

Supplementary Materials for

Inverse-designed subwavelength metamaterial 3×3 reconfigurable power distributor for high-density programmable photonic computing

Hangming Fan,^{1,†} Zhenyu Hu,^{1,†} Chengying Yu,¹ Xiaoyang Liu,¹ Xiao Ren,¹ Mengfan Cheng,^{1,2}
Qi Yang,¹ Ming Tang,¹ Deming Liu,¹ and Lei Deng^{1,2,*}

¹ *Huazhong University of Science and Technology, Wuhan National Laboratory for Optoelectronics and School of Optical and Electronic Information, Wuhan 430074, China*

² *Shenzhen Huazhong University of Science and Technology Research Institute, Shenzhen 518000, China*

[†] *These authors contributed equally: Hangming Fan, Zhenyu Hu*

^{*} *Corresponding author E-mail address: denglei_hust@mail.hust.edu.cn*

This PDF file includes:

Supplementary Text

Figs. S1 to S8

Note 1 to Note 3

Supplementary Text

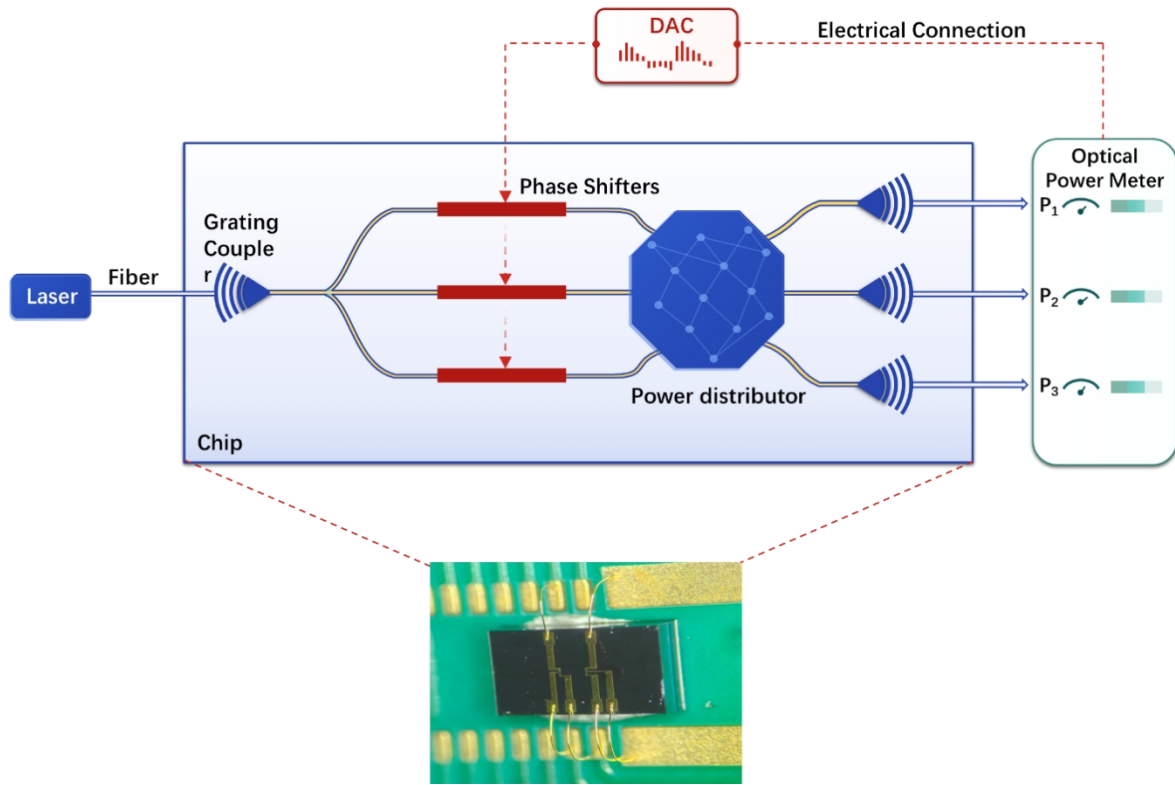


Fig. S1 Schematic of device measurement. a 1550 nm continuous - wave (CW) laser is utilized as the input light source. The light emitted from the laser traverses a PC prior to being coupled into a grating coupler. At the output end, the light is coupled into an optical fiber through another grating coupler and subsequently connected to an optical power meter for the measurement of optical power.

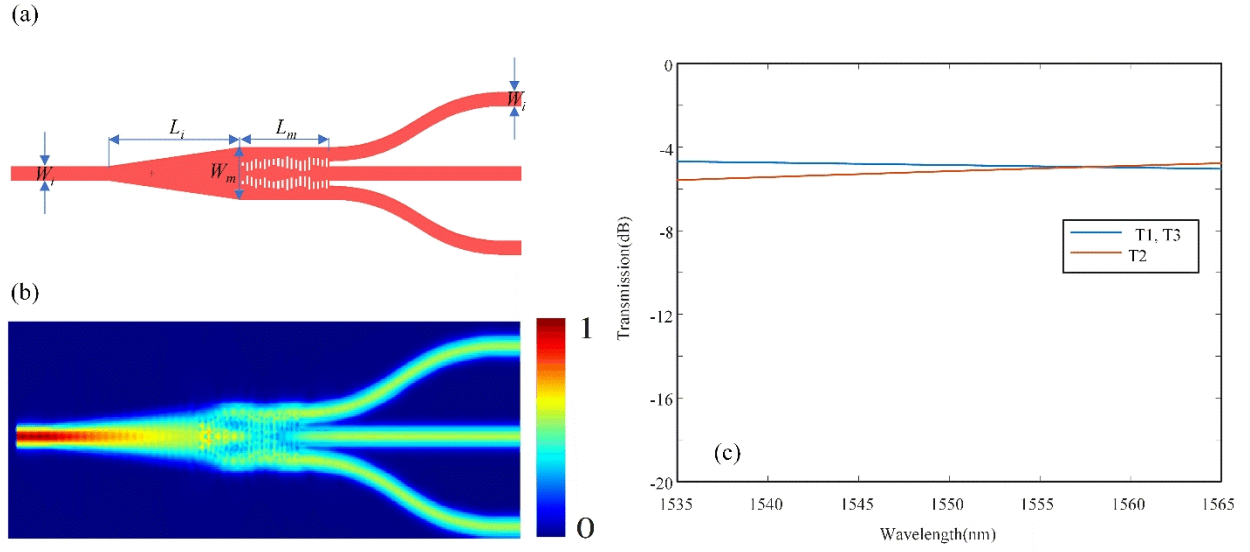


Fig. S2 Simulation of 1×3 equal power splitter. a, Structure of the 1×3 Splitter. b, Electric-field transmission map. c, Transmission spectrum of the 1×3 Splitter. The 1×3 equal-ratio power splitter equally divides the input light into three beams at the input port, which are then fed into the 3×3 power distributor. Simulation results reveal that nearly no extra loss is introduced. The phase difference originating from the splitter can be eliminated by tuning the voltages applied to the three electrodes.

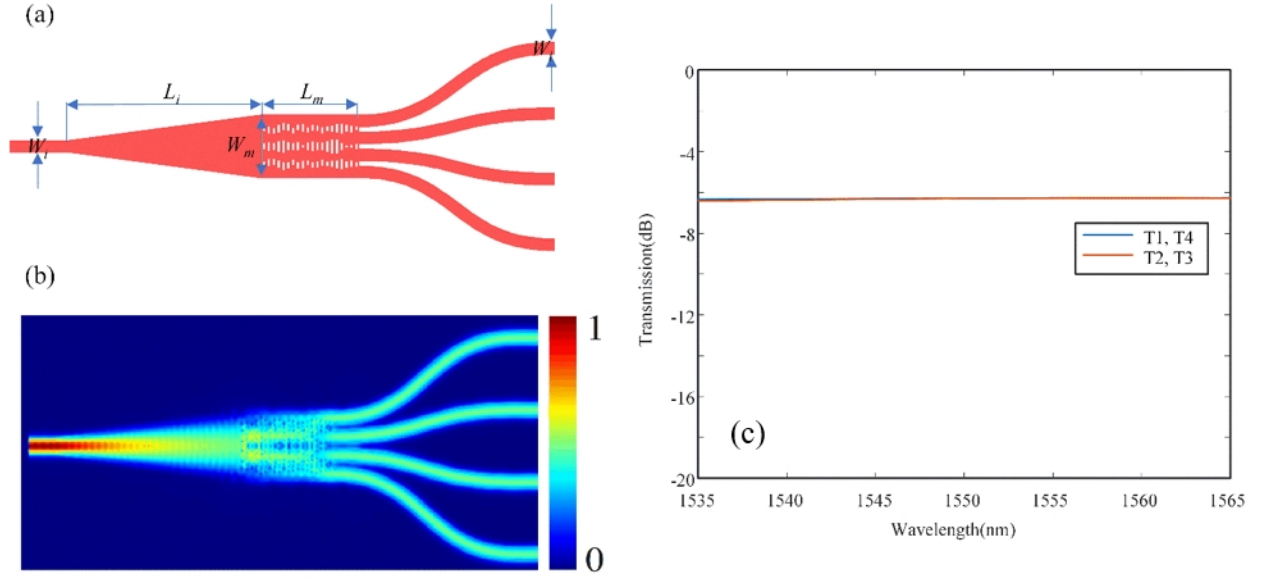


Fig. S3 Simulation of 1×4 equal power splitter. a, Structure of the 1×4 Splitter. b, Electric-field transmission map. c, Transmission spectrum of the 1×4 Splitter. The 1×4 equal-ratio power splitter equally divides the input light into four beams at the input port, which are then fed into the 4×4 power distributor. Simulation results reveal that nearly no extra loss is introduced.

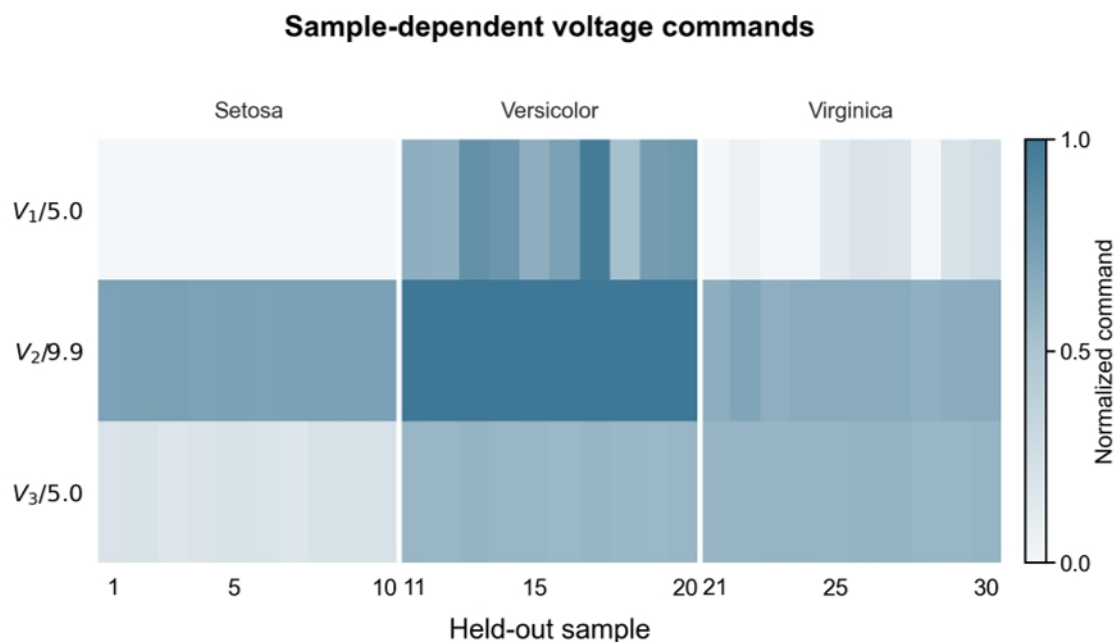


Fig. S4 Continuous voltage encoding of the fixed test set. Control-voltage triplets for 30 held-out Iris samples, grouped as Setosa, Versicolor and Virginica, with ten samples per species. Rows show V_1 , V_2 and V_3 divided by their calibrated upper limits of 5.0, 9.9 and 5.0 V, respectively. All 30 voltage triplets were distinct. These commands correspond to the separate fixed-split model, not the fivefold out-of-fold evaluation.

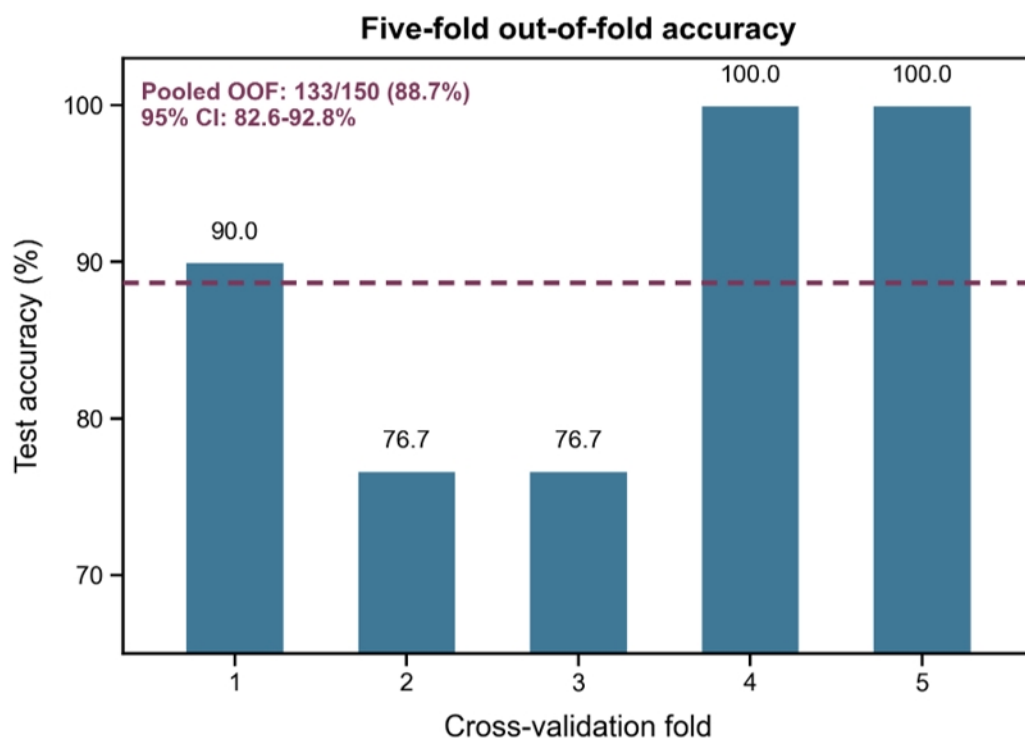


Fig. S5 Fold-resolved evaluation of the three-port classifier. Classification accuracy in each of five stratified folds, with 30 out-of-fold test samples per fold and ten samples per class. Each of the 150 Iris samples is evaluated once outside its training fold. Fold accuracies are 90.00%, 76.67%, 76.67%, 100% and 100%. The mean is 88.67%, with a sample standard deviation of 11.69 percentage points. Predictions are obtained from the calibrated hardware-response model.

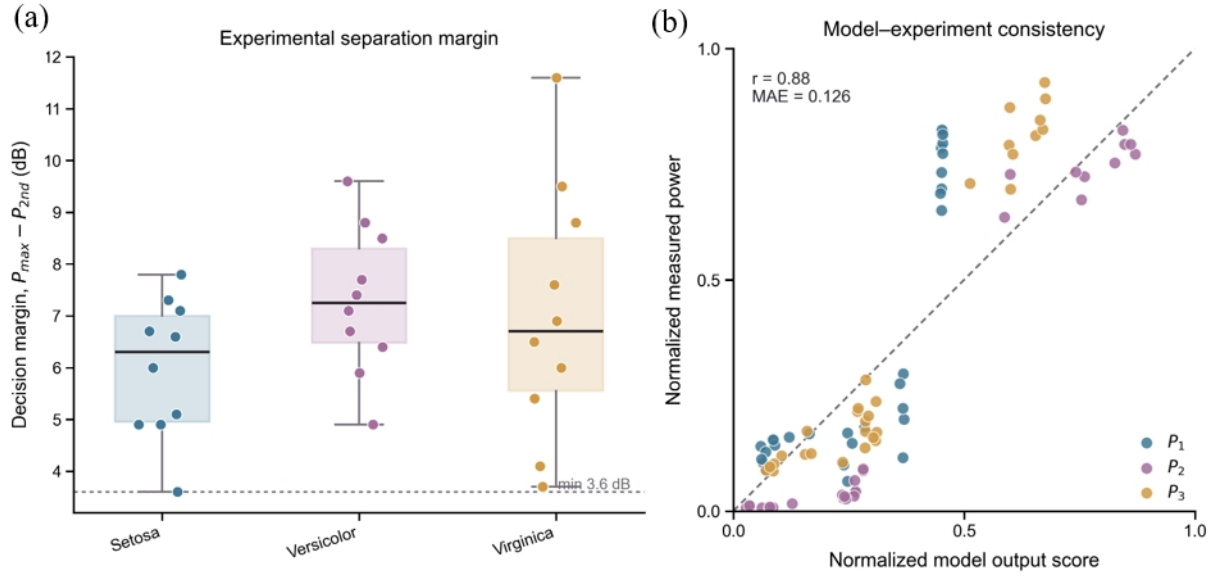


Fig. S6 Additional analysis of experimental three-port outputs. a, Decision margins for 30 measured samples, grouped by species. The margin is the difference in dB between the largest and second-largest output powers. Points represent individual samples; boxes indicate the median and interquartile range, with whiskers extending to the most extreme points within 1.5 times the interquartile range. b, Comparison of normalized offline-model scores and normalized measured power fractions for all three ports, giving 90 port-sample pairs from the same 30 samples. Pearson $r = 0.880$ and mean absolute error = 0.126. Each sample was measured once.

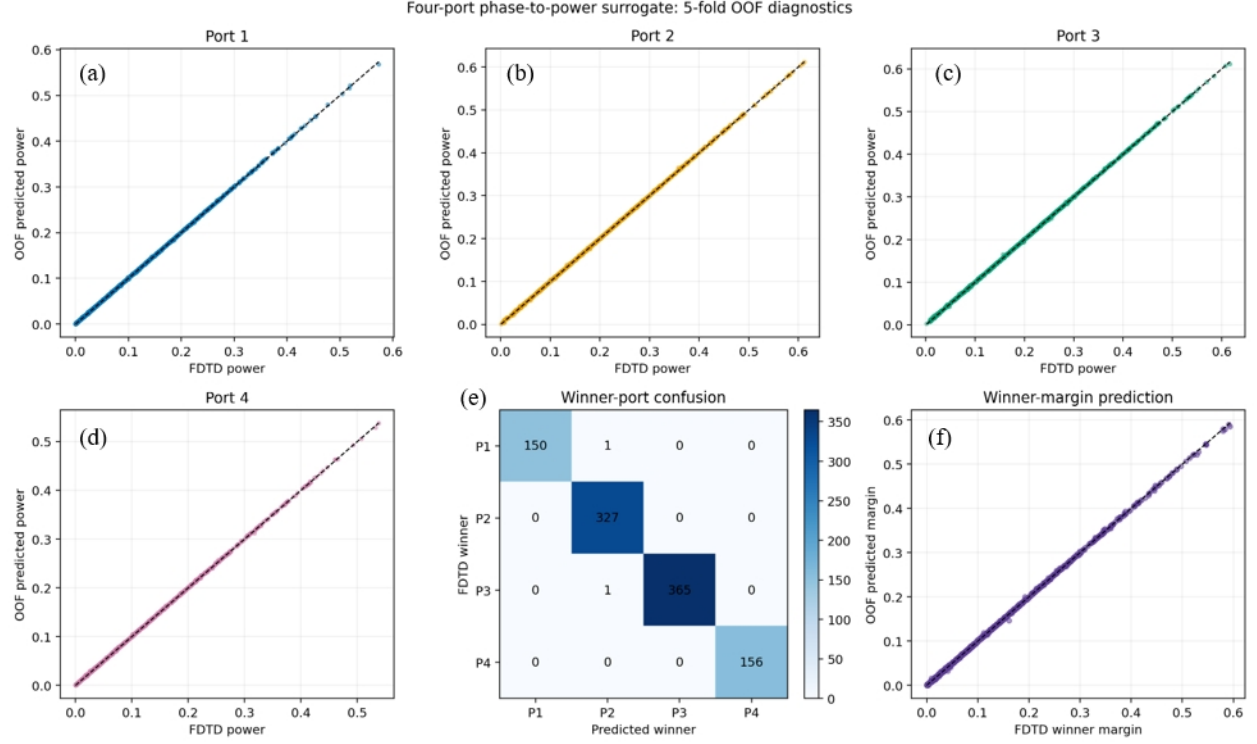


Fig. S7 Cross-validated fidelity of the four-port response surrogate. Fivefold out-of-fold predictions for 1,000 FDTD response states. a-d. The four port-resolved scatter plots compare predicted and FDTD output powers; dashed lines indicate equality. e. The confusion matrix compares the ports with the highest predicted and FDTD powers, with agreement for 998 of 1,000 states. f. compares the winner margin, defined as the difference between the largest and second-largest port powers. The mean port-wise coefficient of determination is 0.999944, and the mean absolute power error is 0.000562. These diagnostics assess the device-response surrogate independently of the 400-image digit-classification test.

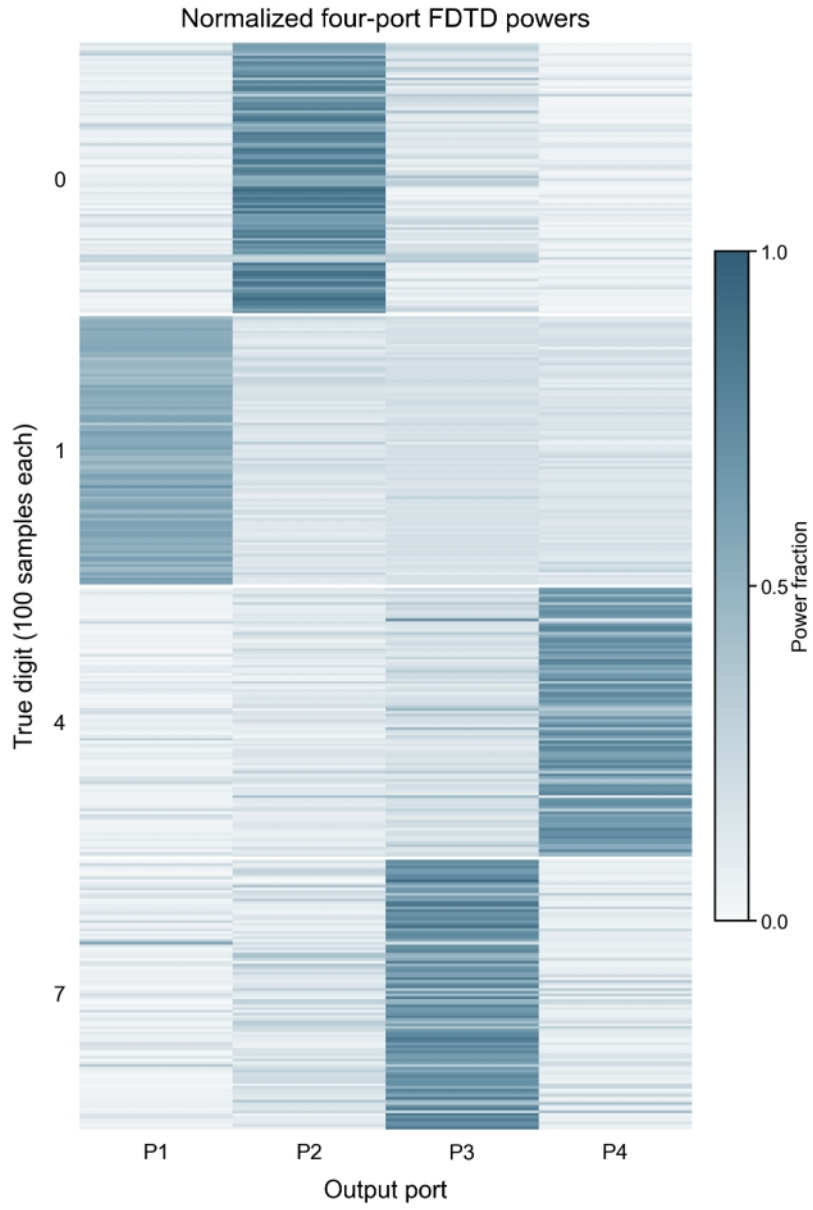


Fig. S8 Per-sample four-port outputs for digit classification. FDTD output powers for 400 test images of digits 0, 1, 4 and 7, with 100 images per digit. Powers are normalized by their sum across the four ports for each image, and samples are grouped by the true digit. The target ports are P2, P1, P4 and P3 for digits 0, 1, 4 and 7, respectively. All test phase states are generated by the fixed encoder and evaluated directly in FDTD.

Supplementary Note 1

Device response modelling and fivefold classification

The three-port classifier combines an electronic feature encoder with the calibrated voltage-to-output response of the fabricated device. Each Iris sample is represented by four morphological features, which are standardized using the mean and standard deviation of the corresponding training fold. An encoder with four inputs, eight hidden units and three outputs converts these features into continuous control voltages. A hyperbolic tangent activation is used in the hidden layer, and sigmoid outputs are scaled to the calibrated voltage range of each channel. Setosa, Versicolor and Virginica are assigned to ports P1, P2 and P3, respectively.

The device response is represented by trilinear interpolation of 1,224 measured voltage states. In the calibration readout convention, the lowest recorded photocurrent identifies the output port with the highest optical power. Classification therefore selects the port with the minimum interpolated photocurrent. For visualization, the resulting scores are normalized by their sum across the three ports. This transformation preserves the winning port assignment; the normalized scores describe the model response and do not correspond to measured optical-power fractions. Continuous voltage commands for the separate fixed set of 30 held-out samples are shown in Supplementary Fig. 4, normalized to the calibrated range of each control channel.

Training begins with masked feature reconstruction using only the 120 training samples per fold. The autoencoder adopts a 4–8–3–8–4 architecture and is trained with a feature-masking probability of 0.35 and additive Gaussian noise with a standard deviation of 0.04 in standardized feature units. Its first encoder layer initializes the voltage encoder. Subsequent fine-tuning against the device response uses 30 labelled samples (ten per class). The remaining 90 training samples contribute to reconstruction pretraining, while their labels are not used for fine-tuning. The objective function combines classification cross-entropy, a target-port margin penalty, and regularization of the encoder parameters. Gradients are propagated through the interpolated device response, while the calibration grid remains fixed.

Stratified five-fold cross-validation is performed across all 150 samples. Within each outer training fold, three of the ten labelled samples per class are reserved for internal validation during the search over initialization and softmax temperature. The selected settings are then used to fine-tune the encoder with all 30 labelled training samples. Neither feature standardization nor model selection uses the outer test fold. Each sample consequently yields one out-of-fold prediction from an encoder trained without that sample.

The five folds yield 27, 23, 23, 30 and 30 correct predictions out of 30 samples, respectively (Supplementary Fig. 5). The mean accuracy is 88.67%, with a sample standard deviation of 11.69 percentage points across folds; pooled predictions give 133 correct assignments out of 150 total samples. Six Versicolor samples are assigned to Virginica and ten Virginica samples to Versicolor, accounting for 16 of the 17 errors. The fold-resolved results describe variation across data

partitions and selected model initializations. They are obtained with the calibrated response model, separate from the chip measurements described below.

Supplementary Note 2

Experimental readout and output separation

The experimental test uses a separate, fixed set of 30 held-out Iris samples, with ten samples from each species. Their encoded voltage triplets are applied to the three thermo-optic control channels, and the output powers are measured at a wavelength of 1550 nm. For each sample, the power at each port is divided by the sum of the three output powers, and the port with the largest measured power determines the predicted class. These measurements evaluate the voltage commands of the fixed-split model, rather than the 150 out-of-fold commands from the fivefold evaluation.

Output separation is quantified by the decision margin, defined as ten times the base-10 logarithm of the ratio between the largest and second-largest measured output powers. All 30 samples are assigned to their target ports in this acquisition. The mean margin is 6.77 dB and the minimum is 3.60 dB (Supplementary Fig. 6, left). Averaging the target-port power fraction over the ten samples in each class gives 75.8%, 74.2% and 81.4% for Setosa, Versicolor and Virginica, respectively. These values quantify the distribution of the measured output power among ports, not the fraction of input power transmitted through the device.

The relative output distributions are also compared with the corresponding offline-model predictions (Supplementary Fig. 6, right). Across 90 port-sample pairs, the normalized model scores and measured power fractions have a Pearson correlation coefficient of 0.880 and a mean absolute error of 0.126. This comparison assesses agreement in the relative port responses; classification itself is determined from the measured powers. The reported margins and agreement statistics describe one acquisition per sample and do not quantify temporal drift or measurement repeatability.

Supplementary Note 3

Four-port image encoding and full-wave validation

Four-class image classification uses MNIST digits 0, 1, 4 and 7. The original 28×28 images are reduced to 8×8 pixels by area averaging and scaled to the range of zero to one. The training and validation sets contain 500 and 100 images per digit, respectively, sampled without overlap from the official training split. An additional 100 images per digit are drawn from the official test split, resulting in 2,000 training, 400 validation and 400 test images in total.

A differentiable surrogate represents the four-phase-to-four-power response using 1,000 finite-difference time-domain (FDTD) simulations. Its inputs are the sine and cosine of the three phase differences relative to the first input, providing a periodic representation of the relative

phases. Two hidden layers of 64 units map these features to four non-negative port powers. Fivefold out-of-fold evaluation of these response samples yields a coefficient of determination of 0.999944 and a mean absolute error of 0.000562; the predicted winning port agrees with FDTD results for 998 of the 1,000 states (Supplementary Fig. 7). This evaluation characterizes the fidelity of the response model, independent of the digit-classification task.

The image encoder applies an affine projection to the 64 pixels standardized using training-set statistics, followed by a sigmoid scaled by π to constrain each phase between zero and π . The surrogate parameters are fixed during encoder optimization. Classification cross-entropy is combined with a target-port margin penalty and a phase-boundary penalty. Encoder training uses AdamW with an initial learning rate of 0.008 and weight decay of 0.00002. The class-to-port assignment and encoder checkpoint are selected using validation performance, with the final checkpoint obtained at epoch 220. The resulting mapping is digit 0 to P2, digit 1 to P1, digit 4 to P4, and digit 7 to P3.

The frozen encoder is then applied to the 400 test images, and all resulting phase states are evaluated directly in FDTD. Classification uses the port with the largest simulated output power. The per-sample distributions, normalized across the four ports for display, are shown in Supplementary Fig. 8. FDTD correctly classifies 388 images, including 95, 100, 97 and 96 images for digits 0, 1, 4 and 7, respectively. Across the 1,600 port-sample pairs, surrogate and FDTD powers have a coefficient of determination of 0.99978 and a mean absolute error of 0.00203, with matching winning ports for 399 of 400 images. Power-agreement statistics are calculated before normalization across the output ports; coefficients of determination are calculated for each port and averaged with equal weighting. These results provide full-wave numerical validation of the encoded test states.