Probing a theoretical framework for a Photonic Extreme Learning Machine

Vicente Rocha[1,2], Duarte Silva[3], Felipe C. Moreira[1], Catarina S. Monteiro[1], Tiago D. Ferreira[1], Nuno A. Silva[1]

[1] INESC TEC, Centre of Applied Photonics, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

[2] Faculty of Sciences of the University of Porto, Department of Physics and Astronomy, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

[3] Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, the Netherlands

E-mail: nuno.a.silva@inesctec.pt

August 2024

Abstract.

The development of computing paradigms alternative to von Neumann architectures has recently fueled significant progress in novel all-optical processing solutions. In this work, we investigate how the coherence properties can be exploited for computing by expanding information onto a higher-dimensional space in the photonic extreme learning machine framework. A theoretical framework is provided based on the transmission matrix formalism, mapping the input plane onto the output camera plane, resulting in the establishment of the connection with complex extreme learning machines and derivation of upper bounds for the hidden space dimensionality as well as the form of the activation functions. Experiments using free-space propagation through a diffusive medium, performed in low-dimensional input space regimes, validate the model and the proposed estimator for the dimensionality. Overall, the framework presented and the findings enclosed have the potential to foster further research in a multitude of directions, from the development of robust general-purpose all-optical hardware to a full-stack integration with optical sensing devices toward edge computing solutions.

Keywords: Optical Computing, Interferometry, Machine Learning, Extreme Learning Machine

1. Introduction

Optical computing has long been promising applications for faster information processing and lower power consumption [1]. Today, photonic hardware contributes to modern systems in multiple tasks, from accelerating data transmission [2] and implementing specific linear operations such as Fourier transforms [3], to more complex optical correlators [4], integral transforms [5], matrix-vector multiplication [6, 7], and

reconfigurable linear operators [8, 9, 10]. Despite these successes, scalable and energy-efficient optical transistors and memories to support von Neumann general-purpose photonic processors remain ellusive due to weak light interaction and optical field confinement [11]. One way to bypass these obstacles is to seek alternative computing frameworks, such as neuromorphic-inspired architectures, a research line that recently seeded the development of computing paradigms closer to analog computing, where light has long been known to excel.

In short, neuromorphic architectures offer a promising path for optical processors by leveraging a distributed memory framework through weights [12, 13] and approximation capabilities of neural networks (NN) enabled by the expressivity of non-linear maps [14]. Together with native light speed and broad bandwidth of photonic hardware, it thus enables massive parallelism and throughput in compact, low-power devices [15]. Recent advances in photonic neuromorphic computing have followed several promising directions, with one of the most compelling paths being the realisation of all-optical deep neural networks through a sequential series of fabricated [16, 17] or reprogrammable diffractive layers [18, 19]. Although fabricated layers have the shortfall of fixing the setup for a specific task, using reprogrammable layers unlocks hardware change between tasks, giving rise to a versatile family of task-specific accelerators that take advantage of physics-aware training and even structural nonlinearities [20]. However, as the size scale of the network scales, accumulated optical loss, noise, fabrication imperfections, and the time and energy overhead of training these devices result in limited practical scalability.

A greater degree of flexibility is offered by reservoir computing (RC) and, in particular, its memoryless counterpart, the Extreme Learning Machine (ELM) [21, 22, 23]. In these architectures, the combination of random weights and the nonlinear activation function gives the framework a universal function approximation capability for both regression and classification problems [22]. The photonic implementation of ELMs, coined Photonic Extreme Learning Machine (PELM), has been demonstrated in free-space architectures [24, 25], optical fibers [26], and integrated optics [27], typically exploiting the inherent randomness of light scattering to realise a fixed random weight matrix. Programmable optical hardware has also been employed, enabling ensemble techniques and evolutionary optimisation strategies [28]. These systems have already tackled tasks ranging from dimensionality reduction [29] and kernel approximation [30] to natural language processing [25]. However, the overall comprehension of the ultimate limits and opportunities of these optical computing devices would benefit from a more comprehensive theoretical study on the dimensionality of the output space and its activation functions, in particular when the input space has low dimensionality.

Taking this context as the starting point, this work analyzes a simplified free-space optical computing configuration (depicted in Figure 1) that serves as an experimental PELM test bed. First, in section 2, we develop a theoretical model by describing both the input information encoding, the random linear propagation using the transmission matrix formalism, and the detection step, finishing with the derivation

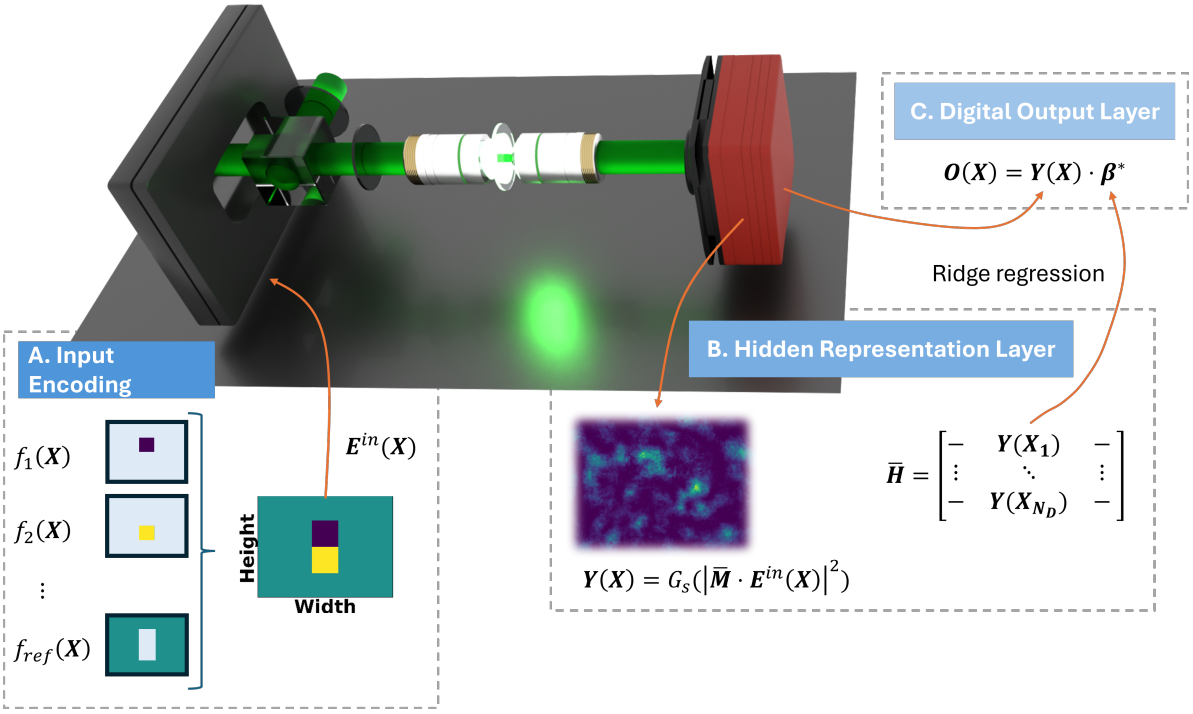

Figure 1. Overview of the optoelectronic PELM experimental implementation and connection to the theoretical framework. **A.** The information of each input state $\boldsymbol{X}$ is encoded on the incident optical beam in phase or amplitude using an SLM. **B.** The beam is then focused on an optical media (here a diffusive media) inducing a random transmission matrix $\bar{\boldsymbol{M}}$ modeling the propagation before being collimated again resulting in a speckle patter then collected with a CMOS sensor, resulting in the hidden layer representation $\boldsymbol{Y}(\boldsymbol{X})$. **C.** The imaged speckle pattern is then used as input to the linear model $\boldsymbol{O}(\boldsymbol{X}) = \boldsymbol{Y}(\boldsymbol{X})\bar{\boldsymbol{\beta}}^*$, obtaining a prediction $\boldsymbol{O}(\boldsymbol{X})$ by leveraging on a $\boldsymbol{\beta}^*$ trained with Ridge regression using the hidden layer output matrix $\bar{\boldsymbol{H}}$ built from N_D training samples.

of the dimensionality upper bounds of the hidden space for linear camera response. Then, we proceed by proposing a methodology to estimate the dimensionality of the space in the presence of experimental noise, via the analysis of the spectrum of singular values. In section 4 we employ the experimental setup in two low-dimensional spaces to validate the theoretical model and observe the enrichment of the hidden space by the introduction of the camera nonlinear response. To conclude, in section 5 we discuss future prospect of growth for the expressivity of this technology based on the theoretical analysis provided.

2. Theoretical Framework

The optical computing setup used for this work depicted in Figure 1 permits a simplified, stage-by-stage analysis of encoding, transmission, and detection within a typical PELM implementation. Briefly, the input features of a dataset are encoded on the wavefront of a coherent optical beam using a spatial light modulator (SLM) either with phase

or amplitude modulation, as described in the next subsections. As the information in the wavefront propagates, it is randomly mixed by the tight focusing of a microscope objective onto a diffusive medium, resulting in a speckle pattern. A second objective collimates the light that is then collected by a CMOS camera. The integrated intensity across N_{out} macropixel regions of the sensor composes the hidden layer used to train a final digital linear transformation to solve a task (regression or classification problem).

In the discussion that follows, we model how information is transformed as the optical signal propagates through each stage. This theoretical model aims to clarify how the interplay of input encoding, scattering-induced randomness, and measurement collectively determines the activation functions and effective dimensionality of the hidden layer, imposing limitations on its computing capacity for complex computations.

2.1. Input Layer and feature encoding

In the proposed system, an input vector $\boldsymbol{X} \in \mathbb{R}^{N_I}$ containing N_I independent and non-constant features is encoded on the wavefront by modulating the amplitude and phase of N_{enc} spatial modes $\{\boldsymbol{e}_j^{(in)}, j = 1, \cdots, N_{enc}\}$ (see Figure 1 for details), plus an additional tunable reference mode. Each spatial channel is assigned a complex encoding function $f_j(\boldsymbol{X}) = a_j(\boldsymbol{X}) \exp[ib_j(\boldsymbol{X})]$ with amplitude $a_j(\boldsymbol{X}) \in [0, 1]$ and $b_j(\boldsymbol{X}) \in [0, 2\pi]$. With this encoding, we must highlight that PELM performs the computation in the codomain of $a(\boldsymbol{X}) \exp[ib_j(\boldsymbol{X})]$ rather than $\boldsymbol{X}$ itself.

The information in $\boldsymbol{X}$ resides now in a continuous optical field, but the dimensionality of this space remains to be determined because it depends on the specific choice of the encoding functions and their spatial distribution. Collecting the spatial mode basis $\bar{\boldsymbol{e}}^{(in)} = \left[\boldsymbol{e}_1^{(in)} \ldots \boldsymbol{e}_{N_{enc}}^{(in)}\right]$ and the corresponding encoding vector $\boldsymbol{F} = [f_1(\boldsymbol{X}), \ldots, f_{N_{enc}}(\boldsymbol{X})]^T$, allows us to express the input field as $\boldsymbol{E}^{(in)}(\boldsymbol{X}) = \bar{\boldsymbol{e}}^{(in)} \boldsymbol{F}(\boldsymbol{X})$. To simplify implementation and avoid pre-processing that mixes feature information, we adopt a one-to-one mapping between input dimensions and spatial modes, i.e. f_i depends only on the i-th entry of $\boldsymbol{X}$. With this separable encoding, the number of encoded modes is $N_{enc} = N_I + 1$.

2.2. Linear Propagation in Scattering Media

The propagation of the optical beam in the scattering media is well described using the transmission matrix formalism [31] establishing a linear mapping between the encoding plane and the detection plane. Discritising the input and output apertures into N_{enc} and N_{out} channels (number of pixels or macro-pixels used for detection) we may define a complex transmission matrix $\bar{\boldsymbol{M}} \in \mathbb{C}^{N_{out} \times N_{enc}}$ that links each input mode to every output mode. The optical field reaching the camera plane is therefore $\boldsymbol{E}^{(out)} = \bar{\boldsymbol{M}} \cdot \boldsymbol{F}(\boldsymbol{X})$, a linear combination of the encoding functions.

By discretising the input and output plane using the transmission matrix formalism, we cast the optical mapping in linear algebra terms. For diffusive media, the matrix $\bar{\boldsymbol{M}}$

obeys random-matrix statistics [31] and thus can be assumed full-rank, i.e. $rank\left(\bar{\boldsymbol{M}}\right) = \min\left(N_{out}, N_{enc}\right)$. This leads to two distinct regimes. The first, dimensionality reduction, arises when $N_{enc} > N_{out}$. Here, each output feature is independent of the others, but the rank of the transmission matrix induces a bottleneck in the dimensionality of the field. Conversely, when $N_{enc} \leq N_{out}$, the scattering stage mixes the input features while preserving the dimensionality of the field and is capable of supporting a higher-dimensional space. For this reason in this manuscript, we focus on the second scenario.

2.3. Hidden Layer Output

The next stage of the implemented optical PELM is the detection at the camera plane, which will be the source of nonlinearity of the system. Indeed, each pixel will integrate contributions from the entire input aperture combined via the transmission matrix, meaning that a square-law detector records an interferometric speckle pattern of the input field. Formally, the optical read-out intensity is obtained by applying a composite function G to the output field in an element-wise manner,

$$\boldsymbol{Y}\left(\boldsymbol{X}\right) = G\left(\boldsymbol{E}^{(out)}\right) = G_s\left(I\left(\boldsymbol{E}^{(out)}\right)\right), \tag{1}$$

where the interference of coherent waves $I(\boldsymbol{E}^{(out)}) = |\boldsymbol{E}^{(out)}|^2$ results in an optical speckle that is separated from the camera optoelectronic response G_s, which can potentially introduce further non-linearity.

For a direct comparison with the formalism of ELMs, we reinterpret equation 1 by partitioning the feature vector into a data-dependent part and a constant reference, $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{F}^{'}(\boldsymbol{X}) + \boldsymbol{C}$, where $\boldsymbol{F}^{'} = \left[f_1(x_1), \ldots, f_{N_{input}}(x_{N_{input}}), 0\right]^T$ and $\boldsymbol{C} = [0, \ldots, 0, f_{ref}]^T$. Substituting into the optical read-out gives $\boldsymbol{Y}\left(\boldsymbol{X}\right) = G\left(\bar{\boldsymbol{M}}\boldsymbol{F}'\left(\boldsymbol{X}\right) + \boldsymbol{b}\right)$ which is has the conventional form of neural networks by explicitly separating the linear mapping of the input features from the bias $\boldsymbol{b} = \bar{\boldsymbol{M}}\boldsymbol{C}$ vector contributed by the tunable reference mode and randomised by the transmission matrix.

In the study of ELMs, the central object for analysis is the hidden layer output matrix, which we build from the optical read-out as

$$\bar{\boldsymbol{H}} = \begin{bmatrix} — & G\left(\bar{\boldsymbol{M}}\boldsymbol{F}'\left(\boldsymbol{X}_1\right) + \bar{\boldsymbol{M}}\boldsymbol{C}\right) & — \\ & \vdots & \\ — & G\left(\bar{\boldsymbol{M}}\boldsymbol{F}'\left(\boldsymbol{X}_{N_D}\right) + \bar{\boldsymbol{M}}\boldsymbol{C}\right) & — \end{bmatrix}, \tag{2}$$

with N_{out} columns and N_D rows. In short, from ELM theory (see Appendix A and details in reference [22]) provided that:

- the activation function G in $\bar{\boldsymbol{H}}$ is a nonlinear, non-polynomial function that is infinitely differentiable over the input domain; and
- the weights drawn from a continuous distribution,

then the matrix $\bar{\boldsymbol{H}}$ has full rank and the ELM satisfies the universal approximation property (i.e. may arbitrarily approximate a given continuous function to a desired level of accuracy).

2.4. Digital Output Layer

The final stage of the architecture is the output prediction $\boldsymbol{O}(\boldsymbol{X}) \in \mathbb{R}^{N_y}$ obtained from the hidden-layer read-out via the linear model $\boldsymbol{O}(\boldsymbol{X}) = \boldsymbol{Y}(\boldsymbol{X})\bar{\boldsymbol{\beta}}$, where $\bar{\boldsymbol{\beta}} \in \mathbb{R}^{N_{out} \times N_y}$ is the output weight matrix and with N_y being the dimensionality of the target $\boldsymbol{T}(\boldsymbol{X}) \in \mathbb{R}^{N_y}$ associated to each input vector. In the formal ELM framework, the weights are learned in a supervised manner by minimising the ℓ_2 norm featuring a unique global minimum obtained from inversion of the hidden layer output matrix. To preserve all native optical computing advantages, the digital readout involves only linear operations. Thus, it does not introduce further nonlinear transformations, and the expressive power of the overall model remains rooted in the optical stages described above.

To formalise the supervised training step for obtaining the weight matrix, we gather N_D labels into a target matrix, ordered row-wise as

$$\bar{\boldsymbol{T}} = \begin{bmatrix} — & \boldsymbol{T}(\boldsymbol{X}_1) & — \\ & \vdots & \\ — & \boldsymbol{T}(\boldsymbol{X}_{N_D}) & — \end{bmatrix},$$

so that each sample corresponds to the respective hidden layer row in $\bar{\boldsymbol{H}}$. For regression tasks, we can minimise a regularised empirical metric that balances the fit of the model with the norm of the weight matrix. Given the hidden-layer output matrix $\bar{\boldsymbol{H}}$ and the target matrix $\bar{\boldsymbol{T}}$, the optimal weights are obtained from

$$\bar{\boldsymbol{\beta}}^* = \underset{\bar{\boldsymbol{\beta}}}{\operatorname{argmin}} \left[\left\| \bar{\boldsymbol{H}}\bar{\boldsymbol{\beta}} - \bar{\boldsymbol{T}} \right\|_q^{\alpha_2} + \lambda \left\| \bar{\boldsymbol{\beta}} \right\|_p^{\alpha_1} \right],$$

where $\alpha_1, \alpha_2 > 0$, $p, q = 0, 1/2, 1, 2, \ldots, +\infty$ specify the matrix norms, while $\lambda > 0$ controls the strength of the regularisation. Among the available regularisation schemes, Ridge regression with $\alpha_1 = \alpha_2 = p = q = 2$ is particularly attractive because it admits a closed-form solution [32]

$$\bar{\boldsymbol{\beta}}^* = \left(\bar{\boldsymbol{H}}^T \bar{\boldsymbol{H}} + \lambda \bar{\mathbf{1}} \right)^{-1} \bar{\boldsymbol{H}}^T \bar{\boldsymbol{T}}, \tag{3}$$

and the Ridge parameter λ improves generalisation by penalising the ℓ_2 norm of the weights [33], being $\bar{\mathbf{1}}$ the identity matrix. This analytic expression yields the unique global minimiser in a single matrix inversion, avoiding the convergence issues of iterative solvers.

In classification settings, one can repurpose Ridge regression by treating the output $\boldsymbol{O}(\boldsymbol{X})$ as a vector of class scores. The predicted label can then be obtained as the largest component, allowing the model to serve as a multi-class classifier without additional post-processing. For binary tasks, a single output dimension suffices, with targets encoded as -1 and 1, and the sign determines the predicted class.

3. Dimensionality of the hidden space and Universal approximation capability of the PELM

As established before, the rank of the hidden-layer matrix $\bar{\boldsymbol{H}}$ determines the dimensionality of the hidden feature space having consequences on the computing capabilities of ELMs. To correctly estimate the dimensionality with the rank of $\bar{\boldsymbol{H}}$, the matrix must capture the full range of degrees of freedom in this space. A practical bottleneck arises when the training set undersamples the hidden layer, $N_D < N_{out}$. Therefore, for now, we assume the minimal condition $N_D \geq N_{out}$, ensuring each output dimension is supported by at least one observation.

To expose the combinatorial nature of the hidden space, we expand equation 2 by inserting the interferometric interaction and pulling the camera response G_s outside the matrix as its action is element-wise. Expressing the transmission matrix $\bar{\boldsymbol{M}}$ and the encoding matrix $\bar{\boldsymbol{F}}$ entry-wise reveals the row-wise repetition of transmission coefficients and a column-wise repetition of products of encoding functions. After straightforward algebra, it can be shown that the interferometric process is given by the sum of rank-1 matrices with the form

$$\bar{\boldsymbol{H}} = G_s\left(\sum_{j=1}^{N_{enc}}\sum_{k=1}^{N_{enc}}\begin{bmatrix} f_i\left(\boldsymbol{X}_1\right) f_j^*\left(\boldsymbol{X}_1\right) \\ \vdots \\ f_i\left(\boldsymbol{X}_{N_D}\right) f_j^*\left(\boldsymbol{X}_{N_D}\right) \end{bmatrix} \cdot \left[m_{1,j}m_{1,k}^*, \ldots, m_{N_{out},j}m_{N_{out},k}^*\right]\right),$$

where $m_{i,j}$ denotes the j-th element of row i of $\bar{\boldsymbol{M}}$. Each element of the interferometric hidden layer is therefore a randomly weighted sum of all pairwise products of the N_{enc} encoded modes.

In the special case of a linear camera response ($G_s(x) = x$), the expression above simplifies to a sum of rank-1 matrices in which the encoding functions are fully decoupled from the random transmission matrix. Assuming the maximum rank of the transmission matrix, the hidden layer matrix is then limited by the number of independent combinations of encoding functions and their conjugates, leading to the maximal bound $rank\left(\bar{\boldsymbol{H}}\right) \leq \min\left(N_{enc}^2, N_{out}\right)$ when both amplitude and phase encoding is used. Analogously to the output-field analysis, a bottleneck can still arise from an insufficient number of output channels.

These thresholds can be derived analytically for the specific schemes of amplitude-only or phase-only modulation, resulting in smaller upper bounds for the hidden space dimensionality. In the case of amplitude-only modulation, the encoded functions are real-valued, therefore satisfying the identity $f_i\left(\boldsymbol{X}\right) = f_i^*\left(\boldsymbol{X}\right)$. Pairwise products satisfy $f_i f_j^* = f_i^* f_j$, leading to redundancy. The number of linearly independent terms equals the unordered pairs of encoding functions (including self-products), giving the rank bound as the number of combinations

$$rank\left(\bar{\boldsymbol{H}}\right) = C_2^{N_{enc}+1}. \tag{4}$$

Besides, it is easy to conclude that for a linear response detection, amplitude modulation will lead to quadratic-polynomial activation functions.

In the case of phase-only modulation, the encoded functions are complex-valued so that $f_i(\boldsymbol{X}) \neq f_i^*(\boldsymbol{X})$. Self-products reduce to a constant, $f_i(\boldsymbol{X}) f_i^*(\boldsymbol{X}) = constant$, whereas cross-products $f_i f_j^*$ for $i \neq j$ and their conjugates $f_j f_i^*$ are distinct functions. The maximum rank is therefore

$$rank\left(\bar{\boldsymbol{H}}\right) = 2C_2^{N_{enc}} + 1. \tag{5}$$

representing all distinct cross-combinations plus one constant term. Following straightforward calculations, it is also possible to demonstrate that phase-only modulation will lead to cosine or sine-type activation functions, with a fixed frequency defined in the input space.

It is then possible to conclude that the linear camera model leads to an analytic upper bound on the hidden layer dimensionality. Under either modulation scheme, the hidden activations reduce to polynomials of the encoding functions. As a result, $\bar{\boldsymbol{H}}$ may be rank deficient if the number of input features is not sufficiently high (e.g. for low-dimensional input spaces) and thus the resulting PELM fails to satisfy the requirements set forth by ELM universal approximation theorem.

Departing from the linear detector response, it is easy to conclude also that a saturated camera regime can indeed increase the dimensionality and help in increasing the performance of the PELM. For example, considering a regime modeled by $G_{sat}(x) = x/(I_{sat} + x)$ as suggested in previous works, e.g. [24], can lead to the appearance of higher-order polynomial expansion of the input functions. Yet, two limitations may be noted in this case: first, for amplitude modulation, the activation function will still be of polynomial type, which is incompatible with the ELM requirements; second, by expanding G_{sat}, it is straightforward to conclude that higher-order corrections will become smaller and smaller, meaning that in real world conditions, may be obscured by experimental noise.

3.1. Singular Value Decomposition: Methodology for Dimensionality Estimation and Noise Robustness

In practice, one way to quantify the dimensionality of the hidden space is to compute the numerical rank of $\bar{\boldsymbol{H}}$ via a singular value decomposition (SVD) [34]

$$\bar{\boldsymbol{H}} = \bar{\boldsymbol{U}}\bar{\boldsymbol{\Sigma}}\bar{\boldsymbol{V}}^T, \tag{6}$$

where $\bar{\boldsymbol{U}}$ and $\bar{\boldsymbol{V}}$ are orthogonal matrices and $\bar{\boldsymbol{\Sigma}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_{N_{out}})$ contains the singular values ordered as $\sigma_1 \geq \ldots \geq \sigma_{N_{out}}$. In theory, the number of non-zero singular values corresponds to the rank of the matrix $\bar{\boldsymbol{H}}$. Yet, in real-world conditions, experimental noise perturbs the data, introducing small singular values along directions that do not carry information.

To circumvent this limitation, we need to establish a threshold that defines a noise level for which the number of singular values exceeding it yields the estimated

dimensionality. For this, we introduce the Weyl inequality that relates the SVD decomposition of the original matrix with one perturbed under additive noise $\bar{\boldsymbol{N}}$ [35] as

$$\|\sigma_i\left(\bar{\boldsymbol{H}}+\bar{\boldsymbol{N}}\right)-\sigma_i\left(\bar{\boldsymbol{H}}\right)\| \leq \sigma_1\left(\bar{\boldsymbol{N}}\right),\ i=1,\ldots,N_{out}.$$

where $\sigma_1\left(\bar{\boldsymbol{N}}\right)$ is the largest singular value of the noise matrix. As for indices $i > rank\left(\bar{\boldsymbol{H}}\right)$ the true singular values vanish, we have that $\sigma_i\left(\bar{\boldsymbol{H}}+\bar{\boldsymbol{N}}\right) \leq \sigma_1\left(\bar{\boldsymbol{N}}\right)$, indicating that these modes lie below a noise threshold and can be discarded. This expression highlights $\sigma_1\left(\bar{\boldsymbol{N}}\right)$ as a viable threshold to identify singular values that belong to both true and noisy matrices. To estimate this threshold, we may then average several independent realisations of the dataset to compute the residual noise, and compute its SVD, see Appendix B.

An additional feature of using the SVD decomposition is that it may also clarify the robustness of Ridge regression in the presence of noisy features. Indeed, substituting the decomposition of equation 6 into the closed form Ridge solution in equation 3 and multiplying on the right by $\bar{\boldsymbol{H}}$ yields [32]

$$\boldsymbol{\beta}^*\bar{\boldsymbol{H}} = \bar{\boldsymbol{T}}\bar{\boldsymbol{V}}\left(\sum_{j=1}^{N_{out}}\frac{\sigma_j^2}{\sigma_j^2+\lambda}\right)\bar{\boldsymbol{V}}_j^T$$

where $\bar{\boldsymbol{V}}_j$ is the j-th column of $\bar{\boldsymbol{V}}$. The quantity in parentheses is often referred to as the effective degrees of freedom as the regularisation parameter λ suppresses components with $\sigma_j^2 \ll \lambda$, effectively filtering noise, while directions with $\sigma_j^2 \gg \lambda$ are preserved with minimal distortion. Therefore, when optimising the λ for best generalisation, we expect it to achieve values above the singular values induced by noise.

4. Experimental Results and Model Validation

Having set the theoretical model, we proceed to its validation using two proof-of-concept experiments on deliberately low-dimensional input spaces: a single-parameter regression task and a two-dimensional classification task involving non-linearly separable regions. Each experiment was executed both under amplitude and phase-only encoding to validate the lower bounds extracted from the model and demonstrate the limitations of the approach.

4.1. Experimental Setup and Data Acquisition

A simplified representation of the experimental setup is presented in Figure 1. A $532nm$ diode laser with near-Gaussian wavefront is encoded by an SLM (Holoeye Pluto 2), enabling continuous feature encoding as described previously. A first polariser before the SLM ensures vertical polarisation, while a second polariser (an analyser) after the SLM enables either amplitude or phase modulation with a pre-calibrated linear response. The modulated beam is focused onto a diffusive medium with a $10\times$ microscope objective, generating a speckle pattern that is collected by a second $10\times$ objective and imaged onto a CMOS camera (Ximea MQ013MG-E2).

To investigate how the camera nonlinear response influences the hidden space dimensionality, we vary the sensor exposure time so that the detector operates either in a nearly linear regime or in saturation. For amplitude-only encoding, we set exposure times to $2ms$ and $30ms$, whereas for phase-only encoding, we use $0.5ms$ and $7ms$. This controlled variation allows us to isolate the effect of sensor nonlinearity on the effective dimensionality of the hidden layer. In total, for each experiment - regression or classification task - there are four datasets with the settings amplitude or phase encoding with low or high exposure.

The hidden layer output matrix dataset is obtained by encoding one input $\boldsymbol{X}_i$ at a time, propagating it through the system, and subsequently recording the associated speckle at the camera plane corresponding to one row of $\bar{\boldsymbol{H}}$. To characterise the hidden space dimensionality in the presence of environmental noise according to the method described in section 3.1, we record ten independent datasets for each configuration. This approach serves two purposes. First, it allows for the estimation of the Weyl threshold for identifying relevant singular values. Second, training on one dataset and evaluating with another allows the analysis of the robustness of the model to experimental noise.

Model evaluation is performed in the following way. The ten independent datasets acquired are distributed into one for training the model, four for validation, and five for testing. Within each dataset, we perform a 5-fold cross-validation, corresponding to an $80-20\%$ split of feature points into train-validation subsets. The train, validation, and test root-mean-square error (RMSE) and accuracy metrics presented are the results of the average over the 5-folds and respective dataset distribution. Ridge regression is implemented with scikit-learn, and its regularisation hyperparameter λ is tuned on the four validation sets by minimising RMSE (for regression) or maximising accuracy (for classifications). The final test scores are obtained from the validation points of the datasets withheld from both training and validation, preventing any data leakage to bias the results.

Performance in Regression Task

To evaluate the system on a simple regression task, we try to approximate the nonlinear function $g(x) = \mathrm{sinc}(x)$ in a one-dimensional input space ($N_I = 1$), corresponding to $N_{enc} = 2$ spatial modes. According to equations 4 and 5, this setting yields a theoretical hidden space dimensionality of $rank\left(\bar{\boldsymbol{H}}\right) = 3$ for both amplitude-only and phase-only encoding schemes. The input domain $x \in [-8, 8]$ is normalised to match the SLM modulation range and discretised into 64 equally spaced values.

In Figure 2.a1, for amplitude encoding under low exposure (first row and column), the camera response is approximately linear, resulting in three dominant singular components above the minimum value of the Weyl threshold (red region) as expected. For improved observation of the single value distribution and Weyl threshold we plotted their values in the Appendix B, where it can be seen that the third largest SVD is already in the bottom region of the estimated threshold. These suggest that not only

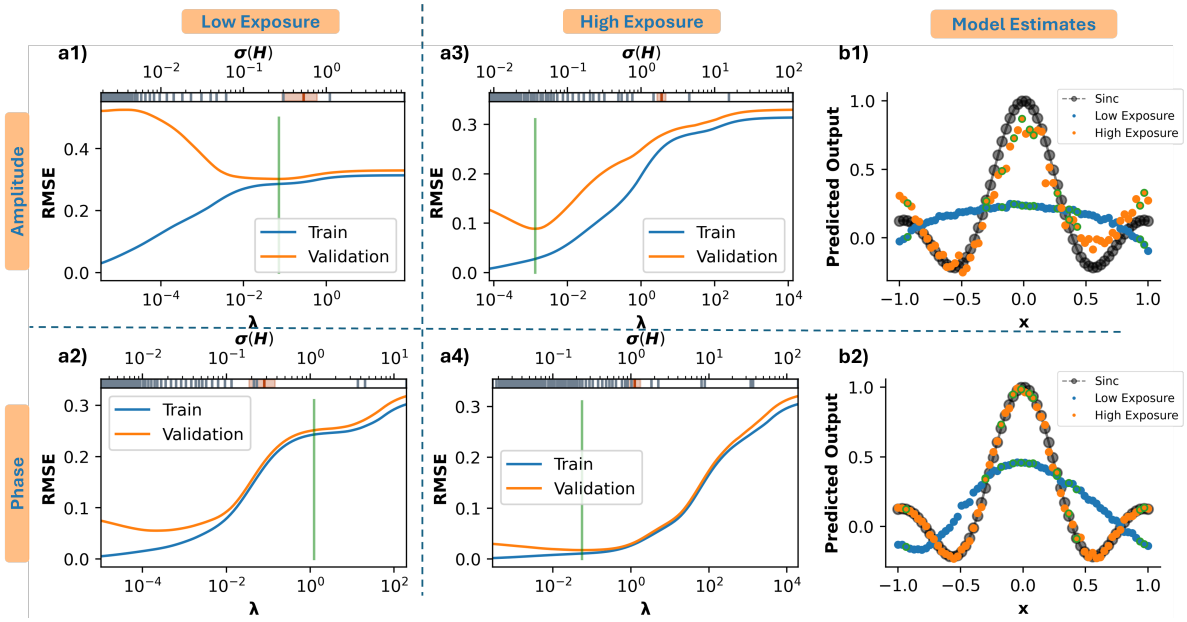

Figure 2. Regression results. (a) Train and validation RMSE versus ridge parameter λ for amplitude (top row) and phase (bottom row) encoding under low (left) and high (right) exposure settings. Gray bars above each plot show the singular value spectrum of the hidden layer such that $\sigma^2 = \lambda$, the red band marks the Weyl noise threshold. The green marker denotes the λ value chosen for testing. (b) Single fold predictions for the low and high exposure settings compared with the ground truth sinc function (dashed black) for amplitude (top row) and phase (bottom row) modulations. A green edge denotes the test points while the training points have no edge. For both modulation settings, the low exposure configuration showcases the limited expressivity of the model to the underlying encoding functions while the high exposure nonlinearity enriches the effective hidden space, resulting in higher expressivity, as reflected in an improved fit between prediction and target function.

the approximation capabilities may be affected (due to low dimensionality of the hidden space) but that noise may also strongly affect performance. Indeed, in top panel of Figure 2.b1 it is easy to observe that the predictions of the PELM in low exposure regime (blue markers) cannot approximate anything beyond second-order polynomials, as expected by their activation functions.

A similar behavior is obtained for phase-only encoding with low exposure, Figure 2.a2, with the empirical Weyl noise level estimate suggesting now a five-dimensional hidden-space rank instead of the expected three. This may indicate that some nonlinearity may be introduced by residual sensor nonlinearity caused by higher intensities throughout the variation of the input phase, or even due to the deadband of the camera in lower intensity regions. However, we shall also note that the gap between the fourth and third singular value is of one order of magnitude, reflecting the much higher participation of the expected three components. Despite the presence of sensor-induced higher-order contributions, we can still tune λ to filter the smaller contributions and still study the system in an approximation to the linear regime. For this, we set the

hyperparameter $\lambda \approx 4$ penalising the smaller components, to obtain the results present in Figure 2 b2). Again, it is easy to see that the predictions cannot approximate the target function, which is again connected to the lack of dimensionality of the output space.

Finally, to increase the dimensionality of the output space what we can do is to increase the camera exposure time, entering a nonlinear saturable regime. In Figure 2.a3, for amplitude, the optimal λ now includes a higher number of singular components, while for phase in Figure 2.a4, this can also be seen in the growth of singular components identified by the Weyls threshold to 8. Quantitatively, this lowers the test RMSE from 0.31 for low exposure to 0.13 for amplitude encoding, and from 0.25 to 0.02 for phase-only modulation. Qualitatively, in Figure 2.b1 for amplitude and Figure 2.b2 for phase modulation, the models expressivity is seen to allow improved fitting of predictions to the target function, thus leaving their polynomial-like limitation.

Performance in Classification Tasks

The classification experiment explores a two-dimensional point distribution along two distinct spirals, resulting in an encoding space with $N_{enc} = 3$. According to equations 4 and 5, we expect a hidden space dimensionality of 6 for amplitude-only encoding and 7 for phase-only encoding. Input coordinates $x, y \in [-3, 3]$ are generated with fixed noise along two spiral functions with 32 sampled points for each spiral.

Under amplitude encoding and low exposure, the singular value distribution and Weyl-based noise level estimate are shown in Figure 3.a indicating the presence of 5 singular components, one below the theoretical prediction. A corresponding jump in classification accuracy occurs when λ is chosen between the fifth and sixth singular values, corroborating an effective rank of five. The missing component remains unobserved due to the lack of variability in the mixed term xy, i.e. when either coordinate is small, the signal may be low and indistinguishable from noise. As a consequence of the limited dimensionality and expressivity, the test and validation accuracies for the optimal λ are 67% and 68%, respectively, and the model cannot generalize properly.

Under phase-only encoding and low exposure, the empirical Weyl noise level is shown in Figure 3.a2 (bottom row, see Appendix B for a clearer representation) and the cross-validation correctly estimate the expected seven significant singular components. Unlike amplitude-only encoding, phase modulation redirects the light rather than controlling the intensity, rendering the speckle intensity approximately invariant to input features, resulting in improved observation of mixed terms. Nevertheless, despite this higher hidden-space dimensionality, due to limited expressivity, the optimal test and validation accuracies remain only at 65%, similar to the amplitude-only case.

Finally, when increasing the camera exposure to induce stronger nonlinearities, we observe improved classification performance in both encoding schemes. In the case of amplitude encoding, the rank grows from 5 in low exposure to 7 for high exposure,

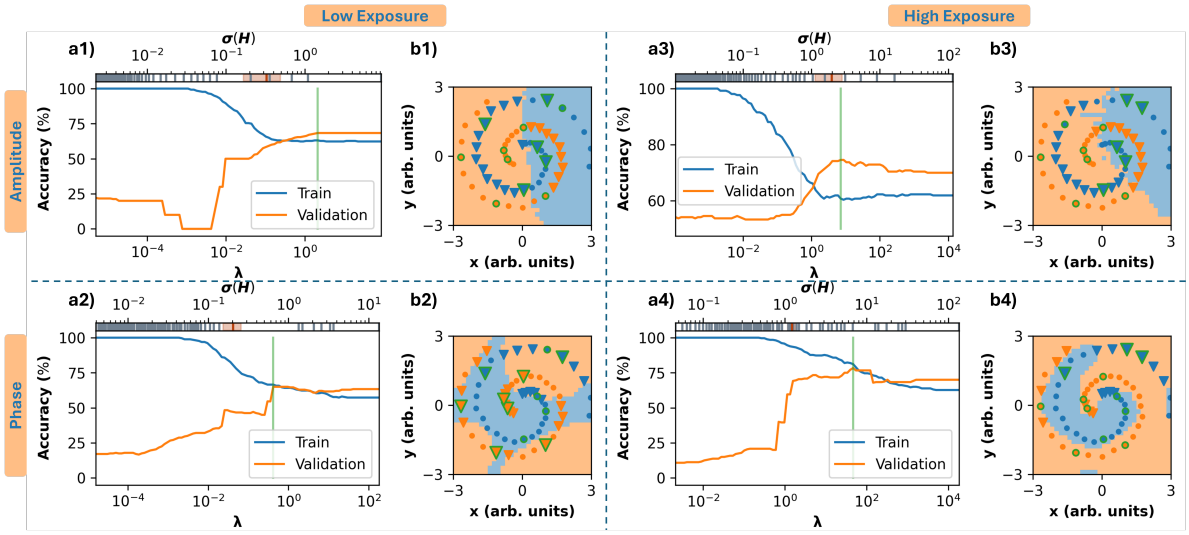

Figure 3. Classification results. Low and high exposure versus amplitude and phase encoding table with each entry showing an accuracy versus λ plot. The green marker denotes the optimal λ used for evaluation. Alongside each plot is an example of the predictions of the train (no edges) and test (green edge) data points from a single fold. Correct predictions are marked with a dot, while incorrect predictions have a triangular shape. The background of these figures is the prediction map over a 32×32 grid showcasing the prediction field used to analyse the generalisation capabilities of the model.

while for phase, it grows from 7 to 22. The cross-validation optimal model yields test and validation accuracies of 75% for amplitude-only encoding and 78% for phase-only encoding. The generalisation map in Figure 3.b4 reveal the key difference. Under amplitude-only encoding in Figure 3.b1 and 3.b3 or phase encoding with low exposure in Figure 3.b2, predictions fail to capture the spiral geometry. In contrast, phase-only predictions with high exposure are shown in Figure 3.b4 begin to conform to the spiral distribution, indicating a more accurate representation of the underlying structure.

5. Discussion and Concluding Remarks

Overall, the model and results presented provide a quantitative link between coherent light propagation and the theory of ELMs. By describing each stage of PELM separately, input encoding, complex-valued linear propagation, and detection, we can isolate the contribution of each block and how they interplay. Casting the free-space optical processor as an ELM with a complex-valued linear layer places this architecture in the family of C-ELMs, bringing the analytical tools of that framework to photonic hardware.

However, the model shows that the function space of the hidden space is ultimately limited by the input encoding functions. With a purely linear camera response, the network can only form polynomials of the pairwise combinations of input encoding functions. As a result, the hidden layer dimensionality is upper-bounded and depends solely on the number of input modes, failing to show the expected

universal approximation capabilities predicted for ELM. For amplitude and phase modulation, taken together or separately, we derived closed-form expressions for that bound, capturing the underlying combinatorial process. Operating the experiment at low exposures confirmed these predictions, where we estimate the dimensionality of the system using a Weyl estimation of noise level, used to exclude residual components. Increasing the detector exposure, in turn, introduced additional non-linearity, enriching the hidden space and lifting several more singular values above the noise floor.

These results imply that enriching the input space of functions, through pre-processing or multiple encodings, cannot overcome the dimensionality limitation and therefore falls short of the universal approximation property. Modifying the detector function, here achieved by raising the exposure time, does increase the effective dimensionality, pointing to a practical route towards a more expressive PELM framework. However, the experimental observation of the full-rank condition for saturable nonlinearity on the detection side was not possible due to the presence of noise.

In summary, this manuscript clarifies the relationship between the standard ELM algorithm and its photonic implementation. While PELMs have shown limitations, they remain attractive thanks to their versatility and energy efficiency for tasks that profit from rapid, hardware-level feature expansion, in particular if one adds a nonlinear propagation layer instead of a simple diffusive layer[36, 37].

Acknowledgements

This work is co-financed by Component 5 - Capitalization and Business Innovation, integrated in the Resilience Dimension of the Recovery and Resilience Plan within the scope of the Recovery and Resilience Mechanism (MRR) of the European Union (EU), framed in the Next Generation EU, for the period 2021 - 2026, within project HfPT, with reference 41, and by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the support UID/50014/2023 (https://doi.org/10.54499/UID/50014/2023). Nuno A. Silva acknowledges the support of FCT under the grant 2022.08078.CEECIND/CP1740/CT0002 (https://doi.org/10.54499/2022.08078.CEECIND/CP1740/CT0002).

Author contributions

V.R. theoretical model, analysed results and wrote manuscript; N.A.S. conceived experiments, theoretical model, and wrote manuscript; D.S. theoretical model; F.C.M. conducted the experiments, analysed results and reviewed the manuscript; T.D.F. conducted the experiments and reviewed the manuscript; C.S.M. reviewed the manuscript.

Data Availability

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

[1] Ravi Athale and Demetri Psaltis. Optical computing: past and future. *Optics and Photonics News*, 27(6):32–39, 2016.

[2] David J Richardson, John M Fini, and Lynn E Nelson. Space-division multiplexing in optical fibres. *Nature photonics*, 7(5):354–362, 2013.

[3] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005.

[4] A Vander Lugt. Signal detection by complex spatial filtering. *IEEE Transactions on information theory*, 10(2):139–145, 1964.

[5] RA Athale, HH Szu, and JN Lee. Optical implementation of integral transforms with bessel function kernels. *Optics Letters*, 7(3):124–126, 1982.

[6] Yichen Shen, Nicholas C Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, et al. Deep learning with coherent nanophotonic circuits. *Nature photonics*, 11(7):441–446, 2017.

[7] James Spall, Xianxin Guo, Thomas D Barrett, and AI Lvovsky. Fully reconfigurable coherent optical vector–matrix multiplication. *Optics Letters*, 45(20):5752–5755, 2020.

[8] Maxime W Matthès, Philipp Del Hougne, Julien De Rosny, Geoffroy Lerosey, and Sébastien M Popoff. Optical complex media as universal reconfigurable linear operators. *Optica*, 6(4):465–472, 2019.

[9] Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific reports*, 8(1):12324, 2018.

[10] Changming Wu, Heshan Yu, Seokhyeong Lee, Ruoming Peng, Ichiro Takeuchi, and Mo Li. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nature communications*, 12(1):96, 2021.

[11] Nikolay L Kazanskiy, Muhammad A Butt, and Svetlana N Khonina. Optical computing: Status and perspectives. *Nanomaterials*, 12(13):2171, 2022.

[12] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[13] Nabil H Farhat, Demetri Psaltis, Aluizio Prata, and Eung Paek. Optical implementation of the hopfield model. *Applied optics*, 24(10):1469–1475, 1985.

[14] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[15] Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David AB Miller, and Demetri Psaltis. Inference in artificial intelligence with deep optics and photonics. *Nature*, 588(7836):39–47, 2020.

[16] Xing Lin, Yair Rivenson, Nezih T Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008, 2018.

[17] Xuhao Luo, Yueqiang Hu, Xiangnian Ou, Xin Li, Jiajie Lai, Na Liu, Xinbin Cheng, Anlian Pan, and Huigao Duan. Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible. *Light: Science & Applications*, 11(1):158, 2022.

[18] Che Liu, Qian Ma, Zhang Jie Luo, Qiao Ru Hong, Qiang Xiao, Hao Chi Zhang, Long Miao, Wen Ming Yu, Qiang Cheng, Lianlin Li, et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nature Electronics*, 5(2):113–122, 2022.

[19] Minjia Zheng, Lei Shi, and Jian Zi. Optimize performance of a diffractive neural network by controlling the fresnel number. *Photonics Research*, 10(11):2667–2676, 2022.

[20] Loubnan Abou-Hamdan, Emil Marinov, Peter Wiecha, Philipp del Hougne, Tianyu Wang, and Patrice Genevet. Programmable metasurfaces for future photonic artificial intelligence. *Nature Reviews Physics*, pages 1–17, 2025.

[21] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, volume 2, pages 985–990. Ieee, 2004.

[22] Guangbin Huang, Qin-Yu Zhu, and Chee Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70:489–501, 2006.

[23] Ming-Bin Li, Guang-Bin Huang, Paramasivan Saratchandran, and Narasimhan Sundararajan. Fully complex extreme learning machine. *Neurocomputing*, 68:306–314, 2005.

[24] Davide Pierangeli, Giulia Marcucci, and Claudio Conti. Photonic extreme learning machine by free-space optical propagation. *Photonics Research*, 9(8):1446–1454, 2021.

[25] Carlo M. Valensise, Ivana Grecco, Davide Pierangeli, and Claudio Conti. Large-scale photonic natural language processing. *Photonics Research*, 10(12):2846–2853, 2022.

[26] Brandon Redding, Joseph B Murray, Joseph D Hart, Zheyuan Zhu, Shuo S Pang, and Raktim Sarma. Fiber optic computing using distributed feedback. *Communications Physics*, 7(1):75, 2024.

[27] Stefano Biasi, Riccardo Franchi, Lorenzo Cerini, and Lorenzo Pavesi. An array of microresonators as a photonic extreme learning machine. *APL Photonics*, 8(9), 2023.

[28] José Roberto Rausell-Campo, Antonio Hurtado, Daniel Pérez-López, and José Capmany Francoy. Programmable photonic extreme learning machines. *Laser & Photonics Reviews*, 19(9):2400870, 2025.

[29] Antoine Liutkus, David Martina, Sébastien Popoff, Gilles Chardon, Ori Katz, Geoffroy Lerosey, Sylvain Gigan, Laurent Daudet, and Igor Carron. Imaging with nature: Compressive imaging using a multiply scattering medium. *Scientific reports*, 4(1):5552, 2014.

[30] Alaa Saade, Francesco Caltagirone, Igor Carron, Laurent Daudet, Angélique Drémeau, Sylvain Gigan, and Florent Krzakala. Random projections through multiple optical scattering: Approximating kernels at the speed of light. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6215–6219. IEEE, 2016.

[31] Sébastien M Popoff, Geoffroy Lerosey, Rémi Carminati, Mathias Fink, Albert Claude Boccara, and Sylvain Gigan. Measuring the transmission matrix in optics: An approach to the study and control¡? format?¿ of light propagation in disordered media. *Physical review letters*, 104(10):100601, 2010.

[32] Trevor Hastie, Robert Tibshirani, Jerome Friedman, et al. The elements of statistical learning, 2009.

[33] Peter L Bartlett. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE transactions on Information Theory*, 44(2):525–536, 2002.

[34] Gilbert Strang et al. *Linear algebra and learning from data*, volume 4. Wellesley-Cambridge Press Cambridge, 2019.

[35] Roger A Horn and Charles R Johnson. *Topics in matrix analysis*. Cambridge university press, 1994.

[36] Nuno A Silva, Vicente Rocha, and Tiago D Ferreira. Optical extreme learning machines with atomic vapors. *Atoms*, 12(2):10, 2024.

[37] Pierre Azam and Robin Kaiser. Optically accelerated extreme learning machine using hot atomic vapors. *Physical Review Applied*, 22(3):034041, 2024.

Appendix A. ELM framework

The ELM algorithm is summarised in three steps [21], described as follows. First, it assigns an arbitrarily distributed weight matrix $\bar{\boldsymbol{w}}$ and bias vector $\boldsymbol{b}$ which can be real or complex valued in the case of complex ELMs [23], producing a random, nonlinear mapping of the input features $\boldsymbol{X} = [x_1, \ldots, x_N]^T$ into the hidden space via the transformation $g\left(\bar{\boldsymbol{w}}\boldsymbol{X} + \boldsymbol{b}\right)$ where g is the element-wise nonlinear activation function.

Second, it constructs the hidden layer output matrix with N_D training samples along the rows and N hidden nodes along the columns,

$$\bar{\boldsymbol{\mathcal{H}}} = \begin{bmatrix} g\left(\bar{\boldsymbol{w}}_1 \boldsymbol{X_1} + \boldsymbol{b}\right) & \dots & g\left(\bar{\boldsymbol{w}}_N \boldsymbol{X_1} + \boldsymbol{b}\right) \\ \vdots & \dots & \vdots \\ g\left(\bar{\boldsymbol{w}}_1 \boldsymbol{X_{N_D}} + \boldsymbol{b}\right) & \dots & g\left(\bar{\boldsymbol{w}}_N \boldsymbol{X_{N_D}} + \boldsymbol{b}\right) \end{bmatrix} \tag{A.1}$$

where the index of $\bar{\boldsymbol{w}}$ is the row of the weight matrix. In the final step, the hidden layer output matrix is used to compute the linear readout $\boldsymbol{\beta}$ solving $\bar{\boldsymbol{\mathcal{H}}}\boldsymbol{\beta} = \boldsymbol{T}$, where $\boldsymbol{T}$ contains the training labels. This makes the hidden layer output matrix the core subject in the universal approximation proof for ELMs [22]. By guaranteeing that it has full rank independently of the network size, the inverse of $\bar{\boldsymbol{H}}$ always exists, guaranteeing the linear regression has a solution. The proof of this theorem leverages the randomness of the weights, such that $\bar{\boldsymbol{w}}_{\boldsymbol{j}} \boldsymbol{X}_k \neq \bar{\boldsymbol{w}}_{\boldsymbol{j}} \boldsymbol{X}'_k$ for all $k \neq k'$ and the infinitely differentiable activation function resulting in a proof where the columns of the matrix cannot belong to any space with fewer dimensions than N.

Appendix B. Rank Distributions

To estimate the singular value spectrum and the Weyls threshold we adopt the following procedure. Let $\mathcal{D} = \{\bar{\mathbf{H}}^{(i)}\}_{i=1}^{M}$ be the set of hidden layer output matrices obtained from M independent experimental realisations. For each realisation we compute the residual by subtracting the ensemble mean,

$$\bar{\boldsymbol{N}}^i = \bar{\boldsymbol{H}}^i - \frac{1}{N} \sum_{j=1}^{M} \bar{\boldsymbol{H}}^j,$$

where $i = 1, \dots, N$. A SVD of each $\bar{\boldsymbol{N}}^i$ and $\bar{\boldsymbol{H}}^i$ yield the largest singular value of the residual $\sigma_1^i\left(\bar{\boldsymbol{N}}\right)$ and the spectra of singular values $\sigma_j^i\left(\bar{\boldsymbol{H}}^{\boldsymbol{i}}\right)$ for $j = 1, \dots, min\left(N_{out}, N_D\right)$, respectively. By averaging $\sigma_1^i\left(\bar{\boldsymbol{N}}\right)$ and $\sigma_j^i\left(\bar{\boldsymbol{H}}^{\boldsymbol{i}}\right)$ for $j = 1, \dots, min\left(N_{out}, N_D\right)$ over the M realisations we obtain the Weyl threshold and spectra used for analysis.

The results for the regression task in Figure B1 are obtained with this procedure. For amplitude encoding with low or high exposure we denote 3 singular values above the Weyl threshold while for phase we identify 5 and 8, respectively.

In the classification task with results presented in Figure B2, low exposure yields 5 and 7 significant singular values for amplitude and phase encoding, while high exposure increases these numbers to 6 and 22, respectively.

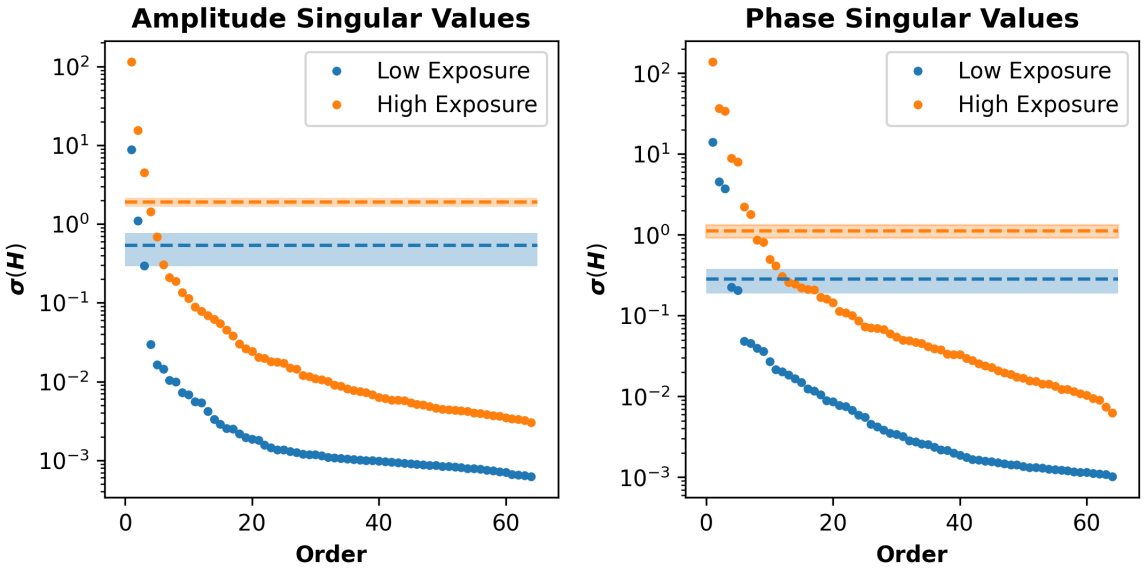

Figure B1. Singular value spectra and Weyl threshold for the regression configuration. In the left amplitude encoding spectra for low and high exposure, identifying 3 singular values above the Weyl threshold (highlighted region). Similar setup on the right for phase encoding we identify 5 and 8 singular values above the threshold.

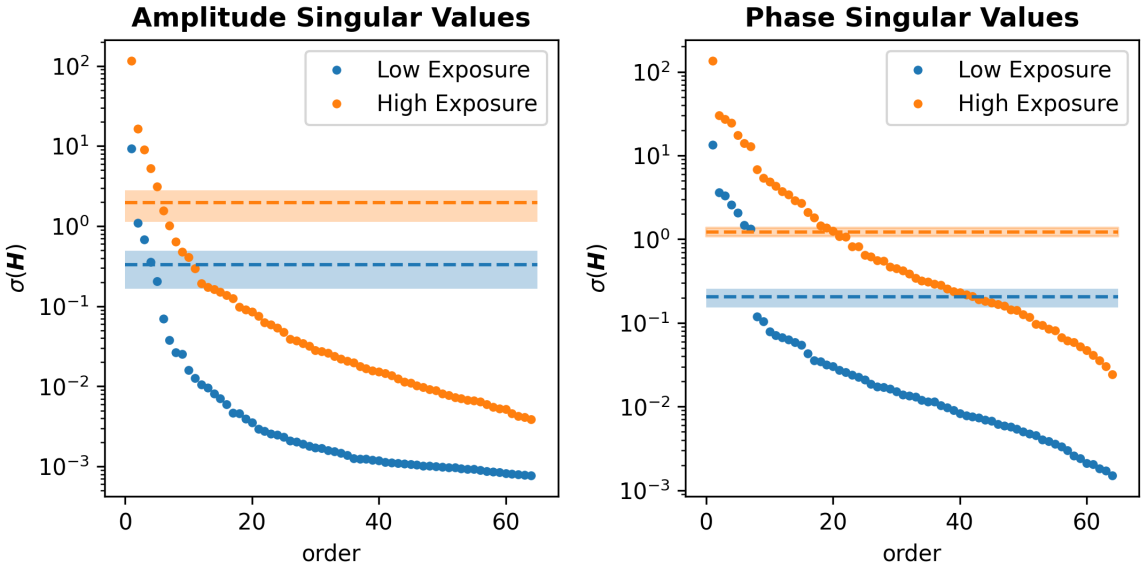

Figure B2. Singular value spectra and Weyl threshold for the classification configuration. In the left amplitude encoding spectra for low and high exposure, identifying 5 and 6 singular values above the Weyl threshold (highlighted region). Similar setup on the right for phase encoding we identify 7 and 22 singular values above the threshold.