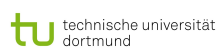PHOTONIC AND NANOMETRIC HIGH-SENSITIVITY BIO-SENSING

DELIVERABLE 2.4 [UNIFI]

REPORT ON THE EFFECTS OF THE PROPOSED NOISE FILTERS ON COMPLEX/BIOLOGICAL SYSTEMS

EFFECTS OF THE PROPOSED NOISE FILTERS ON COMPLEX/BIOLOGICAL SYSTEMS
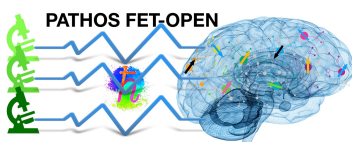
Work Package 2

Contents

PATHOS FET-OPEN

European Commission

1 Deep learning enhanced noise spectroscopy

1.1 NV sensing

Quantum sensing combines theoretical results with experimental and engineering techniques to carry out inference of signals with improved accuracy and/or less computation time by making use of quantum physics. In the past years different classes of applications have emerged, employing quantum systems as sensors for various physical quantities ranging from magnetic and electric fields, to time and frequency, rotations, temperature and pressure. Quantum sensors take advantage of the central weakness of quantum systems, i.e. their strong sensitivity to external disturbances, to improve the detection of external perturbations with higher accuracy compared to any classic sensor. Notable examples include atomic clocks, atomic vapor magnetometers, and superconducting quantum interference devices. However, this same property implies that the quantum sensor is subjected to detrimental noise stemming from the coupling with its environment. For this reason, it is desirable to fully characterize the sensor's environment, either to filter out its detrimental effect, or to take it into account when detecting external signals, for example, in algorithms using quantum optimal control. On the other side, neural networks (NNs) could be a powerful tool to infer the sensor's environment. In this context, deep learning has been already proposed theoretically for the classification and detection of quantum noise features, and employed experimentally for different tasks. NNs are algorithmic models provided by the interconnection of a group of nodes commonly called neurons whose weights are trained to optimize the model output given an input in form of features values. The optimization is usually performed with gradient descent techniques and the outputs can be categorical for classification tasks or continue for regression tasks. In our work (recently published as S. Martina, S. Hernández-Gómez, S. Gherardini, F. Caruso, N. Fabbri. Deep learning enhanced noise spectroscopy of a spin qubit environment. Mach. Learn.: Sci. Technol. 4 02LT01, 2023), we theoretically and experimentally demonstrated that NNs can be used to process the data obtained by a qubit, operating as a quantum sensor, and then reconstruct the noise spectrum that induces dephasing into the qubit itself. In particular, we focused on a qubit under dynamical decoupling (DD) control sequences in the presence of classical random noise with an unknown power density spectrum, usually denoted as noise spectral density (NSD).
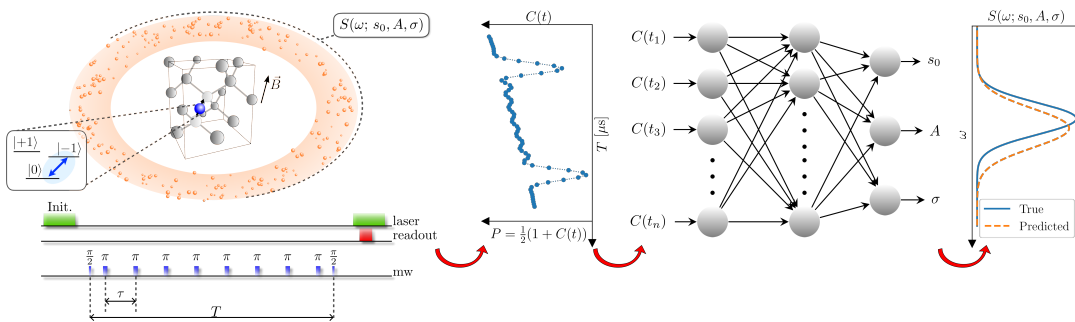

Figure 1: NV center and Neural Networks for noise spectroscopy.

We numerically tested our machine learning (ML) algorithms and we also used a single NV center in diamond as a spin qubit sensor to theoretically and experimentally perform a spectroscopic reconstruction of the magnetic noise of its local environment. The NV center represented in the left part of Figure 1, surrounded by a depiction of the environment, is constituted by a lattice of ^{13}C carbon atoms with a defect in the structure where a carbon atom is substituted by a nitrogen atom and a vacancy who act as a qubit. The environment comprises ^{13}C nuclear spins randomly distributed in the diamond lattice. The undesired interaction of a quantum system with its environment generally leads to a coherence decay of superposition states in time. The dephasing affecting the qubit sensor is analyzed by applying a set of DD control pulses that realize filter functions in the frequency domain.

PATHOS FET-OPEN
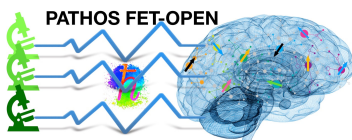
A widely used DD control pulse is the Carr–Purcell (CP) sequence that is given by N equidistant π pulses, performed between an initial and a final $\pi/2$ pulse. With this protocol, the requirement to achieve high values of the noise reconstruction accuracy is to perform sequences with a high number of pulses meaning $N \in [30, 120]$ or even higher. This usually leads to long experiments to reconstruct the whole spectrum of the noise. Other techniques using non-equidistant or even more sophisticated DD sequences have proved to be effective for noise sensing, but sometimes at the price of a higher computational burden.

For our sensing task, NNs are designed to solve a multi-regression problem, i.e. the reconstruction of the NSD. We assume that the NSD of the bath of spins has a Gaussian profile. The Gaussian NSD is thus parametrized as a function of key parameters, i.e. the mean value, variance, offset and noise power that we aim to predict. The NNs were trained over a set of synthetic data generated by simulating how the coherence of the qubit sensor decays over time under the influence of both the CP control pulses and the NSD. Moreover, to make the measurement statistics as close as possible to the ones obtained from the experiments, extra artificial errors sampled from a normal distribution were added.

Our approach using NNs entails the following advantages that we have proven experimentally:

(i)   NNs have the capability to predict never-before-seen experimental data, and they can work with a better reconstruction accuracy (even up to seven times better) than standard noise spectroscopy, as the Álvarez–Suter method, by making use at the same time of DD control sequence with a much smaller number of pulses.

(ii)   The training dataset, which can contain both synthetic and experimental data, is generated just once and then it can be applied several times, as long as the new collected data reproduce the physical context under analysis. The amount of data used as input to the NNs can be smaller than the one needed to resolve the NSD by means of standard noise spectroscopy methods.

To the best of our knowledge, our work is the first experimental proof of enhanced reconstruction performance with NNs for carrying out noise spectroscopy in single color centers in diamond. We thus expect that the techniques discussed there could fast become a novel standard spectroscopy tool both for such quantum systems and other quantum platforms in which regression problems have to be solved.

We carried out noise spectroscopy with a quantum sensor using DD sequences with a much smaller number of π pulses and, at the same time, achieving a higher reconstruction accuracy than standard methods. This means that with our proposal the noise spectroscopy procedure will take less time and give better results. More in detail, we experimentally demonstrate the capability of NNs to reconstruct the NSD of the collective nuclear spin bath that surrounds an electronic spin qubit, i.e. the ground state of a single NV center in bulk diamond at room temperature. We are also confident that the extent of our results can be quite easily replicated in other experimental settings, as e.g. superconducting flux qubits, trapped ions, cold atoms, quantum dots, nuclear magnetic resonance (NMR) experiments in molecules, and nanoelectronic devices. For such a purpose, one might slightly adapt the deep learning techniques used in our work to methods tailored for time series. To facilitate this task we shared via GitHub the source code along with the (gold open-access) paper, available at https://github.com/trianam/noiseSpectroscopyNV.

1.2   Results and analysis

The training dataset is composed of synthetic data that are originated by simulating the coherence decay of the qubit sensor in a noise spectroscopy experiment based on DD, as the one depicted in Figure 1. This standard sensing procedure, which stems from Ramsey interferometry, maps information about the quantum coherence of the sensor into the population in $|0\rangle$ that is then effectively
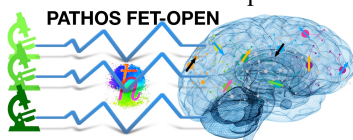
recorded. The probability that the state of the quantum sensor is $|0\rangle$, which corresponds to the observable population, equals to

$$P = \frac{1}{2}\big(1 + C(\tau, N)\big),$$

where N is the number of π pulses and τ is the time between them. The coherence function $C(\tau, N)$ was simulated numerically, for a set of different values of τ and N, to generate the training dataset. In order to simulate the data, the NSD $S(\omega)$ is parameterized as

$$S(\omega) = s_0 + A exp\left(\frac{-(\omega - \omega_c)^2}{2\sigma^2}\right).$$

Thus, being a Gaussian distribution, the NSD is fully described by the offset s_0, amplitude A, width σ and center ω_c. For the training dataset in our work, the values of these parameters were taken randomly from defined intervals. Instead, ω_c was kept constant because it depends on the known static magnetic field aligned with the NV quantization axis z. The training dataset was generated by uniformly sampling 10^4 sets of parameters within the chosen intervals, each one used to simulate a different choerence curve $C(\tau, N)$ for a different set of τ and N. Finally, in order to make the synthetic data used to train the NNs closer to the experimental setting, extra artificial errors sampled from a normal distribution, comparable with the expected error in our experimental measurements, were added to every point of the generated coherence decay curves. From the 10000 simulated curves, 6000 are used for the training of the NNs and 2000 for their validation. Instead, the test step is performed either by using the remaining 2000 simulated curves, or by using experimental data.

As first analysis, we trained ML models to infer the value of the NSD parameters s_0, A, σ from the coherence functions and we tested such models on simulated data with different parameters to evaluate the Mean Squared Error (MSE) between the latters and the predicted values. In order to determine the smallest amount of data required to reconstruct the NSD, we perform the training, validation and test of the NNs with sub-sets of the simulated curves. These sub-sets are defined by introducing the variable $\bar{N}$ that denotes the upper bound for the number of pulses $N \leqslant \bar{N}$ considered during the whole process. The sub-sets defined for each value of $\bar{N}$ contain the curves for all the times τ and for all the different NSD parameters in the training, validation and test sets. In detail, the input of the NN is defined as the concatenation of all the values of $C(\tau, N)$, for all values of τ and for $N = 1, \ldots, \bar{N}$.
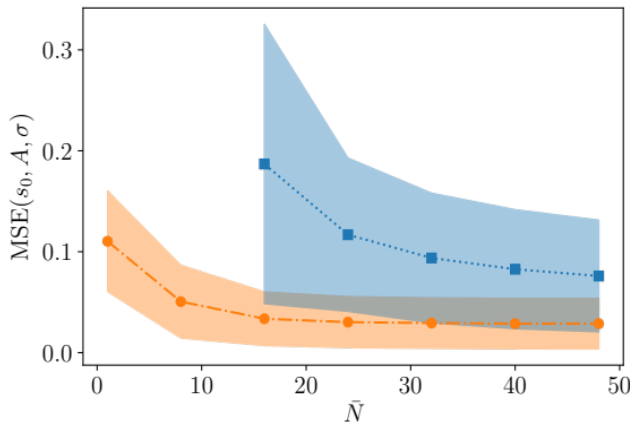

Figure 2. Mean-square-errors (MSEs) between original and estimated NSD parameters for a set of 2000 simulated test cases. Orange bullets with dash-dotted line are the mean values returned by NNs. Blue squares with dotted line are the mean values provided by the HS method. Shaded areas denote the standard deviation of all the cases.

The results of this analysis is shown in Figure 2 with the orange circles. Remarkably, the MSE seems to achieve its minimum value after $\bar{N} = 16$. This entails that the NNs do not significantly improve their precision on the reconstruction of the NSD by using more data to train them beyond this point.
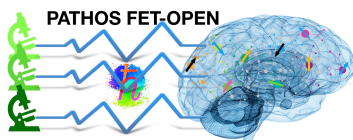
To establish how accurately an NN reconstructs the NSD, we compared the corresponding results with those of a different commonly adopted method denoted as harmonics spectroscopy (HS). We have analyzed the same 2000 different curves C(τ, N) used to test the ML models also with the HS method and the results are shown in Figure 2 with the blue squares. The MSE values for the HS method are always above the MSE values for the NN method, especially for lower values of $\bar{N}$. These results demonstrate that the NN method can predict the parameters of the NSD with an improved accuracy (up to five times larger) with respect to the HS method.

As second analysis, we used the NNs, trained and validated with noisy simulated data, to reconstruct the NSD from experimental data. In contrast with the tests using simulated data, in the experimental case we do not know the exact values of the NSD parameters. Therefore, we cannot calculate the MSE to quantify the accuracy of the reconstructed parameters. In order to estimate such accuracy we have used the following procedure: from the inferred NSD, the coherence curves C(τ, N) are simulated and then compared with the curves collected from experimental results. An example of this comparison is shown in Figure 3(a), where C(τ, N) is simulated under the assumption that the NSD parameters are inferred either by the ML models (orange) or by the HS method (red), both for $\bar{N} = 16$. Qualitatively, it is clear that the orange curves are much closer to the experimental data, than the red curves. To quantitatively compare the experimental data and the simulation results, we used both the reduced chi-squared and the MAE between the experimental data and the predicted coherence functions C(τ, N). The results of this comparison are shown in Figure 3(b). Remarkably, the NSD reconstructed by the NN for $\bar{N} = 16$ behaves better that any case using the HS method.
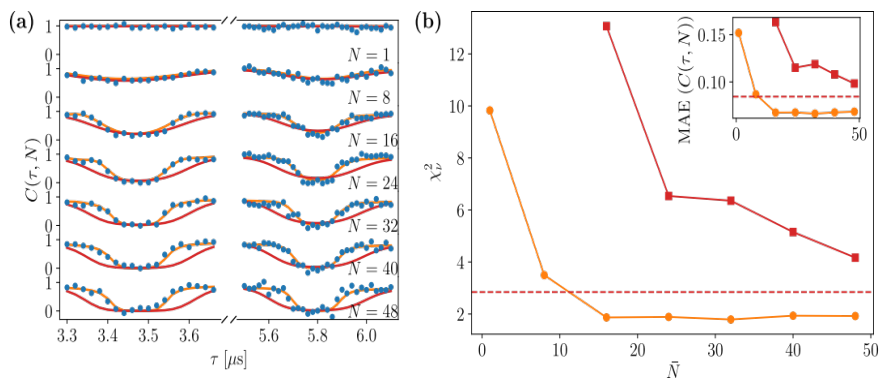

Figure 3: (a) Coherence function C(τ, N). The experimental data (blue bullets) are shown together with the simulated ones using the NSD predicted respectively by the HS method (red lines) and ML models (orange lines), both for $\bar{N} = 16$. (b) Reduced chi-squared χ_ν^2, obtained by comparing simulation and experimental data, as a function of $\bar{N}$. orange and red curves refer to the NN and HS method, respectively. Instead, the dashed line denotes the value of the reduced chi-squared for the HS method when we employ additional measurements. In the inset: same results but quantified by the mean-absolute-error (MAE).

We demonstrated that NNs can be used to reconstruct the NSD affecting a quantum sensor, **achieving higher precision and with considerable less data** than the standard HS method. Improved values of the reconstruction accuracy have been obtained with simulated and experimental data. Both the HS and NN methods are comparable—in terms of NSD reconstruction accuracy—for high values of $\bar{N}$, but not for small ones, where NNs give significantly better results. The main result of our investigation is that NNs trained with data obtained for $\bar{N} = 16$ reconstruct the NSD more accurately than the best estimate provided by the HS method with $\bar{N} = 48$. This improvement is remarkable by itself, but it becomes more significant when we consider that the time required to complete these experiments has a growth faster than a linear function with respect to $\bar{N}$, following an arithmetic progression.
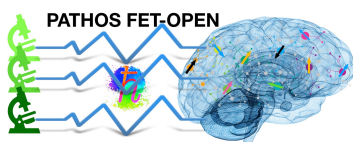
2 Machine-learning based high-bandwidth magnetic sensing

2.1 Magnetic Sensing

Quantum technologies have emerged as an important platform relevant for a broad range of fields, such as quantum communications and quantum sensing. These advances have been driven by the development of experimental realizations exhibiting useful properties. In the context of quantum sensing, one of the leading systems is based on Nitrogen Vacancy (NV) color centers in diamond, which provide a versatile platform for diverse quantum sensing, notably magnetic sensing. NVs have found important applications in magnetic sensing, covering paleomagnetometry, biosensing, nuclear magnetic resonance and more. Quantum sensing with NVs is realized through spin resonance measurements, usually detected optically. While this approach achieves quantitative vectorial information with high sensitivity and spatial resolution, it suffers from a trade-off between sensitivity and bandwidth, specifically in the high dynamic range regime. In fact, that working with small fields (small dynamic range) enables an optimal sensing strategy, which relies on precise measurements at a predetermined high-sensitivity point (the point of maximal signal gradient). However, this is not possible in the regime of large dynamic range signals.

In our work (G. Haim, S. Martina, N. Bar-Gill, F. Caruso, paper in preparation), being a collaboration between two PATHOS partners (UNIFI and HUJI), we addressed this limitation and introduced a ML algorithm which significantly improves this trade-off. We demonstrate experimentally that training an appropriate neural network using a combination of real data and simulated data enables a clear improvement in measurement bandwidth for a given sensitivity goal, in the large dynamic range scenario.
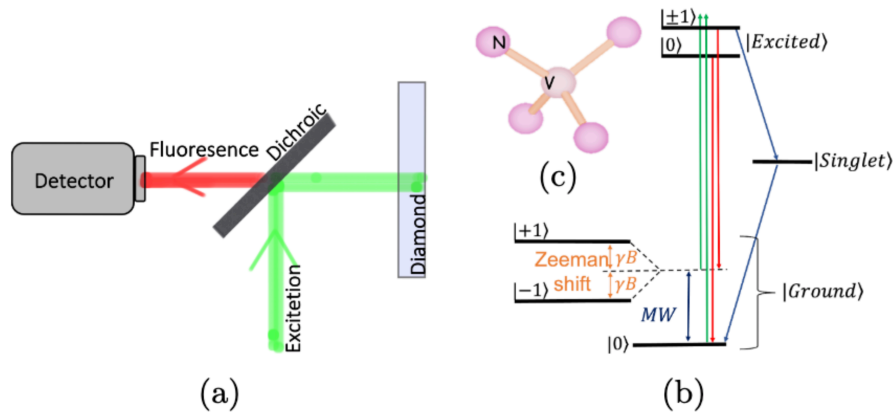
2.2 Experimental setup

Figure 4: (a) Experimental setup schematic; Green excitation laser illuminating a diamond sample, fluorescence from NV centers is collected and read by a detector. (c) NV can hold four possible orientations within the diamond. (b) NV energy level schematic.

The ground state of NV centers (Figure 4) is an effective two-level quantum system. Under green laser excitation, it is possible to initialize the spin to the ms = 0 ground state. In detail, the population occupying ms = 0 would reach the excited state manifold and decay back to the ground state ms = 0, emitting a red photon. The population occupying ms = ±1 is more likely to decay through the singlet state, to ms = 0 in a non-radiative way. Within the ground state, spin manipulation is possible with resonant Microwave (MW) pulses, population transfer to ms = ±1 would lead to a drop in measured fluorescence.

In the presence of an external magnetic field, degeneracy is lifted off the $m_s = \pm 1$ due to Zeeman shift, the shift is given by $\gamma B_\parallel$, where γ is the gyromagnetic ratio of the NV and $B_\parallel$ is the external magnetic field component parallel to the NV axis. In the diamond lattice, there are four possible crystallographic orientations the NV can take, and so, in the presence of a magnetic field that is not aligned with any of the orientations, there will be eight resonance frequencies, two for each orientation. Once these frequencies are known, the vectorial magnetic field can be calculated.

2.3   Results

Our basic approach relies on training a ML model to enable efficient identification of the positions of the resonances in the measured signal, from which the magnetic field information is extracted. We employ a relevant Multi Layer Perceptron (MLP) model, and compare it to regular raster scanning as a function of the number of scan points (subsampling). To evaluate the proposed approach in an experimental setting, we trained ML models with real experimental data augmented with synthetic samples. Ninety six full raster scans were measured in an epi-illumination wide field setup, each one under a different magnetic field (i.e. different resonance frequencies).

To generate synthetic data we simulated measurements based on a simplified NV Hamiltonian considering Zeeman shift, projecting varying magnetic fields of $\sim$ 100G on the four NV orientations, with added Gaussian noise, mimicking contrast and line shape (eight Lorentzians) as in the real (measured) data. We simulate a range of magnetic field strengths at various angles (θ and ϕ, respectively longitudinal and azimuthal, with respect to the diamond surface), and we use the resulting simulated data for which we produce full ESR spectrums with six or eight Lorentzians.

The ML models are trained on subsamplings of the full ESR spectrum. To do that, starting from the full 600-measures spectrum, we take every other point, every third point and so on, obtaining at the end spectrums with 300 measures, 200 measures and so on. The error for the ML prediction is defined as the averaged absolute value of the distances between the ML output to the resonance locations extracted from the full scans of the real data or from the ground truth used to generate the simulated data. To provide a fair comparison, a similar process was done for raster scans: the resonances extracted from the subsampled data are compared to the ones extracted from the full real data or with the ground truth used in the simulations.
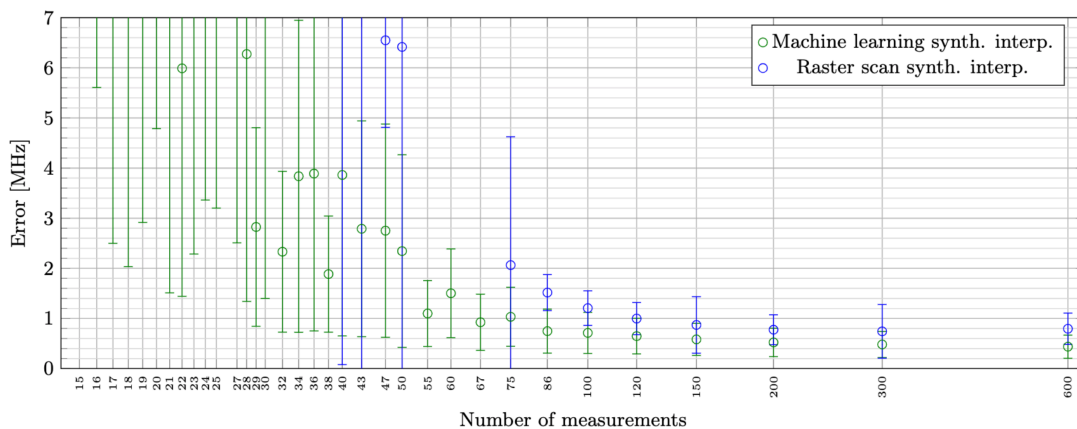

Figure 5: Neural network (green) and raster scan (blue) error as a function of the number of measurement points in the data. The error is calculated with the Mean Absolute Error (MAE) for the neural networks and normalized dividing by the square root of the success probability for the raster scan.
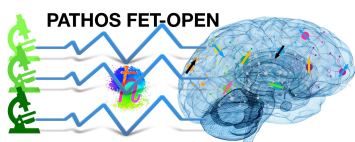
Figure 5 depicts as a function of the number of measurements the error and the normalized error of ML and raster scan respectively. The neural network (green) was trained on a full length dataset (600 frequency points) with 10000 samples and validated with 2000 samples to find the optimal hyperparameters. After the training, the network was introduced with other 2000 subsampled data to calculate the values of the plot. The same 2000 samples were used to test raster scanning as a function of number of data points. The subsampled data used for the neural network validation and for the test of either the network and the raster scanning were first linearly interpolated to 600 data points before its use. The network provides a better result compared to the raster scan, even for the full number of data points the ML averaged error is 20 KHz better. As the number of points decreases, the average error does increase and so does the standard deviation; however, the ML error scales better and performs impressively well with an **error of about 1 MHz with just 10 percent of the data points.**
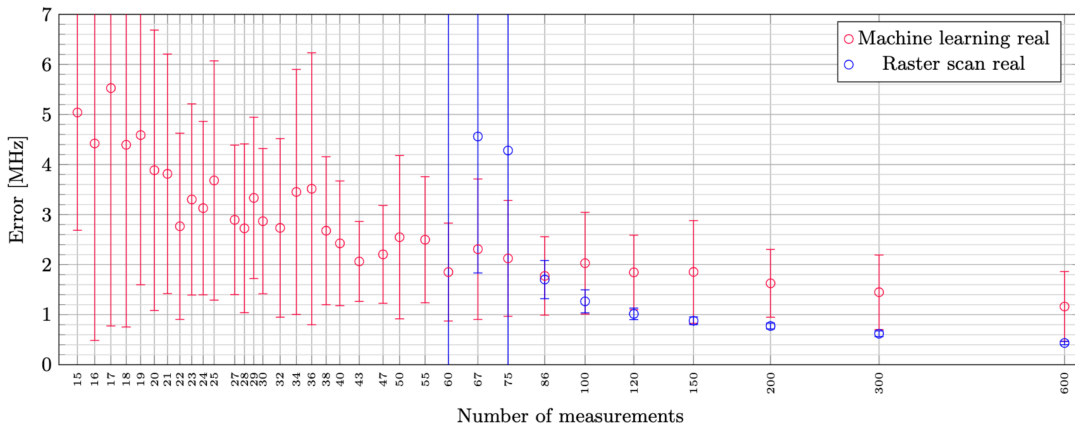

Figure 6: Comparison between MAE of neural networks and normalized error of raster scans for different number of data points. In red the dataset contained 1000 synthetic samples and 50 real data samples, validation was done with 46 real data samples. The same 46 raster were subsampled, the averaged error for those is depicted in blue.

We trained another neural network as a comparison to subsampling of the real data. The same 46 samples that were used for validation, were subsampled and the normalized mean error is depicted in blue in Figure 6. By contrast, in red in the same figure we report the evaluation of the network trained on the different subsampled datasets. These results show again the better scaling of the network error, although for a high number of points the network error is about 0.5 MHz worse than the raster. The raster scan of Figure 6 shows slightly better results respect to the raster results in Figure 5, this can be explained with the fact that all the real data in the former had eight distinguishable Lorentzians, and had high SNR, where in the simulated data in Figure 5 not all samples have eight Lorentzians, and the SNR is randomly sampled within a range.

2.4   Discussion and Conclusions

In this work we demonstrated that is possible to adopt ML models to retrieve a full ESR spectrum. We present the limitations of sub sampled raster scan and how the measurement error behaves as a function of SNR and linewidth. We show the flexibility and advantages of training the network in different ways: interpolating and using data with a range of noise and width. The ML models are more robust to subsampling the data respect to the raster scan. In this way, they can be effectively used as a method **to reduce the number of experimental measurements**. Moreover the ML model works even in the presence of overlapping Lorentzian peaks. We observed that training the neural network with more data reduces the error and the standard deviation.
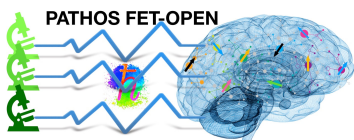
3   ODMR spectroscopy on complex/biological systems

The fundamental principle underlying Optically Detected Magnetic Resonance (ODMR) spectroscopy lies in the phenomenon of electron spin resonance (ESR), where the spin angular momentum of electrons interacts with an external magnetic field. This interaction leads to characteristic resonance frequencies, which ODMR spectroscopy exploits to extract detailed information about the spin dynamics and magnetic properties of the system under investigation. ODMR spectrum analysis has emerged as a powerful technique in the study of biological systems, offering unique insights into the behavior of biomolecules and their interactions with magnetic fields. Overall, ODMR spectrum analysis holds great promise for advancing our understanding of biological systems at the molecular and cellular levels. By probing the magnetic properties of biomolecules with precision and sensitivity, this technique opens up new avenues for research in areas such as biophysics, biochemistry, and biomedical engineering, with potential applications ranging from drug discovery to neurobiology. This work, being a collaboration between two PATHOS partners (INRIM and UNIFI), goes beyond the proof-of-principle results discussed in Sec. 1 and 2 of this report, towards real tests in biological environments.

A common limit of ODMR analysis is the necessity of an high number of measurement and this traduces in an increased experimental cost. In our work we are interested in characterizing ODMR spectra with the aid of ML tools and improve the current state of the art technique by reducing the experimental measurement time.
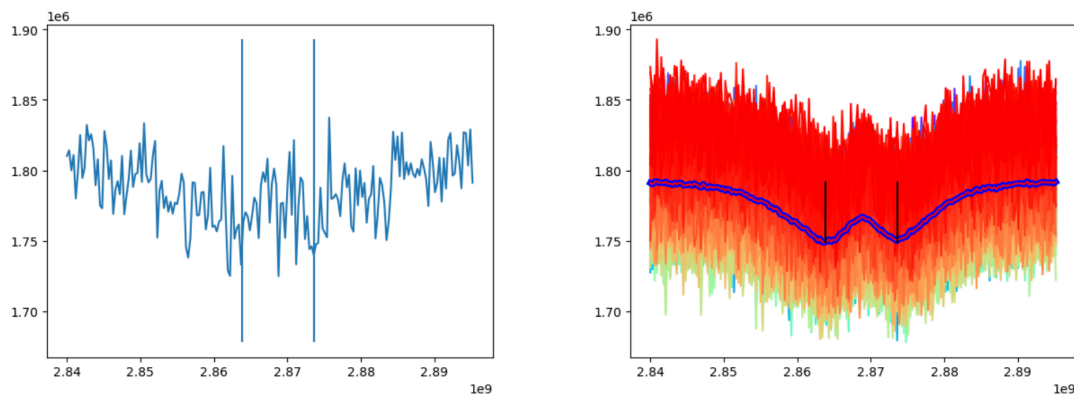

Figure 7: Real single sweep of the ODMR spectrum (on the left) and the average of different sweeps (on the right).

A Single ODMR spectrum has the form of the one in Figure 7. A set of around 600 single sweeps are averaged to reconstruct one ODMR spectrum. To calculate the external magnetic field, we are interested in the position of the two minimum. In order to train ML models for this task, it is necessary to have an high number of training samples. The available experimental data As the experimental measurements are expensive to obtain, thus we decided to integrate the real data with synthetic one in the form of mixture of two Gaussians plus a shift and with an added noise similar to the one observed in the real data.
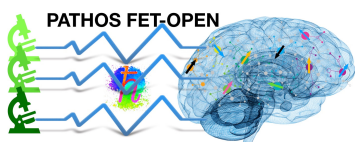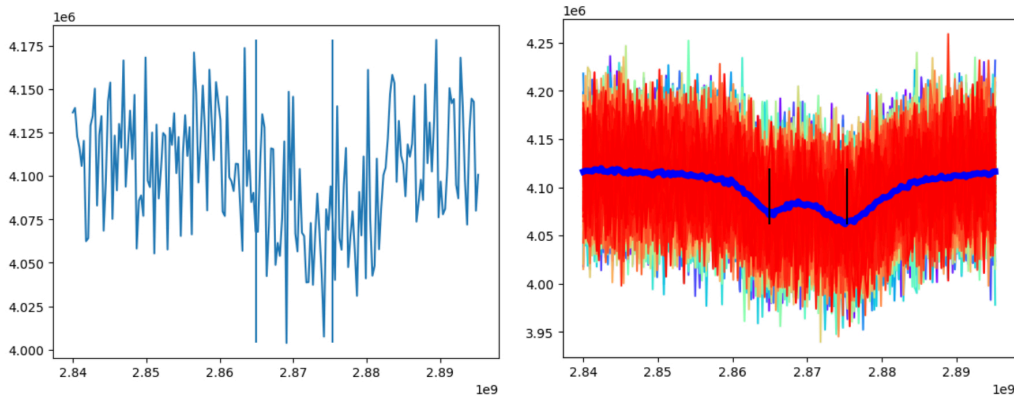
Figure 8: Example of a synthetic single sweep (on the left) and the average of different synthetic sweeps (on the right).

In Figure 8 we show one of the simulated spectra.

We trained different linear regression models on both simulated and real data to predict the two minimum in the spectrum. In order to evaluate our approach in the context of reducing the experimental cost, we scaled the capabilities of the model on less sweeps for the spectrum reconstruction.
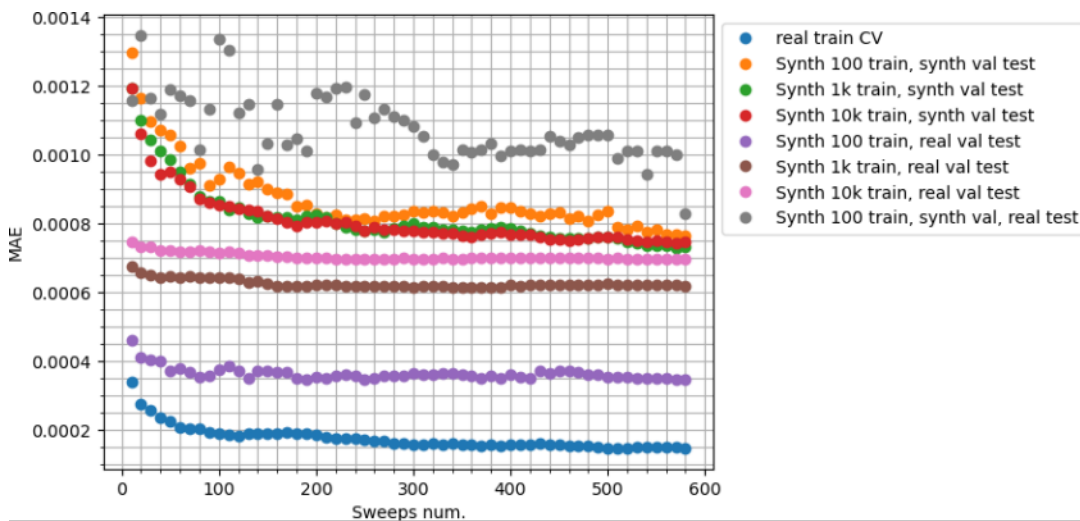

Figure 9: Results of the linear regression training for different kinds of dataset.

In Figure 9 we report the results, in the form of Mean Absolute Error of linear regression models predictions compared to the number of sweeps that were used to reconstruct a single ODMR spectrum. The models were trained on different datasets with both real and synthetic data in different proportions. Also the composition of the test set is different for the different curves. It can be completely real or synthetic. We can observe that the model with the less error is the one trained only on the real data. Regarding this, we should consider that the real data do not present an high variance in the peaks position. This means that the possible overfitting of the ML model is not visible with this data. Conversely, we can observe, both on real and synthetic data, that the ML models are robust to sweeps reductions. In other words, it is possible to use the ML model to predict the peak values using less measurements respect to the standard fitting procedure. This can be a useful property in an experimental context and can be associated with a reduction in measurement time and experiment cost.
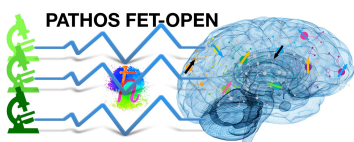
Moreover this is even more important in the frame of the NV-assisted biosensing, which is in the objectives of the PATHOS project, as the possibility to reduce as much as possible the laser exposure time emerges as a very relevant feature. In fact, in order to keep the data acquisition as biocompatible as possible it is desirable that the delicate biological samples are subjected to low laser intensity and limited to short exposure time. In this sense the use of ML algorithms represents an advantage not only in terms of experiment cost and resources but specifically it opens more interesting scenarios as it allows performing measurements of systems that would not be possible to probe more conventionally without altering the cells' functionality.

Collaboration has, indeed, started between INRIM and UNIFI units in order to implement this ML assisted sensing in biological environment. In particular INRIM adapted their single-photon sensitive confocal facility for the study of optical emitters in diamond (as the one represented in Figure 4) in order to render it compatible to ODMR measurements of biological samples (e.g. including incubator, inverted microscope, perfusion system, thermocouples, wide-field camera, Faraday cage, etc). The sensing activity is currently focused on mice hippocampal neurons and RAMOS cells and the goal is to perform biocompatible intracellular ODMR-based thermometry to investigate models of temperature dysregulation in cells in relation to cell disfunctions. A picture of the experimental ODMR set-up dedicated to biological sensing is reported in Figure 10.
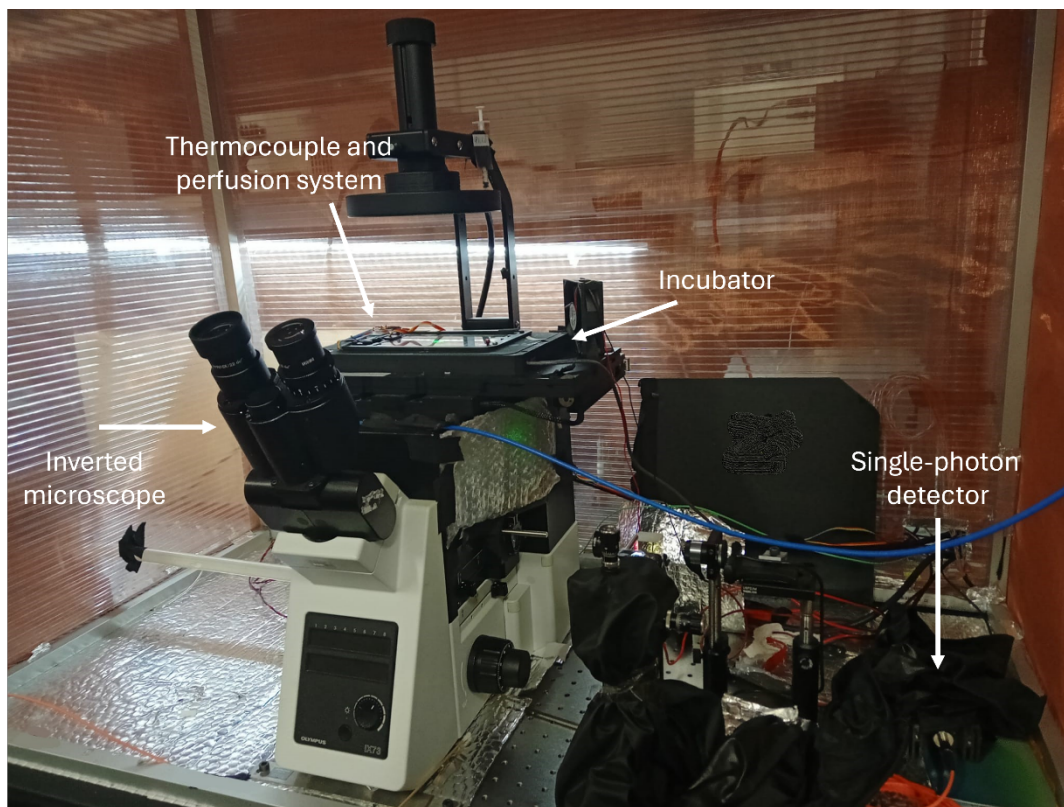

Figure 10: Picture of ODMR bio-sensing set-up built at INRiM laboratories.