Research Article

Vahid Nikkhah, Mario Junior Mencagli and Nader Engheta*

Reconfigurable nonlinear optical element using tunable couplers and inverse-designed structure

https://doi.org/10.1515/nanoph-2023-0152
Received March 3, 2023; accepted May 17, 2023;
published online June 13, 2023

Abstract: In recent years, wave-based analog computing has been at the center of attention for providing ultrafast and power-efficient signal processing enabled by wave propagation through artificially engineered structures. Building on these structures, various proposals have been put forward for performing computations with waves. Most of these proposals have been aimed at linear operations, such as vector-matrix multiplications. The weak and hardly controllable nonlinear response of electromagnetic materials imposes challenges in the design of wave-based structures for performing nonlinear operations. In the present work, first, by using the method of inverse design we propose a three-port device, which consists of a combination of linear and Kerr nonlinear materials, exhibiting the desired power-dependent transmission properties. Then, combining a proper arrangement of such devices with a collection of Mach–Zehnder interferometers (MZIs), we propose a reconfigurable nonlinear optical architecture capable of implementing a variety of nonlinear functions of the input signal. The proposed device may pave the way for wave-based reconfigurable nonlinear signal processing that can be combined with linear networks for full-fledged wave-based analog computing.

Keywords: inverse design; Mach–Zehnder interferometer; metamaterial; nonlinear element; nonlinearity; reconfigurable

***Corresponding author: Nader Engheta**, Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA, E-mail: engheta@ee.upenn.edu. https://orcid.org/0000-0003-3219-9520
Vahid Nikkhah, Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA, E-mail: vnikkhah@seas.upenn.edu. https://orcid.org/0000-0003-0666-2614
Mario Junior Mencagli, Department of Electrical and Computer Engineering, University of North Carolina, Charlotte, NC 28223, USA, E-mail: mmencagl@uncc.edu. https://orcid.org/0000-0001-7208-036X

1 Introduction

Optical nonlinearity plays a crucial role in several exciting areas of research [1–4]. Among these areas, emerging optical neural networks (ONNs) [5, 6] which are a physical implementation of a standard artificial neural network (ANN) with optical components, have received growing interest. ONNs have several advantages over their electronic counterparts. Among these advantages, it is worth emphasizing their higher computational speed with lower power consumption. These features make them appealing for several applications that require handling large data sets such as real-time image processing [7], language translation [8], decision-making problems [9], and more [10, 11]. One of the most important units in the ANN structure is the activation function [12]. This function determines the neural network output, its accuracy, and the computational efficiency of a training model. Activation functions can be described by simple mathematical nonlinear functions. In recent years, different nonlinear functions, acting as activation functions, have been investigated in the ANN community. It turned out that the nonlinear activation function's choice is closely connected with the ANN application [12]. Diverse ANN applications require the use of different nonlinear activation functions. For example, the sigmoid function, which is the most common nonlinear activation function, is particularly suitable for applications that produce output values in the range of [0,1]. Despite the mathematical simplicity of nonlinear activation functions, it is still unclear how to perform arbitrary nonlinear functions on waves physically or even in principle whether such functions are generally possible. Current nonlinear optical components, such as bistable and saturable absorber devices [5, 6, 13], can implement only a subset of all the possible nonlinear activation functions. Moreover, most of them suffer from a lack of reconfigurability, which means that once they have been realized, the form of their nonlinear response cannot be changed. As a result, they can only be used to implement a single activation function limiting the ONN application range. In the present work, we introduce an idea for a *reconfigurable* architecture that can implement arbitrary nonlinear functions (within certain constraints) between the input signal,

and the output signal (see Figure 1). The proposed architecture is composed of two networks: (1) the nonlinear signal divider and (2) the linear optical network composed of a mesh of Mach–Zehnder Interferometers (MZIs). The nonlinear optical network consists of a set of identical three-port devices functioning as specialized optical power limiters. These limiters employ a Kerr nonlinear material and a linear dielectric for limiting the optical signal/intensity of the light that is coupled to one of the output ports and the rest of the energy is coupled to another output port. The specific composition and spatial distributions of the Kerr material and the linear dielectric are obtained through the method of the inverse design [14–19]. This method has opened up enormous opportunities both in linear and nonlinear optics for designing non-intuitive optical devices with complex functionalities. The optically-linear MZI portion of the proposed device consists of a properly arranged collection of MZIs, as shown in Figure 1. Photonic MZI meshes, which commonly consist of waveguide-based MZI array laid out on a flat silicon substrate, have attracted a great deal of attention in recent years [13, 20–25]. This interest mostly focuses on MZI networks that can be programmed to implement any linear transformations with electromagnetic waves [26–28]. Since such networks exploit the well-known computational capabilities of electromagnetic waves and are suitable for on-chip integration, it is playing an important role in several applications, including optical machine learning [5], quantum-information processing [25], forward scattering problems [29], and optical analog equation solving platform [30, 31]. Our proposed device here, thanks to its versatility and reconfigurability, may be useful for various applications in nonlinear photonics, such as ONNs mentioned above, in which reconfigurable nonlinear optical elements are needed.

Throughout the paper, we assume $e^{j\omega t}$ as the time dependence of EM fields.

2 Inverse-designed nonlinear power limiter for the nonlinear network

In this section, the required characteristics and design of the constitutive element of the nonlinear network, i.e., the nonlinear power limiter is discussed. As schematically illustrated in Figure 2(a), the nonlinear element is a three-port device that is required to exhibit the following transmission features. When a monochromatic signal, which is injected into the input port (Port 1) carries a power below a certain value, denoted as P_{sat}, this signal should appear at the designated output port on the right (Port 2). If the power of the input signal exceeds P_{sat}, the power of the signal at Port 2 should stay at its "saturated" value, P_{sat}, and the spillover of the input signal should be directed to the other output port on the bottom (Port 3). As the power of the input signal increases beyond P_{sat}, the power reaching Port 3 keeps increasing, while the one going to Port 2 remains

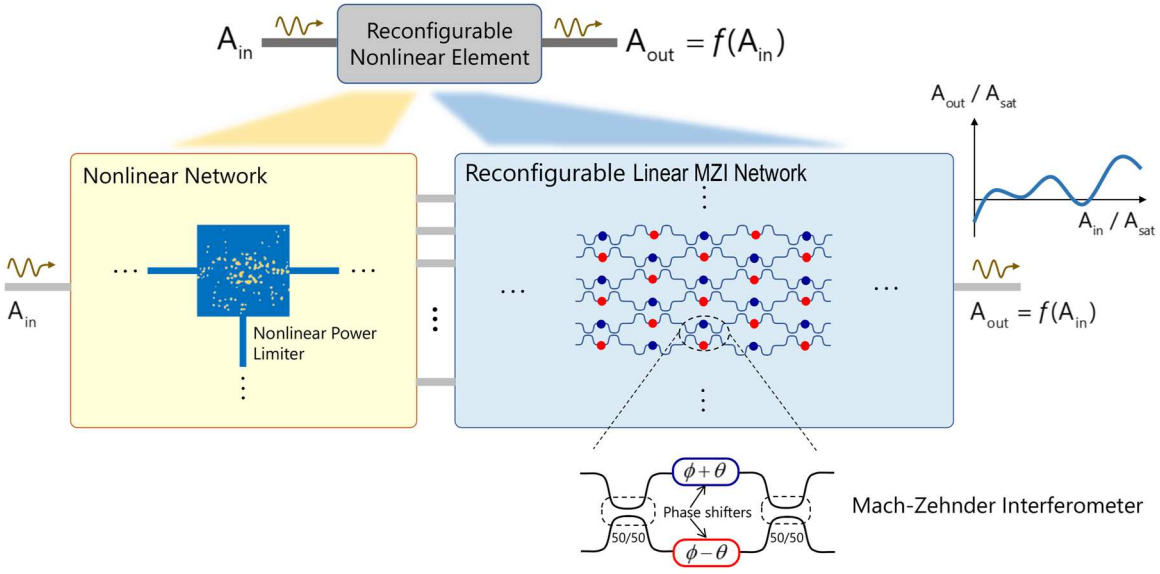

Figure 1: General reconfigurable optical architecture for generating tunable nonlinear functions. The architecture is composed of a nonlinear and a linear network. The nonlinear network (yellowish-filled rectangular box) consists of a set of identical nonlinear three-port devices, each of which acts as a nonlinear power limiter. The linear network (bluish-filled rectangular box) consists of a collection of linear MZIs. The inset sketches a generic MZI mesh configuration.

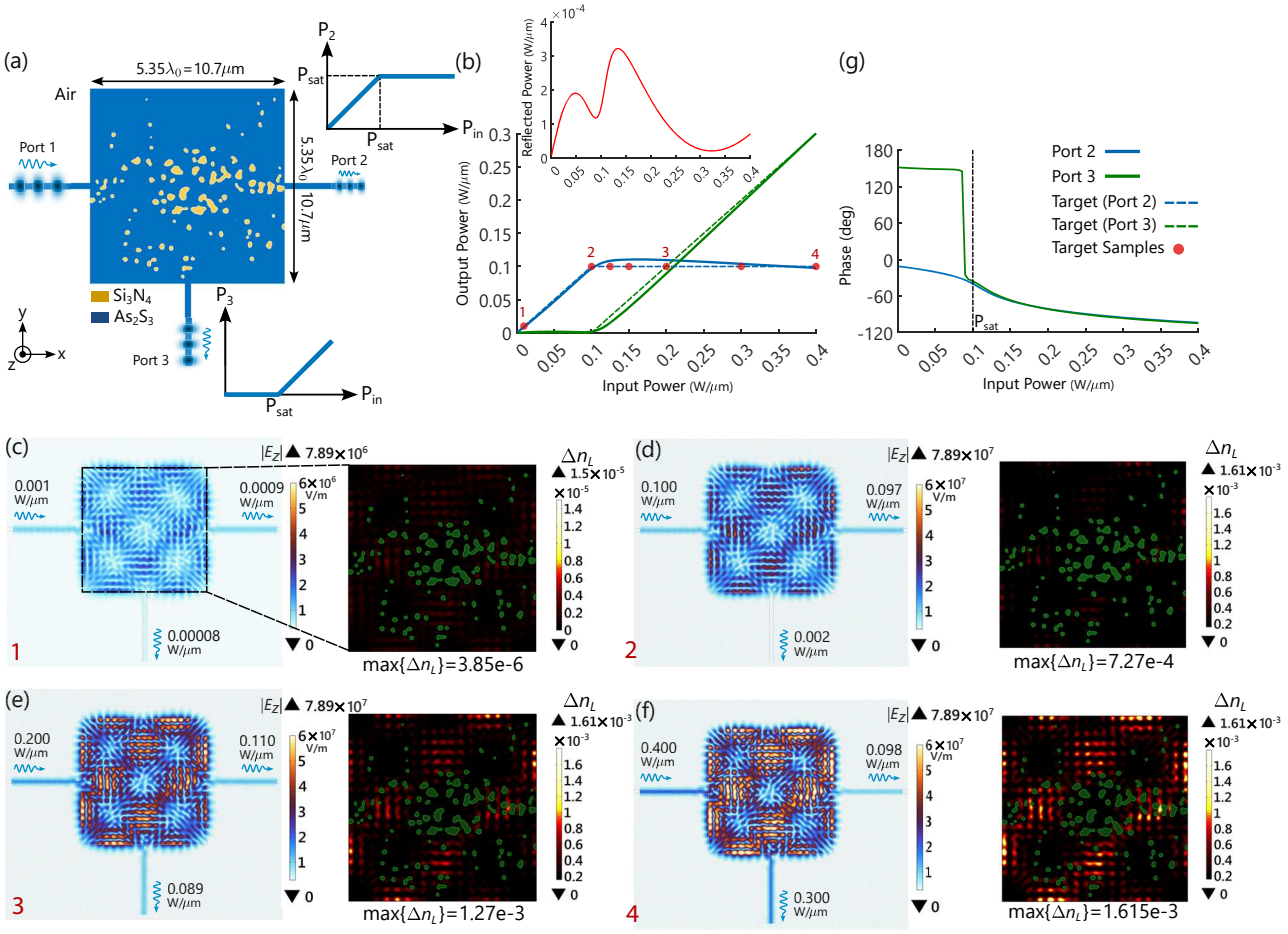

Figure 2: Inverse-designed nonlinear photonic structure as power limiter. (a) A transverse-electric (TE)-polarized optical signal is the input to the waveguide denoted as Port 1. The power of the output signal at Port 2 follows the power of the input signal but is limited when the input power increases beyond P_{sat}. The rest of the energy is then directed toward Port 3. The optimized distribution of the Kerr nonlinear material (As_2S_3) and the linear dielectric (Si_3N_4) are, respectively, shown by blue and yellow regions. (b) The blue and green solid curves, respectively, show the transmitted power of the signals at Port 2 and Port 3 of the proposed element in panel (a) versus the input power. The dashed curves are the corresponding desired transmission plots. The red solid curve illustrates the reflected power back to Port 1. The eight red circles are the target samples of the input power for which the cost function is defined for optimization. (c)–(f) The magnitude of the electric field distribution (left side of each panel) and the nonlinear refractive index shift inside the Kerr medium (right side of each panel) simulated for different input power values, $P_{in} = 0.01P_{sat}$, P_{sat}, $2P_{sat}$, $4P_{sat}$. (g) The phases of the output signals at Port 2 (blue solid curve) and Port 3 (green solid curve) relative to the input phase.

constant and equal to P_{sat}. Also, as will become clear shortly, it is required that the phases of the output signals at Ports 2 and 3 be the same (and, with an additional part discussed in the MZI section, to be the same as the phase of the input signal). To achieve these transmission characteristics, we consider a design region containing a distribution of Kerr nonlinear and linear dielectric materials connected to the input and output ports through slab waveguides as depicted in Figure 2(a). The cladding surrounding the structure is assumed to be air. Without loss of generality, we also assume the structure to be two-dimensional (2D) meaning it is infinitely extended in the out-of-plane direction (z axis). The materials of the structure are invariant along the out-of-plane coordinate (z) and, as a result, the distribution of the electromagnetic (EM) fields is computed only over the in-plane geometry (x–y plane). Our proposed approach here is general and the choice of materials depends on the frequency of operation. Here, as a nonlinear material, we used a realistic high-index Kerr material such as As_2S_3 (arsenic sulfide), which is modeled with the following intensity-dependent permittivity [32]:

$$\varepsilon_r^{\mathrm{NL}} = \varepsilon_r^{\mathrm{L}} + 3\chi^{(3)}|\mathbf{E}|^2 \tag{1}$$

where $\varepsilon_r^{\mathrm{L}}$ and $\chi^{(3)}$ denote the linear relative permittivity and the Kerr coefficient, respectively. At an operating free-space wavelength of $\lambda_0 = 2\,\mu\mathrm{m}$, As_2S_3 presents a linear

permittivity of $\varepsilon_r^{\mathrm{L}} = 5.76$, and a relatively high Kerr coefficient of $\chi^{(3)} = 4.1 \times 10^{-19}\ \mathrm{V^2/m^2}$. As shown in Figure 2(a), the slab waveguides' cores are assumed to be As_2S_3 with a width of $\lambda_0/6$. This width only allows the propagation of the fundamental TE (transverse electric) mode [$\mathbf{E} = (0, 0, E_z)$]. Within the design region, the distribution of As_2S_3 and a linear dielectric material such as Si_3N_4 (Silicon nitride with $\varepsilon_r^{\mathrm{Si_3N_4}} = 4$ at $\lambda_0 = 2\ \mu\mathrm{m}$) is optimized to achieve the desired transmission features. Note that this choice of materials is just an example and other materials can also be used. For example, at $\lambda_0 = 1.55\ \mu\mathrm{m}$ CMOS-compatible a-SiC (amorphous silicon carbide) [33] and SiO_2 (silica) can also be considered for the design, respectively, as the Kerr and linear dielectric materials. The distribution of the two materials (As_2S_3 and Si_3N_4) in the design region is obtained by the density-based topology optimization technique known as the inverse design method [15–18]. The goal of the optimization process is to determine the proper distribution of the two materials to provide the device with the transmission characteristics discussed above with *a priori* chosen saturation power e.g., $P_{\mathrm{sat}} = 0.1\ \mathrm{W/\mu m}$. (The choice of P_{sat} value here is an example, and in general, this value depends on the choice of the Kerr nonlinear material. Moreover, since we are dealing with a two-dimensional structure, the power is given as W/μm where μm is along the *z*-axis. We choose $P_{\mathrm{sat}} = 0.1\ \mathrm{W/\mu m}$ which provides a large enough refractive index shift inside As_2S_3 for our proposes and it is small enough that maintains a safe margin from damaging the materials within the design region.) Namely, the power injected at Port 1 is coupled to Ports 2 and 3 according to the blue and green dashed curves, respectively, as shown in Figure 2(b). The optimization procedure, which is discussed in detail in Supplementary Information, was carried out to match the transmission from Port 1 to Ports 2 and 3 at a discrete number of input power values (see the red circles in Figure 2(b)), resulting in the optimized distribution of As_2S_3 and Si_3N_4 shown in Figure 2(a). The distribution of the magnitude of the electric field ($|E_z|$) and the nonlinear refractive index shift ($\Delta n_{\mathrm{NL}} = \sqrt{\varepsilon_r^{\mathrm{L}} + 3\chi^{(3)}|E_z|^2} - \sqrt{\varepsilon_r^{\mathrm{L}}}$) that are simulated for the optimized device at four different input power values are shown in Figure 2(c)–(f). For instance, Figure 2(c) shows $|E_z|$ and Δn_{NL} for $P_{\mathrm{in}} = 0.01P_{\mathrm{sat}} = 0.001\ \mathrm{W/\mu m}$ which is well below the saturation power. With this level of input power, the nonlinear response of Kerr material induced by the local field intensity is very weak. The maximum observed nonlinear index shift inside the Kerr medium is 3.85×10^{-6} (see the right panel), which is basically negligible, and as a result, the device operates in the linear regime. As can be seen on the left panel, for $P_{\mathrm{in}} = 0.01P_{\mathrm{sat}}$, most of the input power is coupled to Port 2, $P_2 = 0.0009\ \mathrm{W/\mu m}$, as desired. Only a small amount of leakage is observed at Port 3, $P_3 = 0.00008\ \mathrm{W/\mu m}$. By increasing the input power up to the edge of the saturation i.e., $P_{\mathrm{in}} = P_{\mathrm{sat}}$, the nonlinear response of the Kerr material kicks in, as can be observed on the right panel of Figure 2(d). The maximum shift of the refractive index is 7.27×10^{-4}, which is two orders of magnitude larger than the one observed with $P_{\mathrm{in}} = 0.01P_{\mathrm{sat}}$ (see the right panels of Figure 2(c) and (d)). Nevertheless, most of the input power is still coupled to Port 2 ($P_2 = 0.097\ \mathrm{W/\mu m}$) as desired and there is a small leakage of $P_3 = 0.002\ \mathrm{W/\mu m}$ at Port 3. Also, a negligible portion of the input power equal to $P_{\mathrm{in}} - (P_2 + P_3) = 0.001\ \mathrm{W/\mu m}$ is either scattered away from the device to the air cladding or reflected back towards the input port, as shown in the inset of Figure 2(b). By bringing the input power above the saturation level, i.e., $P_{\mathrm{in}} = 2P_{\mathrm{sat}}$, the nonlinear response becomes more pronounced as indicated by the right panel of Figure 2(e) showing a maximum $\Delta n_{\mathrm{NL}} = 1.27 \times 10^{-3}$. As can be observed in the right panel of Figure 2(e), the collective power-dependent response is strong enough to limit the power at Port 2 ($P_2 = 0.110\ \mathrm{W/\mu m}$) around the level of the saturation power ($P_{\mathrm{sat}} = 0.100\ \mathrm{W/\mu m}$). The remaining power, $P_3 = 0.089\ \mathrm{W/\mu m}$, appears at Port 3, as expected. As the input power increases further beyond P_{sat}, i.e., $P_{\mathrm{in}} = 4P_{\mathrm{sat}}$, the increase of the nonlinear effect enables maintaining the power at Port 2 almost constant around P_{sat} and directing the extra power at Port 3 (see Figure 2(f)).

So far, we have discussed the performance of the three-port inverse-designed nonlinear power limiter for a discrete number of input power values. In Figure 2(b), the solid blue and green lines, respectively, show the transmitted power from Port 1 to Ports 2 and 3 which are calculated by continuously sweeping the input power from 0 to $4P_{\mathrm{sat}}$. As observed from the transmission plots, The agreement with the target responses (dashed curves) is quite good. For $P_{\mathrm{in}} \le P_{\mathrm{sat}}$, the power reaching Port 2 increases linearly. For the input powers beyond P_{sat}, the output power at Port 2 saturates around P_{sat} and the power going to Port 3 starts increasing almost linearly with P_{in}. Figure 2(g) shows the relative phase variation between the output signals (Ports 2 and 3) and the input signal (Port 1) versus the input power. For $P_{\mathrm{in}} < P_{\mathrm{sat}}$, the phase at Port 3 is immaterial, as the amount of power reaching this port is very small. For $P_{\mathrm{in}} > P_{\mathrm{sat}}$, the phase of the signals at Ports 2 and 3 assumes a similar behavior decreasing quasi-linearly as the input power increases. This phase variation is expected with Kerr nonlinear materials with a positive $\chi^{(3)}$. An increase in the input power implies an increase in the refractive index resulting in a decrease in the output signal phase (note that we are using $e^{j\omega t}$ as the time dependence). Considering that a properly arranged

set of these inverse-designed nonlinear three-port elements feed the MZI mesh that requires proper phase relations for its input, the phase behavior of the three-port output signals needs to be properly adjusted for the MZI mesh. This issue will be addressed in the following section. The MZI mesh works with complex-valued signals (A) containing both magnitude and phase information. The phases of the output signals leaving the inverse-designed nonlinear element are available in Figure 2(g). Their magnitudes ($|A|$) can be retrieved from the output power values (see Figure 2(b)) through the expression $P = \frac{1}{2Z_t}\int |A\psi(l)|^2 \mathrm{d}l$. Z_t denotes the transverse impedance of the fundamental TE mode supported by the input/output waveguides (For the lossless waveguide, which we assumed here, this transverse impedance is real-valued.) $\psi(.)$ is the eigenmode profile that is normalized such that $\int |\psi(l)|^2 \mathrm{d}l = 1$, with l denoting the coordinate along a slab waveguide's cross-section. Note that, in doing so, $\psi(.)$ is dimensionless whereas A has the dimension of the electric field [V/m]. The magnitude of the signals leaving the nonlinear element can be calculated from the power as $|A| = \sqrt{2Z_t P}$.

3 Adding reconfigurable MZI network

In the previous section, the design of the nonlinear power limiter with the specific relationship between the input and the two output signals was presented. As shown in Figure 3(a), we cascade $N-1$ such nonlinear elements with identical functionality into a special 1-to-N nonlinear signal divider (NSD). Then we feed the outputs of the NSD to a properly arranged collection of MZIs. The MZI network processes and combines the complex-valued signals coming from the NSD to provide an output signal that is a tunable nonlinear function of the input signal i.e., $A_{\mathrm{out}} = f(A_{\mathrm{in}})$. Without loss of generality, we consider A_{in}, which denotes the input signal to the proposed architecture, to be a real positive number. On the other hand, the output signal (A_{out}) can be a positive or negative real number. As will be clear shortly, the proposed architecture enables generating an output signal (A_{out}) with 0 or π phase despite the positive input signal, which is a feature to implement a large set of nonlinear functions. Now, let us discuss the functionality of the proposed reconfigurable nonlinear architecture beginning with the nonlinear signal divider. As can be seen in Figure 3(a), Port 3 of each nonlinear element is connected to the input port of the next one, and Ports 2 are considered the outputs of the NSD. As the magnitude of the input signal (A_{in}) increases from zero, Port 2 of the first nonlinear element gets activated, and the magnitude of its signal increases up to the saturation level $|A_{\mathrm{sat}}|$, which is the magnitude corresponding to P_{sat}. As A_{in} goes beyond $|A_{\mathrm{sat}}|$, the spillover goes to Port 3, which is connected to the input port of the second nonlinear element. Therefore, Port 2 of the second nonlinear element gets activated next. This process continues until the magnitudes of all the output signals are saturated at $|A_{\mathrm{sat}}|$. For instance, let us assume the magnitude of the input signal is between $\sqrt{2}|A_{\mathrm{sat}}| \le A_{\mathrm{in}} \le \sqrt{3}|A_{\mathrm{sat}}|$ (corresponding to $2P_{\mathrm{sat}} < P_{\mathrm{in}} < 3P_{\mathrm{sat}}$). Within this interval, the magnitudes of the signals at the first two output ports of the NSD are fixed at $|A_{\mathrm{sat}}|$, whereas the difference between the input power and the power of the saturated signals appears at the third output port i.e., $A_3 = \sqrt{|A_{\mathrm{in}}|^2 - 2|A_{\mathrm{sat}}|^2}$. In general, for the input amplitude in the range $\sqrt{m-1}|A_{\mathrm{sat}}| \le |A_{\mathrm{in}}| \le \sqrt{m}|A_{\mathrm{sat}}|$ ($m \le N$), the magnitudes of the outputs A_k, with $k = 1, 2, \ldots, m-1$, are fixed at $|A_{\mathrm{sat}}|$, the magnitude of the mth output signal is $A_m = \sqrt{|A_{\mathrm{in}}|^2 - (m-1)|A_{\mathrm{sat}}|^2}$, and the rest are zero, $A_k = 0, k = m+1, m+2, \ldots, N$. Figure 3(b) shows the magnitude of the first five output signals as a function of A_{in}. The signals leaving the NSD can be seen as a set of nonlinear basis functions that are properly processed upon entering the MZI mesh to generate a continuous nonlinear output signal (A_{out}) as a reconfigurable function of the input signal (A_{in}).

Each MZI consists of two 50 % beam splitters and two phase shifters parametrized by θ, the differential phase, and ϕ, the common phase [5, 13, 20, 27], as shown in the inset at the bottom of Figure 3(a). In silicon photonics, beam splitters are usually realized by directional couplers that transform complex-valued input signals $\mathbf{a}_1$ and $\mathbf{a}_2$ into complex-valued output signals $\mathbf{b}_1$ and $\mathbf{b}_2$ according to $\mathbf{b}_{1,2} = \frac{1}{\sqrt{2}}\left(\pm j\mathbf{a}_{1,2} + \mathbf{a}_{2,1}\right)$ [27]. Assuming lossless beam splitters and phase shifters, the MZI implements a 2×2 unitary transformation between the input and output signals that can be mathematically described by the following transmission matrix [27]:

$$\mathbf{T}_{\mathrm{MZI}} = \begin{pmatrix} -j\sin(\theta)e^{j\phi} & -j\cos(\theta)e^{j\phi} \\ -j\cos(\theta)e^{j\phi} & j\sin(\theta)e^{j\phi} \end{pmatrix} \tag{2}$$

As depicted in Figure 3(a), the linear network of MZI mesh can be broken up into two sub-networks: (1) the MZI collection that sets the "slopes" (denoted as the "slope screen" in Figure 3(a)), which are the coefficients that would be multiplied by the $A_m = \sqrt{|A_{\mathrm{in}}|^2 - (m-1)|A_{\mathrm{sat}}|^2}$, and (2) the MZI collection that coherently combines the signals (depicted as the "adder" section). Namely, it performs the addition of N input signals into one output signal. In setting up the phases of each MZI to achieve such adding functionality, the phases of the N input signals are assumed to be in

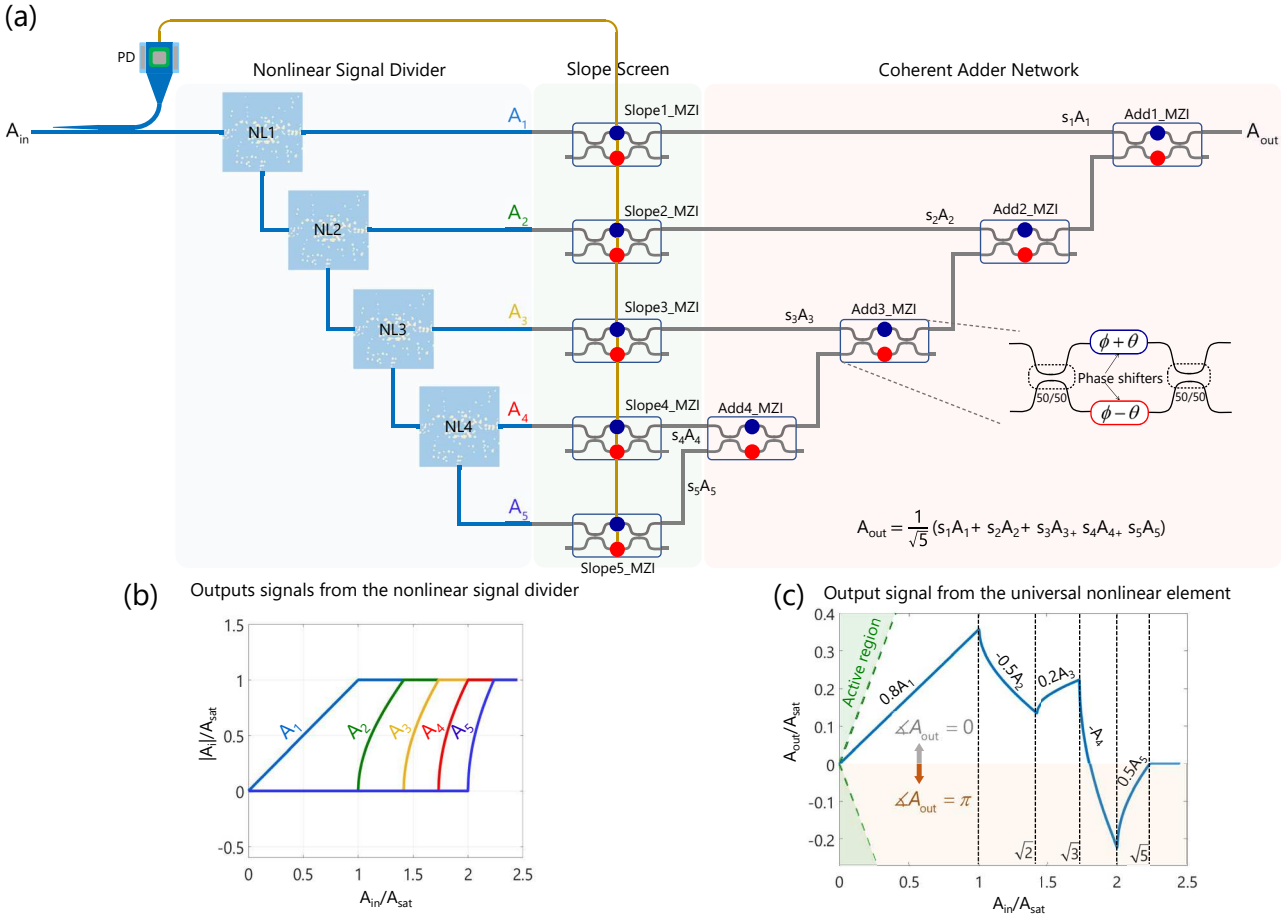

Figure 3: Schematic illustration of our proposed nonlinear architecture composed of a nonlinear signal divider (NSD) and a linear network of MZIs. The 1-to-N nonlinear signal divider is a cascaded architecture of $N-1$ individual nonlinear power limiters that distribute the input signal with amplitude A_{in} to N output ports in a particular way, as the input power increases. Two MZI sets identified as "slope screen" and "coherent adder network" are the sub-parts of the linear MZI network. The slope screen scales the outputs of NSD each with a coefficient $-1 \le s \le 1$. The adder network adds the scaled signals with an equal weight of $1/\sqrt{N}$ to generate the output signal with amplitude A_{out}. A tap waveguide connected to a photodetector produces a power-dependent voltage signal that modifies the common phases of the MZIs in the slope screen for compensating the phases of the nonlinear limiters. (b) The amplitudes of the signals at the 5 outputs of NSD. These signals form a set of basis functions for generating a tunable function of the input signal i.e., $A_{out} = f(A_{in})$ over a specific range of the input power. (c) An example of the output signal as a function of the input signal for the slopes $s = [0.8, -0.5, 0.2, -1, 0.5]$ of the slope-MZIs that are set from top to bottom. A negative output means a π phase shift relative to the input.

phase. The phase shifters of those MZIs are set progressively from the top to the bottom when one is to run the adder backward: sending a signal from the output port and imposing the power carried by that signal to be equally divided and appear at the input ports with the same phase. [27, 34]. As discussed in the previous section, the input and outputs of each nonlinear power limiter experience a phase difference that decreases quasi-linearly with the input power as shown in Figure 2(g). Cascading the nonlinear power limiters as shown in Figure 3(a) implies that the output phases of the NSD also decrease quasi-linearly with the input power but with a slope that increases progressively with the number of activated nonlinear elements. For example, let us assume the power of the input signal is within the range such that the first two nonlinear elements are activated. The phase of the first NSD output (A_1) will decrease quasi-linearly with the input power as shown in Figure 2(g). The phase of the second output (A_2) will decrease as A_1 but with a double slope because the second nonlinear limiter is fed with a signal whose phase profile is identical to the one of A_1. A similar argument can be repeated for the other nonlinear elements. To compensate for this phase difference and equalize the phases of the NSD outputs with the phase of the input signal (A_1) as required by the adder, a voltage-driven phase shifter [5, 13, 24, 25] can be used to change the common phases of the MZIs in the slope screen based on the power of the input signal. Specifically, on the input side of the proposed architecture (see Figure 3(a)),

a tap waveguide connected to a photodetector producing a voltage proportional to the input power can be used to modify the common phases of the MZIs to compensate for the phase variation of the NSD outputs. As an aside, it is worth mentioning that for the sake of simplicity and just to describe the concept, we assume that the mode propagation in all waveguides in each MZI experiences the same phase shift when they get to the adder network. If that is not the case the phase differences due to the different lengths of these waveguides can be easily compensated for by the common phase shifter in each MZI. Following this approach, the common phase of mth slope-MZI is set to $\phi_m(P_{\text{in}}) = \pi/2 + m\Delta\psi(P_{\text{in}})$ where $\Delta\psi(P_{\text{in}})$ is the required phase adjustment produced by the photodetector signal to make the first NSD output (A_1) coherent with the input signal (A_{in}). After setting the common phases according to $\phi_m(P_{\text{in}})$, the differential phase (θ_m) determines the coupling between the mth NSD output and the mth input to the adder network according to $s_m = \sin(\theta_m)$ with s_m (θ_m) ranging from -1 ($-\pi/2$) to 1 ($\pi/2$) (see the first entry of the matrix in Eq. (2)). The adder combines the signals that are leaving the slope screen into the output signal, which will be of the form $A_{\text{out}} = \frac{1}{\sqrt{N}}\sum_{m=1}^{N} s_m A_m$. The details regarding the design of the adder network are provided in the Supplementary Information. As can be observed from the previous expression, the differential phase θ_m sets the slope of the portion of A_{out} for the input amplitude between $\sqrt{m-1}A_{\text{sat}}$ and $\sqrt{m}A_{\text{sat}}$. By tuning these differential phases of the MZIs in the slope screen, the architecture depicted in Figure 3(a) can approximately provide a variety of different nonlinear transformations, $A_{\text{out}} = f(A_{\text{in}})$, through a continuous piecewise function (for any multiplicative coefficient between -1 and $+1$ before the A_m, and then multiplied by $1/\sqrt{N}$), defined on intervals of input magnitude $[\sqrt{m-1}A_{\text{sat}}, \sqrt{m}A_{\text{sat}}]$. Assuming the phase shifters of the MZIs in the slope screen are implemented with voltage-driven phase shifters, as suggested above, the differential phases can be tuned by combining the photodetector signal with an extra electrical signal. Thus, the phase shifters in the slope screen allow controlling the phases of the NSD outputs and setting the slopes (s_m) of the output signal (A_{out}). Figure 3(c) shows an example of a nonlinear function realized with the proposed architecture assuming $N = 5$. For this example, the differential phases of the slope screen are arbitrarily chosen to be $\mathbf{s} = [0.8, -0.5, 0.2, -1, 0.5]$. During the first interval of the input magnitude, corresponding to ($[0, A_{\text{sat}}]$), all the signals leaving the NSD are zero except for A_1, and, as a result, the output signal is given by $A_{\text{out}} = \frac{1}{\sqrt{5}}\mathbf{s}(1)A_1$. Since the first MZI of the slope screen sets a positive slope ($\mathbf{s}(1) = 0.8$), the output signal increases with A_{in} as can be observed in Figure 3(c). When A_{in} goes beyond A_{sat}, entering the second interval of the input magnitude ($[A_{\text{sat}}, \sqrt{2}A_{\text{sat}}]$), the second nonlinear limiter in the NSD is activated in addition to the first one. The output signal is now given by $A_{\text{out}} = \frac{1}{\sqrt{5}}(\mathbf{s}(1)A_1 + \mathbf{s}(2)A_2)$. As discussed above, the second nonlinear limiter in the NSD is activated when the output of the first one (A_1) is saturated, that is, $|A_1|$ is constant with respect to the input power. Thus, the slope of the output during the interval $[A_{\text{sat}}, \sqrt{2}A_{\text{sat}}]$ is dictated by the differential phase (θ_2) of the second MZI of the slope screen. In the example under consideration, θ_2 was selected to be $-\pi/6$ and the output signal decreases with the slope $\mathbf{s}(2) = -0.5$ when the input magnitude goes from A_{sat} to $\sqrt{2}A_{\text{sat}}$ as illustrated in Figure 3(c). A similar argument can be repeated when the input power increases further activating all the five nonlinear limiters in the considered example. As can be observed in Figure 3(c), interestingly, the proposed architecture can generate positive and negative nonlinear output signals enabling the implementation of a large set of nonlinear functions. As the proposed architecture does not include any active components, the output signals are always confined in the region bounded by the bisector of the first and fourth quadrants of the $A_{\text{out}} - A_{\text{in}}$ Cartesian diagram (see Figure 3(c)). We consider four different arbitrarily chosen nonlinear functions as examples for demonstrating the capabilities of the proposed architecture and its *reconfigurability* in implementing nonlinear functions of the input signal. The first one is a quadratic function of the form $f(x) = x^2/4$. The target function of the input signal magnitude considering $x = A_{\text{in}}/A_{\text{sat}}$ is shown in Figure 4(a) (solid red line). Assuming $N = 5$, the differential phases of the five MZIs in the slope screen have been optimized to match the realized nonlinear output signal with the target quadratic function. The blue solid line in Figure 4(a) shows the output signal obtained by our proposed architecture with the optimized slopes $\mathbf{s} = [0.42, 0.61, 0.56, 0.49, 0.87]$. As observed in Figure 4(a), the agreement of the realized output signal with the target quadratic function is acceptable. The same architecture can also implement, for instance, a third-degree polynomial function (Figure 4(b)), a sigmoid function (Figure 4(c)), and a transcendental function (Figure 4(d)), just by properly tuning the differential phases of the MZIs in the slope screen. Also in these three examples, the realized output signal by the proposed architecture closely follows the target nonlinear functions. Just like the quadratic function, the required differential phases of the MZIs in the slope screen for implementing the other three nonlinear functions have been obtained through an optimization process. The resulting slopes are shown in the insets of Figure 4(b)–(d). We point out that the optimization process to characterize the

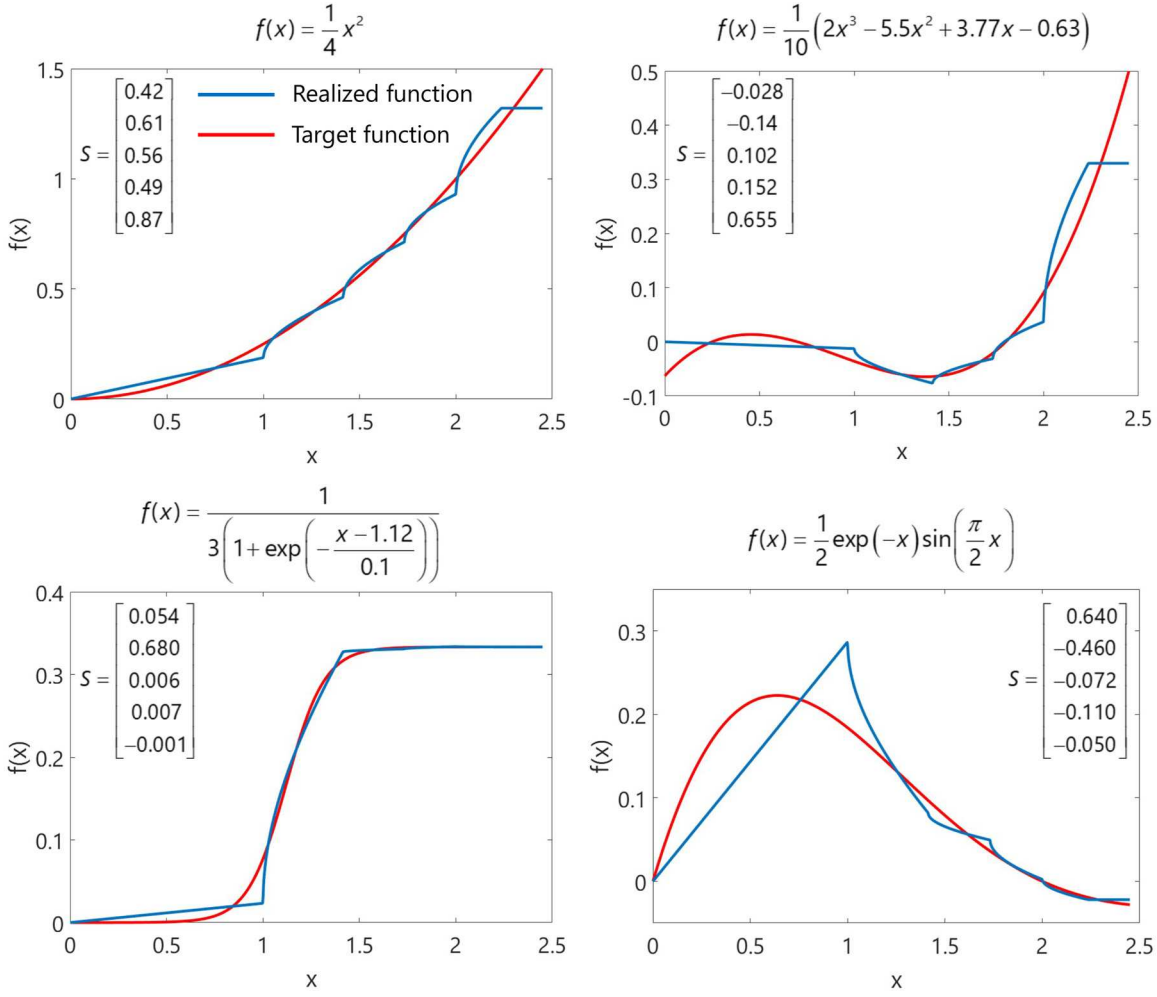

Figure 4: Examples of different target nonlinear functions of the input signal by our proposed reconfigurable nonlinear element shown in Figure 2(a). Red curves are the target functions of the input signal the expression of which are shown at top of the plots. The blue curves are the resulting functions from our proposed element, by optimizing the scales of the slope screen using minimization of the MSE error between target and realized functions. The inset in each panel shows the optimized scales of the corresponding realized function.

slope screen for implementing a nonlinear function needs to be run only once and usually takes little time (e.g., less than a second).

4 Summary and conclusions

In this work, we theoretically proposed and numerically demonstrated a device that can exhibit reconfigurable nonlinear dependence between the input and output signals. It consists of two networks, the first of which is a set of several identical inverse-designed three-port structures with an optimized mixture of linear and nonlinear materials, providing desired power limiting and power dividing characteristics. The second network is a linear mesh of MZI elements that consists of two sections; the slope screen and the adder networks. The reconfigurability of the MZI mesh enables one to change and tune the functional dependence of the entire device at will. Several examples illustrating salient features of this device were given and discussed. The fact that the proposed architecture is reconfigurable makes it useful for numerous applications such as optical neural networks and wave-based analog computing architectures with nonlinearity.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Research funding: This work is supported in part by the US Air Force Office of Scientific Research (AFOSR) Multidisciplinary University Research Initiative (MURI) grant numbers FA9550-17-1-0002 and FA9550-21-1-0312, and in part by

the US National Science Foundation (NSF) MRSEC program under award No. DMR-1720530.

Conflict of interest statement: The authors declare no conflicts of interest.

Data Availability Statement: The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

[1] X. Guo, C.-L. Zou, H. Jung, and H. X. Tang, "On-chip strong coupling and efficient frequency conversion between telecom and visible optical modes," *Phys. Rev. Lett.*, vol. 117, no. 12, p. 123902, 2016.

[2] Y. Jiang, P. T. DeVore, and B. Jalali, "Analog optical computing primitives in silicon photonics," *Opt. Lett.*, vol. 41, no. 6, pp. 1273—1276, 2016.

[3] M. Lapine, I. V. Shadrivov, and Y. S. Kivshar, "Colloquium: nonlinear metamaterials," *Rev. Mod. Phys.*, vol. 86, no. 3, p. 1093, 2014.

[4] A. Krasnok, M. Tymchenko, and A. Alù, "Nonlinear metasurfaces: a paradigm shift in nonlinear optics," *Mater. Today*, vol. 21, no. 1, pp. 8—21, 2018.

[5] Y. Shen, N. C. Harris, S. Skirlo, et al., "Deep learning with coherent nanophotonic circuits," *Nat. Photonics*, vol. 11, no. 7, pp. 441—446, 2017.

[6] R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," *Phys. Rev. X*, vol. 9, no. 2, p. 021032, 2019.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84—90, 2012.

[8] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55—75, 2018.

[9] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, pp. 1—21, 2015.

[10] D. Silver, T. Hubert, J. Schrittwieser, et al., "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, no. 6419, pp. 1140—1144, 2018.

[11] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep learning for identifying metastatic breast cancer," *arXiv preprint arXiv:1606.05718*, 2016.

[12] B. Karlik and A. V. Olgac, "Performance analysis of various activation functions in generalized mlp architectures of neural networks," *Int. J. Artif. Intell. Expert Syst.*, vol. 1, no. 4, pp. 111—122, 2011.

[13] N. C. Harris, J. Carolan, D. Bunandar, et al., "Linear programmable nanophotonic processors," *Optica*, vol. 5, no. 12, pp. 1623—1631, 2018.

[14] M. P. Bendsoe and O. Sigmund, *Topology Optimization: Theory, Methods, and Applications*, 2nd ed. Berlin/Heidelberg, Springer Science & Business Media, 2003.

[15] J. Lu, S. Boyd, and J. Vučković, "Inverse design of a three-dimensional nanophotonic resonator," *Opt. Express*, vol. 19, no. 11, pp. 10563—10570, 2011.

[16] A. Y. Piggott, J. Lu, K. G. Lagoudakis, J. Petykiewicz, T. M. Babinec, and J. Vučković, "Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer," *Nat. Photonics*, vol. 9, no. 6, pp. 374—377, 2015.

[17] A. Y. Piggott, J. Petykiewicz, L. Su, and J. Vučković, "Fabrication-constrained nanophotonic inverse design," *Sci. Rep.*, vol. 7, no. 1, pp. 1—7, 2017.

[18] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez, "Inverse design in nanophotonics," *Nat. Photonics*, vol. 12, no. 11, pp. 659—670, 2018.

[19] T. W. Hughes, M. Minkov, I. A. Williamson, and S. Fan, "Adjoint method and inverse design for nonlinear nanophotonic devices," *ACS Photonics*, vol. 5, no. 12, pp. 4781—4787, 2018.

[20] D. A. Miller, "Sorting out light," *Science*, vol. 347, no. 6229, pp. 1423—1424, 2015.

[21] D. A. Miller, "Establishing optimal wave communication channels automatically," *J. Lightwave Technol.*, vol. 31, no. 24, pp. 3987—3994, 2013.

[22] D. A. Miller, "Self-aligning universal beam coupler," *Opt. Express*, vol. 21, no. 5, pp. 6360—6370, 2013.

[23] C. Taballione, T. A. Wolterink, J. Lugani, et al., "8× 8 reconfigurable quantum photonic processor based on silicon nitride waveguides," *Opt. Express*, vol. 27, no. 19, pp. 26842—26857, 2019.

[24] W. Bogaerts, D. Pérez, J. Capmany, et al., "Programmable photonic circuits," *Nature*, vol. 586, no. 7828, pp. 207—216, 2020.

[25] N. C. Harris, G. R. Steinbrecher, M. Prabhu, et al., "Quantum transport simulations in a programmable nanophotonic processor," *Nat. Photonics*, vol. 11, no. 7, pp. 447—452, 2017.

[26] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, no. 1, p. 58, 1994.

[27] D. A. Miller, "Self-configuring universal linear optical component," *Photonics Res.*, vol. 1, no. 1, pp. 1—15, 2013.

[28] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers," *Optica*, vol. 3, no. 12, pp. 1460—1465, 2016.

[29] V. Nikkhah, D. C. Tzarouchis, A. Hoorfar, and N. Engheta, "Inverse-designed metastructures together with reconfigurable couplers to compute forward scattering," *ACS Photonics*, vol. 10, no. 4, pp. 977—985, 2022.

[30] M. J. Mencagli, N. M. Estakhri, B. Edwards, and N. Engheta, "Solving equations with waves in collections of mach-zehnder interferometers," in *2018 Conference on Lasers and Electro-Optics (CLEO)*, IEEE, 2018, pp. 1—2.

[31] D. C. Tzarouchis, M. J. Mencagli, B. Edwards, and N. Engheta, "Mathematical operations and equation solving with reconfigurable metadevices," *Light: Sci. Appl.*, vol. 11, no. 1, pp. 1—13, 2022.

[32] R. W. Boyd, *Nonlinear Optics*, 3rd ed. Cambridge, Massachusetts, Academic Press, 2008.

[33] P. Xing, D. Ma, K. J. Ooi, J. W. Choi, A. M. Agarwal, and D. Tan, "Cmos-compatible pecvd silicon carbide platform for linear and nonlinear optics," *ACS Photonics*, vol. 6, no. 5, pp. 1162—1167, 2019.

[34] D. A. Miller, "Perfect optics with imperfect components," *Optica*, vol. 2, no. 8, pp. 747—750, 2015.

Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/nanoph-2023-0152).