

An Optical Neural Network Based on Nanophotonic Optical Parametric Oscillators

Midya Parto^{1,2,*}, Gordon H.Y. Li^{3,*}, Ryoto Sekine¹, Robert M. Gray¹, Luis L. Ledezma¹, James Williams¹, Alireza Marandi^{1,3}

¹Department of Electrical Engineering, California Institute of Technology, 1200 E California Blvd., Pasadena, CA 91125, USA.

²Physics and Informatics Laboratories, NTT Research, Inc., Sunnyvale, California 94085, USA.

³Department of Applied Physics, California Institute of Technology, 1200 E California Blvd., Pasadena, CA 91125, USA.
marandi@caltech.edu
mparto@caltech.edu

Abstract: We experimentally demonstrate a recurrent optical neural network based on a nanophotonic optical parametric oscillator fabricated on thin-film lithium niobate. Our demonstration paves the way for realizing optical neural networks exhibiting ultra-low latencies. © 2024 The Author(s)

The rise of artificial intelligence (AI) and its widespread implementation in the past decade has led to disruptive advances in a variety of fields ranging from fundamental sciences to everyday life technologies. In particular, deep neural networks [1] have enabled a plethora of exciting functionalities such as self-driving cars and natural language processing. Nevertheless, costly and time-consuming computations required for training is known to impede the scalability of the majority of AI models. In this regard, certain models such as extreme learning machines (ELMs) and reservoir computers (RCs) offer major advantages, since only the output layer is trained while other connection weights in the network remain untrained. This makes these models more suitable for implementation in analog physical arrangements, where the internal parameters of the underlying physical system can be left unaltered. Yet, despite their simple training, ELMs and RCs can provide powerful tools when dealing with sequential data [2, 3].

Here, we experimentally demonstrate an ultrafast recurrent optical neural network based on a nanophotonic optical parametric oscillator (OPO) fabricated on a thin-film lithium niobate (TFLN) chip [4], which we call OPO-RNN. Remarkably, we demonstrate an operating clock rate of $\sim 10\text{GHz}$, a quantity that is only limited by the electronic sources responsible for generating the input signal. In our scheme, the masked sequential data is generated by modulating the optical pulses emitted from an electro-optic (EO) frequency comb, which is used to synchronously pump the OPO. In response, the OPO, which is designed to operate under degenerate conditions, emits signal pulses at half-harmonic frequency with respect to the input pump. The resulting signal pulses that are coupled to one another via the nonlinear delayed dynamics of the OPO, form the internal nodes of the network, similar to the neurons within a nonlinear reservoir network (Fig. 1). The output of the OPO-RNN is then determined by taking a linear weighted sum of the electronically detected signal pulses at the output of the TFLN chip. We use the OPO-RNN for chaotic time series prediction, compensating for nonlinear distortions in a noisy communication channel, as well as demonstrating waveform classification.

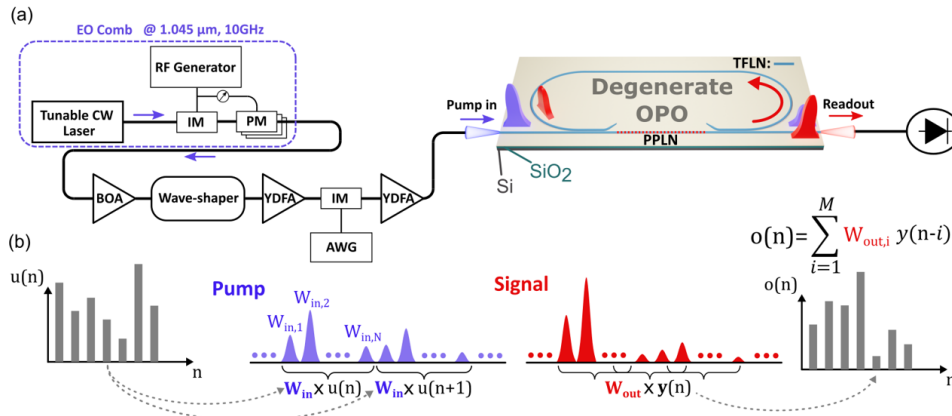


Fig. 1: **RNN based on on-chip OPO.** (a) Schematic diagram of the experimental setup. (b) Implemented recurrent neural network architecture, where only the output weights $\mathbf{W}_{\text{out}}$ are trained.

Figure 1 shows a schematic of our experimental setup. We use an EO comb to generate optical pulses at 10 GHz repetition rate and center wavelength of 1045 nm with a pulsewidth of $\sim 2\text{ps}$ to synchronously pump the on-chip OPO. These pulses are then passed through a booster optical amplifier (BOA), wave shaper, and Ytterbium-doped

fiber amplifier (YDFA) for amplification and dispersion compensation. The sequential data at the input layer is prepared by modulating the pump pulses using an intensity modulator driven by an arbitrary waveform generator (AWG) operating at 10 GSa/s. The modulated pump pulses are then amplified using a second YDFA before being coupled into the chip. Our on-chip OPO is fabricated on an x-cut, 700-nm-thick TFLN wafer with a silica buffer layer (Fig. 1(a)). It comprises two input/output adiabatic couplers before and after the periodically poled section in the middle which are designed to only allow the degenerate signal/idler modes to resonate within the cavity. The emitted signal pulses at 2090 nm are resolved using a fast detector, thus forming the output layer (Fig. 1(b)).

We first perform time-series prediction on the chaotic signal associated with the Lorenz model. The input sequence $u(n)$ is provided by the modulated pump and the goal is to predict $u(n+1)$ at each timestep using previous instances $u(n-i)$ where $i \in 0, 1, \dots, M$. Figure 2(a) shows the target sequence and the experimentally measured predicted one. Here, we use the first 750 timesteps to train the output weights and the following 750 timesteps to test the network, experimentally observing a normalized error of $\text{NMSE} = 0.07 \pm 0.017$.

The next task involves equalizing nonlinear distortions that typically occur in a noisy communication channel [5]. Here, the original signal $S(n)$ comprises symbols chosen randomly from four discrete levels $\{-3, -1, 1, 3\}$. While propagating through the communication channel, the signal experiences inter-symbol interference (ISI), nonlinear effects, as well as random additive noise. Hence, the transmitted signal is given by $T(n) = Q(n) + 0.036 \times Q(n)^2 - 0.011 \times Q(n)^3 + R(n)$, where $R(n)$ is a random Gaussian noise and $Q(n) = \sum_{i=-2}^7 \alpha_i S(n-i)$. Without any equalization and assuming a signal to noise ratio (SNR) of 30dB, our experimental measurements indicate a symbol error rate (SER) of $\text{SER} = 19.3 \pm 1.6\%$. In contrast, by using the OPO-RNN to compensate for the nonlinear distortions, we were able to suppress the error to $\text{SER} = 7.1 \pm 2.3\%$.

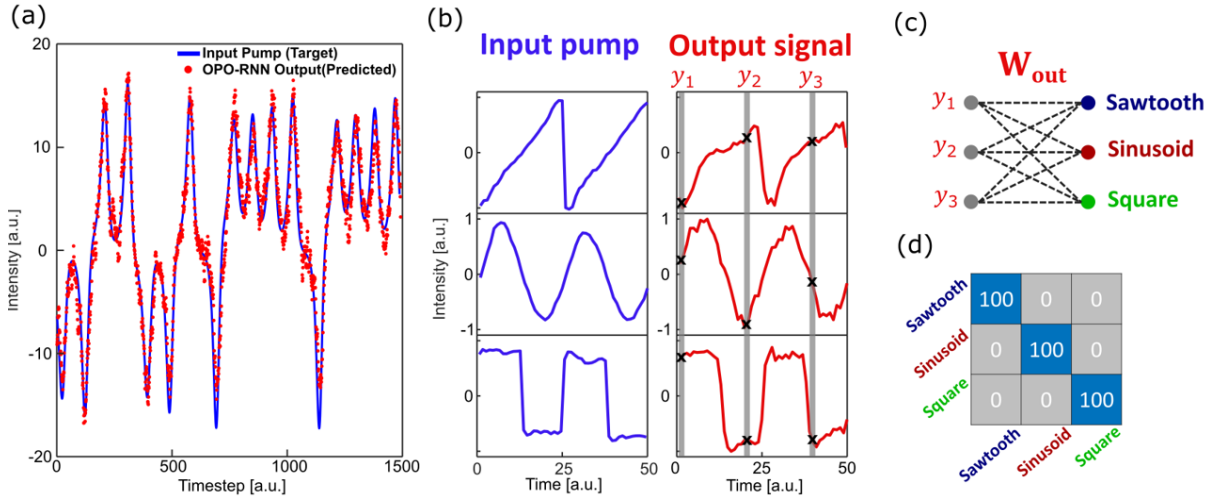


Fig. 2: **Experimental results for different benchmark tasks using OPO-RNN.** (a) Chaotic time series prediction of the Lorenz model. (b) Waveform classification involving three different noisy waveforms, the figure shows typical waveforms used as input pump (left) and detected at the output (right). (c) Using only three output nodes, we achieved 100% accuracy (d).

Finally, we utilize OPO-RNN to classify noisy waveforms randomly chosen from three different classes of sinusoid, sawtooth and square signals. Figure 2(b) shows three instances of input signals belonging to each of these classes that we experimentally measured. Every waveform consists of $N = 50$ samples constituting two consecutive cycles. At the output, the sub-harmonic signal generated by the OPO is sampled at three equidistant points, forming $K = 3$ output neurons (Fig. 2(c)). We use 300 waveforms (100 from every class) to train the output weights according to a winner-takes-all approach, and then use a different set of 300 waveforms for testing. Using this scheme, the OPO-RNN was able to achieve 100% accuracy for this classification task (Fig. 2(d)).

In conclusion, we demonstrated an ultrafast recurrent neural network based on an on-chip OPO operating at 10GHz clock rates. This paves the way for all-optical neural networks using nonlinear photonic integrated circuits. The authors acknowledge support from ARO grant no. W911NF-23-1-0048, NSF grant no. 1846273 and 1918549, Center for Sensing to Intelligence at Caltech, NASA/JPL, and NTT Research for financial and technical support.

References

1. Y. LeCun et al, Nature **521**, 436–444 (2015).
2. G.-B. Huang et al, Neurocomputing **70**, 489–501 (2006).
3. G. Sande et al, Nanophotonics **6**, 561–576 (2017).
4. L. Ledezma et al, Sci. Adv. **9**, eadf9711 (2023).
5. A. Lupo et al, Optica **10**, 1478–1485 (2023).