Optically accelerated extreme learning machine using hot atomic vapors

Pierre Azam and Robin Kaiser*
Université Côte d'Azur, CNRS, Institut de Physique de Nice, Nice F-06200, France

(Received 3 April 2024; revised 6 June 2024; accepted 30 August 2024; published 17 September 2024)

Machine learning is becoming a widely used technique with impressive growth due to the diversity of problems of societal interest for which it can offer practical solutions. This increase of applications and required resources start to become limited by present-day hardware technologies. Indeed, novel machine learning subjects such as large language models or high-resolution image recognition raise the question of large computing time and energy cost of the required computation. In this context, optical platforms have been designed for several years with the goal of developing more efficient hardware for machine learning. Among different explored platforms, optical free-space propagation offers various advantages: parallelism, low energy cost, and computational speed. Here, we present a new design combining the strong and tunable nonlinear properties of a light beam propagating through a hot atomic vapor with an extreme learning machine model. We numerically and experimentally demonstrate the enhancement of the training using such free-space nonlinear propagation on an MNIST image classification task. We point out different experimental hyperparameters that can be further optimized to improve the accuracy of the platform.

DOI: 10.1103/PhysRevApplied.22.034041

I. INTRODUCTION

Machine learning is commonly used at a global scale in an extremely wide range of applications, offering powerful tools in many domains. Because of their adaptability and efficiency, artificial neural network (ANN) models have been widely implemented in most machine learning applications. The performances provided by such models can be explained by the large number of training parameters that can be implemented, reaching several billions for the biggest models (175 billions for GPT-3 [1]). These models, even though powerful in terms of accuracy, require important computational resources to perform training, which yields a heavy cost in both time and energy. The increase of complexity in machine learning tasks due to the democratization and spread of such technologies leads to new global issues, such as the excessive energy consumption of these models with an impact on global warming [2]. In addition, specific applications require high computation speed and have intrinsic energy consumption constraints [3,4].

Alternative methods have therefore been developed to tackle these issues: One of them aims to use physical systems as fast and energy-efficient platforms to process the data transformation required for machine learning. In this context photonic platforms appear to be a good candidate offering a fast, energy-efficient, and scalable solution [5] in the time domain [6–8] as well as in the spatial domain [9–13]. Methods going beyond those used by the usual neural network models, such as transformer models [14], have been developed in order to parallelize information processing and perform training on larger datasets and are widely used in natural language processing [15] and computer vision [16]. This approach is well adapted for optical platforms, in particular in the spatial domain, where large datasets can be processed in parallel using simple devices such as digital micromirror devices or spatial light modulators (SLMs).

Several optical platforms are already used for machine learning tasks, with present investigations focusing on the identification of specific potential to be competitive for precise user cases [17]. Among the various possibilities, a powerful method to adapt optical platforms for such applications is based on reservoir computing [18,19]. Unlike ANNs, where data are transformed by a trained nonlinear system and then read, the reservoir computing model consists in the transformation of data by a fixed and unknown nonlinear system while the training occurs at the readout. This conceptual difference with ANNs, despite being applicable to optical platforms, is also highly efficient in terms of cost, the number of required training parameters being several orders of magnitude smaller. Optical-based reservoir computing models have shown promising performances using silicon chips [20,21], laser networks [22], disordered media [23], Kerr nonlinear media [24–29], or delay-based systems [30–32].

Having a similar structure, extreme learning machines (ELMs) [33,34] are composed of a single hidden feed-forward layer performing a nonlinear transformation

*Contact author: robin.kaiser@univ-cotedazur.fr

 © 2024 American Physical Society

of the data with training done at the readout. Unlike reservoir computing, however, the reservoir used in ELMs is not designed to process dynamical memory. ELM models are based on support vector machines [35] and kernel methods [36] and can be implemented on various optical platforms. Among these different platforms, hot atomic vapors are a good candidate offering strong nonlinearities and are suitable for free-space propagation. We note that this nonlinear medium is well known in academic research and used in multiple fields of research [37,38], including machine learning [39,40]. An important advantage using hot atomic vapors is the possibility to tune the strength of the nonlinear transformation, giving access to a critical parameter to optimize the learning for a given problem.

In this paper we present the use of such hot atomic vapor as a tunable nonlinear optical medium for machine learning tasks through numerical and experimental results highlighting the impact of such a nonlinear medium on the learning efficiency. We then compare our system to a usual convolutional neural network (CNN), before concluding and providing a route for further improvement.

II. OPTICAL ELM ARCHITECTURE

A sketch of our hardware device is shown in Fig. 1(a): An SLM is used to encode the data onto the incoming laser beam, which then propagates through a hot atomic vapor (natural mixture of rubidium) undergoing a nonlinear spatial transformation. The modified data are then recorded by a camera where the learning is numerically processed. The equivalence with a regular ELM scheme is illustrated in Fig. 1(b): The SLM encodes N samples $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_N]$ (inputs) in the optical system and the atomic vapor plays the role of the hidden layer (reservoir) that performs the data transformation $\mathbf{H}(\mathbf{X})$. Finally, the camera records the resulting field, and a linear combination with a set of M weights $\boldsymbol{\alpha}$ is performed to get the outputs $\mathbf{Y} = \mathbf{H}\boldsymbol{\alpha}$. The training of the system is then done by optimizing these weights by digitally solving a ridge regression problem [41].

The detail of our hidden-layer matrix $\mathbf{H}$ can be explained as follows: It projects input data into a feature space necessary to properly learn. In our system this projection is done completely optically and can be written as

$$\mathbf{H} = \mathbf{D}[\mathbf{A}(p(\mathbf{X}, \mathbf{W}))]. \tag{1}$$

Here $\mathbf{D}$ is the detection function of the camera, $\mathbf{A}$ is a complex function modeling the spatial evolution of the beam along its propagation, p is the encoding function of the SLM, and $\mathbf{W}$ is an embedding matrix added to the encoding and creating biases on the input that leads to an enhancement of transformations during the propagation of the data.

The embedding $\mathbf{W}$ in our optical system is implemented by adding to each sample the same random matrix as presented in Fig. 1(c). The encoding as mentioned above uses an SLM to imprint the information onto the phase profile of the laser beam: $p(\mathbf{x}) = \exp(i\mathbf{x})$. Considering a Gaussian beam of waist w propagating along z, the laser field is then

$$\psi(\mathbf{r}, z) = \sqrt{I}\exp\left(-\frac{r^2}{w^2}\right)\exp[i(\mathbf{X}(\mathbf{r}) + \mathbf{W}(\mathbf{r}))], \tag{2}$$

where I is the beam intensity at the center, $\mathbf{r} = (x, y)$ ($r = \sqrt{x^2 + y^2}$) is the coordinate in the transverse plane of propagation of the beam, and w is the waist of the laser beam. It is important to note that the use of complex values transformed in a continuous space largely enhances the performance of the model compared to a real-valued binary solution.

The detection function $\mathbf{D}$ can also increase the complexity of the hidden layer by using the saturation properties of the camera, with, for example, a linear response up to a threshold intensity value given by I_s. The detection function can be approximated by $\mathbf{D}(\mathbf{I}) = \min(\mathbf{I}, I_s)$ with $\mathbf{I}$ the intensity matrix incident on the camera. The hidden-layer matrix of our optical system can thus be rewritten as

$$\mathbf{H} = \mathbf{D}(\mathbf{I})[\mathbf{A}[\psi(\mathbf{r}, z)]]. \tag{3}$$

Finally, the propagation of the data through the nonlinear vapor is defined by the transformation $\mathbf{A}$ and is described by the nonlinear Schrödinger (NLS) equation. In the paraxial approximation for the optical field propagating along z through the nonlinear medium, the slowly varying complex amplitude $\psi(\mathbf{r}, z)$ of the laser field is described by

$$i\frac{\partial\psi(\mathbf{r}, z)}{\partial z} = \left(-\frac{1}{2k_0}\nabla_{\mathbf{r}}^2 - k_0 n_2|\psi(\mathbf{r}, z)|^2\right)\psi(\mathbf{r}, z), \tag{4}$$

where k_0 corresponds to the wave vector of the laser beam. The last term of the equation shows that the nonlinear propagation of the light inside the vapor depends on the intensity of our beam $I = |\psi(\mathbf{r}, z)|^2$. Here $n_2(\rho, \Delta)$ is the first-order nonlinear refractive index, ρ is the atomic density of our vapor (directly linked to its temperature), and Δ is the frequency detuning between the laser and the atomic resonance (with respect to the $F = 2 \rightarrow F'$ transition of ^{87}Rb). It is thus possible to control the nonlinear transformation seen by the data by modifying the beam intensity, its frequency, or the temperature of the vapor cell. In the context of machine learning, these control parameters can be defined as hyperparameters.

Once the optical transformation has been performed and recorded by the camera, the learning step can be processed digitally. As presented in Fig. 1(d), M channels (yellow squares) are randomly selected on this output matrix,

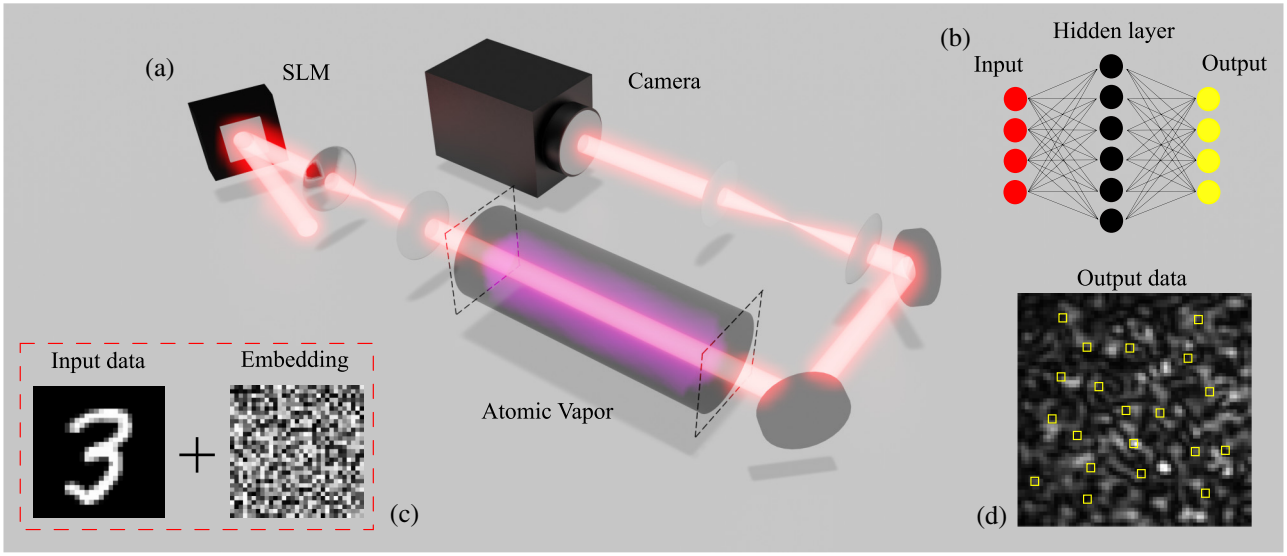


FIG. 1. Scheme and description of the optical system and its use as an extreme learning machine device. (a) The optical setup. The input data are encoded onto the beam with an SLM imaged at the entrance of the vapor cell. The data then propagate through the tunable nonlinear atomic vapor cell. The output image is recorded on a camera, conjugating the output plane of the vapor cell. (b) Extreme learning machine model. (c) Typical input data of MNIST dataset and a random embedding matrix added to the data. (d) Typical image recorded after nonlinear propagation of MNIST data. Yellow squares are randomly distributed areas used as training parameters with associated weights (α).

corresponding to training parameters, and a weight α is assigned to each of them. These channels are assimilated to training parameters. In order to reduce technical noise, each channel is an average over a square of 8×8 pixels. The digital learning is a simple ridge regression task, in our case processed using Scikit-learn [42] on Python.

III. RESULTS

We first illustrate the impact of a nonlinear propagation by a numerical investigation. We therefore use samples from the Abalone dataset, which we imprint onto the phase profile of a Gaussian beam whose propagation is simulated numerically. We compare two different situations: one with a purely linear propagation and one using the nonlinear transformation as described by the NLS equation. The learning task is then performed identically for the linear and nonlinear propagation.

We present the corresponding regression accuracy by plotting the root-mean-square error (RMSE) in Fig. 2(a) on a test set as a function of the number of training parameters obtained in these two cases. This numerical simulation illustrates the positive impact that a nonlinear propagation can have on the learning efficiency of such an ELM model. We find that a weak nonlinearity is already sufficient to observe an enhancement of the inference accuracy from 91.8% to 92.5%. It is also interesting to notice a particular strength of ELM models, with the regression converging toward a high accuracy with only relatively few training parameters ($\approx$1000). We stress that the efficiency of ELM models to learn from a small number of parameters leads to a potentially important gain in terms of time and energy of training.

We have, however, realized that the computation time needed to simulate our optical system is already extremely large even for a relatively small dataset like Abalone. We therefore continued our investigation on larger datasets directly on our optical platform. Indeed, as our free-space optical system is well suited for image data processing and ELM is a good model for classification, the experimental study has been performed for a classification task using the larger MNIST (Modified National Institute of Standards and Technology) dataset composed of images of handwritten digits comprising 28×28 pixels.

The results of the corresponding classification accuracy are shown in Fig. 2(b). First, we show as a reference the case of purely digital learning, in the absence of any optical encoding and propagation. The corresponding black line in Fig. 2(b) illustrates the fast learning (with fewer than 1000 trained parameters), which, however, comes along with a limited accuracy of 86.0%. Using an optical free-space propagation already allows for an increased accuracy of 87.4%, underlying the impact of the encoding and recording optical devices [i.e., encoding and detection function p and $\mathbf{D}$ in Eq. (1)]. We note that, as discussed earlier, a random embedding matrix has been added to the sample to enhance interactions during the free-space beam propagation. We have checked that using different random matrices leads to significant fluctuations on the classification error, implying that the embedding matrix needs to be

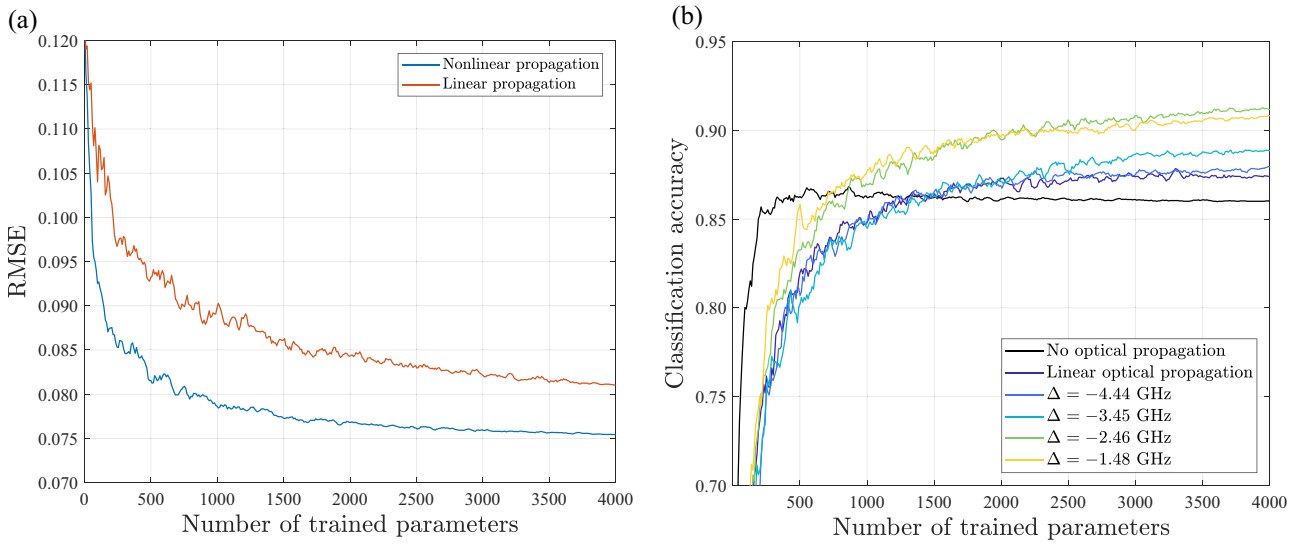


FIG. 2. Classification error and accuracy as functions of the number of training parameters. (a) Classification error (RMSE) for a numerical simulation with the Abalone dataset (4000 data) comparing linear (red) and nonlinear (blue) propagation. (b) Classification accuracy for MNIST datasets. Numerical results for a linear regression on the generated input mask, corresponding to no optical propagation (black line), and experimental results in the presence of optical propagation: linear optical propagation (violet) and optical propagation with different laser frequency detunings (i.e., nonlinearities) as shown. Reducing $|\Delta|$ increases the nonlinearity. Curves have been smoothed for better readability.

optimized for specific tasks. We also noticed that the number of pixels per channel has an impact on the accuracy, with a smaller number of pixels per channel improving the learning efficiency.

Preliminary experimental classification tests with different values of saturation parameter confirmed that a lower saturation intensity I_s of the camera enhances the accuracy of the model. It is important to note that the enhancement due to the camera saturation is additional to the learning properties obtained only with the atomic vapor. The results presented in Fig. 2 have been done without any saturation of the detector. After optical propagation and recording on a CCD, the learning protocol is the same as for the purely digital process and involves associating a weight to each training parameter, consisting of a group of pixels (channel) of the CCD. We note that both numerical and experimental studies present a similar increase of accuracy with the number of training parameters. Even though the fully digital ELM model stays at lower accuracy, it converges with fewer trained parameters. We associate this difference to a spreading of the information of the input matrix during free-space propagation requiring more output channels to access all the information spread over more output channels.

We now turn to the characterization of the nonlinear beam propagation through the atomic vapor, the main novelty of our optical platform. As explained earlier, the nonlinear strength of the atomic vapor depends on three experimental hyperparameters, namely laser intensity, laser frequency (detuning), and atomic vapor density. In the results presented in this paper, we keep the laser intensity fixed at $I \approx 1.3$ W cm^{-2}. The linear beam propagation regime is easily reached by using the atomic vapor cell at room temperature, where the low atomic density results in no noticeable impact of the atoms on the index of refraction. The nonlinear propagation regime is obtained when heating the atomic vapor cell. We then control the nonlinearity more conveniently via the laser detuning Δ, as tuning the laser frequency closer to the atomic transition frequency increases the nonlinear strength of the medium. We note that, in all studies presented in this paper, we have used a negative detuning, corresponding to a defocusing nonlinearity.

In Fig. 2(b) we observe an enhancement of the learning capacities when increasing the nonlinearity (i.e., reducing detuning $|\Delta|$): The classification accuracy with 4000 training parameters for increasing nonlinearity is 88.0%, 88.8%, 91.2%, and 90.7%. It is interesting to notice a decrease of efficiency between the strongest nonlinearity $\Delta = -1.48$ GHz and the one before $\Delta = -2.46$ GHz. It is well known that optical propagation in a strong nonlinear medium can exhibit beam cleaning characteristics [43], which can smoothen out some of the information imprinted onto the beam, resulting in a decrease of learning efficiency. This effect, connected to condensation of photons under nonlinear propagation, is more prominent in optical fibers [27] and appears after longer nonlinear propagation for free-space devices as used in our

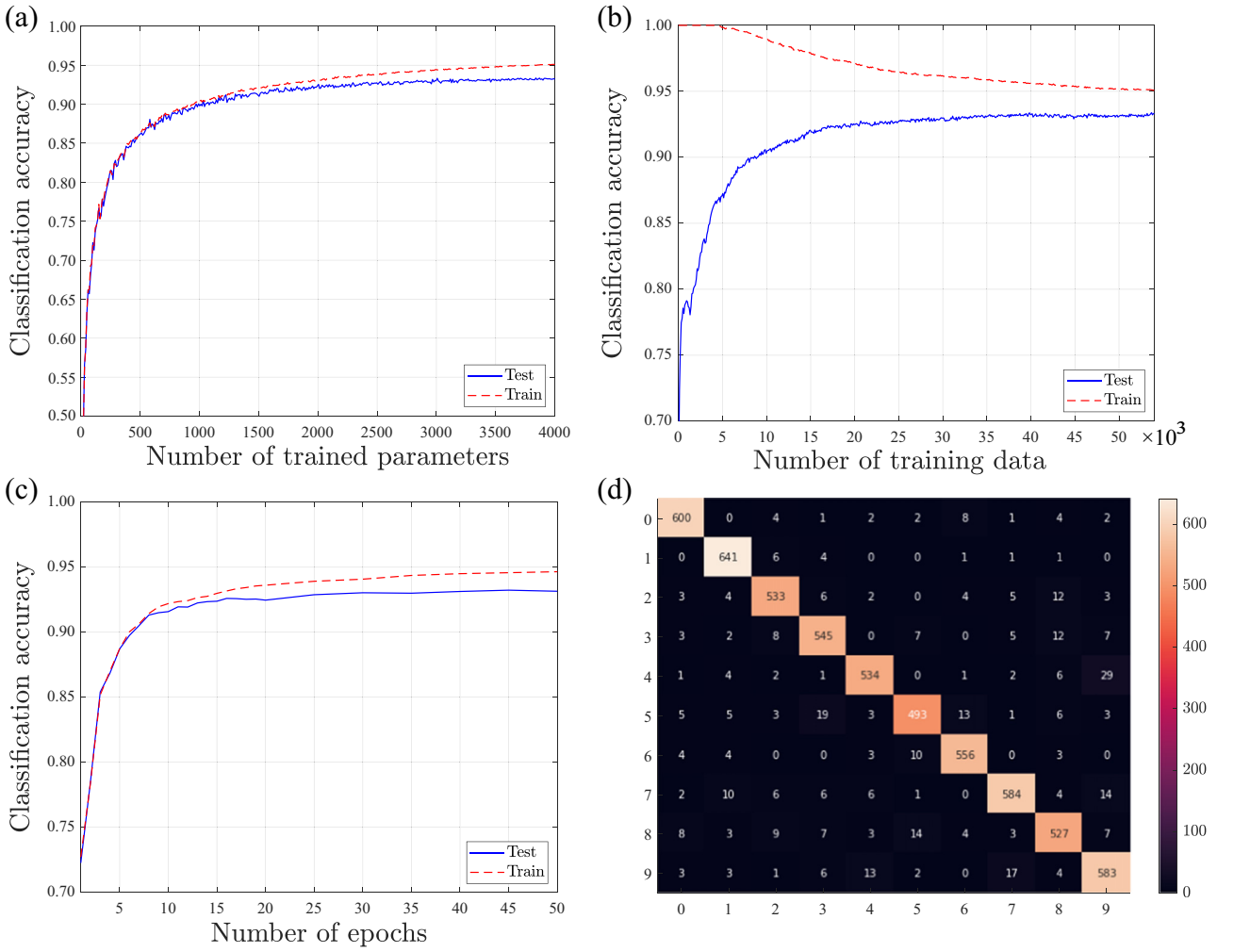


FIG. 3. Experimental training and inference classification accuracy for different hyperparameters on 60 000 samples of the MNIST dataset (54 000 for training and 6000 for testing). (a) Accuracy as a function of the number of trained parameters with the full dataset. (b) Accuracy as a function of the number of training data of the MNIST dataset with 4000 trained parameters. The number of testing data is fixed at 4000 samples. (c) Accuracy as a function of epochs with the full dataset and 4000 trained parameters. (d) Confusion matrix obtained in the best case: full dataset, 4000 trained parameters, and 50 epochs.

platform [44]. Additional effects have been observed in hot atomic vapors, such as light absorption, nonlocality and saturation of the nonlinearity [45], and remain to be explored in the context of machine learning efficiency. A full exploration to determine optimal experimental hyperparameters giving access to a nonlinear regime that maximizes the learning efficiency thus remains to be performed.

The study presented in Fig. 3 has been performed on the full MNIST dataset with nonlinear propagation ($\Delta = -2.46$ GHz) to better understand the training and testing behavior of the model for larger datasets. Similarly to the previous discussion, Fig. 3(a) shows the evolution when increasing the number of trained parameters. Once again, it highlights that good accuracies can be achieved with very low number of training parameters, an intrinsic advantage of ELM models. We can also observe in Fig. 3(b) that the accuracy of our optically accelerated ELM converges toward its final value with only a third of the dataset.

The last study presented in Fig. 3(c) corresponds to the evolution of the accuracy as a function of the number of epochs (i.e., times the digital training has been performed on each sample, after their optical propagation). This curve shows that tens of epochs are enough to reach a final value. Finally, Fig. 3(d) presents the confusion matrix of the best case using the full dataset, 4000 training parameters, and 50 epochs, leading to a classification accuracy of 93.25% for the testing and 95.10% for the training. On top of presenting the best accuracy reached so far with our ELM

photonic platform, Fig. 3 highlights the low requirements needed by the model to achieve a good testing accuracy.

IV. COMPARISON WITH CNN AND FUTURE PROSPECTS

With this first benchmarking for machine learning performed with our hot atomic vapor as an extreme learning machine, we now compare the potential of this hardware against other standard machine learning systems. We therefore present in Table I a comparison between a simple convolutional neural network (Lenet-5) [46] and our system. For this CNN model, we have used three convolutional plus two average pooling layers, well suited to perform MNIST classification. Here, we have expanded the samples and used an adapted version of Lenet-5 to study the impact of sample size on CNN performance. By resizing MNIST samples up to 128×128, both the number of training parameters required by the digital model as well as the time it takes to process a sample through the five layers increase drastically. The accuracy presents also a tendency to decrease when increasing the sample size. Even though the CPU we have used in this example is not the most efficient, this test highlights a common problem of CNN, where an increase of sample size has critical consequences in terms of time and energy of the required computation.

In the middle two columns of the table we show how the number of required training parameters and the required computation time scale with the sample size in the case of learning using an adapted version of Lenet-5 (CNN) and for our ELM model. Both the number of required training parameters and the computation time scale as the size $N \times N$. We stress that the reduced number of training parameters for our approach is an inherent advantage of all ELM models. The third column shows the required computation time. Again, for the Lenet-5 (CNN) model, one can see that the computation time scales again linearly with the size size $N \times N$. However, when using our nonlinear optical approach to implement the ELM model, we can see that the computation is independent of the system size! This presents a major advantage for our nonlinear optical acceleration on top of the use of the ELM model. We note that the ELM model requires fewer training parameters, allowing one to scale the size of the system to 1000×1000, whereas for the CNN-based approach this larger number of required training parameters put very hard constraints on the computational effort.

The last column of the table indicates the accuracy of the different situations, showing the reduced accuracy for increasing system size in our implementation of the Lenet-5 (CNN) model. We note that, due to restricted computation time, we were not able to perform the full training on the 1000×1000 sample size. Similarly, for our nonlinear optical approach, we did not yet implement a 1000×1000 sample (which required further development in our hardware). We note that in the Lenet-5 model the number of required training parameters scales linearly with the input size whereas in our platform the input size is dependent on the SLM resolution while the number of training parameters depends on the CCD resolution. We stress that, in our system, the computation time is the same for large and small samples, an important asset for any optical solver.

In summary, Table I thus shows that a simple CNN is more efficient than our present system for small samples (such as the MNIST dataset used in this paper). We note that the limitation in computational time of our optical platform stems from the electrical-to-optical (SLM) and optical-to-electrical (camera) transformations. However, the main advantage of using a free-space optical platform is the possibility of processing very large samples without supplementary cost in computation time. Indeed, these devices are able to process at the same speed of 50 Hz samples of size up to 10^6 pixels (with commercially available SLMs already reaching sampling times of 100 Hz/10^7 pixels).

In order to explore the potential of our nonlinear optics platform, let us assume that each pixel of the encoding device can be assimilated to an optical mode propagating along the optical path and is interacting with the other modes via the nonlinear atom-photon interaction

TABLE I. MNIST classification performances (number of trained parameters, speed, and accuracy) using Lenet-5 (CNN) with different sample sizes and our nonlinear optics (NLO) device as an ELM. "Input size" corresponds to the image resolution (in pixels) sent into the model. The accuracies have been obtained after training over 10 000 samples. The CPU used to perform this comparison is an Intel core i5-4690 3.5 GHz.

Platform / model (input size)	Number of required parameters	Time per sample (ms)	Testing accuracy (%)
CPU / Lenet-5 (28×28)	60 000	1.6	94.50
CPU / Lenet-5 (64×64)	340 000	5.0	94.40
CPU / Lenet-5 (128×128)	1740 000	19.0	89.20
CPU / Lenet-5 (1000×1000)	118 100 000	1100.0	unknown
NLO / ELM (28×28)	4000	20	91.20
NLO / ELM (1000×1000)	$\sim$ 1000 000	20	unknown

occurring inside the hot atomic vapor. Such interaction between two pixels being equivalent to a mathematical operation in a digital processing unit, we estimate that our system (with the actual encoding technology) is able to perform 10^{12} eight-bit operations per frame at a rate of 50 Hz, leading to potentially 5×10^{13} OPS (operations per second). With an energy consumption of our entire system (including ELM processed digitally and the heating of the vapor cell) of about 500 W, the efficiency of our platform can be estimated to be on the order of 10^{11} OPS/W (operations per second per watt). We note that the cost (in time and energy) of the learning part of this protocol itself remains negligible. Indeed the inherent properties of ELM models need only a simple pass through only one layer of training parameters, whose number is always kept low.

Based on the results of our first experimental study with the MNIST dataset, we estimate that a number of training parameters (channels on the output image) equivalent to the input size ($L \times L$) should be sufficient to reach an accuracy of at least 90%. It is also worth noting that this time and energy advantage of our nonlinear optical platform does not come at the expense of increased fabrication cost. If the expected performances of our platform become competitive with present-day supercomputers (used for machine learning tasks), our platform can be realized with a hugely reduced fabrication cost and hardware size. Indeed, while the actual best supercomputer Frontier [47] cost 600 million US dollars to build and consists of 74 nineteen-inch rack cabinets (over 680 m^2), we estimate that our solution could roughly cost a few hundred thousands of US dollars and the entire system could be contained in one nineteen-inch rack cabinet. This small size also provides an important advantage for tasks requiring on-site or offline work.

We stress that the results presented in this paper consist in a very first proof of principle, and our device is far from being optimized for machine learning purposes. The embedding matrix, for instance, plays a key role on how the data are projected into the feature space, and it will be interesting to explore how different matrices or encoding methods would impact the learning efficiency. We had, for example, implemented randomly distributed embedding matrices to maximize the connections into the sample along its optical propagation. Even if we did not observe any significant difference between adding or multiplying the sample with this embedding matrix, we noticed an important difference of accuracy from one random matrix to another: A variation of almost 10% has been observed just by changing the seed of the random matrix. This difference when using various random distributions of this embedding matrix suggests a potential for optimization, which is likely to depend on the task and the dataset on which the system is used.

Similarly as in Refs. [48,49], a learning step has to be implemented on the optical encoding to optimize the impact of the embedding matrix. Even though the above technique is already used by other optical platforms, our nonlinear atomic vapor gives the possibility to develop and enhance this technique. For the study presented in this paper, we encoded the information onto the phase profile of our laser beam alone, keeping its intensity, which is one of the nonlinear hyperparameters, as a constant. It is, however, possible to use an SLM to simultaneously control both the intensity and the phase profile of a beam [50] via the use of a blazed grating, allowing one to simultaneously encode the sample onto the phase beam with the embedding matrix and locally tune the nonlinear strength of the atomic vapour via the use of the intensity profile. Therefore, the supplementary step needed to optimize the embedding matrix (phase profile) will also be used to determine the appropriate intensity profile enhancing the nonlinear transformation for a specific task or dataset.

V. CONCLUSION

Optical platforms have been used in the past to demonstrate the various advantages that they offer in performing complex calculation tasks such as machine learning. At present, many new systems are being developed in order to determine which ones will reach competitive performances against fully digital solutions. With this objective in mind, we have presented in this work the use of optical hardware based on nonlinear light propagation through a hot atomic vapor and its integration into an extreme learning machine model. Combining the large scalability of optical free-space propagation with the strong and tunable nonlinearity of our optical hardware with the specific design of ELMs optimized for fast learning on large and heavy datasets could lead to important savings in terms of time and energy of the computation.

In this paper, we have presented the first evidence of learning enhancement via the use of a hot atomic vapor as a nonlinear medium combined with an extreme learning machine model in the context of an image classification task. This study, performed on the MNIST dataset as a benchmark tool, shows that reasonable accuracies can be achieved quite easily. More importantly, it highlights the potential of optical platforms in terms of both scalability and efficiency regarding the time and energy cost of large samples. We note that other promising optical platforms using, for example, multimode fiber propagation to process the nonlinear transformation [27,51] are limited on this aspect due to the low number of modes they can inject into the fibers.

Various applications using large data (large language model and image recognition, for example) are currently being developed, and thus there is a necessity to create new machine learning platforms in order to efficiently reduce the cost of such applications. When fully developed, such platforms could become competitive with the actual best

supercomputers in terms of speed and energy cost of calculation along with lower fabrication cost and smaller hardware size. Finally, future implementations are envisioned such as the learning of the phase embedding matrix and a local tuning of the nonlinear medium via an adapted intensity matrix to drastically enhance the accuracy of the platform for various kinds of tasks.

ACKNOWLEDGMENTS

We thank M. Morisse for fruitful discussions on machine learning and C. Conti and D. Pierangeli for discussions on the optical extreme learning machine. We also thank C. Jurczak for his support throughout this project and CDL (Creative Destruction Lab) for their advice and mentorship. We also acknowledge funding from CNRS Innovation via the prematuration project SQVAC and Provence-Alpes-Côte d'Azur région via a "Jeune docteur innovant" project.

APPENDIX A: EXPERIMENTAL SETUP

The platform uses a continuous laser beam at 780 nm with a Gaussian profile of waist $w = 10$ mm and a power fixed at 1.3 W. Once linearly polarized, this beam illuminates a spatial light modulator (SLM) from Hamamatsu (model X15223), encoding eight-bit values on 1024×1242 pixels with a pitch of 12.5 μm and at a frame rate of 50 Hz. The SLM encodes onto the beam the desired matrix by imposing local phase delay of the beam from 0 to 2π with 255 values. The signal displayed by the SLM is the sum of the sample and the embedding matrix. The plane of the SLM is then imaged via an afocal system at the entrance of the 7-cm vapor cell. The vapor, a natural mixture of rubidium (72% of 85Rb and 28% of 87Rb), is heated at 120 °C to increase the atomic density (to $\sim 10^{19}$ atoms/m^3), enhancing photon-atom interactions.

The laser frequency is detuned by several GHz toward the red from the atomic resonance in order to optimize the effect of the Kerr-like nonlinearity. In such a regime the vapor behaves as a defocusing medium; this effect being enhanced by the beam intensity itself, we talk about the self-defocusing phenomenon. Once the beam has propagated through the vapor, the output window of the cell is imaged onto a CCD camera. In this first study, to avoid noise, the image displayed by the SLM is magnified by a factor 10 and the decoding channels used to learn are an average over 10×10 camera pixels.

All the samples are processed one by one through the optical system with fixed hyperparameters (detuning, vapor temperature, beam power, camera saturation, etc.) and embedding matrix. Once recorded by the camera, they are stored on the computer controlling the system. Randomly distributed channels are chosen to perform the learning on the training dataset by determining the weight of all these channels. When the training is finished, the channels with adequate weights are used to process the classification on the testing dataset.

APPENDIX B: DATASETS

The Abalone dataset [52] is a commonly used dataset to benchmark machine learning hardwares and models on regression tasks and is well suited to start working on new systems due to its low number of samples (8177). This dataset concerns the identification of sea snails, each of them having eight attributes (age, length, diameter, height, rings, ...).

The MNIST dataset [53] concerns a classification task on a multiclass problem: handwritten digits. The dataset includes 10 classes from 0 to 9 over 60 000 samples for the training and 10 000 for the testing. Each of these samples are matrices of 28×28 pixels with filled eight-bit values. This widely used dataset is a necessary step to benchmark a machine learning platform on classification tasks. The two-dimensional format of the samples makes it a perfect first playground for systems aiming to develop image-based tasks.

[1] L. Floridi and M. Chiriatti, GPT-3: Its nature, scope, limits, and consequences, Minds Mach. **30**, 681 (2020).
[2] E. Strubell, A. Ganesh, and A. McCallum, Energy and policy considerations for modern deep learning research, Proc. AAAI Conf. Artif. Intell. **34**, 13693 (2020).
[3] L. Hanzlik, Y. Zhang, K. Grosse, A. Salem, M. Augustin, M. Backes, and M. Fritz, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Nashville, TN, USA, 2021), pp. 3295–3304.
[4] S. Islam, J. Deng, S. Zhou, C. Pan, C. Ding, and M. Xie, in *IEEE/DATE Design, Automation & Test in Europe Conference & Exhibition* (Antwerp, Belgium, 2022), pp. 921–926.
[5] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. Miller, and D. Psaltis, Inference in artificial intelligence with deep optics and photonics, Nature **588**, 39 (2020).
[6] A. Lupo, L. Butschek, and S. Massar, Photonic extreme learning machine based on frequency multiplexing, Opt. Express **29**, 28257 (2021).
[7] M. Yildirim, I. Oguz, F. Kaufmann, M. R. Escale, R. Grange, D. Psaltis, and C. Moser, Nonlinear optical data transformer for machine learning, arXiv preprint arXiv:2208.09398.
[8] B. Fischer, M. Chemnitz, Y. Zhu, N. Perron, P. Roztocki, B. MacLellan, L. Di Lauro, A. Aadhi, C. Rimoldi, T. H. Falk *et al.*, Neuromorphic computing via fission-based broadband frequency generation, Adv. Sci. **10**, 2303835 (2023).
[9] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, Reinforcement learning in a large-scale photonic recurrent neural network, Optica **5**, 756 (2018).

[10] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, All-optical machine learning using diffractive deep neural networks, Science **361**, 1004 (2018).

[11] Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, All-optical neural network with nonlinear activation functions, Optica **6**, 1132 (2019).

[12] D. Pierangeli, G. Marcucci, and C. Conti, Photonic extreme learning machine by free-space optical propagation, Photonics Res. **9**, 1446 (2021).

[13] T. Bu, H. Zhang, S. Kumar, M. Jin, P. Kumar, and Y. Huang, Efficient optical reservoir computing for parallel data processing, Opt. Lett. **47**, 3784 (2022).

[14] U. Kamath, K. Graham, and W. Emara, *Transformers for Machine Learning: A Deep Dive* (Chapman and Hall/CRC, New York, 2022).

[15] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, Natural language processing: An introduction, J. Am. Med. Inform. Assoc. **18**, 544 (2011).

[16] B. Jähne, H. Haußecker, and P. Geissler, in *Handbook of Computer Vision and Applications* (Citeseer, 1999), Vol. 2.

[17] P. L. McMahon, The physics of optical computing, Nat. Rev. Phys. **5**, 717 (2023).

[18] B. Schrauwen, D. Verstraeten, and J. V. Campenhout, in *Proceedings of the 15th European Symposium on Artificial Neural Networks* (ESANN, Bruges, Belgium, 2007), pp. 471–482.

[19] G. V. der Sande, D. Brunner, and M. C. Soriano, Advances in photonic reservoir computing, Nanophotonics **6**, 561 (2017).

[20] 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, Nat. Commun. **5**, 3541 (2014).

[21] S. Bandyopadhyay, A. Sludds, S. Krastanov, R. Hamerly, N. Harris, D. Bunandar, M. Streshinsky, M. Hochberg, and D. Englund, Single chip photonic deep neural network with accelerated training, arXiv preprint arXiv:2208.01623.

[22] A. Röhm, L. Jaurigue, and K. Lüdge, Reservoir computing using laser networks, IEEE J. Sel. Top. Quantum Electron. **26**, 1 (2020).

[23] F. Xia, K. Kim, Y. Eliezer, L. Shaughnessy, S. Gigan, and H. Cao, Deep learning with passive optical nonlinear mapping, arXiv preprint arXiv:2307.08558.

[24] J. Pauwels, G. Verschaffelt, S. Massar, and G. Van der Sande, Distributed Kerr non-linearity in a coherent all-optical fiber-ring reservoir computer, Front. Phys. **7**, 138 (2019).

[25] G. Marcucci, D. Pierangeli, and C. Conti, Theory of Neuromorphic Computing by Waves: Machine Learning by Rogue Waves, Dispersive Shocks, and Solitons, Phys. Rev. Lett. **125**, 093901 (2020).

[26] X. Xu, M. Tan, B. Corcoran, J. Wu, T. G. Nguyen, A. Boes, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell *et al.*, Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks, Laser Photonics Rev. **14**, 2000070 (2020).

[27] U. Teğin, M. Yıldırım, İ. Oğuz, C. Moser, and D. Psaltis, Scalable optical learning operator, Nat. Comput. Sci. **1**, 542 (2021).

[28] D. Moss, Ultra-high bandwidth optical neuromorphic processing with Kerr soliton crystal microcombs: Scaling the network to the PetaOp regime, Authorea Preprints (2023).

[29] I. Boikov, D. Brunner, and A. De Rossi, Evanescent coupling of nonlinear integrated cavities for all-optical reservoir computing, New J. Phys. **25**, 093056 (2023).

[30] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, All-optical reservoir computing, Opt. Express **20**, 22783 (2012).

[31] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, Sci. Rep. **2**, 287 (2012).

[32] Y. K. Chembo, Machine learning based on reservoir computing with time-delayed optoelectronic and photonic systems, Chaos **30**, 1054 (2020).

[33] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, Extreme learning machine: Theory and applications, Neurocomputing **70**, 489 (2006). neural Networks.

[34] J. Wang, S. Lu, S.-H. Wang, and Y.-D. Zhang, A review on extreme learning machine, Multimed. Tools. Appl. **81**, 41611 (2022).

[35] W. S. Noble, What is a support vector machine?, Nat. Biotechnol. **24**, 1565 (2006).

[36] T. Hofmann, B. Schölkopf, and A. J. Smola, Kernel methods in machine learning, Ann. Stat. **36**, 1171 (2008).

[37] K. W. Martin, G. Phelps, N. D. Lemke, M. S. Bigelow, B. Stuhl, M. Wojcik, M. Holt, I. Coddington, M. W. Bishop, and J. H. Burke, Compact optical atomic clock based on a two-photon transition in rubidium, Phys. Rev. Appl. **9**, 014019 (2018).

[38] Q. Glorieux, T. Aladjidi, P. D. Lett, and R. Kaiser, Hot atomic vapors for nonlinear and quantum optics, New J. Phys. **25**, 051201 (2023).

[39] A. Ryou, J. Whitehead, M. Zhelyeznyakov, P. Anderson, C. Keskin, M. Bajcsy, and A. Majumdar, Free-space optical neural network based on thermal atomic nonlinearity, Photonics Res. **9**, B128 (2021).

[40] N. A. Silva, V. Rocha, and T. D. Ferreira, Optical extreme learning machines with atomic vapors, Atoms **12**, 10 (2024).

[41] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst., Man, Cybern., B (Cybernetics) **42**, 513 (2012).

[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

[43] K. Krupa, A. Tonello, B. M. Shalaby, M. Fabert, A. Barthélémy, G. Millot, S. Wabnitz, and V. Couderc, Spatial beam self-cleaning in multimode fibres, Nat. Photonics **11**, 237 (2017).

[44] N. Šantić, A. Fusaro, S. Salem, J. Garnier, A. Picozzi, and R. Kaiser, Nonequilibrium Precondensation of Classical Waves in Two Dimensions Propagating through Atomic Vapors, Phys. Rev. Lett. **120**, 055301 (2018).

[45] P. Azam, A. Fusaro, Q. Fontaine, J. Garnier, A. Bramati, A. Picozzi, R. Kaiser, Q. Glorieux, and T. Bienaimé, Dissipation-enhanced collapse singularity of a nonlocal fluid of light in a hot atomic vapor, Phys. Rev. A **104**, 013515 (2021).

[46] Y. LeCun *et al.*, Lenet-5, convolutional neural networks, URL: http://yann.lecun.com/exdb/lenet **20**, 14 (2015).

[47] D. Schneider, The exascale era is upon us: The Frontier supercomputer may be the first to reach 1,000,000,000,000,000,000 operations per second, IEEE Spectr. **59**, 34 (2022).

[48] I. Oguz, J.-L. Hsieh, N. U. Dinc, U. Teğin, M. Yildirim, C. Gigli, C. Moser, and D. Psaltis, Programming nonlinear propagation for efficient optical learning machines, Adv. Photonics **6**, 016002 (2024).

[49] L. Wu and Z. Zhang, Only-train-electrical-to-optical-conversion (OTEOC): Simple diffractive neural networks with optical readout, Opt. Express **30**, 28024 (2022).

[50] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, Exact solution to simultaneous intensity and phase encryption with a single phase-only hologram, Opt. Lett. **38**, 3546 (2013).

[51] D. Ancora, M. Negri, A. Gianfrate, D. Trypogeorgos, L. Dominici, D. Sanvitto, F. Ricci-Tersenghi, and L. Leuzzi, Multi-mode fiber reservoir computing overcomes shallow neural networks classifiers, arXiv preprint arXiv:2210.04745.

[52] (https://www.kaggle.com/datasets/rodolfomendes/abalone-dataset).

[53] (https://www.kaggle.com/datasets/hojjatk/mnist-dataset).