Supplementary Information for A Neural Network Assisted ^{171}Yb^{+} Quantum Magnetometer

Yan Chen,[1,2,3] Yue Ban,[4,5] Ran He,[1,2,3] Jin-Ming Cui,[1,2,3] Yun-Feng Huang,[1,2,3] Chuan-Feng Li,[1,2,3] Guang-Can Guo,[1,2,3] and Jorge Casanova[4,5,6]

[1]*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China*

[2]*CAS Center For Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China*

[3]*Hefei National Laboratory, Hefei 230088, China*

[4]*Department of Physical Chemistry, University of the Basque Country UPV/EHU, Apartado 644, 48080 Bilbao, Spain*

[5]*EHU Quantum Center, University of the Basque Country UPV/EHU, Barrio Sarriena, s/n, 48940 Leioa, Biscay, Spain*

[6]*IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, 48009 Bilbao, Spain*

A. HAMILTONIAN

For a ^{171}Yb^{+} ion in an external static magnetic field B_z, external drivings $B_x^j \cos(\omega_j t + \phi_j)$ can be applied, leading to the Hamiltonian

$$H = A\,\mathbf{J}\cdot\mathbf{I} + \gamma_e B_z J_z - \gamma_n B_z I_z + \sum_j (\gamma_e B_x^j J_x - \gamma_n B_x^j I_x)\cos(\omega_j t + \phi_j), \tag{A.1}$$

where $A \approx (2\pi)\times 12.643$ GHz, the electronic/nuclear gyromagnetic ratios are $\gamma_e = (2\pi)\times 2.8024$ MHz/G and $\gamma_n \equiv \gamma_{^{171}\mathrm{Yb}^+} = (2\pi)\times 4.7248$ kHz/G. $\mathbf{J}$ and $\mathbf{I}$ are electron and nuclear spin-1/2 operators that can be written in a basis $\{|11\rangle, |10\rangle, |01\rangle, |00\rangle\}$ such that $J_z|0\,m\rangle = -\frac{1}{2}|0\,m\rangle$, $J_z|1\,m\rangle = \frac{1}{2}|0\,m\rangle$ and $I_z|m\,0\rangle = -\frac{1}{2}|m\,0\rangle$, $I_z|m\,1\rangle = \frac{1}{2}|m\,1\rangle$ for $m = 0, 1$. In the basis $\left\{|1\rangle, |\acute{0}\rangle, |-1\rangle, |0\rangle\right\}$ obtained after diagonalizing $A\,\mathbf{J}\cdot\mathbf{I} + \gamma_e B_z J_z - \gamma_n B_z I_z$ we have

$$H = \omega_1|1\rangle\langle 1| + \omega_{\acute{0}}|\acute{0}\rangle\langle\acute{0}| + \omega_{-1}|-1\rangle\langle -1| + \omega_0|0\rangle\langle 0| + \sum_j \Omega_j\Big[|1\rangle\langle\acute{0}| + |1\rangle\langle 0| + |\acute{0}\rangle\langle -1| + |0\rangle\langle -1| + \mathrm{H.c.}\Big]\cos(\omega_j t) \tag{A.2}$$

with $\omega_1 = \frac{A}{4} + (\gamma_e - \gamma_n)\frac{B_z}{2}$, $\omega_{-1} = \frac{A}{4} - (\gamma_e - \gamma_n)\frac{B_z}{2}$, $\omega_{\acute{0}} \approx \frac{A}{4} + \frac{(\gamma_e+\gamma_n)^2}{4A}B_z^2$, and $\omega_0 \approx -\frac{3A}{4} - \frac{(\gamma_e+\gamma_n)^2}{4A}B_z^2$.

Our scheme includes two MW driving fields ($j = 1, 2$) resonant with the $|0\rangle \leftrightarrow |1\rangle$ and $|0\rangle \leftrightarrow |-1\rangle$ hyperfine transitions with phases $\phi_1 = \phi_2 = 0$ and Rabi frequencies $\Omega_1 = \Omega_2$ (where $\Omega = \Omega_{1,2} = B_x^j\gamma_e/2\sqrt{2} = 2\pi\times 5.5$kHz) that cancel magnetic field fluctuations.

When an RF target field near to the $|\acute{0}\rangle \leftrightarrow |1\rangle$ transition is included, we get the the following Hamiltonian in the dressed state basis $|u\rangle, |d\rangle, |D\rangle, |0\rangle$

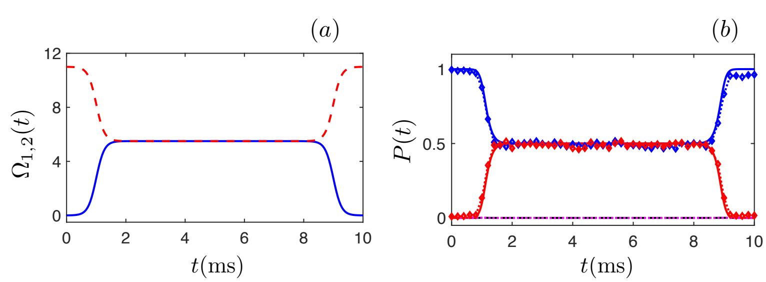

FIG. S1. (a) To generate the dark state $|D\rangle$, two MW drivings $\Omega_1(t)$ (solid blue) and $\Omega_2(t)$ (dashed red) are applied on the $|0\rangle \leftrightarrow |1\rangle$ and $|0\rangle \leftrightarrow |-1\rangle$ transitions. During the time interval that comprises $2-8$ ms we have $\Omega = \Omega_1 = \Omega_2 = 2\pi\times 5.5$kHz. (b) Numerical simulation of population transfer of the states $|1\rangle$ (solid blue), $|-1\rangle$ (solid red) and $|\acute{0}\rangle$ (dot-dashed black), $|0\rangle$ (dotted purple), where the dark state is maintained in the interval $[2, 8]$ ms. For comparison, experimentally obtained population of the states $|1\rangle$ (blue diamond) and $|-1\rangle$ (red diamonds) are shown.

$$
\begin{aligned}
H(t) = & \frac{\Omega}{\sqrt{2}}(|u\rangle\langle u| - |d\rangle\langle d|) + \left[\frac{\Omega}{4}(|u\rangle\langle \acute{0}| + |d\rangle\langle \acute{0}|) - \frac{\Omega_{\rm tg}}{2\sqrt{2}}|D\rangle\langle \acute{0}|\right] e^{-i\xi t} + \text{H.c.} \\
& - \left[\frac{\Omega}{2\sqrt{2}}(|u\rangle\langle u| - |d\rangle\langle d|) + \frac{\Omega}{4}(|u\rangle\langle D| + |D\rangle\langle d|) - \frac{\Omega}{4}(|D\rangle\langle u| + |d\rangle\langle D|)\right] e^{i\gamma_e B_z t} + \text{H.c.} \\
& + \frac{\Omega_{\rm tg}}{2}\left(\frac{1}{2}|u\rangle\langle \acute{0}| + \frac{1}{2}|d\rangle\langle \acute{0}| - \frac{1}{\sqrt{2}}|D\rangle\langle \acute{0}|\right) e^{2i(\frac{\gamma_e B_z}{2} - \frac{\gamma_e^2}{4A}B_z^2)t} e^{i\xi t} + \text{H.c.} \\
& + \frac{\Omega_{\rm tg}}{2}\left(\frac{1}{2}|\acute{0}\rangle\langle u| + \frac{1}{2}|\acute{0}\rangle\langle d| + \frac{1}{\sqrt{2}}|\acute{0}\rangle\langle D|\right) e^{i\gamma_e B_z t} e^{i\xi t} + \text{H.c.} \\
& + \frac{\Omega_{\rm tg}}{2}\left(\frac{1}{2}|\acute{0}\rangle\langle u| + \frac{1}{2}|\acute{0}\rangle\langle d| + \frac{1}{\sqrt{2}}|\acute{0}\rangle\langle D|\right) e^{i\frac{\gamma_e^2}{2A}B_z^2 t} e^{-\xi t} + \text{H.c.},
\end{aligned}
\tag{A.3}
$$

where $\xi = \omega_{\rm tg} - (\omega_1 - \omega_{\acute{0}})$ is a potential detuning between the target frequency and the sensor hyperfine transition.

B. PULSE DESIGN TO GENERATE THE STATE $|D\rangle$.

Stimulated Raman adiabatic passage (STIRAP) pulses applied on the $|0\rangle \leftrightarrow |1\rangle$ and $|0\rangle \leftrightarrow |-1\rangle$ transitions are used to create and detect the dark state $|D\rangle = (|-1\rangle - |1\rangle)/\sqrt{2}$. In particular, these are designed in the following shape

$$
\begin{aligned}
\Omega_1(t) &= \frac{A}{2}\left[\tanh\left(\frac{t-b_1}{c}\right)+1\right] + \frac{A}{2}\left[\tanh\left(\frac{t-b_2}{c}\right)+1\right], \\
\Omega_2(t) &= -\frac{A}{2}\left[\tanh\left(\frac{t-b_1}{c}\right)+1\right] - \frac{A}{2}\left[\tanh\left(\frac{t-b_2}{c}\right)+1\right] + 2A
\end{aligned}
\tag{B.1}
$$

We may slightly tune the parameters b_1, b_2, c to adjust the time duration of keeping $\Omega_1 = \Omega_2$. For instance, for a process lasting $t_f = 10$ms, with $A = 5.5 \times 2\pi$ kHz, $b_1 = 1$ms, $b_2 = 9$ms, and $c = 3\pi/A$. it takes about 2 ms to generate the state $|D\rangle$. This is maintained from $t = 2$ ms to $t = 8$ ms with $\Omega = \Omega_1 = \Omega_2 = 2\pi \times 5.5$kHz. In Fig. S1 (a) we show the controls that adiabatically evolve the initial state $|\Psi(0)\rangle = |1\rangle$. The state population evolution of theoretical and experimental data is shown in Fig. S1 (b).

C. TRAINING OF THE NEURAL NETWORK AND ESTIMATION RESULTS FOR $\Omega_{\rm tg}$ FROM A FINITE NUMBER OF MEASUREMENT

The data generated from numerical simulations is used to establish the NN. During the training stage, all the inputs $\mathbf{X}$ and the targets $\mathbf{A}$ are rescaled into the range $[0, 1]$ before being used in the NN (note the notation $\mathbf{A}^r$ and $\mathbf{X}^r$). Correspondingly, From the outputs $\mathbf{Y}^r$ (also in the range $[0, 1]$) obtained from the NN, we can get the results of the estimation $\mathbf{Y}$ with the real units. The training result for estimating $\Omega_{\rm tg}$ via standard gradient descent are shown in Fig. S2. The fit line of the outputs of the NN of the whole (including training/validation/test) dataset with respect to the targets is $y_1^r = \alpha a_1^r + \beta$, where $1 - \alpha \approx 10^{-5}$ and $\beta \approx 7.7 \cdot 10^{-6}$, as well as the correlation coefficient R = 0.99999, prove the regression with high fidelity.

D. ESTIMATION RESULTS FOR $\Omega_{\rm tg}$ AND ξ WITH A FINITE NUMBER OF MEASUREMENT

We can also train a NN to recognize two parameters, namely potential detunings ξ between the target frequency and the sensor hyperfine transition, and the Rabi frequency of the target field $\Omega_{\rm tg}$. In particular, we inspect two ranges for the parameters $\Omega_{\rm tg}$ and ξ: Firstly, we consider $\Omega_{\rm tg}/(2\pi) \in [0.5, 10]$ kHz and $\xi/(2\pi) \in [-0.6, 0.6]$ kHz by extracting 96 values of $\Omega_{\rm tg}$ with a step 0.1 kHz and 51 values of ξ separated by 0.024kHz. The input data $\mathbf{X} = \{P_1, P_2, ..., P_{N_p}\}$ in each example consists of $N_p = 201$ values collected in the range $t \in [0, t_f]$, where $t_f = 2.828$ ms. To enable the NN learning statistical fluctuation from shot noise, 100 repetitions are done for each numerical acquisition. Secondly, we keep the same range of detuning $\xi/(2\pi) \in [-0.6, 0.6]$ kHz, and set another range for $\Omega_{\rm tg}/(2\pi) \in [6.9, 10]$ kHz. In both cases, 70%, 15%, 15% of examples in each dataset mentioned above are randomly extracted to create the training/validation/test set in order to build the respective NN. The regression plots of the outputs from the above two NNs and the targets are illustrated in Fig. S3 (a-b).

We observe that the regression relation experiences less fluctuations in the regime in which $\Omega_{\rm tg}$ is far away from harmonic shape. In particular, at small $\Omega_{\rm tg}$ shot noise hinders the identification of the sensor response, as the slight tuning in ξ already mixes the responses. Meanwhile, the accuracy to estimate $\Omega_{\rm tg}$ is always high. This is proved in the numerical simulation in

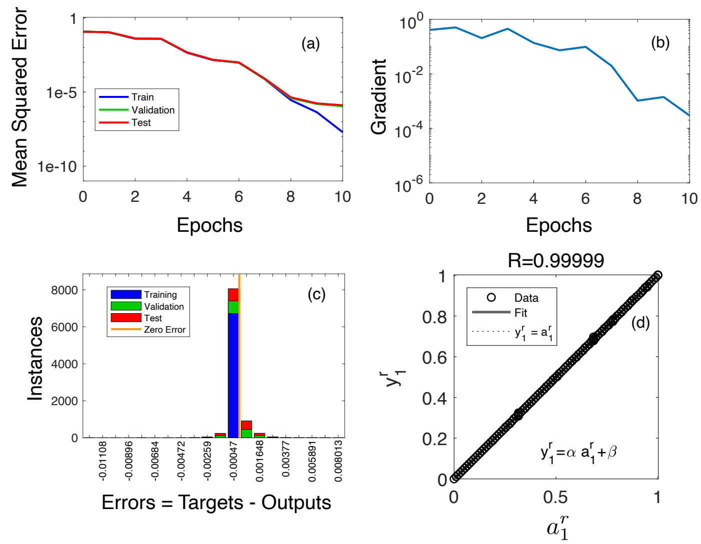

FIG. S2. Training results of the NN to estimate $\Omega_{\rm tg}$ where the input data $\mathbf{X} = \{P_1, P_2, ..., P_{N_p}\}$ is numerically derived from 96 different values for $\Omega_{\rm tg}/(2\pi) \in [0.5, 10]$ kHz with 100 repetitions. (a) Cost function values for training (blue) / validation (green) / test (red) datasets and (b) Gradient of the training set at each epoch. At the 10th epoch, we show the error histogram in panel (c), and (d) the comparison between the NN outputs y^1_r and the targets a^r_1. The fit in (d) shows the linear relation between the outputs y^r_1 and the target $a^r_1 = \Omega_{\rm tg}$, i.e. $y^r_1 = \alpha a^r_1 + \beta$ (solid), where $1 - \alpha \approx 10^{-5}$ and $\beta \approx 7.7 \cdot 10^{-6}$. The fit, the linear (ideal) case $y^r_1 = a^r_1$ (dotted) almost coincide. The correlation coefficient is R = 0.99999.

targets	$a_1/(2\pi) = 7.3920$ kHz	$a_2/(2\pi) = -0.6$ kHz
outputs	$y_1/(2\pi) = 7.5614$ kHz	$y_2/(2\pi) = -0.5865$ kHz
targets	$a_1/(2\pi) = 7.3920$ kHz	$a_2/(2\pi) = 0$ kHz
outputs	$y_1/(2\pi) = 7.4520$ kHz	$y_2/(2\pi) = 0.019$ kHz
targets	$a_1/(2\pi) = 7.3920$ kHz	$a_2/(2\pi) = 0.6$ kHz
outputs	$y_1/(2\pi) = 7.4575$ kHz	$y_2/(2\pi) = 0.3969$ kHz

TABLE I. Comparison of the outputs $\mathbf{Y} = \{y_1, y_2\}$ from the NN established for the range $\Omega_{\rm tg}/(2\pi) \in [6.9, 10]$ kHz and $\xi/(2\pi) \in [-0.6, 0.6]$ kHz with the targets $\mathbf{A} = \{a_1, a_2\}$, while the input data are experimentally obtained response where $a_1 = \Omega_{\rm tg}$ and $a_2 = \xi$.

Fig. S3 (c-f): when adopting the same ξ, the responses obtained numerically under 100 times of shots, can be identified in the presence of a small variation in $\Omega_{\rm tg}$ both at small values $\Omega_{\rm tg}$ (c) and large ones (d). Provided by the same $\Omega_{\rm tg}$ and a small variation in ξ, the responses show little difference at a small $\Omega_{\rm tg}$, whereas large difference at large $\Omega_{\rm tg}$. Here, we choose the response with large $\Omega_{\rm tg}$ from experiments as test examples, when fixing $\omega_{\rm tg} = 10.03$ MHz and $N_m = 50$. Inputing it into the NN established in the range $\Omega_{\rm tg}/(2\pi) \in [6.9, 10]$ kHz and $\xi/(2\pi) \in [-0.6, 0.6]$ kHz, we can get the results demonstrated in TABLE I.

It is possible to separate the responses with different ξ under shot noise by allowing for a longer evolution time and an increased number of shots. Consequently, the detection scheme via NNs can be improved. However, high accuracy comes at the cost of the increased time to generate the datasets.

E. PRECISION ANALYSIS.

The measurement precision for a parameter θ encoded in a state $|\Psi\rangle$ is upper bounded by the quantum Fisher information (QFI) [1] defined as $I_\theta = 4\left[\langle\partial_\theta\Psi|\partial_\theta\Psi\rangle - |\langle\Psi|\partial_\theta\Psi\rangle|^2\right]$. Consequently, the variance of the parameter estimator is lower bounded as

$$\Delta^2_\theta \geq \Delta^2\theta(t_0)^{\rm QFI} = \frac{1}{N_T I_\theta(t_f)}. \tag{E.1}$$

The term at the right hand side of Eq. (E.1) is accesible by computing the state at the final time t_f, where $N_T = N_p \times N_m$ with N_p the time instants at which the state is interrogated, and N_m the number shots at each time instant. For instance, taking $\Omega_{\rm tg} = (2\pi) \times 4.2265$ kHz from the scheme of continuous data acquisition, we get the lower bound for the variance $\Delta\Omega_{\rm tg}{}^{\rm QFI} =$

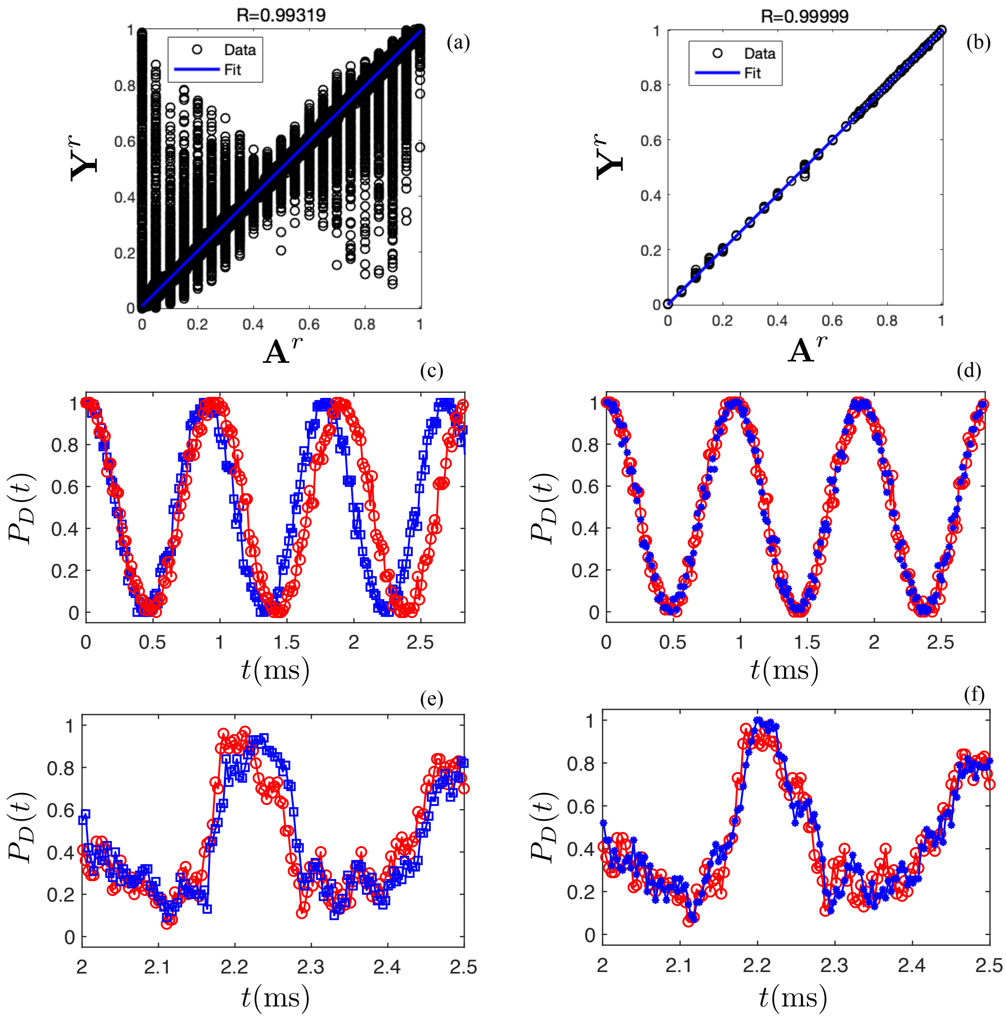

FIG. S3. (a-b) Comparison between the rescaled outputs of the NN $\mathbf{Y} = (y_1^r, y_2^r)$ and the targets $\mathbf{A} = (a_1^r, a_2^r) = (\Omega_{\rm tg}^r, \xi^r)$, with different range of $\Omega_{\rm tg}$, namely: (a) $\Omega_{\rm tg}^r \in [0,1]$, (b) $[0.67, 1]$, while keeping the same detuning range $\xi^r \in [0,1]$ (i.e. $\xi/(2\pi) \in [-0.6, 0.6]$ kHz). All the data is renormalized to $[0,1]$ (note the superscript r). In both plots, the fit $y_1^r = \alpha a_1^r + \beta$ and $y_2^r = \alpha a_2^r + \beta$ (solid blue) coincide with the ideal dependences $y_1^r = a_1^r$ and $y_2^r = a_2^r$ (dashed red) for both parameters where (a) $\alpha = 0.9858, \beta = 0.007377$, (b) $\alpha = 0.9997, \beta = 0.0002507$ and the correlation coefficient (a) R = 0.99319, (b) R = 0.99999. (c-d) Numerical simulation of the responses $P_D(t)$ under 100 shots. Comparisons of responses when (c) $\Omega_{\rm tg}/(2\pi) = 1.5$ kHz (red circle), $\Omega_{\rm tg}/(2\pi) = 1.6$ kHz (blue squared) at the same $\xi = 0$, (d) $\xi = 0$ (red circle), $\xi/(2\pi) = 0.05$ kHz (blue star) at the same $\Omega_{\rm tg}/(2\pi) = 1.5$ kHz, (e) $\Omega_{\rm tg}/(2\pi) = 7.4$ kHz (red circle), $\Omega_{\rm tg}/(2\pi) = 7.5$ kHz (blue squared) at the same $\xi = 0$, (f) $\xi = 0$ (red circle), $\xi/(2\pi) = 0.05$ kHz (blue star) at the same $\Omega_{\rm tg}/(2\pi) = 7.4$ kHz.

$(2\pi)\times 0.003294$ kHz. Our NN gives 20 results for y_1 based on 20 strings of experimental responses leading to $\bar{y}_1 = (2\pi)\times 4.2761$ kHz, with SD = $(2\pi)\times 0.0594$ kHz, see Table II in the main text. This proves that the estimation precision via the NN from the experimentally acquired responses is, as expected, below the limit imposed by QFI.

[1] Pang, S. & Jordan, A. N., Optimal adaptive control for quantum metrology with time-dependent Hamiltonians, *Nat. Commun.* **8**, 14695 (2016).