Multidimensional Convolution Operation with Synthetic Frequency Dimensions in Photonics

Lingling Fan,[1] Zhexin Zhao,[1] Kai Wang,[1] Avik Dutt,[2] Jiahui Wang,[3] Siddharth Buddhiraju,[1] Casey C. Wojcik,[1] and Shanhui Fan[1,*]

[1] *Department of Electrical Engineering, Ginzton Laboratory, Stanford University, Stanford, California 94305, USA*

[2] *Department of Mechanical Engineering and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA*

[3] *Department of Applied Physics, Stanford University, Stanford, California 94305, USA*

(Received 26 May 2022; revised 17 July 2022; accepted 22 July 2022; published 30 September 2022)

The convolution operation is widely used in signal and image processing and represents the most computationally intensive step in convolutional neural networks. We introduce a scheme to achieve arbitrary convolution kernels in the synthetic frequency dimension with a simple setup consisting of a ring resonator incorporating a phase and an amplitude modulator. This scheme can be used to perform multidimensional convolutions. We provide an analytic approach that determines the required modulation profile for any convolution kernel. Our work points to a direction of using optical computing to remove the computational bottleneck in traditional electronic circuits and may be useful in improving machine-learning hardware in artificial-intelligence applications.

DOI: 10.1103/PhysRevApplied.18.034088

I. INTRODUCTION

Artificial neural networks [1] have demonstrated state-of-the-art performance in machine-learning tasks such as image, video, speech, and text processing. Among these networks, convolutional neural networks (CNNs) [2] play a particularly important role in extracting hierarchical features from complex raw data, as they mimic characteristics of biological neural perception. In addition, CNNs are capable of making correct predictions based on unseen data, without increasing parameter complexities [3,4].

In CNNs, an important class of tasks, including spatiotemporal perception [5], require the convolution of large-scale data encoded in multidimensional matrices, which is energy consuming using conventional electronic hardware due to the data-movement bottleneck [6]. To overcome this bottleneck, optical neural networks (ONNs) [7] perform linear algebra tasks more energy efficiently by simply propagating the optical signals through a structure [8,9]. ONNs can also increase computing speed and lower energy consumption. For example, Mach-Zehnder interferometers (MZIs) have been employed in integrated photonic circuits to achieve linear transformations [10–12]. Microring resonators have been used as reservoir computing neurons [13,14]. Diffractive [15,16] and scattering media [17] have been used as analog hardware platforms for image and vowel classification tasks. Recently, state-of-the-art ONNs with high parallelism [18] and high-speed [19–21] operations have been demonstrated, with the speed reaching 10^{12} operations per second.

For many computational tasks, ONNs need to be compact and scalable to process input data and encode parameters on large scales. Linear transformation of N input signals is described by a $N \times N$ matrix with $\mathcal{O}(N^2)$ degrees of freedom. In the MZI implementation, the area of the device also scales as $\mathcal{O}(N^2)$ in order to provide the degrees of freedom in the $N \times N$ matrix [22]. Recently, there have been efforts to realize more scalable devices for linear transformation, by employing the internal degrees of freedom of photons [23]. Frequency is an important intrinsic degree of freedom of light and its manipulation based on the concept of the synthetic frequency dimension in dynamically modulated ring resonators has attracted growing interest for both the explorations of fundamental physics and optical information processing [24–28]. Compared with spatial encoding, the synthetic frequency dimension enables a compact spatial footprint for manipulating photons in both classical and quantum domains [29]. Using the photonic synthetic frequency dimension, a recent work shows that it is possible to realize an arbitrary linear transformation [23] with multiple rings connected in series. In Ref. [23], to realize the linear transformation of N input frequencies, the number of rings scales as N. Part of the required N^2 degrees of freedom is now compactly encoded in the modulation tones, as opposed to the spatial coupling constants as in the MZI configuration.

*shanhui@stanford.edu

The work of Ref. [23] has implemented a linear transformation described by a dense $N \times N$ matrix. For this purpose, small auxiliary rings have been introduced to break the natural translational symmetry in frequency space for a dynamically modulated ring. Here, we note that for convolution tasks, it is not necessary to break such translational symmetry. Instead, the natural translational symmetry along the frequency dimension in modulated ring resonators can be harnessed to perform convolutions, resulting in a configuration that is far simpler for practical implementations. Moreover, in synthetic frequency dimensions, modulations at higher multiples of the free spectral range (FSR) of a resonator enable long-range couplings between farther-apart frequency modes. Such long-range coupling has been used in Ref. [26,30] to synthesize a multidimensional Hamiltonian. It should be of interest to extend this approach to multidimensional convolutions and hence accelerate signal processing.

In this work, we propose a scheme for convolution based on synthetic frequency dimensions [31,32] using a single optical ring resonator undergoing dynamic modulations. The convolution is achieved using the scattering matrix of such a modulated system with a discrete frequency input matching the free spectral range of the ring resonator. We use both a phase modulator and an amplitude modulator to obtain both unitary and nonunitary scattering matrices, analogous to recent experiments demonstrating non-Hermitian physics in synthetic dimensions [27]. We analytically develop a deterministic closed-form expression to directly obtain the modulation parameters for the desired convolution kernels. We show that the kernel implemented can perform multidimensional convolutions, analogous to the working principles of synthesizing higher dimensions using multiple orders of couplings developed in synthetic dimensions [30,33]. Specifically, we verify such convolution with two-dimensional (2D) images. We introduce an approach to performing the convolution on large-scale images by judiciously slicing the input data, without the need for high modulation frequencies. We also extend our scheme to higher-dimensional convolution cases where the input and output data contain several channels, such as videos and LIDAR scans. Our scheme provides a means of achieving multidimensional convolution in a compact and configurable manner.

This paper is organized as follows. In Sec. II, we present the working principles for convolution by using the photonic synthetic frequency dimension. In Sec. III, we demonstrate 2D convolution in images, highlighting some of the detailed considerations in modulation for symmetric and asymmetric kernel matrices. In Sec. IV, we discuss an approach that slices the image in order to reduce the required modulation bandwidth. This slicing approach is of interest for convolution on a larger image. In Sec. V, we demonstrate a three-dimensional (3D) convolution case. In Sec. VI, we provide concluding remarks.

II. THEORY

A. The synthetic frequency dimension

1. Modulated ring resonator

This work uses a dynamically modulated ring resonator sketched in Fig. 1(a). The ring resonator and the coupling waveguide are both formed by a single-mode waveguide. In the absence of group-velocity dispersion and modulation, the ring resonator supports equally spaced longitudinal modes $\omega_n = \omega_0 + n\Omega$, where n is an integer indexing the modes that are separated by the FSR as given by $\Omega/2\pi = c/n_g\ell$, with ω_0, c, n_g, and ℓ being the central frequency, the speed of light, the group index, and the circumference of the ring, respectively. Inside the ring resonator, we place a phase modulator and an amplitude modulator. Both modulators are assumed to be spatially compact and, together, the two modulators produce a time-dependent

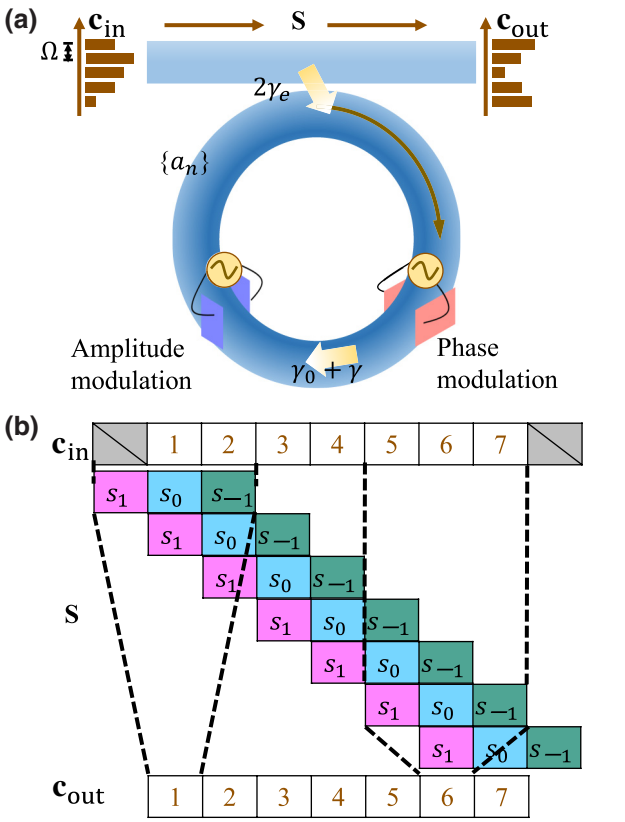

FIG. 1. (a) A schematic of the modulated ring with simultaneous modulation in amplitude and phase at the frequency of free spectral range $\Omega/2\pi$ and its integer multiples. The ring supports resonant modes $\{a_n\}$ and has an input-output coupling rate γ_e and an intrinsic decay rate γ_0. The scattering matrix $\mathbf{S}$ from the modulated ring resonator converts the input $\mathbf{c}_{in}$ to the output $\mathbf{c}_{out}$. (b) A one-dimensional (1D) convolution operation with the kernels s_{-1}, s_0 and s_1 maps the input $\mathbf{c}_{in}$ to the output $\mathbf{c}_{out}$, which can be completed with a scattering matrix $\mathbf{S}$ of translational symmetry.

transmission factor $T(t)$ [27],

$$T(t) = T_{\mathrm{Ph}}(t)T_{\mathrm{Am}}(t), \tag{1}$$

where

$$T_{\mathrm{Ph}}(t) = \exp\left[-j\sum_{m\geq 1} A_m \cos(m\Omega t + \alpha_m)\right], \tag{2}$$

$$T_{\mathrm{Am}}(t) = \exp\left[\sum_{m\geq 1} B_m \sin(m\Omega t + \beta_m) - \gamma t_R\right], \tag{3}$$

correspond to the time-dependent transmission factors for the phase and amplitude modulators, respectively. A_m (B_m) and α_m (β_m) describe the magnitude and phase angle of the mth order of the frequency components in the phase (amplitude) modulations, respectively. The time-independent term γt_R in the exponent of Eq. (3), where $\gamma > 0$ and $t_R = 2\pi/\Omega$ denotes the round-trip time of the ring, describes a background loss due to the amplitude modulator. This loss is important in order to ensure the passivity of the device, i.e., a device without the need of amplification, as we discuss in more detail in Sec. II C. We choose the modulation signal to have the same period as t_R such that $T(t) = T(t + t_R)$, so that a large number of modes can be resonantly coupled together.

In our discussion, we assume that all the modes of interest in the ring resonator in the absence of modulation have the same intrinsic decay rate γ_0, which accounts for all sorts of internal losses including, but not limited to, waveguide bending loss and material loss. We also assume that the input-output coupling rate γ_e between the coupling waveguide and the ring resonator is the same for all the modes of interest in the ring resonator. To ensure that the neighboring resonant modes are well separated, the line width of each mode $\gamma_e + \gamma_0$ in the absence of modulation is required to be much smaller than the FSR. Throughout the paper, we assume that we use an in-coupling beam splitter with a power splitting ratio of $z = 50\%$ between the input port and the cavity [27], where the corresponding input-output coupling rate is $\gamma_e = -\ln(1-z)/2t_R \approx 0.0552\ \Omega$.

2. Input and output

We compute the input-output relation for the setup discussed in the previous section. For this purpose, we denote the amplitude of the nth mode in the ring resonator as $a_n(\tau)\exp(j\omega_n t)$, where τ is a slow time variable depending on the number of round trips. Similarly, we denote the amplitude of the modes in the coupling waveguide with a frequency centered around ω_n at the input and output ports as $c_{\mathrm{in},n}(\tau)\exp(j\omega_n t)$ and $c_{\mathrm{out},n}(\tau)\exp(j\omega_n t)$, respectively. Thus, the dynamics of the modulated ring resonator coupling to a waveguide can be described by the formalism of the temporal coupled-mode theory [23,32]:

$$\frac{d}{d\tau}a_n = -j\sum_{m\geq 1}[\kappa_m a_{n-m} + \kappa_{-m}a_{n+m}] - (\gamma + \gamma_0 + \gamma_e)\,a_n + j\sqrt{2\gamma_e}c_{\mathrm{in},n}, \tag{4}$$

$$c_{\mathrm{out},n} = c_{\mathrm{in},n} + j\sqrt{2\gamma_e}a_n, \tag{5}$$

where the coupling coefficients induced by the dynamic modulation are given by

$$\kappa_{+m} = \frac{1}{2t_R}(A_m e^{+j\alpha_m} + B_m e^{+j\beta_m}), \tag{6}$$

$$\kappa_{-m} = \frac{1}{2t_R}(A_m e^{-j\alpha_m} - B_m e^{-j\beta_m}), \tag{7}$$

where $m \geq 1$. In obtaining Eqs. (6) and (7), it is assumed that the modulation magnitudes $|A_m|$ and $|B_m|$ are small. We observe that Eqs. (4) and (5) have a translational symmetry along the frequency axis, which is desirable for convolution operation.

We further assume that the input wave consists of a sequence of equally spaced frequency components, where the frequency separation is Ω and the frequency detuning with respect to the resonant frequencies of the ring is $\Delta\omega$, i.e., $c_{\mathrm{in},n}(\tau) = c_{\mathrm{in},n}\exp(j\,\Delta\omega\tau)$. In this case, the steady-state amplitudes of the modes in the ring resonator and at the output port take the similar forms $a_n(\tau) = a_n\exp(j\,\Delta\omega\tau)$ and $c_{\mathrm{out},n}(\tau) = c_{\mathrm{out},n}\exp(j\,\Delta\omega\tau)$, respectively. As we consider the on-resonance coupling in the system throughout the paper, the frequency detuning $\Delta\omega = 0$. In the representation of discrete frequency modes, $\mathbf{a} = [\cdots, a_{-1}, a_0, a_1, \ldots]^T$, $\mathbf{c}_{\mathrm{in}} = [\cdots, c_{\mathrm{in},-1}, c_{\mathrm{in},0}, c_{\mathrm{in},1}, \ldots]^T$, and $\mathbf{c}_{\mathrm{out}} = [\cdots, c_{\mathrm{out},-1}, c_{\mathrm{out},0}, c_{\mathrm{out},1}, \ldots]^T$, we obtain the scattering matrix $\mathbf{S}$ [23], where $\mathbf{c}_{\mathrm{out}} = \mathbf{S}\mathbf{c}_{\mathrm{in}}$, from Eqs. (4) and (5):

$$\mathbf{S} = \mathbf{I} + j2\gamma_e\,[\mathbf{K} - j\,(\gamma + \gamma_0 + \gamma_e)\,\mathbf{I}]^{-1}, \tag{8}$$

where $\mathbf{K}$ is the matrix that contains the coupling coefficients induced by the modulation with the matrix element as given by $K_{m,n} = \kappa_{m-n}(n \neq m)$ and $\mathbf{I}$ is the identity matrix. Therefore, the matrix elements of $\mathbf{S}$ satisfy $S_{m,n} = s_{m-n}$ and thus have a translational symmetry along the frequency axis—as expected, since the system described by Eqs. (4) and (5) is translationally invariant along the frequency axis.

B. Convolution-kernel generation

Equation (8) describes a convolution operation, since

$$c_{\mathrm{out},m} = \sum_n S_{m,n}c_{\mathrm{in},n} = \sum_n s_{-n}c_{\mathrm{in},m+n} \tag{9}$$

with s_n being the convolution kernel [34]. From Eq. (9), we illustrate a simple example of one-dimensional (1D)

convolution in Fig. 1(b), where each frequency site of $\mathbf{c}_{\text{out}}$ is given by a corresponding frequency site of $\mathbf{c}_{\text{in}}$ with its local neighbors, averaged with the weights given by the kernel $\mathbf{s}$. The 1D convolution is widely used in a number of applications including natural-language processing [35] and time-series modeling [36].

Equations (8) and (9) allow us to determine the convolution kernel s_n from the modulation profile as described in Eqs. (2) and (3). For an infinite-dimensional matrix $\mathbf{A}$ having translational symmetry, i.e., $A_{m,n} = a_{m-n}$, its inverse $\mathbf{A}^{-1}$ also has translational symmetry, i.e., $(A^{-1})_{m,n} = \tilde{a}_{m-n}$, where $\tilde{a}_m$ is given by

$$\tilde{a}_m = \frac{1}{2\pi}\int_0^{2\pi} e^{-jm\theta}\left(\sum_n e^{jn\theta}a_n\right)^{-1} d\theta. \tag{10}$$

Applying Eq. (10) to Eq. (8), we obtain

$$s_m = \delta_{m,0} + \frac{j\gamma_e}{\pi}\int_0^{2\pi} e^{-jm\theta}\left(\sum_n e^{jn\theta}\kappa_n\right)^{-1} d\theta, \tag{11}$$

where we define $\kappa_0 = -j(\gamma + \gamma_0 + \gamma_e)$ and $\kappa_m (m \neq 0)$ is defined in Eqs. (6) and (7).

Equation (11) enables us to determine the convolution kernel from the modulation profile. On the other hand, in typical applications, the convolution kernel s_n is prescribed and the task is then to choose the modulation profile, as well as other parameters of the device, to achieve the desired kernel. For this purpose, we derive the corresponding modulation parameters from Eq. (11) as

$$\kappa_m = \frac{j\gamma_e}{\pi}\int_0^{2\pi} e^{-jm\theta}\left[\sum_n e^{jn\theta}\left(s_n - \delta_{n,0}\right)\right]^{-1} d\theta. \tag{12}$$

From Eq. (12), and using Eqs. (6) and (7), we find the parameters for the amplitude and phase modulations, as well as the decay rate of the resonator as

$$A_m e^{j\alpha_m} = t_R(\kappa_{+m} + \kappa_{-m}^*), \tag{13}$$

$$B_m e^{j\beta_m} = t_R(\kappa_{+m} - \kappa_{-m}^*), \tag{14}$$

$$\gamma + \gamma_0 + \gamma_e = j\kappa_0. \tag{15}$$

Equations (12)–(15) provide an analytic approach to finding the required modulation wave forms and decay rates for any desired kernel. The implementation of a kernel with N nonzero elements requires N modulation frequencies in both the amplitude and the phase modulation.

C. Passivity constraint

For convolution operations in digital signal processing, the norm of the kernel $\mathbf{s}$ in Eq. (9) does not play a significant role. In our physical implementation, however, the norm of the kernel is important, since one typically prefers to use a passive system without net energy gain. As a sufficient condition for a passive system, the time-dependent transmission

$$s(t) = \sum_m s_m e^{jm\Omega t} \tag{16}$$

is required to satisfy

$$|s(t)| \leq 1 \tag{17}$$

for every t. We note that Eq. (17) is not a necessary condition for a passive system—as, for example, noted in Ref. [37]. For a given prescribed kernel s'_m, we define

$$\eta = \max_{t\in(0,2\pi/\Omega]}\left|\sum_m s'_m \exp(jm\Omega t)\right|, \tag{18}$$

and implement the kernel with the scaling factor 1.1η:

$$s_m = s'_m/(1.1\eta). \tag{19}$$

Here, the factor of 1.1 is introduced so that the implemented system is slightly lossy.

D. Two-dimensional convolution

Equation (9) has the form of a 1D convolution. Here, we establish how we can perform higher-dimensional convolutions in such a discrete frequency system, by judiciously arranging input higher-dimensional matrices into a vector and accordingly converting the higher-dimensional kernel into a 1D kernel. We illustrate this by considering 2D convolution first.

A convolution between a 2D matrix $\mathbf{A}$ of size $H \times W$ and a kernel $\mathbf{F}$ of size $(2P_1 + 1) \times (2P_2 + 1)$ produces an output matrix $\mathbf{O}$ of size $(H - 2P_1) \times (W - 2P_2)$. In many applications, it is desirable that the output matrix has the same size as $\mathbf{A}$. For this purpose, it is common to pad the matrix $\mathbf{A}$ with zero-valued elements. The entire input matrix $\mathbf{X}$ with paddings that ensure same size of the output as $\mathbf{A}$ is therefore of size $L_1 \times L_2$ with $L_1 = H + 2P_1$ and $L_2 = W + 2P_2$. Here, the kernel sizes are chosen as odd numbers, as is typical in convolutional neural networks and image processing [34]. For the kernel matrix $\mathbf{F}$, we index the first and second dimensions as $[-P_1, \ldots, P_1]$ and $[-P_2, \ldots, P_2]$, respectively [38]. We index the first and second dimensions of $\mathbf{X}$ as $[-P_1, \ldots, H + P_1 - 1]$ and $[-P_2, \ldots, W + P_2 - 1]$, respectively. The matrix $\mathbf{A}$ occupies a block in $\mathbf{X}$ indexed from 0 to $H - 1$ for the first dimension and 0 to $W - 1$ for the second dimension. The rest of the matrix $\mathbf{X}$ is padded with zero-valued elements. The output data matrix $\mathbf{Y}$ by convolving $\mathbf{X}$ with $\mathbf{F}$ is of

size $H \times W$. Mathematically, this 2D convolution can be described as

$$Y_{i,j} = \sum_{p=-P_1}^{P_1} \sum_{q=-P_2}^{P_2} F_{-p,-q} X_{i+p,j+q}, \tag{20}$$

where $i \in [0, \ldots, H-1]$ and $j \in [0, \ldots, W-1]$. For illustration, Fig. 2(a) presents an example with an input $\mathbf{A}$ matrix of size $H = 2$, $W = 3$, padded with $P_1 = P_2 = 1$ zero-valued elements, convolving with $\mathbf{F}$ matrix of size 3×3 and generating an output matrix $\mathbf{Y}$ of size 2×3.

We now show that the 2D convolution as described by Eq. (20) can be achieved using the dynamically modulated ring resonator with the input-output relation described by Eq. (9). The input data $\mathbf{X}$ are flattened into the input vector $\mathbf{c}_{\text{in}}$ as given by

$$c_{\text{in},iL_2+j} = X_{i,j}, \tag{21}$$

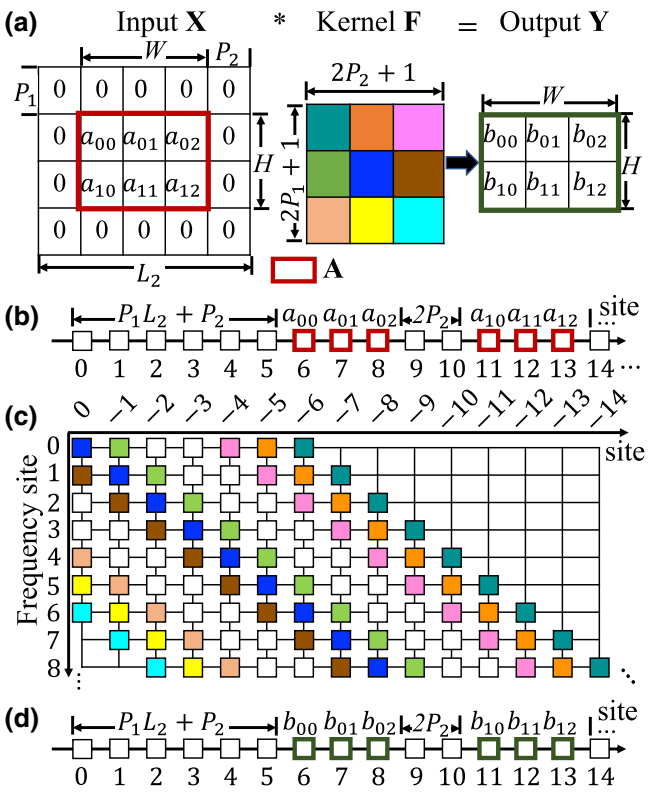

FIG. 2. (a) A schematic of convolution for a two-dimensional (2D) kernel $\mathbf{F}$ of size 3×3 with the zero-padded input matrix $\mathbf{X}$ of size 4×5 and the output matrix $\mathbf{Y}$ of size 2×3. (b) The input matrix $\mathbf{X}$ is vectorized into a 1D vector $\mathbf{c}_{\text{in}}$ as a frequency comb. (c) The scattering matrix generated by the modulated ring resonator maintains translational symmetry among frequency sites, which is equivalent to a convolution operation. (d) The output vector $\mathbf{c}_{\text{out}}$ after multiplication between the scattering matrix in (c) and the input vector in (b), which recovers the convolution output matrix $\mathbf{Y}$.

where we choose $i \in [0, \ldots, L_1 - 1]$ and $j \in [0, \ldots, L_2 - 1]$, as shown in Fig. 2(b). We also reshape the convolution kernel in Eq. (20) accordingly to the 1D kernel embedded in the scattering matrix element in Eq. (9), as

$$s_{-pL_2-q} = F_{-p,-q}, \quad p \in \{-P_1, \ldots, P_1\}, \quad q \in \{-P_2, \ldots, P_2\}, \tag{22}$$

$$= 0, \quad \text{otherwise}, \tag{23}$$

as shown in Fig. 2(c). The length of the converted 1D kernel is $(2P_1 + 1)L_2 + 2P_2 + 1$. Here, we note that the nonzero elements of s_m form blocks that are not contiguous, due to the flattening of the input image into a 1D array.

From Eq. (9), and using Eqs. (21)–(23), the convolution process in the modulated ring resonator can be described as

$$c_{\text{out},m} = \sum_{p=-P_1}^{P_1} \sum_{q=-P_2}^{P_2} s_{-pL_2-q} c_{\text{in},m+pL_2+q} \tag{24}$$

$$= \sum_{p=-P_1}^{P_1} \sum_{q=-P_2}^{P_2} F_{-p,-q} X_{i+p,j+q}. \tag{25}$$

By using the relation $m = iL_2 + j$, we can map back the elements of $c_{\text{out},m}$ obtained from Eq. (20) with the 2D output data as

$$Y_{i,j} = c_{\text{out},m}, \tag{26}$$

as illustrated in Fig. 2(d). In deriving Eq. (24), we keep only the nonzero components of s_{-n} in the summation of Eq. (9). Therefore, we show that the 2D convolution can be achieved with a single dynamically modulated ring resonator. This process can also be generalized for higher-dimensional convolution in Sec. V.

III. DEMONSTRATION OF TWO-DIMENSIONAL CONVOLUTIONS

In this section, we employ a few kernels to show how our proposed approach applies in 2D convolutions. The input data consist of a 2D image taken from the Modified National Institute of Standards and Technology (MNIST) database [39] and cropped with central 22×24 pixels as shown in Fig. 3(a). Together with padding $P_1 = P_2 = 1$, the size of the input matrix $\mathbf{X}$ is $L_1 = 24$ and $L_2 = 26$ in the first and second dimensions, respectively. Using Eq. (21), we represent this image with a 1D input vector in the frequency space.

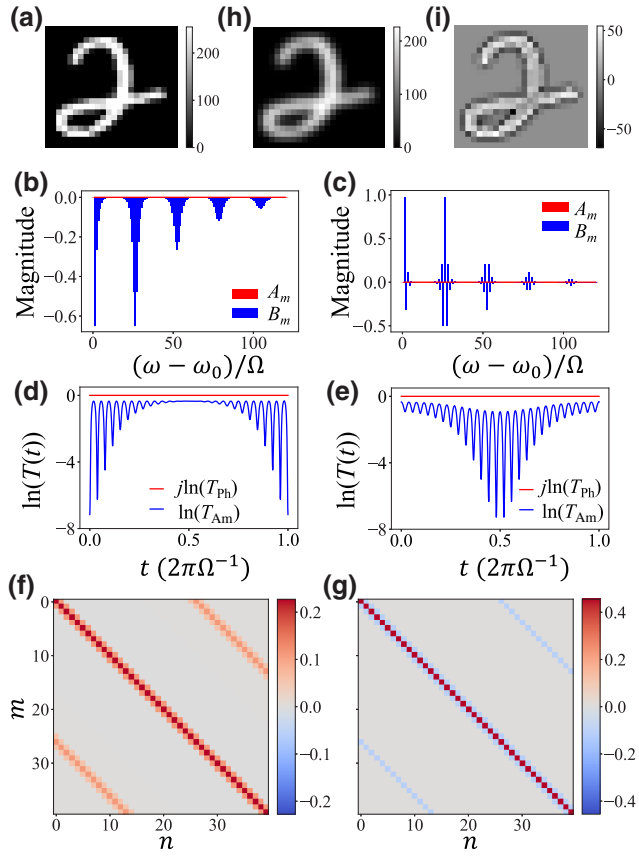

FIG. 3. A demonstration of image convolution for a symmetric Gaussian blurring kernel **G** and a Laplacian kernel **L**. (a) The original 2D image data from MNIST representing the digit 2 with a size of 22×24. (b),(c) The modulation magnitude for the phase and amplitude modulation for **G** and **L**, respectively. (d),(e) The time-dependent transmission factors for **G** and **L**, respectively. $j\ln(T_{\rm Ph})$ and $\ln(T_{\rm Am})$, the respective logarithm of transmission factors due to the phase and amplitude modulation, are plotted. (f),(g) Generated scattering matrices with elements $S_{m,n}$, which are unitless, that correspond to **G** and **L**. (h),(i) The convolution output images from **G** and **L** kernels generated from the modulated ring system, where the image is blurred and highlighted with edges, respectively. In (a), (h), and (i), the color map represents unitless pixel values.

As the first set of examples, we consider two kernels: a Gaussian kernel **G** and a Laplacian kernel **L**:

$$\mathbf{G} = \frac{1}{\eta(\mathbf{G})}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \quad \mathbf{L} = \frac{1}{\eta(\mathbf{L})}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}. \tag{27}$$

G and **L** are widely used in digital image processing, for image blurring [40] and edge detection [41], respectively.

We follow the procedure as outlined in the previous section to implement these kernels in synthetic frequency space. For each kernel in Eq. (27), we construct the corresponding 1D kernels **s** using Eqs. (22) and (23). We then use Eqs. (12)–(15) to determine the appropriate modulation parameters and cavity decay rates. Since both **G** and **L** are real-valued symmetric matrices, the corresponding 1D kernels satisfy $s_k = s_{-k}$, as can be seen from Eq. (22). From Eq. (12), this implies that $\kappa_m + \kappa_{-m}^* = 0$. With Eq. (13), we can see that the phase modulation has zero magnitude, i.e., $A_m = 0$ for all positive integers m. Therefore, only amplitude modulation is required to implement such symmetric kernels.

Based on the discussion in Sec. II C on the passivity constraint, we determine the scaling factors $\eta(\mathbf{G}) = 17.6$ and $\eta(\mathbf{L}) = 8.78$ in Eq. (27). Under this scaling factor, we obtain $\gamma + \gamma_0 = 0.1290\ \Omega$ for **G** and $\gamma + \gamma_0 = 0.2125\ \Omega$ for **L**, respectively. For $m > 0$, we obtain the magnitude of the phase and amplitude modulation for the mth-order modulation, i.e., A_m and B_m in Eqs. (2) and (3), for **G** and **L**, as shown in Figs. 3(b) and 3(c), respectively, where we confirm that $A_m = 0$ as expected above from the symmetry argument.

From the A_m and B_m as determined above, the time-dependent transmission factors $T(t)$ of the modulator, as determined by Eqs. (1)–(3), are presented in Figs. 3(d) and 3(e) for **G** and **L**, respectively, over a period of time from 0 to $2\pi/\Omega$. In this plot, and in similar plots in the rest of the paper, we assume that $\gamma_0 = 0$ and plot $j\ln(T_{\rm Ph})$ and $\ln(T_{\rm Am})$ using Eqs. (2) and (3). Under this temporal modulation, the frequency-domain scattering matrices for **G** and **L** are shown in Figs. 3(f) and 3(g), respectively. The scattering matrix is sparse. Within each row, the nonzero matrix elements are separated by zero-valued gaps, the size of which is given by the difference between the size of the kernel and input data. By multiplying this scattering matrix with the 1D input vector as generated from the image in Fig. 3(a), we obtain the output images. Figure 3(h) shows the output image for the Gaussian blurring kernel. We see that the output image is smoothened as compared with the input image. Figure 3(i) shows the output image for the Laplacian kernel. Here, the edges of the handwritten digit are highlighted in the output images.

As a second set of examples, in Fig. 4 we consider the Sobel x kernel $\mathbf{I}_x$ and the Sobel y kernel $\mathbf{I}_y$:

$$\mathbf{I}_x = \frac{1}{\eta(\mathbf{I}_x)}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{I}_y = \frac{1}{\eta(\mathbf{I}_y)}\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \tag{28}$$

$\mathbf{I}_x$ and $\mathbf{I}_y$ are commonly used for edge detection along the horizontal and vertical directions, respectively [42].

We follow the same procedure as outlined above to implement these kernels in synthetic frequency space. $\mathbf{I}_x$ and $\mathbf{I}_y$ are not symmetric matrices. To implement these matrices, both phase and amplitude modulations are required. Based on the discussion in Sec. II C on the passivity constraint, we determine the scaling factors

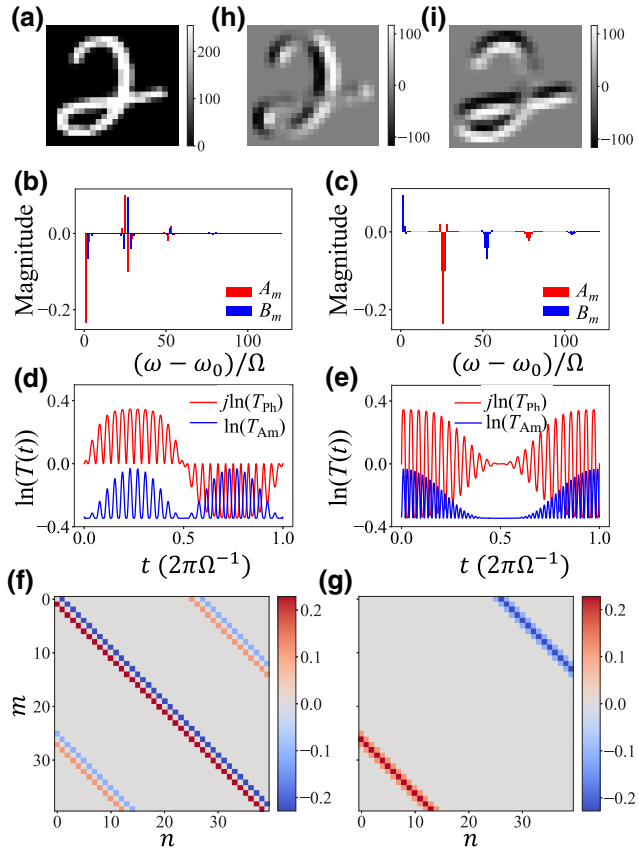

FIG. 4. A demonstration of image convolution for an asymmetric Sobel x kernel $\mathbf{I}_x$ and the Sobel y kernel $\mathbf{I}_y$. (a) The original 2D image data from MNIST [the same as in Fig. 3(a)]. (b),(c) The magnitudes for the phase and amplitude modulation of $\mathbf{I}_x$ and $\mathbf{I}_y$, respectively. (d),(e) The time-dependent transmission factors for $\mathbf{I}_x$ and $\mathbf{I}_y$, respectively. $j\ \ln(T_{\rm Ph})$ and $\ln(T_{\rm Am})$, the respective logarithms of transmission factors due to the phase and amplitude modulation, are plotted. (f),(g) Generated scattering matrices with elements $S_{m,n}$, which are unitless, that correspond to $\mathbf{I}_x$ and $\mathbf{I}_y$. (h),(i) The convolution output images from $\mathbf{I}_x$ and $\mathbf{I}_y$ kernels generated from the modulated ring system, where the image is highlighted with the horizontal and vertical edges. In (a), (h), and (i), the color map represents unitless pixel values.

$\eta(\mathbf{I}_x) = 8.7355$ and $\eta(\mathbf{I}_y) = 8.7730$ in Eq. (28). Under this scaling factor, we obtain $\gamma + \gamma_0 = 0.04294\ \Omega$ for $\mathbf{I}_x$ and $\gamma + \gamma_0 = 0.04301\ \Omega$ for $\mathbf{I}_y$, respectively. For $m > 0$, we obtain the magnitude of the phase and amplitude modulation for the mth-order modulation, i.e., A_m and B_m in Eqs. (2) and (3), for $\mathbf{I}_x$ and $\mathbf{I}_y$ as in Figs. 4(b) and 4(c), respectively. In contrast to the symmetric case, we confirm that the magnitude of the phase modulation is generally nonzero, as expected above from the symmetry argument.

From the A_m and B_m as determined above, the time-dependent transmission factors $T(t)$ of the modulator, as determined using Eq. (1)–(3), are presented in Figs. 4(d) and 4(e), for $\mathbf{I}_x$ and $\mathbf{I}_y$, respectively, over a period of time from 0 to $2\pi/\Omega$. Under this temporal modulation, the frequency-domain scattering matrices for $\mathbf{I}_x$ and $\mathbf{I}_y$ are shown in Figs. 4(f) and 4(g), respectively, and we observe that these matrices are sparse, similar to those shown in Figs. 3(f) and 3(g). By multiplying this scattering matrix with the 1D input vector as generated from the image in Fig. 4(a), we obtain the output images. Figure 4(h) shows the output image for the Sobel x kernel. We see that the horizontal edges of the handwritten digit are highlighted in the output image. Figure 4(i) shows the output image for the Sobel y kernel. Here, the vertical edges of the handwritten digit are highlighted in the output image.

We now proceed to analyze the maximum modulation frequency required to generate a target convolution kernel. From Eq. (22), in the kernel **s**, among all frequency sites that have nonzero amplitudes, the maximum index of the sites corresponds to a frequency shift of

$$\Omega_m = (P_1 L_2 + P_2)\Omega. \tag{29}$$

To generate such a kernel, the required maximum modulation frequency is typically a few times Ω_m. As illustrations, for the examples considered in this section, $\Omega_m = 27\ \Omega$. As we can see in Figs. 3(b) and 3(c) as well as in Figs. 4(b) and 4(c), the computed modulation magnitude becomes quite small when the order of modulation m exceeds 100. Thus, typically, the required maximum modulation frequency is about 3 to 4 times Ω_m.

To conclude this section, we realize 2D convolution using one modulated ring resonator. Our approach should be applicable to all convolution kernels used in digital image processing.

IV. LARGE-SIZE IMAGE CONVOLUTION

In this section, we discuss issues associated with the limited modulation bandwidth Ω_b (i.e., the maximum modulation frequency) of the modulator. Again, we consider an image described by a matrix of the size $H \times W$ convolving with a kernel of the size $(2P_1 + 1) \times (2P_2 + 1)$. In our original approach as described in the previous section, we generate a padded matrix $\mathbf{X}$ of the size $L_1 \times L_2$, where $L_1 = H + 2P_1$, and $L_2 = W + 2P_2$. Based on the analysis in Sec. III, the required maximum modulation frequency is approximately proportional to Ω_m as given by Eq. (29). Therefore, the required maximum modulation frequency scales linearly with one of the dimensions L_2 of the input images. Such a scaling is undesirable for large images when the modulation bandwidth is limited.

Here, we provide an approach to reducing the required modulation bandwidth by judiciously slicing the image. We illustrate the working principle in Fig. 5(a). We slice the image into several nonoverlapping subimages of the size $H \times W'$ with $W' < W$. For each subimage, we choose a submatrix of $\mathbf{X}$ with the size $L_1 \times L$, where

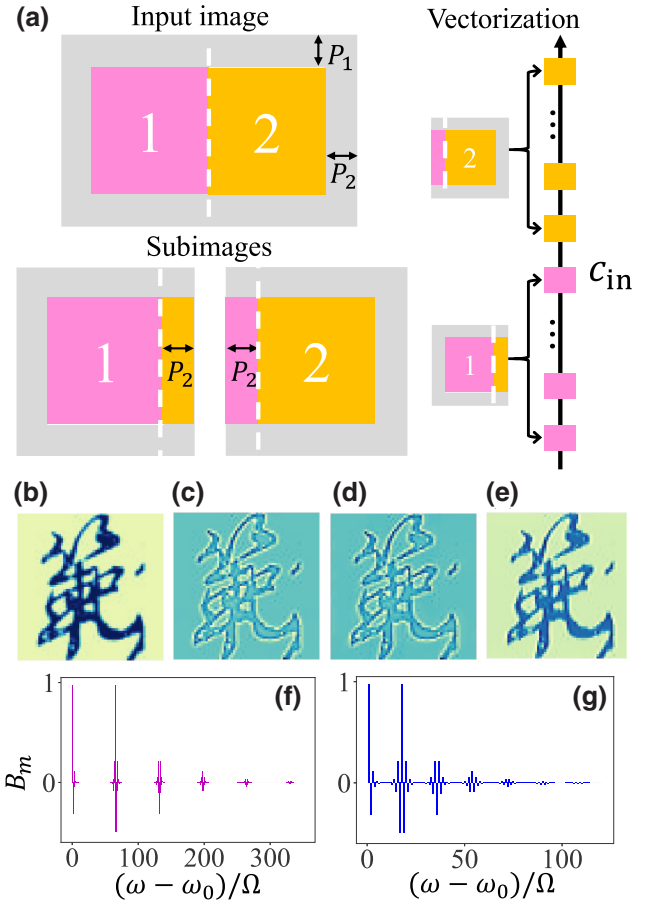

FIG. 5. (a) A schematic for large-scale 2D convolution, where the input is sliced using the bandwidth-saving technique to efficiently utilize the modulator strength. (b) An example input image with a size of 64 × 64 pixels. (c) Output images from the original input with cutoff modulation orders at 500 Ω. (d) An output image from the sliced input as generated with cutoff modulation orders at 50 Ω. (e) An output image from the original input as generated by the modulation $B_{m,o}$ with cutoff modulation orders at 50 Ω. (b)–(e) The color maps represent unitless pixel values. (f) The magnitudes of the amplitude modulation B_m required to generate the kernel using the original input. (g) The magnitudes of the amplitude modulation B_m required to generate the kernel using the sliced inputs.

$L = W' + 2P_2$, such that the subimage is located at the center of the submatrix and the padded region contains sufficient information so that the convolution operation on the subimage can be carried out [Fig. 5(a)]. The convolution of such a submatrix with the kernel can then proceed in the same way as we have described in the previous section, with the frequency shift that corresponds to the maximum index in the kernel reduced to

$$\Omega'_m = (P_1 L + P_2)\,\Omega. \tag{30}$$

Since $L < L_2$, $\Omega'_m < \Omega_m$, and consequently the required maximum modulation frequency is also reduced. We also note that the convolution of multiple subimages can be performed in parallel. For this purpose, we form a 1D array consisting of a concatenation of all the flattened submatrices as described above and proceed with the same convolution operation, as shown in Fig. 5(a).

In this following, we provide an illustration. The input data consist of an image chosen from the Kuzushiji-Kanji data set [43]. It is of size $H = W = 64$, as shown in Fig. 5(b). The convolution kernel is chosen as the Laplacian kernel $\mathbf{L}$ in Eq. (27) of size 3×3, so we have $P_1 = P_2 = 1$ and $L_1 = L_2 = 66$. For comparison, we represent this image with a 1D input vector in the frequency space via either the original approach as discussed in Sec. II D, with $\Omega_m = 67\,\Omega$ given by Eq. (29), or the slicing approach, where the image is sliced into four subimages with $W' = 16$ corresponding to $\Omega'_m = 19\,\Omega$ as determined using Eq. (30). For both approaches, using the method as discussed in Sec. II B, we obtain the magnitude B_m of the amplitude modulation, as shown in Fig. 5(f) for the original approach and in Fig. 5(g) for the slicing approach. For the slicing approach, B_m decreases more rapidly as m increases, as compared with the original approach. For both cases, we also confirm that $A_m = 0$, which is consistent with the previous observation.

We now show that the slicing approach can produce the desired output but with lower requirements on the modulation bandwidth. Figure 5(c) shows the output image with the original approach, with a maximum modulation frequency of 500 Ω. Figure 5(d) shows the output image with the slicing approach, with a maximum modulation frequency of 50 Ω. We see that the output images in Figs. 5(c) and 5(d) are very similar to each other, as both highlight the edges of the input image. In contrast, in Fig. 5(e), we show the output image with the original approach but with a maximum modulation frequency of 50 Ω. The output image resembles the original image and no longer highlights the edges. Our results indicate that the slicing approach can indeed significantly reduce the requirement on the modulation bandwidth as compared with the original approach.

V. HIGHER-DIMENSIONAL CONVOLUTION

In the above discussions, we consider 2D convolution, which is useful for extracting spatial features within a single 2D image. However, many applications produce higher-dimensional input data sets. For example, LIDAR scans produce an array of images at various spatial depths [44–46] and a video consists of an array of images at different temporal frames [47,48]. For processing these data sets, higher-dimensional convolution is important. For processing of LIDAR data sets, 3D convolution is useful in identifying 3D objects [49,50]. For video processing, 3D convolution is useful in recognizing and predicting

motion [35,51]. Thus there have been emerging interests in creating specialized hardware for such computations.

Higher-dimensional convolutions are more computationally demanding as compared with 2D convolutions [51]. Here, we show that higher-dimensional convolutions can be accomplished using the same modulated ring cavity as we have discussed above. Our approach for higher-dimensional convolution closely follows that of the 2D case. Here, as an illustration, we consider the 3D case. The input data are represented by a matrix of size $H \times W \times D$. The kernel matrix $\mathbf{F}$ is of the size $(2P_1+1)\times(2P_2+1)\times(2P_3+1)$. We again generate an input matrix $\mathbf{X}$ by padding the input data along three dimensions so that the convolution output has the same dimension as the input. The resulting input matrix $\mathbf{X}$ has dimensions of $L_1 \times L_2 \times L_3$, where $L_1 = H + 2P_1$, $L_2 = W + 2P_2$, and $L_3 = D + 2P_3$. For $i = [0,1,\ldots,H-1]$, $j = [0,1,\ldots,W-1]$, and $k = [0,1,\ldots,D-1]$, the 3D convolution can be described by

$$Y_{i,j,k} = \sum_{p=-P_1}^{P_1}\sum_{q=-P_2}^{P_2}\sum_{t=-P_3}^{P_3} F_{-p,-q,-t}X_{i+p,j+q,k+t}. \tag{31}$$

To implement such a 3D convolution in the synthetic dimension, we form a 1D vector as

$$c_{\mathrm{in},iL_3L_2+jL_3+k} = X_{i,j,k}. \tag{32}$$

We also map the convolution kernel in Eq. (31) to the scattering matrix element in Eq. (9) as

$$s_{-pL_3L_2-qL_3-t} = F_{-p,-q,-t},$$
$$\{p,q,t\} \in \{[-P_{1,2,3},\ldots,P_{1,2,3}]\}, \tag{33}$$
$$= 0, \quad \text{otherwise}. \tag{34}$$

In this way, we can achieve the 3D convolution output as

$$c_{\mathrm{out},m} = \sum_{p=-P_1}^{P_1}\sum_{q=-P_2}^{P_2}\sum_{t=-P_3}^{P_3} s_{-pL_3L_2-qL_3-t}c_{\mathrm{in},m+pL_3L_2+qL_3+t} \tag{35}$$

$$= \sum_{p=-P_1}^{P_1}\sum_{q=-P_2}^{P_2}\sum_{t=-P_3}^{P_3} F_{-p,-q,-t}X_{i+p,j+q,k+t} \tag{36}$$

By choosing $m = iL_3L_2 + jL_3 + k$, we recover $Y_{i,j,k} = c_{\mathrm{out},m}$. The required modulation amplitudes and cavity decay rates can be determined from the kernel $\mathbf{s}$ in the same way as discussed in Sec. II B.

As illustrated in Fig. 6(a), we present an example with the input as an array of image frames from a human-motion-recognition database [52] that describe a person waving both arms upward. The total input is cropped to size $50 \times 40 \times 5$ [Fig. 6(d)] and consists of five images of the size 50×40 at five different times. We implement a convolution kernel that corresponds to an operator $\partial_t^2(\partial_x^2 + \partial_y^2)$, where t and x, y correspond to the time and the two spatial dimensions in the input, respectively. This operator is chosen to highlight the motion of the edge of an object. Using a finite-difference approximation, this operator is implemented as a $3 \times 3 \times 3$ kernel matrix $\tilde{\mathbf{L}}$, with three temporal planes denoted as $\tilde{\mathbf{L}}_{1,2,3}$, which are given by

$$\tilde{\mathbf{L}}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \tilde{\mathbf{L}}_2 = \begin{bmatrix} 0 & -2 & 0 \\ -2 & 8 & -2 \\ 0 & -2 & 0 \end{bmatrix}, \tilde{\mathbf{L}}_3 = \tilde{\mathbf{L}}_1. \tag{37}$$

For this kernel ($\tilde{\mathbf{L}}$), the scaling factor is chosen as $\eta(\tilde{\mathbf{L}}) = 1.2\eta = 37.736$, where we use a slightly larger scaling factor compared with Eq. (18) due to the high-bandwidth

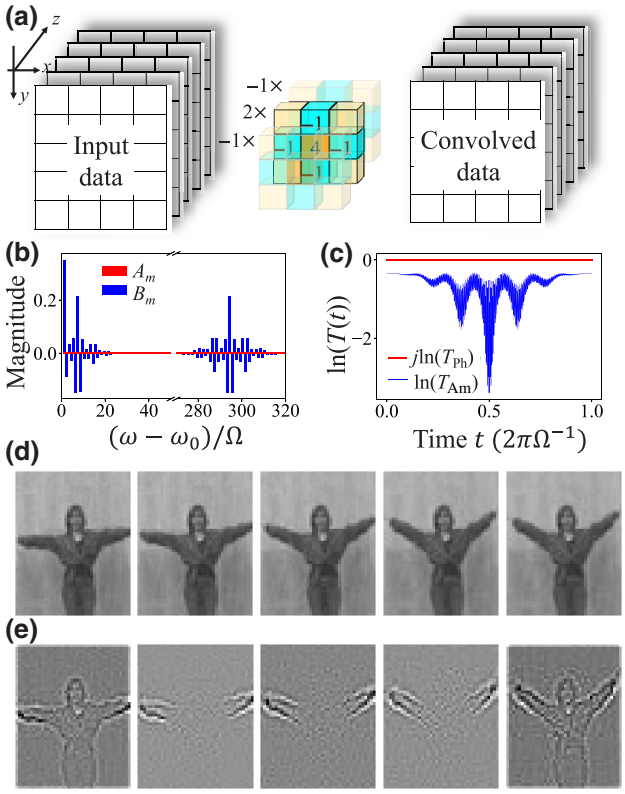

FIG. 6. (a) A schematic for multidimensional convolution, where the input has multiple channels. (b) The modulation magnitude to generate the 3D Laplacian kernel. (c) The time-dependent transmission factors, $j\ \ln(T_{\mathrm{Ph}})$ and $\ln(T_{\mathrm{Am}})$, due to the phase and amplitude modulation. (d) Input data consisting of an array of images at different temporal frames to represent a person waving his or her arms. (e) The output convolution image, consisting of an array of images at different temporal frames that highlight the arm motion. In (d) and (e), the color maps represent unitless pixel values.

modulation for the 3D convolution. The corresponding $\gamma + \gamma_0 = 0.1\Omega$. The modulation profile for generating the convolution kernel given by Eq. (22) is shown in Fig. 6(b). In general, 3D convolution requires a higher modulation bandwidth as compared with 2D convolution. Here, for simplicity, we use the original approach in Sec. II D but the modulation bandwidth can be reduced with the slicing approach as discussed in Sec. IV.

For the modulation thus determined, the time-dependent transmission factors for the amplitude and phase modulations are shown in Fig. 6(c). We note that the phase modulation is constantly zero, similar to the 2D case. Compared with the 2D case as shown in Figs. 4(d) and 4(e), the difference between the largest and smallest modulation frequencies is significantly larger. We show in Fig. 6(d) five input-video clipped images arranged in temporal order. The input images consist of a person waving his or her arms upward, whereas the other parts of the body remain still. Using the convolution kernel shown in Eq. (37), the 3D convolution is expected to detect the motion of arms and highlight the edges. As shown in Fig. 6(e), the output images from convolution with the ring resonator are arranged in the same temporal order as the input. Similar to the 2D convolution case, the first and last frames of the output images highlight the outlines of the person as expected. However, as shown from the second to the fourth frame of the output, the 3D convolution provides additional information that is useful for recognizing human motion. We observe in the central three frames that only the arms are highlighted, whereas the other parts of the person have negligible signals, which indicates that the person is moving his or her arms in the video.

To summarize this section, we show that 3D convolution can be realized in a single dynamically modulated optical ring resonator. The higher-dimensional convolution introduced here has broad applications, such as 3D convolution for edge feature extraction and scene reconstruction [53], four-dimensional (4D) convolution for spatiotemporal detection [5], as well as six-dimensional (6D) convolution for noise-robust geometric pattern recognition [54].

VI. CONCLUSIONS

We describe a scheme for realizing arbitrary convolution kernels in synthetic frequency space using a simple setup with one ring resonator incorporating one phase and one amplitude modulator. This scheme can be used to perform multidimensional convolutions. We provide an analytic approach that determines the required modulation profile for any convolution kernel. In our scheme, the dimension of the input data set that can be processed is limited by the number of equally spaced frequency modes available in the ring, as well as by the loss of the ring. The number of such equally spaced modes is controlled by the group-velocity dispersion of the waveguide forming the ring and the loss may be compensated with the use of an amplifier. Experimentally, nearly one thousand equally spaced frequency modes have been observed in on-chip systems [55]. The group-velocity dispersion in this lithium niobate (LN) system is estimated to be $\beta_2 = -50\ \mathrm{ps}^2/\mathrm{km}$ [56] and $n \approx 2$. Assuming the FSR as 1 GHz, the circumference of the ring is related by $\ell = c/n/\mathrm{FSR} = 0.15$ m. The shift in the FSR is then given by $\Delta\mathrm{FSR} = -2\pi\ell(\mathrm{FSR})^3\beta_2 = 47.1$ Hz. Hence, even in the presence of group-velocity dispersion, nearly one thousand frequency modes are equally spaced within the line width of the resonant modes of the ring. A larger number of modes may be achievable in the fiber-ring system used in Refs. [27,57].

This convolution processing in the synthetic dimension can be implemented for both fiber-loop [27,57] and on-chip [26,55,58] platforms, where a sufficiently fast modulation speed compared to the FSR has been demonstrated. Future advances in the fabrication of high-speed and high-confinement modulators, as well as high-speed photodetectors, may reduce the required energy consumption.

The results demonstrated here can also be extended to complex-valued convolutional neural networks, which have been successfully applied in computer vision, especially in processing magnetic-resonance-imaging data, which are complex in their raw form [59], with the advantages of avoiding overfitting and of robustness to noise. Other than frequency, we anticipate that this convolution scheme can be applied to other internal degrees of freedom of the photon, such as spin, linear momentum, and optical angular momentum [60]. Similar ideas may also be implemented in Rydberg-atom systems [61]. For the realization of convolution, there must be no boundary in the frequency range where the convolution is performed. This requirement is different from the requirement in performing arbitrary finite-dimensional linear transformation as discussed in Ref. [23], where a boundary is required and can be achieved with the use of auxiliary rings [62]. Such a boundary in frequency dimension has been recently demonstrated in Ref. [63]. Our approach to convolution processing points to a direction for removing the computing bottleneck in traditional electronic circuits and may be useful in improving machine-learning hardware for artificial-intelligence applications.

ACKNOWLEDGMENTS

L.F. acknowledges discussions with Dr. Luqi Yuan, Dr. Guangwei Hu, Dr. S. Han, and Alan Luo. This work is supported by Multidisciplinary University Research Initiative (MURI) grants from the U.S. Air Force Office of Scientific Research (Grants No. FA9550-17-1-0002 and No. FA9550-18-1-0379).

[1] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Advances in Neural Information Processing Systems*, Vol. 25, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012).

[3] J. Zhou and O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning–based sequence model, Nat. Methods **12**, 931 (2015).

[4] J. Hou, B. Adhikari, and J. Cheng, DeepSF: Deep convolutional neural network for mapping protein sequences to folds, Bioinformatics **34**, 1295 (2017).

[5] C. Choy, J. Gwak, and S. Savarese, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), p. 3075.

[6] A. Boroumand, S. Ghose, Y. Kim, R. Ausavarungnirun, E. Shiu, R. Thakur, D. Kim, A. Kuusela, A. Knies, P. Ranganathan, and O. Mutlu, Google workloads for consumer devices: Mitigating data movement bottlenecks, SIGPLAN Not. **53**, 316 (2018).

[7] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, and D. Psaltis, Inference in artificial intelligence with deep optics and photonics, Nature **588**, 39 (2020).

[8] V. Bangari, B. A. Marquez, H. Miller, A. N. Tait, M. A. Nahmias, T. F. de Lima, H.-T. Peng, P. R. Prucnal, and B. J. Shastri, Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs), IEEE J. Sel. Top. Quantum Electron. **26**, 1 (2020).

[9] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, Photonics for artificial intelligence and neuromorphic computing, Nat. Photonics **15**, 102 (2021).

[10] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, Deep learning with coherent nanophotonic circuits, Nat. Photonics **11**, 441 (2017).

[11] T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, Training of photonic neural networks through *in situ* backpropagation and gradient measurement, Optica **5**, 864 (2018).

[12] I. A. D. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, Reprogrammable electro-optic nonlinear activation functions for optical neural networks, IEEE J. Sel. Top. Quantum Electron. **26**, 1 (2020).

[13] A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, Neuromorphic photonic networks using silicon photonic weight banks, Sci. Rep. **7**, 7430 (2017).

[14] A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, H. B. Miller, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, Silicon Photonic Modulator Neuron, Phys. Rev. Appl. **11**, 064043 (2019).

[15] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, All-optical machine learning using diffractive deep neural networks, Science **361**, 1004 (2018).

[16] Y. Luo, Y. Zhao, J. Li, E. Çetintaş, Y. Rivenson, M. Jarrahi, and A. Ozcan, Computational imaging without a computer: Seeing through random diffusers at the speed of light, eLight **2**, 4 (2022).

[17] T. W. Hughes, I. A. D. Williamson, M. Minkov, and S. Fan, Wave physics as an analog recurrent neural network, Sci. Adv. **5**, eaay6946 (2019).

[18] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities, Nature **569**, 208 (2019).

[19] X. Xu, M. Tan, B. Corcoran, J. Wu, T. G. Nguyen, A. Boes, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, D. G. Hicks, and D. J. Moss, Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks, Laser Photonics Rev. **14**, 2000070 (2020).

[20] X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, 11 TOPS photonic convolutional accelerator for optical neural networks, Nature **589**, 44 (2021).

[21] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, Parallel convolutional processing using an integrated photonic tensor core, Nature **589**, 52 (2021).

[22] R. Hamerly, A. Sludds, L. Bernstein, M. Prabhu, C. Roques-Carmes, J. Carolan, Y. Yamamoto, M. Soljačić, and D. Englund, in *2019 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2019), p. 22.8.1.

[23] S. Buddhiraju, A. Dutt, M. Minkov, I. A. D. Williamson, and S. Fan, Arbitrary linear transformations for photons in the frequency synthetic dimension, Nat. Commun. **12**, 2401 (2021).

[24] A. Dutt, M. Minkov, Q. Lin, L. Yuan, D. A. B. Miller, and S. Fan, Experimental band structure spectroscopy along a synthetic dimension, Nat. Commun. **10**, 3122 (2019).

[25] A. Dutt, Q. Lin, L. Yuan, M. Minkov, M. Xiao, and S. Fan, A single photonic cavity with two independent physical synthetic dimensions, Science **367**, 59 (2020).

[26] Y. Hu, C. Reimer, A. Shams-Ansari, M. Zhang, and M. Loncar, Realization of high-dimensional frequency crystals in electro-optic microcombs, Optica **7**, 1189 (2020).

[27] K. Wang, A. Dutt, K. Y. Yang, C. C. Wojcik, J. Vučković, and S. Fan, Generating arbitrary topological windings of a non-Hermitian band, Science **371**, 1240 (2021).

[28] H. Zhao, B. Li, H. Li, and M. Li, Scaling optical computing in synthetic frequency dimension using integrated cavity acousto-optics, arXiv:2106.08494 [physics.optics] (2021).

[29] H. Mahmudlu, R. Johanning, A. K. Kashi, A. van Rees, J. P. Epping, R. Haldar, K.-J. Boller, and M. Kues, Fully on-chip photonic turnkey quantum source for entangled qubit/qudit state generation, arXiv:2206.08715 [physics.optics] (2022).

[30] L. Yuan, M. Xiao, Q. Lin, and S. Fan, Synthetic space with arbitrary dimensions in a few rings undergoing dynamic modulation, Phys. Rev. B **97**, 104105 (2018).

[31] L. Yuan, Q. Lin, M. Xiao, and S. Fan, Synthetic dimension in photonics, Optica **5**, 1396 (2018).

[32] L. Yuan, A. Dutt, and S. Fan, Synthetic frequency dimensions in dynamically modulated ring resonators, APL Photonics **6**, 071102 (2021).

[33] K. Wang, B. A. Bell, A. S. Solntsev, D. N. Neshev, B. J. Eggleton, and A. A. Sukhorukov, Multidimensional synthetic chiral-tube lattices via nonlinear frequency conversion, Light: Sci. Appl. **9**, 132 (2020).

[34] B. Jähne, in *Digital Image Processing* (Springer, Berlin, 2002), p. 99.

[35] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, in *2015 IEEE International Conference on Computer Vision (ICCV)* (2015), p. 4489.

[36] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, 1D convolutional neural networks and applications: A survey, Mech. Syst. Signal Process. **151**, 107398 (2021).

[37] S. Sandhu and S. Fan, Lossless intensity modulation in integrated photonics, Opt. Express **20**, 4280 (2012).

[38] Superposition and convolution, in *Digital Image Processing* (John Wiley & Sons, Ltd, 2007) Chap. 7, p. 165.

[39] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE **86**, 2278 (1998).

[40] E. S. Gedraite and M. Hadad, in *Proceedings ELMAR-2011* (2011), p. 393.

[41] P. G. van Dokkum, Cosmic-ray rejection by Laplacian edge detection, Publ. Astron. Soc. Pac. **113**, 1420 (2001).

[42] V. Sanduja and R. Patial, Sobel edge detection using parallel architecture based on FPGA, Int. J. Appl. Inf. Syst. **3**, 20 (2012).

[43] T. Clanuwat, M. Bober-Irizar, A. Kitamoto, A. Lamb, K. Yamamoto, and D. Ha, Deep learning for classical Japanese literature, arXiv:1812.01718 (2018).

[44] M. J. Behrenfeld, Y. Hu, R. T. O'Malley, E. S. Boss, C. A. Hostetler, D. A. Siegel, J. L. Sarmiento, J. Schulien, J. W. Hair, X. Lu, S. Rodier, and A. J. Scarino, Annual boom-bust cycles of polar phytoplankton biomass revealed by space-based LIDAR, Nat. Geosci. **10**, 118 (2017).

[45] A. Hooijer and R. Vernimmen, Global LIDAR land elevation data reveal greatest sea-level rise vulnerability in the tropics, Nat. Commun. **12**, 3592 (2021).

[46] X. Zhang, K. Kwon, J. Henriksson, J. Luo, and M. C. Wu, A large-scale microelectromechanical-systems-based silicon photonics LIDAR, Nature **603**, 253 (2022).

[47] J. Liang, C. Ma, L. Zhu, Y. Chen, L. Gao, and L. V. Wang, Single-shot real-time video recording of a photonic mach cone induced by a scattered light pulse, Sci. Adv. **3**, e1601814 (2017).

[48] D. Schofield, A. Nagrani, A. Zisserman, M. Hayashi, T. Matsuzawa, D. Biro, and S. Carvalho, Chimpanzee face recognition from videos in the wild using deep learning, Sci. Adv. **5**, eaaw0736 (2019).

[49] Z. Liu, H. Tang, Y. Lin, and S. Han, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, USA, 2019), p. 965.

[50] X. Zhu, H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, H. Li, and D. Lin, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2021), p. 9939.

[51] J. Lin, C. Gan, and S. Han, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), p. 7083.

[52] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, in *International Conference on Computer Vision* (2011), p. 2556.

[53] M. Simonovsky and N. Komodakis, in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2017), p. 29.

[54] C. Choy, J. Lee, R. Ranftl, J. Park, and V. Koltun, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2020), p. 11227.

[55] M. Zhang, B. Buscaino, C. Wang, A. Shams-Ansari, C. Reimer, R. Zhu, J. M. Kahn, and M. Lončar, Broadband electro-optic frequency comb generation in a lithium niobate microring resonator, Nature **568**, 373 (2019).

[56] C. Wang, M. Zhang, M. Yu, R. Zhu, H. Hu, and M. Loncar, Monolithic lithium niobate photonic circuits for Kerr frequency comb generation and modulation, Nat. Commun. **10**, 978 (2019).

[57] K. Wang, A. Dutt, C. C. Wojcik, and S. Fan, Topological complex-energy braiding of non-Hermitian bands, Nature **598**, 59 (2021).

[58] Y. Hu, M. Yu, B. Buscaino, N. Sinclair, D. Zhu, R. Cheng, A. Shams-Ansari, L. Shao, M. Zhang, J. M. Kahn, and M. Loncar, High-efficiency and broadband electro-optic frequency combs enabled by coupled micro-resonators, arXiv.2111.14743 (2021).

[59] H. Gudbjartsson and S. Patz, The Rician distribution of noisy MRI data, Magn. Reson. Med. **34**, 910 (1995).

[60] M. Yang, H.-Q. Zhang, Y.-W. Liao, Z.-H. Liu, Z.-W. Zhou, X.-X. Zhou, J.-S. Xu, Y.-J. Han, C.-F. Li, and G.-C. Guo, Topological band structure via twisted photons in a degenerate cavity, Nat. Commun. **13**, 2040 (2022).

[61] S. K. Kanungo, J. D. Whalen, Y. Lu, M. Yuan, S. Dasgupta, F. B. Dunning, K. R. A. Hazzard, and T. C. Killian, Realizing topological edge states with Rydberg-atom synthetic dimensions, Nat. Commun. **13**, 972 (2022).

[62] R. Haldar, A. D. Banik, and S. K. Varshney, Design of CMOS compatible and compact, thermally-compensated electro-optic modulator based on off-axis microring resonator for dense wavelength division multiplexing applications, Opt. Express **22**, 22411 (2014).

[63] A. Dutt, L. Yuan, K. Y. Yang, K. Wang, S. Buddhiraju, J. Vučković, and S. Fan, Creating boundaries along a synthetic frequency dimension, Nat. Commun. **13**, 3377 (2022).