

This is the accepted manuscript made available via CHORUS. The article has been published as:

Utilizing machine learning to improve the precision of fluorescence imaging of cavity-generated spin squeezed states

Benjamin K. Malia, Yunfan Wu, Julián Martínez-Rincón, and Mark A. Kasevich

Phys. Rev. A **105**, L010602 — Published 24 January 2022

DOI: [10.1103/PhysRevA.105.L010602](https://doi.org/10.1103/PhysRevA.105.L010602)

Utilizing machine learning to improve the precision of fluorescence imaging of cavity-generated spin squeezed states

Benjamin K. Malia,¹ Yunfan Wu,² Julián Martínez-Rincón,¹ and Mark A. Kasevich^{1,2,*}

¹*Department of Physics, Stanford University, Stanford, California 94305, USA*

²*Department of Applied Physics, Stanford University, Stanford, California 94305, USA*

(Dated: January 11, 2022)

We present a supervised learning model to calibrate the photon collection rate during the fluorescence imaging of cold atoms. The linear regression model finds the collection rate at each location on the sensor such that the atomic population difference equals that of a highly precise optical cavity measurement. This 192 variable regression results in a measurement variance 27% smaller than our previous single variable regression calibration. The measurement variance is now in agreement with the theoretical limit due to other known noise sources. This model efficiently trains in less than a minute on a standard personal computer's CPU, and requires less than 10 minutes of data collection. Furthermore, the model is applicable across a large changes in population difference and across data collected on different days.

Machine learning (ML) is becoming an increasingly important tool for analysis of scientific data due to its ability to handle large data sets with many dependant variables. In supervised learning, a regression model can be trained on previously collected data to find relationships between an experimental result and potentially dozens of parameters. A successful ML model will be able to accurately predict future data from these many parameters.

A variety of ML algorithms are able to process high resolution images. In experiments involving fluorescence imaging, ML has recently been applied to denoise images [1], classify objects in images [2], classify spectral signatures [3], and quantify fluorescent decay lifetimes [4, 5]. In cold atom experiments, fluorescence imaging is used for atom number measurement to determine the populations of different quantum states. A variety of challenges, including scattered light, sensor read noise, and inhomogeneous photon collection, increase the difficulty of obtaining accurate measurements [6]. The need for accurate imaging is vital in the context of quantum metrology. In these experiments, imaging imperfections fundamentally limit the efficacies of quantum protocols. For example, Qu et al. [7] used principle component analysis to correct for background light in order to fully characterize a quantum state.

Atomic clocks and sensors operate through measurement of population differences between quantum states. For two spin states $|\uparrow\rangle$ and $|\downarrow\rangle$, the collective spin is defined as $J_z = (N_{|\uparrow\rangle} - N_{|\downarrow\rangle})/2$ where $N = N_{|\uparrow\rangle} + N_{|\downarrow\rangle}$ is the total number of atoms in the sensor. The changes in J_z are measured through two consecutive measurements, one before and one after the sensor has detected a field. We define the population difference as $J_z^{(1,2)} = J_z^{(2)} - J_z^{(1)}$.

In the absence of environmental fields, a single measurement of $J_z^{(1,2)}$ should be close to, but not exactly equal to, zero due to the Heisenberg uncertainty principle. In a typical sensor, without spin-squeezing, the quantum projection noise (QPN) limits the variance of

the measurement to $(\Delta J_z^{(2)})^2 = N/4$ [8–10]. If the quantum state is squeezed with a quantum non-demolition (QND) measurement, as in our experiment, then the variance in population difference is reduced such that $\Delta J_z^{(1,2)}$ surpasses the QPN. While $J_z^{(1)}$ is a QND measurement, $J_z^{(2)}$ can be either another QND measurement or a fluorescence measurement. Squeezing allows for increased sensor precision when increasing atom number or interrogation time is not possible.

Here we employ ML to identify and correct for a class of imperfections associated with non-uniform spatial imaging profiles. Potential sources which may contribute to non-uniform spatial imaging are the angle of light entering the objective lens [11], inhomogeneous transparency of vacuum chamber walls, and an inhomogeneous imaging beam power. As we show below, these imperfections, inferred to be 10% across a mm-scale spatially extended cloud, led to a 33% increase in the noise variance in the work of Ref. [12].

To evaluate the performance of the system as a sensor, we define this difference in terms of angular difference $\theta = J_z^{(1,2)}/C(N/2)$ on the multiparticle Bloch sphere, where C is the contrast in a Ramsey sequence. In the case of the data in [12], which is the subject of our analysis in this paper, $C = 0.92$ and the QPN is defined as $\Delta\theta = 1/\sqrt{N}$. As an example for a sensor with $N = 390\,000$, the QPN is $\Delta\theta = 1.6$ mrad. Eliminating extra noise sources is crucial to taking advantage of the noise reduction from spin squeezing.

The relevant details of our experiment [12] are summarized here. ^{87}Rb atoms are cooled and trapped in an optical lattice. The QND measurement is first performed by detecting a cavity-QED dispersive readout phase using an optical light pulse. The lattice is then turned off and the atoms free fall. To perform population spectroscopy at the end of the free fall time, an on-resonance laser pulse imparts a momentum kick to the atoms in the $|\uparrow\rangle$ state. After the two states spatially separate, a

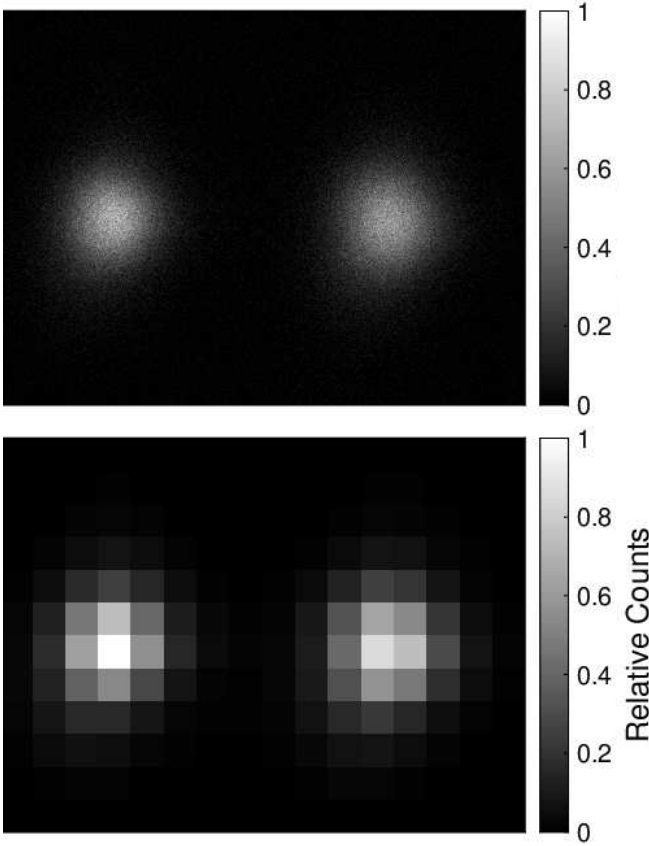


FIG. 1. Relative counts for each pixel. Full resolution image with 3.1×10^6 pixels (upper) and counts binned into 192 superpixels (lower). The left half of the image ($|\downarrow\rangle$) is subtracted from the right half ($|\uparrow\rangle$) to obtain the population difference.

retroreflected laser induces fluorescence in both states. The light is collected with an objective lens with a numerical aperture of 0.25 and is sent to a CMOS camera. Each state takes up approximately half of the sensor region. $J_z^{(2)}$ is determined by the difference in total counts between states $|\uparrow\rangle$ and $|\downarrow\rangle$. As a result of the mechanisms discussed earlier, the average number of counts per atom in each state are several percent different. Due to shot-to-shot fluctuations in the atoms' positions, the spatial inhomogeneity in the scattering rate directly leads to fluctuations in the inferred atomic populations. This additional noise is large enough that, uncorrected, results in ΔJ_z larger than the QPN.

The supervised learning model estimates $J_z^{(2)}$ by first dividing the fluorescence image into 192 “superpixels” (Figure 1). These superpixels are simply 128×128 pixel squares grouped into blocks. Our new regression model now has $n = 192$ variables to fit to. Each superpixel is then assigned a weight β_j which is multiplied by its total counts c_j to estimate the correct atom number for that pixel. The estimated $\tilde{J}_z^{(2),(i)}$ are determined by subtracting the estimated atoms in the left half of the image from

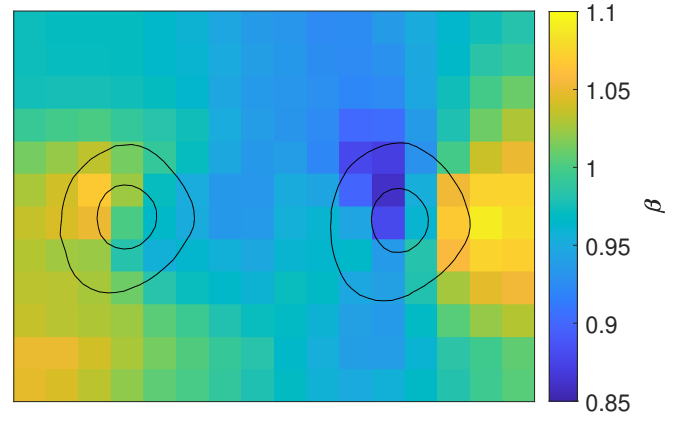


FIG. 2. Relative values for the weights β_j for each superpixel under the parameters $J_z^{\text{cutoff}} = 200$ and $\lambda = 20$. Overlay: inner curves enclose 68% of atoms and outer curves enclose 95% of atoms in a sample image (upper Figure 1).

the ones in the right half:

$$\tilde{J}_z^{(2),(i)} = \beta_0 + \frac{1}{2} \sum_{j=1}^{n/2} c_j^{(i)} \beta_j - \frac{1}{2} \sum_{j=n/2+1}^n c_j^{(i)} \beta_j \quad (1)$$

where i indexes the images and β_0 is a bias term. We can also determine the total atom number N from these β_j :

$$\tilde{N}^{(i)} = \sum_{j=1}^n c_j^{(i)} \beta_j \quad (2)$$

The inhomogeneity must at least vary on length scales smaller than the atom cloud to make a significant impact on the precision of the measurement. The maximum spatial frequency accessible to the model is determined by the resolution of the images. We therefore chose a superpixel resolution large enough to capture most of the spatial variation, but small enough to allow for fast training on small sample sizes. As described below, we enforce a penalty term which puts an upper bound on the spatial frequency of the model's result.

Consecutive low noise cavity measurements show that detection limit of the squeezed state is $\Delta\theta = 310 \mu\text{rad}$, which is lower than the $690 \mu\text{rad}$ theoretical limit of the fluorescence detection (see [12] supplement for details on these noise sources). We therefore take $J_z^{(1)}$ as the target for $J_z^{(2)}$ when training our regression model.

We calculate the optimal β_j by minimizing the cost function, G , with the BroydenFletcherGoldfarbShanno algorithm [13] over m samples. Specifically, the *fminunc* function in MATLAB is supplied with both G and its gradient [14]. G is defined as a weighted least squares with regularization [15]:

$$G = \frac{1}{2m_c} \sum_{i=1}^m w^{(i)} (\tilde{J}_z^{(2),(i)} - \tilde{J}_z^{(1),(i)})^2 + \frac{\lambda}{2m_c} P_{NN}. \quad (3)$$

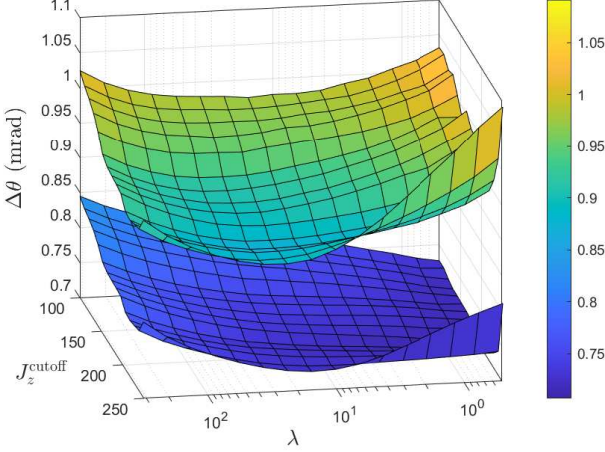


FIG. 3. The evaluation metric $\Delta\theta$ for the validation set (upper surface) and the training set (lower surface). The minimum point on the validation set is used to choose the optimal hyperparameter values, $\lambda = 20$ and $J_z^{\text{cutoff}} = 200$.

Here, the weights $w^{(i)}$ are defined by the Heaviside step function H and the magnitudes of $J_z^{(1),(i)}$:

$$w^{(i)} = H(|J_z^{(1),(i)}| - J_z^{\text{cutoff}}), \quad (4)$$

and m_c is the total number of samples with non-zero weight. $\tilde{J}_z^{(1),(i)}$ is the frequency-corrected cavity measurement ([12] supplement, Equation S8)

$$\tilde{J}_z^{(1),(i)} = J_z^{(1),(i)} + \frac{\delta^{(i)}}{2\Delta} \tilde{N}^{(i)} \quad (5)$$

where $\delta^{(i)}$ is the difference in frequency between the probe laser and the target frequency (detuned Δ from the $|\uparrow\rangle \rightarrow |e\rangle$ transition). Since this correction is relatively small, the uncorrected values are used in Equation 4 to simplify the G 's dependence on the β 's. The nearest neighbor penalty is

$$P_{NN} = \sum_{\langle a,b \rangle} (\beta_b - \beta_a)^2 \quad (6)$$

where $\langle a,b \rangle$ represents the set of all unique nearest neighbor pairs of superpixels and λ is the regularization hyperparameter. The $(\Delta\theta)^2$ predicted by this model is the variance of $\tilde{\theta} = (\tilde{J}_z^{(2)} - \tilde{J}_z^{(1)})/(C\tilde{N}/2)$.

We train the model on a set of 500 images taken from the data collected for Ref. [12]. 120 images remain after J_z^{cutoff} is applied. The hyperparameters, J_z^{cutoff} and λ , are chosen to give the smallest $\Delta\theta$ in a separate validation set of 50 images. Figure 3 shows the results of training with a range of values for each hyperparameter. The optimal values are $J_z^{\text{cutoff}} = 200$ and $\lambda = 20$. Repeating the procedure with a validation set half the size results in the same optimal values. Note that samples with $|J_z^{(1)}|$

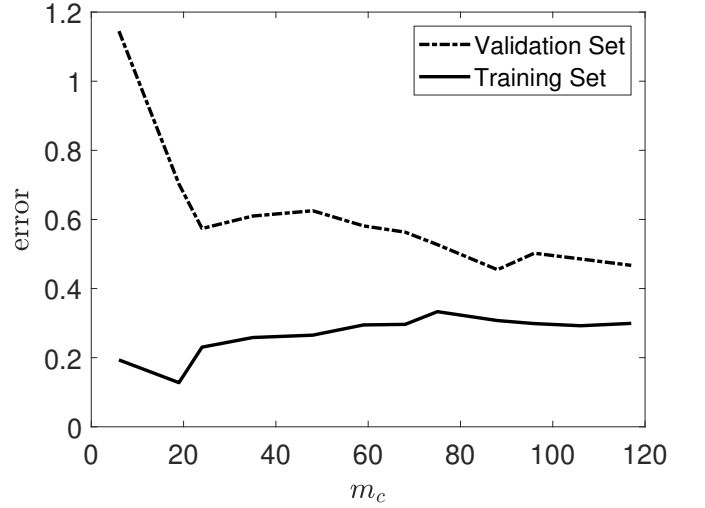


FIG. 4. Mean least squares error (first term of the cost function G) vs number of samples with non-zero weight (m_c). Optimal values of $J_z^{\text{cutoff}} = 200$ and $\lambda = 20$ are used. Final difference between sets is 0.17.

this large are typically excluded when the system is used as a sensor as those $J_z^{(1)}$ closest to zero are most accurate. Therefore, the validation set will have larger $\Delta\theta$ than that of the final result. The β_j determined with these parameters are shown in Figure 2. $\beta_0 < 10^{-3}$ and does not significantly contribute to $\tilde{J}_z^{(2)}$.

Figure 4 shows the learning curves as a function of nonzero weight sample size (m_c). Relatively few data points are needed for the model to minimize the difference of mean least squares error (the first term in Equation 3) between the training and validation sets. The maximum number of points used ($m_c = 117$) corresponds to approximately 8 minutes of data collection and 1 minute of training. Increasing the model complexity by increasing the number of superpixels does not significantly change the gap between the training and validation set errors at maximum m_c . It also does not change the qualitative structure of the β distribution. On the other hand, a model with less complexity fails to minimize the gap.

At the cost of sample size, using the Heaviside function to set the weights prevents the model from optimizing on values near zero and reaching a trivial solution where all β 's approach zero. This eliminates a $\tilde{J}_z^{(1)}$ dependent trend in the residual $\tilde{J}_z^{(2)} - \tilde{J}_z^{(1)}$ and an unreasonably small $\tilde{N}^{(i)}$. The nearest neighbor penalty acts similarly to ridge regularization [15] in that it prevents the model from overfitting and keeps a relatively similar contribution from each β_j . Another benefit is that it does not require the total atom number to be fixed before training the model. When both the weights and nearest neighbor penalty are included, the mean of predicted atom numbers $\tilde{N}^{(i)}$ are within 2% of the previously calibrated mean value. Finally, the optimal λ reveals the character-

mean $\tilde{\theta}^{(2)}$ (mrad)	no correction (mrad)	mean position correction (mrad)	supervised learning (mrad)
0	1.86 ± 0.09	0.81 ± 0.06	0.69 ± 0.06
25	2.23 ± 0.15	1.03 ± 0.06	0.88 ± 0.06

TABLE I. Angular resolution $\Delta\theta$ of the test sets for different analysis methods. The set where the fluorescence measurement has a mean value of $\tilde{\theta}^{(2)} = \tilde{J}_z^{(2)}/C(\tilde{N}/2) = 0$ contains $N = 390\,000$ (QPN limit is $\Delta\theta = 1.60$ mrad). The set with $\theta^{(2)} = 25$ mrad contains $N = 240\,000$ (QPN limit is $\Delta\theta = 2.04$ mrad).

istic spatial frequency of the inhomogeneity.

Despite the model fitting only on samples with $|J_z^{(1),(i)}|$ larger than the cutoff, it extrapolates successfully onto those with values smaller than the cutoff. Applying the model to the remaining data (samples with $|J_z^{(1),(i)}| < J_z^{\text{cutoff}}$) results in an angular resolution of $\Delta\theta = 691 \pm 60$ μrad . This variance is 27% smaller than our previous reported result (see Table I). It is also consistent with our previously predicted limit given the known noise sources. In terms of metrologically useful squeezing, this is 7.2 ± 0.3 dB below the QPN limit.

This collection efficiency map is independent of the mean J_z as well as the mean position of the atoms and the total atom number. It is therefore versatile and can be applied to data sets taken on different days with different parameters. For example, with measurements where the atomic state is rotated along the polar angle of the Bloch sphere by 25 mrad and the total atom number is reduced by 40%, the model now predicts $\Delta\theta = 880 \pm 60$ μrad (7.3 ± 0.3 dB below the QPN), a 27% reduction in variance upon the previous method (see Table I).

The collection inhomogeneity across the cloud can also lead to a degradation in the correlation between a single image and its corresponding QND measurement [16]. In this work, the estimated increase in variance due to this noise source is 0.5% [17], which is negligible. The ML model demonstrated here primarily reduces the variance from shot-to-shot fluctuations in the atom cloud distributions coupled with the detection inhomogeneity.

In our previously reported results we applied a single weight, determined by the mean position of the atoms, to all of the counts from the $|\uparrow\rangle$ state. Although this method provided significant improvement (Table I), the atoms are spread out over most of the image so applying the same factor to all counts is an over-simplification. ML provides the capability of increasing the complexity of our model to accurately calibrate the measurement.

In this work, we have shown that supervised learning is capable of characterizing inhomogeneous fluorescence measurements when prior low noise measurements are available for training the model. This model is robust to changes in spatial arrangement of the atomic populations and correctly predicts the total atom number. This

method could be extended to absorption imaging of cold atoms. Absorption imaging, especially high intensity imaging, requires careful calibration to accurately count atoms [18, 19]. A ML approach could circumvent traditional calibration techniques to convert pixel-by-pixel intensity changes to atom numbers in spin squeezing experiments [20].

Other ML techniques have the potential to further improve our result. In Ness et al. [21], deep learning is used to predict the background noise based on the signal surrounding the atoms. A similar approach may allow our procedure to forgo subtracting the second background image. This would reduce the variance introduced by the background by up to half of its current value.

This work is supported by the Department of Energy (DE-SC0019174-0001) and the Vannevar Bush Faculty Fellowship.

* kasevich@stanford.edu

- [1] Y. Wang, H. Pinkard, E. Khwaja, S. Zhou, L. Waller, and B. Huang, *bioRxiv* (2021), [10.1101/2021.02.01.429188](https://doi.org/10.1101/2021.02.01.429188).
- [2] M. A. K. Sagar, K. P. Cheng, J. N. Ouellette, J. C. Williams, J. J. Watters, and K. W. Eliceiri, *Frontiers in Neuroscience* **14**, 931 (2020).
- [3] L. Ju, A. Lyu, H. Hao, W. Shen, and H. Cui, *Analytical Chemistry* **91** (2019), [10.1021/acs.analchem.9b01315](https://doi.org/10.1021/acs.analchem.9b01315).
- [4] V. Mannam, Y. Zhang, X. Yuan, C. Ravasio, and S. S. Howard, *Journal of Physics: Photonics* **2**, 042005 (2020).
- [5] J. T. Smith, R. Yao, N. Sinsuebphon, A. Rudkouskaya, N. Un, J. Mazurkiewicz, M. Barroso, P. Yan, and X. Intes, *Proceedings of the National Academy of Sciences* **116**, 24019 (2019).
- [6] E. Rocco, R. N. Palmer, T. Valenzuela, V. Boyer, A. Freise, and K. Bongs, *New Journal of Physics* **16**, 93046 (2014).
- [7] A. Qu, B. Evrard, J. Dalibard, and F. Gerbier, *Phys. Rev. Lett.* **125**, 033401 (2020).
- [8] O. Hosten, N. J. Engelsen, R. Krishnakumar, and M. A. Kasevich, *Nature* **529**, 505 (2016).
- [9] K. C. Cox, G. P. Greve, J. M. Weiner, and J. K. Thompson, *Phys. Rev. Lett.* **116**, 093602 (2016).
- [10] I. D. Leroux, M. H. Schleier-Smith, and V. Vuletić, *Phys. Rev. Lett.* **104**, 073602 (2010).
- [11] P. B. Catrysse and B. A. Wandell, *Journal of the Optical Society of America* **19** (2002), [10.1364/JOSA.19](https://doi.org/10.1364/JOSA.19).
- [12] B. Malia, J. Martinez-Rincon, Y. Wu, O. Hosten, and M. A. Kasevich, *Physical Review Letters* **125** (2020), [10.1103/PhysRevLett.125.043201](https://doi.org/10.1103/PhysRevLett.125.043201).
- [13] C. G. Broyden, *Journal of the Institute of Mathematics and Its Appl.*
- [14] See Supplemental Material at [URL will be inserted by publisher] for pseudocode describing the full implementation of the algorithm.
- [15] A. Neumaier, *SIAM Review* **40**, 636 (1998).
- [16] J. Hu, W. Chen, Z. Vendeiro, H. Zhang, and V. Vuletić, *Phys. Rev. A* **92**, 063816 (2015).
- [17] Y. Wu, R. Krishnakumar, J. Martinez-Rincon,

- B. K. Malia, O. Hosten, and M. A. Kasevich, *Phys. Rev. A* **102** (2020), 10.1103/PhysRevA.102.012224.
- [18] C. Gross, *Journal of Physics B: Atomic, Molecular and Optical Physics* **45**, 013004 (2012)-ph].
- [19] K. Hueck, N. Luick, L. Sobirey, J. Siegl, T. Lompe, H. Moritz, L. W. Clark, and C. Chin, *Opt. Express* **25**, 8670 (2017).
- [20] M.-Z. Huang, J. A. de la Paz, T. Mazzoni, K. Ott, A. Sinatra, C. L. G. Alzar, and J. Reichel, (2020), [arXiv:2010.03064 \[quant-ph\]](#).
- [21] G. Ness, A. Vainbaum, C. Shkedrov, Y. Florshaim, and Y. Sagi, *Phys. Rev. Applied* **14**, 014011 (2020).