

## APPLIED PHYSICS

## Wave physics as an analog recurrent neural network

Tyler W. Hughes<sup>1\*</sup>, Ian A. D. Williamson<sup>2\*</sup>, Momchil Minkov<sup>2</sup>, Shanhui Fan<sup>2†</sup>

Analog machine learning hardware platforms promise to be faster and more energy efficient than their digital counterparts. Wave physics, as found in acoustics and optics, is a natural candidate for building analog processors for time-varying signals. Here, we identify a mapping between the dynamics of wave physics and the computation in recurrent neural networks. This mapping indicates that physical wave systems can be trained to learn complex features in temporal data, using standard training techniques for neural networks. As a demonstration, we show that an inverse-designed inhomogeneous medium can perform vowel classification on raw audio signals as their waveforms scatter and propagate through it, achieving performance comparable to a standard digital implementation of a recurrent neural network. These findings pave the way for a new class of analog machine learning platforms, capable of fast and efficient processing of information in its native domain.

## INTRODUCTION

Recently, machine learning has had notable success in performing complex information processing tasks, such as computer vision (1) and machine translation (2), which were intractable through traditional methods. However, the computing requirements of these applications are increasing exponentially, motivating efforts to develop new, specialized hardware platforms for fast and efficient execution of machine learning models. Among these are neuromorphic hardware platforms (3–5), with architectures that mimic the biological circuitry of the brain. Furthermore, analog computing platforms, which use the natural evolution of a continuous physical system to perform calculations, are also emerging as an important direction for implementation of machine learning (6–10).

In this work, we identify a mapping between the dynamics of wave-based physical phenomena, such as acoustics and optics, and the computation in a recurrent neural network (RNN). RNNs are one of the most important machine learning models and have been widely used to perform tasks such as natural language processing (11) and time series prediction (12–14). We show that wave-based physical systems can be trained to operate as an RNN and, as a result, can passively process signals and information in their native domain, without analog-to-digital conversion, which should result in a substantial gain in speed and a reduction in power consumption. In this framework, rather than implementing circuits that deliberately route signals back to the input, the recurrence relationship occurs naturally in the time dynamics of the physics itself, and the memory and capacity for information processing are provided by the waves as they propagate through space.

## RESULTS

## Equivalence between wave dynamics and an RNN

In this section, we introduce the operation of an RNN and its connection to the dynamics of waves. An RNN converts a sequence of inputs into a sequence of outputs by applying the same basic operation to each member of the input sequence in a step-by-step process (Fig. 1A). Memory of previous time steps is encoded into the RNN's hidden state,

which is updated at each step. The hidden state allows the RNN to retain memory of past information and to learn temporal structure and long-range dependencies in data (15, 16). At a given time step,  $t$ , the RNN operates on the current input vector in the sequence,  $\mathbf{x}_t$ , and the hidden state vector from the previous step,  $\mathbf{h}_{t-1}$ , to produce an output vector,  $\mathbf{y}_t$ , as well as an updated hidden state,  $\mathbf{h}_t$ .

While many variations of RNNs exist, a common implementation (17) is described by the following update equations

$$\mathbf{h}_t = \sigma^{(h)}(\mathbf{W}^{(h)} \cdot \mathbf{h}_{t-1} + \mathbf{W}^{(x)} \cdot \mathbf{x}_t) \quad (1)$$

$$\mathbf{y}_t = \sigma^{(y)}(\mathbf{W}^{(y)} \cdot \mathbf{h}_t) \quad (2)$$

which are diagrammed in Fig. 1B. The dense matrices defined by  $\mathbf{W}^{(h)}$ ,  $\mathbf{W}^{(x)}$ , and  $\mathbf{W}^{(y)}$  are optimized during training, while  $\sigma^{(h)}(\cdot)$  and  $\sigma^{(y)}(\cdot)$  are nonlinear activation functions. The operation prescribed by Eqs. 1 and 2, when applied to each element of an input sequence, can be described by the directed graph shown in Fig. 1C.

We now discuss the connection between the dynamics in the RNN as described by Eqs. 1 and 2 and the dynamics of a wave-based physical system. As an illustration, the dynamics of a scalar wave field distribution  $u(x, y, z)$  are governed by the second-order partial differential equation (Fig. 1D)

$$\frac{\partial^2 u}{\partial t^2} - c^2 \cdot \nabla^2 u = f \quad (3)$$

where  $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$  is the Laplacian operator,  $c = c(x, y, z)$  is the spatial distribution of the wave speed, and  $f = f(x, y, z, t)$  is a source term. A finite difference discretization of Eq. 3, with a temporal step size of  $\Delta t$ , results in the recurrence relation

$$\frac{u_{t+1} - 2u_t + u_{t-1}}{\Delta t^2} - c^2 \cdot \nabla^2 u_t = f_t \quad (4)$$

Here, the subscript,  $t$ , indicates the value of the scalar field at a fixed time step. The wave system's hidden state is defined as the concatenation of the field distributions at the current and immediately preceding time steps,  $\mathbf{h}_t \equiv [\mathbf{u}_t, \mathbf{u}_{t-1}]^T$ , where  $\mathbf{u}_t$  and  $\mathbf{u}_{t-1}$  are vectors

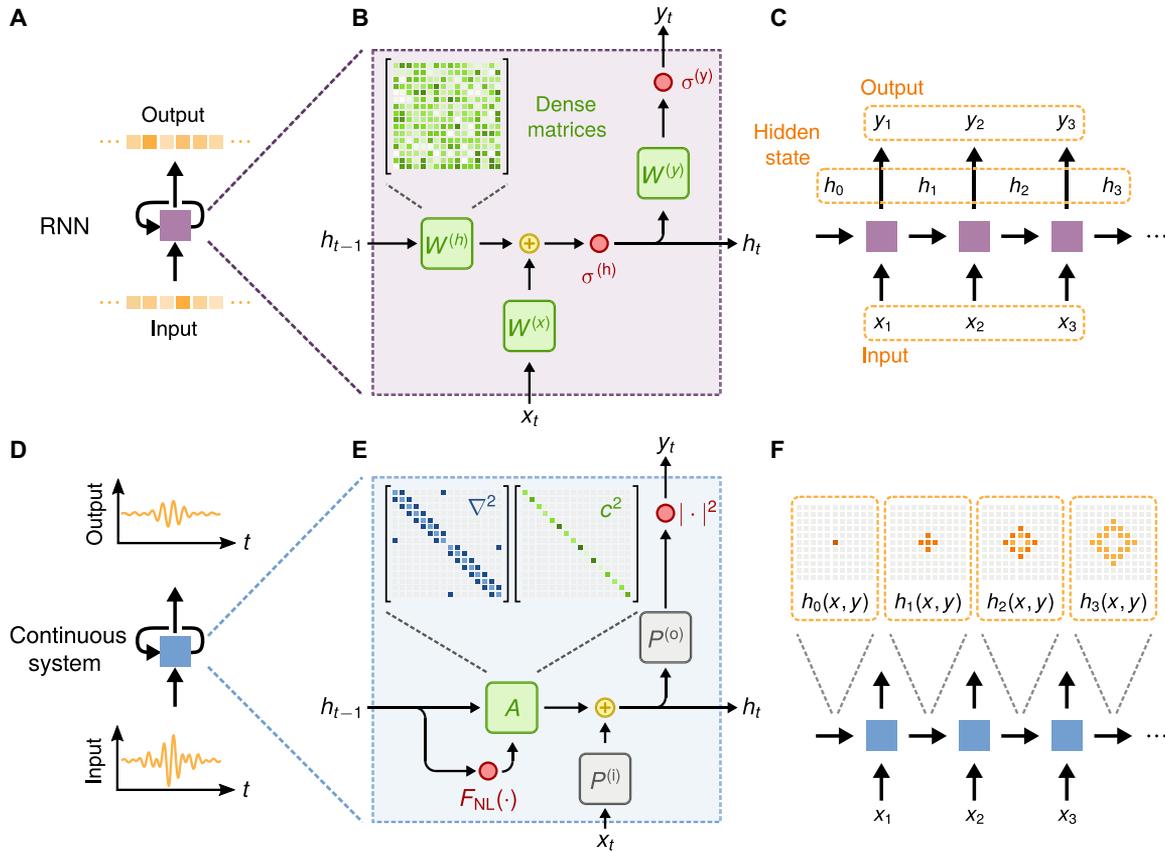
Copyright © 2019  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
NonCommercial  
License 4.0 (CC BY-NC).

<sup>1</sup>Department of Applied Physics, Stanford University, Stanford, CA 94305, USA.

<sup>2</sup>Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: shanhui@stanford.edu



**Fig. 1. Conceptual comparison of a standard RNN and a wave-based physical system.** (A) Diagram of an RNN cell operating on a discrete input sequence and producing a discrete output sequence. (B) Internal components of the RNN cell, consisting of trainable dense matrices  $W^{(h)}$ ,  $W^{(x)}$ , and  $W^{(y)}$ . Activation functions for the hidden state and output are represented by  $\sigma^{(h)}$  and  $\sigma^{(y)}$ , respectively. (C) Diagram of the directed graph of the RNN cell. (D) Diagram of a recurrent representation of a continuous physical system operating on a continuous input sequence and producing a continuous output sequence. (E) Internal components of the recurrence relation for the wave equation when discretized using finite differences. (F) Diagram of the directed graph of discrete time steps of the continuous physical system and illustration of how a wave disturbance propagates within the domain.

given by the flattened fields,  $u_t$  and  $u_{t-1}$ , represented on a discretized grid over the spatial domain. Then, the update of the wave equation may be written as

$$\mathbf{h}_t = \mathbf{A}(\mathbf{h}_{t-1}) \cdot \mathbf{h}_{t-1} + \mathbf{P}^{(i)} \cdot \mathbf{x}_t \quad (5)$$

$$\mathbf{y}_t = (\mathbf{P}^{(o)} \cdot \mathbf{h}_t)^2 \quad (6)$$

where the sparse matrix,  $\mathbf{A}$ , describes the update of the wave fields  $u_t$  and  $u_{t-1}$  without a source (Fig. 1E). The derivation of Eqs. 5 and 6 is given in section S1.

For sufficiently large field strengths, the dependence of  $\mathbf{A}$  on  $\mathbf{h}_{t-1}$  can be achieved through an intensity-dependent wave speed of the form  $c = c_{\text{lin}} + u_t^2 \cdot c_{\text{nl}}$ , where  $c_{\text{nl}}$  is exhibited in regions of material with a nonlinear response. In practice, this form of nonlinearity is encountered in a wide variety of wave physics, including shallow water waves (18), nonlinear optical materials via the Kerr effect (19), and acoustically in bubbly fluids and soft materials (20). Additional discussion and analysis of the nonlinear material responses are provided in section S2. Similar to the  $\sigma^{(y)}(\cdot)$  activation function in the standard RNN, a nonlinear re-

lationship between the hidden state,  $\mathbf{h}_t$ , and the output,  $\mathbf{y}_t$ , of the wave equation is typical in wave physics when the output corresponds to a wave intensity measurement, as we assume here for Eq. 6.

Similar to the standard RNN, the connections between the hidden state and the input and output of the wave equation are also defined by linear operators, given by  $\mathbf{P}^{(i)}$  and  $\mathbf{P}^{(o)}$ . These matrices define the injection and measuring points within the spatial domain. Unlike the standard RNN, where the input and output matrices are dense, the input and output matrices of the wave equation are sparse because they are nonzero only at the location of injection and measurement points. Moreover, these matrices are unchanged by the training process. Additional information on the construction of these matrices is given in section S3.

The trainable free parameter of the wave equation is the distribution of the wave speed,  $c(x, y, z)$ . In practical terms, this corresponds to the physical configuration and layout of materials within the domain. Thus, when modeled numerically in discrete time (Fig. 1E), the wave equation defines an operation that maps into that of an RNN (Fig. 1B). Similar to the RNN, the full time dynamics of the wave equation may be represented as a directed graph, but in contrast, the nearest-neighbor coupling enforced by the Laplacian operator leads to information propagating through the hidden state with a finite velocity (Fig. 1F).

### Training a physical system to classify vowels

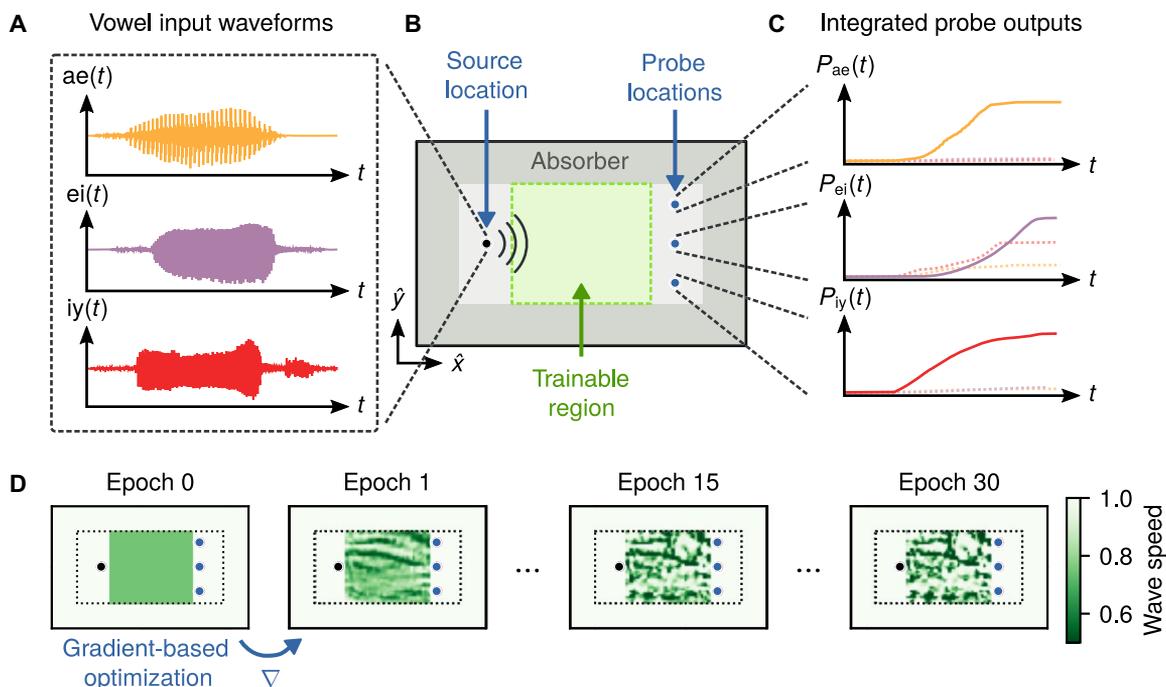
We now demonstrate how the dynamics of the wave equation can be trained to classify vowels through the construction of an inhomogeneous material distribution. For this task, we use a dataset consisting of 930 raw audio recordings of 10 vowel classes from 45 different male speakers and 48 different female speakers (21). For our learning task, we select a subset of 279 recordings corresponding to three vowel classes, represented by the vowel sounds *ae*, *ei*, and *iy*, as contained in the words *had*, *hayed*, and *heed*, respectively (Fig. 2A).

The physical layout of the vowel recognition system consists of a two-dimensional domain in the  $x$ - $y$  plane, infinitely extended along the  $z$  direction (Fig. 2B). The audio waveform of each vowel, represented by  $x^{(i)}$ , is injected by a source at a single grid cell on the left side of the domain, emitting waveforms that propagate through a central region with a trainable distribution of the wave speed, indicated by the light green region in Fig. 2B. Three probe points are defined on the right-hand side of this region, each assigned to one of the three vowel classes. To determine the system's output,  $y^{(j)}$ , we measured the time-integrated power at each probe (Fig. 2C). After the simulation evolves for the full duration of the vowel recording, this integral gives a non-negative vector of length 3, which is then normalized by its sum and interpreted as the system's predicted probability distribution over the vowel classes. An absorbing boundary region, represented by the dark gray region in Fig. 2B, is included in the simulation to prevent energy from building up inside the computational domain. The derivation of the discretized wave equation with a damping term is given in section S1.

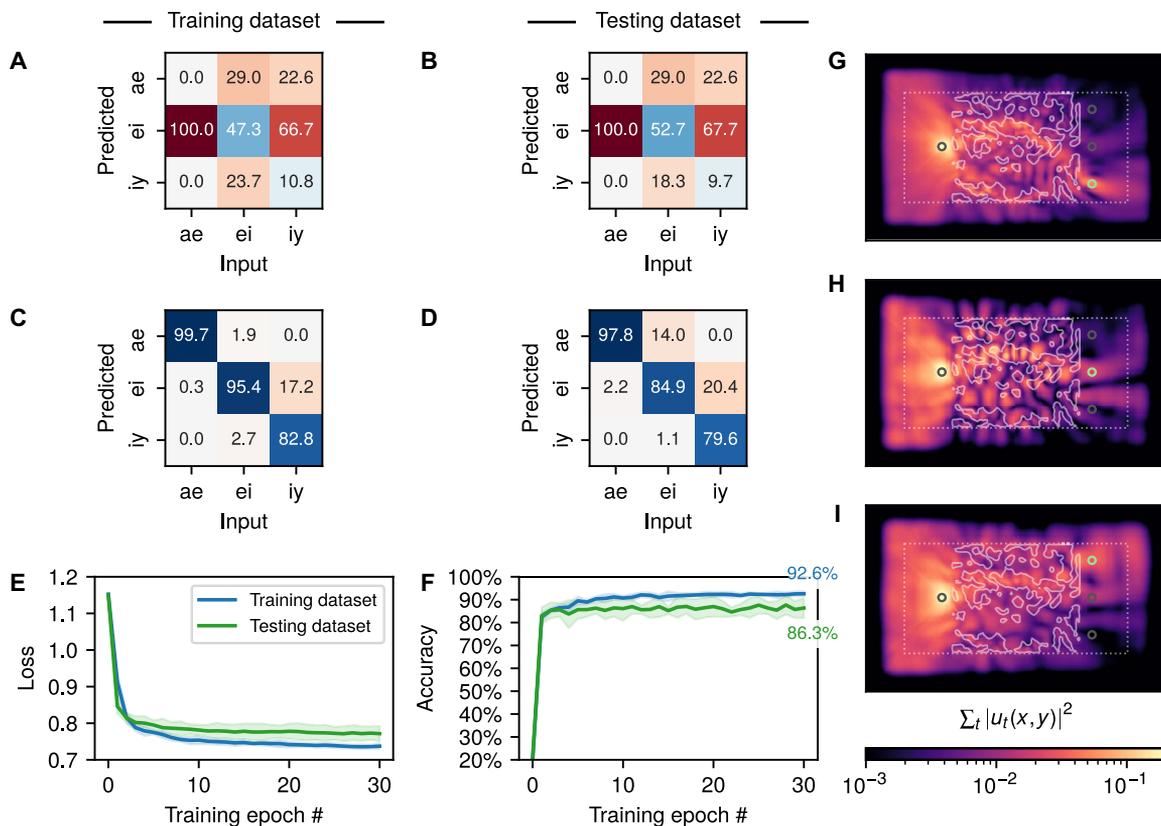
For the purposes of our numerical demonstration, we consider binarized systems consisting of two materials: a background material with

a normalized wave speed  $c_0 = 1.0$  and a second material with  $c_1 = 0.5$ . We assume that the second material has a nonlinear parameter,  $c_{nl} = -30$ , while the background material has a linear response. In practice, the wave speeds would be modified to correspond to different materials being used. For example, in an acoustic setting, the material distribution could consist of air, where the sound speed is 331 m/s, and porous silicone rubber, where the sound speed is 150 m/s (22). The initial distribution of the wave speed consists of a uniform region of material with a speed that is midway between those of the two materials (Fig. 2D). This choice of starting structure allows for the optimizer to shift the density of each pixel toward either one of the two materials to produce a binarized structure consisting of only those two materials. To train the system, we perform back-propagation through the model of the wave equation to compute the gradient of the cross-entropy loss function of the measured outputs with respect to the density of material in each pixel of the trainable region. This approach is mathematically equivalent to the adjoint method (23), which is widely used for inverse design (24–25). Then, we use this gradient information to update the material density using the Adam optimization algorithm (26), repeating until convergence on a final structure (Fig. 2D).

The confusion matrices over the training and testing sets for the starting structure are shown in Fig. 3 (A and B), averaged over five cross-validated training runs. Here, the confusion matrix defines the percentage of correctly predicted vowels along its diagonal entries and the percentage of incorrectly predicted vowels for each class in its off-diagonal entries. The starting structure cannot perform the recognition task. Figure 3 (C and D) shows the final confusion matrices after



**Fig. 2. Schematic of the vowel recognition setup and the training procedure.** (A) Raw audio waveforms of spoken vowel samples from three classes. (B) Layout of the vowel recognition system. Vowel samples are independently injected at the source, located at the left of the domain, and propagate through the center region, indicated in green, where a material distribution is optimized during training. The dark gray region represents an absorbing boundary layer. (C) For classification, the time-integrated power at each probe is measured and normalized to be interpreted as a probability distribution over the vowel classes. (D) Using automatic differentiation, the gradient of the loss function with respect to the density of material in the green region is computed. The material density is updated iteratively, using gradient-based stochastic optimization techniques until convergence.



**Fig. 3. Vowel recognition training results.** Confusion matrix over the training and testing datasets for the initial structure (A and B) and final structure (C and D), indicating the percentage of correctly (diagonal) and incorrectly (off-diagonal) predicted vowels. Cross-validated training results showing the mean (solid line) and SD (shaded region) of the (E) cross-entropy loss and (F) prediction accuracy over 30 training epochs and five folds of the dataset, which consists of a total of 279 total vowel samples of male and female speakers. (G to I) The time-integrated intensity distribution for a randomly selected input (G) ae vowel, (H) ei vowel, and (I) iy vowel.

optimization for the testing and training sets, averaged over five cross-validated training runs. The trained confusion matrices are diagonally dominant, indicating that the structure can indeed perform vowel recognition. More information on the training procedure and numerical modeling is provided in Materials and Methods.

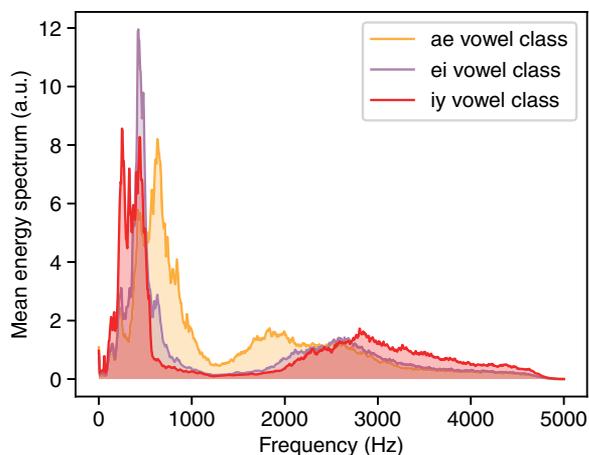
Figure 3 (E and F) shows the cross-entropy loss value and the prediction accuracy, respectively, as a function of the training epoch over the testing and training datasets, where the solid line indicates the mean and the shaded regions correspond to the SD over the cross-validated training runs. We observe that the first epoch results in the largest reduction of the loss function and the largest gain in prediction accuracy. From Fig. 3F, we see that the system obtains a mean accuracy of  $92.6 \pm 1.1\%$  over the training dataset and a mean accuracy of  $86.3 \pm 4.3\%$  over the testing dataset. From Fig. 3 (C and D), we observe that the system attains near-perfect prediction performance on the ae vowel and is able to differentiate the iy vowel from the ei vowel but with less accuracy, especially in unseen samples from the testing dataset. Figure 3 (G to I) shows the distribution of the integrated field intensity,  $\sum_t |u_t|^2$ , when a representative sample from each vowel class is injected into the trained structure. We thus provide visual confirmation that the optimization procedure produces a structure that routes most of the signal energy to the correct probe. As a performance benchmark, a conventional RNN was trained on the same task, achieving comparable classification accuracy to that of the wave equation. However, a larger number of free parameters were required. In addition, we observed that a comparable classification accuracy was obtained

when training a linear wave equation. More details on the performance are provided in section S4.

## DISCUSSION

The wave-based RNN presented here has a number of favorable qualities that make it a promising candidate for processing temporally encoded information. Unlike the standard RNN, the update of the wave equation from one time step to the next enforces a nearest-neighbor coupling between elements of the hidden state through the Laplacian operator, which is represented by the sparse matrix in Fig. 1E. This nearest-neighbor coupling is a direct consequence of the fact that the wave equation is a hyperbolic partial differential equation in which information propagates with a finite velocity. Thus, the size of the analog RNN's hidden state, and therefore its memory capacity, is directly determined by the size of the propagation medium. In addition, unlike the conventional RNN, the wave equation enforces an energy conservation constraint, preventing unbounded growth of the norm of the hidden state and the output signal. In contrast, the unconstrained dense matrices defining the update relationship of the standard RNN lead to vanishing and exploding gradients, which can pose a major challenge for training traditional RNNs (27).

We have shown that the dynamics of the wave equation are conceptually equivalent to those of an RNN. This conceptual connection opens up the opportunity for a new class of analog hardware platforms, in which evolving time dynamics play a major role in both the physics



**Fig. 4. Frequency content of the vowel classes.** The plotted quantity is the mean energy spectrum for the ae, ei, and iy vowel classes. a.u., arbitrary units.

and the dataset. While we have focused on the most general example of wave dynamics, characterized by a scalar wave equation, our results can be readily extended to other wave-like physics. Such an approach of using physics to perform computation (9, 28–32) may inspire a new platform for analog machine learning devices, with the potential to perform computation far more naturally and efficiently than their digital counterparts. The generality of the approach further suggests that many physical systems may be attractive candidates for performing RNN-like computations on dynamic signals, such as those in optics, acoustics, or seismics.

## MATERIALS AND METHODS

### Training method for vowel classification

The procedure for training the vowel recognition system was carried out as follows. First, each vowel waveform was downsampled from its original recording, with a 16-kHz sampling rate, to a sampling rate of 10 kHz. Next, the entire dataset of (3 classes)  $\times$  (45 males + 48 females) = 279 vowel samples was divided into five groups of approximately equal size.

Cross-validated training was performed with four of the five sample groups forming a training set and one of the five sample groups forming a testing set. Independent training runs were performed with each of the five groups serving as the testing set, and the metrics were averaged over all training runs. Each training run was performed for 30 epochs using the Adam optimization algorithm (26) with a learning rate of 0.0004. During each epoch, every sample vowel sequence from the training set was windowed to a length of 1000, taken from the center of the sequence. This limits the computational cost of the training procedure by reducing the length of the time through which gradients must be tracked.

All windowed samples from the training set were run through the simulation in batches of nine, and the categorical cross-entropy loss was computed between the output probe probability distribution and the correct one-hot vector for each vowel sample. To encourage the optimizer to produce a binarized distribution of the wave speed with relatively large feature sizes, the optimizer minimizes this loss function with respect to a material density distribution,  $\rho(x, y)$ , within a central region of the simulation domain, indicated by the green region in Fig. 2B. The distribution of the wave speed,  $c(x, y)$ , was computed by first applying a low-pass spatial filter and then a

projection operation to the density distribution. The details of this process are described in section S5. Figure 2D illustrates the optimization process over several epochs, during which the wave velocity distribution converges to a final structure. At the end of each epoch, the classification accuracy was computed over both the testing and training sets. Unlike the training set, the full length of each vowel sample from the testing set was used.

The mean energy spectrum of the three vowel classes after downsampling to 10 kHz is shown in Fig. 4. We observed that the majority of the energy for all vowel classes is below 1 kHz and that there is a strong overlap between the mean peak energy of the ei and iy vowel classes. Moreover, the mean peak energy of the ae vowel class is very close to the peak energy of the other two vowels. Therefore, the vowel recognition task learned by the system is nontrivial.

### Numerical modeling

Numerical modeling and simulation of the wave equation physics were performed using a custom package written in Python. The software was developed on top of the popular machine learning library, PyTorch (33), to compute the gradients of the loss function with respect to the material distribution via reverse-mode automatic differentiation. In the context of inverse design in the fields of physics and engineering, this method of gradient computation is commonly referred to as the adjoint variable method and has a computational cost of performing one additional simulation. We note that related approaches to numerical modeling using machine learning frameworks have been proposed previously for full-wave inversion of seismic datasets (34). The code for performing numerical simulations and training of the wave equation, as well as generating the figures presented in this paper, may be found online at [www.github.com/fancompute/wavetorch/](http://www.github.com/fancompute/wavetorch/).

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/12/eaay6946/DC1>

Section S1. Derivation of the wave equation update relationship

Section S2. Realistic physical platforms and nonlinearities

Section S3. Input and output connection matrices

Section S4. Comparison of wave RNN and conventional RNN

Section S5. Binarization of the wave speed distribution

Fig. S1. Saturable absorption response.

Fig. S2. Cross-validated training results for an RNN with a saturable absorption nonlinearity.

Table S1. Comparison of a scalar wave model and a conventional RNN on a vowel recognition task.

References (35–43)

### REFERENCES AND NOTES

- O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
- I. Sutskever, O. Vinyals, Q. V. Le, Sequence to sequence learning with neural networks, in *Advances in Neural Information Processing Systems, NIPS Proceedings, Montreal, CA, 2014*.
- J. M. Shainline, S. M. Buckley, R. P. Mirin, S. W. Nam, Superconducting optoelectronic circuits for neuromorphic computing. *Phys. Rev. Appl.* **7**, 034013 (2017).
- A. N. Tait, T. F. D. Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, P. R. Prucnal, Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
- M. Romera, P. Talatchian, S. Tsunegi, F. Abreu Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz, J. Grollier, Vowel recognition with four coupled spin-torque nano-oscillators. *Nature* **563**, 230–234 (2018).

6. 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).
7. J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, S. Lloyd, Quantum machine learning. *Nature* **549**, 195–202 (2017).
8. F. Laporte, A. Katumba, J. Dambre, P. Bienstman, Numerical demonstration of neuromorphic computing with photonic crystal cavities. *Opt. Express* **26**, 7955–7964 (2018).
9. 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).
10. E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, Z. Yu, Nanophotonic media for artificial neural inference. *Photon. Res.* **7**, 823–827 (2019).
11. K. Yao, G. Zweig, M.-Y. Hwang, Y. Shi, D. Yu, Recurrent neural networks for language understanding (Interspeech, 2013), pp. 2524–2528; <https://www.microsoft.com/en-us/research/publication/recurrent-neural-networks-for-language-understanding/>.
12. M. Hüskens, P. Stagge, Recurrent neural networks for time series classification. *Neurocomputing* **50**, 223–235 (2003).
13. G. Dorffner, Neural networks for time series processing. *Neural Net. World* **6**, 447–468 (1996).
14. J. T. Connor, R. D. Martin, L. E. Atlas, Recurrent neural networks and robust time series prediction. *IEEE Trans. Neural Netw.* **5**, 240–254 (1994).
15. J. L. Elman, Finding structure in time. *Cognit. Sci.* **14**, 179–211 (1990).
16. M. I. Jordan, Serial order: A parallel distributed processing approach. *Adv. Physcol.* **121**, 471–495 (1997).
17. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
18. F. Ursell, The long-wave paradox in the theory of gravity waves. *Math. Proc. Camb. Philos. Soc.* **49**, 685–694 (1953).
19. R. W. Boyd, *Nonlinear Optics* (Academic Press, 2008).
20. T. Rossing, *Springer Handbook of Acoustics* (Springer Science & Business Media, 2007).
21. J. Hillenbrand, L. A. Getty, M. J. Clark, K. Wheeler, Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* **97**, 3099–3111 (1995).
22. A. Ba, A. Kovalenko, C. Aristégui, O. Mondain-Monval, T. Brunet, Soft porous silicone rubbers with ultra-low sound speeds in acoustic metamaterials. *Sci. Rep.* **7**, 40106 (2017).
23. T. W. Hughes, M. Minkov, Y. Shi, S. Fan, Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **5**, 864–871 (2018).
24. S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, A. W. Rodriguez, Inverse design in nanophotonics. *Nat. Photonics* **12**, 659–670 (2018).
25. Y. Elesin, B. S. Lazarov, J. S. Jensen, O. Sigmund, Design of robust and efficient photonic switches using topology optimization. *Photonics Nanostruct. Fund. Appl.* **10**, 153–165 (2012).
26. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG] (22 December 2014).
27. L. Jing, Y. Shen, T. Dubcek, J. Peurifoy, S. Skirlo, Y. LeCun, M. Tegmark, M. Soljačić, Tunable Efficient Unitary Neural Networks (EUNN) and their application to RNNs, *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (JMLR.org, 2017), pp. 1733–1741.
28. A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alù, N. Engheta, Performing mathematical operations with metamaterials. *Science* **343**, 160–163 (2014).
29. M. Hermans, M. Burm, T. Van Vaerenbergh, J. Dambre, P. Bienstman, Trainable hardware for dynamical computing using error backpropagation through physical media. *Nat. Commun.* **6**, 6729 (2015).
30. C. Guo, M. Xiao, M. Minkov, Y. Shi, S. Fan, Photonic crystal slab Laplace operator for image differentiation. *Optica* **5**, 251–256 (2018).
31. H. Kwon, D. Sounas, A. Cordaro, A. Polman, A. Alù, Nonlocal metasurfaces for optical signal processing. *Phys. Rev. Lett.* **121**, 173004 (2018).
32. N. M. Estakhri, B. Edwards, N. Engheta, Inverse-designed metastructures that solve equations. *Science* **363**, 1333–1338 (2019).
33. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in PyTorch, Workshop on Autodiff (NIPS, 2017).
34. A. Richardson, Seismic full-waveform inversion using deep learning tools and techniques. arXiv:1801.07232 [physics.geo-ph] (22 January 2018).
35. A. F. Oskooi, L. Zhang, Y. Avniel, S. G. Johnson, The failure of perfectly matched layers, and towards their redemption by adiabatic absorbers. *Opt. Express* **16**, 11376–11392 (2008).
36. W. C. Elmore, M. A. Heald, *Physics of Waves* (Courier Corporation, 2012).
37. M. R. Lamont, B. Luther-Davies, D.-Y. Choi, S. Madden, B. J. Eggleton, Supercontinuum generation in dispersion engineered highly nonlinear ( $\gamma = 10$  /W/m)  $As_2S_3$  chalcogenide planar waveguide. *Opt. Express* **16**, 14938–14944 (2008).
38. N. Hartmann, G. Hartmann, R. Heider, M. Wagner, M. Ilchen, J. Buck, A. Lindahl, C. Benko, J. Grünert, J. Krzywinski, J. Liu, A. A. Lutman, A. Marinelli, T. Maxwell, A. A. Miahnahri, S. P. Moeller, M. Planas, J. Robinson, A. K. Kazansky, N. M. Kabachnik, J. Viefhaus, T. Feurer, R. Kienberger, R. N. Coffee, W. Helml, Attosecond time–energy structure of x-ray free-electron laser pulses. *Nat. Photonics* **12**, 215–220 (2018).
39. X. Jiang, S. Gross, M. J. Withford, H. Zhang, D.-I. Yeom, F. Rotermund, A. Fuerbach, Low-dimensional nanomaterial saturable absorbers for ultrashort-pulsed waveguide lasers. *Opt. Mater. Express* **8**, 3055–3071 (2018).
40. R. E. Christiansen, O. Sigmund, Experimental validation of systematically designed acoustic hyperbolic meta material slab exhibiting negative refraction. *Appl. Phys. Lett.* **109**, 101905 (2016).
41. G. F. Pinton, J. Dahl, S. Rosenzweig, G. E. Trahey, A heterogeneous nonlinear attenuating full-wave model of ultrasound. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56**, 474–488 (2009).
42. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
43. J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555 [cs.NE] (11 December 2014).

#### Acknowledgments

**Funding:** This work was supported by a Vannevar Bush Faculty Fellowship from the U.S. Department of Defense (N00014-17-1-3030), by the Gordon and Betty Moore Foundation (GBMF4744), by a MURI grant from the U.S. Air Force Office of Scientific Research (FA9550-17-1-0002), and by the Swiss National Science Foundation (P300P2\_177721). **Author contributions:** T.W.H. conceived the idea and developed it with I.A.D.W., with input from M.M. The software for performing numerical simulations and training of the wave equation was developed by I.A.D.W. with input from T.W.H. and M.M. The numerical model for the standard RNN was developed and trained by M.M. S.F. supervised the project. All authors contributed to analyzing the results and writing the manuscript. **Competing interests:** T.W.H., I.A.D.W., M.M., and S.F. are inventors on a patent application related to this work filed by Leland Stanford Junior University (no. 62/836328, filed 19 April 2019). The authors declare no other 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. Additional data related to this paper may be requested from the authors. The code for performing numerical simulations and training of the wave equation, as well as generating the figures presented in this paper are available online at [www.github.com/fancompute/wavetorch/](http://www.github.com/fancompute/wavetorch/).

Submitted 10 July 2019

Accepted 29 October 2019

Published 20 December 2019

10.1126/sciadv.aay6946

**Citation:** T. W. Hughes, I. A. D. Williamson, M. Minkov, S. Fan, Wave physics as an analog recurrent neural network. *Sci. Adv.* **5**, eaay6946 (2019).