Abstract: Optical neural networks (ONN) have become the most promising solution to replacing electronic neural networks, which have the advantages of large bandwidth, low energy consumption, strong parallel processing ability, and super high speed. Silicon-based micro-nano integrated photonic platforms have demonstrated good compatibility with complementary metal oxide semiconductor (CMOS) processing. Therefore, without completely changing the existing silicon-based fabrication technology, optoelectronic hybrid devices or all-optical devices of better performance can be achieved on such platforms. To meet the requirements of smaller size and higher integration for silicon photonic computing, the topology of a four-channel coarse wavelength division multiplexer (CWDM) and an optical scattering unit (OSU) are inversely designed and optimized by Lumerical software. Due to the random optical power splitting ratio and incoherency, the intensities of different input signals from CWDM can be weighted and summed directly by the subsequent OSU to accomplish arbitrary multiply–accumulate (MAC) operations, therefore supplying the core foundation for scattering ONN architecture.

Keywords: optical neural network (ONN); multiply–accumulate (MAC) operation; inverse design; silicon-on-insulator (SOI); metastructure; coarse wavelength division multiplexer (CWDM); optical scattering unit (OSU)

1. Introduction

With the widespread use of deep learning techniques, artificial intelligence (AI) has achieved great success in machine vision, autonomous driving, and clinical diagnosis. However, the growth of the computational capability of integrated circuit (IC) chips as a carrier for training and executing AI models is gradually slowing down and failing to meet the demand of artificial neural network (ANN). The IC chip based on the von Neumann architecture has a structural defect that cannot be avoided, as it separates the program from the data when executing operations, resulting in tidal data loads between the memory and the computational unit, which decreases the computational rate and increases the computational power consumption [1]. Researchers have addressed this problem by increasing chip integration and calculating in memory [2]. Nevertheless, for transistors, as the sizes shrink and the quantum effects become more prominent, the performances deteriorate [3]. Besides, silicon-based electronic chips also have the problem of mutual interference of electronic signals. In addition, the existing neural network algorithms are not well matched with in-memory computing [4]. All these problems eventually limit the electronics applications in AI.
ANN based on optical devices, the so-called optical neural network (ONN), can overcome the aforementioned defects. Compared with the electronics counterpart, optical transmission inherently has the advantages of high speed, large bandwidth, low delay, high parallelism, low energy consumption, and immunity to electro-magnetic interference (EMI) [5]. Therefore, ONN can take advantage of photonic technology to break the bottleneck of conventional electrical ANN.

In 2017, Shen et al. [6] successfully developed the first ONN chip based on photonic interferometric units composed of 56 programmable Mach–Zehnder interferometer (MZI) arrays in a silicon photonic integrated circuit (PIC) to achieve linear operations on matrices. They fulfilled linear operations on a layer of fully connected neural networks and obtained 76.7% recognition accuracy in vowel recognition task using this chip twice. In 2018, Lin et al. [7] proposed an all-optical diffractive depth neural network (D²NN) consisting of multiple layers of diffractive surfaces that could perform computational tasks at the speed of light. The training of D²NN was completed by computer. After completing the training, a 3D model of each diffraction layer was established by Poisson surface reconstruction method. Then, the diffraction surface was printed by a 3D printer to classify handwritten digits and images of fashionable items. In 2020, Qu et al. [8] proposed an ONN structure based on light scattering units to carry out high-speed, small-region and low-power deep learning tasks such as high-precision random matrix multiplication with mean square error less than $10^{-4}$ and 97.1% recognition accuracy on the classical image classification dataset MNIST. In 2021, Zhang et al. [9] took advantage of the efficient complex-valued computation in optics and designed a novel ONN chip, realizing a true complex neural network with stronger learning capability than that of real-valued one.

Based on a silicon-on-insulator (SOI) processing platform, we propose a four-channel coarse wavelength division multiplexer (CWDM) and an optical scattering unit (OSU) chip architecture, to inversely design their inner topologies, to forwardly verify such basic metastructure elements, and finally to export mask patterns for fabrication. In this paper, CWDM is used to decompose the input broadband optical signal into four channels with different wavelengths and input them to OSU. The metastructure cascades the CWDM and OSU to implement the arbitrary multiply–accumulate (MAC) operations, realizing the core functionality of ONN.

2. Inverse Design Methodology

In silicon photonics, a CMOS-compatible SOI platform illustrated in Figure 1 is commonly used, where the three-layer structure is given. For the silicon rib waveguide core layer the permittivity $\varepsilon_{Si}$ is 7.84, while for the silicon-oxide cladding layer $\varepsilon_{SiO_2}$ is 2.07. The thickness of the rib waveguide is 220 nm while the width is 400 nm [10]. The CWDM and OSU discussed subsequently are designed and fabricated in SOI.

For optoelectronic devices, the traditional forward design cycles [11] involve systematically varying a small set of characteristic parameters, simulating these devices to determine their performance, and iterating to meet design requirements. Although this process is straightforward, even an experienced designer struggles when more than a few design parameters are to be coordinated simultaneously. To overcome human limitations, researchers employ inverse design methods to begin anticipating strong candidates and more quickly converge on a satisfactory one [12–14]. Molesky et al. [12] reviewed recent advancements in computational inverse design approaches, focusing on algorithms for optimizing optical structures based on desired functionalities to reshape the unconventional structures available to nanophotonics. Lin et al. [13] showed that with careful consideration and reformulation of the design objects, powerful inverse design techniques could be successfully applied to a multitude of interesting problems with rich physical behaviors to address intractable computational problems. Roques-Carmes et al. [14] proposed a framework for the inverse design of multilayer metaoptics via topology optimization and implemented a 3D-printed light concentrator working at five different nonparaxial angles of incidence, paving the way to realize multifunctional ultracompact 3D nanophotonic devices.
While in Lumerical FDTD software (Ansys Canada Ltd., Vancouver, BC, Canada) this paradigm shift is implemented using LumOpt as a python module. LumOpt comes in two different approaches—shape and topology optimization [15]. Rather than shape optimization, topology optimization generally needs more computation time to perform a few more complicated steps to produce physically realizable devices. This numerical algorithm often scales to thousands of parameters, but based on heuristic adjoint method [16,17], we only need two simulations per iteration to calculate the gradient. Furthermore, since the permittivity values are the optimization parameters, there is a direct link to Maxwell’s equations.

Once the optimization is running it proceeds in two or three steps. First comes the greyscale phase where the parameters vary continuously between the core and cladding index. Next, a binarization phase is run to force the mesh cells to take either the core or cladding index. Finally, an optional design for manufacturing (DFM) step is implemented. In this step a penalty function is added to the figure-of-merit (FOM). The magnitude of the penalty function is determined by the degree to which the minimum feature size constraint is violated. Typically, the minimum feature size is set about 500 nm, the finest resolution of mask aligner.

3. Four-Channel CWDM Chip Topology Optimization

3.1. Inverse Design

WDM unit combines a series of different wavelength optical signals into one channel as multiplexer (MUX) and splits them into each wavelength vice versa as demultiplexer (DEMUX). CWDM distinguishes from dense wavelength division multiplexer (DWDM) by broader wavelength channel separation, typically 10–20 nm, loosens the requirements for fine tunability of light source and wavelength division, and therefore could be used simply and cost-effectively.

A four-channel wavelength DEMUX working around a 1550 nm window is depicted in Figure 2. There is one input broadband source channel in the left and also four output channels in the right with the central wavelength of 1550 nm, 1570 nm, 1590 nm, and 1610 nm, individually. For each channel, the spectral width is 10 nm and the separation is also 10 nm. Therefore, the result is four divided transmission channels around [1545 nm, 1555 nm], [1565 nm, 1575 nm], [1585 nm, 1595 nm], [1605 nm, 1615 nm] routing to Port 1, 2, 3, and 4, respectively. In the core region of 6 µm × 6 µm, four S-bends connect and route lightwave from input port to output ports, accomplishing the function of CWDM-DEMUX.

We set FOMs as monitors for the four output ports demonstrating the wavelength division equality, which decrease and converge to a constant value as the iteration number increases, as shown in Figure 3a. Figure 3b depicts the gradient fields due to slight perturbation of permittivity. After optimization, the permittivity distribution and hence the irregular topology of the metastructure is illustrated in Figure 3c prototyping the DEMUX layout with GDSII format file exported for further simulation and fabrication. While
Figure 3d plots the forward direction optical field distribution within the device’s structure. The computing duration lasts about 5–6 h on a Dell server configured with Intel i9-10900X CPU and 128G RAM. It also depends on the structure complexity and iteration numbers.

Figure 2. Initial guess for CWDM-DEMUX planar topology in Lumerical FDTD software.

Figure 3. Inverse design of CWDM-DEMUX. (a) four-port FOM vs. iteration number; (b) sparse perturbation gradient field; (c) permittivity distribution; (d) forward direction field distribution within the device’s structure.

3.2. Forward Verification

To verify the functionality and performance of the inversely designed DEMUX, the optical transmission characteristics (Figure 4) and field trajectory (Figure 5) of the four-
channel window are simulated in the Lumerical FDTD integrated developing environment (IDE). In Figure 4, for each central wavelength of 1550 nm, 1570 nm, 1590 nm, 1610 nm the transmittance approximates to approximately 90% while for the boundary wavelength of each channel it descends to 70–85% due to the DFM treatment of inverse design strategy, which shall be discussed in the following Section 3.3. The less than 1 dB difference for each channel's central wavelength and boundary wavelength demonstrates the rather flat transmittance performance of the device. Figure 5a–d represents the optical field propagation within the device routing to port 1, 2, 3, and 4 for different central wavelength of 1550 nm, 1570 nm, 1590 nm, 1610 nm, respectively.

**Figure 4.** Optical transmission curves for CWDM-DEMUX.

**Figure 5.** Optical field trajectory for CWDM-DEMUX at different central wavelength of (a) 1550 nm; (b) 1570 nm; (c) 1590 nm; (d) 1610 nm.
3.3. DFM Strategy for Mask Design

The inversely designed and subsequently forward-validated metastructure tends to create features that break process design rules. To ensure DFM, Lumerical IDE supplies ways of explicitly enforcing specific minimum feature size constraints [18]. The main idea is to come up with an indicator function which is minimal if the minimum feature size constraints are fulfilled everywhere. DFM treatment implements the manufacturability albeit the device performance degrades to some extent, which can be seen from the transmittance difference between central and boundary wavelength for each divided channel shown in Figure 4 of previous Section 3.2.

Figure 6 compares the performance-priority design and manufacturability-priority design for CWDM-DEMUX’s permittivity distribution. For the former, small holes or disconnected parts may appear in the prototype; while for the latter, due to DFM treatment, the minimum feature speckles are neglected for easy fabrication of the metastructure. Thus, when exporting pattern definition to GDSII file illustrated in Figure 7 from Figure 6b, bigger discontinuous silicon pillars or voids of irregular shapes have replaced smaller ones depicted in Figure 6a.

Figure 6. CWDM-DEMUX’s permittivity distribution for (a) performance-priority design in comparison with (b) manufacturability-priority design.

Figure 7. GDSII pattern definition for CWDM-DEMUX metastructure layout.
4. Inverse Design of OSU Chip

Based on inverse design methodology, we conceive an initial guess of OSU metastructure with multi-input–multi-output (MIMO) on SOI, which has four input ports and four output ports as an instance, illustrated in Figure 8. The square scattering region is $6 \mu m \times 6 \mu m$. The forward TE optical modes with wavelength window of $\lambda_1 \in [1545 \text{ nm}, 1555 \text{ nm}]$, $\lambda_2 \in [1565 \text{ nm}, 1575 \text{ nm}]$, $\lambda_3 \in [1585 \text{ nm}, 1595 \text{ nm}]$ and $\lambda_4 \in [1605 \text{ nm}, 1615 \text{ nm}]$, which come from the aforementioned CWDM-DEMUX device, stimulate into four input ports, respectively. For each light source, the optical intensity is $I_1(\lambda_1)$, $I_2(\lambda_2)$, $I_3(\lambda_3)$ and $I_4(\lambda_4)$ accordingly. Through the OSU, i.e., the MIMO system, we obtain the mixed and redistributed optical signal at output ports with the intensity of $I_{o1}$, $I_{o2}$, $I_{o3}$ and $I_{o4}$. The relationship between the input optical intensity $I_{in}$ and output optical intensity $I_{out}$ is given by Equations (1) and (2). Due to the light source incoherency and scattering behavior, the intensities of different input signals can be weighted and summed directly so that the OSU perfectly fulfills the specific MAC computation represented by the stochastic matrix $P$ defined in Equation (3). Thus, arbitrary MAC operations based on random optical power splitting ratio and summation could be accomplished by OSU.

![Figure 8. Modeling an OSU metastructure on SOI in Lumerical FDTD software.](image)

The FOMs of the OSU can converge when the iteration number reaches 800 as Figure 9a indicates. Figure 9b depicts the sparse perturbation gradient fields. The permittivity distribution and hence the inner topology of the metastructure is illustrated in Figure 9c prototyping the OSU chip layout. While Figure 9d vividly portrays the random scattering process through photons colliding with the metastructure particles such as silicon pillars or voids.

$$I_{out} = P I_{in} = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.25 & 0.33 & 0 & 0 \\ 0.25 & 0.33 & 0.5 & 0 \\ 0.25 & 0.33 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} + I_2 \begin{bmatrix} 0 \\ 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} + I_3 \begin{bmatrix} 0 \\ 0 \\ 0.5 \\ 0.5 \end{bmatrix} + I_4 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} I_{o1} \\ I_{o2} \\ I_{o3} \\ I_{o4} \end{bmatrix} \tag{1}$$

$$I_{o1} = 0.25 I_1(\lambda_1) 
I_{o2} = 0.25 I_1(\lambda_1) + 0.33 I_2(\lambda_2)
I_{o3} = 0.25 I_1(\lambda_1) + 0.33 I_2(\lambda_2) + 0.5 I_3(\lambda_3)
I_{o4} = 0.25 I_1(\lambda_1) + 0.33 I_2(\lambda_2) + 0.5 I_3(\lambda_3) + I_4(\lambda_4) \tag{2}$$
Moreover, we can detail each optical signal propagation characteristics as follows. The optical signal $\lambda_1$ is almost equally routing to the four output ports as the relative electric intensity curve at certain output port indicates in Figure 10a while the field trajectory is depicted in Figure 11a. The optical signal $\lambda_2$ is equally routing to the output port 2, 3, and 4 as the relative electric intensity curve indicates in Figure 10b while the field trajectory is depicted in Figure 11b. The optical signal $\lambda_3$ is equally routing to the output port 3 and 4 as Figure 10c demonstrates, while the field trajectory is depicted in Figure 11c. The optical signal $\lambda_4$ is directly routing to the fourth output port shown in Figure 10d while the field trajectory is depicted in Figure 11d.

\[
P = \begin{bmatrix}
0.25 & 0 & 0 & 0 \\
0.25 & 0.33 & 0 & 0 \\
0.25 & 0.33 & 0.5 & 0 \\
0.25 & 0.33 & 0.5 & 1
\end{bmatrix}
\]  

(3)
Figure 10. The relative electric intensity distributions at certain output port for different optical signals. (a) $\lambda_1 \in [1545 \text{ nm}, 1555 \text{ nm}]$; (b) $\lambda_2 \in [1565 \text{ nm}, 1575 \text{ nm}]$; (c) $\lambda_3 \in [1585 \text{ nm}, 1595 \text{ nm}]$; (d) $\lambda_4 \in [1605 \text{ nm}, 1615 \text{ nm}]$.

For the sake of DFM rule, as with CWDM-DEMUX, we export the OSU layout illustrated in Figure 12 with GDSII file for fabrication.

In the end, as depicted in Figure 13, the silicon-based scattering metastructure cascading CWDM-DEMUX (Figure 7) and OSU (Figure 12) perfectly accomplishes arbitrary MAC operations and hence supplies the sound foundation for scattering ONN architecture.

Furthermore, metastructures over four channels, such as one by nine CWDM and nine by nine OSU, could fulfill nine optical signals’ weighting and summation operation by the stochastic matrix implemented on the random optical power splitting ratio and incoherency.
Figure 11. The optical field trajectory within OSU metastructure. (a) $\lambda_1 \in [1545 \text{ nm}, 1555 \text{ nm}]$; (b) $\lambda_2 \in [1565 \text{ nm}, 1575 \text{ nm}]$; (c) $\lambda_3 \in [1585 \text{ nm}, 1595 \text{ nm}]$; (d) $\lambda_4 \in [1605 \text{ nm}, 1615 \text{ nm}]$.

Figure 12. GDSII pattern definition for OSU metastructure layout.
5. Conclusions

To tackle the obstacles confronted by electronics in AI and meanwhile meet the requirements of smaller size and higher integration for silicon photonic computing, the topologies of four-channel CWDM and OSU are inversely designed and optimized in the state-of-the-art Lumerical IDE, based on which, the scattering metastructures are constructed. Due to the random optical power splitting ratio and incoherency, the intensities of different input signals from CWDM can be weighted and summed directly by the cascaded OSU to accomplish arbitrary MAC operations, therefore supplying the core foundation for scattering ONN architecture. The ongoing research will focus on the optical performance and feasibility of the fabrication on conventional silicon material and universal photolithography/etching process to exploit the potential application of handwritten digit classification in ONN.

Author Contributions: Conceptualization, X.L.; Data curation, X.Z., Y.C. and H.O.; Formal analysis, C.W., Y.C. and H.O.; Funding acquisition, X.L.; Investigation, B.W.; Methodology, X.Z. and C.W.; Project administration, X.L.; Supervision, X.L.; Validation, Y.W., Q.M., J.Z., Z.C. and R.L.; Writing—original draft, X.L.; Writing—review & editing, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Open Project Program of the State Key Laboratory of Mathematical Engineering and Advanced Computing (Grant No. 2020A05).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data presented in this article are available at request from the corresponding author.

Acknowledgments: This work is especially for the 120th anniversary of Southeast University.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zaharia, M.; Chowdhury, M.; Das, T.; Dave, A.; Ma, J.; McCauley, M.; Franklin, M.J.; Shenker, S.; Stoica, I. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation, San Jose, CA, USA, 25–27 April 2012.
2. Sebastian, A.; Le Gallo, M.; Khaddam-Aljameh, R.; Eleftheriou, E. Memory devices and applications for in-memory computing. Nat. Nanotechnol. 2020, 15, 529–544. [CrossRef] [PubMed]
3. Waldrop, M.M. The chips are down for Moore’s law. Nature 2016, 530, 144–147. [CrossRef] [PubMed]
4. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning, 1st ed.; MIT Press: Cambridge, MA, USA, 2016; pp. 267–320.
5. Huang, C.; Sorgor, V.J.; Miscuglio, M.; Al-Qadasi, M.; Mukherjee, A.; Lampe, L.; Nichols, M.; Tait, A.N.; de Lima, T.F.; Marquez, B.A.; et al. Prospects and applications of photonic neural networks. Adv. Phys. X 2022, 7, 1981155. [CrossRef]
6. Shen, Y.C.; Harris, N.C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.J.; Larochelle, H.; Englund, D.; et al. Deep learning with coherent nanophotonic circuits. Nat. Photonics 2017, 11, 441–447. [CrossRef]
7. Lin, X.; Rivenson, Y.; Yardime, N.T.; Veli, M.; Luo, Y.; Jarrahi, M.; Ozcan, A. All-optical machine learning using diffractive deep neural networks. Science 2018, 361, 1004–1008. [CrossRef] [PubMed]
8. Qu, Y.R.; Zhu, H.Z.; Shen, Y.C.; Zhang, J.; Tao, C.N.; Ghosh, P.T.; Qiu, M. Inverse design of an integrated-nanophotonics optical neural network. Sci. Bull. 2020, 65, 1177–1183. [CrossRef]

9. Zhang, H.; Gu, M.; Jiang, X.D.; Thompson, J.; Cai, H.; Paesani, S.; Santagati, R.; Laing, A.; Zhang, Y.; Yung, M.H.; et al. An optical neural chip for implementing complex-valued neural network. Nat. Commun. 2021, 12, 457. [CrossRef] [PubMed]

10. Bozeat, R.J.; Day, S.; Hopper, F.; Payne, F.P.; Roberts, S.W.; Asghari, M. Silicon based waveguides. In Silicon Photonics; Pavesi, L., Lockwood, D.J., Eds.; Topics in Applied Physics; Springer: Berlin, Germany, 2004; Volume 94, pp. 269–294.

11. Chih, H.C. New Design Method in Silicon Photonics with ANSYS Lumerical—Brief Introduction of Inverse Design. In Proceedings of the Ansys Innovation Conference 2020, Online, 17–18 September 2020. Available online: https://hmediastorageansi.blob.core.chinacloudapi.cn/t-1-resource/bbc60e8f-78d1-9dac-f16a-5ff8a6b835be.pdf (accessed on 1 May 2021).

12. Molesky, S.; Lin, Z.; Piggott, A.Y.; Jin, W.L.; Vuckovic, J.; Rodriguez, A.W. Inverse design in nanophotonics. Nat Photonics 2018, 12, 659–670. [CrossRef]

13. Lin, Z.; Johnson, S.G. Freeform inverse design in photonics by re-thinking the questions. In Proceedings of the 19th International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD), University of Ottawa, Ottawa, ON, Canada, 8–12 July 2019; pp. 1–2.

14. Roques-Carmes, C.; Lin, Z.; Christiansen, R.E.; Salamin, Y.; Kooi, S.E.; Joannopoulos, J.D.; Johnson, S.G.; Soljacic, M. Toward 3D-Printed Inverse-Designed Metaoptics. ACS Photonics 2022, 9, 43–51. [CrossRef]

15. Lumerical. Available online: https://support.lumerical.com/hc/en-us/articles/360049853854 (accessed on 1 May 2021).

16. Lalau-Keraly, C.M.; Bhargava, S.; Miller, O.D.; Yablonovitch, E. Adjoint shape optimization applied to electromagnetic design. Opt. Express 2013, 21, 21693–21701. [CrossRef] [PubMed]

17. Keraly, C.L.; Bhargava, S.; Ganapati, V.; Scranton, G.; Yablonovitch, E. Shape Optimization of Nanophotonic Devices Using the Adjoint Method. In Proceedings of the Conference on Lasers and Electro-Optics (CLEO), San Jose, CA, USA, 8–13 June 2014.

18. Zhou, M.D.; Lazarov, B.S.; Wang, F.W.; Sigmund, O. Minimum length scale in topology optimization by geometric constraints. Comput. Meth. Appl. Mech. Eng. 2015, 293, 266–282. [CrossRef]