

OPTICS

Harnessing optoelectronic noises in a photonic generative network

Changming Wu¹, Xiaoxuan Yang², Heshan Yu³, Ruoming Peng¹, Ichiro Takeuchi³, Yiran Chen², Mo Li^{1,4*}

Integrated optoelectronics is emerging as a promising platform of neural network accelerator, which affords efficient in-memory computing and high bandwidth interconnectivity. The inherent optoelectronic noises, however, make the photonic systems error-prone in practice. It is thus imperative to devise strategies to mitigate and, if possible, harness noises in photonic computing systems. Here, we demonstrate a photonic generative network as a part of a generative adversarial network (GAN). This network is implemented with a photonic core consisting of an array of programmable phase-change memory cells to perform four-element vector-vector dot multiplication. The GAN can generate a handwritten number (“7”) in experiments and full 10 digits in simulation. We realize an optical random number generator, apply noise-aware training by injecting additional noise, and demonstrate the network’s resilience to hardware nonidealities. Our results suggest the resilience and potential of more complex photonic generative networks based on large-scale, realistic photonic hardware.

INTRODUCTION

The current rate of improvement in digital electronics’ energy efficiency (1–3) is lagging behind the fast-growing computational load (4, 5) spurred by the widespread implementation of large-scale artificial neural networks for machine learning and artificial intelligence (6–11). Because of its advantages in power efficiency, communication bandwidth, and parallelism (12–19), analog optical computing based on integrated optoelectronic processors (20–25) is once again brought into focus as a hardware accelerator for neural networks. Photonic neural networks reported to date (17, 20, 22, 23, 26, 27) are predominantly hybrid optoelectronic systems in which the photonic components are used for linear multiplication and interconnect, while nonlinear functions and feedback control are implemented electronically. Compared to electronic neural networks using digital processors, photonic neural networks have higher inaccuracy and error rates due to their analog nature and the abundance of optoelectronic noises in the hardware. The accumulation of computational errors in large-scale photonic neural networks could severely impair their performance (28–30), limiting the computation effectiveness and scalability. Although several offline noise-aware training schemes, including injecting noises to layer inputs (29, 31), synaptic weights (28, 32), and preactivations (33, 34), have been proposed to mitigate analog hardware nonidealities (21, 30, 35, 36), those schemes only address discriminative models. In another study, a diffractive optics-based network is trained with carefully drafted parametric randomness to be robust against optical nonidealities (37, 38). Noise in the analog hardware has also been used to facilitate various machine learning algorithms (39–42). In contrast to discriminative models, generative neural network models can automatically discover and learn regularities or patterns from the training data to generate

plausible new instances (43–45). So far, a photonic generative network has not been reported, and the corresponding noise mitigation strategies have not been explored.

Here, we demonstrate a generative network on the basis of a photonic computing core consisting of an array of programmable phase-change metasurface mode converters (PMMCs) (46). This photonic generative network is combined with a discriminator to realize a generative adversarial network (GAN) that is trained to generate handwritten numbers. We show that the photonic GAN can harness and mitigate optoelectronic noises and errors in three ways. First, we use the amplified spontaneous emission (ASE) noise to realize an optical true random number generator (RNG) (47, 48), which is used as the input to the GAN. This optical RNG efficiently generates random numbers at high speed in multiple wavelength channels by slicing the ASE spectrum (48–50). Second, we analyze error sources originating from the components in the photonic GAN and propose noise-aware training approaches by augmenting noises during the training process, which improves the network’s performance and robustness. Last, we validate the training approaches through experiment and simulation and demonstrate that the photonic GAN can benefit from the inevitable random errors in practical implementation. Unexpectedly, the images generated by nonideal photonic hardware show even higher quality than those by ideal, errorless counterparts (i.e., software baseline). Our results demonstrate the feasibility and resilience of more complex photonic GANs using nonideal optoelectronic hardware. Because the proposed noise-aware training approaches are generic, they can be applied to various types of optoelectronic neuromorphic computing hardware.

RESULTS

Photonic generative networks in a GAN architecture

A GAN network consists of two subnetwork models (Fig. 1A): a generator and a discriminator (49–51). These two models compete against each other in a zero-sum game: The discriminator strives to distinguish the instances produced by the generator (labeled as the “fake” instances) from the real instances in the training dataset (labeled as the “real” instances); the generator aims to fool the

Copyright © 2022
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195, USA. ²Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA. ³Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. ⁴Department of Physics, University of Washington, Seattle, WA 98195, USA.

*Corresponding author. Email: moli96@uw.edu

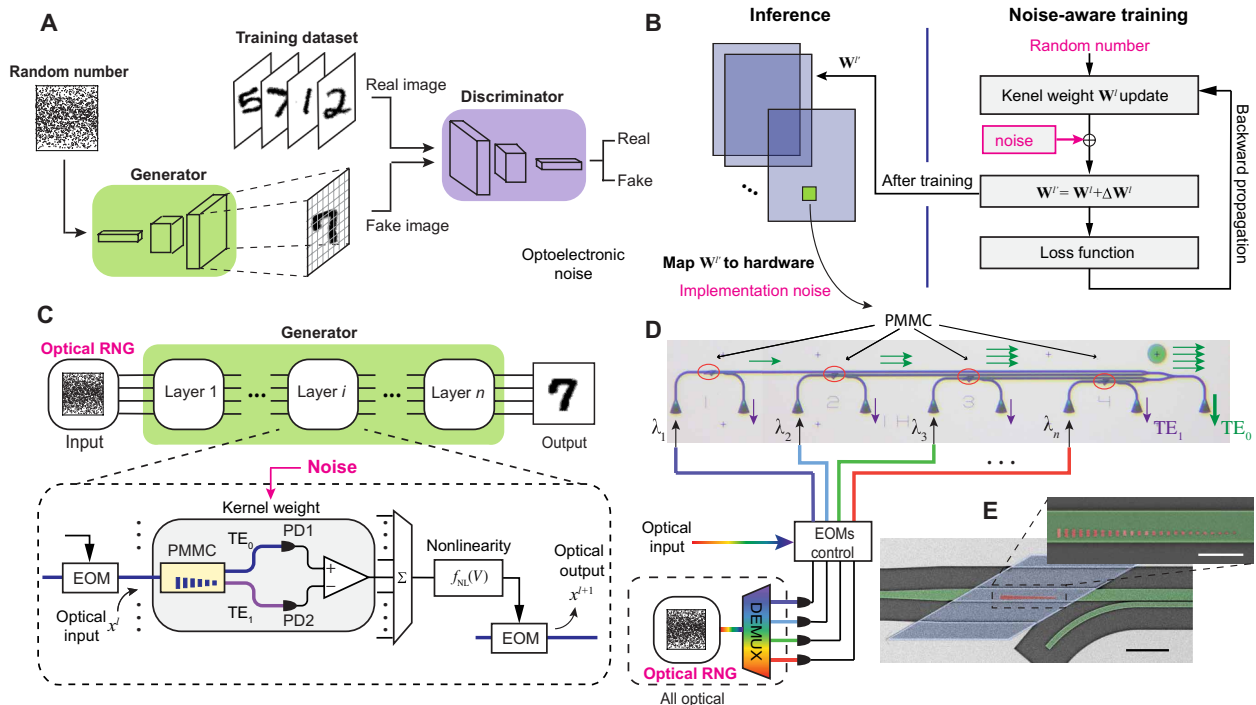


Fig. 1. Photonic GAN network with optoelectronic noises. (A) A GAN architecture is composed of two subnetwork models: a generator and a discriminator. The generator competes with the discriminator during training and produces new instances after it is trained. (B) The offline noise-aware training and inference processes flow of the generator. The process of mapping the trained weight to the hardware during implementation inevitably introduces optoelectronic noise. (C) Decomposition of the generator into individual layers. In each layer, the input signals pass through the photonic tensor core and are converted to the electrical domain by photodetectors (PDs). After postprocessing, the data are converted back into the optical domain and transferred to the next layer. EOM, electro-optic modulator. (D) Optical microscopic image of the photonic tensor core consisting of four input channels. The optical RNG is input to the photonic tensor core through O/E and E/O conversion in our experiment. Potentially, the optical RNG can be directly sent into the tensor core using WDM schemes. DEMUX, demultiplexers. (E) The false-colored scanning electron microscopy (SEM) image of the photonic tensor core. The Si_3N_4 waveguide, the GST metasurface, and the Al_2O_3 protection layer are colored green, red, and blue, respectively. Scale bar, 10 μm . Inset: The zoomed-in SEM image of the phase-gradient metasurface on the waveguide. Scale bar, 2 μm .

discriminator by producing instances that imitate the real instances. The competition drives both networks to improve their capabilities until an equilibrium state is reached, i.e., when the fake instances are indistinguishable from the real instances by the discriminator, so the generator is deemed well trained to generate plausible new instances. In this work, we design a prototype photonic generator to produce images of the handwritten number “7” using a noise-aware offline training configuration: We first train the generator model on a computer (52) and implement it on the photonic platform (Fig. 1B). Here, we only focus on realizing the photonic generator because the photonic discriminator has been demonstrated previously by many groups, including us (16, 17, 20, 46). As shown in Fig. 1C, in each layer of the generator, the input data are encoded in the power of the optical signals through multiple wavelength channels, processed by the PMMC photonic tensor core (Fig. 1D), in which the kernel matrices are stored. The results are detected by the photodetector arrays. Electronic postprocessing is then performed to apply nonlinear functions. The results are reencoded into the optical signals and relayed to the next photonic layer. In such an optical network, various noises—including optical and electrical noises of the optical sources, modulators, and photodetectors—are accumulating through the processes of programming (i.e., writing) the kernel matrices, data encoding, and data transferring (i.e., reading) between the layers of the network.

Optical RNG

One key component of the photonic generator is the optical RNG that produces the random input. To realize it, we use the ASE noise from the erbium-doped fiber amplifiers (EDFAs), which are ubiquitous in fiber-optic communication systems, to generate random optical signals at high rates in four parallel channels as shown schematically in Fig. 2A. Here, the ASE noise is first filtered with wavelength division multiplexing (WDM) demultiplexers and then detected with photodetectors. The generated baseband electrical currents due to beating between different frequency components are the so-called “ASE-ASE beat noises” (53, 54). The DC photocurrent is filtered by a DC block, passing only the stochastic photocurrent variances to a sampling oscilloscope to generate random numbers (see the Supplementary Materials for the theory of the optical RNG). Figure 2 (B and C) plots the statistical histogram and a representative trace of the random numbers (in voltage) generated in a single WDM channel, respectively. The probability density function is well approximated by a zero-mean Gaussian distribution with a standard deviation (SD) of 0.2 V [i.e., $N(0, 0.2)$]. We further calculate the correlation coefficient of an $N = 5 \times 10^4$ number long sequence (Fig. 2D), which reaches the limit of $1/\sqrt{N}$ (red line in Fig. 2D), proving the randomness of the number sequence. Because of the limited size of the photonic tensor core, we need to measure and record the random numbers from the RNG and repeatedly input them to the

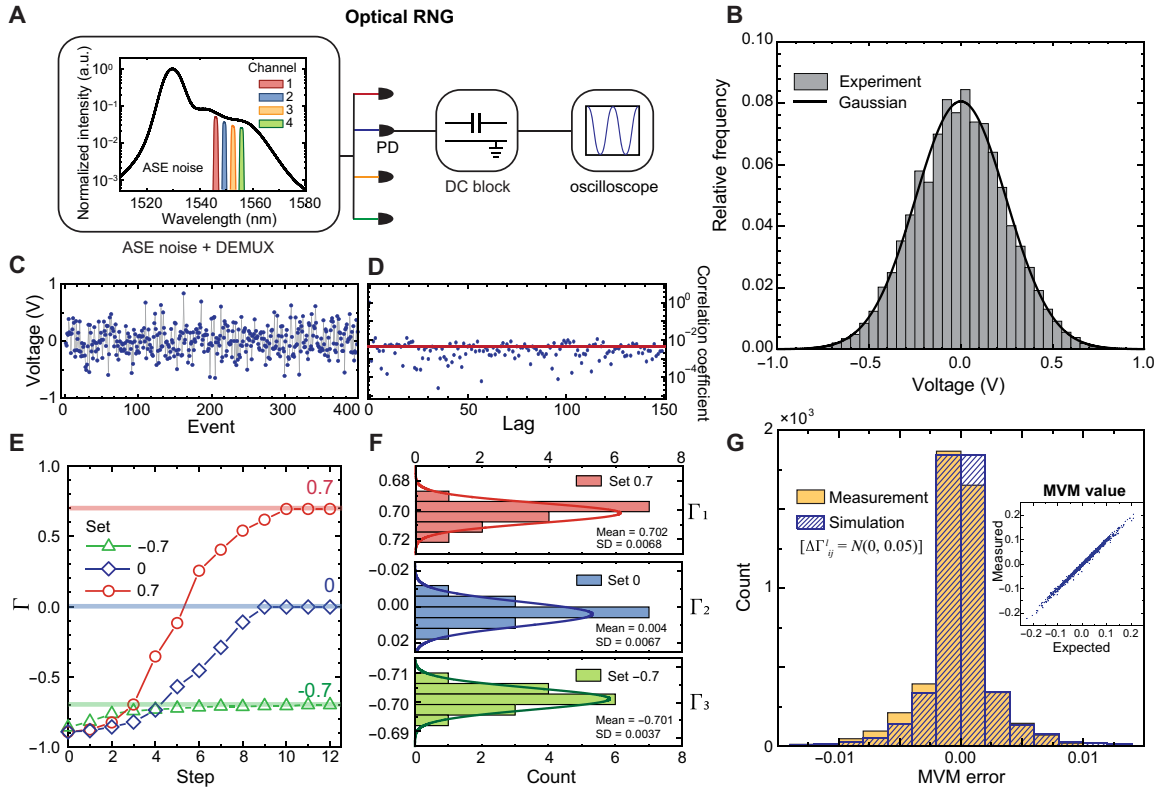


Fig. 2. Optical RNG and kernel programming errors. (A) Schematic of the optical RNG. The ASE noise is spectrally sliced into four wavelength channels using DEMUX and then detected with photodetectors. After a DC block, the random electrical signals are sampled by an oscilloscope. a.u., arbitrary units. (B and C) Statistical histograms (B) and a representative trace (C) of the generated random numbers. The generated random number follows the Gaussian distribution. (D) Correlation coefficient as a function of lag for the random number sequence. A random number sequence with length $N=5 \times 10^4$ has a correlation coefficient (blue dots) around the lower limit $1/\sqrt{N}$ (red line). (E) Process of programming the mode contrast of a kernel element using optical pulses. The target Γ values are -0.7 , 0 , and 0.7 , respectively. (F) Histogram of Γ value distribution when the kernels are repeatedly set to be -0.7 , 0 , and 0.7 , respectively. The SD for each setting is 0.37 , 0.67 , and 0.68% , respectively. (G) Histograms of the error distribution in the experimental measurement (solid) and the simulation (hashed) when assuming the $\Delta\Gamma_{ij}^l$ follow a Gaussian distribution with an SD of 5% . Inset: Measured MVM accuracy for 4900 MVM operations in the first layer of the network.

generator during the experiment (see Fig. 1D). In future full-scale systems, the filtered ASE noise can be directly used as random optical inputs to the GAN without electrical sampling (the dashed box in Fig. 1D) and detected after the first layer of the network is performed.

Photonic tensor core error analysis

The other key component of the photonic generator is the photonic tensor core, which optically performs matrix-vector multiplication (MVM). The inset in Fig. 1C shows the schematic of one PMMC kernel element of the core that computes multiply accumulate (MAC): $x \rightarrow x \cdot w + b$, the fundamental operation of MVM. The PMMC consists of an array of $\text{Ge}_2\text{Sb}_2\text{Te}_5$ (GST) nanoantennas with tapering widths (see Fig. 1E for the scanning electron microscopy images), forming a phase-gradient metasurface patterned on a silicon nitride waveguide (46). The input vector element x is encoded in the power of the input optical signal. The corresponding kernel element weight w is represented using the TE_0/TE_1 mode contrast $\Gamma = \beta_{\text{TE}_0} - \beta_{\text{TE}_1}$ at multiple intermediate levels between $[-1, 1]$, where $\beta_{\text{TE}_0(\text{TE}_1)} = P_{\text{TE}_0(\text{TE}_1)} / (P_{\text{TE}_0} + P_{\text{TE}_1})$ is the mode purity and P_{TE_0} (P_{TE_1}) is the power of the TE_0 (TE_1) mode component in the waveguide. Thus, the MAC computation is simplified to an incoherent optical transmission measurement and can be performed over a broad bandwidth. Figure 2E

shows the evolution of Γ during the programming process of using optical control pulses to set negative (-0.7), zero (0.0), and positive (0.7) values, respectively. We implement the network model on a 2×2 tensor core with four PMMCs (Fig. 1D). The kernel weight W_{ij}^l value is mapped to the corresponding mode contrast Γ_{ij}^l as $\Gamma_{ij}^l = W_{ij}^l \cdot (|\Gamma|_{\text{max}}^l / |W|_{\text{max}}^l)$, where $|\Gamma|_{\text{max}}^l$ is the maximum absolute mode contrast and $|W|_{\text{max}}^l$ is the maximum absolute kernel weight of layer l . Given the limited number of PMMCs on a chip, we repeatedly reset the kernel elements on the same devices, which bottlenecks the computing speed. With a sufficiently large tensor core in a photonic crossbar array architecture (55–57), one could directly map the full kernel matrices to the hardware so the computing speed will be much accelerated.

The analog nature of weight programming and data encoding and transferring in the photonic neural network limits the precisions of MVM calculations and makes the computation error-prone. The computation errors would accumulate through the layers of the network and impair the final results. Because in realistic experiments, the computation errors stem from various optoelectronic noises in the system, we use the terms of noise and error interchangeably. To quantify the noises and errors in our system, we repeatedly program different fixed Γ values and estimate the short-term inaccuracy by measuring the

variation $\Delta\Gamma$. Figure 2F shows that the SD of 15 programming operations is less than 0.7%, corresponds to 6 bits in resolution, which is one order of magnitude larger than the input encoding error (see the Supplementary Materials for more detailed error analysis). Thus, the short-term programming inaccuracy $\Delta\Gamma$ (write error), limited by the inaccuracy of the programming optical pulses, is one of the dominant error sources. Another error source is the long-term measurement fluctuations (read error), including the noise of photo-detectors, the variation of the O/E and E/O conversions, and the thermo-optic fluctuation of the phase-change material (PCM). These errors collectively contribute to an effective error $\Delta W_{ij}^l = (|\Gamma|_{\max}^l / |W|_{\max}^l) \cdot \Delta\Gamma_{ij}^l$ on the kernel element weight W_{ij}^l , where $\Delta\Gamma_{ij}^l$ is the total write error. To estimate the computation error of the overall system, Fig. 2G compares the measured MVM error distributions with the simulation, which assumes a Gaussian distribution of error. The result estimates the overall error $\Delta\Gamma_{ij}^l$ to be 5%, corresponding to more than 3 bits in resolution, which we subsequently use in the noise-aware training and simulation.

Unlike the discriminative network, where the input regularities or patterns are well defined, the generator network takes random numbers as the input. It would be more susceptible to the effective weight setting noise ΔW_{ij}^l , which could degrade the quality of the generated new instances (58, 59). To reveal the noise effect on the GAN, we emulate the noisy hardware on a GAN model that is trained using a noiseless offline training approach but add a random error ΔW_{ij}^l [introduced by $\Delta\Gamma_{ij}^l$ with a Gaussian distribution $N(0, 0.05)$] when using it to generate images. Figure 3A plots 49 images of 14×14 pixels

generated from simulation using random inputs produced by the optical RNG. These images show the handwritten number 7 but with very noisy backgrounds, demonstrating that the noise-free training algorithm is impaired by the practical weight setting noise (see the Supplementary Materials for the detailed comparison between inference results using accurate and inaccurate kernels).

Therefore, it is necessary to consider hardware noise during training to realize a GAN that is resilient to realistic noises. Theoretically, it has been proven that adding noises to the training data of a neural network is equivalent to an extra regularization added to the error function (31), which can significantly improve hardware noise tolerance in a discriminative neural network. Meanwhile, it was shown that introducing noise on kernel weights during training enhances the robustness against weight perturbations of multilayer perceptrons (28), such that inference accuracy close to the software baseline could be achieved. However, previous demonstrations of noise-aware solutions are limited to discriminative networks. For GAN, theoretical, simulation, and experimental validations of effective noise-aware solutions are still lacking and require further investigation.

Noise-aware training of the photonic generative model

For our photonic GAN, we propose and experimentally validate two noise-aware training approaches, namely, the input-compensatory approach (IC-GAN) and the kernel weight-compensatory approach (WC-GAN), to improve the network’s tolerance to the effective weight setting noise ΔW_{ij}^l . The IC-GAN approach inflates the SD of the random signal input from the experimental value of 0.2 to 0.5 V

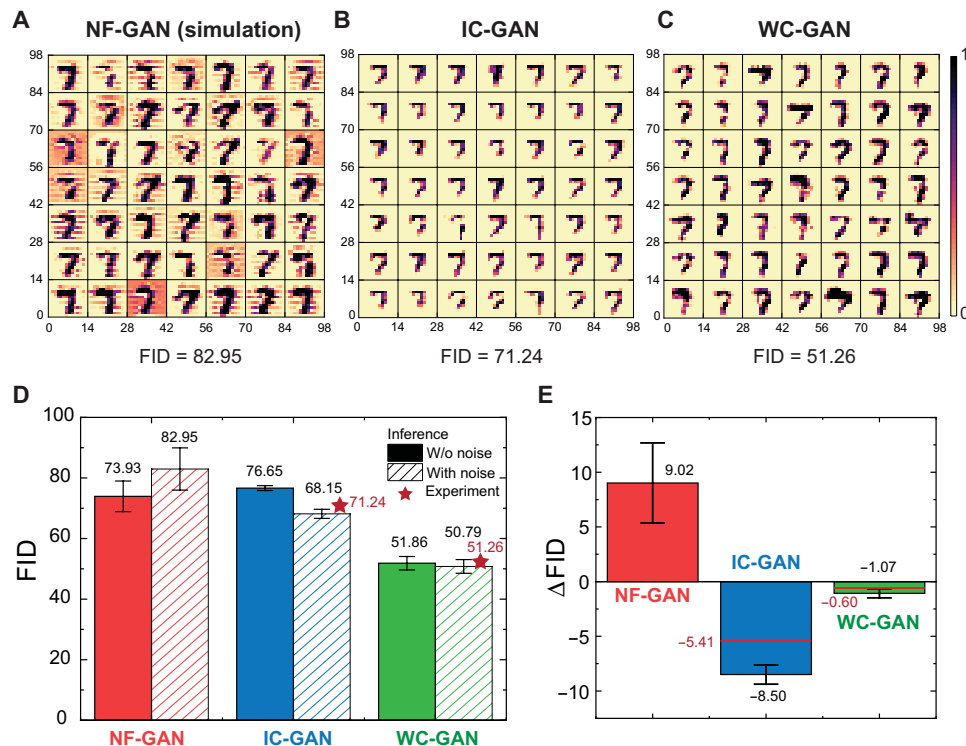


Fig. 3. Generating handwritten numbers with GAN. (A to C) Forty-nine images (size, 14×14 pixels) generated by (A) NF-GAN, (B) IC-GAN, and (C) WC-GAN under effective kernel weight setting error (introduced by 5% Gaussian random error $\Delta\Gamma_{ij}^l$) and using random inputs $\sim N(0, 0.2)$ produced by the optical RNG. (A) is generated by simulation, and (B) and (C) are from the experiments. (D) The FIDs of the generated images, assuming the network is trained using various approaches and is implemented either on the ideal (solid bars) or noisy hardware (hatched bars). The FIDs obtained from the experimental results are labeled as stars. (E) The Δ FIDs from the experimentally generated images are denoted by the red lines.

during training. The WC-GAN approach adds $\Delta\Gamma_{ij}^l$ with 5% SD to the corresponding weight at each forward-propagation pass but performs noiseless gradient descent in the back-propagation pass (see Fig. 1B and the Supplementary Materials for the training procedure of these noise-aware training approaches). Figure 3 (B and C) shows the experimentally generated images of handwritten number 7 by the photonic GAN trained using both approaches. For a fair comparison, the random number inputs used for inferences are produced by the same optical RNG. Compared to the images generated by the noise free-trained GAN (NF-GAN) (Fig. 3A), the images generated using both noise-aware approaches display much clearer patterns with lower background noise, validating the noise tolerance of the IC-GAN and WC-GAN. Furthermore, we observe that the images generated by the WC-GAN (Fig. 3C) have richer handwritten-like features than those by the IC-GAN (Fig. 3B), with more diverse variations in styles. Therefore, we conclude that the WC-GAN is advantageous for practical implementation using nonideal analog hardware.

DISCUSSION

To quantitatively compare the GAN performance, we use the standard metric of Frechet inception distance (FID), which evaluates both the fidelity and diversity of the generated images by comparing the feature distribution in the generated images with images from the training dataset. The lower the FID score, the better performance of the GAN (51). In Fig. 3D, the FIDs of the images generated by the NF-GAN (36, 51, 60), the IC-GAN, and the WC-GAN, respectively, are compared, assuming either ideal (FID_{ideal}) or noisy (FID_{noisy}) hardware (see the Supplementary Materials for detailed steps to calculate the FID). The FID_{noisy} (hashed bars in Fig. 3D) is the lowest for the WC-GAN and the highest for the NF-GAN, consistent with the observation in Fig. 3 (A to C). The impact of hardware noise $\Delta FID = FID_{noisy} - FID_{ideal}$ is plotted in Fig. 3E. The noise-aware WC-GAN and IC-GAN show two notable benefits. First, the FID_{ideal} (solid bars in Fig. 3D) for the WC-GAN is lower than the NF-GAN [e.g., the software baseline (61)], indicating that introducing noises during training helps GAN to learn better. Such a gain is absent in discriminative networks, where the inference accuracies of the noise-aware trained model cannot exceed the software baseline (29, 30, 33, 34). Second, unexpectedly, the noise impact results (Fig. 3E) show that, unlike the NF-GAN, the WC-GAN and IC-GAN implemented on the photonic hardware with practical noise (hashed bars in Fig. 3D) perform even better in inference than the noiseless hardware (solid bars in Fig. 3D). In contrast, a discriminative network's inference accuracy always decreases with more noisy hardware (37, 38). This unexpected gain in performance suggests photonic neural networks' potential in generative models despite the inevitable optoelectronic noises and errors.

Optical computation in this work is performed at a low speed of 4000 operations per second (4 KOPS), limited by the use of the low-speed variable optical attenuators (VOAs) to encode data and the small-scale 2×2 tensor core. However, the state-of-the-art integrated photonic transmitters and photodetectors can drive the system at many 10s of Gbits/s. The size of the photonic core can be further scaled up to a much larger array. Assuming a moderate data rate of 10 Gbits/s and four WDM channels, the computing density of a photonic tensor core can reach an upper-bound value of 25 TOPS/mm² (tera operations per second per square millimeter), significantly higher than that of the state-of-the-art digital electronics. To predict whether the

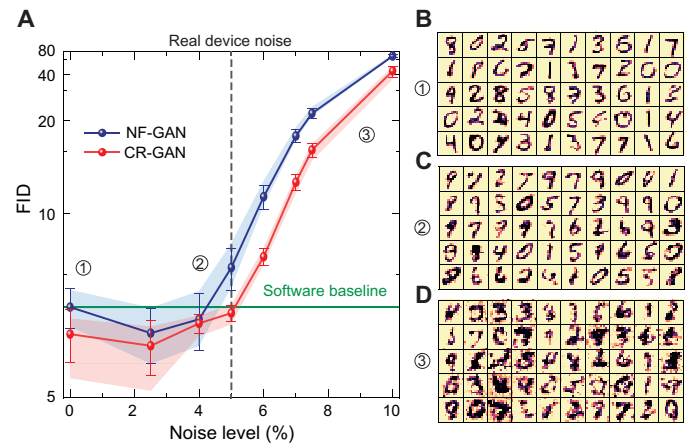


Fig. 4. Scalability of noise-aware training. (A) The FID of the generated images by the NF-GAN and the CR-GAN, respectively, under various effective mode contrast setting noise $\Delta\Gamma_{ij}^l$ with SD ranging from 0 to 10%. The shaded region indicates the range of FID over five individual tests. The FID is lower for CR-GAN at every noise level. At the measured noise level of 5% (black dashed line), the FID for CR-GAN is below the software baseline (solid green line), while the FID for the NF-GAN is above it. (B to D) 50 images (size, 14×14) generated by CR-GAN assuming effective mode contrast setting noise of (B) 0%, (C) 5%, and (D) 10%.

noise-aware approaches performance gain is scalable, in simulation, we train a larger-scaled GAN to generate images of all 10 number digits using ideal or noise-aware approaches under various levels of writing errors. Figure 4A shows the FID score of the results as a function of $\Delta\Gamma_{ij}^l$. Here, the curvature regularization approach (CR-GAN), which evolves from the WC-GAN, is used to improve the GAN robustness further (see the Supplementary Materials for more details about the CR-GAN). The comparison shows that the CR-GAN performs better than the NF-GAN at every error level. Note that, under our present realistic noise level of 5% (Fig. 3G), the FID of CR-GAN is still below the software baseline, whereas the NF-GAN's FID is higher than the baseline. For both approaches, with the increasing noise level, the FID first drops until reaching a minimum at $\sim 2.5\%$ noise and then increases. To explain this, we further examine the images generated by CR-GAN at three noise levels: 0, 5, and 10% in Fig. 4 (B to D). The comparison shows that the increasing hardware noise in GAN would improve the diversity [evaluated by the SD of the percentage of each number classes in the generated images (58); see the Supplementary Materials for more details] but, at the same time, reduce the fidelity of the generated images (58). The trade-off results in a minimal FID at $\sim 2.5\%$ noise, as shown in Fig. 4A. Throughout the full range of noise levels, the noise-aware approach consistently improves the GAN over the noiseless approach.

In conclusion, we demonstrate a photonic generative network on the basis of phase-change photonics, which is used to form a GAN network and harnesses the intrinsic noise sources in the photonic system. Unlike the previously demonstrated discriminative networks that suffer from the hardware noise, our experimental and simulation results show that the photonic generative network not only can tolerate but also can benefit from a certain level of hardware noise after training by noise-aware training approaches. Our finding expands the current implementation of photonic neural networks to generative models (62), in which the inevitable and ubiquitous optoelectronic noises and errors can be mitigated and even leveraged in intelligent

ways. We emphasize that the proposed noise-aware training approaches are generic and thus applicable to various types of optoelectronic neuromorphic computing hardware. The improved noise resilience of the model also implies their scalability in large-scale photonic neural networks with tightly cointegrated electronics and photonics.

MATERIALS AND METHODS

PMMC design and fabrication

The PMMC consists of a phase-gradient metasurface made of GST thin film on silicon nitride waveguides. The metasurface is designed to convert the incident TE_0 mode into the TE_1 mode when GST is in the crystalline phase while maintaining the TE_0 mode when GST is in the amorphous phase. The PMMC is fabricated by depositing a 30-nm-thick GST film using a sputtering tool on an oxidized silicon substrate with 330-nm-thick silicon nitride film. The GST film is then patterned into the metasurface using standard electron beam lithography and inductively coupled plasma etching processes. A 218-nm-thick Al_2O_3 layer is deposited with atomic layer deposition to cap the GST conformally.

Measurement setup

The measurement set up to operate the photonic tensor core is shown in fig. S2. The input optical signals are carried by four different wavelengths using four tunable continuous wave lasers. The signal amplitudes are controlled by VOAs with a 1-kHz operation speed. An additional control laser coupled with a 1×4 optical switch is used to optically program the kernel weight into each GST PMMC. The control pulses are generated with a 12-GHz electro-optical modulator and amplified by a low noise EDFA. The energy of each control pulse is further tuned using another VOA. The input signals and the control pulses are coupled into the photonic device via integrated grating couplers with a coupling efficiency of $\sim 20\%$. The input signals propagate forward through each input channel, while the control pulses propagate in the opposite direction through the TE_1 detection waveguides. The optical power in TE_0 mode is combined on-chip using integrated Y junctions and detected. The optical power in the TE_1 mode is collected and combined off-chip. The mode power contrast is measured to give the MVM results.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abm2956>

REFERENCES AND NOTES

- B. Marr, B. Degnan, P. Hasler, D. Anderson, Scaling energy per operation via an asynchronous pipeline. *IEEE Trans. Very Large Scale Integration (VLSI) Syst.* **21**, 147–151 (2013).
- M. M. Waldrop, The chips are down for Moore's law. *Nature* **530**, 144–147 (2016).
- T. N. Theis, H.-S. Philip Wong, The end of Moore's law: A new beginning for information technology. *Comput. Sci. Eng.* **19**, 41–50 (2017).
- N. Jones, How to stop data centres from gobbling up the world's electricity. *Nature* **561**, 163–166 (2018).
- X. Xiaowei, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, Y. Shi, Scaling for edge inference of deep neural networks. *Nat. Electron.* **1**, 216–222 (2018).
- K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [preprint] (2014) [cs.CV].
- V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, D. Hassabis, Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
- Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
- D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
- Y. H. Chen, T. Krishna, J. S. Emer, V. Sze, Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE J. Solid State Circuits* **52**, 127–138 (2017).
- P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, H. Qian, Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).
- D. Brunner, M. C. Soriano, C. R. Mirasso, I. Fischer, Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **4**, 1364 (2013).
- D. A. B. Miller, Attjoule optoelectronics for low-energy information processing and communications. *J. Lightwave Technol.* **35**, 346–396 (2017).
- R. Hamerly, L. Bernstein, A. Sludde, M. Soljačić, D. Englund, Large-scale optical neural networks based on photoelectric multiplication. *Phys. Rev. X* **9**, 021032 (2019).
- Z. Ying, C. Feng, Z. Zhao, S. Dhar, H. Dalir, J. Gu, Y. Cheng, R. Soref, D. Z. Pan, R. T. Chen, Electronic-photonic arithmetic logic unit for high-speed computing. *Nat. Commun.* **11**, 2154 (2020).
- X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, D. J. Moss, 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
- 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).
- T. Wang, S.-Y. Ma, L. G. Wright, T. Onodera, B. Richard, P. L. McMahon, An optical neural network using less than 1 photon per multiplication. arXiv:2104.13467 [preprint] (2021) [physics.optics].
- L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, P. L. McMahon, Deep physical neural networks enabled by a backpropagation algorithm for arbitrary physical systems. arXiv:2104.13386 [preprint] [cs.LG].
- 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).
- C. Ríos, N. Youngblood, Z. Cheng, M. le Gallo, W. H. P. Pernice, C. David Wright, A. Sebastian, H. Bhaskaran, In-memory computing on a photonic platform. *Sci. Adv.* **5**, eaau5759 (2019).
- J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, W. H. P. Pernice, All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
- G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, D. Psaltis, Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
- L. Bernstein, A. Sludde, R. Hamerly, V. Sze, J. Emer, D. Englund, Freely scalable and reconfigurable optical hardware for deep learning. *Sci. Rep.* **11**, 3144 (2021).
- B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, P. R. Prucnal, Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021).
- X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, A. Ozcan, All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
- T. Zhou, X. Lin, J. Wu, Y. Chen, H. Xie, Y. Li, J. Fan, H. Wu, L. Fang, Q. Dai, Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
- A. F. Murray, S. Member, P. J. Edwards, Enhanced MLP performance and fault tolerance resulting from synaptic weight noise during training. *IEEE Trans. Neural Netw.* **5**, 792–802 (1994).
- S. Moon, K. Shin, D. Jeon, Enhancing reliability of analog neural network processors. *IEEE Trans. Very Large Scale Integration (VLSI) Syst.* **27**, 1455–1459 (2019).
- V. Joshi, M. le Gallo, S. Haefeli, I. Boybat, S. R. Nandakumar, C. Piveteau, M. Dazzi, B. Rajendran, A. Sebastian, E. Eleftheriou, Accurate deep neural network inference using computational phase-change memory. *Nat. Commun.* **11**, 2473 (2020).
- C. M. Bishop, Training with noise is equivalent to Tikhonov regularization. *Neural Comput.* **7**, 108–116 (1995).
- C. Blundell, J. Cornebise, K. Kavukcuoglu, W. Com, G. Deepmind, Weight Uncertainty in Neural Network. *Proc. Int. Conf. Mach. Learn.* **37**, 1613–1622 (2015).
- M. Klachko, M. R. Mahmoodi, D. Strukov, Improving noise tolerance of mixed-signal neural networks, in *International Joint Conference on Neural Networks (IJCNN)* (2019), pp. 1–8.
- A. S. Rekhi, B. Zimmer, N. Nedovic, N. Liu, R. Venkatesan, M. Wang, B. Khailany, W. J. Dally, C. T. Gray, Analog/mixed-signal hardware error modeling for deep learning inference, in *Proceedings - Design Automation Conference* (2019), pp. 81:1–81:6.

35. I. Boybat, M. le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, E. Eleftheriou, Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **9**, 2514 (2018).
36. A. Sebastian, T. Tuma, N. Papandreou, M. le Gallo, L. Kull, T. Parnell, E. Eleftheriou, Temporal correlation detection using computational phase-change memory. *Nat. Commun.* **8**, 1115 (2017).
37. D. Mengü, Y. Zhao, N. T. Yardimci, Y. Rivenson, M. Jarrahi, A. Ozcan, Misalignment resilient diffractive optical networks. *Nano* **9**, 4207–4219 (2020).
38. D. Mengü, Y. Rivenson, A. Ozcan, Scale-, shift-, and rotation-invariant diffractive optical networks. *ACS Photonics* **8**, 324–334 (2021).
39. F. Cai, S. Kumar, T. van Vaerenbergh, X. Sheng, R. Liu, C. Li, Z. Liu, M. Foltin, S. Yu, Q. Xia, J. J. Yang, R. Beausoleil, W. D. Lu, J. P. Strachan, Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks. *Nat. Electron.* **3**, 409–418 (2020).
40. T. Dalgaty, N. Castellani, C. Turck, K. E. Harabi, D. Querlioz, E. Vianello, In situ learning using intrinsic memristor variability via Markov chain Monte Carlo sampling. *Nat. Electron.* **4**, 151–161 (2021).
41. P. Lalanne, Optoelectronic devices for Boltzmann machines and simulated annealing. *Opt. Eng.* **32**, 1904 (1993).
42. A. Dupret, E. Belhaire, J. C. Rodier, P. Lalanne, D. Prévost, P. Garda, P. Chavel, An optoelectronic CMOS circuit implementing a simulated annealing algorithm. *IEEE J. Solid State Circuits* **31**, 1046–1050 (1996).
43. T. Jebara, *Machine Learning: Discriminative and Generative* (Springer Science & Business Media, 2012), vol. 755.
44. J. Lasserre, C. M. Bishop, Generative or discriminative? Getting the best of both worlds. *Bayesian Statistics* **8**, 3–24 (2007).
45. Y. Bengio, É. Thibodeau-Laufer, G. Alain, J. Yosinski, Deep generative stochastic networks trainable by backprop, in *31st International Conference on Machine Learning, ICML* (2014), vol. 2, pp. 1470–1485.
46. C. Wu, H. Yu, S. Lee, R. Peng, I. Takeuchi, M. Li, Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat. Commun.* **12**, 96 (2021).
47. C. R. S. Williams, J. C. Salevan, X. Li, R. Roy, T. E. Murphy, Fast physical random number generator using amplified spontaneous emission. *Opt. Express* **18**, 23584–23597 (2010).
48. X. Li, A. B. Cohen, T. E. Murphy, R. Roy, Scalable parallel physical random number generator based on a superluminescent LED. *Opt. Lett.* **36**, 1020–1022 (2011).
49. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Nets, in *Advances in Neural Information Processing Systems 27* (2014); www.github.com/goodfeli/adversarial.
50. T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training GANs. arXiv:1606.03498 [preprint] (2016) [cs.LG].
51. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. arXiv:1706.08500 [preprint] (2017) [cs.LG].
52. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems (tensorflow, 2015); tensorflow.org.
53. D. M. Baney, P. Gallion, R. S. Tucker, Theory and measurement techniques for the noise figure of optical amplifiers. *Opt. Fiber Technol.* **6**, 122–154 (2000).
54. G.-H. Duan, E. Georgiev, Non-white photodetection noise at the output of an optical amplifier: Theory and experiment. *IEEE J. Quantum Electron.* **37**, 1008–1014 (2001).
55. Z. Yao, K. Wu, B. X. Tan, J. Wang, Y. Li, Y. Zhang, A. W. Poon, Integrated silicon photonic microresonators: Emerging technologies. *IEEE J. Sel. Top. Quantum Electron.* **24**, 1–24 (2018).
56. T. J. Seok, K. Kwon, J. Henriksson, J. Luo, M. C. Wu, Wafer-scale silicon photonic switches beyond die size limit. *Optica* **6**, 490–494 (2019).
57. A. S. P. Khope, M. Saedi, R. Yu, X. Wu, A. M. Netherton, Y. Liu, Z. Zhang, Y. Xia, G. Fleeman, A. Spott, S. Pinna, C. Schow, R. Helkey, L. Theogarajan, R. C. Alferness, A. A. M. Saleh, J. E. Bowers, Multi-wavelength selective crossbar switch. *Opt. Express* **27**, 5203–5216 (2019).
58. Y. Lin, H. Wu, B. Gao, P. Yao, W. Wu, Q. Zhang, X. Zhang, X. Li, F. Li, J. Lu, G. Li, S. Yu, H. Qian, Demonstration of generative adversarial network by intrinsic random noises of analog RRAM device, in *IEEE International Electron Devices Meeting* (2018), pp. 3–4.
59. O. Krestinskaya, B. Choubey, A. P. James, Memristive GAN in Analog. *Sci. Rep.* **10**, 5838 (2020).
60. P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 5967–5976.
61. The code is originated from www.tensorflow.org/tutorials/generative/dcgan, and the size of the network is reduced accordingly.
62. C. G. Turhan, H. S. Bilge, Recent trends in deep generative models: A review, in *2018 3rd International Conference on Computer Science and Engineering (UBMK)* (2018), pp. 574–579.

Acknowledgment

Funding: C.W., R.P., and M.L. acknowledge the funding support provided by the ONR MURI (award no. N00014-17-1-2661) and the National Science Foundation (award no. CCF-2105972). X.Y. and Y.C. acknowledge the funding support by NSF (award no. 1955196) and ARO (award no. W911NF-19-2-0107). H.Y. and I.T. acknowledge the funding support provided by the ONR MURI (award no. N00014-17-1-2661). Part of this work was conducted at the Washington Nanofabrication Facility/Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure (NNCI) site at the University of Washington with partial support from the National Science Foundation via award nos. NNCI-1542101 and NNCI-2025489. **Author contributions:** M.L., Y.C., and C.W. conceived the research and designed the experiments. C.W. designed and fabricated the photonic devices and performed the experiments with the assistance of R.P. X.Y. and C.W. designed the neural network structure and performed simulation. H.Y. and I.T. developed and deposited the GeSbTe thin film. C.W., X.Y., M.L. analyzed the data. C.W., X.Y., and M.L. wrote the manuscript with contributions from all the authors. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The codes of the numerical simulations are available at GitHub (https://github.com/cmwu2021/NoiseAware_Photonic_GAN/tree/main/Curvature_Regularization_GAN and <https://doi.org/10.5281/zenodo.5705808>).

Submitted 7 September 2021

Accepted 29 November 2021

Published 21 January 2022

10.1126/sciadv.abm2956