ReDON: Recurrent Diffractive Optical Neural Processor with Reconfigurable Self-Modulated Nonlinearity

ZIANG YIN, Arizona State University, USA
QI JING, Arizona State University, USA
RAKTIM SARMA, Center for Integrated Nanotechnologies, Sandia National Laboratories, USA
ZHAORAN RENA HUANG, Rensselaer Polytechnic Institute, USA
YU YAO, Arizona State University, USA
JIAQI GU*, Arizona State University, USA

Diffractive optical neural networks (DONNs) have demonstrated unparalleled energy efficiency and parallelism by processing information directly in the optical domain. However, their computational expressivity is constrained by static, passive diffractive phase masks that lack efficient nonlinear responses and reprogrammability. To address these limitations, we introduce Recurrent Diffractive Optical Neural Processor (ReDON), a novel architecture featuring reconfigurable, recurrent self-modulated nonlinearity. This mechanism enables dynamic, input-dependent optical transmission through in-situ electro-optic self-modulation, providing a highly efficient and reprogrammable approach to optical computation. Inspired by the gated linear unit (GLU) in large language models, ReDON senses a fraction of the propagating optical field and modulates its phase or intensity via a lightweight, parametric function, enabling effective nonlinearity with minimal inference overhead. As a non-von Neumann architecture with the main weighting units (metasurfaces) being fixed, we substantially extend the DONN's nonlinear representational capacity and task adaptability via recurrent optical hardware reuse and dynamically tunable nonlinearity. We systematically investigate various self-modulation configurations to uncover the trade-offs between hardware efficiency and expressivity. On image recognition and segmentation tasks, ReDON improves test accuracy and mIoU by up to 20% over prior DONNs with optical or digital nonlinearities at comparable complexity and negligible power overhead. This work establishes a new paradigm for reconfigurable nonlinear optical computing, uniting the benefits of recurrence and self-modulation in non-von Neumann analog processors.

CCS Concepts: • **Hardware → Emerging optical and photonic technologies**; **Emerging architectures**.

ACM Reference Format:
Ziang Yin, Qi Jing, Raktim Sarma, Zhaoran Rena Huang, Yu Yao, and Jiaqi Gu. 2026. ReDON: Recurrent Diffractive Optical Neural Processor with Reconfigurable Self-Modulated Nonlinearity. *ACM Trans. Des. Autom. Electron. Syst.* 1, 1 (August 2026), 18 pages. https://doi.org/10.1145/3840398

1 Introduction

Diffractive optical neural networks (DONNs) have emerged as a promising platform for high-throughput, low-power neuromorphic processing [5, 6, 11, 16, 19, 20, 31]. By encoding neural weights directly into phase elements in the fabricated diffractive masks (e.g., metasurfaces), DONNs

*Corresponding author.

Authors' Contact Information: Ziang Yin, Arizona State University, Tempe, AZ, USA, ziangyin@asu.edu; Qi Jing, Arizona State University, Tempe, AZ, USA, qjing1@asu.edu; Raktim Sarma, Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, USA, rsarma@sandia.gov; Zhaoran Rena Huang, Rensselaer Polytechnic Institute, Troy, NY, USA, huangz3@rpi.edu; Yu Yao, Arizona State University, Tempe, AZ, USA, yuyao@asu.edu; Jiaqi Gu, Arizona State University, Tempe, AZ, USA, jiaqigu@asu.edu.

ACM 1557-7309/2026/8-ART
https://doi.org/10.1145/3840398

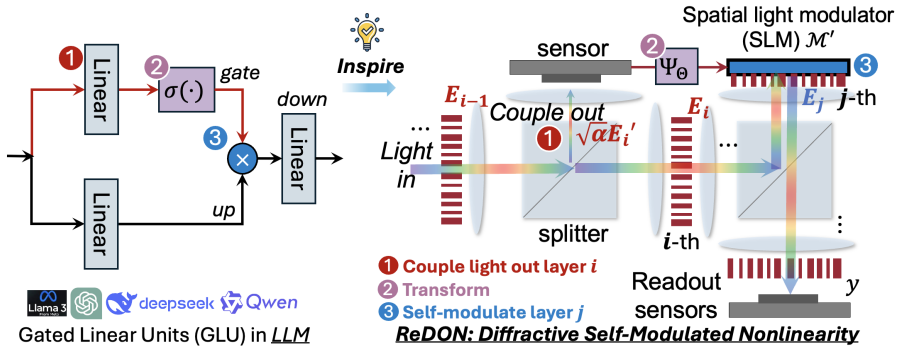

Fig. 1. Inspired by GLU in LLMs, we propose a diffractive self-modulated nonlinear unit in ReDON. A small fraction of the light is sensed and self-modulates downstream metasurfaces.

realize massively parallel in-memory computing, where weights are stored physically in sub-wavelength optical structures and applied to input with sub-nanosecond optical readout. This non-von Neumann computational physical system, where computation and memory are co-located, has enabled ultra-efficient optical inference for applications including computer vision, sensing, holography, generative AI, and scientific computing [5, 16, 19, 25, 31].

Despite these advantages, DONNs remain fundamentally limited by two key challenges: **weak optical nonlinearity and a lack of reconfigurability**, both rooted in the static, passive, and linear nature of diffractive metasurfaces. ❶ **Weak Nonlinearity**: Deep neural networks (DNNs) are fundamentally highly nonlinear functions for feature transformation. Yet DONNs are inherently (near) linear systems: cascaded passive diffractive layers reduce to a single global linear operator, with only weak square-law detection at photodetectors, dramatically restricting expressivity and limiting DONNs to tasks solvable by shallow NNs. Researchers have explored several strategies to introduce nonlinearity. Prior work has explored *all-optical nonlinearity* (e.g., saturable absorbers, $\chi(2)/\chi(3)$ nonlinear materials to build nonlinear optical encoder with enhanced accuracy, but often require unrealistically high optical power to trigger (100 kW/cm^2), have low energy efficiency (0.1%), rooted in the shallow interaction depths (<1 μm) in free-space DONNs [12, 23, 24]. *Structure nonlinearity*, such as repeatedly encoding inputs [32] or using cavity feedback [29], provides high-order nonlinearity but lacks parametric control and acts more like fixed optical reservoirs than programmable activations. To date, no solution offers efficient, programmable, low-latency nonlinear processing while maintaining DONNs' speed and energy advantages.

❷ **Restricted Reconfigurability**: A second major limitation of DONNs is their **reconfigurability**. Traditional DONNs depend on static metasurfaces whose nanostructures encode a fixed set of optical weight banks determined at fabrication time (similar to read-only memory). While this enables extremely fast, low-power inference (similar to memory readout), it prevents multi-channel, multi-layer neural computing and task adaptation in dynamic, evolving AI tasks. Hybrid optical-electronic architectures treat DONNs as static optical encoders and rely on digital backends for task adaptation. Multi-dimensional reconfiguration frameworks [31] and multi-task learning method [14, 34] explore mechanical rotation, permutation, and wavelength/polarization tuning to introduce system-level reconfigurability, but mechanical actuation still shows limited reliability and speed. As an active research topic, programmable metasurfaces can potentially offer per-element programmability but face substantial fabrication complexity, high insertion loss, and limited maturity for large-scale deployment [1]. Overall, existing DONNs lack a scalable mechanism for input-dependent, high-expressivity computation without re-fabricating diffractive layers.

Inspired by the dynamic gating mechanisms of modern gated linear units (GLUs) [22], we introduce ReDON, a diffractive optical neural processor that hybridizes fixed passive metasurfaces with a lightweight electro-optic self-modulation mechanism. ReDON senses a small fraction of the propagating optical field, processes it through a learnable parametric function, and uses the output to modulate the phase or amplitude of downstream diffractive layers. This introduces strong, tunable, input-dependent nonlinearity *far beyond traditional pointwise*, static activations, like ReLU, while requiring negligible inference overhead. Crucially, ReDON applies this modulation **recurrently** to reinforce nonlinearity and expand effective network depth. This in-situ recurrence follows diffusion-style iterative refinement: the system composes multiple simple optical nonlinear steps into highly expressive mappings. By hybridizing non-volatile metasurfaces with dynamic electro-optic self-modulation, ReDON overcomes long-standing limitations in nonlinearity and reconfigurability while preserving the efficiency and massive parallelism of diffractive optical in-memory computing. Our main contributions are summarized as follows:

- **Tunable, Self-Modulated Electro-Optic Nonlinearity**: We introduce a novel diffractive optical nonlinear mechanism that senses intermediate optical fields and applies a learnable, GLU-inspired gating function to modulate downstream transmission, providing input-dependent, reprogrammable nonlinear transmission.
- **Recurrent Diffractive Optical Processing**: We introduce in-situ recurrence to reinforce the DONN nonlinearity by reusing the same hardware system with dynamic parameter tuning, incrementally composing deep neural representations for highly expressive transformations.
- **Hybrid Reconfigurable Optical Processor Architecture:** ReDON unifies static, non-volatile metasurface weight banks with lightweight electro-optic self-modulation to enable dynamic, reconfigurable computation in a non-von Neumann optical setting.
- **Comprehensive Design Space Exploration**: We systematically analyze different architectural design settings and characterize their trade-offs in nonlinear expressivity, parameter efficiency, and hardware complexity. On classification and segmentation applications, our ReDON demonstrates an average of 20% higher accuracy with superior task adaptability and negligible power overhead compared to existing linear/nonlinear DONNs.

2 Preliminary

A diffractive optical neural network (DONN) [16, 19, 31] implements computation through cascaded free-space propagation and metasurface modulation. Let $x_{\text{in}} \in \mathbb{C}^N$ denote the input optical field. Each diffractive layer i applies a diagonal transfer matrix $T_i = \text{diag}(\exp(j\phi_i))$ representing spatially varying phase modulation, and propagation over distance z_i is modeled by a linear operator $U_i(z_i)$ derived from the Fresnel diffraction integral. The end-to-end transformation is

$$h_{\text{out}}(x_{\text{in}}) = \left(\prod_{i=1}^{K} U_i(z_i)\, T_i \right) U_0\, x_{\text{in}}, \qquad y = |h_{\text{out}}|^2, \tag{1}$$

where the square-law detection at the output plane provides the *only* nonlinearity. Thus, once the metasurfaces are fabricated, the DONN implements a single fixed global linear operator in the field domain, followed by a static, uncontrollable intensity nonlinearity. The absence of tunable or distributed nonlinear mechanisms fundamentally limits expressivity and task adaptivity.

Existing Optical Nonlinearity Mechanisms. Recent work has explored several directions to introduce nonlinearity into diffractive or optical neural systems [27, 28, 30]. *Structural nonlinearity* reuses the same linear diffractive elements multiple times so that the optical field is re-encoded and re-scattered through a static medium [15]. Through repeated interaction and interference, the resulting intensity response effectively approximates a polynomial mapping [29]. These methods

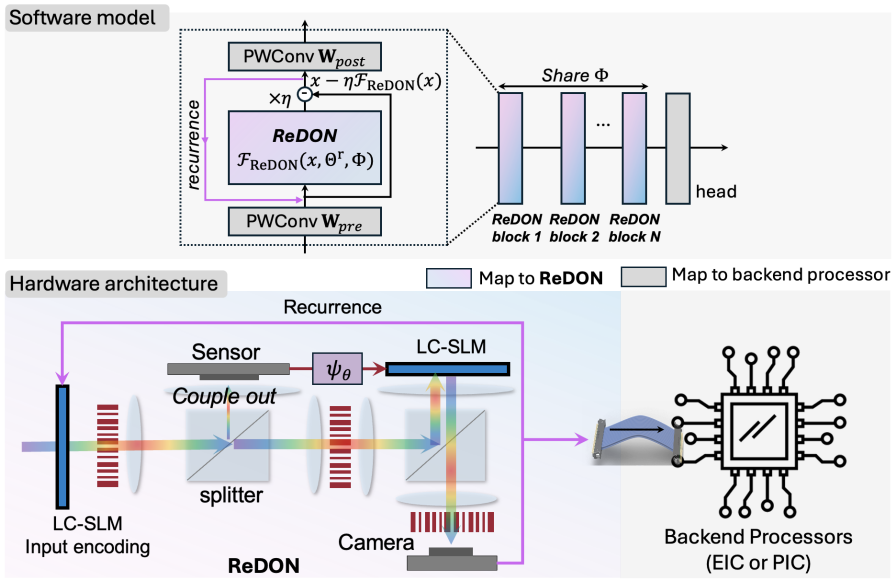

Fig. 2. ReDON software–hardware correspondence. Top: software abstraction of a titlename block, where the optical operator $\mathcal{F}_{\text{ReDON}}(\cdot)$ is integrated with lightweight pointwise layers and a compact digital head; recurrence reuses the same optical core with shared metasurface phases Φ and iteration-dependent modulation parameters Θ_r. Bottom: a corresponding optoelectronic realization using an input encoder (e.g., LC-SLM), passive metasurface stack, intermediate optical sensing (coupling ratio α), and a lightweight electronic processor that computes $\Psi(\cdot, \Theta)$ to drive a modulation plane for self-modulated nonlinearity.

preserve low power and remain fully optical, but the nonlinear mapping is implicitly determined by the fixed geometry, is hard to model analytically due to the chaotic nature of the cavity, or cannot be shaped or tuned after fabrication.

2.1 Gated Linear Units

Modern deep learning models, particularly large language models, increasingly rely on gated linear units (GLUs) [22] and related gating variants to achieve expressive, input-dependent transformations, e.g., SwigLU, SiLU. A GLU augments a linear projection with a learnable multiplicative gating function that modulates the output as

$$z = GLU(x) = (W_{\text{up}}x + b_{\text{up}}) \odot \sigma\big((W_{\text{gate}}x + b_{\text{gate}})\big);\ y = (W_{\text{down}}z + b_{\text{down}}), \tag{2}$$

where $\sigma(\cdot)$ is typically Swish or GELU. The gating branch conditions the main signal, providing a flexible mechanism for selective amplification, suppression, and routing, offering dynamic modulation while maintaining computational and parameter efficiency.

3 Recurrent Diffractive Optical Processor (ReDON)

We present ReDON, a recurrent diffractive optical neural processor. Unlike existing DONNs with fixed functionality and weak nonlinearity only at the photodetector arrays, ReDON introduces an electro-optic self-modulated nonlinearity and recurrent inference mechanism that simultaneously enables strong nonlinear expressivity and reconfigurability for non-von Neumann diffractive optical in-memory computing. As summarized in Fig. 2, ReDON admits a direct software-hardware correspondence: in software, a ReDON block wraps the optical operator $\mathcal{F}_{\text{ReDON}}(\cdot)$ with lightweight pointwise layers and a compact digital head, while in hardware the same computation is realized by sensing an intermediate optical field, computing a parametric transform $\Psi(\cdot, \Theta)$ on a lightweight

electronic backend, and driving a modulation plane (e.g., SLM / tunable metasurface) to self-modulate downstream transmission. Recurrence reuses the same fabricated metasurface stack across iterations by updating only Θ while keeping Φ shared.

3.1 Diffractive Self-Modulated Nonlinearity

Inspired by the powerful and parameter-efficient GLU mechanism in LLMs (Eq. (2)), we design ReDON with a diffractive self-modulation mechanism to introduce a dynamic input-dependent nonlinear response to DONNs. As illustrated in Fig. 1 and Fig. 2, we augment standard multi-layer DONNs with an auxiliary modulation branch that couples out and senses a small fraction (α) of intermediate light field E_i at the i-th metasurface, and convert it into electrical signals $\alpha|E_i|^2$. In the electrical domain, we perform a lightweight parametric transformation function on the sensed signal $\Psi(E, \Theta)$ and use it to control the spatial-light modulator (SLMs) to modulate the phase or intensity for the main branch light field at subsequent metasurface layer(s). As a simplified illustration, we assume the SLM modulates the j-th metasurface. Let E_i denote the complex amplitude of the optical field on the i-th metasurface output plane near-field, and $U_i(\lambda, z_i)$ denotes the linear diffraction matrix parameterized by wavelength λ and diffraction distance z_i between two metasurfaces. Each metasurface is a broadband phase mask that applies phase shifts $e^{j\Phi_i}$ to the incident light field. The formulation of an L-layer ReDON system transmission is as follows

$$\begin{aligned}
&\text{couple-out layer } i: E_i' = U(z_i/2)E_{i-1},\ E_{i-1} = \prod_{k=1}^{i-1}(e^{j\Phi_k}U_{k-1})E_0, E_0 = \mathcal{E}(x)\\
&\text{parametric transform: } \mathcal{M}_j = \Psi(\alpha|E_i'|^2, \Theta),\ \mathcal{M}' = f(\mathcal{M})\\
&\text{self-modulate layer } j: E_j = \sqrt{1-\alpha}\mathcal{M}_j' \odot \prod_{k=i}^{j}(e^{j\Phi_k}U_{k-1})E_{i-1};\\
&\text{transform after layer } j: y = \mathcal{F}_{\texttt{ReDON}}(x) = (1-\alpha)\left|U_L \prod_{k=j+1}^{L}(e^{j\Phi_k}U_{k-1})E_j\right|^2;
\end{aligned} \tag{3}$$

where the input $x \in [0, 1]$ is first encoded as the incident light field E_0 via an input encoding function $\mathcal{E}(\cdot)$, e.g., phase encoding $E_0 = e^{j\pi x}$, amplitude encoding $E_0 = x$, etc. $\alpha|E_i'|^2$ is the intermediate light intensity coupled out between the $(i-1)$-th and i-th metasurfaces. Here, $f(\cdot): \mathcal{M} \to \mathcal{M}'$ maps the generated control signals to either phase shifts or attenuation factors at the modulated metasurface(s). For **phase self-modulation**, $\mathcal{M}'$ represents additional phase shifts within $[-\pi/2, \pi/2]$; for **intensity self-modulation**, $\mathcal{M}'$ is mapped to real-valued attenuation factors.

$$\begin{aligned}
\mathcal{M}'_{\text{phase}} &= f_{\text{phase}}(\mathcal{M}) = e^{j(\mathcal{M}\%\pi - \pi/2)}\\
\mathcal{M}'_{\text{intensity}} &= f_{\text{intensity}}(\mathcal{M}) = \texttt{Clip}(\mathcal{M},\ 0,\ 1).
\end{aligned} \tag{4}$$

3.2 Nonlinear Expressivity Investigation

Input Encoding Investigation. The expressivity of the ReDON system is strongly influenced by the input encoding function $E_0 = \mathcal{E}(x)$. Figure 3 compares several commonly used encoding methods, including phase, amplitude, intensity, and complex-valued encoding. Among them, *phase encoding* consistently achieves the highest accuracy on classification tasks, whereas intensity and complex encodings exhibit noticeable training instability. This observation aligns with prior findings [15], which show that phase encoding preserves more optical energy and introduces a natural nonlinear dependence in the photodetection process. For these reasons, all subsequent experiments adopt the phase-encoding scheme $\mathcal{E}(x) = e^{j\pi x}, x \in [0, 1]$.

Scaled Differential Residual Output. Since the output detection plane measures only optical *intensity*, the resulting features are strictly non-negative, which limits representational richness

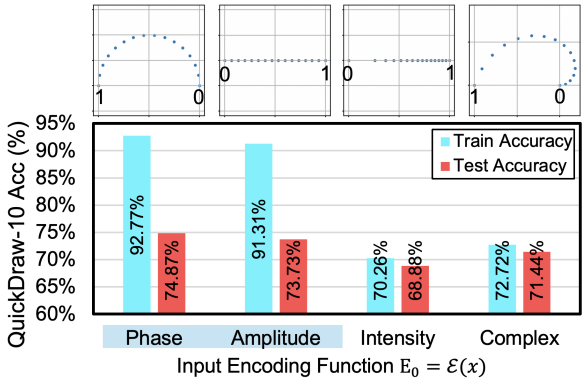

Fig. 3. Compare different input encoding functions for ReDON on QuickDraw-10 classification. Phase and intensity encoding show the best effects.

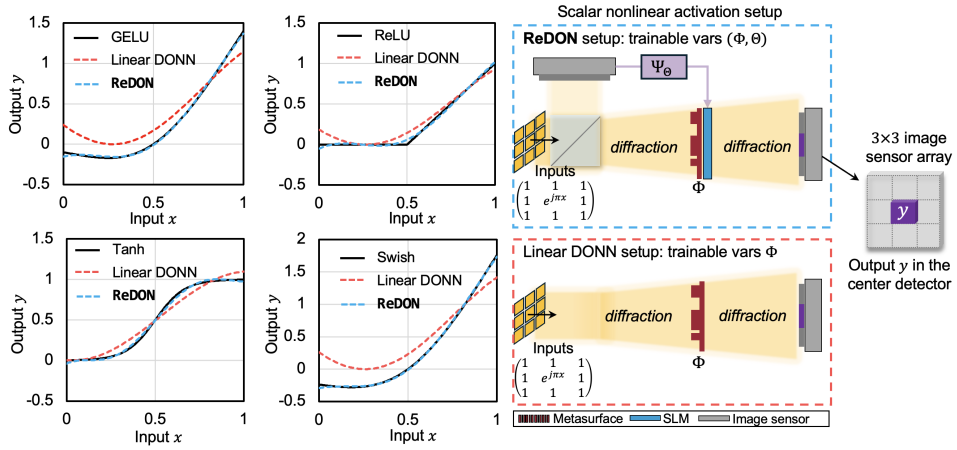

Fig. 4. Evaluation of nonlinear expressivity by fitting popular activation functions. We simulate a 3×3 phase-encoded input where only the center pixel sweeps phase $x \in [0, 1]$ (others fixed at unit amplitude and zero phase) and read out only the center detector intensity to form a scalar mapping. ReDON (with a second-order polynomial Ψ and residual readout $x - \eta\mathcal{F}_{\text{ReDON}}(x)$) approximates common nonlinear activations, while a linear DONN fails.

and causes numerical instability during training. A common workaround in prior optical neural networks is to use two parallel optical paths to form a differential signal, but this doubles the hardware cost. To overcome the non-negative readout and avoid doubling hardware overhead, we instead form *a scaled, differential residual path* between the input image and the detector readout, $y = x - \eta\mathcal{F}_{\text{ReDON}}(x)$. The scaling factor η controls the overall magnitude(norm) of the residual branch and compensates for the limited modulation range of phase-encoded diffractive propagation. This formulation restores sign flexibility to the features while avoiding additional optical paths.

Nonlinear Activation Expressivity. The proposed self-modulation mechanism is fully learnable, enabling input-dependent nonlinear responses. To probe its expressivity, Fig. 4 considers a minimal single-layer ReDON setup with 3×3 input resolution and asks whether it can fit common nonlinear activation curves (compared to a linear DONN baseline). We synthesize controlled interference patterns by sweeping only the center input phase, $x \in [0, 1]$ (field $e^{j\pi x}$), while keeping the remaining eight inputs at unit amplitude and zero phase. We then form a scalar output by reading the *center* photodetector intensity (normalized), yielding a 1D mapping $x \mapsto y$; in contrast, aggregating all

detector intensities is uninformative in our (approximately) lossless simulation because total optical power is conserved and thus nearly constant. By jointly optimizing the metasurface phases Φ and the coefficients Θ of a second-order polynomial modulation function $\Psi(\cdot, \Theta)$, ReDON accurately reproduces a variety of widely used nonlinear activations in Fig. 4. Existing linear DONNs, despite weak inherent nonlinearity, fail to approximate these functions and are constrained to nonnegative outputs under intensity-only readout. With the residual-output formulation, ReDON can realize sign-changing nonlinear mappings and, more generally, is not restricted to canonical pointwise activations; it can discover task-adaptive nonlinear functions during training.

3.3 Expressivity Augmentation with Multi-Layer Self-Modulation and Recurrent Architecture

Multi-Layer Self-Modulation. In a standard GLU, only a single multiplicative gating operation is applied. To further strengthen the nonlinear capacity of our diffractive system, we extend this idea by allowing the same sensed intermediate signal to modulate *multiple* downstream metasurface layers. As illustrated in Fig. 5(b), the optical field is coupled out at layer i, processed through individual Ψ functions, and used to control any subsequent metasurface layer(s) j, where $j \geq i$. For example, one may couple out at $i = 1$ to modulate layers $j = 2, 3, 5$. Increasing the number of modulated layers naturally strengthens nonlinear expressivity but also incurs higher hardware cost. In Sec. 4.2.3, we systematically evaluate this design space and quantify the trade-off between hardware complexity and expressivity gains achieved through multi-layer self-modulation.

Recurrent Network Architecture. Figure 2(top) shows the overall model architecture using ReDON as the basic building block. Lightweight pointwise convolutions are used for channel mixing and dimension matching, while the ReDON module performs per-channel spatial transformation. Multiple ReDON blocks can be stacked to form a deeper backbone (with N blocks), followed by a compact digital head for downstream tasks. To further enhance nonlinear expressivity and maximize the utility of the non-volatile passive metasurfaces, we introduce a **recurrent inference mechanism**. Each feature map is passed through the same ReDON system for R iterations, effectively composing multiple nonlinear transformations in situ. During recurrence, the metasurface phases Φ remain entirely shared and fixed, while we update per-channel Θ coefficients per recurrence iteration, which are stored in local SRAM and streamed to the modulation plane. That is, **each hidden channel receives a distinct set of parameters at each recurrence iteration**, i.e., $\{\Theta^{r,c}\}, r \in [R], c \in [C_{in}]$. With a fixed layer depth (i.e., constant $R \times N$), we find that recurrence of R=2 achieves the best expressivity and latency, while more recurrence gives saturated gain.

3.4 Parameter-Efficient Coefficient Sharing

The parametric function Ψ introduces dynamically tunable weights Θ, which must be stored in local SRAM and reloaded every inference. This incurs additional data movement, memory footprint, and compute overhead. To reduce these costs while preserving high nonlinear expressivity, we explore two simple yet effective parameter-sharing strategies for Ψ. Figure 5 summarizes the two approaches.

Spatial Group-wise Coefficient Sharing. Here, the $H \times W$ pixels on the modulation plane are divided into $p \times q$ spatial groups, each of size $r \times c$. If $p = H, q = W$, each pixel has its own independent set of coefficients Θ in the Ψ function, yielding maximal flexibility but the highest parameter count. At the opposite extreme, when $p = q = 1$, a single set of coefficients is shared across all modulated pixels. By adjusting the group setting p and q, we can smoothly trade expressivity for reduced memory and compute complexity, achieving up to rc times reduction in parameters.

Cross-layer Coefficient Sharing. This strategy becomes relevant when the sensed signal from layer i modulates multiple downstream metasurface layers j. Instead of assigning distinct parameters

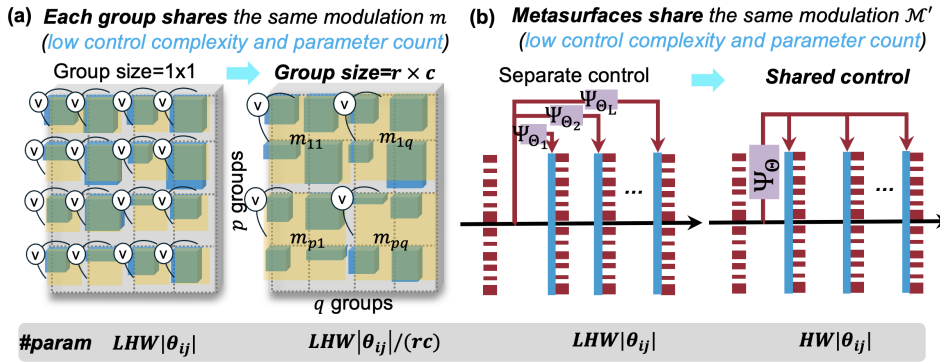

Fig. 5. (a) Group-wise and (b) layer-wise parameter sharing on metasurface modulation.

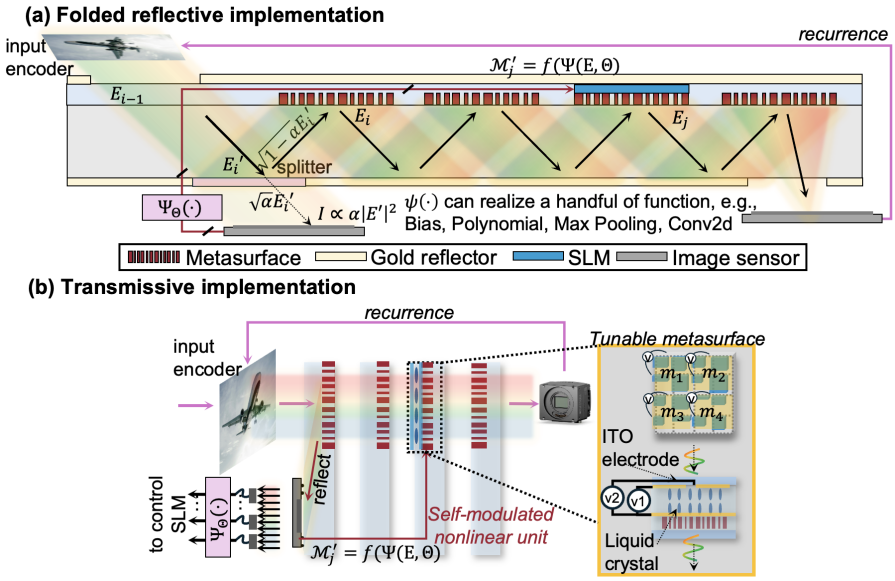

Fig. 6. (a) Folded reflective and (b) transmissive example implementations of ReDON architecture. The self-modulation mechanism can be realized by image sensors and SLM.

Θ for multiple modulated layers, we can potentially share the Θ coefficients for all layers to reduce the parameter count and compute cost by L times, where L is the number of modulated layers.

3.5 Implementation Practicality Justification

Figure 6 illustrates two feasible implementations of the proposed ReDON architecture. In the reflective design (Fig. 6(a)), both the diffractive layers and light path are integrated within a single substrate. The back plane employs a gold reflector, while the intermediate optical field at layer i is partially coupled out through a beam splitter and detected by an on-chip image sensor array. A small number of high-speed ADCs can sequentially read out the sensed intensities in a time-multiplexed fashion, amortizing ADC cost. The sensed signal $\alpha|E|^2$ can be transformed by $\Psi(\cdot, \Theta)$ in situ, either through compact analog circuits (for simple operations such as scaling, shifting, ReLU, or small-kernel convolutions) or lightweight digital units. The resulting control signal directly drives the modulation applied at the downstream metasurface layer(s) j. Alternatively, a transmissive implementation

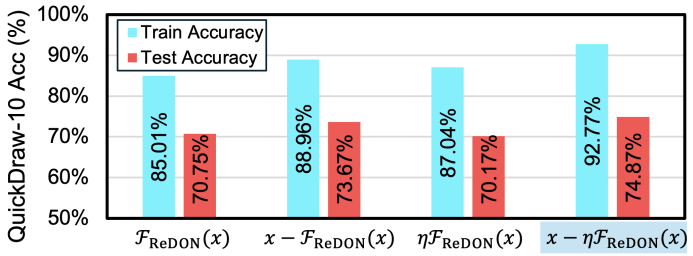

Fig. 7. Compare different readout strategies. Our differential, scaled residual shows the best accuracy on QuickDraw-10.

(Fig. 6(b)) can be constructed by engineering metasurface i to redirect a small fraction of the incident light (α%) toward an intermediate image sensor, while allowing the remaining light to propagate forward. Multiple metasurfaces can be fabricated on separate substrates with precise optical alignment.

Modulation Plane. A key hardware component is the tunable modulation plane at layer(s) j. Commercial liquid-crystal spatial light modulators (LC-SLMs) already support reprogramming rates near 10 kHz, phase modulation exceeding π, and more than 8-bit precision, sufficient for many proof-of-concept real-time AI systems [18]. Looking forward, emerging technologies in actively programmable metasurfaces [1, 21], promise significantly higher speeds and tighter integration. Electro-optic modulation in particular can reach modulation rates in the **gigahertz range** with per-element tunability with compact footprints, which in principle, *raise the achievable throughput by 2–3 orders of magnitude over LC-SLM-based prototypes*. Indium-tin-oxide (ITO) transparent electrodes can be used to minimize optical loss.

4 Evaluation

4.1 Evaluation Setup

4.1.1 Model and Dataset. As a case study, we assume our ReDON system has up to 5 cascaded metasurfaces and 1 ReDON block. We couple out α=5% of light. The hidden dimension is 3. For classification on CIFAR-10 [13] and QuickDraw-10/50 [8], the head is FC512ReLU-FC128ReLU-FC32ReLU-FC10. For segmentation on binarized Stanford Background [9], the head is Conv2DC128K3ReLU-Conv2DC64K3ReLU-Conv2DC2K1. All 5 metasurfaces Φ **are shared** across all blocks. Others parameters, including $\mathbf{W}_{pre}$, $\mathbf{W}_{post}$, η, Θ **are independently learned** for each ReDON block. Each metasurface has 32×32 meta-atoms. We use wavelength $\lambda = 532$ nm, meta-atom size $s = 400$ nm, and metasurface spacing $z = 8.42$ μm as used in the literature [19, 31].

4.2 Ablation Study

4.2.1 Scaled, Differential Residual Output. Figure 7 compares our proposed residual formulation $x - \eta\mathcal{F}(x)$ against several alternatives on the QuickDraw-10 classification task. Both the explicit scaling factor and the differential residual path substantially improve expressivity and training stability.

4.2.2 Parametric Function Ψ Selection. The choice of the parametric function Ψ provides a tunable trade-off between nonlinear expressivity and hardware efficiency. Figure 8 presents a comprehensive comparison across a wide range of candidates, including polynomials, common nonlinear operators (Tanh, ReLU, MaxPool), and lightweight operators such as 3×3 convolutions, evaluated under both phase and intensity self-modulation. The results reveal a general trend between achievable accuracy and computational complexity. Importantly, all candidates are implementable using compact digital

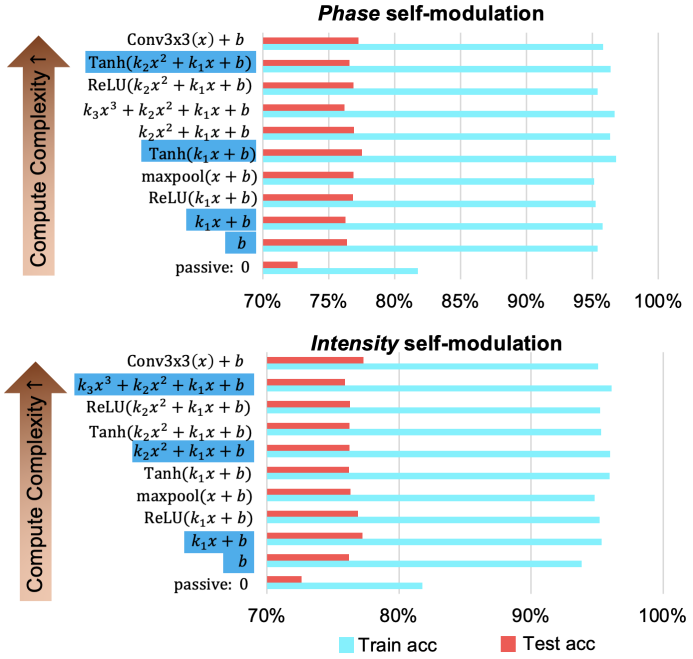

Fig. 8. Compare the expressivity of different $\Psi(E, \Theta)$ functions on phase and intensity self-modulation. $\Psi(\cdot, \Theta)$ are ordered by computing complexity. Each highlighted function represents the highest-accuracy model at different complexity levels.

logic, while several simple operations, such as scaling, bias, or small convolutions, can potentially be realized directly in the analog electrical domain without ADCs. A key observation is that even the *simplest affine transformations (scale + bias) significantly outperform the passive baseline (i.e., no self-modulation)*. The strong performance of even an affine Ψ confirms that **the dominant nonlinear mechanism arises from the intermediate field detection and input-dependent self-modulation. Ψ itself does not need to be nonlinear. Increasing the complexity of Ψ provides additional but secondary gains in expressivity.** Therefore, the GLU analogy in ReDON refers primarily to its input-conditioned multiplicative gating mechanism, rather than requiring an explicit nonlinear activation within Ψ.

For further evaluation, we select two nonlinear functions (third-order polynomial and scaled Tanh) and two linear functions (affine transform and simple bias) as representatives spanning a range of hardware costs.

4.2.3 Modulation Layer Exploration. We next examine how the choice of modulation layers affects expressivity and hardware cost. Figure 9 enumerates all combinations of coupling-out layers $i \in [1, 5]$ and modulated layers $j \in [i, 5]$. For **phase self-modulation**, we observe two conclusions. (1) **Later modulation layers yield higher accuracy**: For any fixed couple-out index i, accuracy increases as the modulation target j moves deeper into the stack, with j=5 consistently offering the best performance. This suggests that a longer linear transformation path between layers i and j provides richer interference structure for modulation, whereas transformations occurring after modulation have comparatively limited impact. (2) **Modulating more layers increases expressivity**: The strongest performance is achieved when all five layers are modulated, i.e., $i = 1$ and $j = 1, 2, 3, 4, 5$.

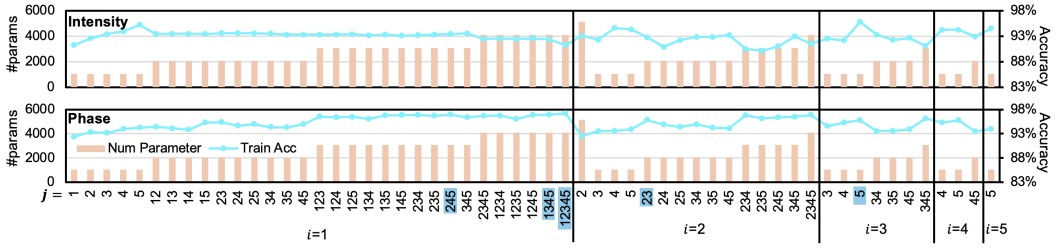

Fig. 9. Parameter count and accuracy comparison with different sensing layer i to modulated layers j using phase and intensity self-modulation. Phase self-modulation shows better accuracy than intensity. The blue markup shows the best setting under different #layers in j. These results motivate our choice of phase self-modulation and multi-layer modulation in the main experiments.

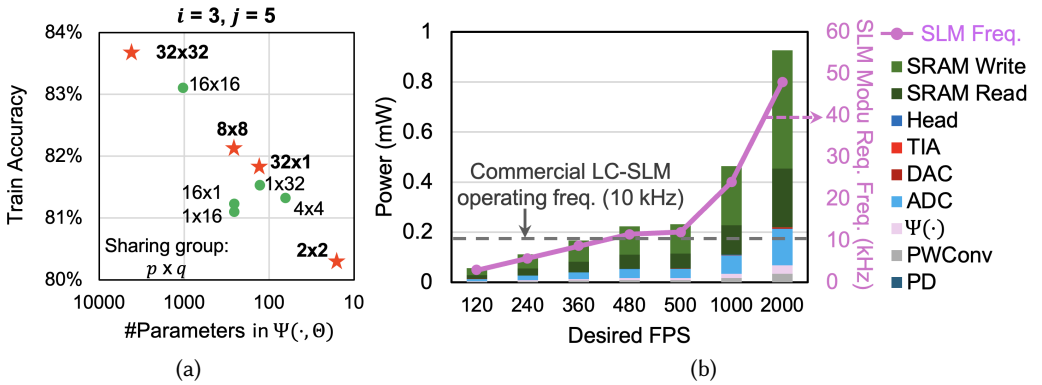

Fig. 10. (a) Accuracy and #Params Pareto-front selection with different parameter sharing group sizes $p \times q$ on phase self-modulation with i=1, j=1,2,3,4,5 on QuickDraw-50. (b) Power breakdown and scaling with higher FPS.

While for **intensity self-modulation**, however, increasing the number of modulated layers drastically reduces accuracy. Unlike phase modulation, intensity modulation inherently attenuates optical power, and repeated attenuation across layers significantly limits expressivity due to compounded energy loss (and consequently low optical efficiency). Based on these observations, we select representative configurations across different hardware budgets for later experiments: 3 to 5, 2 to (2,3), 1 to (2,4,5), 1 to (1,3,4,5), and 1 to (1,2,3,4,5).

4.2.4 Parameter Sharing Strategy. Figure 10 evaluates different spatial parameter-sharing group sizes in the accuracy-parameter-count trade-off, using i=3, j=5 as a representative configuration. Along the Pareto front, we identify four representative operating points, 32×32, 8×8, 32×1, and 2×2, that provide well-balanced trade-offs between expressivity and model compactness. Depending on the available on-chip memory budget for storing the Θ parameters, one can choose an appropriate group size to meet system-level constraints.

Figure 11 further combines the optimal group-sharing configurations (from Fig. 10) with the best-performing modulation-layer settings (from Fig. 9) to reveal the overall accuracy-parameter-count landscape. A clear trend emerges: *modulating more layers has a stronger positive impact on expressivity than simply reducing the parameter-sharing group size.* For instance, comparing 32×32 3-to-5 vs. 16×16 1-to-12345, with similar parameter counts, the latter consistently achieves

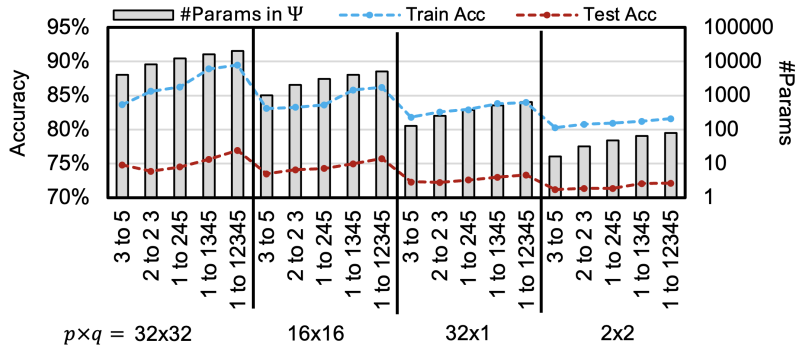

Fig. 11. With the best modulation layer settings (i to j) from Fig. 9, we explore #Params vs. accuracy trade-off with different sharing group p, q.

higher accuracy. It indicates that **modulating more diffractive layers contributes more to expressivity than increasing per-pixel parameter granularity**, suggesting a practical design rule for future nonlinear DONNs.

4.2.5 System Hardware Cost Analysis. **Throughput Analysis**. The primary throughput limitation of the proposed system stems from the modulation speed of commercially available spatial light modulators (SLMs), which currently operate at a maximum of $F_{\text{SLM,max}} \approx 10\,\text{kHz}$. Although our self-modulation mechanism substantially enhances nonlinear expressivity, its reliance on SLM reprogramming makes it the dominant inference speed bottleneck. On the digital side, the detection head contains a total of 814,058 trainable parameters, corresponding to only 1.6×10^6 FLOPs, negligible compared with the DONN part.

To characterize the SLM modulation speed requirement, we model the required modulation frequency as $F_{\text{SLM,req}} = R \cdot N \cdot C_{\text{mid}} \cdot \text{FPS}$, where R is the number of recurrent iterations, N is the number of ReDON blocks, C_{mid} denotes the hidden-channel dimension, and FPS is the target application frame rate. The system, using a commercial liquid-crystal SLM, remains feasible when $F_{\text{SLM,req}} \leq F_{\text{SLM,max}} \approx 10\,\text{kHz}$.

Given the SLM speed limit and a spatial resolution of 32×32, the maximum digital throughput after photodetection and readout from the CMOS sensor is bounded by $f_{\text{sample}} = S \cdot F_{\text{SLM,req}}$, where S is the total number of pixel readout samples. Under our typical configuration, this corresponds to $f_{\text{sample}} \approx 2.95$ MS/s, which is well within the capability of modern ADC/DAC hardware, commonly supporting sampling rates on the order of 10–14 GS/s [4, 17].

Overall, with four ReDON blocks (N=4) and fewer than two recurrent iterations (R=2) per block, the system can process a 32×32 spatial input with 3 hidden channels at up to 416 FPS, or 78 FPS when using 16 hidden channels. These results demonstrate that ReDON remains compatible with real-time throughput. As mentioned in the next section, one can potentially enable much higher throughput using a programmable metasurface that supports GHz modulation speed.

Power Breakdown. We further analyze the power overhead related to the self-modulated circuitry. The introduced power overhead is overall negligible (<1 mW) compared to the laser power (>100 mW) of the DONN system. We estimate data-converter power using a standard figure-of-merit style scaling where power grows approximately linearly with sampling rate and exponentially with resolution (number of quantization levels). For a b_{in}-bit DAC and b_{out}-bit ADC operating at sampling rate f, we model:

$$P_{DAC}(b_{in}, f) = P_{0,DAC} \frac{2^{b_{in}}}{b_{in}+1} f, \tag{5}$$

Table 1. Comparison across different nonlinearity designs on 3 benchmarks, our proposed ReDON with scaled Tanh as Ψ consistently achieves the best performance. ReDON shares Θ across all 5 modulated layers. With more recurrence (R) and blocks (N), ReDON shows higher expressivity. We replace our self-modulation with standard DONN or DONN with a digital Tanh activation function, which shows much worse results than our self-modulation.

Different Nonlinear Strategy	CIFAR-10		QuickDraw-50		Stanford Background Seg. Binarized Seg	
	Train Acc.	Test Acc.	Train Acc.	Test Acc.	Train mIoU	Test mIoU
Saturable Absorption [33]	61.9%	61.1%	80.1%	71.3%	72.2%	47.1%
Reflection Coating [7]	58.4%	58.3%	76.6%	68.5%	68.3%	41.3%
Encoding (Phase) [15]	61.2%	60.2%	79.4%	70.4%	76.1%	53.4%
Encoding (Intensity) [15]	58.3%	58.6%	76.5%	68.7%	66.4%	46.9%
Encoding (Complex) [15]	59.7%	59.6%	77.9%	69.8%	68.2%	44.3%
Digital activation (Log)	60.1%	58.1%	78.3%	68.3%	69.5%	38.2%
Digital activation (Square)	61.4%	60.4%	79.6%	70.6%	72.1%	51.7%
Digital activation (Tanh)	62.6%	60.7%	80.7%	70.9%	77.2%	55.6%
ReDON i=1, j=1,2,3,4,5 Tanh(k_1x+b) cross-layer share $p\times q$=32×32, $R=1$, $N=1$	74.4%	64.8%	86.7%	76.7%	83.3%	68.7%
ReDON Diffractive Only $R=2$, $N=4$	70.7%	61.2%	82.9%	73.1%	70.9%	58.2%
ReDON Digital Tanh $R=2$, $N=4$	77.4%	66.4%	91.8%	77.6%	85.4%	63.8%
ReDON i=1, j=1,2,3,4,5 Tanh(k_1x+b) cross-layer share $p\times q$=32×32, $R=2$, $N=4$	83.6%	74.5%	98.8%	81.33%	89.2%	72.4%

$$P_{ADC}(b_{out}, f) = P_{0,ADC}\frac{2^{b_{out}}}{b_{out}+1}f, \tag{6}$$

where $P_{0,DAC}$ and $P_{0,ADC}$ are technology- and architecture-dependent constants derived from representative high-speed converter designs.

We estimate the power of the lightweight digital blocks (PWConv, Ψ when implemented digitally, and the classification/segmentation head) by converting their required MAC throughput into power using an energy-per-MAC model. Using a 45 nm INT8 MAC energy of $E_{\mathrm{MAC}} \approx 0.23$ pJ/MAC [10], the corresponding efficiency is $\eta_{\mathrm{MAC}} = 1/E_{\mathrm{MAC}} \approx 4.35\times10^{12}$ MAC/J. Therefore,

$$P_{\mathrm{digital}} = \frac{\mathrm{MAC/s}}{\eta_{\mathrm{MAC}}} = (\mathrm{MAC/s})\cdot E_{\mathrm{MAC}}. \tag{7}$$

In Fig. 10(b), we show the power scaling with increased target FPS. With a commercial liquid-crystal SLM that supports 10 kHz modulation rate, our ReDON can achieve over 400 FPS throughput. A higher FPS can be realized using advanced electro-optic tunable metasurfaces [2, 3, 26]. We can see SRAM (45 nm) access takes the largest proportion of power, while sensor array, TIA, ADC/DAC, and digital (PWConv/head) (45 nm), due to time-multiplexing and low cost, only take less than 3% of the power overhead. Our ReDON system (N=1) can potentially support over 1000 FPS end-to-end inference with only ~1 mW electrical power overhead compared to passive DONNs, showing its potential for low-power edge deployment on real-time inference tasks.

4.3 Main Results

Using the optimal settings identified through our ablation studies, Table 1 compares our self-modulated nonlinearity with prior optical and hybrid nonlinear mechanisms. We report both training accuracy (reflecting expressivity) and test accuracy (reflecting generalization). For fully

Table 2. Transferability evaluation of ReDON (R=2, B=4). On Fashion-MNIST→QuickDraw-10, we show classification test accuracy. On Darcy flow→Navier-Stokes PDE solving, we show the test mean-square error (MSE). We train on the initial task and adapt to the second task, reusing fabricated metasurfaces.

Design	Accuracy ↑		Mean-Square Error (MSE) ↓	
	FMNIST	QuickDraw-10 adapt from FMNIST	Darcy	Navier-Stokes adapted from Darcy
Saturable Absorption [33]	80.1%	53.1%	0.095	0.1822
ReDON	95.1%	87.2%	0.015	0.1035

optical nonlinearities, we benchmark against saturable absorbers and reflection-based nonlinear coatings. We also compare against natural nonlinearities arising from input encoding [15], as well as digital nonlinear activations (logarithmic, square, Tanh). **All baselines use the same metasurface stack, input resolution, and digital head; only the nonlinearity mechanism differs**. Across CIFAR-10, QuickDraw-50, and Stanford Background segmentation, our ReDON achieves substantially higher accuracy (**+20%** on average) and mIoU, even with a single block (N=1) and a lightweight scaled Tanh Ψ function. Notably, prior DONNs with a single diffractive stage typically achieve <60% accuracy on CIFAR-10 in the literature. To our knowledge, **ReDON is the first to reach ~65% accuracy using only one block**, highlighting its markedly improved expressivity.

In the lower half of Table 1, we perform a controlled ablation by keeping all architectural components identical but replacing our self-modulation mechanism with either (1) a purely linear DONN or (2) a DONN augmented with digital nonlinear activation. Compared to simply adding digital activations, the self-modulation learns task-specific nonlinear functions with much higher accuracy, explaining the **clear gain from self-modulation over the digital activation function** at the same recurrence/block depth. It also shows its **benefits compound when scaling to deeper architectures**. Our evaluation focuses on modest-scale inputs for proof-of-concept evaluation, also due to training resource limitations. Extending ReDON to high-resolution vision tasks is an important direction for future work.

Task Adaptability of ReDON. We assess ReDON 's task adaptability on both classification and PDE solving, as summarized in Table 2. After training on a source task, we **freeze all metasurfaces** and adapt only the head, PWConv, and the tunable $\Psi(\cdot, \Theta)$ functions on new tasks. Under this constraint, ReDON achieves substantially better transfer performance: on QuickDraw, it improves accuracy by 34% over the baseline, and on Navier-Stokes it reduces the adapted solution error by 40%. We further compare ReDON to a nonlinear DONN based on saturable absorption [33], which can only finetune the head with a fixed diffractive optical encoder, which severely limits its transferability.

ReDON Inference Visualization. Figure 12 visualizes the inference results of ReDON on advanced learning tasks, including image segmentation and PDE solving, demonstrating the high expressivity and inference quality of ReDON in real-world complex tasks beyond simple classification.

ReDON Robustness Study. Practical diffractive optical systems inevitably suffer from physical non-idealities such as diffractive layer misalignment, random perturbations on the readout due to input/sensor noises, and fabrication-induced metasurface phase response deviations. We evaluate robustness on Fashion-MNIST classification using a 5-layer ReDON system *without recurrence* to isolate the physical sensitivity of the optical front-end. As illustrated in Fig. 13, we consider three non-ideality sources: (i) *misalignment*, modeled as a random lateral shift of each diffractive plane, where the shift direction (horizontal or vertical) and sign are randomly selected, and the shift magnitude is set to the specified misalignment level (in pixel units with pixel size s=400nm); (ii) *readout noise*, modeled as additive i.i.d. Gaussian noise $\mathcal{N}(0, \sigma^2)$ applied to the sensed intensities; and (iii) *fabrication error*, modeled as i.i.d. Gaussian phase perturbations $\delta\phi \sim \mathcal{N}(0, \sigma^2)$ added to each meta-atom phase response. For e-beam lithography fabricated metasurfaces, the average phase

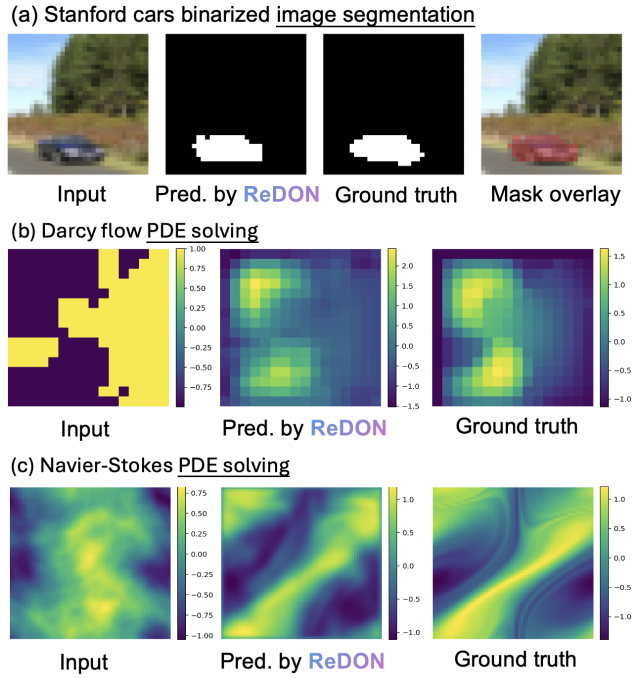

Fig. 12. Visualization of representative inference results of ReDON on (a) Stanford cars image segmentation, (b) Darcy flow, and (c) Navier-Stokes PDE solving tasks.

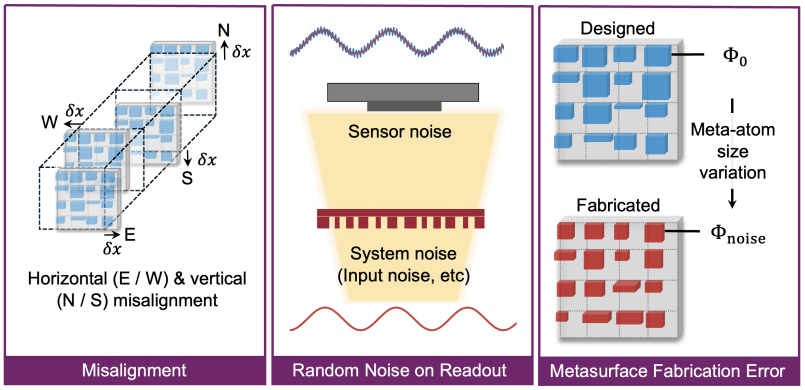

Fig. 13. Robustness evaluation settings under practical non-idealities. *Left*: random horizontal/vertical misalignment of diffractive planes up to ±1 pixel/meta-atom size. *Middle*: additive Gaussian noise with $\sigma = 0.01$ applied to sensed light intensities. *Right*: fabrication-induced phase error modeled as i.i.d. Gaussian phase perturbations with $\sigma_\phi = 0.01\ rad$ on metasurface phase responses.

error experimentally observed is typically around 5 to 10 degrees, which justifies the practicality of phase error injection strength later in Table 3.

Table 3 summarizes the resulting test accuracy under inference-time noise injection. For each noise setting in Table 3, we perform 5 independent evaluation iterations over the test set. Within an iteration, the injected noise realization is held fixed, while a new noise realization is resampled

Table 3. Robustness of a 5-layer ReDON system (no recurrence) on Fashion-MNIST under misalignment (MA), readout noise (RN), fabrication error (FE), and their combination. The noise-free test accuracy is 95.1%. The noise-aware model is trained with combined system errors (MA 2 + RN 0.1 + FE 0.8) injected during training; we report the average and std. of test accuracy over 5 iterations under inference-time noise injection.

Noise-injected inference (Standard Training)	Misalignment (MA)				
	Misaligned pixel size	0.5	1	1.5	2
	Averaged test accuracy	92.6%	91.3%	90.1%	88.2%
	Std. of test accuracy	0.0048	0.0051	0.0049	0.0053
	Readout Noise (RN)				
	Injected Gaussian noise (σ)	0.04	0.06	0.08	0.1
	Averaged test accuracy	93.1%	92.4%	91.2%	89.1%
	Std. of test accuracy	0.0029	0.0028	0.0029	0.0033
	Fabrication Error (FE)				
	Injected phase error $\sigma(\delta\phi)$	0.2	0.4	0.6	0.8
	Averaged test accuracy	90.3%	88.2%	82.3%	76.3%
	Std. of test accuracy	0.0046	0.0052	0.0054	0.0062
	Combined: MA + RN + FE				
	Combined system error	MA 0.5 + RN 0.04 + FE 0.2	MA 1 + RN 0.06 + FE 0.4	MA 1.5 + RN 0.08 + FE 0.6	MA 2 + RN 0.1 + FE 0.8
	Averaged test accuracy	88.1%	86.2%	79.1%	71.8%
	Std. of test accuracy	0.0052	0.0054	0.0051	0.0059

Noise-injected inference (**Noise-aware Training**)	Misalignment (MA)				
	Misaligned pixel size	0.5	1	1.5	2
	Averaged test accuracy	91.9%	91.7%	91.6%	91.7%
	Std. of test accuracy	0.0041	0.0049	0.0042	0.0044
	Readout Noise (RN)				
	Injected Gaussian noise (σ)	0.04	0.06	0.08	0.1
	Averaged test accuracy	91.1%	91.0%	92.7%	92.0%
	Std. of test accuracy	0.0028	0.0029	0.0031	0.0034
	Fabrication Error (FE)				
	Injected phase error $\sigma(\delta\phi)$	0.2	0.4	0.6	0.8
	Averaged test accuracy	91.8%	91.2%	91.3%	91.2%
	Std. of test accuracy	0.0051	0.0057	0.0055	0.0048
	Combined: MA + RN + FE				
	Combined system error	MA 0.5 + RN 0.04 + FE 0.2	MA 1 + RN 0.06 + FE 0.4	MA 1.5 + RN 0.08 + FE 0.6	MA 2 + RN 0.1 + FE 0.8
	Averaged test accuracy	91.6%	91.1%	91.5%	91.1%
	Std. of test accuracy	0.0049	0.0044	0.0051	0.0052

for the next iteration. Therefore, the evaluation is stochastic across iterations, and Table 3 reports the mean and standard deviation of accuracy over 5 iterations.

Under standard (noise-free) training, ReDON remains relatively resilient to moderate MA and RN, but becomes increasingly sensitive to FE, and especially to compound system-level errors. In the most severe combined setting (MA 2 + RN 0.1 + FE 0.8), accuracy drops substantially. To mitigate this, we perform noise-aware training by injecting the combined system noise (MA 2 + RN 0.1 + FE 0.8) during training to encourage a more robust model. This noise-aware model slightly reduces noise-free accuracy (92.9%), but markedly improves robustness, maintaining 91-92% accuracy across the tested ranges of MA/RN/FE and recovering most of the accuracy under harsh combined-noise conditions.

5 Conclusion

We present ReDON, a recurrent diffractive optical neural processor that overcomes the weak nonlinearity and rigidity of conventional DONNs through a self-modulated electro-optic nonlinearity. By sensing a small portion of the optical field and dynamically modulating downstream metasurfaces, ReDON provides a strong, tunable, input-dependent nonlinear response with minimal overhead while preserving the speed and parallelism of optical in-memory computing. Through extensive design-space exploration, we show that ReDON consistently delivers far higher nonlinear expressivity with balanced efficiency compared to prior nonlinear mechanisms. This makes ReDON a promising candidate for adaptive front-end encoders in optical machine learning systems. It establishes a new direction toward reconfigurable, recurrent, nonlinear diffractive in-memory computing.

Acknowledgments

R. Sarma and Y. Yao acknowledge support from the Laboratory Directed Research and Development program at Sandia National Laboratories and Sandia University Partnerships Network (SUPN) program. This work was performed in part at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. DOE's National Nuclear Security Administration under Contract No. DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. DOE or the United States Government.

References

[1] Loubnan Abou-Hamdan, Emil Marinov, Peter Wiecha, Philipp del Hougne, Tianyu Wang, and Patrice Genevet. 2025. Programmable metasurfaces for future photonic artificial intelligence. 7 (2025), 331–347. doi:10.1038/s42254-025-00721-3 Perspective article.

[2] Thomas Becker, Thomas G Bifano, Hocheol Lee, Michele Miller, Paul A Bierden, and Steven Cornelissen. 2003. MEMS spatial light modulators with integrated electronics. In *MOEMS and Miniaturized Systems III*, Vol. 4983. SPIE, 193–203.

[3] Ileana-Cristina Benea-Chelmus, Maryna L. Meretska, Delwin L. Elder, Michele Tamagnone, Larry R. Dalton, and Federico Capasso. 2021. Electro-optic spatial light modulator from an engineered organic layer. *Nature Communications* 12, 1 (11 Oct 2021), 5928. doi:10.1038/s41467-021-26035-y

[4] Pietro Caragiulo, Oscar Elisio Mattia, Amin Arbabian, and Boris Murmann. 2020. A Compact 14 GS/s 8-Bit Switched-Capacitor DAC in 16 nm FinFET CMOS. In *2020 IEEE Symposium on VLSI Circuits*. 1–2. doi:10.1109/VLSICircuits18222.2020.9162776

[5] Minho Choi and Arka Majumdar. 2025. Free-space optical encoder for computer vision. 2, 36 (2025). doi:10.1038/s44286-025-00036-3 Review, Open access.

[6] Minho Choi, Jinlin Xiang, Anna Wirth-Singh, Seung-Hwan Baek, Eli Shlizerman, and Arka Majumdar. 2025. Transferable polychromatic optical encoder for neural networks. 16, 5623 (2025). doi:10.1038/s41467-025-56192-0

[7] Yiying Dong, Bohan Zhang, Ruiqi Liang, Wenhe Jia, Kunpeng Chen, Junye Zou, Futai Hu, Sheng Liu, Xiaokai Li, and Yuanmu Yang. 2025. Scalable multilayer diffractive neural network with all-optical nonlinear activation. arXiv:2504.13518 [physics.optics] https://arxiv.org/abs/2504.13518

[8] Google. 2017. The Quick, Draw! Dataset. https://quickdraw.withgoogle.com/data

[9] Stephen Gould, Richard Fulton, and Daphne Koller. 2009. Decomposing a scene into geometric and semantically consistent regions. In *2009 IEEE 12th International Conference on Computer Vision*. 1–8. doi:10.1109/ICCV.2009.5459211

[10] Mark Horowitz. 2014. 1.1 Computing's energy problem (and what we can do about it). In *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. 10–14. doi:10.1109/ISSCC.2014.6757323

[11] Jingtian Hu, Deniz Mengu, Dimitrios C. Tzarouchis, Brian Edwards, Nader Engheta, and Aydogan Ozcan. 2024. Diffractive optical computing in free space. 15, 1525 (2024). doi:10.1038/s41467-024-45867-3 Perspective, Open access.

[12] Alexander Krasnok, Mykhailo Tymchenko, and Andrea Alù. 2017. Nonlinear metasurfaces: a paradigm shift in nonlinear optics. 21, 1 (2017), 8–21. doi:10.1016/j.mattod.2017.06.007 Research Review, Open access.

[13] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning Multiple Layers of Features from Tiny Images. (2009).

[14] Yingjie Li, Weilu Gao, and Cunxi Yu. 2023. Rubik's Optical Neural Networks: Multi-task Learning with Physics-aware Rotation Architecture. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI 2023)* (2023-08-19). International Joint Conferences on Artificial Intelligence Organization, 7197–7206. doi:10.24963/ijcai.2023/847

[15] Yuhang Li, Jingxi Li, and Aydogan Ozcan. 2024. Nonlinear encoding in diffractive information processing using linear optical materials. *Light: Science & Applications* 13, 1 (23 Jul 2024), 173. doi:10.1038/s41377-024-01529-8

[16] Xing Lin, Yair Rivenson, Nezih T. Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. 2018. All-optical machine learning using diffractive deep neural networks. 361, 6406 (2018), 1004–1008. doi:10.1126/science.aat8084

[17] Juzheng Liu, Mohsen Hassanpourghadi, and Mike Shuo-Wei Chen. 2022. A 10GS/s 8b 25fJ/c-s 2850um2 Two-Step Time-Domain ADC Using Delay-Tracking Pipelined-SAR TDC with 500fs Time Step in 14nm CMOS Technology. In *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, Vol. 65. 160–162. doi:10.1109/ISSCC42614.2022.9731625

[18] Xialin Liu, Boris Braverman, and Robert W. Boyd. 2023. Using an acousto-optic modulator as a fast spatial light modulator. 31, 2 (2023), 1501–1515. doi:10.1364/OE.471910

[19] Xuhao Luo, Yueqiang Hu, Xiangnian Ou, Xin Li, Jiajie Lai, Na Liu, Xinbin Cheng, Anlian Pan, and Huigao Duan. 2022. Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible. *Light: Science & Applications* 11, 1 (27 May 2022), 158. doi:10.1038/s41377-022-00844-2

[20] Arka Majumdar. 2025. Meta-optical encoders for high-resolution computer vision. In *Image Sensing Technologies: Materials, Devices, Systems, and Applications XII* (2025-05-28), Vol. PC13454. SPIE, Orlando, Florida, United States, PC134540J. doi:10.1117/12.3055225

[21] Cosmin-Constantin Popescu, Maarten Robbert Anton Peters, Oleg Maksimov, et al. 2025. 2D Addressable Mid-infrared Metasurface Spatial Light Modulator. (2025).

[22] Noam Shazeer. 2020. GLU Variants Improve Transformer. (2020).

[23] Simon Stich, Jewel Mohajan, Domenico de Ceglia, Luca Carletti, Hyunseung Jung, Nicholas Karl, Igal Brener, Alejandro W. Rodriguez, Mikhail A. Belkin, and Raktim Sarma. 2025. Inverse Design of an All-Dielectric Nonlinear Polaritonic Metasurface. 19, 18 (2025). doi:10.1021/acsnano.xxxxxx

[24] Simon Stich, Jewel Mohajan, Domenico de Ceglia, Luca Carletti, Jaeyeon Yu, Igal Brener, Alejandro W. Rodriguez, Mikhail A. Belkin, and Raktim Sarma. 2025. Inverse Design of an All-Dielectric Polaritonic Nonlinear Metasurface. In *CLEO 2025 Technical Digest Series*. Optica Publishing Group. doi:10.1364/CLEO_FS.2025.FF147_4 Paper FF147_4.

[25] Yingheng Tang, Ruiyang Chen, Minhan Lou, Jichao Fan, Cunxi Yu, Andrew Nonaka, Zhi Yao, and Weilu Gao. 2025. Optical neural engine for solving scientific partial differential equations. 16, 4603 (2025). doi:10.1038/s41467-025-54985-1 Open access.

[26] Texas Instruments. 2012. *DLP® Discovery™ 4100 Digital Controller (Rev. A).* https://datasheet.octopart.com/DLPC410ZYR-Texas-Instruments-datasheet-12093769.pdf Datasheet. Originally published Aug. 2012; revised Sept. 2012..

[27] Tianyu Wang, Mandar M. Sohoni, Logan G. Wright, Martin M. Stein, Shi-Yuan Ma, Tatsuhiro Onodera, Maxwell G. Anderson, and Peter L. McMahon. 2023. Image sensing with multilayer nonlinear optical neural networks. 17 (2023), 408–415. doi:10.1038/s41566-023-01090-4

[28] Fei Xia. 2025. On-chip programmable nonlinearity. 19 (2025), 662–663. doi:10.1038/s41566-025-01510-0 News & Views.

[29] Fei Xia, Kyungduk Kim, Yaniv Eliezer, SeungYun Han, Liam Shaughnessy, Sylvain Gigan, and Hui Cao. 2024. Nonlinear optical encoding enabled by recurrent linear scattering. *Nature Photonics* 18, 10 (01 Oct 2024), 1067–1075. doi:10.1038/s41566-024-01493-0

[30] Mustafa Yildirim, Niyazi Ulas Dinc, Ilker Oguz, Demetri Psaltis, and Christophe Moser. 2024. Nonlinear processing with linear optics. 18 (2024), 1076–1082. doi:10.1038/s41566-024-01391-3 Open access.

[31] Ziang Yin, Yu Yao, Jeff Zhang, and Jiaqi Gu. 2025. CHORD: Composable Hybrid Optical Reconfigurable Diffractive Framework For Optical Neural Network. In *ACM/IEEE Design Automation Conference (DAC)*. https://arxiv.org/abs/2411.05748

[32] Xiaoyun Yuan, Yong Wang, Zhihao Xu, Tiankuang Zhou, and Lu Fang. 2023. Training large-scale optoelectronic neural networks with dual-neuron optical-artificial learning. *Nature Communications* 14, 1 (04 Nov 2023), 7110. doi:10.1038/s41467-023-42984-y

[33] Fan Zhang, Wuyang Shi, Xixiao Li, Yigang Wang, Leilei Si, Wentao Gao, Meng Qi, Minjie Zhou, Jiajun Ma, Ao Li, Zhiqiang Li, Hongming Wang, and Bing Jin. 2025. Nonlinear Absorption Properties of Phthalocyanine-like Squaraine Dyes. *Photonics* 12, 8 (2025). doi:10.3390/photonics12080779

[34] Shanglin Zhou, Yingjie Li, Weilu Gao, Cunxi Yu, and Caiwen Ding. 2025. Automating multi-task learning on optical neural networks with weight sharing and physical rotation. 15, 14419 (2025). doi:10.1038/s41598-025-62222-1