A photonic chip-based machine learning approach for the prediction of molecular properties

Jonathan Wei Zhong Lau,1,* Hui Zhang,2,* Lingxiao Wan,2 Liang Shi,3 Hong Cai,4 Xianshu Luo,5 Patrick Lo,5 Chee-Kong Lee,6 † Leong-Chuan Kwek,1,7,8,9,‡ and Ai Qun Liu2,§

1Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543

2Quantum Science and Engineering Centre (QSec), Nanyang Technological University, 50 Nanyang Ave, 639798 Singapore, Singapore.

3Chemistry and Biochemistry, University of California, Merced, California 95343, United States

4Institute of Microelectronics, A*STAR (Agency for Science, Technology and Research), 138634, Singapore

5Advanced Micro Foundry, 11 Science Park Road, 117685 Singapore

6Tencent America, Palo Alto, CA 94306, United States

7MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore

8National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616

9School of Electrical and Electronic Engineering Block S2.1, 50 Nanyang Avenue, Singapore 639798
Abstract

Machine learning methods have revolutionized the discovery process of new molecules and materials. However, the intensive training process of neural networks for molecules with ever increasing complexity has resulted in exponential growth in computation cost, leading to long simulation time and high energy consumption. Photonic chip technology offers an alternative platform for implementing neural network with faster data processing and lower energy usage compared to digital computers. Here, we demonstrate the capability of photonic neural networks in predicting the quantum mechanical properties of molecules. Additionally, we show that multiple properties can be learned simultaneously in a photonic chip via a multi-task regression learning algorithm, which we believe is the first of its kind, as most previous works focus on implementing a network for the task of classification. Photonics technology are also naturally capable of implementing complex-valued neural networks at no additional hardware cost and we show that such neural networks outperform conventional real-valued networks for molecular property prediction. Our work opens the avenue for harnessing photonic technology for large-scale machine learning applications in molecular sciences such as drug discovery and materials design.

INTRODUCTION

Data-driven approaches, in particular machine learning methods, have in the past decade become an indispensable tool for material design and molecular discovery [1, 2]. The availability of large datasets, both experimental and computational, and the advancements in algorithms and computer capability significantly accelerate the screening and identification of molecules and materials with desired properties in vast chemical space. Among the machine learning methods, deep learning methods involving large and sophisticated neural networks have emerged as the leading candidates for molecular property prediction. However, despite their promising performance, training of such deep neural networks inadvertently leads to substantial computational cost associated with expensive hardware and high power consumption. For example, the highly successful recent work, Alphafold [3], for protein folding typically requires training times in the order of weeks using hundreds of costly graphic processing units (GPUs) [4]. Furthermore, while the computational costs of machine learning models continue to double every few months [5], the efficiency of digital computational
chips has not managed to improve at a similar pace [6]. Thus, training of ever more sophisticated machine learning models leads to exponential growth of energy consumption [7], which comes with a huge environmental cost [8].

Optical computing offers an alternative paradigm to conventional digital computing for computation-intensive tasks, such as implementing deep neural networks. Optical computing implementations of machine learning methods have been realized in neuromorphic photonics [9, 10], all-optical neural networks [11, 12] and photonic reservoir computing [13, 14]. In fact, optical implementation of neural networks has recently been demonstrated to surpass cutting-edge GPUs in terms of speed and energy consumption. Optical computing offers several advantages over conventional digital computers, such as low power usage [15–17], ultrafast optoelectronics with high noise robustness [18], inherent parallelism [19], and large information storage [20]. Furthermore, optical neural networks implementing complex-valued neural networks have been developed recently [21]. Despite the promises of optical neural network, its potential in molecular sciences such as chemistry and biology has not been explored. Additionally, most current works on optical neural networks focus on single-task learning (STL) method. Yet, humans typically acquire knowledge and arrive at generalization through multi-task learning (MTL). In the study of molecules, we often encounter the need to predict the values of many properties (e.g. free energy and enthalpy) as the most desirable molecules or materials typically optimize multiple properties simultaneously. This sort of multi-task learning differs fundamentally from single-task classification and requires modified algorithms. Instead of producing a category code to categorize a set of input data into a specific class, we seek to produce and optimize a continuous real-valued output [22]. It also requires us to be able to learn multiple properties (tasks) simultaneously in hope that we can leverage the knowledge contained in one task to generalize and improve performance of other tasks [23, 24]. With this motivation in mind, we hope to implement multi-task regression neural network within the optical scheme.

Furthermore, molecular properties are inherently quantum mechanical and its calculations involve complex-valued arithmetics, it would therefore be appealing to attempt molecular property prediction with complex-valued neural networks, a function that can be accomplished on an optical chip with no additional hardware cost.

In this work, we apply an optical neural chip to the task of implementing a complex-valued neural network to predict multiple quantum mechanical properties of molecules.
Specifically, the contributions of our work are of three folds. Firstly, it constitutes the first application of optical machine learning methods to molecular sciences. Secondly, previous works of optical neural networks only consider single-task classification learning, here we extend the optical neural networks implementation to a different class of machine learning tasks, namely multi-task regression learning, and obtain satisfactory accuracy. Lastly, we demonstrate that complex-value neural networks offer superior performance for molecular property prediction as compared to the conventional real-value neural networks.

RESULTS

Molecular Database and Machine Learning Model

In this work, we use a database consisting of computed molecular properties, the QM9 database [25], to train a photonic neural network. The database contains approximately 134k molecules that consist of up to 9 non-hydrogen heavy atoms (carbon, oxygen, nitrogen and fluorine). Some notable entries in this database are small amino acids, such as GLY and ALA, and pharmaceutically relevant organic materials, such as pyruvic acid, piperazine and hydroxy urea. The molecular geometries and chemical properties in the database are calculated by density functional theory methods.

For this dataset, we encoded the chemical structure of the molecule in a Coulomb matrix (see Methods for details). Since the largest molecules in the database contain up to 29 atoms (including hydrogen), the Coulomb matrix for each molecule is of size $29 \times 29$, and zero padding is applied to the molecules with fewer than 29 atoms. As we will elaborate later, for our current optical chip we are limited to 16 real feature inputs, which are then subsequently encoded in 8 complex inputs on the chip. To reduce the dimensionality of the Coulomb matrix to 16, we adopted the eigenspectrum representation of the Coulomb matrix. For each Coulomb matrix, we diagonalized it to find its 29 eigenvalues, and only took the largest 16 eigenvalues as our feature set for each molecule. Our simulations show that even with a restricted subset of the eigenvalues, the dataset is still rich enough to support accurate learning of the chosen chemical properties (See Figure 4 in our Methods). We focus on the predictions of three molecular properties: enthalpy, internal energy, and free energy. More details about our model, such as label scaling and model validation, can
be found in Methods.

**Details of Optical Chip**

Diagrams of the optical chip are shown in Figure 1. Similar to a typical neural network implemented on a classical computer, it comprises of an input layer, multiple hidden layers, and an output layer. In an optical neural chip (ONC), light signals are encoded and manipulated by controlling their magnitude and phase. The Maxwell equations that describe such systems of interacting wave guides are inherently complex in nature, naturally giving rise to a complex arithmetic. Usually an arbitrarily complex transformation matrix $W$ can be decomposed into $W = UDV$ via Singular value decomposition, where $U$ and $V$ are unitary and $D$ is a diagonal matrix. For this work, we consider the transformations that can be expressed as $W = UD$. This is not universal, but has a large searching space compared to just a single unitary matrix, and experimentally meets our requirements for the prediction of molecular properties.

The ONC chip implements this transformation by acting as a multiport interferometer. Multiport interferometers implement linear transformation between several optical channels by arranging many Mach-Zehnder interferometers (MZIs) in a specific pattern (a triangular mesh) [26, 27], and it can be shown that any discrete unitary operator can be decomposed and implemented in such a way [28]. Each MZI consists of two beam splitter (BS) - phase shifter (PS) pairs. By setting the transmissivity of both BS to be fixed at 50 : 50, one MZI implements the complex unitary transformation between the 2 neighbouring wave guides indexed by $p$ and $q$:

$$
T_{p,q}(\theta, \phi) = \begin{bmatrix}
1 & 0 & \ldots & \ldots & \ldots & \ldots \\
0 & 1 & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & e^{i\phi} \sin \theta & e^{i\phi} \cos \theta & \ldots & \ldots \\
\ldots & \ldots & \cos \theta & -\sin \theta & \ldots & \ldots \\
\ldots & \ldots & \ldots & 1 & 0 & \ldots \\
\ldots & \ldots & \ldots & 0 & 1 & \ldots 
\end{bmatrix}, \quad (1)
$$

which is essentially a rotation matrix. The PS can be thermally modulated to tune their phase shifts, allowing us to reconfigure the chip parameters without requiring complicated
modifications. In our case, all the PSs are thermally tuned with integrated titanium nitride (TiN) heaters bonded to a PCB.

The MZIs on a chip have different functionalities, which are related to the 3 different layers of a typical neural network. A coherent laser is used to generate the input signals by injecting it into one of the wave guides. The first (left most) layer of MZIs are then used to divide and modulate the signal into the rest of the wave guides, and are used to encode the input data into the ONC. For complex-valued inputs, this amounts to modulating both the magnitude and relative phases of the light signals in each wave guide. After the encoding of the input data, the bulk of the MZIs are next used to implement the optical neural network, or in other words, the trainable transformation matrix. A \( N \)-mode network transforms the input state into an output state by implementing an arbitrary unitary transformation \( U \) with complex weights. \( U \) is implemented with multiple rotation matrices \( \{ T_{p,q} \} \) and a diagonal matrix \( D \),

\[
U = \prod_{p=2}^{N} \prod_{q=1}^{p-1} T_{p,q} D. 
\]

\( D \) is a diagonal matrix with complex elements with a modulus equal to one on the diagonal, implementing a bias on the outputs, and can be implemented in an interferometer with individual phase shifts on the wave guides at the output end.

The optical neural is thus capable of implementing a complex-valued neural network, which is similar to a conventional real-valued neural network that is implemented on classical computers, with the only difference being that it relies on complex arithmetic. For a system with \( N \) wave guides and an arrangement of MZIs described in [26–28], this has the effect of implementing a hidden layer of \( N \) complex-valued neurons. The output of such a neuron can be expressed as:

\[
y = f \left( \sum_{i=1}^{N} \omega_i x_i + b \right), \tag{2}
\]

where the weights \( \omega_i \) and bias \( b \) are complex numbers and are determined by the specific choices of \( \theta_s \) and \( \phi_s \) of the MZIs in the ONC chip.

The last (right most) layer of MZIs can be used to implement the output layer by enabling coherent detection. The light signals at the end of the neural network contain both magnitude and phase information, and we would optimally want to access both, rather than just the magnitude information. By additional two columns of MZIs that interfere any two adjacent output ports, we are able to determine the phase of the output signals by varying
FIG. 1. Design and fabrication of the optical chip. a, Architecture of an 8-mode optical neural network chip. The optical neural network is divided into a few layers. The first layer is the input layer, which comprises of a laser followed by a number of MZIs that distribute the reference light among the wave guides and encodes the input into the network by manipulating the phase and magnitude of the light in each waveguide. The next layer comprising the bulk of the ONC is the layer that implements the trainable transformation matrix. The last layer is the output layer that in our cases comprises of photon detectors, which enable us to implement the non-linear absolute activation function by detecting the intensity of light in each output wave guide. b, The packaging of the optical chip. c, A false-color micrograph of the optical chips MZI network with integrated heaters. All 3 layers required for the ONC to function are fabricated on one chip. This chip comprises of 8 wave guides and 56 PS for the trainable transformation matrix. The chip is wire-bonded to a circuit board that provides independent control of each PS by an electronic current driver.
the phase shifter placed at one of the paths. The choice of our detection method is related to
the activation function we want to choose to implement. In this work, we implement a type
of absolute activation function \( M(z) = ||z|| \), and thus we only require intensity detection of
the output light signals. If we instead chose to implement a type of complex Rectified Linear
Unit \([29]\) activation function \( \text{ModReLU}(z) = \text{ReLU}(||z|| + b)e^{i\theta z} \), we would instead adopt
coherent detection. The detection is versatile by having more channels, such that ONCs are
extremely amenable to multitask learning, or the practice of training a network on multiple
labels simultaneously.

In our case, we used an ONC comprising of 8 modes and 56 PS, to simultaneously learn
3 labels (enthalpy, free energy, and internal energy). This limits us to 16 real inputs, and
the reason we only used the largest 16 eigenvalues as our feature set for our data. Our
model essentially implements a single complex-valued \( 8 \times 8 \) weight matrix representing 8
complex-valued neurons (given in Equation 2), followed by an absolute activation function
when reading out. The Supplementary Information contains results on using a neural net-
work on a digital computer to learn these 3 labels when only given access to the largest 16
eigenvalues, and they indicate that the 16 largest eigenvalues are sufficiently rich in informa-
tion about the underlying system to support accurate learning of these chemical properties.
The Supplementary Information also contains comparisons of a similar complex-valued and
real-valued neural network on the data set, and suggests that a complex-valued network is
superior at learning than a similar real-valued network.

**Experimental result**

First, we discuss the neural networks trained independently for enthalpy, free energy,
and internal energy. The neural network training process with the genetic algorithm (GA)
is illustrated in Fig.2, using free energy as the example. The entire training process is
visualised in Fig.2a, and show that the optical neural network is capable of generating good
successive generations by crossover (see Methods for more details). The evolution of the
best and mean cost function values with the GA iteration are shown in Fig. 2b and c,
respectively. It is clear that both values reach plateaus after 20 iterations, indicating the
high efficiency of the GA algorithm in optimizing optical neural networks. From the parity
plot in Fig. 2d, good correlations are seen between the reference and predicted values of free
|                  | Enthalpy | Free Energy | Internal Energy |
|------------------|----------|-------------|-----------------|
| Train MAE (Ha)   | 23.22    | 25.49       | 23.23           |
| Train $R^2$      | 0.9472   | 0.9411      | 0.9299          |
| Test MAE (Ha)    | 127.9    | 50.13       | 56.0            |
| Test $R^2$       | 0.9225   | 0.9325      | 0.9303          |

TABLE I. Results from the experimental chip. Labels were learnt independently. As can be seen, even for small testing and training data sets, the ONC is able to obtain high coefficients of determination.

energy for both training and test datasets, even with such small neural network and training dataset. The training processes are similar for enthalpy and internal energy, as shown in the Supplementary Information.

The quantitative performances of the neural networks for the three thermodynamic properties are summarized in Table I. While they are greater than MAEs in the literature, such comparison are not as meaningful given the differences in model complexity and training data size. In the table, we compare the MAEs to a similarly complex neural network on a digital computer, with a similarly small data size. Even with this small data set, our results demonstrate that our ONC possesses a strong ability to implement a complex valued neural network model. Already they are able to get impressive coefficients of determination ($R^2$) values, and indicate that the model has strong generalization as the coefficients of determinations between the test and train sets are comparable.

Table II and Figure 3 show the results when we learned the labels concurrently (multitask learning). We note that our results for the multitask learning are similar to the case where we learned the labels separately. Furthermore, we emphasize that no additional resources were required to perform multitask learning, as the chip inherently implements parallel learning of labels, as long as we have enough photon detectors.
FIG. 2. Experimental regression results of the free energy. A total of 100 training instances and 30 testing instances are used for the chip experiment. An initial population of 50 individuals is randomly generated. 

a, The visualization of the on-chip training process. The x-axis represents training iterations, i.e., the generations in the Genetic Algorithm. The y-axis represents the 50 individuals in each generation. On the y-axis, the rows 1-3 are the elites directly inherited from the previous generation, the rows 4-41 are the children created by crossover, and the rows 42-50 are children created by the mutation operator. Different colors represent different cost function values, while blue represents smaller cost value and yellow the opposite. From this figure, we can see the evolution of performance of chip configurations.

b, The evolution of the best cost function value in each generation, with increasing generations.

c, The evolution of the mean cost function value in each generation, with increasing generations.

d, The regression curve of the trained model on training and testing samples. The 100 training instances (blue data points) are evaluated on chip after training, to validate whether the training is successful, and the 30 test instances (orange data points) measure the generalizability of the trained model. The regression quality is measured by the coefficient of determination $R^2$ score (the best regression has a $R^2$ score of 1). The training $R^2$ of our trained model on chip is 0.9411, and the testing score is 0.9325.
|                      | Enthalpy | Free Energy | Internal Energy |
|----------------------|----------|-------------|-----------------|
| ONC Train MAE (Ha)   | 27.52    | 25.33       | 23.33           |
| ONC Train $R^2$      | 0.9449   | 0.9307      | 0.9294          |
| ONC Test MAE (Ha)    | 130.70   | 45.67       | 59.06           |
| ONC Test $R^2$       | 0.9237   | 0.9051      | 0.8974          |
| DCC Train MAE (Ha)   | 17.34    | 11.90       | 13.38           |
| DCC Train $R^2$      | 0.9624   | 0.9838      | 0.9812          |
| DCC Test MAE (Ha)    | 176.92   | 165.07      | 160.41          |
| DCC Test $R^2$       | 0.9123   | 0.8912      | 0.9214          |

TABLE II. Results from the experimental chip, compared to results from a digital classical computer. The same data set was used for comparison between the ONC and the digital classical computer. Labels were learnt concurrently (multitask learning). It is worth noting that to perform multitask learning on the chip, no additional computational resources were needed as the chip inherently implements parallel learning of labels. The only requirement was for more photon detectors (using absolute activation function, required one detector per label). The ONC rows indicate the results obtained experimentally from the optical neural chip, while the DCC rows indicate results from a digital classical computer. Interestingly, while the digital classical computer was more accurate in this case on the training set, it was much worse at generalizing the results for the test set.
FIG. 3. Experimental results obtained from the chip, when regressing on all 3 properties at once (multitask learning). a, The visualization of the training process. b, The regression curve for enthalpy. c, The regression curve for free energy. (d) The regression curve for internal energy.
DISCUSSIONS AND CONCLUSIONS

The use of machine learning models, in particular deep neural networks, has been demonstrated to be effective in learning the hidden relationships between complex, high dimensional data sets [22, 30]. These techniques provide us with a more efficient and economical way to predict molecular properties and hold enormous promise to accelerate material design and drug discovery. [31]. Despite their excellent performance, the complicated structures necessary for such networks, especially for performing more complex learning tasks will require increasingly more computing power and higher energy consumption.

Furthermore, the vast majority of such models rely on a real-valued neural network. Recent studies suggest that using complex-valued neural networks could significantly improve the performance of similar models [32] by offering richer representational capacity [33], faster convergence [34], strong generalization [35] and noise-robust memory mechanisms [36]. It has also been shown that complex valued neural networks have potential in domains where the data is naturally represented with complex numbers, or the problem is complex by design [37]. Thus, using such complex networks could also potentially be better in dealing with quantum-mechanical problems, as quantum mechanics is inherently a complex-valued theory. However, it is typically not efficient to implement a complex-valued network on a classical digital computer as complex numbers have to be represented by two real numbers on the digital computer [38, 39], which increases the computationally expensive components of the neural network algorithms [40, 41].

To overcome this, optical computing has been proposed as an alternate computing platform. The applications of optical computing to run neural networks provides various advantages over classical digital computers, ranging from low electrical power usage, being more energy efficient and robust to noise, and its inherent parallelism allowing us to break down the computation into small steps that are performed in parallel. Most interestingly, it is capable of truly complex-valued arithmetic, allowing us to implement complex-valued neural networks with no additional cost. Such complex-valued networks have been shown to also hold advantages over similar real-valued networks. However, most studies on applying optical computing to neural networks have focused on single-task classification learning. This is a different class of machine learning tasks as compared to what is usually required for applications in chemistry, which is multi-task regression learning, and it was not known
previously if such promising prior results in single-task classification would carry over.

Our work is the first known application of optical neural networks to chemistry. This is the first known application of an optical neural network on chip to optimizing a continuous valued regression task. It is also the first work to show that such a setup is capable of learning multiple properties at once (multi-task learning) at no additional cost. This also suggests that they would fare well in similar learning paradigms in machine learning, such as transfer learning and multi-output regression. It also demonstrates that complex-valued optical neural networks can achieve performance that is similar to classical real-valued neural networks implemented on digital computers, for the specific task of learning chemical properties of molecules. Although at a small scale right now, there are no physical reasons preventing such devices from being scaled up. It is also reasonable to assume that similar models will be effective in regressions tasks for other types of data sets.

By utilizing such optical chips in a manner where we offload the majority of the computational tasks of a neural network off a classical digital computer to them, we also show the potential of creating future hybrid computing systems, blending the advantages of optical and neuromorphic computing with those of classical digital computers. It is also possible to extend our ONC into a fully fledged multi-layer neural network with multiple hidden layers by cascading the optical circuits. Such cascaded photonic neural networks have also been demonstrated to be effective in performing various complicated machine learning tasks [42].

ACKNOWLEDGEMENTS

This work is supported by the Singapore Ministry of Education (MOE) Tier 3 grant (MOE2017-T3-1-001), and the Singapore National Research Foundation (NRF) National Natural Science Foundation of China (NSFC) joint grant (NRF2017NRF-NSFC002-014). JL and KLC are supported by the Ministry of Education and the National Research Foundation under the Center for Quantum Technologies (CQT).

* Equal contribution
† cheekonglee@tencent.com
‡ cqtklc@nus.edu.sg
1 Know the basics of machine learning for molecular and materials science.

2 Understand the integration of quantum-based machine learning in chemical compound space.

3 Explore high-accuracy protein structure prediction with AlphaFold.

4 Integrate AI and compute concepts for future technological advancements.

5 Conduct a survey of machine learning accelerators.

6 Learn about the systematic reporting of energy and carbon footprints in machine learning.

7 Analyze energy and policy considerations for deep learning in NLP.

8 Study machine learning with neuromorphic photonics for enhanced computing performance.

9 Investigate all-optical spiking neurosynaptic networks.

10 Examine neuromorphic photonic networks using silicon photonic weight banks.

11 Research all-optical spiking neurosynaptic networks with self-learning capabilities.

12 Explore all-optical machine learning using diffractive deep neural networks.

13 Advance in photonic reservoir computing with nanophotonic technologies.
[14] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, Experimental demonstration of reservoir computing on a silicon photonics chip, Nature communications 5, 1 (2014).

[15] H. J. Caulfield and S. Dolev, Why future supercomputing requires optics, Nature Photonics 4, 261 (2010).

[16] J. Ahn, M. Fiorentino, R. G. Beausoleil, N. Binkert, A. Davis, D. Fattal, N. P. Jouppi, M. McLaren, C. M. Santori, R. S. Schreiber, et al., Devices and architectures for photonic chip-scale integration, Applied Physics A 95, 989 (2009).

[17] A. N. Tait, Quantifying power use in silicon photonic neural networks, arXiv preprint arXiv:2108.04819 (2021).

[18] P. R. Prucnal, B. J. Shastri, T. F. de Lima, M. A. Nahmias, and A. N. Tait, Recent progress in semiconductor excitable lasers for photonic spike processing, Advances in Optics and Photonics 8, 228 (2016).

[19] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutiérrez, L. Pesquera, C. R. Mirasso, and I. Fischer, Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing, Optics express 20, 3241 (2012).

[20] M. T. Hill, H. J. Dorren, T. De Vries, X. J. Leijtens, J. H. Den Besten, B. Smalbrugge, Y.-S. Oei, H. Binsma, G.-D. Khoe, and M. K. Smit, A fast low-power optical memory based on coupled micro-ring lasers, Nature 432, 206 (2004).

[21] H. Zhang, M. Gu, X. Jiang, J. Thompson, H. Cai, S. Paesani, R. Santagati, A. Laing, Y. Zhang, M. Yung, et al., An optical neural chip for implementing complex-valued neural network, Nature Communications 12, 1 (2021).

[22] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning (MIT press, 2016).

[23] R. Caruana, Multitask learning, Machine learning 28, 41 (1997).

[24] Y. Zhang and Q. Yang, A survey on multi-task learning, arXiv preprint arXiv:1707.08114 (2017).

[25] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Scientific data 1, 1 (2014).

[26] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, Optica 3, 1460 (2016).
[27] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, et al., Universal linear optics, Science 349, 711 (2015).

[28] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Experimental realization of any discrete unitary operator, Physical review letters 73, 58 (1994).

[29] J. Brownlee, A gentle introduction to the rectified linear unit (relu), Machine learning mastery 6 (2019).

[30] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, nature 521, 436 (2015).

[31] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman, Accelerated search for materials with targeted properties by adaptive design, Nature communications 7, 1 (2016).

[32] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, Deep complex networks, arXiv preprint arXiv:1705.09792 (2017).

[33] D. P. Reichert and T. Serre, Neuronal synchrony in complex-valued deep networks, arXiv preprint arXiv:1312.6115 (2013).

[34] M. Arjovsky, A. Shah, and Y. Bengio, Unitary evolution recurrent neural networks, in International Conference on Machine Learning (PMLR, 2016) pp. 1120–1128.

[35] A. Hirose and S. Yoshida, Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence, IEEE Transactions on Neural Networks and learning systems 23, 541 (2012).

[36] I. Danihelka, G. Wayne, B. Uria, N. Kalchbrenner, and A. Graves, Associative long short-term memory, in International Conference on Machine Learning (PMLR, 2016) pp. 1986–1994.

[37] J. Bassey, L. Qian, and X. Li, A survey of complex-valued neural networks, arXiv preprint arXiv:2101.12249 (2021).

[38] I. Aizenberg, Complex-valued neural networks with multi-valued neurons, Vol. 353 (Springer, 2011).

[39] A. Yadav, D. Mishra, S. Ray, R. Yadav, and P. Kalra, Representation of complex-valued neural networks: a real-valued approach, in Proceedings of 2005 International Conference on Intelligent Sensing and Information Processing, 2005. (IEEE, 2005) pp. 331–335.
[40] H.-T. Peng, M. A. Nahmias, T. F. De Lima, A. N. Tait, and B. J. Shastri, Neuromorphic photonic integrated circuits, IEEE Journal of Selected Topics in Quantum Electronics 24, 1 (2018).

[41] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, Efficient processing of deep neural networks: A tutorial and survey, Proceedings of the IEEE 105, 2295 (2017).

[42] S. Zarei, M.-r. Marzban, and A. Khavasi, Integrated photonic neural network based on silicon metalines, Optics Express 28, 36668 (2020).

[43] L. C. Blum and J.-L. Reymond, 970 million druglike small molecules for virtual screening in the chemical universe database gdb-13, Journal of the American Chemical Society 131, 8732 (2009).

[44] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, Physical review letters 108, 058301 (2012).

[45] L. Ruddigkeit, R. Van Deursen, L. C. Blum, and J.-L. Reymond, Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17, Journal of chemical information and modeling 52, 2864 (2012).

[46] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, Machine learning of molecular electronic properties in chemical compound space, New Journal of Physics 15, 095003 (2013).

[47] A. B. Tchagang and J. J. Valdés, Prediction of the atomization energy of molecules using coulomb matrix and atomic composition in a bayesian regularized neural networks, in International Conference on Artificial Neural Networks (Springer, 2019) pp. 793–803.

[48] P. Reddy and A. B. Bhattacherjee, A hybrid quantum regression model for the prediction of molecular atomization energies, Machine Learning: Science and Technology 2, 025019 (2021).

[49] G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, A. Lilienfeld, and K.-R. Müller, Learning invariant representations of molecules for atomization energy prediction, Advances in neural information processing systems 25, 440 (2012).

[50] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques (Elsevier, 2011).

[51] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in International conference on machine learning (PMLR, 2015) pp. 448–456.
Much work has been done on figuring out how to efficiently and effectively encode the physical properties of a molecule in a form that can be fed as an input to a machine learning model. One method is to encode it in a form called the Coulomb matrix, and such methods have been routinely applied to the QM7 \[43, 44\] and QM9 \[25, 45\] data sets. While the Coulomb matrix is a low-level molecular descriptor, it is extremely simple to calculate and generate, does not require any domain knowledge, and is already able to obtain promising results \[46–48\].

The elements of the Coulomb matrix are defined by Equations 3 and 4:

\[
\begin{align*}
c_{ii} &= \frac{1}{2} Z_i^2, \\
c_{ij, i \neq j} &= \frac{Z_i Z_j}{|R_i - R_j|},
\end{align*}
\]  

where \(Z_i\) is the atomic number of atom \(i\), and \(R_i\) is its position in atomic units. Note that the Coulomb matrix is symmetric and is invariant to translation and rotation. The labeling of the atom indexes is also not unique, thus for any given molecule, many different Coulomb matrices could be used to represent it, all related to each other by permuting
rows and columns. This is commonly seen as an issue that prevents the Coulomb matrix representation from being used out-of-the-box, as many different Coulomb matrices can be associated with the same molecule. 3 workarounds are usually used for this [49]:

1. Eigenspectrum representation of the Coulomb matrix. For a $d \times d$ Coulomb matrix, the eigenspectrum representation can be obtained by solving for the eigenvalues of the Coulomb matrix under the constraint $\{\lambda_k\} > 0$, where $\{\lambda_k\}$ is the set of eigenvalues. The spectrum $(\lambda_1, \lambda_2, \ldots, \lambda_d)$, can then be used as the representation. This also conveniently reduces the dimension of the feature set, although it could potentially remove unrecoverable information about the structure of the molecule.

2. Sort the Coulomb matrix. We can choose the permutation of atoms whose associated Coulomb matrix $C$ satisfies $||C_i \geq C_{i+1}||$, where $C_i$ denotes the $i^{th}$ row of the Coulomb matrix. By doing so, we ensure that each molecule has an unique Coulomb matrix associated with it.

3. Extend the data set with randomly sorted Coulomb matrices. This involves generating a set of valid Coulomb matrices for each molecule by randomly permuting rows and columns (equivalent to randomly indexing the molecules) and extending the data set with them. Thus, in the extended data set, each molecule (and thus each set of property labels) has multiple, equally valid, Coulomb matrices associated with it.

All 3 methods have been shown to work quite effectively for regression in neural networks on the QM7 data base, which is a smaller subset of the QM9 data base which we will be using in this work.

Furthermore, as mentioned in our main text, our current experimental setup is limited to 16 real feature inputs, encoded into 8 complex inputs. Simulations indicate that the data set still retains enough information about the molecule to permit accurate learning. A comparison of how the accuracy of the machine learning model changes with the number of eigenvalue inputs is shown in Figure 4.

Finally it is worth mentioning that while the data set provides other properties like the Highest Occupied Molecular Orbital (HOMO) and the heat capacity, from our numerical simulations running similar sized real-valued and complex-valued neural networks on the data set, we found that if we are limited to the eigenvalue spectrum of the Coulomb matrix
for each molecule, the data set was not rich enough to learn those other properties accurately. However, we would like to stress that our optical implementation is general. Any descriptor of the molecule can be used, not just the Coulomb matrix. In our experiment, we utilize the Coulomb matrix as an example, but other, higher dimensional descriptors of the molecule like the molecular matrix and the 3D-MoRSE descriptors could be used too.

FIG. 4. Comparison of different number of eigenvalues used. Simulations were conducted with a simple model built using PyTorch, and assessed using Root Mean Square Error (RMSE). As can be seen, by using only 16 eigenvalues, we are already able to capture most of the details of the system, and using more eigenvalues does not drastically improve the results of the model. a, Enthalpy. b, Internal energy. c, Free energy.

**Scaling of data**

It is known that many machine learning and data-driven analytical methods perform better when the numerical features and labels are scaled to a standard range [50, 51]. In this work, we are dealing with eigenvalue features that numerically range from a few hundred to single digit numbers, and property labels that are of the order $10^5$. While the inputs features do not pose too much of a problem due to the way we construct our optical chip, the output labels do need to be scaled to a smaller range. We also want to avoid our labels having negative values, due to the method we are detecting outputs on our optical chip. Thus, we used a simple Min-Max normalization scaling technique on each label separately, so that all the labels are scaled into the range [0, 1]. This was done with this map:

$$y_{\text{scaled}} = \frac{y - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}$$  (5)
where for each label, $y_{\min}(y_{\max})$ is the minimum (maximum) value for that label in the entire data set.

**Model validation**

For our model we divided up the data set randomly into a train group consisting of $\frac{4}{5}$ths of the data, and a test group consisting of the remaining $\frac{1}{5}$th of the data. While a more systematic manner of distributing the data could be used (for example, stratified 5-fold cross validation with identical cross validation folds, used in [44]), from our simulations over many random divisions of data, this did not play a major part in affecting the accuracy of our model. We used a simple mean square error objective function for our model, and relied on easily available commercial optimization software to train the parameters in our model [52].

**Experimental set up**

The chip is fabricated at Advanced Micro Foundry, Singapore. The chip is wire-bonded to a printed circuit board, providing independent control of each phase shifter by an electronic current driver with 1-kHz frequency and 12-bit resolution (Qontrol Devices, Inc.). Laser pulses are generated by an Ultrafast Optical Clock device (PriTel Inc.) with a repetition rate of 500 MHz, central wavelength of 1550.12 nm and bandwidth of 2 nm. The input single pump light is coupled into the chip by a one-dimensional subwavelength grating coupler, and detected off chip by 8-mode grating coupler arrays. A polarization controller is utilized to maximize the coupling efficiency of the fibre to chip. A Peltier controlled by Thorlabs TED200C is used to stabilize the temperature of the chip and reduce the heat fluctuations caused by the ambient temperature and the heat crosstalk within the chip.

**Genetic algorithm**

While our previous simulations on a digital computer utilized variations of gradient descent methods, for the experiment we utilized a genetic algorithm to optimize the parameters for the ONC, to cut down on the time needed to conduct the experiment.

A genetic algorithm is a global optimization algorithm that can be summarized as so:
1. The genetic algorithm first starts out by creating a random initial population.

2. The algorithm then iteratively creates a sequence of new populations, known as generations. It does so with the following steps:

   (a) Calculates the scores of the current population with the cost function. These scores are known as the fitness scores.

   (b) Chooses a certain number of the current population with the best fitness scores. These are known as the elite and are passed on to the next generation.

   (c) Produces children from the current generation (parents), with crossovers (combining the entries of a pair of parents) and mutations (randomly making changes to parents). These children are then also passed on to the next generation.

3. The current generation is replaced with the next generation.

4. This is repeated until the stopping criteria is met. In our case, the stopping criteria is a set number of iterations/generations.

For our genetic algorithm, we conducted it over 50 training iterations (otherwise known as generations), with each generation having a population of 50. Between each generation, we selected the top 3 to survive into the next generation (the elite), then generated 47 additional samples, 39 by crossover, and 8 by mutation. Also due to time constraints, instead of utilizing the whole data set, we only utilized 100 randomly chosen data points in the data set to serve as our training data, and another 30 randomly chosen data points to serve as our test data. This implies that our results will be less accurate than what we expect. However, we believe that even with this small data set, our results demonstrate that our ONC possesses a strong ability to implement a complex valued neural network model.

Useful references can be found in [53–57].