

Supplementary Material

Optimising Complexity and Learning for Photonic Reservoir Computing with Gain-Controlled Multimode Fibres

Giulia Marcucci¹ Luana Olivieri² and Juan Sebastian Totoro Gongora^{2,*}

¹LumiAlres Ltd, Glasgow, G12 0DG, United Kingdom

²Emergent Photonics Research Centre, Dept. of Physics, Loughborough University, Loughborough, LE11 3TU, United Kingdom

Correspondence*:

Juan Sebastian Totoro Gongora

Emergent Photonics Research Centre, Dept. of Physics, Loughborough University, Loughborough, LE11 3TU, United Kingdom

j.totoro-gongora@lboro.ac.uk

S1 MODEL DERIVATION

In the presence of a low refractive index contrast, we express the signal and pump fields in the fibre as a superposition of LP-modes $\psi_m(\mathbf{x})$ and $\phi_n(\mathbf{x})$, respectively (1, 2). In the following, we will consider a singly polarised electric field distribution, but our results can be extended to a full-vector case straightforwardly. We denote the signal and pump mode amplitudes as $S_m(z)$ and $P_n(z)$, with z the propagation direction, and the propagation constants of the signal and pump modes as β_m and β'_n , respectively. With this position, the signal and pump fields are expressed as follows:

$$\begin{aligned} E_{signal}(\mathbf{x}, z) &= \sum_m S_m(z) \psi_m(\mathbf{x}) \exp(i\beta_m z), \\ E_{pump}(\mathbf{x}, z) &= \sum_j P_j(z) \phi_j(\mathbf{x}) \exp(i\beta'_j z), \end{aligned} \quad (S1)$$

where the slowly varying amplitudes are expressed in units of $[V/m]$. The total transverse intensity distributions (expressed in $[W/m^2]$) at the signal λ_{signal} and pump λ_{pump} wavelengths read as follows:

$$\begin{aligned} I_{signal}(\mathbf{x}, z) &= \frac{1}{2} \epsilon_0 c \sum_{mn} S_m(z) S_n^*(z) \psi_m(\mathbf{x}) \psi_n^*(\mathbf{x}) \exp[i(\beta_m - \beta_n)z] \\ I_{pump}(\mathbf{x}, z) &= \frac{1}{2} \epsilon_0 c \sum_{jk} P_j(z) P_k^*(z) \phi_j(\mathbf{x}) \phi_k^*(\mathbf{x}) \exp[i(\beta'_j - \beta'_k)z]. \end{aligned} \quad (S2)$$

In the presence of doping ions, the dynamics of the electromagnetic field propagating and interacting in the fibre depend on the spatially varying population operators $N_1(\mathbf{x}, z)$ and $N_2(\mathbf{x}, z)$ for the ground and excited levels of the ions, whose steady-state expressions read as follows:

$$\begin{aligned} N_2(\mathbf{x}, z) &= \frac{N_d(\mathbf{x}) [I_p(\mathbf{x}, z) \sigma_a(\lambda_p) + I_s(\mathbf{x}, z) \sigma_a(\lambda_s)]}{\tau_{21} + I_p(\mathbf{x}, z) \sigma_a(\lambda_p) + I_s(\mathbf{x}, z) [\sigma_a(\lambda_s) + \sigma_e(\lambda_s)]}, \\ N_1(\mathbf{x}, z) &= N_d(\mathbf{x}) - N_2(\mathbf{x}, z), \end{aligned} \quad (S3)$$

where $N_d(\mathbf{x})$ is the spatial distribution of dopants (z-invariant), $\sigma_{e,a}(\lambda)$ are the emission and absorption scattering cross-sections, and $I_p(\mathbf{x}, z)$, $I_s(\mathbf{x}, z)$ denote the total intensity profiles of

the signal and pump fields (2, 3, 4, 5). The evolution equations for the system read as follows:

$$\begin{aligned}\partial_z S_m(z) &= i\beta_m S_m(z) + (1 + iK_{ric}) \sum_n S_n(z) G_{mn}^s(z), \\ \partial_z P_j(z) &= i\beta'_j P_j(z) + (1 + iK_{ric}) \sum_k P_k(z) G_{jk}^p(z),\end{aligned}\tag{S4}$$

where K_{ric} is the refractive index change constant (4). We define the nonlinear coupling matrices

$$\begin{aligned}G_{mn}^s &= \exp [i(\beta_m - \beta_n)z] \int d\mathbf{x} \psi_m(\mathbf{x}) \psi_n^*(\mathbf{x}) [\sigma_e(\lambda_s) N_2(\mathbf{x}, z) - \sigma_a(\lambda_s) N_1(\mathbf{x}, z)], \\ G_{jk}^p &= -\exp [i(\beta'_j - \beta'_k)z] \int d\mathbf{x} \phi_j(\mathbf{x}) \phi_k^*(\mathbf{x}) [\sigma_a(\lambda_p) N_1(\mathbf{x}, z)].\end{aligned}\tag{S5}$$

In our simulations, we consider the case $\sigma_e(\lambda_p) = 0$, corresponding to the realistic conditions for an Er-doped fibre (5). Eqs (S4)-(S5) capture the main nonlinear effects originating from the interaction of the signal and pump fields with the dopant ions, and include saturation and refractive index effects. Due to the explicit dependence of $N_1(\mathbf{x}, z)$ and $N_2(\mathbf{x}, z)$ on the local intensity distributions, the signal and pump modes are coupled in a nonlinear fashion. The coupling matrices $G_{mn}^s(z)$ and $G_{jk}^p(z)$ depend on the local population densities, which in turn depend on the local intensity distributions. This self-consistent coupling is crucial for accurately capturing the dynamics of the system.

S2 SYSTEM PARAMETERS

We considered an Er-doped, step-index multi-mode fibre of length $L=1$ m and a core diameter of $d=30 \mu\text{m}$. In our system, the numerical aperture of the fibre is set as $NA = 0.32$ and the core index is set as $n_{core} = 1.45$, leading to a number of 94 and 227 supported modes per polarisation at the signal and pump wavelengths, respectively. For the gain medium, we assumed a dopant density $N_d = 2 \times 10^{25}$, a slow decay time $\tau_{12} = 1 \text{ ms}$, and we used the following values for the absorption and emission cross-sections:

- $\sigma_a(\lambda_p)=1.7\text{e-}25 \text{ m}^2$
- $\sigma_e(\lambda_s)=4\text{e-}25 \text{ m}^2$
- $\sigma_a(\lambda_s)=2.25\text{e-}25 \text{ m}^2$

Without loss of generality, we assumed the spatial density of dopants $N_d(\mathbf{x}, z)$ within the fibre as constant and uniform across the fibre core. As discussed in the main text, we consider a total illumination area $A = 31.8\mu\text{m} \times 31.8\mu\text{m}$ corresponding to approximately $1.5/\sqrt{2}$ times the fibre core diameter and a pixel size $\Delta x = 1.06\mu\text{m}$. The pixel size corresponds to roughly one internal wavelength (λ_s/n_{core}) for the signal field.

S3 NUMERICAL APPROACH

We solved Eqs. (S4)-(S5) using a modal evolution approach based on a Runge-Kutta 45 ODE solver (as implemented in MATLAB ode45). The evolution equations were defined and solved in cylindrical coordinates with radial and angular grids of 200 points each, ensuring accurate representation of the transverse field profiles. At each propagation step, the population densities $N_1(\mathbf{x}, z)$ and $N_2(\mathbf{x}, z)$ were updated self-consistently using the local intensities $I_s(\mathbf{x}, z)$ and $I_p(\mathbf{x}, z)$. The nonlinear coupling matrices $G_{mn}^s(z)$ and $G_{jk}^p(z)$ were computed at each step by numerically integrating over the transverse coordinates. The initial modal amplitudes $S_m(0)$ and $P_j(0)$ were set by projecting the input and pump field distributions onto the corresponding mode profiles, as detailed in Eqs. (8) in the main manuscript. The resulting set of coupled ODEs was integrated along the fibre length to obtain the evolution of the modal amplitudes and the output field distributions under a leading order approximation to optimise the computational workload.

S4 DETAILS ON THE CALCULATION OF SAMPLE AND FEATURE KOLMOGOROV COMPLEXITY (KC)

While the main manuscript primarily discusses the sample KC, we also computed an additional metric, denoted as feature KC, that provides complementary insight into the variability and structure of each pixel's response across the full dataset.

As described in the main text, we employ the Kolmogorov Complexity (KC), approximated numerically via the Lempel–Ziv compression algorithm, to quantify the structural complexity of the reservoir outputs. Below, we describe the approach used to compute two variants of KC.

The Sample KC quantifies the complexity across features (pixels) for each reservoir output (sample). For each output sample (a single CCD image, represented as a column of the readout matrix \mathbf{H}), we binarized the pixel intensities by comparing each pixel to the mean intensity across all pixels within that sample. Specifically, pixel values above the mean are set to 1, and those below are set to 0. This produces a binary string representing each reservoir output, capturing its spatial complexity. We then apply the Lempel–Ziv algorithm (as described by Kaspar and Schuster (6)) to measure the compressibility of this binary sequence. The resulting complexity value is normalized by the complexity of a randomly generated binary sequence of identical length, ensuring that a normalized $KC = 1$ corresponds to fully random (incompressible) output, and $KC < 1$ indicates structured (compressible) outputs. This procedure is repeated independently for each sample, and the final sample KC reported in the manuscript figures (main text Fig. 8a–c) is the mean KC averaged across all samples in the dataset. In Fig. 8d–f of the main text, we include the standard deviation of the KC statistics normalised to the mean to assess the degree of variability of the sample KC across the dataset.

The Feature KC, on the contrary, quantifies complexity across samples for each feature (CCD pixel). For each pixel (a row in the readout matrix \mathbf{H}), we binarize the pixel's intensity values by comparing each pixel intensity across all samples to the pixel's mean intensity across the dataset. Pixel values above this mean are set to 1 and those below to 0, yielding a binary sequence per pixel across the entire dataset. Again, we compute the complexity of each binary sequence via the Lempel–Ziv algorithm and normalize it against the complexity of a random binary sequence of equivalent length. Feature KC thus captures the input-driven complexity of each spatial feature (pixel). The results are shown in Supplementary Figure S9. Note that, while the sample KC is an intrinsic feature of the reservoir output, the feature KC depends on the specific ordering of the dataset. For the regression case, we sorted the input data u_n . For the classification case, the input data is two-dimensional and randomly shuffled, leading to KC close to or above one naturally. Still, some regions are visible even in this case, highlighting the remarkable capability of the KC to identify structural complexity even within outputs associated with random inputs. Finally, for the time-series case, the data has a specific ordering dictated by the temporal evolution, leading to significantly low sample KC across the whole range of parameters.

REFERENCES

1. Buck JA. *Fundamentals of Optical Fibers*. Wiley Series in Pure and Applied Optics (Hoboken, NJ: Wiley-Interscience), 2. ed edn. (2004).
2. Trinel JB, Cocq GL, Quiquempois Y, Andresen ER, Vanvincq O, Bigot L. Theoretical study of gain-induced mode coupling and mode beating in few-mode optical fiber amplifiers. *Optics Express* **25** (2017) 2377–2390. doi:10.1364/OE.25.002377.
3. Sperber T, Billault V, Sebbah P, Gigan S. Wavefront Shaping of the Pump in Multimode Fiber Amplifiers; The Gain-Dependent Transmission Matrix. *arXiv:1911.07812 [physics]* (2019).
4. Sperber T, Billault V, Dussardier B, Gigan S, Sebbah P. Gain As Configurable Disorder-Adaptive Pumping for Control of Multimode Fiber Amplifiers and Lasers. *arXiv:2008.04085 [physics]* (2020).
5. Agrawal GP, Premaratne M. *Light Propagation in Gain Media_ Optical Amplifiers* (Cambridge University Press) (2011).

6 .Kaspar F, Schuster HG. Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review A* **36** (1987) 842–848. doi:10.1103/PhysRevA.36.842. Publisher: American Physical Society.

SUPPLEMENTARY FIGURES

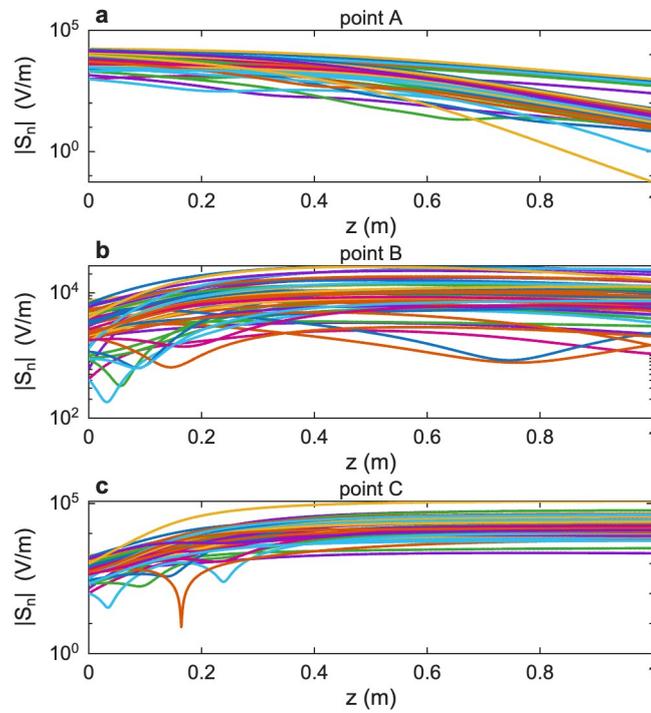


Figure S1. Modal amplitudes a-c Semi-logarithmic plot of the modal amplitudes $\log_{10} |S_n(z)|$ as a function of the propagation distance z for the same cases shown in Fig. 2 of the main text. Here, panel a corresponds to Fig. 2c, panel b corresponds to Fig. 2d, and panel c corresponds to Fig. 2e.

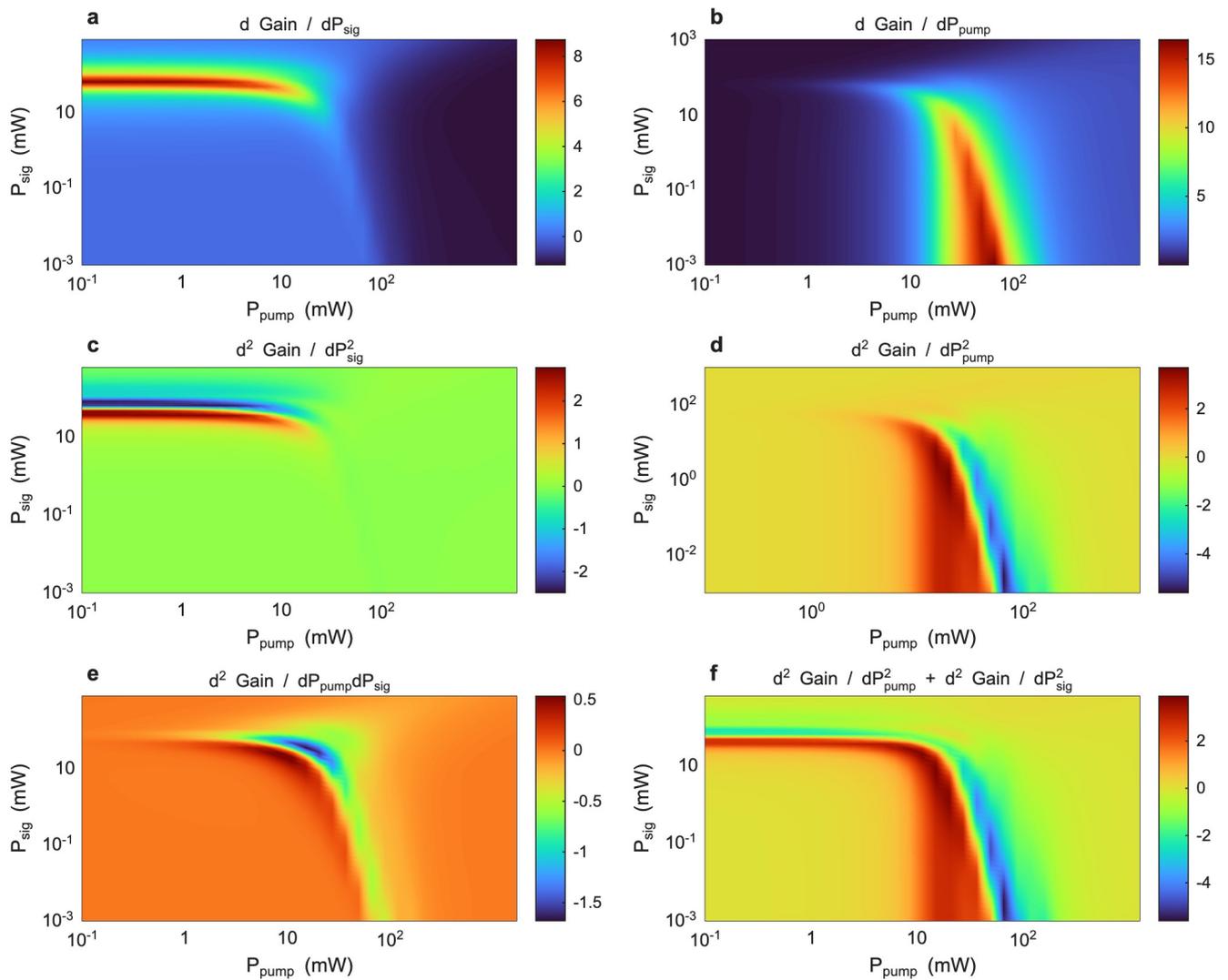


Figure S2. Gain variation analysis Each panel represents a different derivative of the total gain across the signal and pump powers. **a** First derivative with respect to the pump power dG/dP_p (same as Fig. 2b in the main text). **b** First derivative of the total gain versus signal power dG/dP_s . **c** Second derivative of the total gain versus pump power d^2G/dP_p^2 . **d** Second derivative derivative of the total gain versus signal power d^2G/dP_s^2 . **e** Mixed derivative of the total gain versus signal and pump powers d^2G/dP_sP_p . **f** Laplacian of the total gain versus signal and pump powers $d^2G/dP_s^2 + d^2G/dP_p^2$.

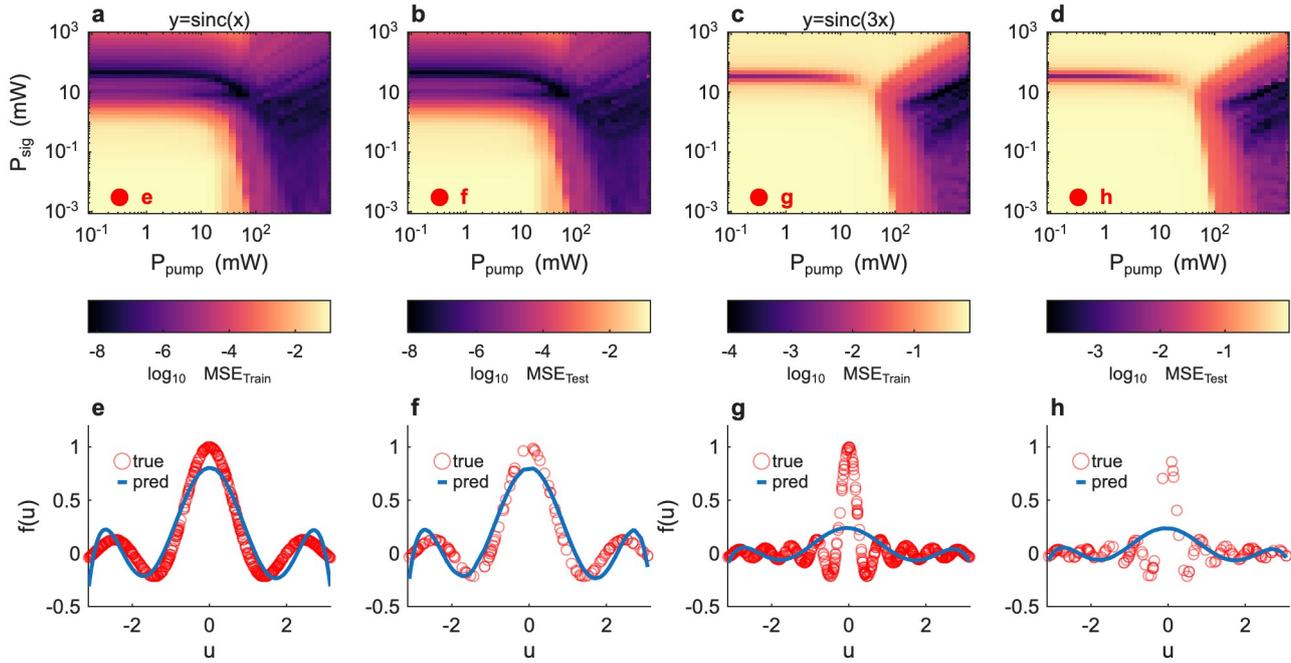


Figure S3. Nonlinear Regression. a,b Training (panel a) and Testing (panel b) MSE error for the task $u^{(k)} \rightarrow \text{sinc}(u^{(k)})$ as a function of the signal P_s and pump P_p powers. The MSE is plotted on a logarithmic scale (i.e., $\log_{10} MSE$). The red dots correspond to the prediction results shown in panels e and f, respectively. c,d Same as panels a and b, but for the case $u^{(k)} \rightarrow \text{sinc}(3u^{(k)})$. e-h Plot of the predicted regression output (solid blue line) and the ground truth values (red circles) for the testing datasets. Each panel corresponds to the operating points highlighted in panels a-d.

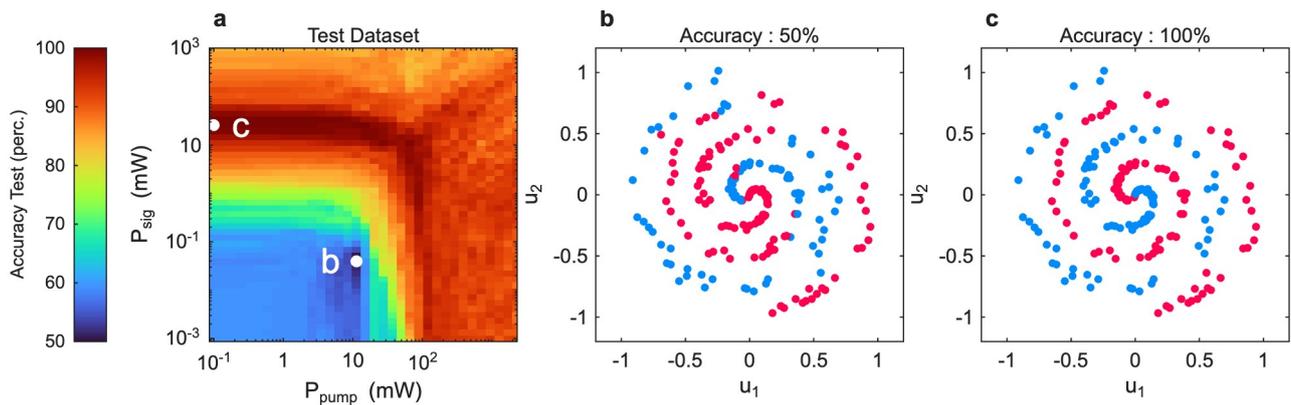


Figure S4. Spiral classification examples a Testing classification accuracy (%) as a function of the signal P_s and pump P_p powers. The accuracy is defined as the percentage of correct classifications. The white dots correspond to poor (50% accuracy, panel b) and good (100% accuracy, panel c) classification accuracy. b, c Plot of the spiral points (u_1, u_2) for the two configurations highlighted in panel a. The blue dots correspond to a predicted class $c^{(k)} = 0$ and the purple dots to a predicted class $c^{(k)} = 1$.

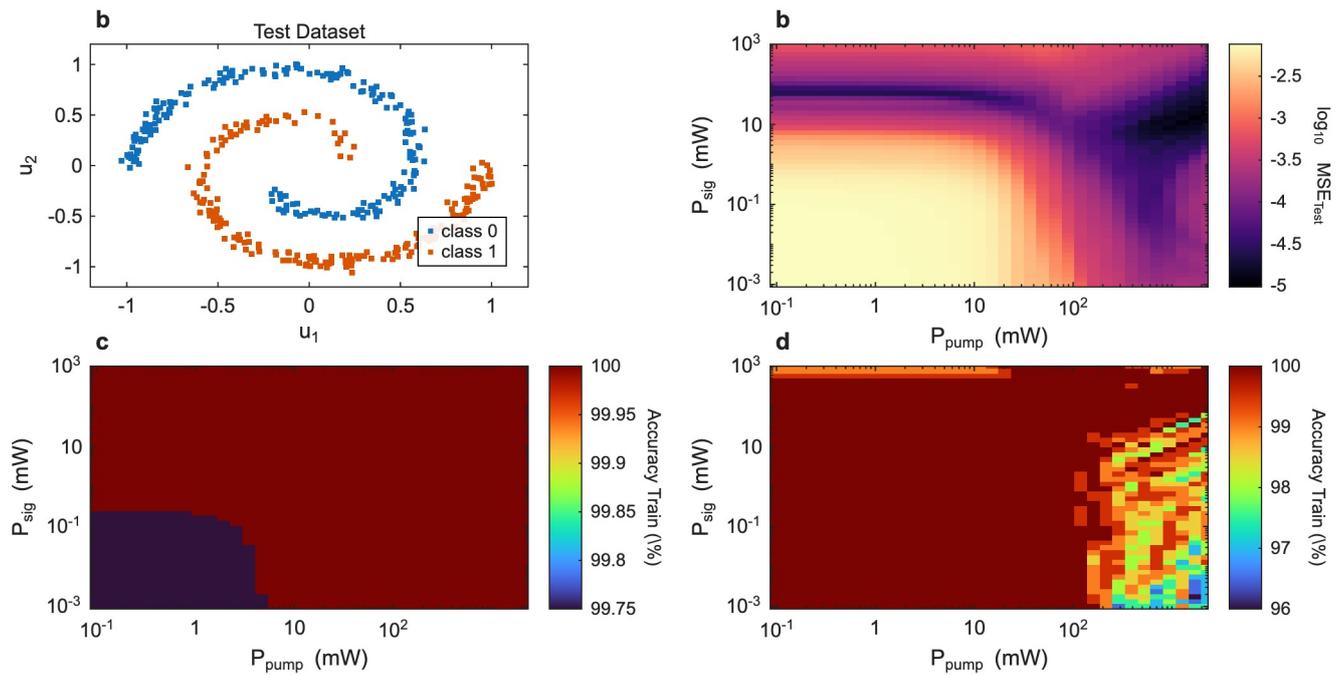


Figure S5. Spiral dataset classification **a** Test dataset for the spiral dataset as defined by Eqs. (13) of the main text. The two spiral arms are denoted as blue dots (class $c^{(k)} = 0$) and orange dots (class $c^{(k)} = 1$). Differently from the case reported in the main text, here the winding number is set as $\chi = 1$, leading to an easily classifiable dataset. **b** MSE for the training dataset for the task $(u_1^{(k)}, u_2^{(k)}) \rightarrow c^{(k)}$ as a function of the signal P_s and pump P_p powers. The MSE is plotted on a logarithmic scale (i.e., $\log_{10} MSE$). Here, we report the raw MSE obtained via ridge regression before winner-takes-all classification is implemented, essentially comparing the raw prediction outputs to the target values. **c,d** Training (panel c) and Testing (panel d) classification accuracy as a function of the signal P_s and pump P_p powers. The accuracy is defined as the percentage of correct classifications. The white crosses denote points with 100% accuracy. In the highly nonlinear regimes, the testing accuracy drops below 90%, suggesting overfitting.

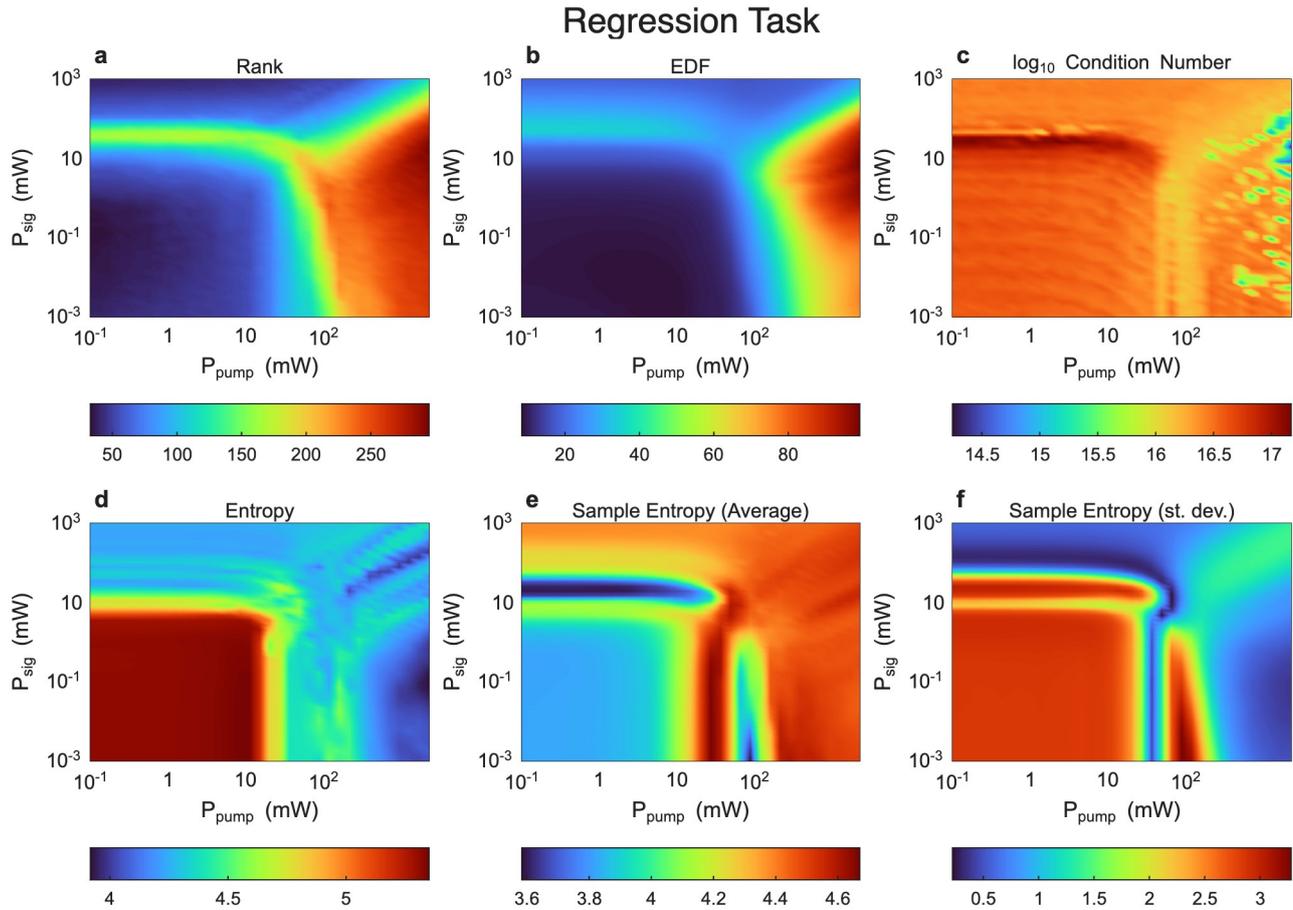


Figure S6. Linear Algebra metrics of the readout matrix \mathbf{H} - nonlinear regression task.. **a-c** Matrix rank (panel a), *edf* (panel b) and condition number (panel c) of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers for the nonlinear sinc regression. The size of the readout matrix is $N \times M_{train}$ with $M_{train} = 500$. **d** Total Shannon entropy of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers. The entropy is computed across all entries of the readout matrix. **e** Average sample Shannon entropy of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers. For each combination (P_s, P_p) , the sample entropy is computed for each CCD output and averaged across all samples. **f** Relative variability of sample entropy across the dataset as a function of the signal P_s and pump P_p powers. The relative variability is the ratio of the standard deviation associated with each average in panel e divided by the corresponding average.

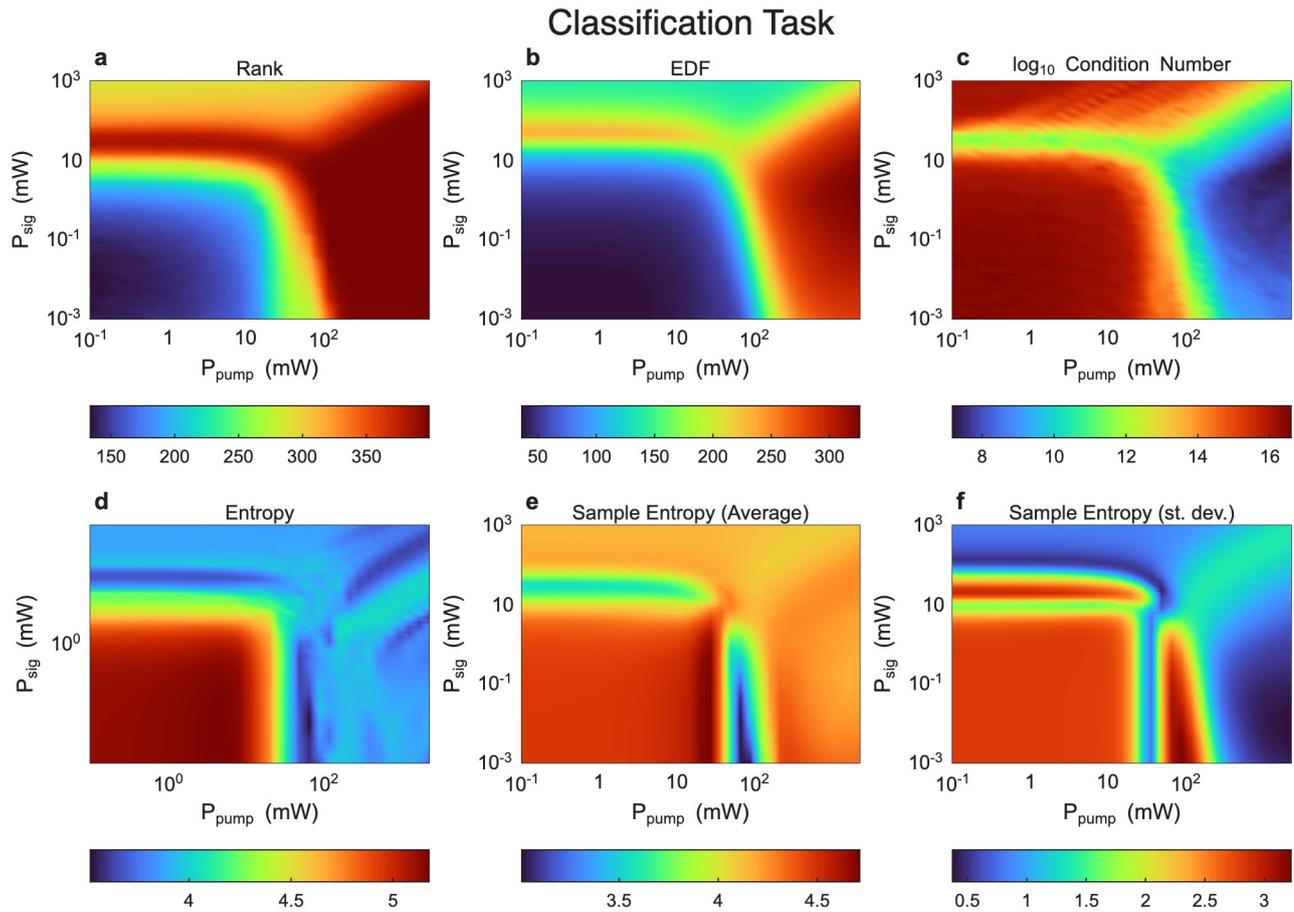


Figure S7. Linear Algebra metrics of the readout matrix \mathbf{H} - spiral classification task..
a-c Matrix rank (panel a), *edf* (panel b) and condition number (panel c) of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers for the nonlinear sinc regression. The size of the readout matrix is $N \times M_{train}$ with $M_{train} = 400$. **d** Total Shannon entropy of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers. The entropy is computed across all entries of the readout matrix. **e** Average sample Shannon entropy of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers. For each combination (P_s, P_p) , the sample entropy is computed for each CCD output and averaged across all samples. **f** Relative variability of sample entropy across the dataset as a function of the signal P_s and pump P_p powers. The relative variability is the ratio of the standard deviation associated with each average in panel e divided by the corresponding average.

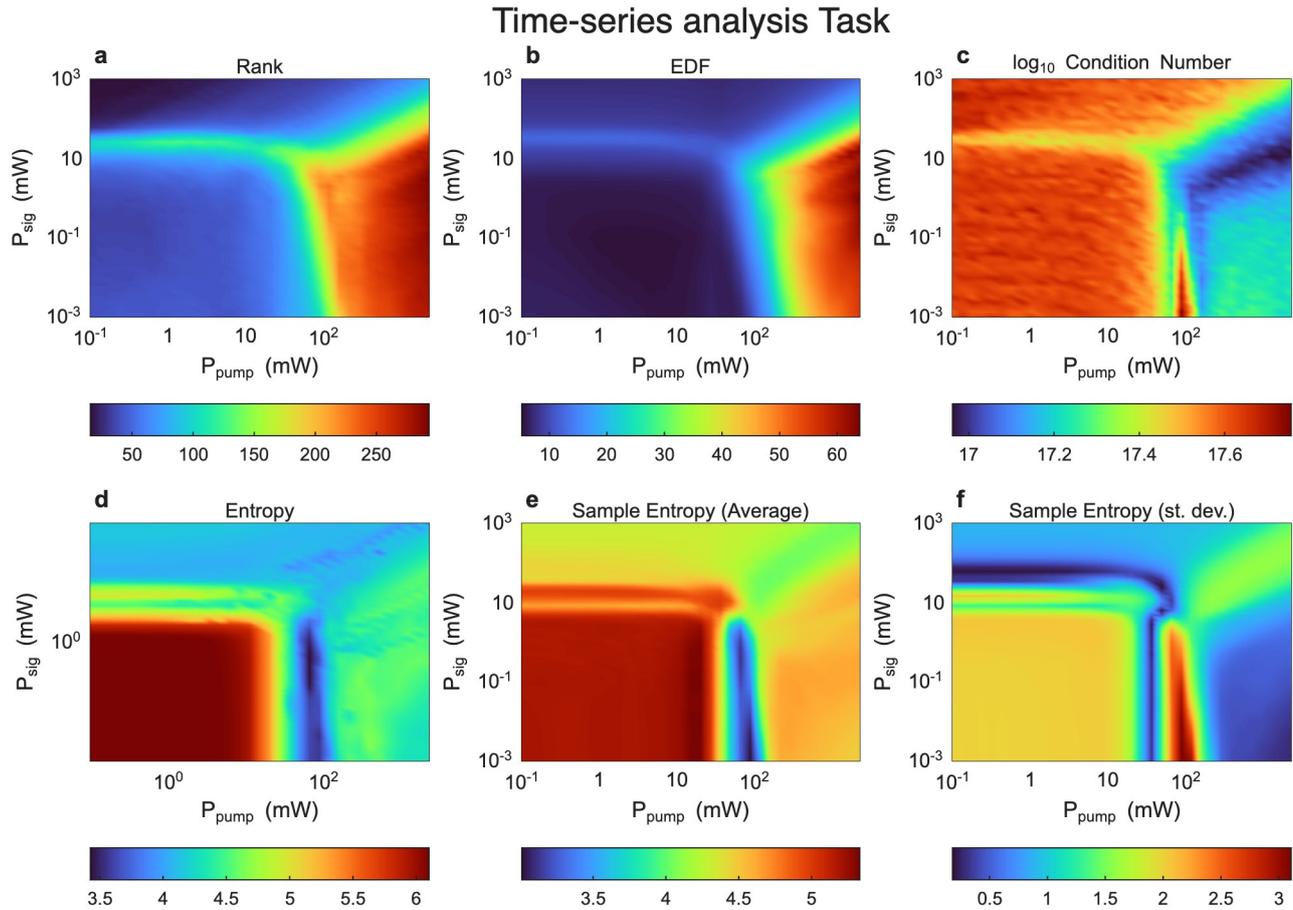


Figure S8. Linear Algebra metrics of the readout matrix \mathbf{H} - time-series prediction tasks..

In this case, all the metrics are computed on the raw readout matrix \mathbf{H} before applying the delay embedding postprocessing. **a-c** Matrix rank (panel a), *edf* (panel b) and condition number (panel c) of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers for the nonlinear sinc regression. The size of the readout matrix is $N \times M_{train}$ with $M_{train} = 1000$. **d** Total Shannon entropy of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers. The entropy is computed across all entries of the readout matrix. **e** Average sample Shannon entropy of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers. For each combination (P_s, P_p) , the sample entropy is computed for each CCD output and averaged across all samples. **f** Relative variability of sample entropy across the dataset as a function of the signal P_s and pump P_p powers. The relative variability is the ratio of the standard deviation associated with each average in panel e divided by the corresponding average.

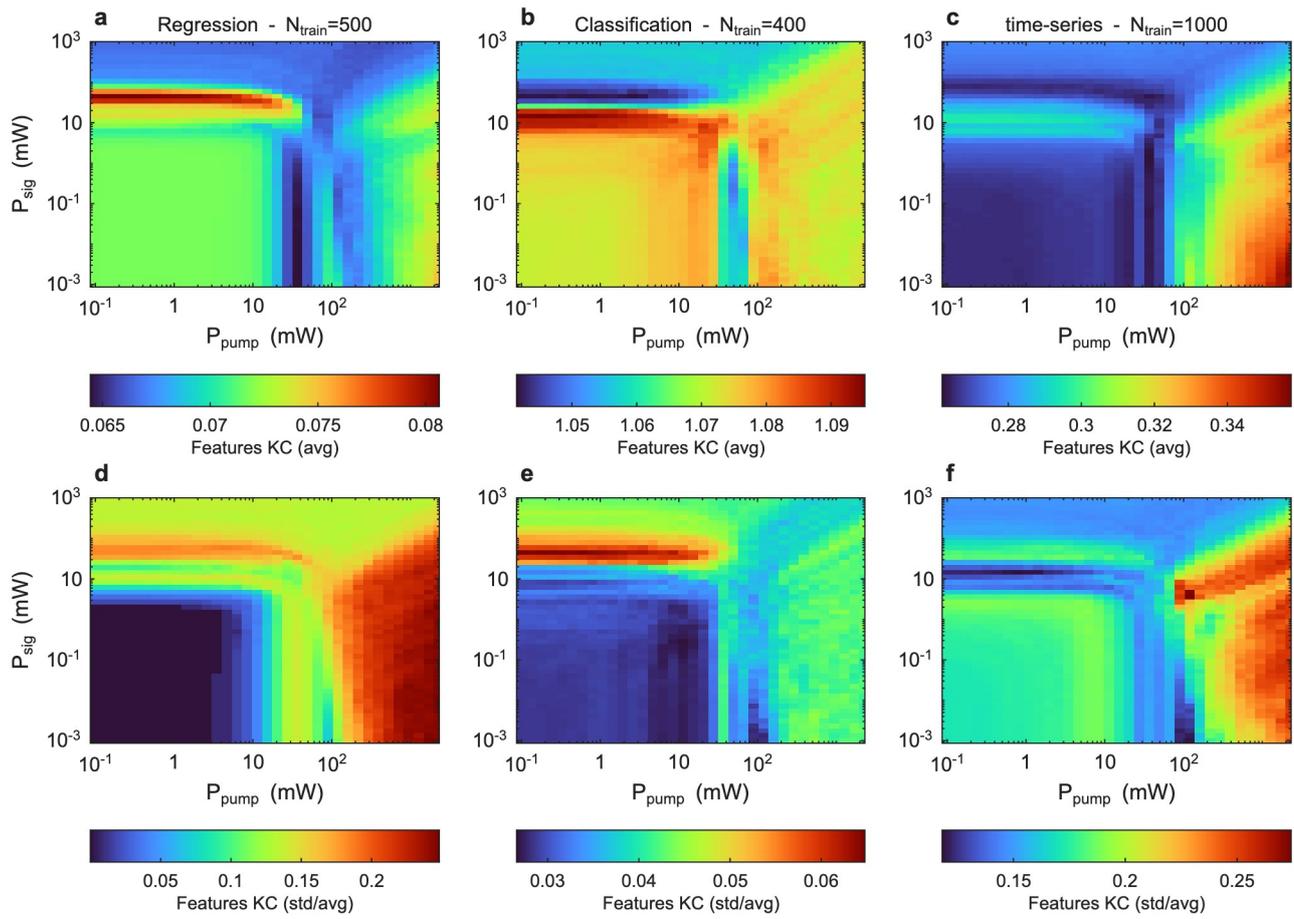


Figure S9. Feature Kolmogorov Complexity (KC) of the readout matrix \mathbf{H} . **a-c** Average feature KC of the readout matrix \mathbf{H} as a function of the signal P_s and pump P_p powers for the regression (panel a), classification (panel b) and time-series prediction (panel c) tasks. For each combination (P_s, P_p) , the feature KC is computed for each CCD pixel and averaged across all features. The number of features is $N = 900$ for all datasets. **d-f** Relative variability of feature KC across the dataset as a function of the signal P_s and pump P_p powers for the regression (panel d), classification (panel e), and time-series prediction (panel f) tasks. The relative variability is the ratio of the standard deviation associated with each average in panels (a-c) divided by the corresponding average.