OPTICS

End-to-end all-optical in-sensor computing system using photonic integrated circuits

Zian Xiao[1,2,3], Zhihao Ren[1,2], Yangyang Zhuge[1,2], Zixuan Zhang[1], Yan Yang[4,5]*, Bowei Dong[3]*, Chengkuo Lee[1,2]*

Photonic sensors play an increasingly important role in chemical analysis, but conventional systems rely on sequential wavelength scanning and extensive electronic post-processing, leading to large data redundancy, high latency, and excessive energy consumption. Here, we introduce an end-to-end all-optical in-sensor computing system based on photonic integrated circuits that merges sensing and computing in the optical domain. The system integrates a photonic waveguide sensor and a microring weight bank on a single chip for linear processing, while nonlinear activation is implemented using an erbium-doped fiber amplifier. By performing spectral compression, this system enables real-time optical-domain computing with up to 6-bit precision. The system achieves a classification accuracy of 94.2% across 27 liquid-mixture classes and concentration prediction for mixture chemicals. Compared with conventional photonic sensing systems, the proposed architecture reduces inference latency by a factor of 4.48 and energy consumption by a factor of 11.47. These results establish a compact, low-redundancy, and energy-efficient paradigm for intelligent photonic sensing at the edge.

INTRODUCTION

Photonic sensors have become essential tools in chemical, biological, and environmental analysis, where they enable high sensitivity, broad spectral coverage, and rapid signal acquisition (*1*–*3*). Their compatibility with integrated photonic platforms has further expanded their impact, allowing compact, stable, sensitive, and energy-efficient sensing systems far beyond those of electronic counterparts (*4*–*6*). As applications require faster responses and lower power consumption, photonic sensors are expected to play an even greater role in next-generation intelligent sensing technologies.

Despite these advantages, conventional photonic sensing architectures still typically follow a sensing-storage-then compute workflow as shown in Fig. 1A (*7*–*9*). In a standard implementation, a tunable laser injects light into the photonic sensor across different wavelengths, and a photodetector (PD) records the corresponding transmitted intensities to reconstruct the spectrum for subsequent electronic analysis. Machine-learning-assisted spectral analysis has been widely adopted to enhance feature extraction from such photonic sensor outputs (*10*, *11*). However, in most existing approaches, sensing, spectral readout, and computation remain physically and functionally decoupled, such that the measured optical information must first be converted into electronic signals and then transferred to downstream processors for digital inference. This architecture increases intermediate data handling and repeated optoelectronic processing, which in turn contributes to higher power consumption and additional latency. Moreover, optical-to-electrical conversion and off-chip data communication further reduce overall system efficiency.

To alleviate these bottlenecks, in-sensor computing has emerged as a promising paradigm that brings computation closer to, or directly into, the sensing front end (*12*–*15*). By performing partial data processing near or at the sensor, in-sensor computing reduces off-chip data movement and mitigates the redundancy inherent in raw spectral readout (*16*–*19*). Co-locating sensing and computation suppress redundant data generation at the source, thereby lowering power consumption and end-to-end latency while improving overall system efficiency

In parallel, recent advances in integrated photonic computing have enabled high-speed computing using microring resonator (MRR) based weight banks (*20*–*24*), crossbar array (*25*–*27*), reconfigurable waveguide meshes (*28*–*30*), cascaded Mach-Zehnder interferometer (MZI) arrays (*31*–*33*) and diffractive optics (*34*–*36*). Recent works have demonstrated end-to-end photonic neural network inference by implementing both linear transformations and nonlinear activations directly in the optical domain (*37*, *38*). It should be noted that this use of "end-to-end" refers to direct optical inference/computing, which is distinct from end-to-end optimization strategies that jointly optimize photonic structures and reconstruction algorithms at the system level (*39*). However, these photonic computing architectures are typically developed independently of the sensing front end, resulting in a physical and functional separation between sensing and computation. This separation limits the ability of photonic computing hardware to fully exploit the advantages of in-sensor computing.

To address this gap, photonic in-sensor computing has recently been explored to more tightly integrate photonic sensing and computation. Existing demonstrations typically combine photodetectors with analog or digital electronic circuits to perform in-sensor processing (*40*–*42*). More recently, photonic sensitizers have been integrated with photonic neural-network processors for multidimensional light-field sensing, including intensity, polarization, and wavelength decoupling (*43*). However, the monolithic integration of a photonic waveguide sensor with task-specific photonic computing for molecular absorption-based mixture analysis remains largely unexplored. As a result, optical signals cannot be transformed in a continuous, task-specific manner during acquisition, preventing truly end-to-end optical in-sensor computing.

[1]Department of Electrical & Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, Singapore. [2]National Centre for Advanced Integrated Photonics (NCAIP), National University of Singapore, Singapore 119077, Singapore. [3]Institute of Microelectronics (IME), Agency for Science, Technology and Research (A*STAR), 2 Fusionopolis Way, Innovis #08-02, Singapore 138634, Singapore. [4]Key Laboratory of Fabrication Technologies for Integrated Circuits, Chinese Academy of Sciences, Beijing 100029, China. [5]Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China.
*Corresponding author. Email: yyang10@ime.ac.cn (Y.Y.); dong_bowei@a-star.edu.sg (B.D.); elelc@nus.edu.sg (C.L.)

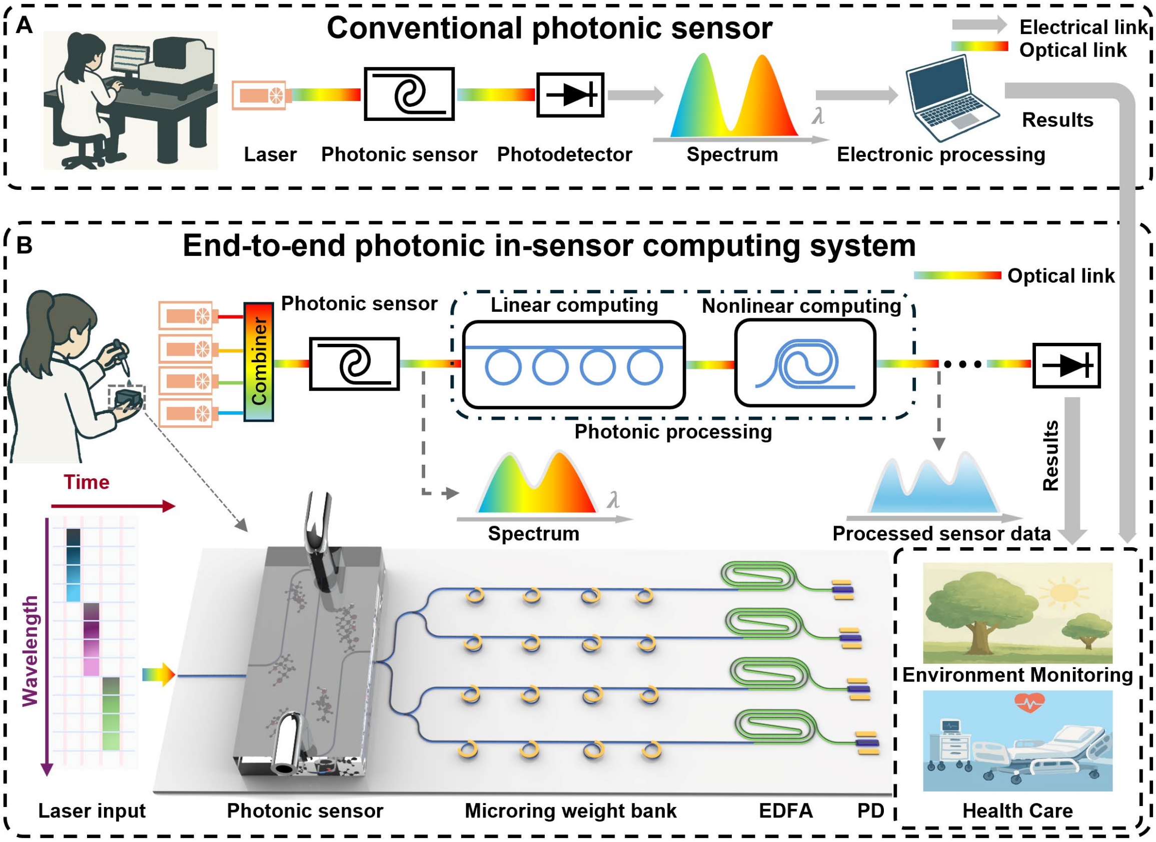

Fig. 1. Concept of end-to-end photonic in-sensor computing system. (**A**) Conventional photonic sensor workflow. A tunable laser illuminates the photonic sensor at a specific wavelength, where molecular absorption modulates the transmitted light. A PD records the corresponding intensity point-by-point, and the full spectrum is reconstructed by repeating this process across wavelengths. The obtained spectrum is then transferred to electronic processors for further computation and decision making. (**B**) End-to-end photonic in-sensor computing system. The proposed in-sensor computing chip integrates a photonic sensing layer with a programmable convolution layer, followed by an optical nonlinear activation function implemented using an EDFA. The PD collects the final output for downstream decision making. Together, these components form a fully optical computing pipeline that produces task-specific optical outputs directly during wavelength interrogation. The system can be used for environmental monitoring, healthcare, and industrial sensing.

Here, we introduce an end-to-end all-optical in-sensor computing system based on photonic integrated circuits, that integrates a photonic sensing front end with a MRR weight bank for linear optical processing as shown in Fig. 1B. By integrating a photonic sensor and microring weight bank on the same chip, this architecture reduces the intermediate conversion and data movement compared to traditional photonic sensor system. Rather than relying on the conventional workflow of full spectral readout followed by optoelectronic conversion and downstream digital processing, this system performs sensing and computation simultaneously. This architecture enables task-specific feature extraction and data compression directly during light–matter interaction. By effectively executing the neural-network inference in the optical domain before any electronic storage occurs, the system inherently reduces intermediate data movement and avoids redundant signal-processing steps. In this way, the proposed architecture highlights a sensor–computing co-design paradigm in which measurement and computation are directly coupled in the optical domain on the same chip. To further support end-to-end operation, nonlinear activation is implemented using an erbium-doped fiber amplifier (EDFA) as a proof-of-principle optical-domain component in the present system. This architecture achieves a classification accuracy of 94.2% across 27 classes of chemical mixtures and enables concentration prediction for the mixed analytes. Compared with conventional photonic sensing systems, the proposed end-to-end in-sensor computing system reduces latency by a factor of 4.48 and energy consumption by a factor of 11.47. By eliminating the traditional separation between measurement and computation, this all-optical end-to-end processing framework provides a compact and energy-efficient pathway toward real-time intelligent photonic sensing for next-generation edge-computing applications.

RESULTS

Detailed working principle

The detailed working principle of the end-to-end photonic in-sensor computing system is illustrated in Fig. 2. As shown in Fig. 2A, the laser module consists of four tunable lasers whose wavelength ranges are selected according to the sensing requirements. The outputs of the four lasers are combined into a single optical path through an optical combiner and injected into the chip. The multiplexed light propagates through the rib waveguide sensor, whose cross section is shown in fig. S1. The evanescent field of the guided mode interacts with the liquid mixture in the sensing chamber, and wavelength-dependent molecular absorption modulates the transmitted intensities through

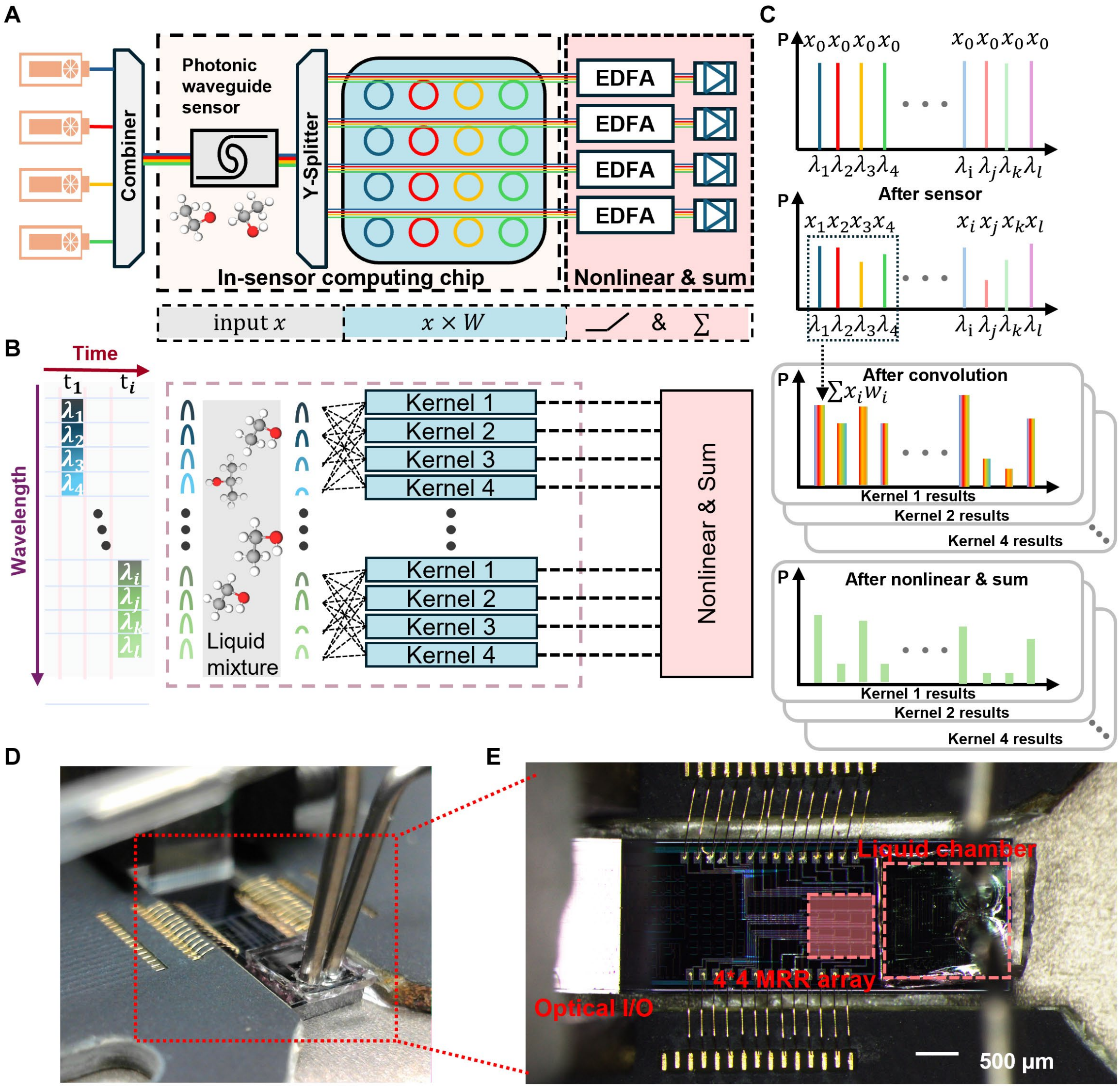

Fig. 2. Working principle of the end-to-end photonic in-sensor computing chip. (**A**) Implementation of the end-to-end in-sensor computing framework. The photonic sensor loads the optical signal input, after which a microring weight bank performs the linear convolution operation. The EDFA provides nonlinear activation through gain saturation, and the PD performs final readout and summation. (**B**) Operation within a single time slot. Four lasers emit light at four different wavelengths simultaneously. After propagating through the photonic sensor, the multiplexed signals enter the microring weight bank, where wavelength-dependent weighting is applied according to the pretrained neural network kernel. The signals then pass through the nonlinear activation and summation stages. (**C**) Evolution of signal intensity. Initially, each wavelength has an input intensity x_0. Molecular absorption in the sensing region modifies the transmitted intensity to x_i. The MRR-based convolution further transforms the intensity profile, and final nonlinear activation and summation are performed at the PD output. (**D**) Photograph of the packaged photonic in-sensor computing chip. (**E**) Zoom-in view of the chip region showing the sensing interface and integrated components.

vibrational overtone transitions in the near-infrared region (*44–47*). These intensity variations encode the spectral fingerprint of the analyte, enabling the photonic sensor to directly load the optical input signal x_i at wavelength λ_i for subsequent in-sensor computation. In our demonstration, four optical signals are generated that go into the following computing parts. Following the sensing stage, a microring weight bank performs the linear convolution operation. Each MRR provides a wavelength-specific transmission weight determined by the pretrained neural-network kernel. Since four wavelengths propagate in each optical path, four MRRs contribute to one kernel operation, and a total of 16 MRRs are tuned for the four rows in the network. Voltage control adjusts the resonance condition of each MRR, thereby implementing the linear weighted sum $\sum_i x_i w_i$. After the linear operation, the weighted optical signal passes through an EDFA. The gain saturation of the erbium ions introduces a nonlinear response, as characterized in fig. S2. Together, the EDFA and the PD provide nonlinear activation and summation, generating the final processed output.

To illustrate the time-domain operation of the system, Fig. 2B shows the optical inputs at successive time slots. At time slot 1, four wavelengths, $\lambda_1, \lambda_2, \lambda_3,$and λ_4, are injected into the chip. As these wavelengths propagate through the photonic waveguide sensor filled with a liquid mixture, wavelength-dependent molecular absorption modifies their optical intensities, with each wavelength experiencing a distinct attenuation. The sensor-modulated optical signals are then routed to the computing unit, where the microring weight bank is programmed with the corresponding pretrained kernel values, followed by nonlinear activation and summation. In the following time slot, the tunable lasers switch to a new wavelength set while the MRR weights are simultaneously updated according to the pretrained kernel. By scanning successive wavelength sets across the target spectral range, the time-multiplexed operation completes the in-sensor computation.

The evolution of optical intensity through the sensing and computing stages is illustrated in Fig. 2C. For the wavelength channels λ_1 to λ_4, each optical signal starts with an initial intensity x_0 and is first attenuated to x_1, x_2, x_3 and x_4 by wavelength-dependent absorption in the photonic sensor. The optical signals after sensing then undergo linear weighting and nonlinear transformation through the microring weight bank and the nonlinear activation stage. The microring weight bank performs the convolution operation, yielding an output of $\sum_{i=1}^{4} x_i w_i$, in which four spectral data points are compressed into a single output value. By employing four parallel rows of microring weight banks, the system simultaneously implements four convolution kernels, producing the final output intensities after nonlinear activation. At subsequent time slots, the system operates on successive wavelength sets $\lambda_i, \lambda_j, \lambda_k$, and λ_l, generating the corresponding sensor-modulated intensities x_i, x_j, x_k, and x_l. After the full wavelength scan is completed, the in-sensor computing chip outputs the final computing results, thereby completing the end-to-end in-sensor computation.

A photograph of the packaged photonic in-sensor computing chip is shown in Fig. 2D. A polydimethylsiloxane (PDMS) chamber is bonded onto the chip and serves as the liquid chamber. The liquid chamber has a length of 2.1 mm, and the fabrication steps are detailed in fig. S3. A zoomed-in view of the device is presented in Fig. 2E, showing the monolithic integration of the photonic sensor and the microring weight bank on the same chip. Wire bonding is employed to supply external voltages for MRR tuning. A fiber array is used as the optical input/output interface to deliver the optical signals and collect the processed outputs.

The computing module of the in-sensor computing chip is fabricated using a standard 220-nm silicon-on-insulator (SOI) photonic integration process. A dry etching process is used to define the window region, followed by wet etching to ensure complete exposure of the silicon waveguide.

Sensor performance

After introducing the working principle, experiments were conducted to evaluate the feasibility of the proposed in-sensor computing chip for detecting different organic analytes in water using the photonic in-sensor computing system. Isopropyl alcohol (IPA), acetone, and ethanol were selected as representative targets because they are widely used industrial solvents whose high volatility and flammability pose substantial safety risks. Even at low concentrations, these solvents can rapidly generate hazardous vapor levels, making accurate and real-time detection essential for chemical safety and leak monitoring.

Photonic waveguide sensors can be broadly categorized based on whether they rely on changes in the real part (Δn) or the imaginary part (Δk) of the refractive index (*9*). In this work, Δk-based sensing mechanism is used, where the spectral response arises from the intrinsic absorption characteristics of the analytes interacting with the evanescent field of the waveguide. Figure 3A shows the chemical sensing setup, comprising a syringe pump for precise liquid delivery and a PD for optical signal acquisition. The sensor output is divided into two paths by an on-chip splitter, with one path used for signal monitoring to obtain training data for the neural network and the other directed to the computing module of the photonic in-sensor computing chip as shown in the fig. S4. The length of the waveguide sensor is 1.5 mm. The insertion loss of this waveguide sensor is 2.666 dB after cladding opening processing. Figure 3B presents the dynamic monitoring results of ethanol at different concentrations in water measured at a wavelength of 1550 nm. As the ethanol concentration increases, the transmitted optical intensity correspondingly increases, consistent with the expected absorption behavior of the liquid mixture. The time-trace signals exhibit excellent stability and repeatability, demonstrating the robustness of the photonic in-sensor computing chip under continuous operation. The dynamic responses of acetone and IPA measured at 1550 nm are provided in fig. S5.

Based on these real-time monitoring results and according to the Beer–Lambert law, the optical intensity after transmission can be described as

$$T = \frac{I_{analyte}}{I_{ref}} = \exp(-\alpha \Gamma \mathrm{LC}) \quad (1)$$

where T is the transmittance of the optical path, I_{analyte} represents the optical intensity measured in the presence of the analyte, and I_{ref} represents the reference intensity. Here I_{ref} taken as the intensity measured in pure water. Γ is the external confinement factor, α is the absorption coefficient of the analyte, and L is the physical length of the waveguide sensing region.

In our case, because the reference medium is water, we evaluate the absorbance difference between pure water and the analyte solution. The change in absorbance, ΔAbsorbance, as a function of analyte concentration is calculated as

$$\Delta Absorbance = \left(\alpha_{analyte} - \alpha_{water}\right)\Gamma L C_{analyte} \times \lg e = -\lg\left(\frac{I_{analyte}}{I_{water}}\right) \quad (2)$$

The extracted ΔAbsorbance values for the three analytes, together with their corresponding linear fits, are shown in Fig. 3, C to E. All three analytes exhibit excellent linearity with coefficients of determination $R^2 > 0.98$, confirming the good linear response of the photonic sensor part in in-sensor computing platform. Because acetone has a lower absorption than water at the measured wavelength, its fitted slope is negative, in contrast to the positive sensitivity observed for IPA and ethanol.

To evaluate the limit of detection (LoD), we characterize the noise level of the system under illumination. Based on the measured noise and the linear calibration curves, which is shown in fig. S6, the LoDs for IPA, acetone, and ethanol in water are estimated to be 6.6, 5.1, and 9.8 vol%. The details are shown in text S6. Building upon the successful validation of photonic sensor part in photonic in-sensor computing chips for single-analyte chemical sensing, we next investigate its capability for liquid-mixture analysis, which represents one of the most common sensing scenarios in practical environments. To emulate complex mixture conditions, we prepare ternary solutions composed of IPA, acetone, and ethanol in water, yielding a

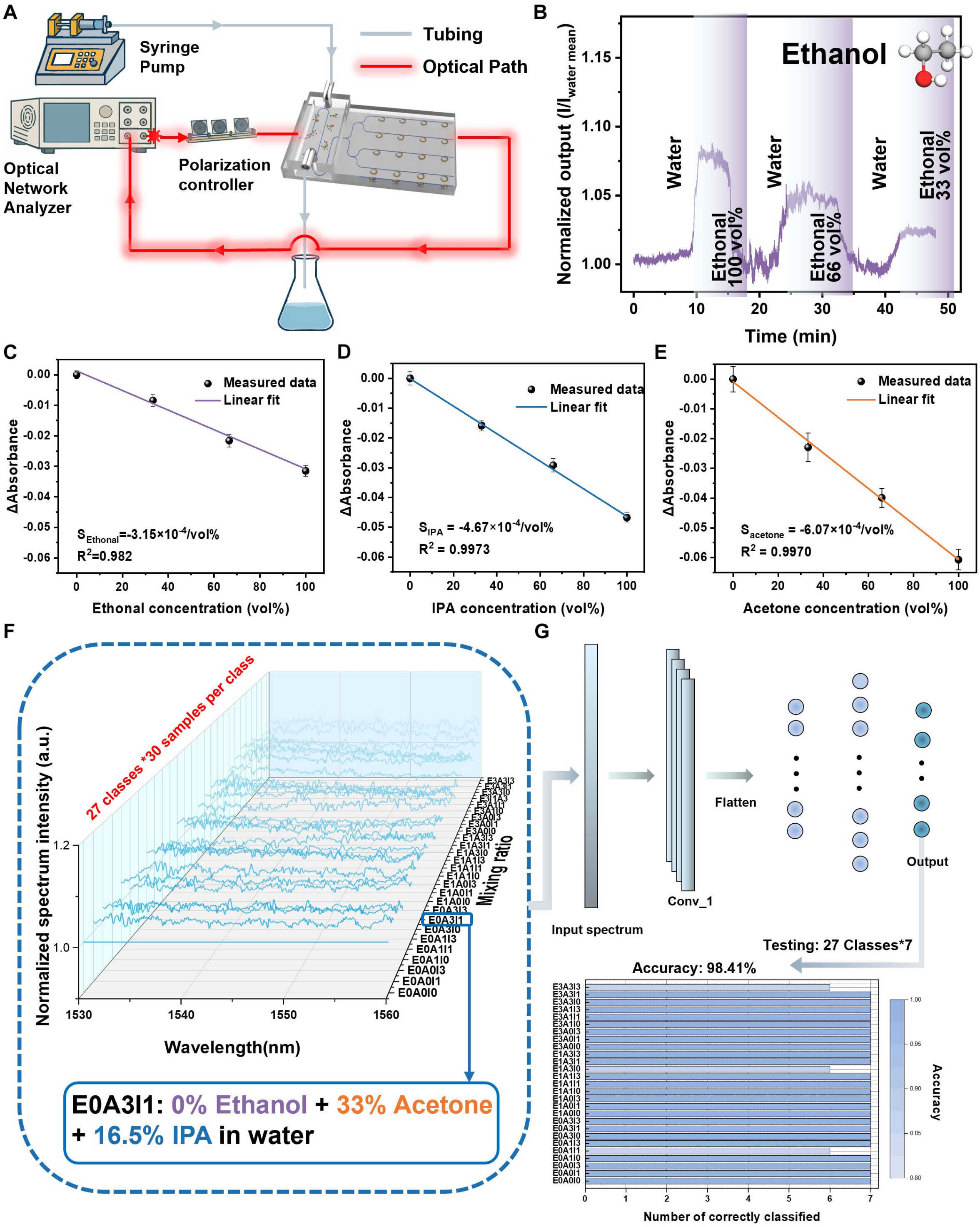

Fig. 3. Sensor performance. (**A**) Diagram of chemical sensing testing setup, comprising optical characterization and liquid flow control modules. (**B**) Real-time monitoring of ethanol solution under various concentrations (**C** to **E**) The absorbance difference between ethanol and IPA solution, acetone solution with respect to the analyte concentration. The sensitivity of ethanol, IPA, and acetone is -3.15×10^{-4}/vol%, -4.67×10^{-4}/vol%, and -6.07×10^{-4}/vol%, respectively. (**F**) 27 kinds of mixture spectrum visualization. Each mixing ratio randomly selects one spectrum as an example. (**G**) The construction of the CNN model and the bar chart of testing accuracy in recognizing mixture spectra with 27 mixing ratios. The classification accuracy is 98.41%.

total of 27 distinct mixing ratios. Each component was tested at three concentration levels: 0%, 16.5%, and 33%. The intermediate concentration of 16.5% was selected as approximately half of the maximum concentration, allowing uniform sampling of low, medium, and high concentration regimes. For each mixture, the tunable laser is swept from 1530 nm to 1560 nm, generating 30 spectra per mixture class. In total, 27×30 spectra are collected to construct the mixture-spectrum dataset. With a tuning step of 20 p.m., each spectrum contains 1501 wavelength-dependent data points. To better match the input dimension of the in-sensor computing chip, the spectral data were down-sampled to 41 wavelength points for machine-learning training. To minimize the influence of experimental instabilities, we additionally acquire reference spectra during pure-water injection between successive mixture measurements. This procedure helps suppress slow-varying background noise and enhances the reliability of the dataset. Each mixture spectrum is normalized using the nearest-in-time water reference spectrum to suppress slow-varying experimental drift. For visualization, one representative spectrum from each mixture class is randomly selected, as shown in Fig. 3F.

A major challenge in distinguishing these mixtures arises from the high chemical similarity among IPA, ethanol, and acetone. These analytes share similar functional groups and exhibit closely overlapping absorption features within the measured wavelength range, making reliable discrimination particularly difficult. To address these challenges, a convolutional neural network (CNN) with strong pattern-recognition capability is introduced for mixture-spectrum classification. The architecture of the constructed CNN is shown in Fig. 3G, consisting of three convolution layers followed by two fully connected layers. For each mixture class, 21 spectra are used for training, 2 for validation, and the remaining 7 spectra for testing. CNN can effectively learn multi-dimensional spectral information that is not accessible in single-wavelength sensing. The per-class testing accuracy is summarized in Fig. 3G, and an overall accuracy of 98.41% is achieved using a digital CNN on a PC for distinguishing 27 mixture classes, demonstrating the strong capability of the sensor parts in this end-to-end in-sensor computing system framework for complex liquid-mixture analysis, further confirming the robustness of the trained model and the consistency of the on-chip in-sensor computing.

Computing performance

After validating the performance of the sensing part, the computing part of the in-sensor computing chip is also characterized. The computing module occupies a compact footprint of 2.6 mm by 2.0 mm and incorporates a 4 by 4 microring weight bank that forms the core of the in-sensor computing system. Each MRR is thermally tunable and integrated with TiN microheaters, serving as the computational control element that enables precise resonance tuning for accurate linear weighting. The insertion loss of the MRR weight bank is 1.03 dB, which is shown in fig. S7.

Precise control of microring weight bank is achieved using a multiport voltage driver equipped with a high-resolution 16-bit digital-to-analog converter, allowing fine tuning of programmable voltage. To ensure operational stability, the chip is mounted on a thermoelectric cooler (TEC) to suppress thermal drift during computation. A temperature sensor attached to the TEC provides real-time temperature feedback, and a closed-loop temperature controller is employed to maintain a stable operating temperature. Figure 4A shows the transmission spectrum of a representative MRR. As the heater voltage increases, due to the thermo-optic effect, the refractive index of waveguide is changed. So, the resonance condition of the MRR changes, causing its resonant wavelength to shift accordingly. The quantitative relationship between the applied voltage and the resonance wavelength shift is provided in fig. S8.

To perform multiplexed operations, when a fixed-wavelength laser with input intensity x_i passes through the MRR, tuning the heater voltage modulates the optical transmission to a weight w_i at that wavelength. This mechanism implements the optical equivalent of a multiplication operation between the input intensity $x_i$and the corresponding synaptic weight w_i. When four wavelengths are simultaneously injected, each wavelength is assigned to a dedicated MRR within the weight bank. By independently tuning the four MRRs, the transmission of each wavelength channel is weighted according to its corresponding synaptic value. As a result, the output of the weight bank realizes the linear weighted summation $\sum_i x_i w_i$, which constitutes the core linear operation of the photonic in-sensor computing architecture. The computational performance of the photonic computing module is validated through an optical multiplication experiment. As shown in Fig. 4B, the measured results closely match the exact theoretical outputs. The standard deviation of the residual error is 0.0118, corresponding to a computational precision of approximately 6.4 bits. A Gaussian fit of the error distribution yields a mean error of 0.0001, as illustrated in Fig. 4C, confirming the high fidelity of the MRR-based linear operations implemented in the in-sensor computing chip. To evaluate the convolutional computing capability of the photonic computing module, a proof-of-concept experiment was performed using the letter 'a' from the A*STAR logo as a standard test input. The overall experimental setup is illustrated in Fig. 4D. In this demonstration, a 4 by 4 MRR array was configured to load the weights of four 2 by 2 convolutional kernels. The input image, a 96 by 96 grayscale frame, was first flattened into a one-dimensional vector. To enable high-speed operation, a parallel data-loading strategy was employed where every four consecutive elements of the vector were grouped and encoded onto an electro-optic modulator (EOM) operating at 1.25 GHz. The detailed look-up table for the EOM is shown in fig. S9. This strategy allows the serialized data stream to be synchronously injected into the photonic computing module.

Within the photonic computing part, each MRR was programmed to represent a single weight element of the convolution kernel. The encoded input signals were routed through the pass-through port, where multiplication-and-accumulation (MAC) operations were executed along each row of the MRR array. The resulting convolution outputs were collected at the drop port and detected by a PD for downstream analysis.

Four Prewitt filters were implemented to evaluate how different convolutional kernels extract directional edge features. These filters extract horizontal and vertical edge features, enabling a clear evaluation of the photonic processor's response to distinct convolution operators. Figure 4E compares the ideal convolution results with the experimentally measured outputs, showing strong agreement. Figure 4F presents the corresponding edge maps generated by representative kernels, validating that the photonic computing module can successfully perform parallel convolutional operations within an optoelectronic system.

To evaluate the computing speed of the photonic computing module in the in-sensor computing system, throughput is used as the performance metric. Throughput is defined as the number of operations

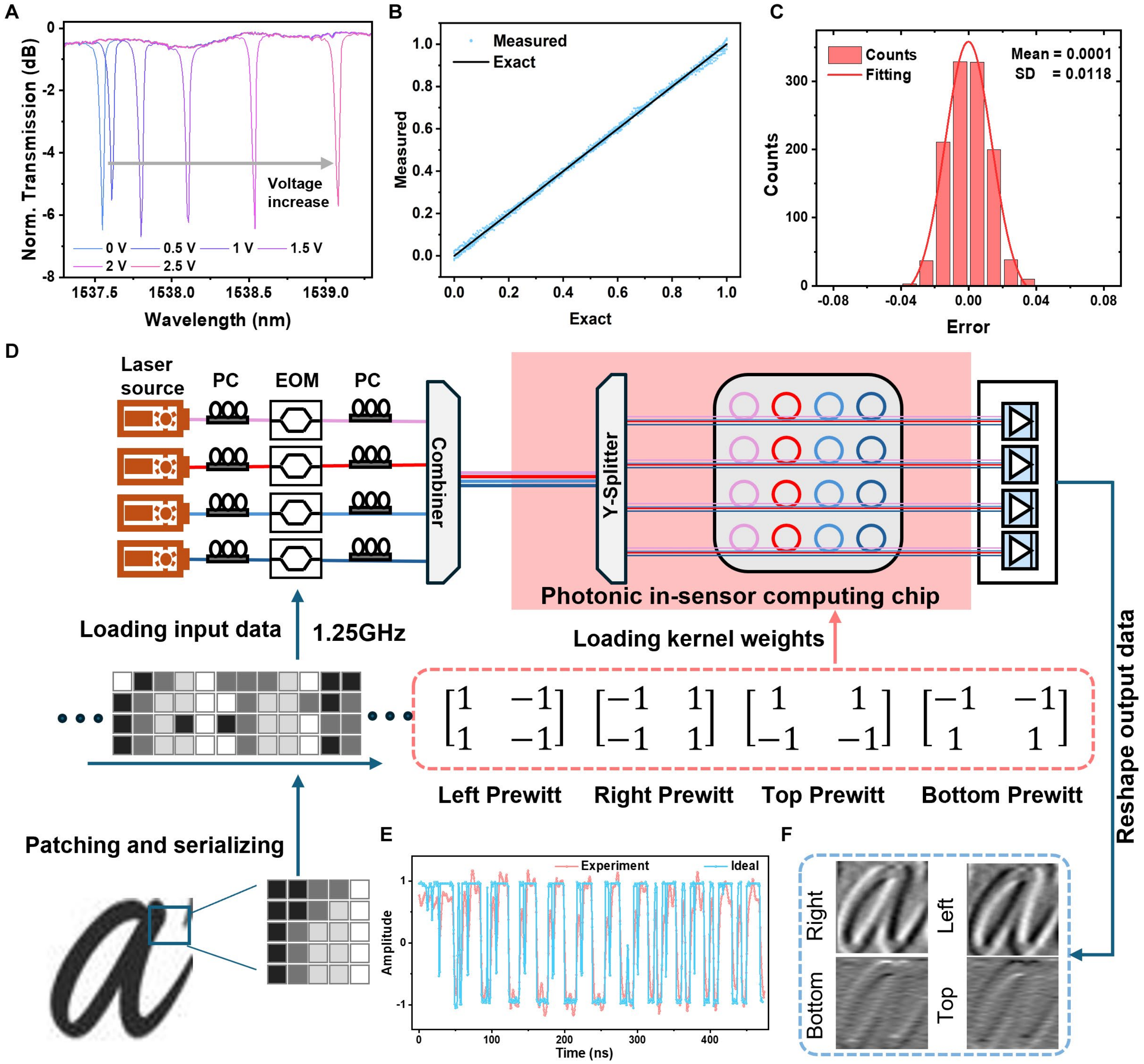

Fig. 4. Characterization of computing performance. (**A**) The transmission spectra of the MRR array. Different voltages are applied to the one MRR. Similar results can be obtained when the voltage is applied to other MRRs. (**B**) Measured scalar multiplication results versus exact results. (**C**) A histogram of computational error calculated by subtracting the measured scalar multiplication from the exact. The histogram is fitted by a Gaussian distribution (red solid curve). SD standard deviation. (**D**) Experimental setup of the photonic in-sensor computing chip for performing convolutional operations. (**E**) Experimental results during the image convolution result. (**F**) Corresponding convolution image results.

per second (OPS) performed by the processor. The throughput of photonic computing hardware can be calculated using Eq. 3 (*32*)

$$T(OPS) = 2m \times N^2 \times r \tag{3}$$

here, T represents the computational throughput in OPS, m is the number of computing layers implemented by the photonic hardware, N^2 represents the size of the on-chip weight bank, and r is the detection rate of the PD. In the proposed end-to-end photonic in-sensor computing system, MAC operations are naturally executed by the photonic hardware and are treated as fundamental computing operations. The detection rate $r = 1.25$GHz is obtained from real-time experimental operation of the end-to-end photonic in-sensor computing system, rather than a projected or theoretical value. With a 4 by 4 weight bank, the computing module achieves a throughput of 40 GOPS. To synchronize the four high-speed signal paths and compensate for their phase differences, a trigger pulse is inserted at the beginning of the data stream. To keep the weight stable and away from the thermal drift, the DC component of the PD output is used as a feedback signal for weight update via a calibration-free negative-feedback algorithm (*48*), while the AC component is simultaneously used for high-speed data acquisition and inference, as illustrated in text S10. This separation of DC and AC signals enables real-time hardware-level weight optimization without interrupting the computing process, allowing the photonic sensor and computing module to be trained and evaluated directly in the end-to-end system. To further assess long-term stability, a cross-day drift test is conducted to evaluate reliability. The system is monitored over 50,000 s, and the results are presented in fig. S11A. A Gaussian fit is applied to the measured data in fig. S11B.

In-sensor computing performance

Following the individual characterization of the photonic sensing and computing modules, we evaluated the complete end-to-end in-sensor computing system, in which spectral sensing, optical weighting, nonlinear activation, and final readout are performed in a synchronized workflow. Four lasers generate four distinct wavelengths, which are combined off-chip and coupled into the in-sensor computing chip via a fiber array and on-chip grating couplers. The photonic sensor, a 1 by 4 splitter, and a 4 by 4 MRR weight bank are monolithically integrated on the same chip. After coupling into the chip, the four wavelengths propagate through the photonic waveguide sensor, where molecular absorption modulates their intensity. The 1 by 4 splitter then distributes the signal evenly into four optical paths, which are subsequently processed by the microring weight bank. The weighted signals are routed to the nonlinear activation and summation stage to complete the in-sensor computing process.

The overall propagation and component loss of the end-to-end photonic in-sensor computing chip is estimated to be approximately 16.8 dB, excluding the splitter loss. Including the ideal 6 dB splitting loss from the two cascaded Y-splitters, the effective loss for each wavelength channel is approximately 22.8 dB. Under an input laser power of 10 dBm for each wavelength channel, the optical power delivered to the off-chip EDFA-based nonlinear activation module remains within its operating range. The detailed optical loss budget and output power estimation are provided in text S11. Figure 5A presents the measured transmission spectrum of a 10% IPA solution, providing a representative example of the optical signal before in-sensor computing. A wavelength range of 1530–1560 nm is employed, yielding 41 data points per spectrum. During operation, each laser is tuned by 0.75 nm between successive time slots, enabling sequential acquisition of the spectral components. The resulting outputs after convolution and nonlinear activation are shown in Fig. 5B. To improve both the computational accuracy and the energy efficiency of microring weight bank when applying the kernel values to the microring weight bank, a hardware-aware regularization term is introduced into the loss function during training (*49*), which can be expressed as

$$L = f\left(y, \hat{y}\right) + \alpha \parallel W - 1 \parallel^2 \tag{4}$$

where L is the defined loss function, $f\left(y, \hat{y}\right)$ is the cross-entropy loss, y is the neural network output, $\hat{y}$ is the target label. α is the regularization coefficient, and $\parallel W - 1 \parallel^2$ is the regularization to make the W as close to 1. This regularization encourages the trained weights to approach 1. In this operating region, the MRR transfer function exhibits minimal sensitivity to temperature-induced resonance drift, leading to significantly reduced weight fluctuations. Meanwhile, weights close to unity require nearly zero tuning power and provide a higher signal-to-noise ratio at the PD output. As a result, the proposed loss function simultaneously enhances robustness, reduces power consumption, and improves overall inference accuracy. The kernel values are illustrated in Fig. 5C.

These convolutional kernels are initially obtained through offline training based on the photonic sensor characterization data and are first loaded onto the microring weight bank. The look-up table for one ring in the microring weight bank is shown in Fig. 5D. Nonlinear activation is implemented using an EDFA, whose wavelength-dependent and power-dependent gain saturation provides a tunable nonlinear response. The extracted nonlinear activation function used in the neural network is shown in Fig. 5E, derived from the EDFA look-up table. Additional experiments were performed to demonstrate the importance of EDFA-induced nonlinearity in the in-sensor computing system. The corresponding training accuracy and loss curves are shown in Fig. 5F. With nonlinear activation enabled, the testing accuracy reaches 94.2%, whereas removing the nonlinear activation reduces the accuracy to approximately 88%. Incorporating optical nonlinearity not only improves classification accuracy but also accelerates convergence, highlighting its essential role in end-to-end in-sensor computing inference. The classification results for the liquid-mixture detection task are summarized in Fig. 5G. Finally, the trained neural network is applied to concentration prediction. Figure 5H presents the prediction results for IPA, while the corresponding results for acetone and ethanol are provided in fig. S13. Because the sensing and computing operations before the final readout are performed in the optical domain, intermediate optoelectronic conversion and digital post-processing are avoided. This architecture fully exploits the inherent advantages of photonic computing, including high computational speed and low power consumption. Compared with conventional photonic sensing systems, the proposed end-to-end in-sensor computing system reduces latency by a factor of 4.48 and energy consumption by a factor of 11.47. Detailed comparisons are provided in text S13.

End-to-end in-sensor computing system for industry application

Owing to its low power consumption and high data throughput, the proposed end-to-end photonic in-sensor computing system is well suited for industrial sensing applications, as illustrated in Fig. 6A. In industrial environments, accidental liquid leakage may occur, and it is often critical to rapidly determine whether the leaked substance is a hazardous chemical or a benign liquid such as rainwater. The photonic in-sensor computing system can be deployed for this purpose, as shown in Fig. 6B.

A simulated industrial environment was constructed to assess the practical applicability of the end-to-end photonic in-sensor computing system. Three representative chemical mixtures: 10% IPA, 50% IPA + 50% ethanol, and 100% ethanol, were used for real-time concentration prediction. The original optical spectra of these mixtures are shown in Fig. 6C. After processing by the photonic in-sensor computing system, which incorporates both linear convolution and nonlinear activation, the transformed data are presented in Fig. 6D. The corresponding concentration prediction results are shown in Fig. 6E. Real-time demonstrations of the system are provided in Movies S1 to S3.

Identification of blind samples under perturbations

The performance of the proposed system was further evaluated using blind samples consisting of mono, binary, and ternary mixtures

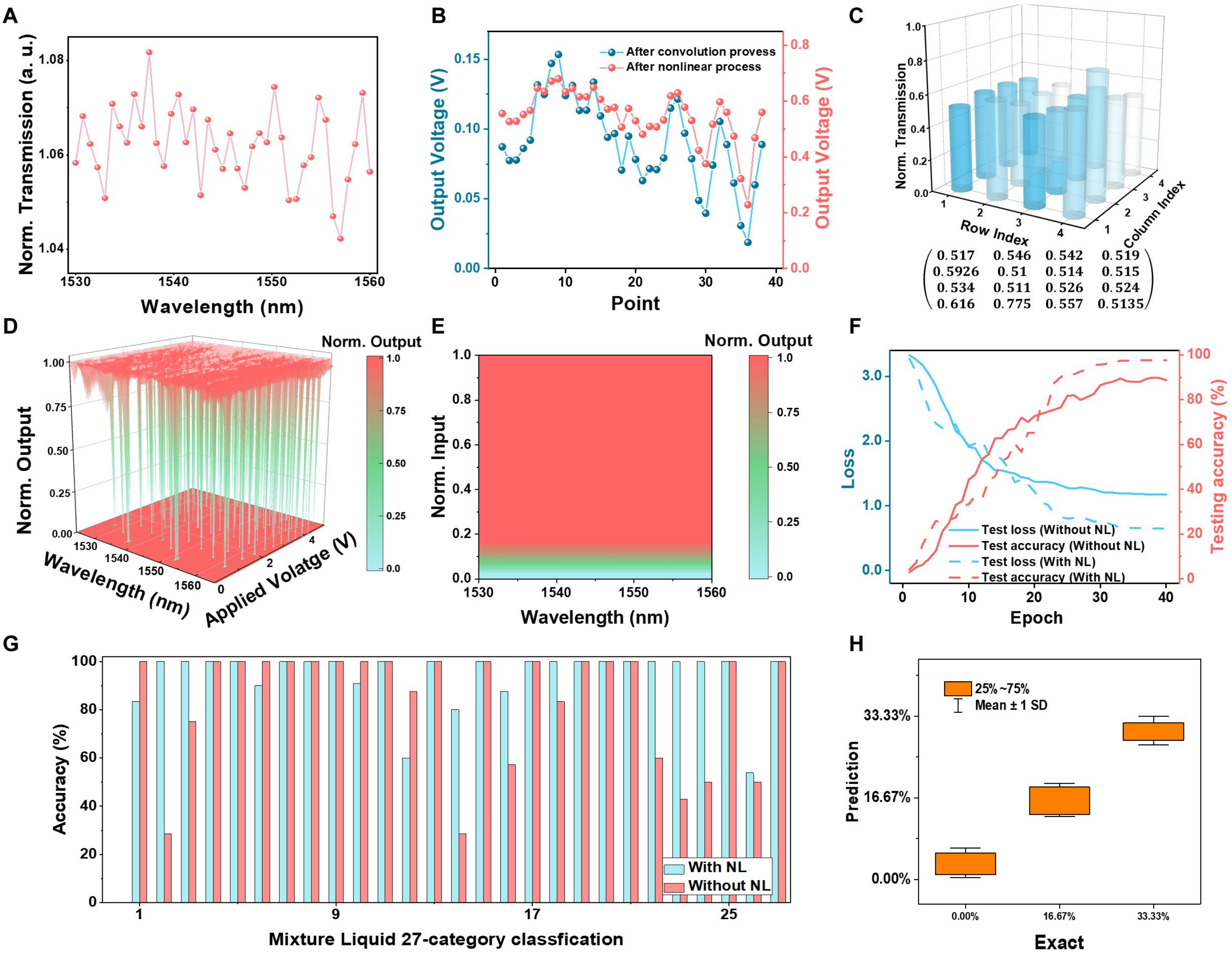

Fig. 5. In-sensor computing performance. (**A**) Spectrum for 10% IPA. (**B**) Data after convolution and nonlinear operation. (**C**) Pre-trained kernel. (**D**) 3D look-up table of MRR array. (**E**) 3D look-up table of EDFA. (**F**) Loss and testing accuracy with or without EDFA. With the EDFA, the testing accuracy can achieve 94.2%. (**G**) Classification accuracy with and without EDFA for different mixture classes. (**H**) Prediction of IPA through end-to-end in-sensor computing system.

of IPA, ethanol, and acetone at different concentration ratios. To simulate a perturbed sensing environment, tap water is used as background interference. These blind samples were not included in the training dataset. During the blind testing stage, a total of 27 samples were tested. Their ground-truth concentrations were only used after prediction to calculate the RMSE values. Therefore, all predictions were performed without prior knowledge of the ground-truth values. As shown in fig. S15, A and B, the system accurately identified all blind samples, demonstrating reliable discrimination capability for single-component analytes as well as binary and ternary mixtures under perturbed conditions. The calculated RMSE values for ethanol, acetone, and IPA were 2.00%, 1.32%, and 2.20%, respectively.

DISCUSSION

In summary, we demonstrate an end-to-end photonic in-sensor computing system that combines the photonic sensor part and photonic linear and nonlinear computing part entirely in the optical domain. The photonic sensor and photonic linear computing part is integrated in the same chip sharing the same CMOS fabrication. By performing task-specific feature extraction and spectral compression directly in the optical domain, this architecture reduces the redundancy, latency, and energy overheads inherent to conventional sequential sensing-storage-then compute workflows. Compared with conventional photonic sensing systems, the proposed end-to-end in-sensor computing system achieves a comparable classification accuracy of approximately 94%, while reducing inference latency by a factor of 4.48 and energy consumption by a factor of 11.47. Unlike prior photonic in-sensor computing approaches that are limited to linear operations or rely on optical–electrical–optical conversion to implement nonlinearity, our system realizes fully optical inference before the final photodetection readout synchronized with wavelength multiplexing. The experimental demonstrations in mixture classification and concentration prediction validate the effectiveness of nonlinear optical activation in

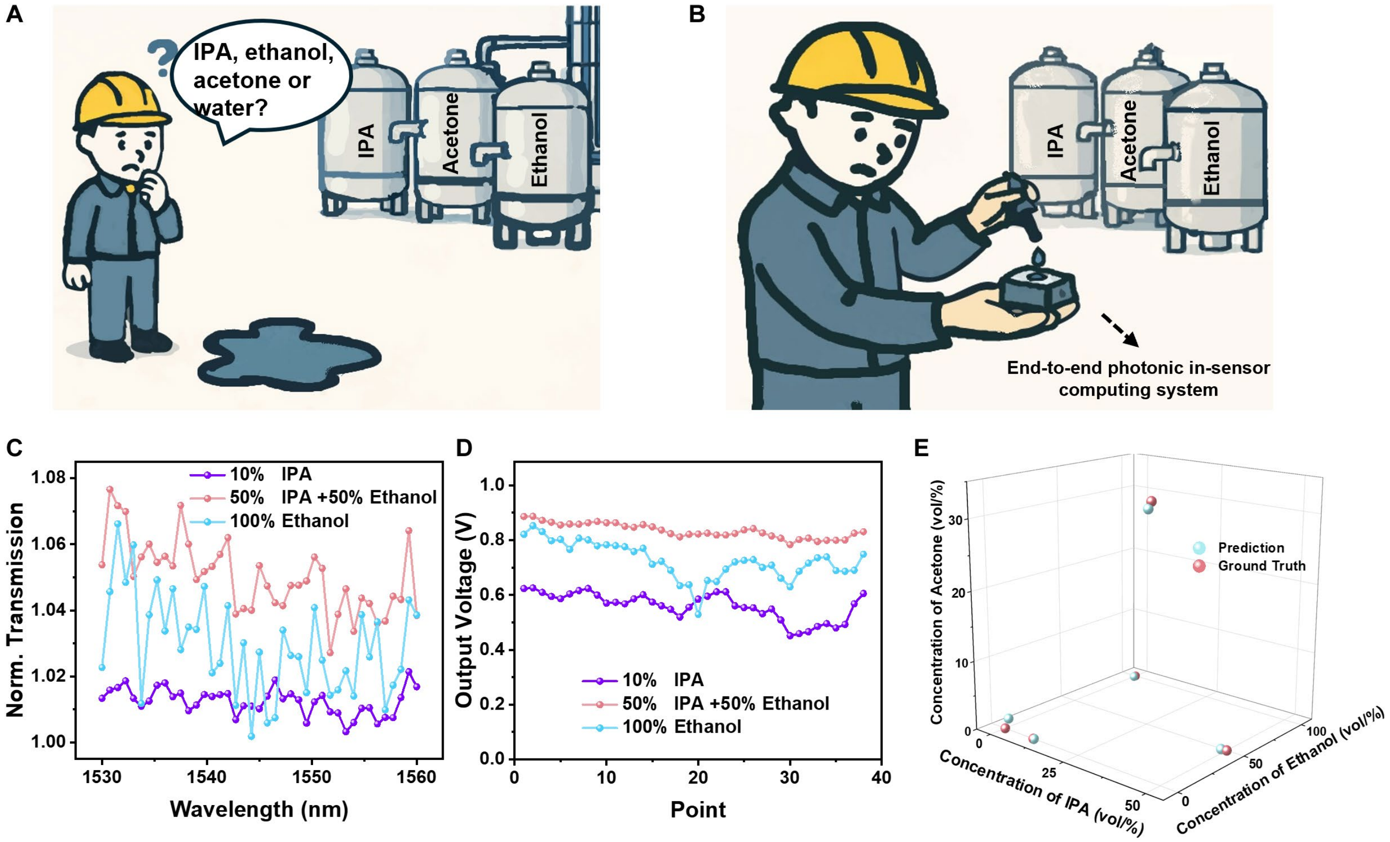

Fig. 6. Industrial chemical liquid concentration prediction using the end-to-end photonic in-sensor computing system. (**A**) Schematic illustration of a simulated chemical leakage scenario in an industrial environment, where the leaked liquid may be IPA, acetone, ethanol, or water. (**B**) Concentration prediction performed using the end-to-end photonic in-sensor computing system. (**C**) Original optical spectra of three representative liquid mixtures. (**D**) Processed data after end-to-end in-sensor computing. (**E**) Final concentration prediction results.

enhancing accuracy and convergence speed. This work establishes a scalable paradigm for end-to-end photonic intelligence, offering a general framework for low-power, high-throughput, and latency-critical chemical sensing and analysis at the sensor edge.

Looking forward, continued advances in photonic integrated circuits will enable additional functions in the proposed end-to-end photonic in-sensor computing system to be monolithically integrated on chip, further reducing reliance on off-chip components. For example, on-chip integration of the laser source and nonlinear elements could substantially reduce power consumption and communication latency compared with the current off-chip implementation (*50*, *51*), while also minimizing optical losses associated with fiber-to-chip coupling (*52*). Future optimization of the sensing waveguide may further improve system performance. For absorption-based sensing on SOI platforms, properly engineered TM-polarized strip waveguides can offer high sensing figures of merit by balancing modal overlap and propagation loss (*53*). Further, the MRR calibration and programming overhead could be further reduced in future implementations. For example, the broadband absorption spectrum could be spectrally separated by an arrayed waveguide grating and then weighted in parallel using electro-absorption modulators (*54*). Alternatively, a programmable waveshaper could directly impose wavelength-dependent attenuation on different spectral bins (*55*). In these architectures, the optical weights are defined by the fixed or programmed spectral transfer function rather than by sequentially aligning MRR resonances. Therefore, the weights can remain pre-defined for different spectral channels as the sensing wavelength varies, avoiding sequential ring tuning. In addition, more computing layers can be integrated to further enhance the computational capacity for more complicated tasks. With these developments, the proposed end-to-end photonic in-sensor computing system provides a scalable and effective solution for complex, multi-scenario sensing tasks in applications such as environment monitoring, healthcare and industrial sensing.

MATERIALS AND METHODS

Chip fabrication

The in-sensor computing chip was fabricated on a 220-nm silicon-on-insulator (SOI) wafer with a 2-μm buried oxide (BOX) layer. Silicon waveguides were patterned using standard lithography and reactive-ion etching (RIE), enabling both fully etched and partially etched structures to realize strip and rib geometries. TiN heaters were deposited and patterned to provide thermal tuning of the MRRs, followed by tungsten contacts and copper interconnects for electrical routing.

To expose the sensing region, the upper silica cladding above the photonic waveguide sensor was selectively removed using an anisotropic RIE step to define the sensing window, followed by an isotropic vapor-phase hydrogen fluoride (HF) etch to fully clear the oxide.

This process allows the guided optical mode's evanescent field to interact directly with the liquid analytes.

Experimental setup

Photonic sensor characterization

The photonic sensor characterization was conducted using a customized near-infrared (NIR) fiber-optic alignment system. NIR light from a tunable laser (Keysight 81960A) was coupled into NIR optical fibers (FS OS2). A 6-axis manual alignment stage (Kouzu GXM07S) was used to align the chip with the optical fiber arrays, ensuring efficient coupling of NIR light into the photonic waveguide and transmission to the output fiber. The output spectrum was measured by a power meter (Keysight 81636B), synchronized with the tunable laser source.

Photonic computing characterization

Photonic computing experiments were carried out using a tunable laser source (Cobrite DX) providing four independently tunable output channels. Each laser output was coupled into a near-infrared single-mode fiber (FS OS2) and passed through a polarization controller (Thorlabs FPC564) to ensure proper polarization alignment. The optical signals were then modulated by four electro-optic modulators (Optilab IM-1550-10-B), with one modulator assigned to each wavelength channel, followed by another polarization controller to make the polarization suitable for the grating coupler. The four modulated optical channels were subsequently combined using a fiber-based photonic combiner and coupled into the in-sensor computing chip via a fiber array for optical computation. After on-chip sensing and linear computing, the output signals were collected by a fiber array and detected using a photodetector (Newport 1811-FC) to measure the computing results.

Photonic in-sensor computing characterization

Photonic in-sensor computing experiments were performed using four tunable laser sources (Cobrite DX). Each laser output was coupled into near-infrared single-mode fibers (FS OS2) and passed through a polarization controller (Thorlabs FPC564) to ensure optimal polarization alignment. The four optical channels were subsequently routed through a variable optical attenuator (VOA) array to balance the input power levels before being combined by a photonic fiber-based combiner. The multiplexed signals were then coupled into the in-sensor computing chip via a fiber array. After on-chip sensing and computing, the output signals were collected by an output fiber array and amplified using an erbium-doped fiber amplifier (Thorlabs EDFA100P), followed by detection with a photodetector (Newport 1811-FC) to measure the output power.

Supplementary Materials

This PDF file includes:
Supplementary Text S1 to S14
Figs. S1 to S15
Tables S1 to S3
Legends for movies S1 to S3
References

Other Supplementary Material for this manuscript includes the following:
Movies S1 to S3

REFERENCES

1. H. Tang, B. Lou, F. Du, G. Gao, M. Zhang, X. Ni, E. Hu, A. Yacoby, Y. Cao, S. Fan, E. Mazur, An adaptive moiré sensor for spectro-polarimetric hyperimaging. *Nat. Photonics* **19**, 463–470 (2025).
2. Z. Yao, S. Liu, Y. Wang, X. Yuan, L. Fang, Integrated lithium niobate photonics for sub-ångström snapshot spectroscopy. *Nature* **646**, 567–575 (2025).
3. L. Ranno, Y. Z. Tan, C. S. Ong, X. Guo, K. N. Koo, X. Li, W. Wang, S. Serna, C. Liu, Rusli, C. G. Littlejohns, G. T. Reed, J. Hu, H. Wang, J. X. B. Sia, Crown ether decorated silicon photonics for safeguarding against lead poisoning. *Nat. Commun.* **15**, 3820 (2024).
4. Z. Li, J. Zou, H. Zhu, B. T. T. Nguyen, Y. Shi, P. Y. Liu, R. C. Bailey, J. Zhou, H. Wang, Z. Yang, Y. Jin, P. H. Yap, H. Cai, Y. Hao, A. Q. Liu, Biotoxoid Photonic Sensors with Temperature Insensitivity Using a Cascade of Ring Resonator and Mach–Zehnder Interferometer. *ACS Sens.* **5**, 2448–2456 (2020).
5. X. Liu, Z. Zhang, J. Zhou, W. Liu, G. Zhou, C. Lee, Development of Photonic In-Sensor Computing Based on a Mid-Infrared Silicon Waveguide Platform. *ACS Nano* **18**, 22938–22948 (2024).
6. H. Zhou, D. Li, Z. Ren, C. Xu, L.-F. Wang, C. Lee, Surface plasmons-phonons for mid-infrared hyperspectral imaging. *Sci. Adv.* **10**, eado3179 (2024).
7. F. Hsiao, C. Lee, Computational Study of Photonic Crystals Nano-Ring Resonator for Biochemical Sensing. *IEEE Sens. J.* **10**, 1185–1191 (2010).
8. W. Liu, Y. Ma, X. Liu, J. Zhou, C. Xu, B. Dong, C. Lee, Larger-Than-Unity External Optical Field Confinement Enabled by Metamaterial-Assisted Comb Waveguide for Ultrasensitive Long-Wave Infrared Gas Spectroscopy. *Nano Lett.* **22**, 6112–6120 (2022).
9. Y. Ma, B. Dong, C. Lee, Progress of infrared guided-wave nanophotonic sensors and devices. *Nano Converg.* **7**, 12 (2020).
10. J. Zhou, X. Liu, H. Zhou, S. Xu, J. Xie, C. Xu, W. Liu, Z. Zhang, C. Lee, Artificial-Intelligence-Enhanced Mid-infrared Lab-on-a-Chip for Mixture Spectroscopy Analysis. *Laser Photon. Rev.* **19**, 2400754 (2025).
11. Z. Li, H. Zhang, B. T. T. Nguyen, S. Luo, P. Y. Liu, J. Zou, Y. Shi, H. Cai, Z. Yang, Y. Jin, Y. Hao, Y. Zhang, A.-Q. Liu, Smart ring resonator–based sensor for multicomponent chemical analysis via machine learning. *Photonics Res.* **9**, B38 (2021).
12. Y. Chai, In-sensor computing for machine vision. *Nature* **579**, 32–33 (2020).
13. F. Zhou, Y. Chai, Near-sensor and in-sensor computing. *Nat. Electron.* **3**, 664–671 (2020).
14. Y. Yang, C. Pan, Y. Li, X. Yangdong, P. Wang, Z.-A. Li, S. Wang, W. Yu, G. Liu, B. Cheng, Z. Di, S.-J. Liang, F. Miao, In-sensor dynamic computing for intelligent machine vision. *Nat. Electron.* **7**, 225–233 (2024).
15. H. Huang, S. Shi, J. Zha, Y. Xia, H. Wang, P. Yang, L. Zheng, S. Xu, W. Wang, Y. Ren, Y. Wang, Y. Chen, H. P. Chan, J. C. Ho, Y. Chai, Z. Wang, C. Tan, In-sensor compressing via programmable optoelectronic sensors based on van der Waals heterostructures for intelligent machine vision. *Nat. Commun.* **16**, 3836 (2025).
16. J. Chen, Z. Zhou, B. J. Kim, Y. Zhou, Z. Wang, T. Wan, J. Yan, J. Kang, J.-H. Ahn, Y. Chai, Optoelectronic graded neurons for bioinspired in-sensor motion perception. *Nat. Nanotechnol.* **18**, 882–888 (2023).
17. B. Ouyang, J. Wang, G. Zeng, J. Yan, Y. Zhou, X. Jiang, B. Shao, Y. Chai, Bioinspired in-sensor spectral adaptation for perceiving spectrally distinctive features. *Nat. Electron.* **7**, 705–713 (2024).
18. N. T. Duong, Y. Shi, S. Li, Y.-C. Chien, H. Xiang, H. Zheng, P. Li, L. Li, Y. Wu, K.-W. Ang, Coupled Ferroelectric-Photonic Memory in a Retinomorphic Hardware for In-Sensor Computing. *Adv. Sci.* **11**, 2303447 (2024).
19. Z. Xiao, Z. Ren, Y. Zhuge, Z. Zhang, J. Zhou, S. Xu, C. Xu, B. Dong, C. Lee, Multimodal In-Sensor Computing System Using Integrated Silicon Photonic Convolutional Processor. *Adv. Sci.* **11**, e2408597 (2024).
20. S. Ohno, R. Tang, K. Toprasertpong, S. Takagi, M. Takenaka, Si Microring Resonator Crossbar Array for On-Chip Inference and Training of the Optical Neural Network. *ACS Photonics* **9**, 2614–2622 (2022).
21. B. Bai, Q. Yang, H. Shu, L. Chang, F. Yang, B. Shen, Z. Tao, J. Wang, S. Xu, W. Xie, W. Zou, W. Hu, J. E. Bowers, X. Wang, Microcomb-based integrated photonic processing unit. *Nat. Commun.* **14**, 66 (2023).
22. B. Zhao, J. Cheng, B. Wu, D. Gao, H. Zhou, J. Dong, Integrated photonic convolution acceleration core for wearable devices. *Opto-Electron. Sci.* **2**, 230017–230017 (2023).
23. C. Huang, S. Fujisawa, T. F. de Lima, A. N. Tait, E. C. Blow, Y. Tian, S. Bilodeau, A. Jha, F. Yaman, H.-T. Peng, H. G. Batshon, B. J. Shastri, Y. Inada, T. Wang, P. R. Prucnal, A silicon photonic–electronic neural network for fibre nonlinearity compensation. *Nat. Electron.* **4**, 837–844 (2021).
24. A. N. Tait, T. F. de Lima, M. A. Nahmias, B. J. Shastri, P. R. Prucnal, Multi-channel control for microring weight banks. *Opt. Express* **24**, 8895–8906 (2016).
25. B. Dong, F. Brückerhoff-Plückelmann, L. Meyer, J. Dijkstra, I. Bente, D. Wendland, A. Varri, S. Aggarwal, N. Farmakidis, M. Wang, G. Yang, J. S. Lee, Y. He, E. Gooskens, D.-L. Kwong, P. Bienstman, W. H. P. Pernice, H. Bhaskaran, Partial coherence enhances parallelized photonic computing. *Nature* **632**, 55–62 (2024).
26. B. Dong, S. Aggarwal, W. Zhou, U. E. Ali, N. Farmakidis, J. S. Lee, Y. He, X. Li, D.-L. Kwong, C. D. Wright, W. H. P. Pernice, H. Bhaskaran, Higher-dimensional processing using a photonic tensor core with continuous-time data. *Nat. Photonics* **17**, 1080–1088 (2023).
27. J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg,

W. H. P. Pernice, H. Bhaskaran, Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
28. D. Pérez-López, A. Gutierrez, D. Sánchez, A. López-Hernández, M. Gutierrez, E. Sánchez-Gomáriz, J. Fernández, A. Cruz, A. Quirós, Z. Xie, J. Benitez, N. Bekesi, A. Santomé, D. Pérez-Galacho, P. DasMahapatra, A. Macho, J. Capmany, General-purpose programmable photonic processor for advanced radiofrequency applications. *Nat. Commun.* **15**, 1563 (2024).
29. D. Pérez, I. Gasulla, L. Crudgington, D. J. Thomson, A. Z. Khokhar, K. Li, W. Cao, G. Z. Mashanovich, J. Capmany, Multipurpose silicon photonics signal processor core. *Nat. Commun.* **8**, 636 (2017).
30. D. P. Lopez, Programmable Integrated Silicon Photonics Waveguide Meshes: Optimized Designs and Control Algorithms. *IEEE J. Sel. Top. Quantum Electron.* **26**, 8301312 (2020).
31. H. Zhang, M. Gu, X. D. Jiang, J. Thompson, H. Cai, S. Paesani, R. Santagati, A. Laing, Y. Zhang, M. H. Yung, Y. Z. Shi, F. K. Muhammad, G. Q. Lo, X. S. Luo, B. Dong, D. L. Kwong, L. C. Kwek, A. Q. Liu, An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021).
32. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, M. Soljačić, Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
33. X.-K. Li, J.-X. Ma, X.-Y. Li, J.-J. Hu, C.-Y. Ding, F.-K. Han, X.-M. Guo, X. Tan, X.-M. Jin, High-efficiency reinforcement learning with hybrid architecture photonic integrated circuit. *Nat. Commun.* **15**, 1044 (2024).
34. Z. Xu, T. Zhou, M. Ma, C. Deng, Q. Dai, L. Fang, Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence. *Science* **384**, 202–209 (2024).
35. H. H. Zhu, J. Zou, H. Zhang, Y. Z. Shi, S. B. Luo, N. Wang, H. Cai, L. X. Wan, B. Wang, X. D. Jiang, J. Thompson, X. S. Luo, X. H. Zhou, L. M. Xiao, W. Huang, L. Patrick, M. Gu, L. C. Kwek, A. Q. Liu, Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **13**, 1044 (2022).
36. T. Fu, Y. Zang, Y. Huang, Z. Du, H. Huang, C. Hu, M. Chen, S. Yang, H. Chen, Photonic machine learning with on-chip diffractive optics. *Nat. Commun.* **14**, 70 (2023).
37. S. Zhang, X. Jiang, B. Wu, H. Zhou, W. Xu, H. Zhou, Z. Ruan, J. Dong, X. Zhang, Photonic edge intelligence chip for multi-modal sensing, inference and learning. *Nat. Commun.* **16**, 10136 (2025).
38. B. Wu, C. Huang, J. Zhang, H. Zhou, Y. Wang, J. Dong, X. Zhang, Scaling up for end-to-end on-chip photonic neural network inference. *Light Sci. Appl.* **14**, 328 (2025).
39. G. Arya, W. F. Li, C. Roques-Carmes, M. Soljačić, S. G. Johnson, Z. Lin, End-to-End Optimization of Metasurfaces for Imaging with Compressed Sensing. *ACS Photonics* **11**, 2077–2087 (2024).
40. B. Shao, T. Wan, F. Liao, B. J. Kim, J. Chen, J. Guo, S. Ma, J.-H. Ahn, Y. Chai, Highly Trustworthy In-Sensor Cryptography for Image Encryption and Authentication. *ACS Nano* **17**, 10291–10299 (2023).
41. J. Meng, T. Wang, H. Zhu, L. Ji, W. Bao, P. Zhou, L. Chen, Q.-Q. Sun, D. W. Zhang, Integrated In-Sensor Computing Optoelectronic Device for Environment-Adaptable Artificial Retina Perception Application. *Nano Lett.* **22**, 81–89 (2022).
42. Z. Zhang, X. Zhao, X. Zhang, X. Hou, X. Ma, S. Tang, Y. Zhang, G. Xu, Q. Liu, S. Long, In-sensor reservoir computing system for latent fingerprint recognition with deep ultraviolet photo-synapses and memristor array. *Nat. Commun.* **13**, 6590 (2022).
43. Z. Gu, Y. Shi, Z. Zhu, Z. Li, M. Zou, C. Yang, Y. Liu, Y. Yu, X. Zhang, All-integrated multidimensional optical sensing with a photonic neuromorphic processor. *Sci. Adv.* **11**, eadu7277 (2025).
44. A. Katiyi, A. Karabchevsky, Si Nanostrip Optical Waveguide for On-Chip Broadband Molecular Overtone Spectroscopy in Near-Infrared. *ACS Sens.* **3**, 618–623 (2018).
45. C. Yao, W. Zhang, P. Bao, J. Ma, W. Zhuo, M. Chen, Z. Shi, J. Zhou, Y. Ye, L. Ming, T. Yan, R. Penty, Q. Cheng, Chip-scale sensor for spectroscopic metrology. *Nat. Commun.* **15**, 10305 (2024).
46. W.-C. Lai, S. Chakravarty, X. Wang, C. Lin, R. T. Chen, On-chip methane sensing by near-IR absorption signatures in a photonic crystal slot waveguide. *Opt. Lett.* **36**, 984–986 (2011).
47. T. H. Stievater, M. W. Pruessner, D. Park, W. S. Rabinovich, R. Andrew McGill, D. A. Kozak, R. Furstenberg, S. A. Holmstrom, J. B. Khurgin, Trace gas absorption spectroscopy using functionalized microring resonators. *Opt. Lett.* **39**, 969–972 (2014).
48. S. Liu, T. Xu, B. Wang, D. Wang, Q. Xiao, L. Fan, C. Huang, Calibration-free and precise programming of large-scale ring resonator circuits. *Optica* **12**, 1113 (2025).
49. T. Xu, W. Zhang, J. Zhang, Z. Luo, Q. Xiao, B. Wang, M. Luo, X. Xu, B. J. Shastri, P. R. Prucnal, C. Huang, Control-free and efficient integrated photonic neural networks via hardware-aware training and pruning. *Optica* **11**, 1039–1049 (2024).
50. B. Zhao, B. Wu, H. Zhou, J. Dong, X. Zhang, Cascadable optical nonlinear activation function based on Ge–Si. *Opt. Lett.* **49**, 6149–6152 (2024).
51. Z. Tao, H. Wang, H. Feng, Y. Guo, B. Shen, D. Sun, Y. Tao, C. Han, Y. He, J. E. Bowers, H. Shu, C. Wang, X. Wang, Ultrabroadband on-chip photonics for full-spectrum wireless communications. *Nature* **645**, 80–87 (2025).
52. W. Deng, L. Chen, Z. Wang, G. Dong, Y. Yu, X. Wu, X. Zhang, "The design of terahertz two-dimensional grating coupler" in *International Photonics and OptoElectronics Meeting 2019* (OFDA, OEDI, ISST, PE, LST, TSA) (OSA, Washington, D.C., 2019), p. JW4A.60.
53. D. M. Kita, J. Michon, S. G. Johnson, J. Hu, Are slot and sub-wavelength grating waveguides better than strip waveguides for sensing? *Optica* **5**, 1046–1054 (2018).
54. H. Agarwal, B. Terrés, L. Orsini, A. Montanaro, V. Sorianello, M. Pantouvaki, K. Watanabe, T. Taniguchi, D. Van Thourhout, M. Romagnoli, F. H. L. Koppens, 2D-3D integration of hexagonal boron nitride and a high-κ dielectric for ultrafast graphene-based electro-absorption modulators. *Nat. Commun.* **12**, 1070 (2021).
55. W. Liu, S. Xu, C. Lee, Ultracompact on-chip spectral shaping using pixelated nano-opto-electro-mechanical gratings. *Science* **389**, 806–810 (2025).
56. Z. Xiao, Complete data and code (Zenodo, 2026). https://doi.org/10.5281/zenodo.20227459.
57. M. F. S. Ferreira, "Basic Concepts of Optical Amplifiers" in *Encyclopedia of Modern Optics* (Elsevier, 2005), pp. 351–354.
58. A. Karabchevsky, U. Sheintop, A. Katiyi, Overtone Spectroscopy for Sensing—Recent Advances and Perspectives. *ACS Sens.* **7**, 2797–2803 (2022).
59. J. Zhou, Z. Zhang, B. Dong, Z. Ren, W. Liu, C. Lee, Midinfrared Spectroscopic Analysis of Aqueous Mixtures Using Artificial-Intelligence-Enhanced Metamaterial Waveguide Sensing Platform. *ACS Nano* **17**, 711–724 (2023).
60. A. S. Raja, S. Lange, M. Karpov, K. Shi, X. Fu, R. Behrendt, D. Cletheroe, A. Lukashchuk, I. Haller, F. Karinou, B. Thomsen, K. Jozwik, J. Liu, P. Costa, T. J. Kippenberg, H. Ballani, Ultrafast optical circuit switching for data centers using integrated soliton microcombs. *Nat. Commun.* **12**, 5867 (2021).
61. F. Hu, F. Yu, X. Liu, A. Wang, X. Hu, H. Cai, W. Chu, 400-Gbps/λ Ultrafast Silicon Microring Modulator for Scalable Optical Compute Interconnects. *Laser Photon. Rev.* **20**, e02840 (2026).
62. S. Saeedi, A. Emami, "A 25Gb/s 170μW/Gb/s optical receiver in 28nm CMOS for chip-to-chip optical communication" in *2014 IEEE Radio Frequency Integrated Circuits Symposium* (IEEE, 2014), pp. 283–286.
63. N. Mehta, Z. Su, E. Timurdogan, J. Notaros, R. Wilcox, C. Poulton, C. Baiocco, N. Fahrenkopf, S. Kruger, T. Ngai, Y. Timalsina, M. Watts, V. Stojanovic, "An Optically Sampled ADC in 3D Integrated Silicon-Photonics/65nm CMOS" in *2020 IEEE Symposium on VLSI Technology* (IEEE, 2020), pp. 1–2.
64. Y. Kim, C. Ong, A. M. Pillai, H. Jagadeesh, G. Baek, I. Rajwani, Z. Guo, E. Karl, Energy-Efficient High Bandwidth 6T SRAM Design on Intel 4 CMOS Technology. *IEEE J. Solid-State Circuits* **58**, 1087–1093 (2023).
65. M. Bakhshalipour, H. Zare, F. Samandi, F. Golshan, P. Lotfi-Kamran, H. Sarbazi-Azad, "Blenda: Dynamically-Reconfigurable Stacked DRAM" in *2024 57th IEEE/ACM International Symposium on Microarchitecture (MICRO)* (IEEE, 2024), pp. 1323–1337.
66. H. Wang, R. Liu, R. Dorrance, D. Dasalukunte, D. Lake, B. Carlton, A Charge Domain SRAM Compute-in-Memory Macro With C-2C Ladder-Based 8-Bit MAC Unit in 22-nm FinFET Process for Edge Inference. *IEEE J. Solid-State Circuits* **58**, 1037–1050 (2023).
67. S. Chunarkar, S. R. Chiluveru, "Efficient Hardware Implementation of Nonlinear Activation Function For Inference Model" in *2024 21st International SoC Design Conference (ISOCC)* (IEEE, 2024), pp. 350–351.
68. C. St-Arnault, S. Bernal, R. Gutiérrez-Castrejn, E. Berikaa, W. Li, Z. Wei, Y. Hu, M. S. Alam, J. Rautert, S. V. Poltavtsev, A. E. Gubenko, V. V. Belykh, V. S. Mikhrin, A. R. Kovsh, D. V. Plant, Performance and Characterization Comparison of QD SOA, QW SOA, Bulk SOA and PDFA for Multi-Tbps O-Band WDM Links. *J. Light. Technol.* **43**, 1915–1925 (2025).
69. V. Sze, Y.-H. Chen, T.-J. Yang, J. S. Emer, *Efficient Processing of Deep Neural Networks* (Springer International Publishing, Cham, 2020), 10.1007/978-3-031-01766-7.
70. A. Maharry, H. Andrade, S. Misak, J. Liu, Y. Xia, A. Wissing, G. Movaghar, V. Arrunategui-Norvick, E. D. Chansky, X. Du, A. A. M. Saleh, J. F. Buckwalter, L. Coldren, C. L. Schow, Integrated SOAs enable energy-efficient intra-data center coherent links. *Opt. Express* **31**, 17480–17493 (2023).

Acknowledgments
Funding: This work was supported by: Ministry of Education (MOE), Singapore, through the National University of Singapore under grant T2EP50125–0030, awarded to C.L.; Agency for Science, Technology and Research (ASTAR), Singapore, through the National Centre for Advanced Integrated Photonics (NCAIP), Singapore, under grant NRF-MSG-2023-0002, awarded to C.L.; Agency for Science, Technology and Research (ASTAR), Singapore, through the National Semiconductor Translation and Innovation Centre (NSTIC), Singapore, under grant M24W1NS005, awarded to C.L.; National Research Foundation, Singapore (NRF), Singapore, under grants NRF-F-CRP-2024-0006 and NRF-NRFF17–2025-0003; Agency for Science, Technology and Research (ASTAR), Singapore, through the ASTAR Start-up Fund for the project "Hybrid Photonic-Electronic Computing Platform for Accelerating Artificial Intelligence," awarded to B.D.; Natural Science Foundation of China under grant 62274179, awarded to Y.Y. and International Partnership Program of Chinese Academy of Sciences (02GJHZ2023026FN) awarded to Y.Y. **Author contributions:** Conceptualization: Z.X., Z.R., Y.Y., B.D., C.L., Data Curation: Z.X., Y.Z., Formal Analysis: Z.X., Z.R., Z.Z., Y.Z., C.L., Funding Acquisition: Z.R., Y.Y., B.D., C.L., Investigation: Z.X., Z.R., Methodology: Z.X., Z.R., B.D. Z.Z., Y.Z., Project Administration: Z.R., B.D., C.L., Resources: Y.Y., B.D., C.L.,Software: Z.X. Z.R., Z.Z., Supervision: B.D., Y.Y., C.L., Validation: Z.X., Z.R., Y.Z. C.L., Visualization: Z.X., Z.R., C.L., Writing – original draft: Z.X., Z.R., B.D., Writing – review & editing: Z.X., Z.R., B.D., C.L. **Competing interests:** The authors declare that they have no competing interests. **Data, code and materials availability:** All data and code needed to evaluate and reproduce the results in the paper are present in the

paper and/or the Supplementary Materials. The source data supporting the findings of this study has been deposited in Zenodo (https://doi.org/10.5281/zenodo.20227459) (*56*) and GitHub repository at https://github.com/zianxiao/All-optical-end-to-end-in-sensor-computing-system. This study did not generate new materials.

Submitted 27 January 2026
Accepted 15 July 2026
Published 21 August 2026
10.1126/sciadv.aef8657

Science Advances

End-to-end all-optical in-sensor computing system using photonic integrated circuits

Zian Xiao, Zhihao Ren, Yangyang Zhuge, Zixuan Zhang, Yan Yang, Bowei Dong, and Chengkuo Lee

Sci. Adv. **12** (34), eaef8657. DOI: 10.1126/sciadv.aef8657

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.