Laser & Photonics Reviews

LASER & PHOTONICS REVIEWS

www.lpr-journal.org

RESEARCH ARTICLE

Optical Kernel Machine With Programmable Nonlinearity

SeungYun Han[1] | Fei Xia[2,3] | Sylvain Gigan[2] | Bruno Loureiro[4] | Hui Cao[1]

[1]Department of Applied Physics, Yale University, Connecticut, USA | [2]Laboratoire Kastler Brossel, ENS-Universite PSL, CNRS, Sorbornne Université, Collège de France, Paris, France | [3]Department of Electrical Engineering and Computer Science, University of California, Irvine, California, USA | [4]Département d'Informatique, École Normale Supérieure - PSL & CNRS, Paris, France

Correspondence: Hui Cao (hui.cao@yale.edu)

Received: 18 November 2025 | **Revised:** 2 June 2026 | **Accepted:** 3 June 2026

Keywords: kernel regression | nonlinear operations | optical computing | reservoir computing

ABSTRACT

Nonlinear optical kernel machines enable parallel processing of complex, nonlinear data, achieving optimal performance when the kernel nonlinearity is tailored to a specific task. However, in conventional nonlinear optical approaches, adjusting the kernel nonlinearity typically requires varying the optical power. To overcome this issue, we present an optical kernel with structural nonlinearity that can be continuously tuned at low power. It is implemented in a linear optical scattering cavity with a reconfigurable micro-mirror array. By tuning the degree of nonlinearity through multiple scattering, we vary the kernel sensitivity and information capacity. We further optimize the kernel nonlinearity to best approximate parity functions from first order to fifth order for binary variables. Our scheme offers potential applicability across photonic platforms, providing programmable kernels with high performance and low power consumption.

1 | Introduction

Kernel machine learning is a powerful approach to learning complex, nonlinear patterns by mapping data to higher-dimensional spaces [1–5]. It is highly effective with limited data, offering strong performance, robustness, and good generalization [6–8]. Kernel methods are especially advantageous when the relationships in the data are not linearly separable. However, they often scale poorly with large datasets because kernel matrices grow quadratically with the number of samples, leading to high memory and computation costs [3, 5, 9].

Optical implementation of kernel machine learning involves the realization of kernel-based computations directly with light. These implementations can achieve extremely high-speed parallel processing ("high throughput") with significantly reduced energy consumption [10–17]. This makes them ideal for edge computing scenarios where latency and power efficiency are critical, such as high-speed signal classification [13, 18–23], chaotic system prediction [11, 12, 24, 25], signal recovery in optical communication [26–28], and ultrafast image or spectrum recognition [29, 30]. However, free-space optical systems often suffer from alignment sensitivity, and on-chip ones from fabrication imperfections and limited reconfigurability.

To build a feature-space embedding best suited for a specific task, we explore an optical kernel machine with tunable nonlinearity. In particular, we utilize the structural nonlinearity in a linear multiple-scattering cavity because it allows tuning the nonlinear order without the need to changing optical power [31, 32]. By encoding the input to the scattering structure rather than the illumination light, the output light of the cavity has a nonlinear relation with the input even without material or detection nonlinearities. The order of structural nonlinearity can be tuned continuously with cavity parameters at constant low power.

We present two methods of tuning the structural nonlinearity and show that stronger nonlinearity leads to higher sensitivity of the kernel output to the input change. This allows us to tune the kernel sensitivity for the learning task. High sensitivity leads to high variance and low bias; thus, the kernel is very flexible and can capture intricate nonlinear patterns in the data. A less sensitive kernel, on the other hand, is more rigid and robust to noise. We find that the information capacity of our optical kernel grows with structural nonlinearity. Information capacity is a measure of the expressive power of the kernel that reflects the complexity of the functions it can represent. By tuning the structural nonlinearity, we show that the nonlinear optical kernel can best approximate target nonlinear functions of varying complexity, e.g., parity functions of different orders for binary inputs.

2 | Tunable Optical Kernel

2.1 | Multiple-Scattering Cavity

There are different ways of realizing structural nonlinearity, which have been used for optical implementation of artificial neural networks [32–41]. We use a multiple-scattering cavity to achieve a nonlinear mapping between the scattering potential and the output light [31].

Our experimental setup is shown schematically in Figure 1a. A continuous-wave linearly-polarized light from a single-frequency laser (Keysight Technologies 81940A in an 8163B lightwave mainframe) at wavelength $\lambda = 1550$ nm and power 21.3 mW is injected into a spherical cavity of diameter 3.75 cm through a single-mode fiber. The opposite side of the cavity wall has an open window covered by a digital micro-mirror device (DMD). The DMD (Texas Instruments DLP9000X) modulation area within the window has 1280×800 micro-mirrors, each of which can be flipped to either $+12°$ or $-12°$. The inner surface of the cavity is rough and highly reflective. The input light is scattered multiple times by the cavity wall and the DMD. Part of the light escapes from the cavity through an opening and is directed by a mirror to an InGaAs camera (Xenics Xeva FPA-640). A linear polarizer is placed in front of the camera that records the output intensity pattern.

At low power, the output field E_{out} scales linearly with the input field E_{in} and can be expressed in the Born series:

$$E_{\text{out}} = \left[V + V(G_0V) + V(G_0V)^2 + \cdots\right]E_{\text{in}}\,, \tag{1}$$

where V represents optical scattering potential inside the cavity, and G_0 describes free-space propagation between subsequent bounces from cavity wall or DMD. The different orders of V in Equation (1) describe the number of bounces light experiences before leaving the cavity. V is the sum of the scattering potential of the DMD V_d and that of the cavity wall V_c, $V = V_d + V_c$. The DMD is reconfigurable, each configuration producing a unique output speckle pattern (Figure 1c). With multiple bounces from the DMD, the mapping from V_d to E_{out} is nonlinear. We term such nonlinearity as structural nonlinearity.

There are two ways of varying the degree of structural nonlinearity. One is changing the number of bounces off the DMD. It can be realized by varying the lifetime of light inside the cavity with tunable wall reflectivity. Since this is difficult to realize experimentally, we numerically simulate a two-dimensional (2D) cavity with adjustable wall reflectivity (Figure 1d) and show in the following section the tuning of structural nonlinearity. Another way to tune structural nonlinearity is by varying the effective modulation area on the DMD. This will change the relative weight of the reconfigurable scattering potential to the fixed one (Figure 1b). A larger DMD modulation area leads to a higher probability that light hits it in each bounce, resulting in a stronger structural nonlinearity. This method is adopted in our experiment.

2.2 | Structural Nonlinearity

The simplest way to characterize the strength of structural nonlinearity is to quantify its deviation from linear regression. In our numerical simulation of the 2D scattering cavity, the wall reflectivity is set to $r = 0.51, 0.78, 0.92, 0.97, 1$. For binary inputs ($\pm 1$) to 12 switchable mirrors ($\pm 15°$) in the cavity, the total number of mirror configurations is $K = 2^{12}$. For each configuration, we calculate the one-dimensional output field pattern $E_{\text{out}}(r)$ that contains 42 speckle grains. Then we fit a linear model $\hat{E}_{\text{out}} = W\,x$, where x represents the binary input, and quantify the deviation with the coefficient of determination

$$R^2 = 1 - \frac{\text{Var}(E_{\text{out}} - \hat{E}_{\text{out}})}{\text{Var}(E_{\text{out}})}. \tag{2}$$

Figure 2a shows that R^2 decreases continuously with increasing reflectivity r of the cavity wall. A higher reflectivity extends the lifetime of light inside the cavity, allowing more bounces of light off the mirror array before leaving the cavity, thus the output fields contain more nonlinear orders.

Experimentally, we vary the DMD modulation area. A binary pattern of dimension M is repeatedly written spatially to occupy a certain percentage of the total area of the DMD. Outside this area, the micro-mirrors are set to -1 ($-12°$). When the DMD modulation area increases, the base pattern is fixed in size but is repeated additional times. In Figure 1c, the base pattern contains $M = 10 \times 10$ macropixels, each comprising 16×10 pixels (micro-mirrors). All pixels within a macropixel are set to the same value of 1 ($+12°$) or -1 ($-12°$). The DMD modulation area is set to $12.5\%, 25\%, 50\%, 100\%$. Since the total number of DMD configurations is too large (2^{100}), we randomly sample $K = 2000$ patterns. The output intensity pattern recorded by the camera contains 258 speckle grains.

Since part of the input light is not scattered by the DMD-modulated area at all, its output forms a constant background. This background field E_{bg} contributes to a linear response of the output intensity to the DMD pattern. We denote the output field that varies with the DMD configuration as E_{mod}, and the total output field $E_{\text{out}} = E_{\text{bg}} + E_{\text{mod}}$. Thus, the output intensity is

$$I_{\text{out}} = |E_{\text{bg}}|^2 + |E_{\text{mod}}|^2 + 2\text{Re}[E^*_{\text{bg}}E_{\text{mod}}]. \tag{3}$$

The interference between the background and modulated fields leads to the cross term that scales linearly with E_{mod}. The background intensity $I_{\text{bg}} = |E_{\text{bg}}|^2 + \langle |E_{\text{mod}}|^2 \rangle_K$ is obtained by averaging the output intensity over random inputs to the DMD,

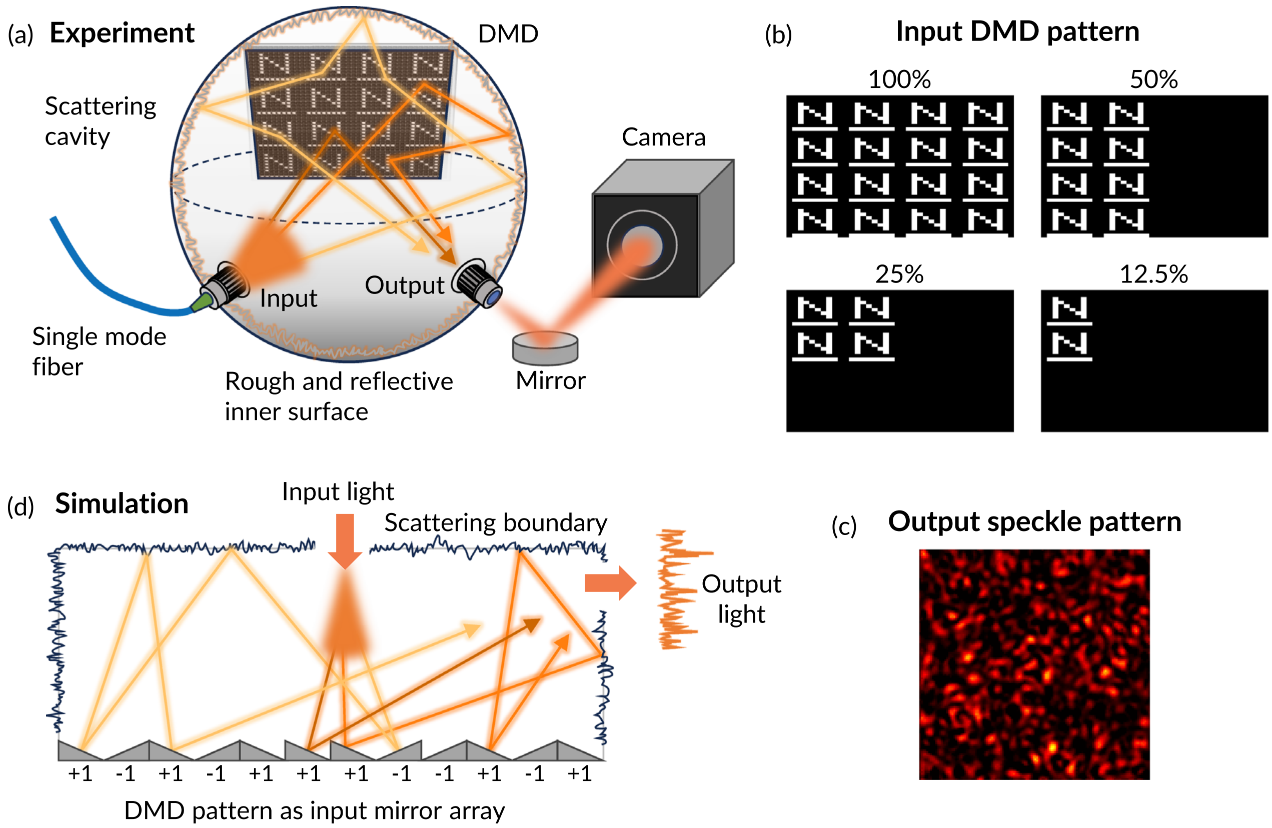

FIGURE 1 | Programmable nonlinear optical kernel. (a) Schematic of experimental implementation of the kernel in a linear scattering cavity. Input light at $\lambda = 1550$ nm is scattered by the rough cavity wall and a reconfigurable micro-mirror array (DMD). Part of the output light is recorded by a CCD camera. (b) Four binary masks written to the DMD. The micro-mirrors (pixels) tilt at angles of $+12°$ and $-12°$, redirecting incident light in 2 directions, each representing 1 and -1 in the input pattern. A representative base pattern with 10×10 macropixels is repeated to occupy 100%, 50%, 25%, 12.5% area of the DMD. (c) Measured speckle intensity pattern of output light from the scattering cavity. (d) Schematic of numerically-simulated 2D scattering cavity with varying reflectivity of the wall. The cavity dimension is $130\lambda \times 60\lambda$. One side of the cavity wall is covered by 12 mirrors, each of which can be flipped to $+15°$ or $+15°$ angle.

and is subtracted from I_{out}. Since E_{mod} contains the single scattering term that scales linearly with V_d (Equation 1), $I_{\text{out}} - I_{\text{bg}}$ has a linear term of V_d.

In Figure 2b, as the DMD modulation area increases, $I_{\text{out}} - I_{\text{bg}}$ shows a stronger deviation from linear regression with DMD inputs. This reflects a stronger structural nonlinearity with a larger DMD modulation area.

A more quantitative assessment of the structural nonlinearity can be done with Boolean function analysis [42]. For binary inputs x of dimension M, any output y can be decomposed into terms of different orders:

$$\begin{aligned}\varphi &= f(x_1, \dots, x_M)\\ &= c_0^{(0)} + \sum_{m_1=1}^{M} c_{m_1}^{(1)} x_{m_1} + \sum_{m_1=2}^{M} \sum_{m_2=1}^{m_1-1} c_{m_1,m_2}^{(2)} x_{m_1} x_{m_2}\\ &\quad + \cdots + c_{1,2,\dots,M}^{(M)} x_1 x_2 \cdots x_M,\end{aligned} \tag{4}$$

where x_m is the m-th element of the input, and $c_{m_1,\dots,m_n}^{(n)}$ is the n-th order coefficient of the term $x_{m_1} \cdots x_{m_n}$. The magnitude of the coefficient represents the contribution of the corresponding term to the output. Therefore, to evaluate the contributions of different orders, we summed the absolute value of the coefficients for each n:

$$S_n = \sum_m \left|\tilde{c}_m^{(n)}\right| \tag{5}$$

where m abbreviates $m_1, \dots, m_n$, and $\tilde{c}_m^{(n)}$ is the normalized coefficient

$$\tilde{c}_m^{(n)} = \frac{c_m^{(n)}}{\sum_n \sum_m \left|c_m^{(n)}\right|}. \tag{6}$$

The results of the Boolean function analysis will be presented in a later section on kernel regression.

2.3 | Kernel Expressivity

Our multiple-scattering cavity performs a nonlinear random projection that embeds the dataset into a high-dimensional feature space, i.e., a *random feature map*. Here, we evaluate how sensitive the output speckle pattern is to the input change.

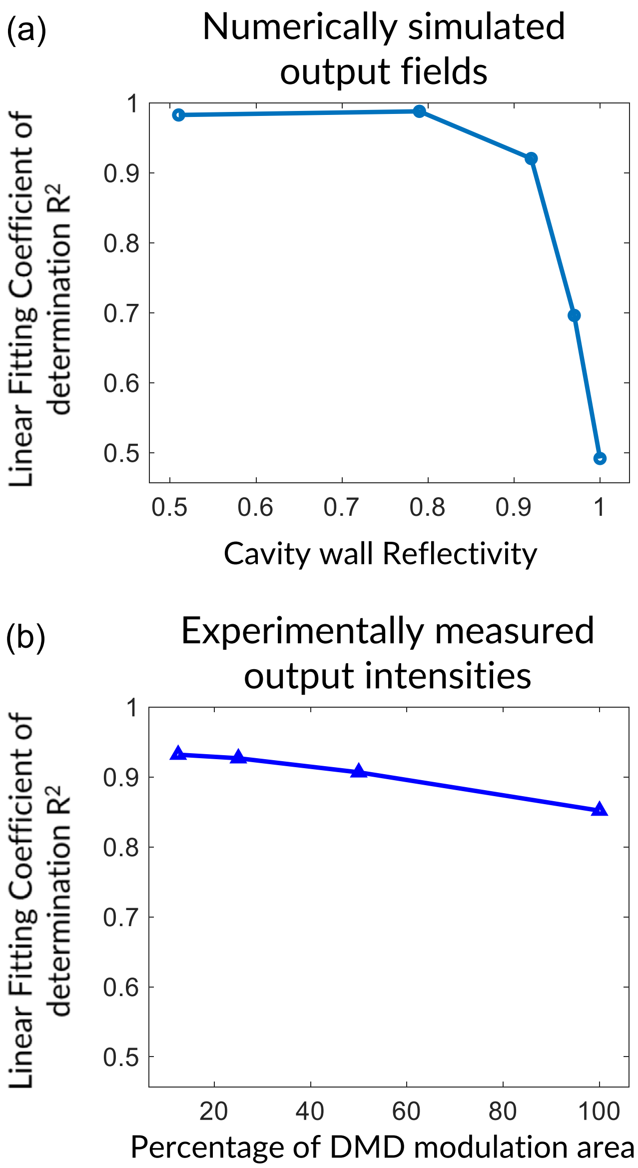

FIGURE 2 | Tunable structural nonlinearity in a linear multiple-scattering cavity. Linear regression of cavity input-output mapping gives the coefficient of determination R^2. (a) In the numerically-simulated 2D cavity, R^2 decreases with increasing cavity wall reflectivity. (b) In the experimental 3D cavity, R^2 drops as the DMD modulated area increases. The decrease of R^2 is a result of enhanced structural nonlinearity.

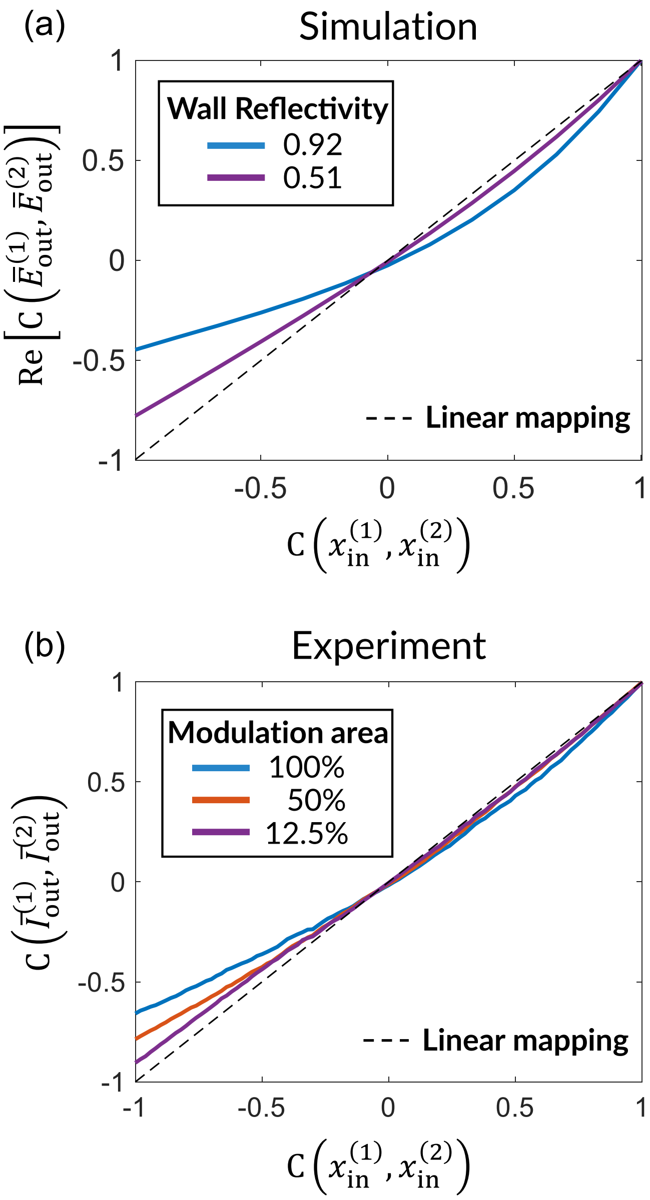

FIGURE 3 | Output sensitivity to input change. (a) Real part of the correlation between output field patterns for different mirror configurations increases with the correlation between input patterns (mirror configurations) in the simulated 2D cavity with wall reflectivity of 0.92 (blue) and 0.51 (purple). The black dashed curve is for linear input-output mapping. For higher wall reflectivity, the curve deviates more from the linear mapping due to stronger structural nonlinearity. (b) Correlation of output intensity patterns from the experimental 3D cavity increases with the correlation of input patterns on the DMD with the modulation area of 100% (blue), 50% (orange), and 12.5% (purple). More deviations from the linear mapping (black dashed line) result from stronger structural nonlinearity at larger DMD modulation area.

The change in input or output is quantified by the normalized correlation function:

$$\mathrm{C}(z_1, z_2) = \frac{z_1^{\dagger} \cdot z_2}{(z_1^{\dagger} \cdot z_1)^{1/2}\,(z_2^{\dagger} \cdot z_2)^{1/2}}, \qquad (7)$$

where z_1, z_2 represent a pair of input or output vectors.

For the simulated 2D scattering cavity, the input pattern on the 12 mirrors is a binary vector of dimension 12. The output field pattern $E_{\mathrm{out}}(r)$, which has 42 speckle grains, is sampled at 549 positions, producing a complex vector of dimension 549. Then the mean output field from $K = 2^{12}$ samples $\langle E_{\mathrm{out}} \rangle_K$ is subtracted from the output field E_{out} to obtain the feature map (output vectors) $\varphi = \bar{E}_{\mathrm{out}} = E_{\mathrm{out}} - \langle E_{\mathrm{out}} \rangle_K$. The cross-correlation of two output vectors corresponding to different mirror configurations (input vectors) is, in general, a complex number, but its imaginary part vanishes when the number of speckle grains is very large. This is because mathematically $C^*_{\mathrm{out}}(\varphi_1, \varphi_2) = C_{\mathrm{out}}(\varphi_2, \varphi_1)$ and statistically $C_{\mathrm{out}}(\varphi_2, \varphi_1) = C_{\mathrm{out}}(\varphi_1, \varphi_2)$ for random output fields. In Figure 3a, we plot the real part of the output field correlation

$\mathrm{Re}[C_{\mathrm{out}}]$ versus the input correlation C_{in}, for weak and strong structural nonlinearity (corresponding to low and high cavity wall reflectivity). Note that a linear feature map would imply a linear relationship between C_{in} and $\mathrm{Re}[C_{\mathrm{out}}]$ (black dashed line). Both curves start at $(1, 1)$ since identical inputs produce identical outputs. As the input correlation decreases, the output correlation falls more rapidly with stronger structural nonlinearity. The accelerated output decorrelation indicates higher output sensitivity to input change, namely, the kernel nonlinearity is enhanced by structural nonlinearity. Both curves pass through (0,0), which means that uncorrelated inputs generate uncorrelated outputs. As the input correlation becomes negative, the output correlation also goes negative, but with a smaller magnitude. This indicates that the nonlinear input-output mapping contains polynomial terms of even order, i.e., x^n with even n. With stronger nonlinearity, there are more terms with even order, and the output correlation further deviates from -1 when the input is anti-correlated ($C_{\mathrm{in}} = -1$).

Next, we switch to the experimental three-dimensional (3D) scattering cavity. For a base pattern of dimension M on the DMD, the input binary vector is of dimension M. The 2D output intensity pattern is converted into a 1D vector of dimension N. The mean intensity of all outputs from K samples $\langle I_{\mathrm{out}} \rangle_K$ is subtracted from the output intensity I_{out}, and the resulting zero-mean intensity is used as the feature map $\varphi = \bar{I}_{\mathrm{out}} = I_{\mathrm{out}} - \langle I_{\mathrm{out}} \rangle_K$. Figure 3b shows the correlation of the output intensity C_{out} as a function of the input correlation C_{in}, for different DMD modulation areas. With a larger modulation area, the stronger structural nonlinearity makes the curve deviate more from the straight line for linear input-output mapping. The output's reduced correlation or memory of the input allows a more diverse response of the kernel machine.

The high sensitivity of output to input change reflects the high expressivity of a kernel machine, which leads to large variance and low bias. While high variance means that the model is very flexible and can capture intricate, nonlinear patterns in the data, it is also highly sensitive to minor fluctuations and noise in the training set. In contrast, low sensitivity leads to high bias and low variance. A less sensitive model is more rigid and makes stronger assumptions about the data, which may cause it to miss important patterns.

Our kernel sensitivity can be tuned and optimized for a specific task, depending on the task's complexity and the noise level of the dataset. This will make the kernel machine accurate for the task without overfitting.

2.4 | Information Capacity

In this section, we consider the information capacity of the nonlinear kernel. Information capacity [43] characterizes the number of independent, noise-resolvable output modes of the kernel [44, 45]. It informs the performance and complexity trade-offs of kernel machines, primarily through the Information Bottleneck principle [46]. By viewing a kernel machine as a communication channel, the concept of capacity helps regulate the amount of relevant information preserved by the kernel, thereby influencing generalization and complexity [47].

Information capacity is defined from the singular value spectrum of responses φ (cavity outputs) under a constant signal-to-noise ratio (SNR). Specifically, the responses (output fields or intensities) $\{\varphi_k\}$ to K random binary inputs are assembled into a response matrix

$$\mathbf{\Phi} = [\, \varphi_1,\ \varphi_2,\ \ldots,\ \varphi_K \,]. \tag{8}$$

The singular value decomposition of the response matrix gives:

$$\mathbf{\Phi} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^T, \qquad \mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_s). \tag{9}$$

where s is the number of singular values. We normalize the singular values

$$\tilde{\sigma}_i = \frac{\sigma_i}{\sqrt{\sum_{j=1}^{s} \sigma_j^2}}. \tag{10}$$

Consider additive white noise and Q the signal-to-noise ratio (SNR), the information capacity is given by [43]

$$P = \sum_{i=1}^{s} \log_2\!\left(1 + Q\,\tilde{\sigma}_i^{\,2}\right). \tag{11}$$

This quantity is the effective number of channels of the kernel above the noise level, which is closely related to the expressive capacity of physical learning machines [47]. This is also closely connected to the *degrees of freedom* of the kernel, a quantity often appearing in the statistical analysis of kernel ridge regression and providing a notion of the effective dimension of the kernel feature space [48, 49].

The information capacity P for the simulated 2D scattering cavity is shown in Figure 4a as a function of the cavity wall reflectivity r for $Q = 10, 100$. P increases with r, indicating that the structural nonlinearity increases the information capacity. Higher SNR leads to larger information capacity, as expected. Figure 4b shows that the information capacity P of the experimental 3D scattering cavity increases with the DMD modulation area. Again, a higher nonlinearity yields a larger information capacity, as more singular modes become significant above the noise level. Here, the SNR of 21.1 is calculated from the ratio between input-induced and noise-induced intensity fluctuation of cavity output (see details in Section S1). These results demonstrate that the tuning of the structural nonlinearity controls the channel capacity and degrees of freedom of the kernel.

3 | Kernel Regression

The key idea of kernel regression is to map inputs to a high-dimensional feature space, where the target can be extracted with a linear model [3, 50]. We employ our multiple-scattering cavity for the physical mapping of the inputs on the DMD to the feature space of output speckles, replacing the digital computation with optical measurements. Then, a linear regression from the feature space (speckle pattern) is performed to estimate the target function. Since only a linear model is applied in the digital domain, the nonlinear capacity of the kernel regression originates from the physical kernel. Instead of using materials with nonlinear

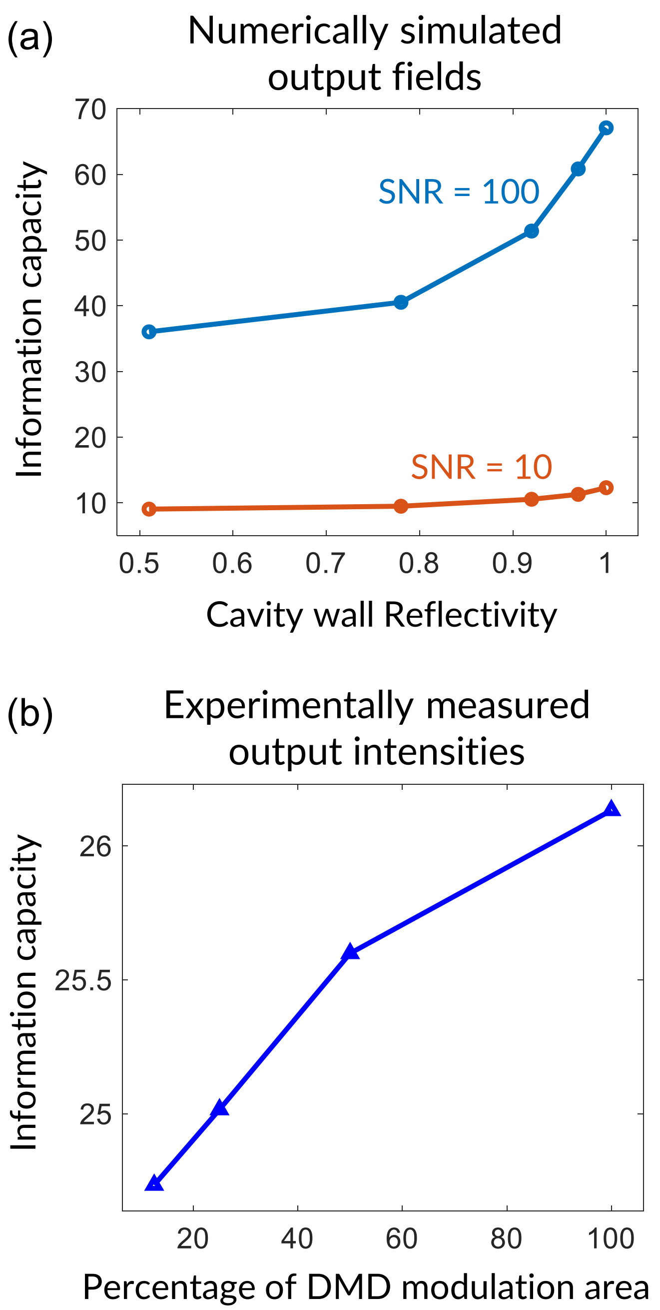

FIGURE 4 | Information capacity of the multiple-scattering cavity. (a) Information capacity increases with the wall reflectivity of the simulated 2D cavity. SNR = 100 (blue) leads to a larger information capacity than SNR = 10 (orange). (b) Information capacity increases with the DMD modulation area in the experimental 3D cavity with SNR = 21.1.

optical responses [51–63] or optical-electronic conversion [64, 65], the programmable nonlinearity in our cavity allows changing the characteristics of the feature space mapping, providing a distinct advantage of optimizing the kernel for different target functions.

As an example, we consider the parity function of varying order, also known as *sparse parity*, as the target function. For binary variables $x_i = -1$ or 1, $i = 1, 2, \ldots M$, the d-th order or d-sparse parity function is $f(x) = \prod_{i=1}^{d} x_i$. This function is highly variable and is notoriously difficult to learn with a kernel method, especially when the order d is equal to the input dimension M [66, 67]. Therefore, it provides an ideal testbed to study the degree of nonlinearity implemented by our optical kernel.

We start with the input dimension $M = 9$ by uploading 3×3 binary patterns to the DMD. We record the output speckle patterns of the 3D scattering cavity for all possible input patterns ($K = 2^9$). During measurement, a reference input pattern is periodically uploaded to the DMD, and the output intensity pattern is recorded to ensure it remains highly correlated during the entire acquisition (see Section S2 for more details). The order of input patterns is randomized to further minimize any influence of temporal drift of the output pattern. The output intensity pattern recorded by our camera contains about 5300 speckles, but we use only 55 speckles in a central square for parity functions of order $d = 1, 2, 3$. For output patterns, we randomly choose 90% for training of the linear regression and the remaining 10% for testing. Performance is quantified by the root-mean-square error (RMSE) on the test set:

$$\mathrm{RMSE}_{\mathrm{test}} = \sqrt{\frac{1}{N_{\mathrm{test}}} \sum_{j=1}^{N_{\mathrm{test}}} \left(y_j - \hat{f}(x_j)\right)^2}, \quad (12)$$

where N_{test} is the number of test patterns, $\hat{f}(x_j)$ is the prediction, and y_j is the ground truth.

To minimize the test error for a parity function of a specific order, we optimize the kernel nonlinearity by varying the DMD modulation area. For the first-order parity function $f(x) = x_1$, the test error $\mathrm{RMSE}_{\mathrm{test}}$ is minimum at the smallest (25%) DMD modulation area (Figure 5a). This is because a smaller area generates a more linear input-output relation, which better matches the linear target function. This is further confirmed in the Boolean function analysis [42]. As shown in Figure 5d, the first-order contribution S_1 to kernel output is larger at a smaller DMD modulation area. When the target function becomes nonlinear, e.g., the second-order parity function $f(x) = x_1 x_2$, the test error $\mathrm{RMSE}_{\mathrm{test}}$ first decreases and then increases with the DMD modulation area (Figure 5b). The Boolean function analysis reveals that the second-order contribution S_2 to the kernel output first increases and then slightly decreases with the modulation area (Figure 5e). With 75% of the DMD area being modulated, S_2 reaches the maximum, and the kernel best approximates the second-order parity function. The non-monotonic change of error with the DMD modulation area illustrates that the kernel performance is maximized when the kernel nonlinearity matches that of the target. The kernel with too weak nonlinearity (smaller DMD area) underfits, while a kernel with too strong nonlinearity (larger DMD area) overfits. This is further confirmed in our numerical simulation presented in Section S3. For a higher-order nonlinear function, e.g., the third-order parity function $f(x) = x_1 x_2 x_3$, the test error $\mathrm{RMSE}_{\mathrm{test}}$ decreases monotonically with increasing modulation area (Figure 5c). According to the Boolean function analysis, the third-order contribution S_3 to the kernel output increases monotonically with the modulation area (Figure 5f). Hence, stronger structural nonlinearity can better approximate higher-order parity functions. We note that our current kernel regression overfits, but in the benign regime, as explained in Section S3.

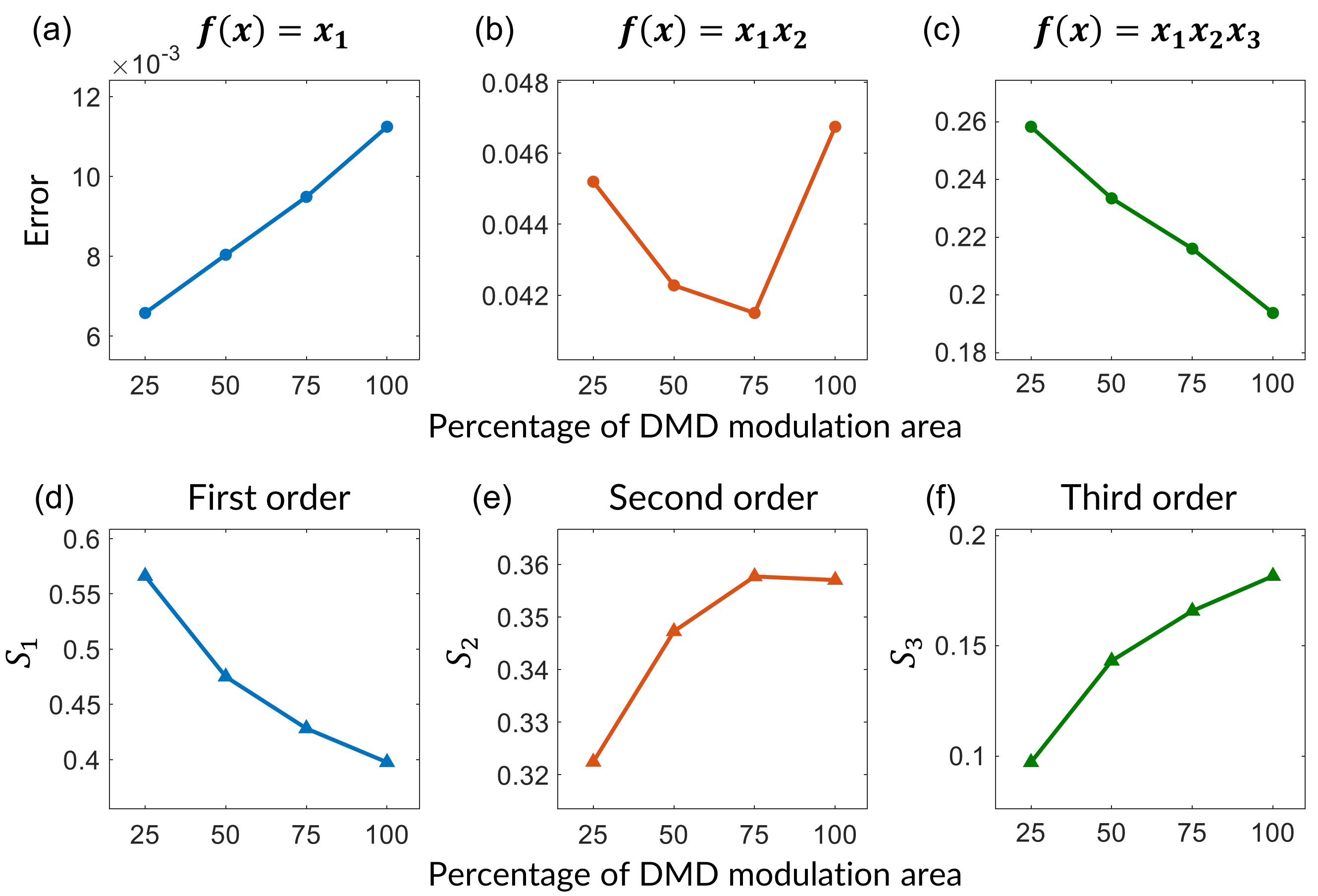

FIGURE 5 | Kernel regression to parity functions of varying order. The experimental 3D scattering cavity serves as a kernel by providing a high-dimensional feature map that maps input patterns on the DMD to optical intensity patterns at cavity output. Binary patterns $(x_1, x_2, \dots, x_9)$ are uploaded to the DMD, and repeated multiple times to occupy a certain percentage of the entire modulation area. Two-dimensional output intensity pattern (consisting of approximately 55 speckles) is linearly regressed (least square regression without regularization) to parity functions of first order $f(x) = x_1$ in (a), second-order $f(x) = x_1x_2$ in (b), and third-order $f(x) = x_1x_2x_3$ in (c). The regression error is plotted as a function of the percentage of DMD modulation area, which controls the kernel nonlinearity. (d-f) Magnitude sum of Boolean function coefficients of first order S_1 in (d), second-order S_2 in (e), and third-order S_3 in (f) are computed for varying percentage of DMD modulation area [42]. The regression error for the parity function of order d reaches the minimum at the DMD modulation area where S_d is maximum. Kernel regression has the best performance when its nonlinearity, tuned by the DMD modulation area, matches the order of the parity function.

In the above analysis, the input and output dimensions of the kernel are fixed when varying the DMD modulation area. Next, we set the DMD modulation area to 100%, and increase the dimension of input patterns on the DMD to approximate higher-order nonlinearity functions. We simultaneously vary the number of output speckles to ensure that the output dimension is not the limiting factor for the kernel performance. This is done by changing the dimension of the central window in the camera screen from which the intensity pattern is used as output.

Figure 6a shows the accuracy of kernel regression to the fourth-order parity function for varying input and output dimensions. Here, accuracy is defined as the percentage of correct binary predictions on the test set. For a fixed dimension M of the input pattern on the DMD, the accuracy increases with the output speckle number and eventually reaches 100%. For a higher M, a larger number of output speckles is needed to reach 100% accuracy. The number of required speckles (output features) scales as M^d for a target function of nonlinear order d (see details in Section S4). We checked experimentally that individual speckle grains are statistically independent, thus using additional speckles increases the dimensionality of optical feature space (details in Section S5). Although a smaller M requires a lower output dimension to approximate the fourth-order parity function, it cannot approximate the fifth-order parity function accurately, no matter how large the output dimension is. This is illustrated in Figure 6b, where the accuracy stays around 50% for $M = 9$. In contrast, the input dimension of $M = 14$ can approximate the fifth-order parity function with 91.8% accuracy when the number of output speckle grains is 4330. The observed trend suggests that the accuracy will further increase with the output dimension. This result confirms that a larger input dimension can approximate more complex functions. Finally, we also try to approximate the sixth-order parity function with $M = 14, 16$, but the accuracy remains low when the number of output speckle grains is increased to ~4000.

We note that a linear scattering system with input encoded in the incident field pattern only has second-order nonlinearity when the output intensity pattern is measured. Hence, it can-

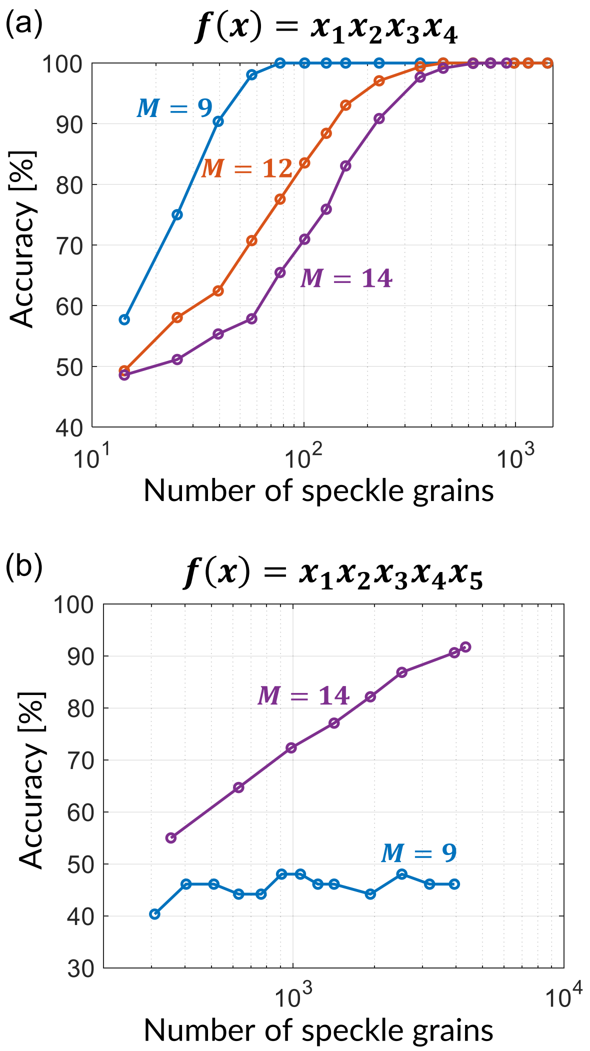

FIGURE 6 | Kernel regression with different input and output dimensions, targeting higher-order parity functions experimentally. (a) Kernel regression to the fourth-order parity function $f(x) = x_1x_2x_3x_4$ with binary inputs $(x_1, x_2, \ldots, x_M)$ of dimensions $M = 9, 12, 14$. As the number of output speckle grains increases, the accuracy of kernel prediction increases, eventually reaching 100%. A larger input dimension requires more speckle grains to perform the target function. (b) Kernel regression to the fifth-order parity function with input dimensions $M = 9, 14$. The kernel accuracy for $M = 14$ reaches 91.8% with 4330 speckle grains (features), but for $M = 9$ the accuracy stays around 50% as the number of output features (speckle grains) increases. In (a,b), the linear regression from output feature space (speckle pattern) is done with ridge regression with a relative ridge factor of 10^{-3}.

not approximate parity functions of order exceeding 2. Prior photonic reservoir and neuromorphic computing systems that used XOR/parity-type benchmarks reached up to the 4-bit with $\sim 80\%$ accuracy in Ref. [68]. Our optical kernel based on a multiple-scattering cavity reaches 5-bit with $> 90\%$ accuracy.

We evaluate the robustness of our optical kernel to experimental noise and perturbations by imposing random modifications of the DMD input patterns (randomly flipping a small fraction of micromirrors). The experimental results, shown in Section S6, reveal that the kernel robustness depends strongly on the target complexity, a linear target tolerating more perturbations than nonlinear ones. We further estimate the operation speed and power consumption of our optical kernel, as detailed in Section S7 and S8.

4 | Discussion and Conclusion

We have presented an optical kernel machine with programmable nonlinearity. It is implemented in a linear multiple-scattering cavity with spatial modulation of the scattering structure. By varying the number of times light is scattered by the modulated structure, the nonlinear strength is tuned continuously at constant low power. The kernel sensitivity and information capacity increase with the nonlinearity. This allows us to optimize the kernel nonlinearity to best approximate the parity functions of varying order for binary inputs. For sufficient input and output dimensions, our kernel can approximate the fifth-order parity function that is highly variable and nonlinear. These results not only show the capability of our kernel in approximating complex nonlinear functions, but also reveal the power of adapting the kernel nonlinearity for diverse learning tasks.

Compared to optical kernels based on material nonlinearity, our kernel with structural nonlinearity can offer higher-order nonlinearity without power increase. Widely used material nonlinearities are of second and third order, mainly limited by the peak power of optical pulses. In our case, the structural nonlinearity is up to the fifth order, even using a continuous wave. A key difference between material and structure nonlinearities is their scaling with power. For the second-order material nonlinearity, the optical signal decreases quadratically with pump power, and for the third-order, the decrease is cubic. However, for the structural nonlinearity, the output power is *independent* of the nonlinear order and scales *linearly* with the input power. Even a single input photon will undergo multiple scattering in our cavity and experience high-order structural nonlinearity. Moreover, $\chi^{(2)}$ and $\chi^{(3)}$ nonlinear materials often generate optical signals at different wavelengths from the input, making it difficult to chain the nonlinear processors to realize full optical deep neural networks. In contrast, the structural nonlinearity does not change the signal wavelength and may achieve comparable expressivity in a much simpler framework.

Our kernel nonlinearity can be further enhanced by optimizing the cavity design to increase the photon lifetime or the DMD modulation area. This will allow the kernel to approximate more complex functions of higher-order nonlinearity, e.g., parity functions of sixth-order and higher. Furthermore, by minimizing the cavity loss and maximizing output collection efficiency, the optical power may be reduced by orders of magnitude. Further power reduction is possible by using a more sensitive camera (see details in Section S8). In the current experiment, the light input to the scattering cavity is monochromatic. If it is replaced by a broadband light or optical pulses, the feature space of the kernel can be further expanded by spectral or temporal multiplexing [69]. This will increase the expressive power of the kernel. Finally, real-time in situ optimization of structural nonlinearity can be incorporated into the training process by taking the modulation area as one of the parameters that can be trained. The training steps will alternate between fitting the readout weights via linear regression and updating the DMD modulation area to match the kernel's nonlinear order to that of the task. This will enable a low-power, programmable optical kernel with target-adaptive nonlinear response.

Acknowledgements

We acknowledge Logan Wright and Florent Krzakala for stimulating discussion. This work is supported partly by the Office of Naval Research under MURI Grant No. N00014-24-1-2548 and by the Swiss National Science Foundation under Sinergia Grant CRSII5_216600 (LION). B.L. is supported by the National Research Agency (ANR) under the France 2030 program with the project references ANR-23-IACL-0008 (PR[AI]RIE-PSAI) and ANR-25-CE23-5660 (MAPLE), as well as the Choose France - CNRS AI Rising Talents program.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. N. Aronszajn, "Theory of Reproducing Kernels," *Transactions of the American Mathematical Society* 68, no. 3 (1950): 337–404.

2. C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning* 20, no. 3 (1995): 273–297.

3. B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, 2002).

4. J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge University Press, 2004).

5. A. Rahimi and B. Recht, "Random Features for Large-Scale Kernel Machines," *Advances in Neural Information Processing Systems* 20 (2007): 1177–1184.

6. S. Arora, S. S. Du, Z. Li, R. Salakhutdinov, R. Wang, and D. Yu, "Harnessing the Power of Infinitely Wide Deep Nets on Small-Data Tasks," *arXiv Preprint arXiv:1910.01663* (2019).

7. L. Brigato and L. Iocchi, "A Close Look at Deep Learning with Small Data," in *2020 25th International Conference on Pattern Recognition (ICPR)* (IEEE, 2021), 2490–2497.

8. A. Mallick, C. Dwivedi, B. Kailkhura, G. Joshi, and T. Y. J. Han, "Deep Kernels with Probabilistic Embeddings for Small-Data Learning," in *Uncertainty in Artificial Intelligence* (Proceedings of Machine Learning Research, 2021), 918–928.

9. Q. V. Le, T. Sarlós, and A. J. Smola, "Fastfood: Approximate Kernel Expansions in Loglinear Time," *arXiv Preprint arXiv:1408.3060* (2014).

10. D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel Photonic Information Processing at Gigabyte per Second Data Rates Using Transient States," *Nature Communications* 4, no. 1 (2013): 1364.

11. M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, "Large-Scale Optical Reservoir Computing for Spatiotemporal Chaotic Systems Prediction," *Physical Review X* 10, no. 4 (2020): 041037.

12. S. Sunada, K. Kanno, and A. Uchida, "Using Multidimensional Speckle Dynamics for High-Speed, Large-Scale, Parallel Photonic Computing," *Optics Express* 28, no. 21 (2020): 30349–30361.

13. D. Pierangeli, G. Marcucci, and C. Conti, "Photonic Extreme Learning Machine by Free-Space Optical Propagation," *Photonics Research* 9, no. 8 (2021): 1446–1454.

14. T. Zhou, F. Scalzo, and B. Jalali, "Nonlinear Schrödinger Kernel for Hardware Acceleration of Machine Learning," *Journal of Lightwave Technology* 40, no. 5 (2022): 1308–1319.

15. F. Hu, G. Angelatos, S. A. Khan, et al., "Tackling Sampling Noise in Physical Systems for Machine Learning Applications: Fundamental Limits and Eigentasks," *Physical Review X* 13, no. 4 (2023): 041020.

16. B. Çarpınlıoğlu and U. Teğin, "Genetically Programmable Optical Random Neural Networks," *Communications Physics* 8, no. 1 (2025): 349.

17. N. Desnos, S. Negrini, D. Labat, et al., "Phase and Amplitude Encoding for All-Optical Kernel Methods with a Photonic Lantern," in *European Conference on Lasers and Electro-Optics* (Optica Publishing Group, 2025), JSI_P_8.

18. K. Vandoorne, P. Mechet, T. Van Vaerenbergh, et al., "Experimental Demonstration of Reservoir Computing on a Silicon Photonics Chip," *Nature Communications* 5, no. 1 (2014): 3541.

19. A. Saade, F. Caltagirone, I. Carron, et al., "Random Projections Through Multiple Optical Scattering: Approximating Kernels at the Speed of Light," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2016), 6215–6219.

20. R. Ohana, J. Wacker, J. Dong, et al., "Kernel Computations from Large-Scale Random Features Obtained by Optical Processing Units," in *ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2020), 9294–9298.

21. G. Cong, N. Yamamoto, T. Inoue, et al., "Implementing Optical Analog Computing and Electro-Optic Hopfield Network by Silicon Photonic Circuits," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 107, no. 5 (2024): 700–708.

22. A. Foradori, A. Lugnan, L. Pavesi, and P. Bienstman, "Neuromorphic Photonic Circuits with Nonlinear Dynamics and Memory for Time Sequence Classification," *arXiv Preprint arXiv:2509.11721* (2025).

23. C. Zhou, P. Mu, Y. Huang, et al., "Deep Photonic Reservoir Computing Based on a Distributed Feedback Laser Array," *APL Photonics* 10, no. 2 (2025): 026103.

24. L. Larger, M. C. Soriano, D. Brunner, et al., "Photonic Information Processing Beyond Turing: An Optoelectronic Implementation of Reservoir Computing," *Optics Express* 20, no. 3 (2012): 3241–3249.

25. U. Paudel, M. Luengo-Kovac, J. Pilawa, T. J. Shaw, and G. C. Valley, "Classification of Time-Domain Waveforms Using a Speckle-Based Optical Reservoir Computer," *Optics Express* 28, no. 2 (2020): 1225–1237.

26. A. Argyris, J. Bueno, and I. Fischer, "Photonic Machine Learning Implementation for Signal Recovery in Optical Communications," *Scientific Reports* 8, no. 1 (2018): 8487.

27. K. Sozos, A. Bogris, P. Bienstman, G. Sarantoglou, S. Deligiannidis, and C. Mesaritakis, "High-Speed Photonic Neuromorphic Computing Using Recurrent Optical Spectrum Slicing Neural Networks," *Communications Engineering* 1, no. 1 (2022): 24.

28. L. C. Silva, P. R. Marciano, M. J. Pontes, M. E. Monteiro, P. S. André, and M. E. Segatto, "Integrated Photonic Neural Networks for Equalizing Optical Communication Signals: A Review," in *Photonics* 12, no. 1 (2025): 39.

29. Z. Denis, I. Favero, and C. Ciuti, "Photonic Kernel Machine Learning for Ultrafast Spectral Analysis," *Physical Review Applied* 17, no. 3 (2022): 034077.

30. T. Zhou, W. Wu, J. Zhang, S. Yu, and L. Fang, "Ultrafast Dynamic Machine Vision with Spatiotemporal Photonic Computing," *Science Advances* 9, no. 23 (2023): eadg4391.

31. Y. Eliezer, U. Rührmair, N. Wisiol, S. Bittner, and H. Cao, "Tunable Nonlinear Optical Mapping in a Multiple-Scattering Cavity," *Proceedings of the National Academy of Sciences of the United States of America* 120, no. 31 (2023): e2305027120.

32. F. Xia, K. Kim, Y. Eliezer, et al., "Nonlinear Optical Encoding Enabled by Recurrent Linear Scattering," *Nature Photonics* 18, no. 10 (2024): 1067–1075.

33. R. Faqiri, C. Saigre-Tardif, G. C. Alexandropoulos, N. Shlezinger, M. F. Imani, and P. Del Hougne, "PhysFad: Physics-Based End-to-End Channel Modeling of RIS-Parametrized Environments with Adjustable Fading," *IEEE Transactions on Wireless Communications* 22, no. 1 (2022): 580–595.

34. A. Momeni, B. Rahmani, M. Malléjac, P. Del Hougne, and R. Fleury, "Backpropagation-Free Training of Deep Physical Neural Networks," *Science* 382, no. 6676 (2023): 1297–1303.

35. M. Yildirim, N. U. Dinc, I. Oguz, D. Psaltis, and C. Moser, "Nonlinear Processing with Linear Optics," *Nature Photonics* 18, no. 10 (2024): 1076–1082.

36. C. C. Wanjura and F. Marquardt, "Fully Nonlinear Neuromorphic Computing with Linear Wave Scattering," *Nature Physics* 20, no. 9 (2024): 1434–1440.

37. Y. Li, J. Li, and A. Ozcan, "Nonlinear Encoding in Diffractive Information Processing Using Linear Optical Materials," *Light: Science & Applications* 13, no. 1 (2024): 173.

38. L. Abou-Hamdan, E. Marinov, P. Wiecha, P. del Hougne, T. Wang, and P. Genevet, "Programmable Metasurfaces for Future Photonic Artificial Intelligence," *Nature Reviews Physics* 7, no. 6, (2025): 331–347.

39. N. Richardson, C. Bosch, and R. P. Adams, "Nonlinear Computation with Linear Optics via Source-Position Encoding," *arXiv Preprint arXiv:2504.20401* (2025).

40. M. You, K. Arai, and S. Sunada, "Nonlinear Time Series Computing Using a Linear Optical Microcavity," *Optics Express* 33, no. 12 (2025): 24982–24994.

41. B. Liu, X. Mei, S. Shafi, et al., "Incoherent Light-Driven Nonlinear Optical Extreme Learner via Data Reverberation," *arXiv Preprint arXiv:2508.08428* (2025).

42. R. O'Donnell, *Analysis of Boolean Functions* (Cambridge University Press, 2014).

43. Z. Kuang, D. A. Miller, and O. D. Miller, "Bounds on the Coupling Strengths of Communication Channels and Their Information Capacities," *IEEE Transactions on Antennas and Propagation* 73, no. 6, (2025): 3959–3974.

44. C. I. Cox and C. Sheppard, "Information Capacity and Resolution in an Optical System," *Journal of the Optical Society of America A* 3, no. 8 (1986): 1152–1158.

45. D. A. Miller, "Communicating with Waves Between Volumes: Evaluating Orthogonal Spatial Channels and Limits on Coupling Strengths," *Applied Optics* 39, no. 11 (2000): 1681–1699.

46. N. Tishby, F. C. Pereira, and W. Bialek, "The Information Bottleneck Method," *arXiv Preprint physics/0004057* (2000).

47. F. Hu, G. Angelatos, S. A. Khan, et al., "Quantifying the Expressive Capacity of Quantum Systems: Fundamental Limits and Eigentasks," *arXiv Preprint arXiv:2301.00042* (2022).

48. T. Zhang, "Learning Bounds for Kernel Regression Using Effective Data Dimensionality," *Neural Computation* 17, no. 9 (2005): 2077–2098.

49. A. Caponnetto and E. De Vito, "Optimal Rates for the Regularized Least-Squares Algorithm," *Foundations of Computational Mathematics* 7, no. 3 (2007): 331–368.

50. T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel Methods in Machine Learning," *The Annals of Statistics* 36, no. 3 (2008): 1171–1220.

51. G. Van der Sande, D. Brunner, and M. C. Soriano, "Advances in Photonic Reservoir Computing," *Nanophotonics* 6, no. 3 (2017): 561–576.

52. A. Ryou, J. Whitehead, M. Zhelyeznyakov, et al., "Free-Space Optical Neural Network Based on Thermal Atomic Nonlinearity," *Photonics Research* 9, no. 4 (2021): B128–B134.

53. U. Teğin, M. Yıldırım, İ. Oğuz, C. Moser, and D. Psaltis, "Scalable Optical Learning Operator," *Nature Computational Science* 1, no. 8 (2021): 542–549.

54. L. G. Wright, T. Onodera, M. M. Stein, et al., "Deep Physical Neural Networks Trained with Backpropagation," *Nature* 601, no. 7894 (2022): 549–555.

55. G. H. Li, R. Sekine, R. Nehra, et al., "All-Optical Ultrafast ReLU Function for Energy-Efficient Nanophotonic Deep Learning," *Nanophotonics* 12, no. 5 (2023): 847–855.

56. P. L. McMahon, "The Physics of Optical Computing," *Nature Reviews Physics* 5, no. 12 (2023): 717–734.

57. T. Wang, M. M. Sohoni, L. G. Wright, et al., "Image Sensing with Multilayer Nonlinear Optical Neural Networks," *Nature Photonics* 17, no. 5 (2023): 408–415.

58. V. Nikkhah, M. J. Mencagli, and N. Engheta, "Reconfigurable Nonlinear Optical Element Using Tunable Couplers and Inverse-Designed Structure," *Nanophotonics* 12, no. 14 (2023): 3019–3027.

59. H. Wang, J. Hu, A. Morandi, et al., "Large-Scale Photonic Computing with Nonlinear Disordered Media," *Nature Computational Science* 4, no. 6 (2024): 429–439.

60. A. Momeni, B. Rahmani, B. Scellier, et al., "Training of Physical Neural Networks," *Nature* 645, no. 8079 (2025): 53–61.

61. S. Saeed, M. Müftüoğlu, G. R. Cheeran, T. Bocklitz, B. Fischer, and M. Chemnitz, "Nonlinear Inference Capacity of Fiber-Optical Extreme Learning Machines," *Nanophotonics* 14, no. 16, (2025): 2749–2760.

62. R. Yanagimoto, B. A. Ash, M. M. Sohoni, et al., "Programmable On-Chip Nonlinear Photonics," *Nature* 649 8096, (2026): 330–337.

63. T. Wu, Y. Li, L. Ge, and L. Feng, "Field-Programmable Photonic Nonlinearity," *Nature Photonics* 19 7, (2025): 725–732.

64. T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of Photonic Neural Networks Through In Situ Backpropagation and Gradient Measurement," *Optica* 5, no. 7 (2018): 864–871.

65. T. Wang, S. Y. Ma, L. G. Wright, T. Onodera, B. C. Richard, and P. L. McMahon, "An Optical Neural Network Using Less Than 1 Photon per Multiplication," *Nature Communications* 13, no. 1 (2022): 123.

66. A. Blum, A. Kalai, and H. Wasserman, "Noise-Tolerant Learning, the Parity Problem, and the Statistical Query Model," *Journal of the ACM* 50, no. 4 (2003): 506–519.

67. Y. Bengio, O. Delalleau, and N. Roux, "The Curse of Highly Variable Functions for Local Kernel Machines," *Advances in Neural Information Processing Systems* 18 (2005): 107–114.

68. A. Aadhi, L. Di Lauro, B. Fischer, et al., "Scalable Photonic Reservoir Computing for Parallel Machine Learning Tasks," *Nature Communications* 17 1, (2025): 1225.

69. X. Dong, K. Lion, F. Xia, et al., "Optical Computing with Spectrally Multiplexed Features in Complex Media," *arXiv Preprint arXiv:2510.04321* (2025).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: lpor71412-sup-0001-SuppMat.pdf.