\sin\theta \end{pmatrix}$$

$$\begin{array}{cc} & \text{Row} \\ T_{k,j}(\theta,\phi) = \begin{bmatrix} 1 & 0 & \cdots & & & 0 \\ 0 & 1 & & & & \vdots \\ \vdots & & e^{i\phi}\sin\theta & e^{i\phi}\cos\theta & & \\ & & \cos\theta & -\sin\theta & & \\ & & & & 1 & 0 \\ 0 & \cdots & & & 0 & 1 \end{bmatrix} & \begin{array}{c} \\ \\ k \\ j \\ \\ \\ \end{array} \\ \text{Column} \qquad k \qquad\qquad j & \end{array}$$

Introduction of MZI VMM

- The Unitary matrix U(N) is multiplied from right with Unitary matrices $T_{N,k}(\omega_{N,k}, \phi_{N,k})$ (for k = N-1,…,1)

$$U(N) \cdot \prod_{k=N-1}^{1} T_{N,k}(\theta_{Nk}, \phi_{Nk}) = \begin{pmatrix} \boxed{U(N-1)} & 0 \\ 0 & e^{i\alpha_N} \end{pmatrix}$$

- Do the same transformation repeatedly

$$U(N) \cdot T_{N,N-1} \cdot T_{N,N-2} \cdots T_{2,1} \cdot D = I(N).$$

I(N): Identity matrix in N dimensions

D:diagonal matrix with element modulus 1 (e.g. $e^{i\alpha}$)

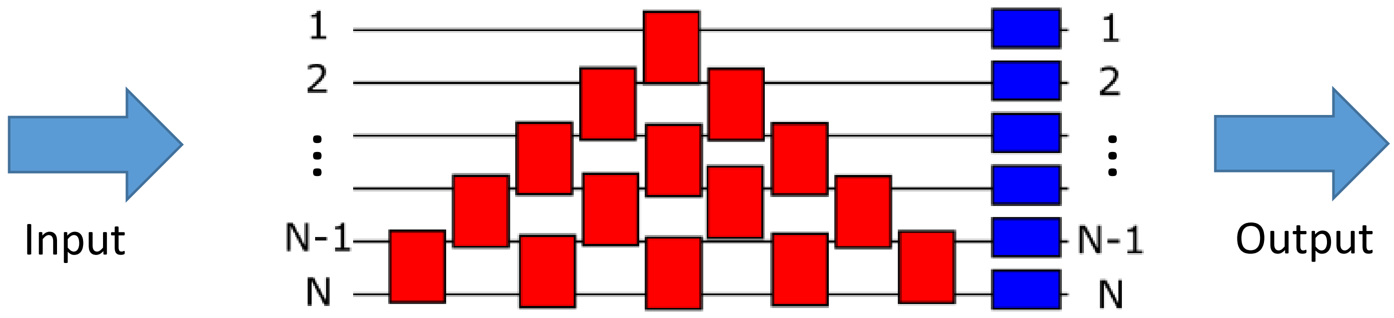
Introduction of MZI VMM

- Unitary matrix can be represented:

$$U(N) = (T_{N,N-1} \cdot T_{N,N-2} \cdots T_{2,1} \cdot D)^{-1} .$$

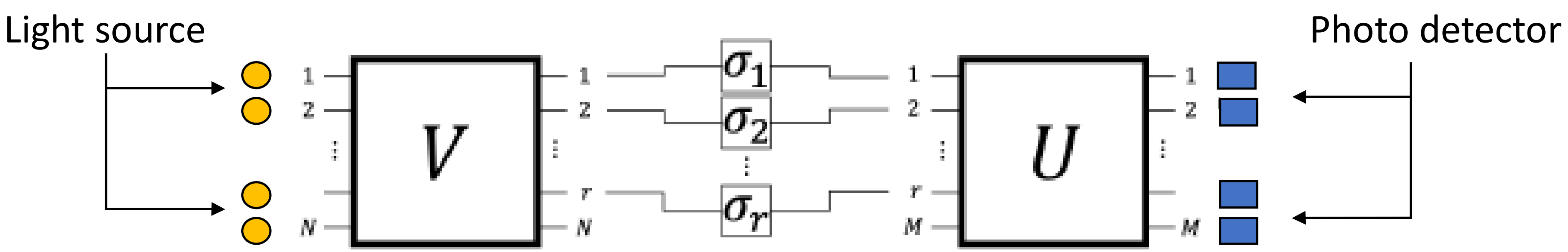
D can be implemented with appropriate phase shifters

- Unitary matrix U(N) can be implemented with MZIs and $\frac{N(N-1)}{2}$ phase shifters!



Introduction of MZI VMM

- Input
  - Light sources generate optical signals
- Output
  - Photo detector detect optical power



Agenda

- Background
- Purpose
- Introduction of optical VMM
- Evaluation
    - Latency
    - Area
- Plan
- Summary

Latency :Method of evaluation

- MZI VMM
  - Calculate from model formula

$$L = \frac{n}{c} l N_{pass} + L_{AMP} + L_{PD}$$

n : refractive index

c : speed of light

l : length of MZI

N_{pass} : the max number of MZI that light must pass

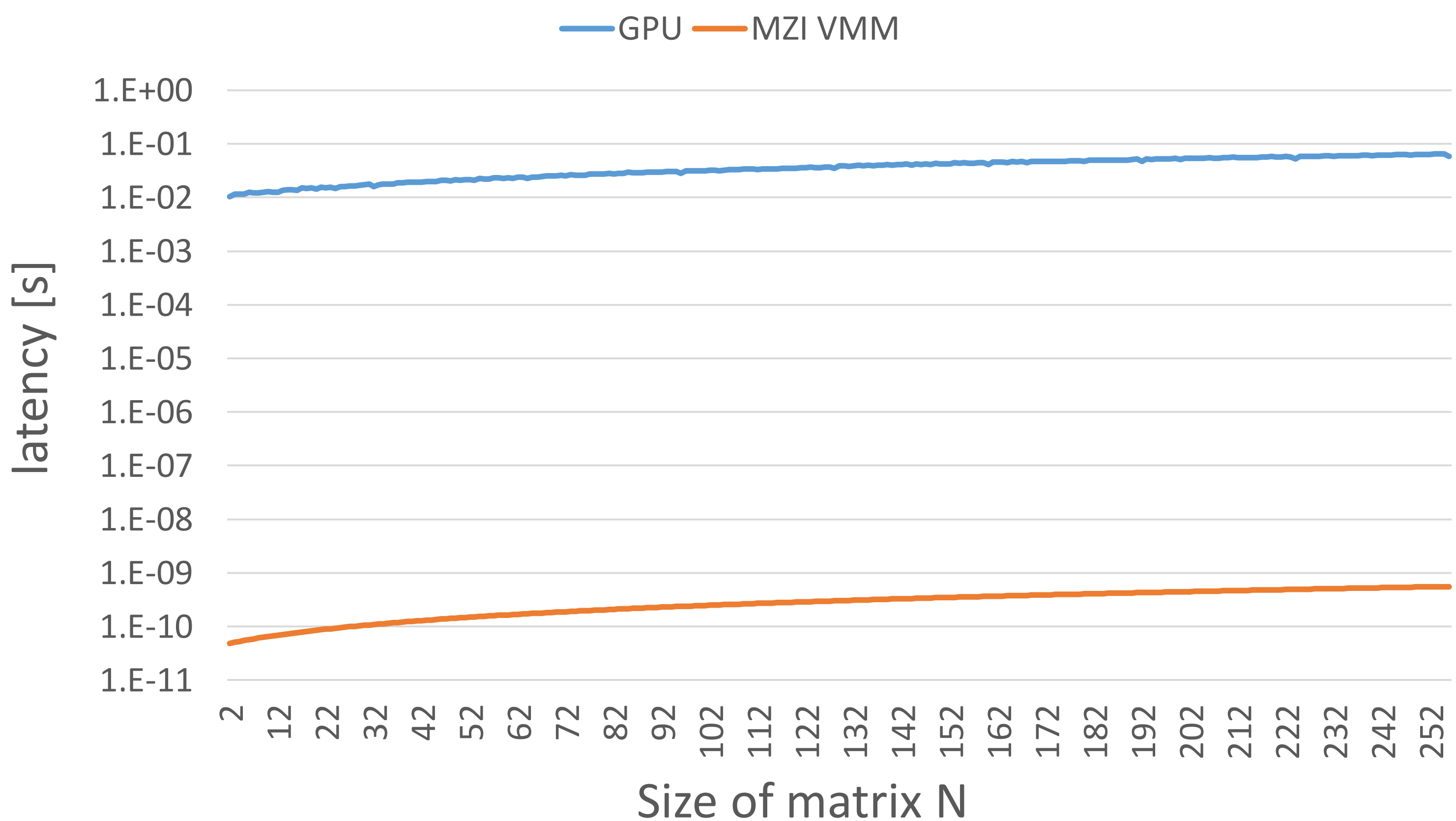
L_{AMP} : latency of amplifier

L_{PD} : latency of photo detector

Latency :Method of Evaluation

- GPU
  - Use library CUBLAS(CUDA Basic Linear Algebra subprograms)
  - Run on NVIDEA Tesla k20m (354nodes)
    - 345.6GFLOPS
    - Memory 128GB
    - Bandwidth 102.4GB
  - compute 400times on each matrix size and get the average latency

Evaluation : latency of VMM



MZI VMM can compute much faster than other VMMs

Area : Method of evaluation

- MZI VMM
  - Calculate from model formula (assuming M=N)

$$S = S_S \times \mathrm{N} + S_{\mathrm{MZI}} \times N(N-1) + S_{AMP} \times N + S_{PD} \times N$$

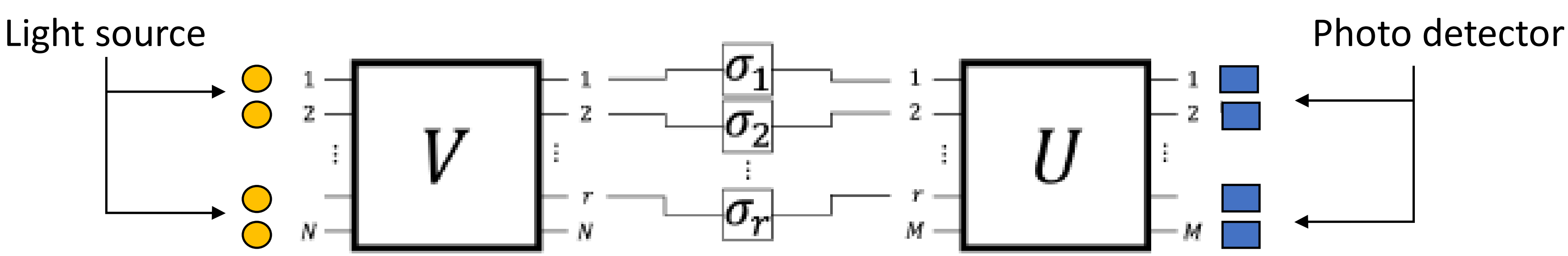
S_S :size of a light source

S_{MZI} : size of a MZI

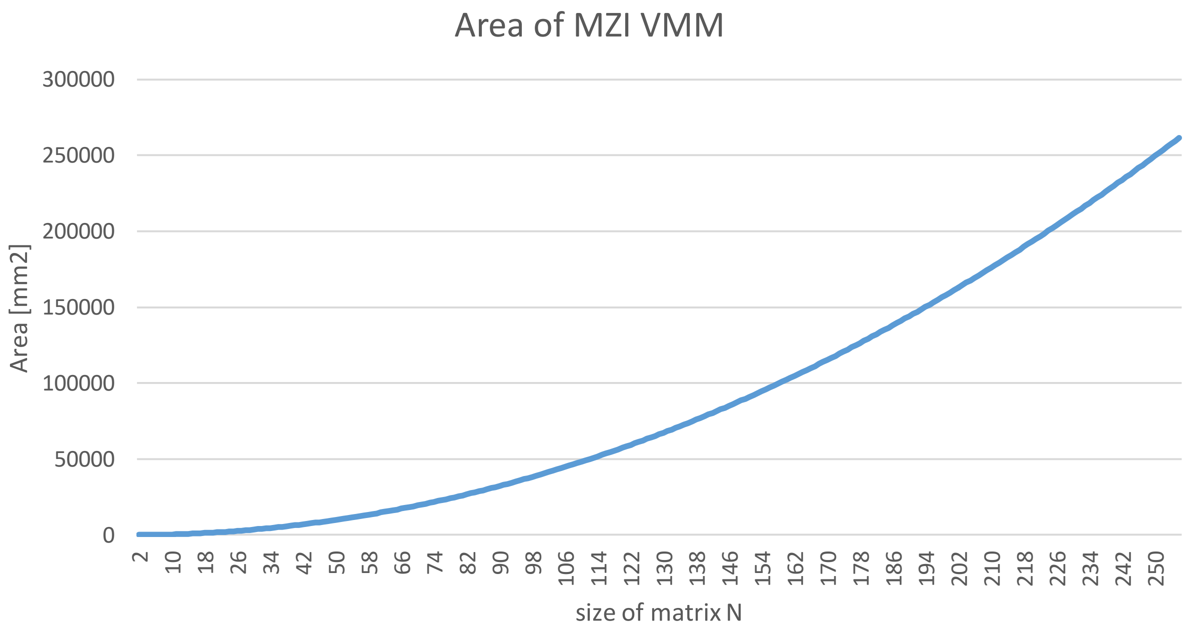
S_{AMP} : size of amplifier

S_{PD} : size of photo detector

N: size of matrix



Area : Evaluation



I will contrast area of MZI VMM with ASIC VMM

Agenda

- Background
- Purpose
- Introduction of optical VMM
- Evaluation
  - Area
  - latency
- Plan
- Summary

Plan

- Evaluate accuracy of MZI VMM
  - Survey about noise of optical devices
    - Optical amplifier
    - Phase shifter
    - Photo detector

- Evaluate performance of other VMM
  - ASIC

Agenda

- Background
- Purpose
- Introduction of optical VMM
- Evaluation
  - Area
  - latency
- Plan
- Summary

Summary

- Introduce MZI VMM
- My work is to compare vector-matrix multiplications
  - MZI VMM
  - GPU
  - ASIC
- Plan to evaluate MZI VMM’s accuracy