

Machine learning enhanced microwave electrometry in many-body Rydberg atoms

Corresponding Author: Professor Jin-Hui Wu

This manuscript has been transferred from another nature journal. This document only contains information relating to versions considered at Communications Physics.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Summary:

This manuscript presents a numerical study combining Rydberg-atom EIT spectroscopy with machine learning to infer extremely weak microwave fields in the presence of many-body interactions. The approach is interesting and the numerical results are intriguing. However, several key issues regarding the physical interpretation of sensitivity, experimental realism, and the specific advantages of machine learning over conventional nonlinear inference remain unclear. These points require substantial clarification before the work can be properly assessed for publication.

Major Comments:

1. Physical sensitivity versus algorithmic resolution

The reported sensitivity enhancement appears to correspond to an algorithmic resolution, i.e., how precisely the model can infer parameters from idealized spectra, rather than a noise-limited physical sensitivity that characterizes sensor performance in realistic environments. In an ideal, noiseless simulation, one could in principle achieve arbitrarily high precision by simply increasing the sampling density of the spectrum. Could the authors clarify whether the reported minimum detectable field $E_{\min}$ is fundamentally limited by the physics of the Rydberg system or instead by the numerical convergence of the neural network?

2. Interpretability and physical insight

To provide deeper physical insight, it would be valuable if the authors could perform an interpretability analysis (e.g., saliency maps or feature-importance ranking) to identify which regions of the EIT spectrum contribute most strongly to the model's inference. For example, does the network primarily rely on peak asymmetry, splitting width, or the amplitude of the emergent central feature? Establishing a connection between machine-learning weights and specific spectral components would significantly enhance the physical depth of the manuscript.

3. Spectral resolution versus experimental laser linewidth

The highest sensitivity is achieved using a very fine detuning step of 0.2 kHz. In practical experiments, however, laser linewidths are typically on the order of kHz or even MHz. Could the authors comment on how laser phase noise and frequency drifts would affect the observability of such fine spectral features, and consequently the achievable sensitivity enhancement under realistic experimental conditions?

4. Robustness against realistic experimental noise

Have realistic experimental noise sources, such as laser frequency drift, intensity fluctuations, or detection noise, been incorporated into the simulated training and testing datasets? Since the machine-learning models appear to be trained exclusively on idealized spectra, it would be helpful to clarify how robust the reported ~3700-fold sensitivity enhancement is expected to be when such noise sources are present.

5. Loss function and bias toward strong fields

The training dataset spans microwave Rabi frequencies from the sub-kHz regime to the MHz regime, while the primary goal is to detect extremely weak fields. Using a standard mean squared error loss (Eq. 12) may bias the model toward minimizing errors at large Rabi frequencies while underweighting errors in the weak-field regime of interest. Did the authors consider alternative loss functions, such as relative-error metrics (e.g., MAPE) or weighted losses, which might better optimize performance for small signals and potentially further improve the inferred sensitivity limit?

6. Comparison with nonlinear fitting and numerical inversion

To clarify whether the reported sensitivity enhancement is a unique advantage of machine learning or a more general consequence of nonlinear inference, it would be informative to compare the ML-based approach with standard nonlinear least-squares fitting or numerical inversion of the underlying mean-field equations applied to the same spectra. Demonstrating superior robustness or accuracy relative to conventional nonlinear optimization would substantially strengthen the claim that machine learning provides specific advantages in this many-body sensing context.

Recommendation:

Major Revision

Reviewer #2

(Remarks to the Author)

This manuscript presents a significant and timely contribution to the field of quantum sensing and microwave electrometry. The authors successfully address a fundamental challenge in Rydberg-atom-based sensing—spectral distortion due to many-body interactions—by leveraging machine learning techniques. The work is well-structured, methodologically sound, and offers impressive results, demonstrating sensitivity enhancements of up to nearly 3,800 times over traditional linear methods. The paper is suitable for publication in Communications Physics and represents a meaningful advance in both quantum metrology and applied machine learning for physical systems. I recommend acceptance after some revisions.

1. In the presented work, the mean-field super-atom model simplifies many-body interactions. At very high densities or strong interaction strengths, could correlated effects beyond mean-field become significant? If so, would the ML model's performance degrade, or could it learn these correlations implicitly?

2. The authors note that Rydberg interactions cause spectral distortion. However, under sufficiently strong interactions, the system may enter a bistable or even non-stationary dynamical regime, which is not addressed in the simulations. It would be valuable to discuss the expected performance and limitations of the proposed ML models under such conditions.

3. The analysis uses fixed experimental parameters (density, beam power). How robust are the ML models to realistic experimental fluctuations in these parameters (e.g., atom number variation, temperature, or thermal fluctuation)?

4. The authors should discuss experimental implementations and associated challenges in more detail. Cold-atom setups are often not ideal for integrated photonics, whereas room-temperature vapor cells, requiring no laser cooling, offer a more compact alternative. Additionally, addressing nonlinearities in Doppler-broadened systems would better connect to recent experiments.

5. Since atomic interactions introduce nonlinearity, the linear formula $\Omega_m = \Delta f/2$ becomes invalid, which explains the large errors shown by the red line in Fig. 2(b2). Therefore, using it as a benchmark is not appropriate. The performance of the ML model should instead be compared against a fit using the correct nonlinear formula.

6. The achieved sensitivity of $242 \text{ nV cm}^{-1} \text{ Hz}^{-1/2}$, while a clear improvement within the paper's own framework, appears modest compared to the $<5 \text{ nV cm}^{-1} \text{ Hz}^{-1/2}$ sensitivities reported with advanced techniques like heterodyne beating.

Could the authors comment on this discrepancy? Furthermore, is there a pathway for this ML-based approach to approach or surpass those experimental benchmarks, perhaps by integrating it with such high-sensitivity detection schemes?

Reviewer #3

(Remarks to the Author)

The manuscript proposes a data analysis technique based on machine learning to infer the microwave (MW) field amplitude from distorted EIT spectra of interacting Rydberg atoms. The authors focus on a regime of non-negligible Rydberg atom density, where interatomic interactions lead to nonlinear spectral distortions, the linear Autler-Townes relation between peak splitting and MW field no longer applies. Using numerically generated spectra based on a mean-field model, the authors trained neural networks to infer the MW Rabi frequency, and report large sensitivity improvements.

While the topic of Rydberg-based MW electrometry is of interest to the community, I do not find that the present work meets the conceptual and methodological standards required for publication in Communications Physics, and I therefore do not recommend publication in Communications Physics. The central issue is that the manuscript focuses on a problem that does not appear to be the limiting factor for weak-field MW detection. Specifically, it addresses a straightforward inverse problem with a machine-learning scheme, without demonstrating a clear advantage over other parameter-estimation techniques, or accounting for realistic experimental uncertainties. These two major comments are detailed below:

1) The inference problem considered in this work is low-dimensional and well-posed: the forward model is explicitly provided in the Supplementary Information [and derived by the authors in Phys. Rev. App. 22, 024034 (2024)], and for fixed parameters (Rabi frequency for control and probe transitions, atomic density, geometry, etc.), the MW Rabi frequency is the only free parameter. In this context, the MW field can be inferred by directly fitting EIT spectra using, for example, a least-squares method. The manuscript does not show that neural networks provide any advantage over such approaches. Instead, the reported enhancements are obtained by comparing the ML-based estimator to the linear Autler-Townes splitting formula, in a regime where this formula is known to fail due to interactions. This naive peak-difference estimator already fails in the non-interacting regime at small MW fields (Figure 2, a1, b1, c1) because it deviates from the actual energy splitting shown in Fig. S1.

2) The manuscript does not include any discussion of experimental noise sources or estimator uncertainty, and the quoted sensitivities (down to the few hundreds $\text{nV / cm} / \sqrt{\text{Hz}}$ level) therefore cannot be interpreted as meaningful metrological limits. In addition, numerical data are generated using a mean-field model for Rydberg interactions, typically valid for weak to moderate interactions, but no discussion is provided regarding the validity regime of this approximation. Interaction-induced

spectral broadening and atom losses, which have been documented experimentally (for instance, in Phys. Rev. A 93, 022709 (2016)), are neglected, and only energy shifts are included.

Addressing these points would require substantial rethinking of both the physical modeling and the benchmarking strategy, and go beyond a minor revision.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The revised manuscript has been substantially improved, and the authors have carefully addressed most of my previous concerns. In particular, the inclusion of realistic noise sources, the comparison with nonlinear least-squares fitting, and the newly added PCA-based interpretability analysis significantly strengthen both the technical completeness and the physical relevance of the work. The manuscript now presents a clearer and more convincing case for machine-learning-assisted microwave electrometry in the interaction-dominated Rydberg regime.

I nevertheless believe that several points would still benefit from further clarification before publication.

1. The revised manuscript demonstrates that the proposed ML schemes remain effective in the presence of several experimentally relevant noise sources. However, both the training and testing datasets are still generated within the same theoretical framework based on the mean-field superatom model and identical stochastic assumptions. Since realistic Rydberg experiments may involve additional model mismatch effects, such as imperfect interaction coefficients, spatial inhomogeneity, correlated technical noise, or beyond-mean-field dynamics, it would be valuable if the authors could discuss more explicitly the expected robustness and generalizability of the trained ML models under such conditions. Even a brief discussion of these limitations would help better contextualize the practical applicability of the method.

2. The newly introduced PCA-based interpretability analysis is a valuable addition and considerably improves the physical transparency of the ML framework. However, the present discussion remains somewhat qualitative. In particular, the manuscript would benefit from a more explicit connection between the dominant principal components and concrete interaction-induced spectral observables, such as peak asymmetry, linewidth broadening, or exchange-induced spectral deformation. Strengthening this connection would further support the central claim that many-body interactions provide additional physically meaningful information beyond the conventional AT splitting picture.

Overall, I find the revised manuscript technically interesting and significantly improved after revision. Subject to the above clarifications, I support publication in Communications Physics.

Reviewer #2

(Remarks to the Author)

The author has addressed all my concerns. Utilizing machine learning is a promising approach to enhancing microwave measurements. Thus, I recommend the publication of this manuscript in Communications Physics.

Reviewer #3

(Remarks to the Author)

I thank the authors for their detailed responses and for revising the manuscript. However, my main concern, also raised by Reviewer 1, remains unresolved and, in my view, prevents a meaningful assessment of the claimed advantages of the machine-learning (ML) approach.

The central issue is the lack of a fair and physically consistent comparison between the ML method and standard nonlinear inference techniques. In the current manuscript, the ML model is trained on spectra generated from the mean-field superatom model, thereby implicitly exploiting detailed knowledge of the underlying physics. In contrast, the “nonlinear least squares” (NLS) benchmark relies on fitting the spectra with Lorentzian functions to extract peak positions, followed by the use of a linear formula. Importantly, details of this procedure cannot be found in the updated manuscript, but only in the rebuttal. Moreover, this procedure is clearly not based on the same physical model, and the comparison is therefore not meaningful. It does not test the actual performance of nonlinear inference methods applied to the same problem, but rather compares a model-informed regression to a naive fitting procedure.

A proper benchmark would require performing parameter inference directly using the forward model employed to generate the data (e.g., the equations described in Supplementary Note 1), with the microwave Rabi frequency as a free parameter, and possibly additional parameters depending on how noise is introduced. This point is particularly important given the nature of the inference problem. As emphasized in my previous report, the task is intrinsically low-dimensional (in its simplest form, a single unknown parameter, and at most a few parameters when accounting for fluctuations). In such a setting, standard nonlinear fitting or numerical inversion of the known forward model should provide near-optimal estimates. Without a comparison to such physically grounded methods, the reported performance gains of the ML approach, including the claimed orders-of-magnitude improvement in sensitivity, cannot be considered established.

In addition, the manuscript remains unclear regarding how noise is incorporated into the datasets. From the rebuttal, noise sources appear to be introduced by varying parameters “in sequence”, which suggests that only a subset of parameters is varied at a time. I could not find clear details about this point in the updated manuscript. It is therefore unclear whether the ML models are trained and evaluated on datasets where all parameters fluctuate simultaneously, as would be expected under realistic experimental conditions. This point directly affects the effective dimensionality and difficulty of the inference problem and should be clarified.

Given these issues, I do not find that the revised manuscript provides a convincing or properly controlled demonstration that the ML approach outperforms standard nonlinear inference methods. As a result, the main claims of the paper are not sufficiently supported in its current form. For these reasons, I do not recommend publication of the manuscript in Communications Physics.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have substantially addressed the concerns raised in my previous review.

In particular, the revised manuscript and the detailed response letter now clarify that the nonlinear least-squares (NLS) benchmark is indeed implemented using the same forward physical model employed to generate the synthetic spectra, rather than relying solely on Lorentzian peak fitting as previously implied. The authors have also provided a much more complete description of the NLS inference procedure, the objective function, and the implementation details, making the comparison considerably clearer and more physically consistent.

The clarification regarding the noise model is also satisfactory. The authors explicitly state that all major noise sources are introduced simultaneously during dataset generation and describe how these fluctuations are incorporated into the simulated spectra. This addresses my previous concern regarding the effective dimensionality of the inference problem and the realism of the training and testing datasets.

Although one could argue that a more general multi-parameter nonlinear optimization framework might constitute an even stronger baseline, I consider that the present comparison is now sufficiently well justified for the scope of this work. The authors have clarified the assumptions underlying their benchmark and have provided a reasonable explanation for the observed performance difference between the NLS estimator and the machine-learning approach.

Overall, I believe that the authors have satisfactorily addressed my previous concerns, and I have no further major comments. I therefore support publication of the manuscript in its present form.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Reviewer 1

Many thanks for a central criticism “*The reported sensitivity enhancement appears to correspond to an algorithmic resolution, ..., rather than a noise-limited physical sensitivity that characterizes sensor performance in realistic environments*” and reminding us to go beyond “*an ideal, noiseless simulation*”. In the resubmitted manuscript, by including most realistic noise sources in calculating the AT-split EIT spectra of a probe field, we have confirmed that minimum detectable field E_{min} and hence the enhanced sensitivity is fundamentally limited by the physics of our Rydberg system, though relying on the numerical convergence of relevant machine-learning (ML) neural networks. Please find below our detailed answers to your specific comments point-by-point, which we hope to be convincing and have resulted in satisfactory revisions in the resubmitted manuscript.

Comment #1:

Physical sensitivity versus algorithmic resolution

The reported sensitivity enhancement appears to correspond to an algorithmic resolution, i.e., how precisely the model can infer parameters from idealized spectra, rather than a noise-limited physical sensitivity that characterizes sensor performance in realistic environments. In an ideal, noiseless simulation, one could in principle achieve arbitrarily high precision by simply increasing the sampling density of the spectrum. Could the authors clarify whether the reported minimum detectable field E_{min} is fundamentally limited by the physics of the Rydberg system or instead by the numerical convergence of the neural network?

Answer #1:

We thank the reviewer for this insightful comment. In the original manuscript, we just considered how precisely two ML schemes can infer microwave amplitudes from the idealized EIT spectra of probe transmission. No realistic noises are included in these spectra though they may be distorted to result in asymmetric and unequal AT-split peaks due to interatomic Rydberg interactions so that the usual linear formula fails to predict microwave amplitudes. In the resubmitted manuscript, we have included a few most important noise sources in real experiments (e.g., fluctuations of probe Rabi frequency Ω_p , coupling Rabi frequency Ω_c , and atomic density N_{at} , as well as Gaussian white noises intrinsic to probe detection T) when calculating more realistic AT-split EIT spectra. With updated calculations and analysis, we have clarified that minimum detectable field E_{min} is fundamentally limited by inevitable noises and Rydberg interactions of our neither too dense nor too dilute atomic sample, but relies also on the numerical convergence of specific neural networks used in different ML schemes. In other words, various physical noises set the lower limit of E_{min} , while an appropriate ML scheme allows us to get much closer to this limit than linear formula and also nonlinear least squares (NLS) fitting – a conventional method of nonlinear optimization.

Comment #2:

Interpretability and physical insight

To provide deeper physical insight, it would be valuable if the authors could perform an interpretability analysis (e.g., saliency maps or feature-importance ranking) to identify which regions of the EIT spectrum contribute most strongly to the model’s inference. For example, does the network primarily rely on peak asymmetry, splitting width, or the amplitude of the emergent central feature? Establishing a connection between machine-learning weights and specific spectral components would significantly enhance the physical depth of the manuscript.

Answer #2:

We thank the reviewer for this important suggestion. In the resubmitted manuscript, we have made an interpretability analysis to identify which regions of noisy AT-split EIT spectra contribute most strongly to the MW inference results of our two ML schemes. This is done by adding a procedure of principal component analysis (PCA) before the input layer of each neural network to transfer a n -dimensional discrete EIT spectrum $\mathbf{T}^{(i)}$ into a K -dimensional PC score vector $\mathbf{S}^{(i)}$ with $K \ll n$ for the i th MW Rabi frequency $\Omega_m^{(i)}$ after removing redundant noisy spectral data. It is found from the new Fig. 3 that Rydberg-distorted noisy EIT spectra exhibit a low-dimensional statistical structure because most explained variances are retained by a few leading PCs and the cumulative variance is already close to 100% before the 10th PC. A back-mapped contribution spectrum (by combining the importance of each PC with its loadings at different coupling detunings) shows that the predicted information is concentrated in several localized spectral intervals but not distributed uniformly over the full spectrum. These spectral intervals overlap with the characteristic dual-peak structure of a mean AT-split EIT spectrum, hence indicating a possibility of extracting physically relevant information from a few dominant PCs. This conclusion is substantiated by a correlation matrix in Fig. S2 of the Supplementary Information, which shows that the most relevant spectral features are positions and separation of two AT-split EIT peaks whereas their heights, widths, and asymmetry are also essential to be included. This PC-related connection between ML weights and specific spectral components has significantly enhanced the physical depth of our work.

Comment #3:

Spectral resolution versus experimental laser linewidth

The highest sensitivity is achieved using a very fine detuning step of 0.2 kHz. In practical experiments, however, laser linewidths are typically on the order of kHz or even MHz. Could the authors comment on how laser phase noise and frequency drifts would affect the observability of such fine spectral features, and consequently the achievable sensitivity enhancement under realistic experimental conditions?

Answer #3

We thank the reviewer for this helpful suggestion. We find from [Opt. Express 31, 19910, (2023)] that laser linewidths can be reduced to 250-350 kHz via the side-band locking technique of a high-finesse optical cavity and may be further down to 20-40 Hz by removing certain noises. Moreover, [Sci. Adv. 10, eado8130, (2024)] shows that the frequency drifts of two laser beams can be as low as 1.0 MHz/day (11 Hz/s) when both locked to a constant-temperature high-finesse optical cavity. Based on this fact, laser phase noises have been taken into account in the resubmitted manuscript by taking $\gamma_{p,c} = 100$ kHz for the probe and coupling fields, while frequency drifts are neglected because each round of probe detection is usually finished in about one second. Our new numerical calculations confirm that spectral features relying on a fine detuning step (especially those around resonance for weaker microwave fields) due to Rydberg interactions remain distinguishable when finite laser linewidths are included. Hence, the higher sensitivity predicted by minimum detectable field E_{min} in our ML schemes are achievable even under realistic experimental conditions.

Comment #4:

Robustness against realistic experimental noise

Have realistic experimental noise sources, such as laser frequency drift, intensity fluctuations, or detection noise, been incorporated into the simulated training and testing datasets? Since the

machine-learning models appear to be trained exclusively on idealized spectra, it would be helpful to clarify how robust the reported ~3700-fold sensitivity enhancement is expected to be when such noise sources are present.

Answer #4

We thank the reviewer for bring us to this important issue. In calculating the updated AT-split EIT spectra, we have included a few important realistic noise sources: *e.g.*, light intensity fluctuations, atomic density fluctuation, and detection white noise via the replacement of $\Omega_{p,c} \rightarrow \Omega_{p,c} + \delta\Omega_{p,c}$, $N_{at} \rightarrow N_{at} + \delta N_{at}$, and $T \rightarrow T + \delta T$, in sequence. More specially, we have set $\delta\Omega_{p,c}/\Omega_{p,c} = \eta \cdot N(0,1)$, $\delta N_{at}/N_{at} = \eta \cdot N(0,1)$, and $\delta T = \eta \cdot N(0,1)$ with $\eta = 2\%$ being a relative fluctuation amplitude while $N(0,1)$ a standard normal random number of mean value $\mu = 0$ and standard deviation $\sigma = 1$ independently sampled for each value of coupling detuning Δ_c . Laser frequency drifts are not considered because they can be made down to 10 Hz/s, when the probe and coupling fields are locked to a temperature-invariant high-finesse optical cavity [Sci. Adv. 10, eado8130, (2024)], hence negligible in each round of probe detection in about one second. Finite linewidths $\gamma_{p,c} = 100$ kHz have been considered, however, for the probe and coupling fields to account for their phase noises. Updated numerical results show that our ensemble ML scheme, upon applied to realistic but not idealized AT-split EIT spectra, still can achieve a 271-fold (354-fold) sensitivity enhancement as compared to linear formula (NLS fitting) in the presence of realistic noises. The essence of this evident advantage of our ensemble ML scheme lies in that it allows us to get much closer to the lower limit (quantifying microwave amplitude uncertainty) of minimum detectable field E_{min} by capturing more and much finer spectral details induced by Rydberg interactions.

PS: A Lorentzian function has been employed to fit the probe transmission spectra and determine the positions of two EIT peaks when the linear formula is used to predict microwave amplitudes.

Comment #5:

Loss function and bias toward strong fields

The training dataset spans microwave Rabi frequencies from the sub-kHz regime to the MHz regime, while the primary goal is to detect extremely weak fields. Using a standard mean squared error loss (Eq. 12) may bias the model toward minimizing errors at large Rabi frequencies while underweighting errors in the weak-field regime of interest. Did the authors consider alternative loss functions, such as relative-error metrics (e.g., MAPE) or weighted losses, which might better optimize performance for small signals and potentially further improve the inferred sensitivity limit?

Answer #5

We thank the reviewer for this helpful suggestion and agree that a loss function based on standard mean squared error (MSE) may bias our ML schemes. Hence, we have updated our calculations in Fig. 5 by considering a weighted-MSE loss function L_{wgh} and a log-MSE loss function L_{log} (for more details on their definitions, please refer to Eqs. (11) and (12) in Methods) in addition to the standard-MSE loss function L_{MSE} . Prediction results based on these loss functions are compared in Figs. 5(b), 5(c), and 5(d) for regression models A, B, and C corresponding to Ω_m in the range of $2\pi \times \{0.5, 5\}$ kHz, $2\pi \times \{5, 50\}$ kHz, and $2\pi \times \{50, 500\}$ kHz, respectively, in our ensemble ML scheme. It is found that loss function L_{log} performs better than loss function L_{wgh} , which in turn outperforms loss function L_{MSE} , consistent with the expectation that standard MSE tends to emphasize samples labeled by larger Ω_m in a broad range. More specifically, loss function L_{log} yields the smallest minimum detectable value $\Omega_m^{min} = 2\pi \times 0.79$ kHz for model A, while always

makes accurate predictions for models B and C. In contrast, $\Omega_m^{\min}$ turns out to be $2\pi \times 1.06$ kHz and $2\pi \times 1.52$ kHz, respectively, for class A when we use loss functions L_{wgh} and L_{log} , which may also make inaccurate predictions for classes B and C. To conclude, we have confirmed that it is viable to improve the inferred sensitivity limit with appropriately chosen loss functions.

Comment #6:

Comparison with nonlinear fitting and numerical inversion

To clarify whether the reported sensitivity enhancement is a unique advantage of machine learning or a more general consequence of nonlinear inference, it would be informative to compare the ML-based approach with standard nonlinear least-squares fitting or numerical inversion of the underlying mean-field equations applied to the same spectra. Demonstrating superior robustness or accuracy relative to conventional nonlinear optimization would substantially strengthen the claim that machine learning provides specific advantages in this many-body sensing context.

Answer #6

We thank the reviewer for this important suggestion. Per your request, we have compared our two ML schemes with standard NLS fitting, in addition to the usually used linear formula, when trying to infer microwave amplitudes from same AT-split EIT spectra distorted by Rydberg interactions. It is easy to observe from updated numerical results that our ML-based approach exhibits superior robustness or accuracy relative to a conventional NLS optimization. To be more specific, our basic and ensemble ML schemes achieve a sensitivity enhancement of 5.5 and 354 folds, respectively, compared to standard NLS fitting when a few important experimental noise sources (characterized by 2% fluctuations of relevant parameters) are included. This indicates that our two ML schemes, especially the ensemble one, allow us to better capture fine spectral features induced by moderate Rydberg interactions no matter in the absence or presence of realistic noises. We didn't consider the case of very strong Rydberg interactions that may result in non-equilibrium dynamics.

Response to Reviewer 2

Many thanks for deeming “*This manuscript presents a significant and timely contribution to the field of quantum sensing and microwave electrometry*” and “*The paper is suitable for publication in Communications Physics and represents a meaningful advance in both quantum metrology and applied machine learning for physical systems*”. More importantly, we appreciate the reviewer for proposing a few specific comments, which have prompted us to note deficiencies of our numerical results so that the original manuscript can be largely amended with new calculations and analysis. Sincerely hope that we have given convincing answers to your specific comments in the following and the resubmitted manuscript has been revised in a satisfactory way as highlighted in RED.

Comment #1:

In the presented work, the mean-field super-atom model simplifies many-body interactions. At very high densities or strong interaction strengths, could correlated effects beyond mean-field become significant? If so, would the ML model's performance degrade, or could it learn these correlations implicitly?

Answer #1:

We thank the reviewer for this important comment. In our work, we just consider atomic samples of moderate densities with weak but nonnegligible interaction strengths to examine their effects on

the Rydberg microwave (MW) electrometry. We agree that many-body correlated effects beyond our mean-field super-atom (SA) model (e.g., non-equilibrium bistabilities, time crystals, and phase transitions) could become significant in the case of very high atomic densities or strong interaction strengths. To examine this case, we should adopt a more refined mean-field method [Nat. Phys. 18, 1447-1452 (2022); Nat. Phys. 20, 1389-1394 (2024)] when calculating the AT-split EIT spectra of probe transmission, which is not of our present interest and may be considered in the future. But we believe that our ML schemes, relying highly on fine spectral features due to weak interatomic interactions, should work well even for strong interaction strengths because they can induce sharp spectral features around critical points of non-equilibrium phase transitions [Nat. Phys. 18, 1447-1452 (2022)]. Hence, the ML model's performance for predicting weak MW electric fields would not degrade (at least too much) by allowing us to learn strong Rydberg correlations implicitly.

Comment #2:

The authors note that Rydberg interactions cause spectral distortion. However, under sufficiently strong interactions, the system may enter a bistable or even non-stationary dynamical regime, which is not addressed in the simulations. It would be valuable to discuss the expected performance and limitations of the proposed ML models under such conditions.

Answer #2:

We thank the reviewer for proposing a more specific issue relevant to strong Rydberg interactions. Yes, as shown in [Nat. Phys. 18, 1447-1452 (2022); Nat. Phys. 20, 1389-1394 (2024)], an atomic system may enter a bistable or even non-stationary dynamical regime when Rydberg interactions become sufficiently strong. Our mean-field SA model used to calculate the AT-split EIT spectra is appropriate only for weak or moderate interactions, and a more refined mean-field method should be adopted in the case of very strong interactions, which goes beyond the goal of our present work and will not be examined here. One main difference between our SA model and the more refined method lies in that the former considers small Rydberg-induced potentials for a weak probe while the latter involves significant Rydberg-dependent nonlinear feedback for a strong probe. Anyway, we expect that our ML schemes should also exhibit rather good performance in predicting (weak) MW amplitudes even for quite strong Rydberg interactions, by recognizing MW-controlled sharp spectral features around critical points of non-equilibrium phase transitions [Nat. Phys. 18, 1447-1452 (2022)] in a way superior to other methods like nonlinear least squares (NLS) fitting. A main limitation may be that our two ML schemes depend highly on the width and resolution of a small spectral region where critical phase transitions occur, and the performance will degrade provided we cannot obtain enough spectral data in this region for training relevant neural networks.

Comment #3:

The analysis uses fixed experimental parameters (density, beam power). How robust are the ML models to realistic experimental fluctuations in these parameters (e.g., atom number variation, temperature, or thermal fluctuation)?

Answer #3

We thank the reviewer for bring us to this important issue. In the resubmitted manuscript, we have included realistic experimental fluctuations of probe Rabi frequency Ω_p , coupling Rabi frequency Ω_c , and atomic density N_{at} when employing the SA model to calculate the AT-split EIT spectra of probe transmission T , which is also assumed to exhibit a white detection noise. These noise sources are described by $\delta\Omega_{p,c}/\Omega_{p,c} = \eta \cdot N(0,1)$, $\delta N_{at}/N_{at} = \eta \cdot N(0,1)$, and $\delta T = \eta \cdot N(0,1)$

with $\eta = 2\%$ denoting a relative fluctuation amplitude while $N(0,1)$ a standard normal random number of mean value $\mu = 0$ and standard deviation $\sigma = 1$ independently sampled for each value of coupling detuning Δ_c . Updated calculations show that our ensemble ML scheme, upon applied to realistic rather than idealized AT-split EIT spectra, still achieves a 354-fold (271-fold) sensitivity enhancement as compared to NLS fitting (usually used linear formula) in the presence of these noise sources, confirming its robustness against random fluctuations of a few important parameters. The underlying physics is that these fluctuations won't eliminate fine spectral features induced by Rydberg interactions, which can be better recognized by our ensemble ML scheme (as compared to both linear formula and NLS fitting) for sensing weaker MW electric fields.

PS: A Lorentzian function has been employed to fit the probe transmission spectra and determine the positions of two EIT peaks when the linear formula is used to predict MW amplitudes.

Comment #4:

The authors should discuss experimental implementations and associated challenges in more detail. Cold-atom setups are often not ideal for integrated photonics, whereas room-temperature vapor cells, requiring no laser cooling, offer a more compact alternative. Additionally, addressing nonlinearities in Doppler-broadened systems would better connect to recent experiments.

Answer #4

We thank the reviewer for this important suggestion and agree that room-temperature vapor cells are more ideal for experimental implementations. However, we are more inclined to consider cold atoms in our work because fine spectral features arising from Rydberg interactions will be largely smoothed out by inhomogeneous Doppler broadening. Our basic and ensemble ML schemes, upon applied to warm atomic vapors, will become not so effective because it relies highly on how fine spectral features are captured by training neural networks. As discussed in [Sci. Adv. 10, eads0683 (2024)], the sensitivity of a warm Rydberg electrometer is primarily degraded by atomic thermal motions and a solution based on cold atoms helps to approach closely the standard quantum limit. Moreover, please notice that it is viable to realize on-chip cold atoms [Chin. Opt. Lett. 22, 060201 (2024); Phys. Rev. Appl. 23, L011001 (2025)] just like on-chip warm atoms [Appl. Phys. Rev. 5, 031302 (2018)], hence cold-atom setups are more favorable for high-sensitivity MW electrometry and can also be used for integrated photonics though basically needing more involved techniques. This point has been clarified in the resubmitted manuscript by stressing that our ML schemes of good performance depend critically on fine spectral features induced by Rydberg interactions and require a cold-atom platform to avoid adverse effects of inhomogeneous Doppler broadening.

Comment #5:

Since atomic interactions introduce nonlinearity, the linear formula $\Omega_m = \Delta f/2$ becomes invalid, which explains the large errors shown by the red line in Fig. 2(b2). Therefore, using it as a benchmark is not appropriate. The performance of the ML model should instead be compared against a fit using the correct nonlinear formula.

Answer #5

We thank the reviewer for this insightful comment and agree that the linear formula $\Omega_m = \Delta f/2$ is not valid again for determining Ω_m when Rydberg interactions introduce obvious nonlinearity. In the resubmitted manuscript, per your request, we have compared our basic and ensemble ML schemes against NLS fitting when trying to infer Ω_m from recalculated AT-split EIT spectra with most experimental noise sources included. Based on this more appropriate comparison, we show

that our ensemble (basic) ML scheme still exhibits a superior performance described by a 354-fold (5.5-fold) sensitivity improvement due to a 354-fold (5.5-fold) reduced minimum detectable field E_{min} relative to NLS fitting as an appropriate benchmark in the nonlinear response regime.

Comment #6:

The achieved sensitivity of $242 \text{ nV cm}^{-1} \text{ Hz}^{-1/2}$, while a clear improvement within the paper's own framework, appears modest compared to the $<5 \text{ nV cm}^{-1} \text{ Hz}^{-1/2}$ sensitivities reported with advanced techniques like heterodyne beating. Could the authors comment on this discrepancy? Furthermore, is there a pathway for this ML-based approach to approach or surpass those experimental benchmarks, perhaps by integrating it with such high-sensitivity detection schemes?

Answer #6

We thank the reviewer for this valuable comment. The sensitivity $242 \text{ nV cm}^{-1} \text{ Hz}^{-1/2}$ estimated via $E_{min} = \mathcal{S}/\sqrt{t}$ in the original manuscript relies on both minimum detectable field E_{min} and detection time t , which has been taken to be 1.0 s just like that used in a few experiments. This value has increased to $1.39 \text{ } \mu\text{V cm}^{-1} \text{ Hz}^{-1/2}$ in the resubmitted manuscript because realistic noise sources (e.g. 2% fluctuations of light intensities, atomic density, and transmission detection) have been considered to calculate the AT-split EIT spectra of probe transmission T . As shown in the resubmitted manuscript, these noises result in a clear uncertainty δE_m between predicted and real values of MW electric fields, which sets a lower limit ($1.28 \text{ } \mu\text{V cm}^{-1} \text{ Hz}^{-1/2}$) for sensitivity $\mathcal{S}$ if we equal δE_m and E_{min} [Sci. Adv. 10, eads0683 (2024)]. Our ensemble ML scheme allows us to achieve a sensitivity ($1.39 \text{ } \mu\text{V cm}^{-1} \text{ Hz}^{-1/2}$) approaching this limit very closely, but a large gap exists between this limit and the sensitivity ($493 \text{ } \mu\text{V cm}^{-1} \text{ Hz}^{-1/2}$) based on NLS fitting.

It is obvious that the sensitivity ($1.39 \text{ } \mu\text{V cm}^{-1} \text{ Hz}^{-1/2}$) based on our ensemble ML scheme is still larger than the sensitivity ($55 \text{ nV cm}^{-1} \text{ Hz}^{-1/2}$) reported by advanced techniques like heterodyne beating [Nat. Phys. 18, 1447-1452 (2022)]. One main reason for this large discrepancy lies in that a relatively larger MW field uncertainty δE_m has been included in our ensemble ML scheme due to 2% fluctuations of atomic density, probe Rabi frequency, coupling Rabi frequency, and probe transmissivity. Another reason may be that the spectra of heterodyne beating is more sensitive to the amplitude variations of MW signal fields than the spectra of direct transmission because the influences of MW signal fields has been largely amplified by a stronger local MW field. Our ML schemes can be integrated, in principle, with alternative detection schemes like heterodyne beating as long as they generate fine spectral features beneficial for training ML neural networks. Such a promising integration must yield higher sensitivities surpassing existent experimental benchmarks, especially when relevant parameters are better stabilized to exhibit less fluctuations.

Response to Reviewer 3

Many thanks for deeming “the topic of Rydberg-based MW electrometry (in our manuscript) is of interest to the community”, while regret for remarking “do not find that the present work meets the conceptual and methodological standards required for publication in Communications Physics”. In particular, we cannot agree “the manuscript focuses on a problem that does not appear to be the limiting factor for weak-field MW detection” because it is evident from Fig. 2 that the usually used linear formula fails to work when not too dilute atomic samples are considered for improving the detection precision by increasing the number of Rydberg atoms participating MW sensing. In this regard, it is meaningful to examine whether our ML schemes help to determine the amplitudes of

weaker MW fields when Rydberg interactions become nonnegligible. This point has been further stressed in the resubmitted manuscript. Anyway, we appreciate the pertinent criticisms and have tried to improve our work by *first* recalculating probe EIT spectra with a few most important noise sources (e.g., fluctuations of light intensity and atomic density) included and *then* comparing our ML schemes to nonlinear least squares (NLS) fitting in terms of minimum detectable field $E_{\min}$ and its related sensitivity. The *methodological* advancement is reflected by the superiority of our ML schemes over NLS fitting while the *conceptual* advancement may be attributed to a proposal of detecting weaker microwave fields with higher precision in the case of nonnegligible Rydberg interactions. Please refer to our answers given below to specific comments for more details, which we hope to be convincing along with satisfactory RED revisions in the resubmitted manuscript.

Comment #1:

The inference problem considered in this work is low-dimensional and well-posed: the forward model is explicitly provided in the Supplementary Information [and derived by the authors in Phys. Rev. App. 22, 024034 (2024)], and for fixed parameters (Rabi frequency for control and probe transitions, atomic density, geometry, etc.), the MW Rabi frequency is the only free parameter. In this context, the MW field can be inferred by directly fitting EIT spectra using, for example, a least-squares method. The manuscript does not show that neural networks provide any advantage over such approaches. Instead, the reported enhancements are obtained by comparing the ML-based estimator to the linear Autler–Townes splitting formula, in a regime where this formula is known to fail due to interactions. This naive peak-difference estimator already fails in the non-interacting regime at small MW fields (Figure 2, a1, b1, c1) because it deviates from the actual energy splitting shown in Fig. S1.

Answer #1:

We thank the reviewer for this insightful comment, and agree that NLS fitting can be used to infer MW amplitudes from distorted EIT spectra when the linear AT splitting formula fails to work. We have improved our theoretical model in the resubmitted manuscript by taking into account finite laser linewidths, Rydberg induced dephasings, fluctuations of light intensities and atomic density, and detection white noise. In this case, MW Rabi frequency is not the only free parameter again in the presence of various experimental noise sources, but NLS fitting also does not work well for determining amplitudes of weaker MW fields due to Rydberg interactions. The advantages of our ML schemes over linear formula and NLS fitting have been confirmed via updated calculations, in which the noise related to parameter $\mathcal{A}$ can be described by $\delta\mathcal{A}/\mathcal{A} = \eta \cdot N(0,1)$ with $\eta = 2\%$ being a relative fluctuation amplitude while $N(0,1)$ a standard normal random number of mean value $\mu = 0$ and standard deviation $\sigma = 1$. Updated numerical results show that in this case our ensemble ML scheme can still achieve a 354-fold (271-fold) sensitivity improvement in terms of minimum detectable field $E_{\min}$ as compared to NLS fitting (linear formula). In this regard, the low-dimensional and well-posed inference problem in our original manuscript has been amended by appropriate comparisons of different methods using noisy AT-split EIT spectra to infer MW amplitudes. Note that NLS fitting performs better than linear formula as expected in the absence of Rydberg interactions especially when two EIT peaks tend to be indistinguishable, but becomes slightly inferior to linear formula unexpectedly in the presence of Rydberg interactions due, e.g. to severely distorted sharp spectral features that hamper the conventional optimization of NLS.

PS: A Lorentzian function has been employed to fit the probe transmission spectra and determine the positions of two EIT peaks when the linear formula is used to predict MW amplitudes.

Comment #2:

The manuscript does not include any discussion of experimental noise sources or estimator uncertainty, and the quoted sensitivities (down to the few hundreds nV/cm/sqrt(Hz) level) therefore cannot be interpreted as meaningful metrological limits. In addition, numerical data are generated using a mean-field model for Rydberg interactions, typically valid for weak to moderate interactions, but no discussion is provided regarding the validity regime of this approximation. Interaction-induced spectral broadening and atom losses, which have been documented experimentally (for instance, in Phys. Rev. A 93, 022709 (2016)), are neglected, and only energy shifts are included.

Answer #2:

We thank the reviewer for this insightful comment. In the resubmitted manuscript, a few important experimental noise sources regarding pump Rabi frequency Ω_p , coupling Rabi frequency Ω_c , and atomic density N_{at} have been included when we calculate the EIT spectra of probe transmission T , which is also assumed to exhibit a Gaussian white noise. That means, we have replaced these constant parameters in the updated calculations by $\Omega_p + \delta\Omega_p$, $\Omega_c + \delta\Omega_c$, $N_{at} + \delta N_{at}$, and $T + \delta T$ with $\delta\Omega_p$, $\delta\Omega_c$, δN_{at} , and δT describing random fluctuations existent in real experiments. New numerical results confirm that our ensemble ML scheme, upon applied to such realistic EIT spectra, still achieves a 354-fold (271-fold) sensitivity enhancement as compared to NLS fitting (linear formula) when all fluctuation amplitudes of $\delta\Omega_p/\Omega_p$, $\delta\Omega_c/\Omega_c$, $\delta N_{at}/N_{at}$, and δT are set as 2%. More specially, the sensitivity achieved with our ensemble ML scheme in the presence of moderate experimental noises is still down to $1.39 \mu\text{Vcm}^{-1}\text{Hz}^{-1/2}$ for a minimum detectable field $E_{min} = 1.39 \mu\text{Vcm}^{-1}$ and a typical experimental detection time $t = 1.0$ s. This sensitivity is very close to the lower limit $1.28 \mu\text{Vcm}^{-1}\text{Hz}^{-1/2}$ set by an uncertainty between predicted and real values of MW electric fields resulted from random fluctuations of Ω_p , Ω_c , N_{at} , and T .

Meanwhile, we agree that our mean-field super-atom (SA) model describing Rydberg interactions is valid only for weak to moderate interactions, while a more refined mean-field model [Nat. Phys. 18, 1447-1452 (2022)] should be employed to calculate probe transmission spectra when Rydberg interactions become sufficiently strong. One main difference lies in that Rydberg interactions are described by small energy shifts and dephasings for a weaker probe in our SA model, but result in significant nonlinear feedback for a stronger probe in the refined mean-field model. Consequently, AT-split EIT spectra are just distorted more or less in our work, while non-equilibrium bistability and phase transition can be observed as in [Nat. Phys. 18, 1447-1452 (2022)]. This point has been briefly discussed in the resubmitted manuscript to clarify the validity regime of our SA model.

We have also read carefully the important work [Phys. Rev. A 93, 022709 (2016)] where spectral broadening and atom losses due to Rydberg interactions are found to be essential. But this work considered cold alkaline-earth-metal Sr gases of very high densities $10^{12} \sim 10^{13} \text{ cm}^{-3}$, a situation clearly different from our alkali-metal ^{87}Rb gases of densities as low as 10^{10} cm^{-3} . As shown in [Phys. Rev. Lett. 107, 213601 (2011)], for our dilute atomic gases, we don't need to consider atom losses due to essential Rydberg excitations but should include small dephasings arising from weak Rydberg interactions between adjacent SAs. In the original manuscript, interaction-induced energy shifts $s_{33,44}$ and coupling rate e_{43} have been calculated in terms of SA density N_{SA} , populations $\Sigma_{33,44}$, and coherence Σ_{43} . In the resubmitted manuscript, we have further calculated interaction-induced dephasings $\sqrt{\vartheta_{33}}$, $\sqrt{\vartheta_{44}}$, and $\sqrt{\vartheta_{43}}$ as the square roots of the variances of s_{33} , s_{44} , and

e_{43} , respectively, by referring to [Sci. Adv. 10, eads0683 (2024)]. They are typically of the order of 10 kHz for the parameters used in our work and relevant variances can be calculated as

$$\vartheta_{33} = 2\pi \int_0^\pi \sin\theta d\theta \int_R^{r_\theta} dr \frac{C_{33}^2}{r^{10}} N_{SA} \Sigma_{33} - s_{33}^2 \approx \frac{4\pi}{9} C_{33}^2 N_{SA} \Sigma_{33} N_{at}^3 - \frac{16\pi^2}{9} C_{33}^2 N_{SA}^2 \Sigma_{33}^2 N_{at}^2,$$

$$\vartheta_{44} = 2\pi \int_0^\pi \sin\theta d\theta \int_R^{r_\theta} dr \frac{C_{44}^2}{r^{10}} N_{SA} \Sigma_{44} - s_{44}^2 \approx \frac{4\pi}{9} C_{44}^2 N_{SA} \Sigma_{44} N_{at}^3 - \frac{16\pi^2}{9} C_{44}^2 N_{SA}^2 \Sigma_{44}^2 N_{at}^2,$$

$$\vartheta_{43} = 2\pi \int_0^\pi \sin\theta d\theta \int_R^{r_\theta} dr \frac{C_{43}^2}{r^4} N_{SA} \Sigma_{43} - e_{43}^2 \approx \frac{4\pi}{3} C_{43}^2 N_{SA} \Sigma_{43} N_{at} - 16\pi^2 f^2(R, r_p) C_{43}^2 N_{SA}^2 \Sigma_{43}^2,$$

with the derivations of s_{33} , s_{44} , and e_{43} already given in the original manuscript.

Response to Reviewer 1

Many thanks for remarking that we have given satisfactory responses to most of your concerns in the previous round of review and supporting publication of our current work in Communications Physics after clarifying two remaining concerns. Please find below our answers in detail to these concerns, regarding a few additional model mismatch effects and a more quantitative PCA-based interpretability analysis, respectively. Hope that these supplementary answers are satisfactory and we have made appropriate revisions in the newly submitted manuscript as you expect.

Comment #1:

The revised manuscript demonstrates that the proposed ML schemes remain effective in the presence of several experimentally relevant noise sources. However, both the training and testing datasets are still generated within the same theoretical framework based on the mean-field superatom model and identical stochastic assumptions. Since realistic Rydberg experiments may involve additional model mismatch effects, such as imperfect interaction coefficients, spatial inhomogeneity, correlated technical noise, or beyond-mean-field dynamics, it would be valuable if the authors could discuss more explicitly the expected robustness and generalizability of the trained ML models under such conditions. Even a brief discussion of these limitations would help better contextualize the practical applicability of the method.

Answer #1:

We thank the reviewer for this insightful comment and agree that “realistic Rydberg experiments may involve additional model mismatch effects, such as imperfect interaction coefficients, spatial inhomogeneity, correlated technical noise, or beyond-mean-field dynamics”. **First of all**, imperfect interaction coefficients will cause slightly different spectral features that, however, won’t degrade the performance of our ML approach, relying on the existence but not specific forms of interaction-induced spectral features. **Second**, spatial inhomogeneity (in atomic density) will lead to stronger interactions in the middle but weaker interactions at both sides of the Rydberg sample, which will also slightly modify interaction-induced spectral features yet without influencing the performance of our ML approach too much since mean Rydberg interactions remain roughly unchanged. **Third**, correlated technique noise in real experiments, when superimposed on others, will degrade (more or less) the performance of our ML approach due to larger uncertainties of detected transmission signals, which is however inevitable for other methods as well. **Last**, our mean-field treatment on Rydberg interactions, though not accurate enough, has been successfully used to recover relevant experimental results by simplifying the involved many-body problem. Going beyond mean-field dynamics, it is viable to observe more subtle spectral features helping to improve the performance of our ML approach. Hence, our ML approach should be robust against the four additional model mismatch effects and can be generalized, in principle, to such more realistic conditions. All above discussions have been briefly stressed by adding a third paragraph in Discussion.

Comment #2:

The newly introduced PCA-based interpretability analysis is a valuable addition and considerably improves the physical transparency of the ML framework. However, the present discussion remains somewhat qualitative. In particular, the manuscript would benefit from a more explicit connection between the dominant principal components and concrete interaction-induced spectral observables, such as peak asymmetry, linewidth broadening, or exchange-induced spectral deformation. Strengthening this connection would further support the central claim that

many-body interactions provide additional physically meaningful information beyond the conventional AT splitting picture.

Answer #2:

We thank the reviewer for approving a newly introduced PCA-based interpretability analysis and suggesting an explicit discussion on the connection between dominant principal components (PCs) and concrete interaction-induced spectral observables. In the absence of Rydberg interactions, two AT-split EIT peaks should be symmetrically positioned on both sides of coupling resonance ($\Delta_c = 0$) with equal heights for a resonant MW field as considered. When Rydberg interactions become essential, positions and heights of the two EIT peaks will not be symmetric and equal, respectively, again due to Δ_c -dependent vdW induced energy shifts ($s_{33,44}$) and DD induced coupling rate (e_{43}) together with their variances ($\vartheta_{33,44,43}$). This then leads to notable variations of left-peak position (LPP), right peak position (RPP), dual-peak separation (DPS), and peak-height asymmetry (PHA) as well as less evident changes of mean peak height (MPH) and mean peak width (MPW). That is, the six dual-peak descriptors of LPP, RPP, DPS, PHA, MPH and MPW are all interaction-induced spectral observables. From Fig. 3(b), it is easy to find that prediction results are mainly governed by PC1, PC2, and PC3 **while weakly depend on PC5, PC6, and PC7, with PC4 and other dominant PCs almost negligible**. Relevance between all dual-peak descriptors and first five most important PCs has been quantified by the correlation matrix/heat map in Fig. S2. This figure clearly shows that **PC1** is most relevant to LPP, RPP, and DPS with their magnitudes of correlation coefficients being $|\rho_{kq}| = 0.97$ but MPH and PHA are also essential to be included with $|\rho_{kq}| = 0.52$ and 0.31 , respectively; **PC2** is mainly associated with the latter three descriptors, *i.e.* MPH, MPW, and PHA with $|\rho_{kq}| = 0.25, 0.41$, and 0.34 , respectively; **PC3** relies homogeneously on the former three descriptors with $|\rho_{kq}| \approx 0.22$ but is connected to MPH, MPW and PHA with $|\rho_{kq}| = 0.43, 0.31$, and 0.39 , respectively. That means, for the whole noisy spectra illustrated in Fig. 2(a2) of our manuscript obtained with different Ω_m , most relevant interaction-induced spectral features are position and separation variations of two EIT peaks while peak asymmetry, linewidth broadening (or narrowing), and spectral deformation account also for a pertinent interpretability analysis. This issue has been clarified by rewriting/expanding the first paragraph, right column, page 4.

Response to Reviewer 2

Many thanks for kindly remarking “*The author has addressed all my concerns*” and especially for recommending the publication of our current work in Communications Physics.

Response to Reviewer 3

Many thanks for the criticism on lacking a fair and physically consistent comparison between the ML method and standard nonlinear inference techniques, which promotes us to further improve or amend the presentation of our current work. The lack of a proper comparison is, however, not the truth due to our careless fault of an insufficient description on the “nonlinear least squares (NLS)” benchmark in the previous rebuttal and manuscript. In fact, as expected by the reviewer, we have considered a NLS benchmark by performing parameter inference directly with the forward model (*e.g.*, equations described in both Mean-Field SA Model of Methods and Supplementary Note 1) employed to generate the transmission spectral dataset. We feel sorry also for not providing details on how different noise sources are incorporated when calculating the transmission spectral dataset

in the previous manuscript, which has been amended accordingly. With above issues clarified, we believe that the claimed advantages of our ML approach are reliable and meaningful.

Comment #1:

The central issue is the lack of a fair and physically consistent comparison between the ML method and standard nonlinear inference techniques. In the current manuscript, the ML model is trained on spectra generated from the mean-field super-atom model, thereby implicitly exploiting detailed knowledge of the underlying physics. In contrast, the “nonlinear least squares” (NLS) benchmark relies on fitting the spectra with Lorentzian functions to extract peak positions, followed by the use of a linear formula. Importantly, details of this procedure cannot be found in the updated manuscript, but only in the rebuttal. Moreover, this procedure is clearly not based on the same physical model, and the comparison is therefore not meaningful. It does not test the actual performance of nonlinear inference methods applied to the same problem, but rather compares a model-informed regression to a naive fitting procedure.

A proper benchmark would require performing parameter inference directly using the forward model employed to generate the data (e.g., the equations described in Supplementary Note 1), with the microwave Rabi frequency as a free parameter, and possibly additional parameters depending on how noise is introduced. This point is particularly important given the nature of the inference problem. As emphasized in my previous report, the task is intrinsically low-dimensional (in its simplest form, a single unknown parameter, and at most a few parameters when accounting for fluctuations). In such a setting, standard nonlinear fitting or numerical inversion of the known forward model should provide near-optimal estimates. Without a comparison to such physically grounded methods, the reported performance gains of the ML approach, including the claimed orders-of-magnitude improvement in sensitivity, cannot be considered established.

Answer #1:

We thank the reviewer for adhering to high standards in evaluating our current work and feel sorry for insufficient or misleading descriptions on linear AT-splitting formula and NLS fitting used for comparison with our ML approach. In the previous rebuttal, what we really want to express is that “*a Lorentzian function is adopted to fit the probe transmission spectra and determine positions of two EIT peaks only when predicting MW amplitudes via linear AT-splitting formula*”. As to NLS fitting, previously we fail to provide any details either in the manuscript or in the rebuttal. Here we would like to stress that our NLS fitting is based on indeed the forward physical model (described in both Mean-Field SA Model of Methods and Supplementary Note 1) employed to generate the transmission spectral dataset as expected by the reviewer. Please find below a detailed description on how NLS fitting is used to infer MW amplitudes from noisy transmission spectra.

First, based on the steady-state solutions given in Supplementary Note 1, we can calculate probe susceptibility χ_p and hence probe transmissivity T given in Mean-Field SA Model of Methods, which are functions of various parameters including $\{\Delta_p, \Delta_c, \Delta_m, \Omega_p, \Omega_c, \Omega_m, N_{at}\}$.

Second, taking into account random fluctuations of a few important parameters $\delta\theta = \{\delta\Omega_p, \delta\Omega_c, \delta N_{at}, \delta T\}$, we simulate real experimental results to obtain a dataset of noisy transmission spectrum $\mathbf{T}^{(i)} = \{T(\Delta_c^{(k)}, \Omega_m^{(i)}, \theta_0, \delta\theta)\}_{k=1}^n$ used to train and test regression neural networks for determining $\Omega_m^{(i)}$. Here, Δ_c has been discretized as $\Delta_c^{(k)}$ while θ_0 refers to all other parameters.

Third, we calculate $\mathbf{T}^{NLS} = \{T(\Delta_c^{(k)}, \Omega_m^{NLS}, \theta_0)\}_{k=1}^n$ independent of $\delta\theta$ to perform NLS fitting by varying Ω_m^{NLS} in an appropriate range to minimize the sum of squared residuals

$$SSR = \sum_{k=1}^n \left[T(\Delta_c^{(k)}, \Omega_m^{(i)}, \boldsymbol{\theta}_0, \delta\boldsymbol{\theta}) - T(\Delta_c^{(k)}, \Omega_m^{\text{NLS}}, \boldsymbol{\theta}_0) \right]^2$$

until it converges to a sufficiently small value ε . Thereby, we can use Ω_m^{NLS} corresponding to $SSR \rightarrow \varepsilon$ to predict the real MW Rabi frequency $\Omega_m^{(i)}$. Note that, in general, we should include $\{s_{33}, s_{44}, e_{43}, \vartheta_{33}, \vartheta_{44}, \vartheta_{43}\}$ due to vdW and DD interactions in above calculations but can take them as zero when these interactions are purposely neglected. Below are a few representative plots on simulated noisy spectra $\mathbf{T}^{(i)} = \{T(\Delta_c^{(k)}, \Omega_m^{(i)}, \boldsymbol{\theta}_0, \delta\boldsymbol{\theta})\}_{k=1}^n$ shown in Black and NLS fitting spectra $\mathbf{T}^{\text{NLS}} = \{T(\Delta_c^{(k)}, \Omega_m^{\text{NLS}}, \boldsymbol{\theta}_0)\}_{k=1}^n$ shown in RED in two cases where vdW and DD interactions are neglected (a1-a3) or included (b1-b3).

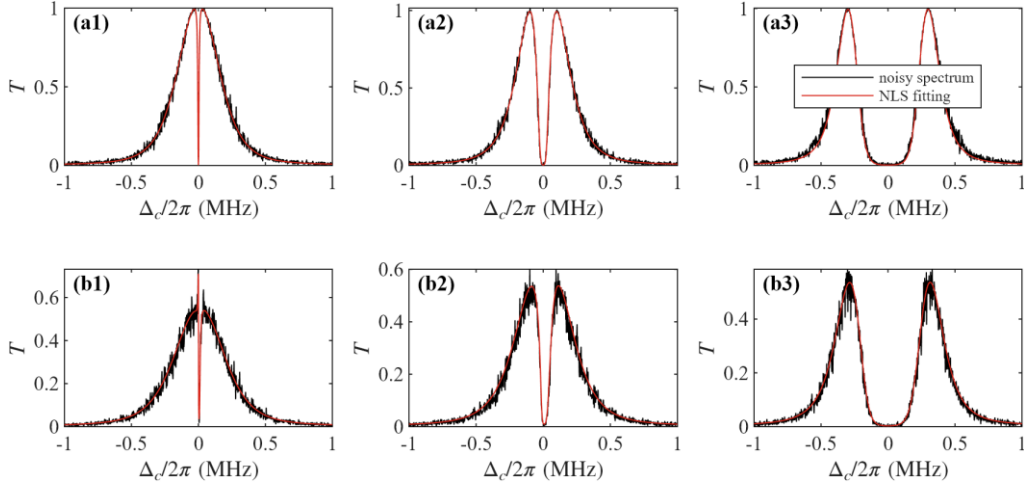


Fig. A1. Probe transmission T against coupling detuning Δ_c in two different cases of non-interacting (a1, a2, a3) or interacting (b1, b2, b3) Rydberg atoms. Panels (a1, b1), (a2, b2), and (a3, b3) are obtained with $\Omega_m^{(i)}/2\pi = 30$ kHz, 100 kHz, and 300 kHz, respectively, by including noise sources (black) or performing NLS fitting (red).

Based on above discussions, we believe that a fair and physically consistent comparison between our ML approach and a standard nonlinear inference technique has been made in our work. In this regard, it can be concluded that the reported performance gains of our ML approach, including the claimed orders-of-magnitude improvement in sensitivity, have been well established. Please note that the tested single-parameter NLS estimator should be distinguished from an ideal noiseless nonlinear inversion built upon a perfectly known forward physical model, which would naturally provide a near-optimal baseline. The performance of this tested NLS estimator must be limited in our noisy mean-field SA model because it is difficult or impossible to well fit noisy transmission spectra $\mathbf{T}^{(i)}$ distorted due to Rydberg interactions by minimizing SSR to a good level when both noise sources $\delta\boldsymbol{\theta}$ and interaction-induced variables $\{s_{33}, s_{44}, e_{43}, \vartheta_{33}, \vartheta_{44}, \vartheta_{43}\}$ depend on Δ_c . In other words, it is difficult or impossible to simultaneously minimize SSR at different values of Δ_c in distorted noisy transmission spectra when $\delta\boldsymbol{\theta}$ and $\{s_{33}, s_{44}, e_{43}, \vartheta_{33}, \vartheta_{44}, \vartheta_{43}\}$ vary with Δ_c independently. This argument is confirmed by numerical results in Fig. A1 shown above and those in Fig. 2 of the manuscript, where NLS fitting behaves clearly better than linear formula in the absence of Rydberg interactions but NOT in the presence of Rydberg interactions.

Comment #2:

In addition, the manuscript remains unclear regarding how noise is incorporated into the datasets. From the rebuttal, noise sources appear to be introduced by varying parameters “in sequence”,

which suggests that only a subset of parameters is varied at a time. I could not find clear details about this point in the updated manuscript. It is therefore unclear whether the ML models are trained and evaluated on datasets where all parameters fluctuate simultaneously, as would be expected under realistic experimental conditions. This point directly affects the effective dimensionality and difficulty of the inference problem and should be clarified.

Answer #2:

We thank the reviewer for reminding us of an unclear issue on the introduction of different noise sources into spectral datasets and feel sorry for an improper use of “*in sequence*” when describing different numerical results obtained by linear formula, NLS fitting, and ML approach, respectively. That is, we didn’t mention anywhere noise sources are introduced by varying relevant parameters “*in sequence*”. The fact is that random fluctuations of a few important parameters (Ω_p , Ω_c , N_{at} , and T) are simultaneously introduced when producing noisy spectral datasets with the Mean-Field SA Model described in both Methods and Supplementary Note 1. To be more specific, we have replaced $\{\Omega_p, \Omega_c, N_{at}\}$ by $\{\Omega_p + \delta\Omega_p, \Omega_c + \delta\Omega_c, N_{at} + \delta N_{at}\}$ when employing Eqs. (4) and (5) in the manuscript to calculate T and meanwhile replaced T by $T + \delta T$ with $\delta\Omega_p$, $\delta\Omega_c$, δN_{at} , and δT being random fluctuations of Ω_p , Ω_c , N_{at} , and T . Detailed procedures for introducing these random fluctuations or noise sources are described as follows: **First**, we can set $\delta\Omega_p/\Omega_p = \eta \cdot N_1(0,1)$, $\delta\Omega_c/\Omega_c = \eta \cdot N_2(0,1)$, $\delta N_{at}/N_{at} = \eta \cdot N_3(0,1)$, and $\delta T = \eta \cdot N_4(0,1)$ with $\eta = 2\%$ being a common (relative) fluctuation amplitude and $N_{1,2,3,4}(0,1)$ four independent normal random numbers; **Second**, for a certain MW Rabi frequency $\Omega_m^{(i)}$, we vary the coupling detuning Δ_c to calculate the probe transmission spectrum $\mathbf{T}^{(i)}(\Delta_c)$ with each value of Δ_c corresponding to a different combination of $N_{1,2,3,4}(0,1)$ so that $\delta\Omega_p/\Omega_p$, $\delta\Omega_c/\Omega_c$, $\delta N_{at}/N_{at}$, and δT must be different and vary also with Δ_c ; **Third**, we repeat the second step N times to obtain all probe transmission spectra $\mathbf{T}^{(i)}(\Delta_c)$ for the N values of $\Omega_m^{(i)}$ to be measured with $N_{1,2,3,4}(0,1)$ being even different for same values of Δ_c as long as $\Omega_m^{(i)}$ changes a little so that $\delta\Omega_p/\Omega_p$, $\delta\Omega_c/\Omega_c$, $\delta N_{at}/N_{at}$, and δT are also $\Omega_m^{(i)}$ -dependent. Moreover, we note that normal random distributions $N_{1,2,3,4}(0,1)$ can be easily generated by calling “randn” separately in our MATLAB code with “0” referring to the mean value while “1” the standard deviation.

Response to Reviewer 1

We thank the reviewer for the careful assessment of our revised manuscript and for confirming that the previous concerns regarding the nonlinear least-squares benchmark and the noise model have been satisfactorily addressed. We appreciate the reviewer's support for publication.