

Channel-noise tracking for sub-shot-noise-limited receivers with neural networks

M. T. DiMario and F. E. Becerra 

Center for Quantum Information and Control, Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA



(Received 27 September 2020; accepted 3 February 2021; published 1 March 2021)

Non-Gaussian receivers for optical communication with coherent states can achieve measurement sensitivities beyond the limits of conventional detection, given by the quantum-noise limit (QNL). However, the amount of information that can be reliably transmitted substantially degrades if there is noise in the communication channel, unless the receiver is able to efficiently compensate for such noise. Here, we investigate the use of a deep neural network as a computationally efficient estimator of phase and amplitude channel noise to enable a reliable method for noise tracking for non-Gaussian receivers. The neural network uses the data collected by the non-Gaussian receiver to estimate and correct for dynamic channel noise in real time. Using numerical simulations, we find that this noise tracking method allows the non-Gaussian receiver to maintain its benefit over the QNL across a broad range of strengths and bandwidths of phase and intensity noise. The noise tracking method based on neural networks can further include other types of noise to ensure sub-QNL performance in channels with many sources of noise.

DOI: [10.1103/PhysRevResearch.3.013200](https://doi.org/10.1103/PhysRevResearch.3.013200)

I. INTRODUCTION

The intrinsic properties of coherent states can enable efficient and practical classical [1–4] and quantum [5–9] communications. When utilizing the phase of coherent states combined with their intensity to encode and transmit information, higher rates of information transfer may be achieved compared to communication schemes using intensity-only encodings [4,10]. However, channel noise can severely limit the advantage of communications with coherent encodings. In conventional coherent communications, the optical receiver performs a heterodyne measurement with shot-noise-limited sensitivity, corresponding to the quantum-noise limit (QNL). This measurement allows for the use of postprocessing methods of the collected data to estimate channel noise and correct the data to recover the transmitted information [10–19]. While current coherent optical communications rely on these conventional approaches, a heterodyne measurement cannot reach the ultimate limit of sensitivity [20] and information transfer [1,21,22].

In contrast to conventional strategies, non-Gaussian receivers can surpass the QNL, providing higher measurement sensitivities for decoding information [23–30]. However, in the presence of channel noise, the benefit of non-Gaussian receivers over conventional strategies critically depends on the ability to perform efficient channel-noise tracking. Recent work demonstrated an efficient method for phase tracking

for non-Gaussian receivers [31]. This phase tracking method estimates and corrects for the phase noise in real time, which is required by the strategies used in non-Gaussian receivers, as opposed to postprocessing of the collected data with heterodyne detection. This method enabled sub-QNL sensitivity in the presence of phase noise, which is particularly damaging for coherent encodings [31,32].

In more realistic situations there may be multiple sources of noise present in the communication channel, such as thermal noise [33,34], phase diffusion [35–37], phase noise, and amplitude noise. In such situations the non-Gaussian receiver must perform efficient high-dimensional parameter estimation and tracking in order to maintain the expected sub-QNL performance. However, current methods for single-parameter noise tracking cannot be efficiently scaled to higher dimensions for tracking and correction of multiple sources of noise. Thus, enabling noise tracking for non-Gaussian receivers in channels with complex and dynamic noise requires novel and efficient methods for multiparameter estimation that scale favorably to higher dimensions. Practical parameter tracking also requires estimation on a timescale which is very small ($\ll 1\%$) compared to the bandwidth of the channel noise. For example, realistic kilohertz scale phase noise [4,17,18] would require estimation on at least megahertz timescales, and a Bayesian estimator may not be compatible with this requirement.

Machine learning has been shown to be a powerful tool for solving many problems in coherent communications where conventional methods may be inefficient or computationally difficult [38–42]. In particular, artificial neural networks [43] have seen broad applications in quantum information [44–53] and optical communications [54–58], and for channel-noise estimation and monitoring [59–61]. While machine learning techniques benefit current communication technologies, their application for parameter tracking for non-Gaussian receivers

*fbecerra@unm.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

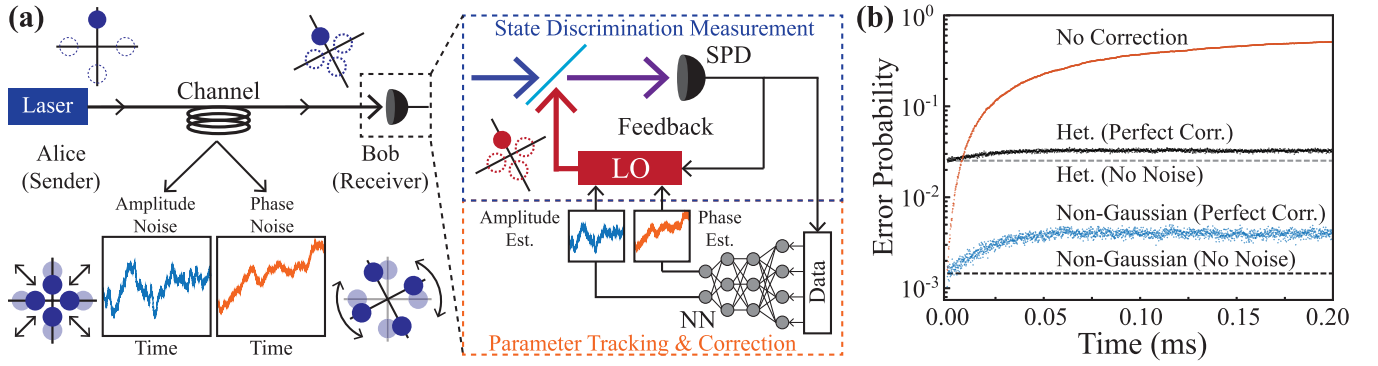


FIG. 1. Channel-noise tracking method. (a) Schematic of two parties, Alice and Bob, communicating over a noisy channel with Bob at the receiver performing phase and amplitude noise tracking. The receiver uses an adaptive non-Gaussian measurement strategy to realize state discrimination below the QNL using interference between the input state and a local oscillator (LO) followed by photon counting with a single photon detector (SPD). A neural network (NN) takes the data collected from the non-Gaussian receiver and outputs estimates for the current phase offset and input intensity. The estimates are fed forward to the LO to match the input intensity and counteract phase noise in real time. (b) Effect of channel noise with (blue) and without (orange) perfect parameter tracking (see main text) for the non-Gaussian receiver. Black points show a perfectly corrected heterodyne receiver, and black and gray dashed lines show the error for a non-Gaussian and a heterodyne receiver in the absence of noise, respectively.

with sub-shot-noise-limited performances have yet to be investigated.

In this work, we numerically investigate a method for multiparameter channel-noise tracking based on a neural network (NN) estimator for a non-Gaussian receiver with sub-QNL sensitivity for state discrimination of quaternary phase-shift-keyed (QPSK) coherent states. We construct a NN as a precise and computationally efficient multiparameter estimator for tracking the time-varying phase and intensity of the input coherent states, and benchmark its performance against a Bayesian estimator, which is expected to be accurate but is computationally expensive to calculate. We find that, across a broad range of channel-noise strengths and input powers, the NN-based method for noise tracking shows similar performance to a Bayesian-based noise tracking approach, and allows the non-Gaussian receiver to maintain sub-QNL sensitivity. This shows that a NN estimator is a viable method for real-time, multiparameter channel-noise tracking in non-Gaussian receivers due to its efficiency and potential scalability to higher dimensions. In Sec. II we describe the non-Gaussian receiver strategy and the NN estimator used for the noise tracking method. In Sec. III we investigate the performance of the channel noise tracking. We discuss the results of the work in Sec. IV.

II. RECEIVER AND ESTIMATION STRATEGY

We numerically study the use of a NN-based method for noise parameter tracking for non-Gaussian receivers based on adaptive measurements and photon counting. As a proof-of-concept, we investigate a NN-based method for tracking phase and amplitude channel noise, that uses only the data collected during the state discrimination measurement. This NN-based method can be easily extended to perform higher dimensional parameter estimation for tracking additional sources of noise in the channel such as thermal noise [33,34] or phase diffusion [35–37]. In this section, we describe (a) the measurement strategy of the non-Gaussian receiver for coherent state dis-

crimination; (b) how the data from the measurement is used by the NN; and (c) the NN estimator, which can be used for estimation of channel noise from multiple sources.

A. State discrimination measurement

Figure 1(a) shows a scenario where a receiver attempts to perform coherent state discrimination with an adaptive photon counting measurement with sensitivity below the QNL [25,26]. Dynamic phase and amplitude noise induced by the communication channel degrades the attainable sensitivity of the receiver. Tracking the phase and amplitude noise of the input states induced in the channel using the data collected during the discrimination measurement can in principle allow the receiver to correct its strategy and maintain sub-QNL sensitivity.

Here, we study a method for channel-noise tracking for a receiver based on an adaptive non-Gaussian strategy [26] for phase-coherent states $|\alpha_k\rangle \in \{|\alpha e^{i2\pi k/M}\rangle\}$, where $k = 0, 1, \dots, M-1$. For $M = 4$, this corresponds to QPSK coherent states. The state discrimination strategy consists of L adaptive measurement steps. Each step performs a hypothesis test of the input state using a local oscillator (LO) to implement a displacement operation $\hat{D}(\beta)$ through interference and single photon counting. In each adaptive step $j = 1, 2, \dots, L$, the receiver attempts to displace the most likely state to the vacuum state by adjusting the LO phase $\arg(\beta) = \theta_j \in \{0, \pi/2, \pi, 3\pi/2\}$ with $|\beta| = |\alpha_k|$, followed by single photon detection. The detector has a finite photon number resolution (PNR) where up to m photons can be resolved, denoted as $\text{PNR}(m)$, before becoming a threshold detector [26]. At the end of the L adaptive steps, the best guess of the receiver θ_{disc} for the true input phase is the state with maximum *a posteriori* probability given the entire detection history. As described in Secs. II B and II C, the photon counting data from the adaptive measurement steps together with θ_{disc} allows the receiver to perform phase and amplitude tracking, where estimates of the

channel noise are fed forward to the LO in order to maintain the sub-QNL performance of the receiver.

Figure 1(b) shows an example of the error probability for the adaptive non-Gaussian receiver for QPSK states for an average input mean photon number of $\langle \hat{n} \rangle_0 = |\alpha|^2 = 5.0$, which is proportional to the intensity, averaged over 5000 noise realizations and obtained through Monte Carlo simulations. For all Monte Carlo simulations in this study, we assume ideal detection efficiency, zero detector dark counts, a photon number resolution of PNR(10), and $L=10$ adaptive steps. To represent a realistic experiment, we use an interference visibility of the displacement operation of 99.7% [26]. The blue (orange) points show the error probability for the non-Gaussian receiver with (without) perfect noise tracking. Perfect tracking refers to a situation where the receiver has complete knowledge of the time-dependent input intensity and phase noise induced by the channel. The black points show the error of an ideal heterodyne measurement, performing at the QNL, with perfect tracking [62]. The dashed lines show the expected error in the absence of noise for a heterodyne (gray) and non-Gaussian (black) receiver. The error for the non-Gaussian measurement remaining below the heterodyne limit (QNL) shows that if the receiver can implement accurate parameter tracking, then its benefit over the QNL can be maintained. Furthermore, any tracking method for the non-Gaussian receiver requires correcting for dynamical noise in real time to ensure sub-QNL performance [31], in contrast to methods for heterodyne receivers, where estimation and correction can be done in postprocessing of the data.

B. Detection matrix

The measurement data collected by the non-Gaussian receiver from the discrimination of N input states is used for parameter estimation. For the discrimination of one input state, this data consists of the L photon detections $\{d_j\}_L$ and relative phases $\{\Delta_j\}_L$ between the LO and input state for each adaptive step j . Due to the low error rate achieved by the non-Gaussian measurement, the guess θ_{disc} of the phase of the input state corresponds to the true input phase with high probability. Thus, θ_{disc} can be used to infer the relative phase Δ_j between the LO and actual input state at every adaptive measurement step j such that $\Delta_j = \theta_j - \theta_{\text{disc}}$, as in [31]. This state discrimination data $\{\Delta_j, d_j\}$ is binned into what we refer to as the detection matrix $\mathbf{D}$, which is a $M \times (m+1)$ matrix, where m is the PNR of the receiver. After each measurement, the matrix elements $D_{k,l}$ are incremented by the total number of times that the number of detected photons in an adaptive step j was $d_j = l$ and the relative phase was $\Delta_j = 2\pi k/M$ for $k \in \{0, 1, \dots, M-1\}$. Thus, the rows of the matrix $\mathbf{D}$ represent the photon number distributions for different relative phases $k\pi/2$ between the LO (θ_j) and final hypothesis (θ_{disc}) for QPSK states [31]. After completing N experiments, the matrix $\mathbf{D}$ contains $N \times L$ pairs $\{d_j, \Delta_j\}$ and it is used for parameter estimation. Once estimation has been performed, the matrix is reset such that $D_{k,l} = 0$ for all l and k . In order to extract information from $\mathbf{D}$ to correct for channel noise affecting the measurement, the receiver must utilize a particular estimator. A Bayesian estimator, which uses the full likelihood functions, will yield estimates for

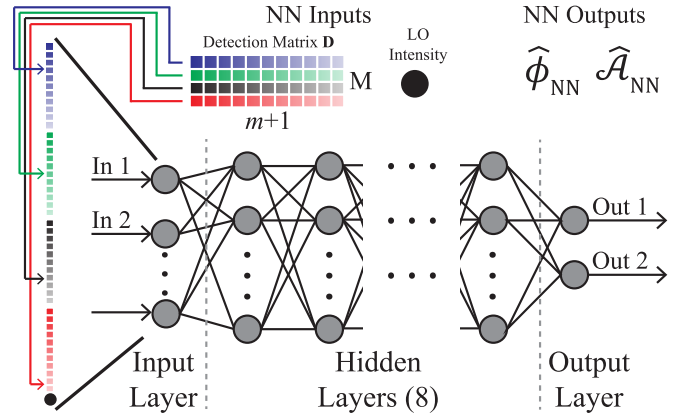


FIG. 2. Neural network. The neural network (NN) for noise estimation has ten layers (eight hidden) with sizes described in Table I. The NN inputs are a flattened version of the detection matrix $\mathbf{D}$ normalized across each row, and the LO intensity $\mathcal{B}$ for the measurements whose data is contained in the detection matrix. The outputs of the NN are estimates for the phase offset $\hat{\phi}_{NN}$ and input intensity $\hat{A}_{NN}$.

the channel noise with small uncertainty [63]. However, this estimator is computationally demanding to calculate. Since the estimation and correction of the channel noise for non-Gaussian receivers must be performed in real time, a Bayesian method may be incompatible with applications requiring high bandwidth sub-QNL receivers. Therefore, to enable practical implementations of non-Gaussian receivers requires an estimator that is both precise and computationally efficient while being easily scalable to higher dimensions to track multiple sources of channel noise. For example, the simple case of single parameter estimation for phase tracking for non-Gaussian measurements has been experimentally demonstrated [31] using a simple estimator, which is calculated in real time with minimal computational resources.

C. Neural network estimator

We construct a NN as a multiparameter estimator which maps the data collected from the state discrimination measurement to estimates for the input intensity and phase offset. We compare the performance of the NN estimator to a Bayesian estimator. The Bayesian-based method for noise tracking serves as a benchmark and is calculated from the same state discrimination measurement data, i.e., the detection matrix $\mathbf{D}$. Although we study phase and amplitude tracking, a properly trained NN can in principle be used as an efficient high-dimensional estimator for tracking many sources of communication channel noise.

Figure 2 shows a diagram of the NN architecture for the proposed noise tracking method, which has ten layers (eight hidden), each with a leaky ReLU activation function [64]. To obtain the input for the NN, the detection matrix $\mathbf{D}$ is first normalized across each row, and then arranged into a one-dimensional vector ($D_{k,l} \rightarrow D_{k(m+1)+l}$). This vector, along with the LO intensity for the previous N measurements, are the inputs to the NN. For ease of notation, we denote the time-dependent input intensity of the QPSK coherent states

TABLE I. Neural network and training parameters.

Network parameters	
Number of layers	Ten (eight hidden)
Hidden layer size (l_i)	{32, 32, 32, 32, 16, 16, 8, 8}
Activation function	Leaky ReLU ($a = 0.1$)
Initialization	Norm($0, \sigma^2 = 1/l_i$) (Xavier)
Training parameters	
Cost function	Weighted mean-squared error
Optimizer	RMSprop (momentum=0.8)
Learning rate	50×10^{-6}
Epochs	2000

as $\mathcal{A}(\tau) = |\alpha|^2(\tau)$ where τ represents time discretized into steps of ΔT , where $1/\Delta T$ is the experimental repetition rate. For a single state discrimination measurement at time τ , the intensity of the LO is denoted as $\mathcal{B}(\tau) = |\beta|^2(\tau)$. The NN outputs, denoted as $\hat{\mathcal{A}}_{NN}$ and $\hat{\phi}_{NN}$, are raw estimates of the input intensity $\mathcal{A}(\tau)$ and relative phase offset $\phi(\tau)$ during the previous N state discrimination measurements. The NN is trained on 5×10^5 samples of the state discrimination measurement generated from Monte Carlo simulations of the experiment in Python. For training the NN, we use the Tensorflow library [65] with a weighted mean-squared-error cost function (see Appendix A for details) [66–69]. The trained NN is then included in the Monte Carlo simulations to perform parameter tracking on the state discrimination data such that the estimates from the NN are fed forward to the LO to correct the measurement.

III. RESULTS

We simulate the performance of the noise tracking method based on the NN estimator for a variety of scenarios with amplitude and phase noise for *average* input intensities $\langle \hat{n} \rangle_0 = \mathcal{A}(0) = \langle \mathcal{A}(\tau) \rangle$ equal to 2, 5, and 10. Here $\langle \cdot \rangle$ denotes the average across all noise realizations at the time step τ . We benchmark the NN against a Bayesian estimator where the prior probability distribution for both parameters is uniform [31]. For all simulations, we use a single NN to perform multiparameter estimation and noise tracking across a range of input powers and noise parameter regimes.

As a model for phase noise $\phi(\tau)$, we simulate a discrete Gaussian random walk in phase [31]. A single step of this walk has a variance of $\sigma_1^2 = 2\pi \Delta\nu \Delta T$ where $\Delta\nu$ is the phase noise bandwidth due to finite laser linewidth [4,17,18] or other phase noise sources [4,70]. The experimental repetition rate is set to $1/\Delta T = 100$ MHz such that $\Delta T = 10$ ns to represent a feasible, near-term communication bandwidth for non-Gaussian receivers [71]. To model amplitude noise of the input states, we simulate noise in the input intensity $\mathcal{A}(\tau)$. As a noise model, we use an Ornstein-Uhlenbeck (OU) process [72,73] whose stochastic differential equation is given by

$$\Delta \mathcal{A}(\tau) = \gamma[\langle \hat{n} \rangle_0 - \mathcal{A}(\tau)]\Delta T + \Sigma dW, \quad (1)$$

where γ is the amplitude noise bandwidth, Σ controls deviation of the walks, and dW denotes a Wiener process. The long-time variance of $\mathcal{A}(\tau)$ is given by $\Sigma_\infty^2 = \Sigma^2/2\gamma$ and the maximum long-time variance we implement is

$\Sigma_\infty^2 = \{0.25, 1.5, 6.0\}$ for $\langle \hat{n} \rangle_0 = \{2, 5, 10\}$, respectively, corresponding to a relative noise level of $\langle \hat{n} \rangle_0 / \Sigma_\infty \approx 0.25$.

After N state discrimination measurements, estimates are calculated from the detection matrix $\mathbf{D}$. To implement correction of the receiver, we set the LO intensity $\mathcal{B}(\tau)$ to the current estimated value $\hat{\mathcal{A}}(\tau)$ of the input intensity $\mathcal{A}(\tau)$. For phase tracking we add a correction $\delta(\tau)$ to the LO phase such that $\arg\{\beta\} = \theta_j + \delta(\tau)$. The correction $\delta(\tau)$ is equal to the cumulative sum of individual estimates $\hat{\phi}$ up to the current time step τ . This is because the receiver is always estimating only the phase shift accumulated in the previous N experiments. The phase and intensity corrections remain fixed at these values for N experiments until new estimates are made and applied to the LO.

To reduce uncertainty in the phase and intensity estimates for noise tracking, we implement a Kalman filter [74] for both estimates (see Appendix B for details). The inputs for the filter are the current raw estimates for the input intensity and phase offset ($\hat{\mathcal{A}}_{NN}, \hat{\phi}_{NN}$), and the filter outputs are updated, filtered estimates for the intensity $\hat{\mathcal{A}}(\tau)$ and phase $\hat{\phi}$. The same procedure is done to obtain filtered Bayesian estimates from the raw estimates ($\hat{\mathcal{A}}_B, \hat{\phi}_B$). To implement the Kalman filter, we assume that the raw NN estimates are Gaussian distributed, and use Monte Carlo simulations with fixed phase offset and input intensity to empirically obtain the variance of the NN estimator. We note that although we study two particular models for phase and amplitude noise, we believe this NN-based tracking method can be applied to a variety of noise forms such as power-law amplitude noise or damping noise. To study different noise models, the NN would need to be retrained using data generated from the new model and the noise dynamics would need to be incorporated into the Kalman filter accordingly.

Figure 3 shows (a) the error probability of the non-Gaussian receiver with noise tracking for 1000 different realizations with both phase ($\Delta\nu = 2$ kHz), and intensity ($\gamma = 25$ kHz, $\Sigma_\infty^2 = 1.5$) noise shown in (b) and (c), respectively, for an input intensity $\langle \hat{n} \rangle_0 = 5.0$ and $N = 10$ experiments per estimation period. The blue (orange) points show the results of the noise tracking method based on NN (Bayesian) estimators. The black points show the error probability with perfect correction, which corresponds to the case where the receiver has complete knowledge of the phase and intensity noise, so that $\mathcal{B}(\tau) = \mathcal{A}(\tau)$ and $\delta(\tau) = \phi(\tau)$. The green points show the error probability of an uncorrected non-Gaussian measurement, and the gray points show that of an ideal heterodyne measurement with perfect phase tracking [equivalent to $\phi(\tau) = 0$]. We note that even though the receiver may have perfect knowledge of the noise, the overall effect of the amplitude noise increases the error probability. This is because input powers smaller than the average power [$\mathcal{A}(\tau) < \langle \hat{n} \rangle_0$] increase the errors more than the reduction of error for larger powers [$\mathcal{A}(\tau) > \langle \hat{n} \rangle_0$]. The dashed black and gray lines show the error for an adaptive non-Gaussian measurement, and an ideal heterodyne measurement with no noise, respectively. By comparing the error of the non-Gaussian measurement with perfect correction (black points) to the black dashed reference line, we observe that non-Gaussian measurements are more sensitive to amplitude noise

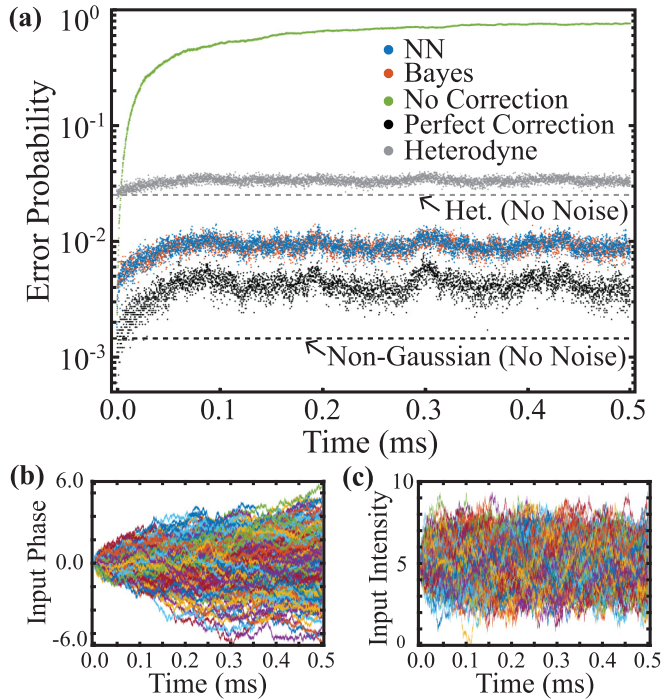


FIG. 3. Probability of error as a function of time. (a) Error probability as a function of time for $\langle \hat{n} \rangle_0 = 5.0$ when both phase (b) and intensity (c) noise are applied and tracked. Here the noise parameters are $\gamma = 25$ kHz, $\Sigma_\infty^2 = 1.5$, and $\Delta\nu = 2$ kHz. Blue (orange) points show the error for the NN- (Bayes-) based estimator. Green and black points show the error with no correction and perfect correction, respectively. Gray points show the effective QNL of a perfectly corrected heterodyne measurement. Black and gray dashed lines show the error for a non-Gaussian and heterodyne receiver, respectively, with no noise.

than a heterodyne measurement (gray points vs gray dashed line), even when they are perfectly corrected. We observe that the NN-based tracking method performs equivalently to the Bayesian method, and both can allow the receiver to maintain an error probability significantly below the QNL. This result demonstrates the capabilities of a NN for efficient and reliable noise tracking for non-Gaussian receivers for state discrimination.

We study the robustness of the NN-based method in scenarios with different noise strengths and bandwidths in the phase and amplitude. For these studies, we use the heterodyne measurement with perfect phase tracking as the limit for conventional strategies, serving as the effective QNL when the same noise is applied to both receivers. In this section, we study (a) the error probability as a function of phase noise with fixed amplitude noise, and (b) the error probability when the amplitude noise levels are varied with phase noise with a fixed bandwidth.

A. Phase noise with different bandwidths

We study the performance of the noise tracking method based on the NN estimator as a function of the phase noise bandwidth $\Delta\nu$ for fixed values of amplitude noise Σ_∞^2 and γ . We compare these results to the tracking method based

on a Bayesian estimator, as well as a perfectly corrected non-Gaussian measurement. We use different amplitude noise parameters for different values of $\langle \hat{n} \rangle_0$ such that the relative amplitude noise strength $\langle \hat{n} \rangle_0 / \Sigma_\infty$ is constant. For an average intensity of $\langle \hat{n} \rangle_0 = 2, 5$, and 10 we simulate 250, 250, and 500 different realizations of the noise, respectively. The simulations are run for 2×10^3 time bins of $N = 10$ experiments each, giving a total of 2×10^4 individual experiments per noise realization. We calculate the average error across all realizations for all time bins.

Figure 4 shows the average error probability as a function of the bandwidth $\Delta\nu$ for intensities $\langle \hat{n} \rangle_0 = 2, 5$, and 10 . Figures 4(a)–4(c) have no amplitude noise ($\gamma = 0$, $\Sigma_\infty^2 = 0$), and 4(d)–4(f) have $\gamma = 2$ kHz, and $\Sigma_\infty^2 = \{0.25, 1.5, 6.0\}$, respectively, corresponding to the relative strength of $\langle \hat{n} \rangle_0 / \Sigma_\infty \approx 0.25$. The performance of the noise tracking method based on the NN estimator (blue) is equivalent to the Bayesian-based method (orange) while being computationally efficient to implement. The purple and gray dashed lines show the average error of the non-Gaussian and ideal heterodyne receivers with perfect parameter tracking, respectively.

We observe that for all the investigated average input intensities $\langle \hat{n} \rangle_0$, the NN-based method performs as well as the Bayesian method both with and without amplitude noise. The NN-based method can enable the non-Gaussian receiver to surpass the QNL up to a phase noise bandwidth of $\Delta\nu \approx 15$ kHz, even in the presence of significant amplitude noise. We note that the situation in Figs. 4(a)–4(c) with no amplitude noise is equivalent to the single parameter problem of phase tracking for non-Gaussian receivers as demonstrated in [31]. For large phase noise bandwidths for $\langle \hat{n} \rangle_0 = 5$, and 10 , we observe that a NN estimator can perform slightly better than a Bayesian estimator. We believe this is due to the relatively small number of samples ($N = 10$) from which estimates are made. In this regime with a few samples for estimation, there may be estimators which perform better than the Bayesian estimator, which is asymptotically optimal in the limit of many samples. Another potential cause of this effect is that in the training process of the NN, the relative weight between error in phase estimates and error in mean photon number estimates can be adjusted. This freedom may allow for fine-tuning of the overall training error to allow for a slightly better overall performance, in terms of error probability, for specific channel models.

B. Amplitude noise with different bandwidths

To investigate the effect of the amplitude noise bandwidth γ , we fix the long-time variance Σ_∞^2 and the phase noise bandwidth $\Delta\nu$. This allows for studying the performance of the NN-based method when the amplitude noise bandwidth γ ranges from much smaller to much larger than the bandwidth for parameter estimation $1/N\Delta T$. Figure 5 shows the average probability of error for different amplitude noise bandwidths γ without and with phase noise of bandwidth $\Delta\nu = 5$ kHz, for $\langle \hat{n} \rangle_0 = \{2, 5, 10\}$ with $\Sigma_\infty^2 = \{0.25, 1.5, 6.0\}$, respectively. Blue (orange) lines show the error rates for the NN- (Bayesian-) based tracking method. Purple and gray dashed lines show the error probability for a non-Gaussian and heterodyne measurement with perfect

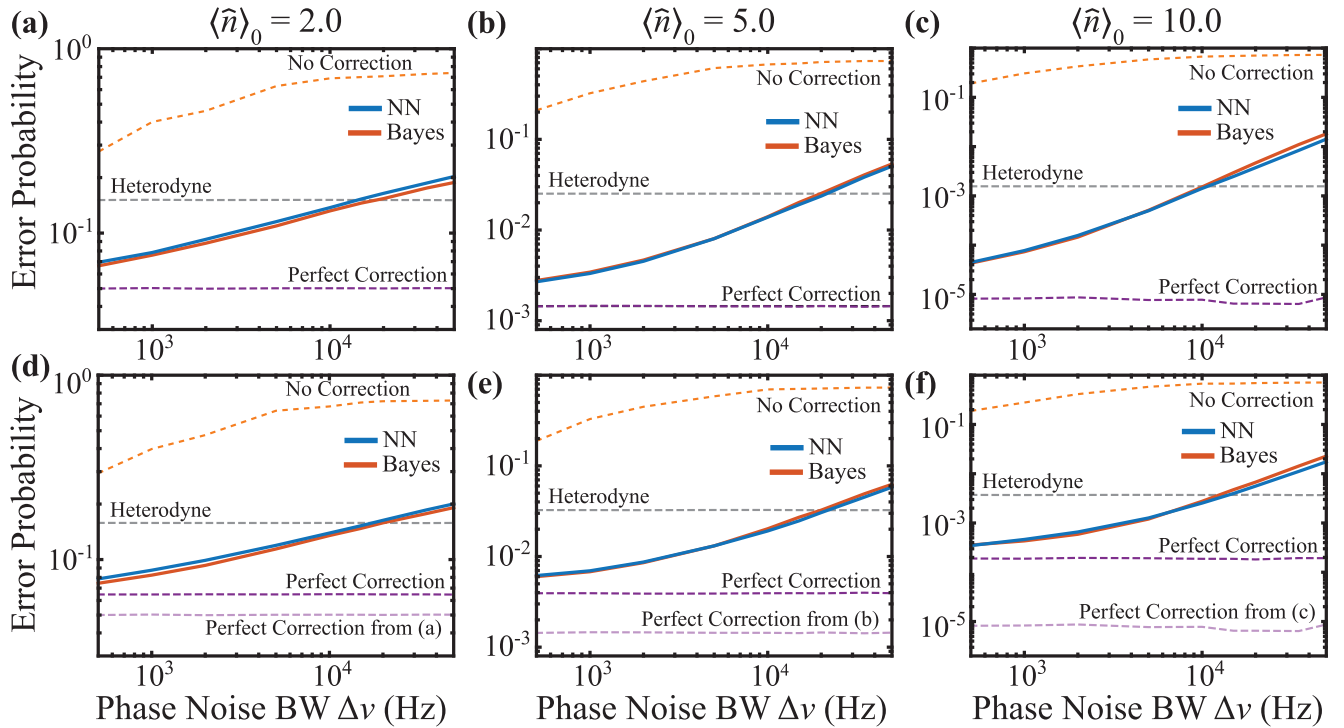


FIG. 4. Phase noise tracking. Error probability as a function of phase noise bandwidth (BW) $\Delta\nu$ without (a)–(c) and with (d)–(f) amplitude noise of $\gamma = 25$ kHz and $\Sigma_\infty^2 = 0.25, 1.5$, and 6.0 for average intensities $\langle\hat{n}\rangle_0 = 2, 5$, and 10 , respectively. Blue and orange lines show the performance of the noise tracking methods based on NN and Bayesian estimators, respectively. The orange dashed line shows the error probability for a non-Gaussian receiver with no correction. The purple and gray dashed lines show the error probability for a non-Gaussian and heterodyne measurement with perfect correction, respectively.

correction, respectively. We find that the NN-based method performs closely to the Bayesian-based method, and enables the receiver to achieve sub-QNL error rates across a broad range of amplitude noise bandwidths even in the presence of phase noise.

We note that for intensity $\langle\hat{n}\rangle_0 = 2$ in Fig. 5(a), the error for both the noise tracking methods is below the perfectly corrected non-Gaussian measurement when $\Delta\nu = 0$. At low input powers, strategies that optimize the LO intensity ($|\beta|^2 > |\alpha|^2$) yield lower error probabilities than when $|\beta|^2 = |\alpha|^2$

[28]. Due to the small number of samples ($N \times L$) used for estimation, the NN and Bayesian estimators have a bias in $\hat{A}_{B,NN}$, such that $\mathcal{B}(\tau) > \mathcal{A}(\tau)$. The effect of this bias in the intensity estimates $\hat{A}_{B,NN}$ is that the corrected measurement unintentionally approximates an optimized strategy [28]. This effect results in error probabilities of the corrected receiver with both NN- and Bayesian-based methods that are below the error of a perfectly corrected nulling receiver where $\mathcal{B}(\tau) = \mathcal{A}(\tau)$, which is due to the bias of the estimators from finite sampling. Further investigation is needed to determine the

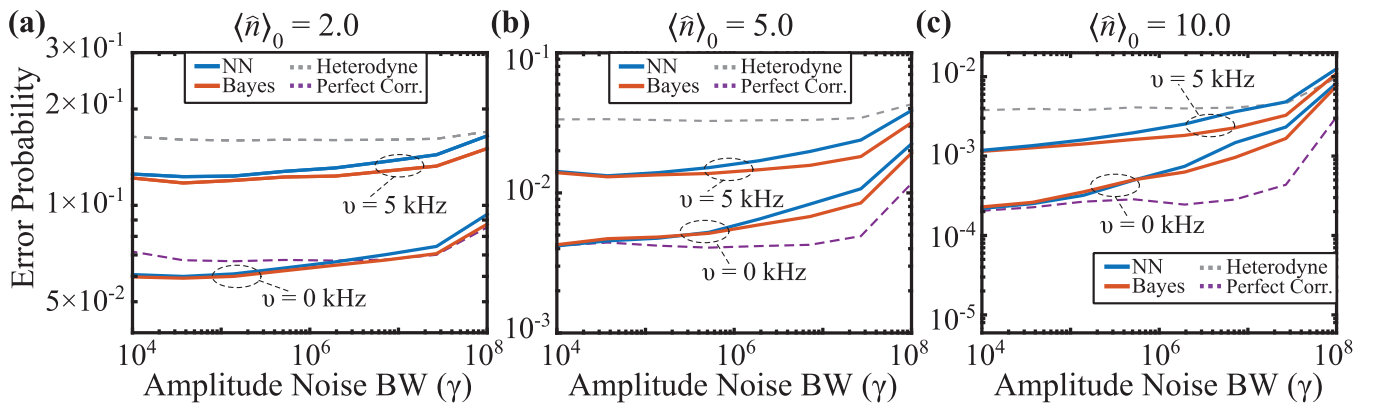


FIG. 5. Amplitude noise tracking. Error probability as a function of amplitude noise bandwidth (BW) γ for $\langle\hat{n}\rangle_0 = \{2, 5, 10\}$ without and with phase noise with bandwidth $\Delta\nu = 5$ kHz. Purple and gray dashed lines show the error for a non-Gaussian and heterodyne measurement with perfect correction, respectively. Beyond $\gamma \approx 10^7$ Hz, the amplitude noise is effectively random across the N experiments.

capabilities of NN-based noise tracking for optimized non-Gaussian receivers [28].

The performance of the non-Gaussian receiver also depends on the long-time variance Σ_∞^2 of the amplitude noise. In our main results, Σ_∞^2 was set to represent a “worst-case” scenario of $\approx 25\%$ relative amplitude noise (see Fig. 3). Appendix C describes our study of noise tracking of amplitude noise with different long-time variance Σ_∞^2 . In our findings, we observe that in the absence of phase noise, both the NN- and Bayesian-based tracking methods enable the receiver to perform below the QNL, and close to the performance of perfect noise correction. In the presence of phase noise with bandwidth $\Delta\nu = 5$ kHz, the sub-QNL performance of the receiver is maintained, and the effect of increasing Σ_∞^2 is small compared to the effects of increasing phase or amplitude bandwidths.

IV. DISCUSSION

The numerical studies in this work show that methods for channel-noise tracking based on NN estimators are able to accurately track dynamic phase and amplitude noise to allow an adaptive non-Gaussian measurement to maintain performance below the QNL. We note that in the asymptotic limit of many samples available for parameter estimation for noise tracking, a Bayesian estimator can achieve minimal mean-square error [63]. However, when noise tracking and correction need to be realized in real time to reduce errors in the state discrimination measurement and generate reliable data for parameter estimation, there is always a limited number of samples from which estimates are made. In these situations, there is a trade-off between estimation precision and noise tracking bandwidth. Other estimators, such as a NN, may balance these two parameters better than a Bayesian estimator for increased precision with finite samples. This property can enable efficient methods for high-dimensional parameter tracking of complex dynamic channel noise.

The computational efficiency of the NN estimator is rooted in the small number of multiplications required to calculate a single estimate, the limited memory requirements, and in the fact that the NN method does not explicitly depend on the value of N or the number of adaptive steps L , as opposed to the Bayesian approach. For example, the NN-based estimator in this work requires ≈ 5500 multiplications. On the other hand, a Bayesian estimator using likelihood functions which are discretized into a 100×100 grid would require $100^2 \times N \times L = 10^6$ multiplications, which may not be compatible with devices such as field-programmable gate arrays [26,31]. While there may be methods to reduce this computational cost, the Bayesian estimator also would require storage in memory of the full photon counting likelihood functions, putting stringent requirements on the device memory. For example, the 100×100 grid for the Bayesian estimator with 16 bit precision would require 800 kB of memory simply to store the likelihood functions. Moreover, to extend the noise tracking method for estimation of three noise parameters, a NN would simply require a single added output and proper retraining, while a Bayesian estimator would require possibly 10^8 multiplications and 80 MB of memory.

The robustness and versatility of the NN-based noise tracking method described here, shows that NN-based methods can be practical and very useful tools for non-Gaussian receivers. In addition, other machine learning techniques, such as reinforcement learning [75,76], could provide further benefits to these nonconventional measurements when the best detection strategy may be unknown or infeasible to calculate. We anticipate that neural networks and machine learning will have a great benefit for non-Gaussian measurements, just as these techniques have proven worthwhile for conventional measurement strategies [54–61].

V. CONCLUSION

We investigate the use of a NN as a computationally efficient multiparameter estimator of dynamic channel noise, enabling robust noise tracking for adaptive non-Gaussian measurements for coherent state discrimination. We study the NN-based tracking method for simultaneous amplitude and phase noise and find that the NN estimator can perform as well as a more complex Bayesian estimator. This performance is observed across a broad range of noise strengths and bandwidths for different average powers of the input coherent states. The non-Gaussian receiver used in this study can have broad applications in classical [26,27] and quantum communication [77,78] due to its ability to attain sensitivities beyond the QNL. Moreover, the proposed method for noise tracking uses only the data collected during the state discrimination measurement without requiring extra resources such as strong reference pulses. This makes the receiver and the proposed method for noise tracking well suited for energy efficient low power communications. Thus, NN-based methods are ideal candidates for real-time tracking of multiple sources of channel noise for non-Gaussian receivers, allowing them to maintain their sub-QNL sensitivity in the presence of complex dynamic channel noise.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) under Grants No. PHY-1653670, No. PHY-1521016, and No. PHY-1630114. See Ref. [79] for the source code.

APPENDIX A: NEURAL NETWORK ESTIMATOR TRAINING

To generate training data, we use Monte Carlo simulations of the adaptive non-Gaussian measurement described in Sec. II A [26]. For a single training data element, the strategy is simulated using a randomly chosen intensity for the input state $\mathcal{A}$ and LO intensity $\mathcal{B}$, both sampled from a uniform distribution $\mathcal{U}(0.05, 25.0)$. A constant value for the phase noise ϕ is then sampled from a Gaussian distribution $\mathcal{N}(0, \sigma = 0.25)$. In addition, we randomly sample the number of experiments N that comprise each sample from a uniform distribution $\mathcal{U}(2, 200)$. This random sampling of the three parameters enables the NN to be used for different values of N depending on the noise characteristics. During the training of the NN, we use true values of the input parameters ($\phi, \mathcal{A}$)

as the target values. This procedure using random sampling of multiple input noise parameters ensures sufficient sampling of the input parameter space, enabling training of a robust NN estimator.

To train the NN we use the Tensorflow framework [65] in Python. Table I summarizes the relevant NN and training parameters. We use the RMSprop optimizer [80] with a weighted mean-squared-error cost function. The weight w_i of each training sample is given by

$$w_i = e^{-(\mathcal{A}_i - \mathcal{B}_i)^2/2} + 0.1, \quad (\text{A1})$$

where $\mathcal{A}_i$ and $\mathcal{B}_i$ are the intensity of the input states and LO of the i th sample, respectively. This allows the NN to accurately estimate the parameters when the LO and input intensities are close to each other, as one would expect in practice, while also being somewhat robust to large amplitude fluctuations.

APPENDIX B: KALMAN FILTER

We implement Kalman filtering [74,81] of both the phase estimate and intensity estimate in order to reduce the uncertainty in the applied corrections to the adaptive non-Gaussian measurement. For the phase estimates, the predicted mean value $\hat{y}_\phi$ and variance $\hat{\sigma}_\phi^2$ are

$$\begin{aligned} \hat{y}_\phi &= 0, \\ \hat{\sigma}_\phi^2 &= \sigma_\phi^2 + N\sigma_1^2, \end{aligned} \quad (\text{B1})$$

where $\hat{y}_\phi$ represents the predicted average value for the phase, σ_ϕ^2 the variance of the current prior probability distribution for the phase, $\sigma_1^2 = 2\pi\Delta\nu\Delta T$, and $\hat{\sigma}_\phi^2$ the predicted phase variance. The filtered estimate $\hat{\phi}$ is then obtained from the raw estimate $\hat{\phi}_{NN}$ and (B1) by

$$\begin{aligned} \hat{\phi} &= K_\phi \hat{\phi}_{NN} + (1 - K_\phi) \hat{y}_\phi, \\ \sigma_\phi^2 &= (1 - K_\phi) \hat{\sigma}_\phi^2, \\ K_\phi &= \frac{\hat{\sigma}_\phi^2}{\hat{\sigma}_\phi^2 + \sigma_{\phi,NN}^2}, \end{aligned} \quad (\text{B2})$$

where K_ϕ is the Kalman gain for the phase estimate, $\sigma_{\phi,NN}^2$ is the variance of the NN phase estimate, and σ_ϕ^2 is the updated variance of the filtered phase estimate.

Similarly, the equations for the predicted input intensity mean value $\hat{y}_\mathcal{A}$ and variance $\hat{\sigma}_\mathcal{A}^2$ are given by

$$\begin{aligned} \hat{y}_\mathcal{A} &= (1 - \gamma\Delta T)^N \mathcal{B}_N + \gamma\langle\hat{n}\rangle_0 \Delta T \sum_{k=0}^{N-1} (1 - \gamma\Delta T)^k, \\ \hat{\sigma}_\mathcal{A}^2 &= (1 - \gamma\Delta T)^{2N} \sigma_\mathcal{A}^2 + \Sigma^2 \Delta T \sum_{k=0}^{N-1} (1 - \gamma\Delta T)^{2k}, \end{aligned} \quad (\text{B3})$$

where γ is the amplitude noise bandwidth, $\langle\hat{n}\rangle_0$ is the average input intensity equal to $\mathcal{A}(0)$, and $\mathcal{B}_N$ is the LO intensity for the previous N experiments. These equations result from propagating the mean and variance of Eq. (1) for N time steps of duration ΔT .

The filtered intensity estimate $\hat{\mathcal{A}}$ and variance $\sigma_\mathcal{A}^2$ are then obtained from the raw estimate $\hat{\mathcal{A}}_{NN}$ and (B3) by

$$\begin{aligned} \hat{\mathcal{A}} &= K_\mathcal{A} \hat{\mathcal{A}}_{NN} + (1 - K_\mathcal{A}) \hat{y}_\mathcal{A}, \\ \sigma_\mathcal{A}^2 &= (1 - K_\mathcal{A}) \hat{\sigma}_\mathcal{A}^2, \\ K_\mathcal{A} &= \frac{\hat{\sigma}_\mathcal{A}^2}{\hat{\sigma}_\mathcal{A}^2 + \sigma_{\mathcal{A},NN}^2}, \end{aligned} \quad (\text{B4})$$

where $K_\mathcal{A}$ is the Kalman gain for the intensity estimate and $\sigma_{\mathcal{A},NN}^2$ is the variance of the NN phase estimate. The initial ($\tau = 0$) variances for both phase and intensity are set to zero.

In order to empirically obtain the variances $\sigma_{\phi,NN}^2$ and $\sigma_{\mathcal{A},NN}^2$ we use Monte Carlo simulations of the experiment and the NN estimator without the Kalman filter. For all average intensities, we fix the input intensity to $\mathcal{A}(t) = \{2, 5, 10\}$ and phase noise to zero, and calculate the variance of 10^6 estimates. The variance is calculated for a range of the number of experiments per estimation N . We fit the variance as a function of N to a power-law function so that the filter may be used for different values of N . For $N = 10$, as used in the simulations, $\sigma_{\mathcal{A},NN}^2 = \{3.37, 5.63, 5.35\} \times 10^{-1}$ and $\sigma_{\phi,NN}^2 = \{4.77, 3.61, 2.66\} \times 10^{-3}$ for $\langle\hat{n}\rangle_0 = \{2, 5, 10\}$, respectively. As discussed in Sec. IV B, the NN estimator is not necessarily unbiased across the range of parameters it is trained on due to the small sample size of only $N \times L$ as well as imperfections in the NN training process.

Employing a Kalman filter allows for construction of the full NN-based parameter tracking method. Algorithm 1 shows the pseudocode for running the NN-based method including the filtering steps. For every time step τ , a single state discrimination measurement is completed, which yields detections $\{d_j\}_L$ and relative phases $\{\Delta_j\}_L$ which populate the detection matrix D , as described in Sec. II B. Every N time steps, i.e., every N measurements, the NN is evaluated to provide raw estimates $\hat{\mathcal{A}}_{NN}, \hat{\phi}_{NN}$ of the intensity and phase offset within the previous N measurements. These raw estimates are then passed through the Kalman filter, which returns the current, filtered estimates for the intensity $\hat{\mathcal{A}}(\tau)$ and phase offset $\hat{\phi}$. These current estimates are then used to update the LO parameters $\mathcal{B}(\tau), \delta(\tau)$ for the next N state discrimination measurements.

APPENDIX C: DIFFERENT MAGNITUDES OF AMPLITUDE NOISE

We further investigate the performance of the NN estimator when varying the long-time strength of the amplitude noise Σ_∞^2 . Figure 6 shows the error probability of the receiver for $\langle\hat{n}\rangle_0 = 5$ across a range of Σ_∞^2 for fixed $\gamma = 25$ kHz without and with phase noise of bandwidth $\Delta\nu = 5$ kHz. We find that the NN-based tracking method performs similar to the one based on the Bayesian estimator, and enables an error probability below that of the ideal heterodyne measurement. We have observed in other studies that the behavior for average input intensities $\langle\hat{n}\rangle_0 = 2$ and 10 is similar to the one for $\langle\hat{n}\rangle_0 = 5$ in Fig. 6.

Algorithm 1. NN Parameter Tracking Algorithm

```

function KALMANFILTER( $\hat{A}_{NN}, \hat{\phi}_{NN}$ )
  Predict:  $\hat{y}_\phi, \hat{\sigma}_\phi^2, \hat{y}_A, \hat{\sigma}_A^2$                                 ▷ Eqs. (B1), (B3)
   $\hat{\phi} \leftarrow K_\phi \hat{\phi}_{NN} + (1 - K_\phi) \hat{y}_\phi$                         ▷ Eq. (B2)
   $\hat{A}(\tau) \leftarrow K_A \hat{A}_{NN} + (1 - K_A) \hat{y}_A$                   ▷ Eq. (B4)
  Update:  $K_\phi, \sigma_\phi^2, K_A, \sigma_A^2$                             ▷ Eqs. (B2), (B4)
  return  $\hat{\phi}, \hat{A}(\tau)$ 
end function

Initial
 $B(0) \leftarrow \hat{A}(0)$ 
 $\delta(0) \leftarrow 0$ 
 $\tau \leftarrow 0$                                 ▷ Time in increments of symbol time
 $n \leftarrow 0$                                 ▷ Number of measurements performed

loop
   $\{d_j\}_L, \{\Delta_j\}_L \leftarrow$  Single discrimination measurement
  Add  $\{d_j\}_L, \{\Delta_j\}_L$  to detection matrix D
   $\tau \leftarrow \tau + \Delta T$ 
   $n \leftarrow n + 1$ 
  if  $n = N$  then                                ▷ Update LO every  $N$  measurements
     $\hat{A}_{NN}, \hat{\phi}_{NN} \leftarrow$  Evaluate NN with inputs D,  $B(\tau)$ 
    Reset D to zeros
     $\hat{\phi}, \hat{A}(\tau) \leftarrow$  KALMANFILTER( $\hat{A}_{NN}, \hat{\phi}_{NN}$ )
     $B(\tau) \leftarrow \hat{A}(\tau)$                                 ▷ Correct LO intensity
     $\delta(\tau) \leftarrow \delta(\tau) + \hat{\phi}$                             ▷ Add estimate to phase correction
     $n \leftarrow 0$                                           ▷ Reset measurement counter
  else                                            ▷ Don't update LO
     $B(\tau) \leftarrow B(\tau - \Delta T)$ 
     $\delta(\tau) \leftarrow \delta(\tau - \Delta T)$ 
  end if
end loop

```

APPENDIX D: ESTIMATION TIME-BANDWIDTH TRADE-OFF

The overall performance of the phase tracking also depends on the number of experiments N used to obtain a single estimate. For this study, we fixed $N=10$ for all simulations

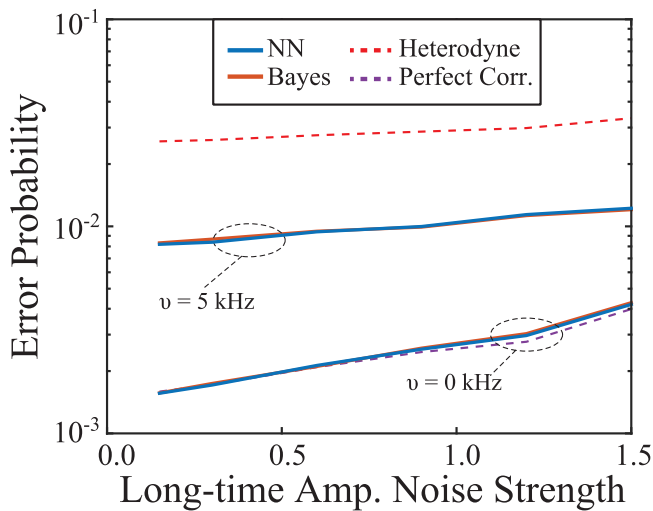


FIG. 6. Error probability as a function of the long-time variance Σ_∞^2 of the amplitude noise for $\langle \hat{n} \rangle_0 = 5$, $\gamma = 25$ kHz without and with phase noise of bandwidth $\Delta\nu = 5$ kHz.

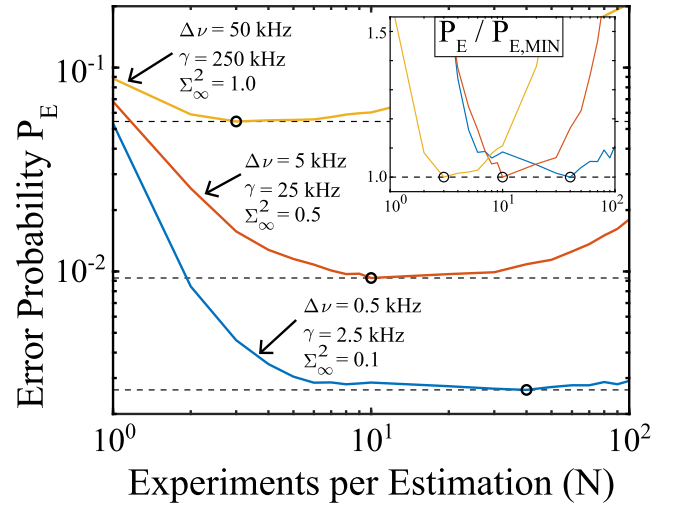


FIG. 7. Error probability as a function of N for three different noise strengths: low noise (blue), moderate noise (orange), and high noise (yellow). The black circles show the minimum error probability $P_{E,MIN}$ and corresponding optimal value for N for each noise regime. The inset shows the error probability normalized by the minimum error given by the black circles in the main figure.

to demonstrate the versatility of the NN estimator. Although, realistic implementations may use a particular channel with specific noise characteristics. In this scenario, the value of N can be fine-tuned to optimize the performance of the phase tracking method to balance the estimation bandwidth (smaller N) and estimation accuracy (larger N). The optimization aims to find a value of N such that the overall error probability is minimized when implementing the noise tracking method. The Kalman filtering attempts to balance the effects of the estimation variance and the noise variance over N measurements in an optimal way through the Kalman gain K . However, there is still a trade-off between these two variances and a value of N which achieves minimal P_E for specific channel-noise conditions.

Figure 7 shows the overall error probability as a function of N when implementing the NN estimator with filtering for three different sets of channel-noise parameters for $\langle \hat{n} \rangle_0 = 5.0$. The blue line corresponds to $\Delta\nu = 0.5$ kHz, $\gamma = 2.5$ kHz, and $\Sigma_\infty^2 = 0.1$. The orange line corresponds to $\Delta\nu = 5$ kHz, $\gamma = 25$ kHz, and $\Sigma_\infty^2 = 0.5$. The yellow line corresponds to $\Delta\nu = 50$ kHz, $\gamma = 250$ kHz, and $\Sigma_\infty^2 = 1.0$. For small noise bandwidth and strength (blue), the optimal value of N (black circles) is approximately $N = 40$, but decreases to $N = 10$ and $N = 3$ as the noise bandwidth and strength increase (orange and yellow). The inset shows the error probability normalized by the minimum for each noise strength for clarity. These optimal values of N for specific channel-noise parameters represent the optimal balance between estimation uncertainty and accumulated uncertainty from the channel noise. Thus, for a channel with known noise characteristics, an optimal value of N can be found. In the studies presented in the main text, we fixed $N=10$, since we found that this value allows the receiver to be versatile and operate well across a wide range of noise bandwidths. In the inset, $N = 10$ is

optimal for moderate noise levels. For small and large noise levels, the error at $N = 10$ is only slightly higher compared to their respective minimums. Thus, for a specific well-known

channel an optimal N can be implemented, but for a robust and versatile implementation a different value of N may be beneficial.

-
- [1] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, Classical Capacity of the Lossy Bosonic Channel: The Exact Solution, *Phys. Rev. Lett.* **92**, 027902 (2004).
 - [2] G. Li, Recent advances in coherent optical communication, *Adv. Opt. Photonics* **1**, 279 (2009).
 - [3] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, Coherent detection in optical fiber systems, *Opt. Express* **16**, 753 (2008).
 - [4] K. Kikuchi and S. Tsukamoto, Evaluation of sensitivity of the digital coherent receiver, *J. Lightwave Technol.* **26**, 1817 (2008).
 - [5] J. M. Arrazola and N. Lütkenhaus, Quantum communication with coherent states and linear optics, *Phys. Rev. A* **90**, 042335 (2014).
 - [6] S. Ghorai, P. Grangier, E. Diamanti, and A. Leverrier, Asymptotic Security of Continuous-Variable Quantum Key Distribution with a Discrete Modulation, *Phys. Rev. X* **9**, 021059 (2019).
 - [7] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Fundamental limits of repeaterless quantum communications, *Nat. Commun.* **8**, 15043 (2017).
 - [8] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, *Rev. Mod. Phys.* **74**, 145 (2002).
 - [9] N. Gisin and R. Thew, Quantum communication, *Nat. Photonics* **1**, 165 (2007).
 - [10] K. Kikuchi, Fundamentals of coherent optical fiber communications, *J. Lightwave Technol.* **34**, 157 (2016).
 - [11] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, Experimental demonstration of long-distance continuous-variable quantum key distribution, *Nat. Photonics* **7**, 378 (2103).
 - [12] A. G. Armada and M. Calvo, Phase noise and sub-carrier spacing effects on the performance of an OFDM communication system, *IEEE Commun. Lett.* **2**, 11 (1998).
 - [13] A. Marie and R. Alléaume, Self-coherent phase reference sharing for continuous-variable quantum key distribution, *Phys. Rev. A* **95**, 012316 (2017).
 - [14] B. Qi, P. Lougovski, R. Pooser, W. Grice, and M. Bobrek, Generating the Local Oscillator ‘Locally’ in Continuous-Variable Quantum Key Distribution Based on Coherent Detection, *Phys. Rev. X* **5**, 041009 (2015).
 - [15] D. B. S. Soh, C. Brif, P. J. Coles, N. Lütkenhaus, R. M. Camacho, J. Urayama, and M. Sarovar, Self-Referenced Continuous-Variable Quantum Key Distribution Protocol, *Phys. Rev. X* **5**, 041010 (2015).
 - [16] T. Wang, P. Huang, S. Wang, and G. Zeng, Carrier-phase estimation for simultaneous quantum key distribution and classical communication using a real local oscillator, *Phys. Rev. A* **99**, 022318 (2019).
 - [17] E. Ip and J. M. Kahn, Feedforward carrier recovery for coherent optical communications, *J. Lightwave Technol.* **25**, 2675 (2007).
 - [18] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation, *J. Lightwave Technol.* **24**, 12 (2006).
 - [19] M. Bina, A. Allevi, M. Bondani, and S. Olivares, Phase-reference monitoring in coherent-state discrimination assisted by a photon-number resolving detector, *Sci. Rep.* **6**, 26025 (2016).
 - [20] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Mathematics in Science and Engineering Vol. 123 (Academic, New York, 1976).
 - [21] V. Giovannetti, R. García-Patrón, N. J. Cerf and A. S. Holevo, Ultimate classical communication rates of quantum optical channels, *Nat. Photonics* **8**, 796 (2014).
 - [22] K. Banaszek, L. Kunz, M. Jachura, and M. Jarzyna, Quantum limits in optical communications, *J. Lightwave Technol.* **38**, 2741 (2020).
 - [23] C. R. Müller, M. A. Usuga, C. Wittmann, M. Takeoka, C. Marquardt, U. L. Andersen, and G. Leuchs, Quadrature phase shift keying coherent state discrimination via a hybrid receiver, *New J. Phys.* **14**, 083009 (2012).
 - [24] C. R. Müller and C. Marquardt, A robust quantum receiver for phase shift keyed signals, *New J. Phys.* **17**, 032003 (2015).
 - [25] F. E. Becerra, J. Fan, G. Baumgartner, J. Goldhar, J. T. Kosloski, and A. Migdall, Experimental demonstration of a receiver beating the standard quantum limit for multiple nonorthogonal state discrimination, *Nat. Photonics* **7**, 147 (2013).
 - [26] F. E. Becerra, J. Fan, and A. Migdall, Photon number resolution enables quantum receiver for realistic coherent optical communications, *Nat. Photonics* **9**, 48 (2015).
 - [27] J. Lee, S.-W. Ji, J. Park, and H. Nha, Gaussian benchmark for optical communication aiming towards ultimate capacity, *Phys. Rev. A* **93**, 050302(R) (2016).
 - [28] A. R. Ferdinand, M. T. DiMario, and F. E. Becerra, Multi-state discrimination below the quantum noise limit at the single-photon level, *npj Quantum Inf.* **3**, 43 (2017).
 - [29] S. Izumi, M. Takeoka, M. Fujiwara, N. D. Pozza, A. Assalini, K. Ema, and M. Sasaki, Displacement receiver for phase-shift-keyed coherent states, *Phys. Rev. A* **86**, 042328 (2012).
 - [30] S. Izumi, J. S. Neergaard-Nielsen, S. Miki, H. Terai, and U. L. Andersen, Experimental Demonstration of a Quantum Receiver Beating the Standard Quantum Limit at Telecom Wavelength, *Phys. Rev. Appl.* **13**, 054015 (2020).
 - [31] M. T. DiMario and F. E. Becerra, Phase tracking for sub-shot-noise-limited receivers, *Phys. Rev. Research* **2**, 023384 (2020).
 - [32] M. Bina, A. Allevi, M. Bondani, and S. Olivares, Homodyne-like detection for coherent state-discrimination in the presence of phase noise, *Opt. Express* **25**, 10685 (2017).
 - [33] J. L. Habif, A. Jagannathan, S. Gartenstein, P. Amory, and S. Guha, Quantum-limited discrimination of laser light and thermal light, [arXiv:1912.06718v1](https://arxiv.org/abs/1912.06718v1).
 - [34] R. Yuan, M. Zhao, S. Han, and J. Cheng, Kennedy receiver using threshold detection and optimized displacement under thermal noise, *IEEE Commun. Lett.* **24**, 1313 (2020).
 - [35] M. T. DiMario, L. Kunz, K. Banaszek, and F. E. Becerra, Optimized communication strategies with binary coherent states over phase noise channels, *npj Quantum Inf.* **5**, 65 (2019).

- [36] M. G. Genoni, S. Olivares, and M. G. A. Paris, Optical Phase Estimation in the Presence of Phase Diffusion, *Phys. Rev. Lett.* **106**, 153603 (2011).
- [37] M. G. Genoni, S. Olivares, D. Brivio, S. Cialdi, D. Cipriani, A. Santamato, S. Vezzoli, and M. G. A. Paris, Optical interferometry in the presence of large phase diffusion, *Phys. Rev. A* **85**, 043817 (2012).
- [38] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [39] J. Walln fer, A. A. Melnikov, W. D r, and H. J. Briegel, Machine learning for long-distance quantum communication, *PRX Quantum* **1**, 010301 (2020).
- [40] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, An optical communication's perspective on machine learning and its applications, *J. Lightwave Technol.* **37**, 493 (2019).
- [41] J. Mata, I. de Miguel, R. J. Dur n, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, Artificial intelligence (AI) methods in optical networks: A comprehensive survey, *Opt. Switch. Netw.* **28**, 43 (2018).
- [42] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, Artificial neural networks-based machine learning for wireless networks: A tutorial, *IEEE Commun. Surveys Tutorials* **21**, 3039 (2019).
- [43] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks* **61**, 85 (2015).
- [44] V. Dunjko and H. J. Briegel, Machine learning and artificial intelligence in the quantum domain: A review of recent progress, *Rep. Prog. Phys.* **81**, 074001 (2018).
- [45] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, Active learning machine learns to create new quantum experiments, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1221 (2018).
- [46] A. Hentschel and B. C. Sanders, Machine Learning for Precise Quantum Measurement, *Phys. Rev. Lett.* **104**, 063603 (2010).
- [47] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, Experimental Phase Estimation Enhanced by Machine Learning, *Phys. Rev. Appl.* **10**, 044033 (2018).
- [48] L. J. Fiderer, J. Schuff, and D. Braun, Neural-network heuristics for adaptive Bayesian quantum estimation, *arXiv:2003.02183*.
- [49] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, and M. Barbieri, Calibration of Quantum Sensors by Neural Networks, *Phys. Rev. Lett.* **123**, 230502 (2019).
- [50] T. Giordani, A. Suprano, E. Polino, F. Acanfora, L. Innocenti, A. Ferraro, M. Paternostro, N. Spagnolo, and F. Sciarrino, Machine Learning-Based Classification of Vector Vortex Beams, *Phys. Rev. Lett.* **124**, 160401 (2020).
- [51] S. Lohani, E. M. Knutson, M. O'Donnell, S. D. Huver, and R. T. Glasser, On the use of deep neural networks in optical communications, *Appl. Opt.* **57**, 4180 (2018).
- [52] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, Quantum optical neural networks, *npj Quantum Inf.* **5**, 60 (2019).
- [53] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, *Nat. Commun.* **11**, 808 (2020).
- [54] J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals, *J. Lightwave Technol.* **35**, 868 (2017).
- [55] X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, Applications of artificial neural networks in optical performance monitoring, *J. Lightwave Technol.* **27**, 3580 (2009).
- [56] D. Zibar, M. Piels, R. Jones, and C. G. Sch effer, Machine learning techniques in optical communication, *J. Lightwave Technol.* **34**, 1442 (2016).
- [57] S. Lohani, E. M. Knutson, and R. T. Glasser, Generative machine learning for robust free-space communication, *Commun. Phys.* **3**, 177 (2020).
- [58] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. B low, D. Lavery, P. Bayvel, and L. Schmalen, End-to-end deep learning of optical fiber communications, *J. Lightwave Technol.* **36**, 4843 (2018).
- [59] D. Zibar, L. H. H. de Carvalho, M. Piels, A. Doberstein, J. Diniz, B. Nebendahl, C. Franciscangelis, J. Estaran, H. Haisch, N. G. Gonzalez, J. C. R. F. de Oliveira, and I. T. Monroy, Application of machine learning techniques for amplitude and phase noise characterization, *J. Lightwave Technol.* **33**, 1333 (2015).
- [60] F. N. Khan, K. Zhong, X. Zhou, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. T. Lau, Joint OSNR monitoring and modulation format identification in digital coherent receivers using deep neural networks, *Opt. Express* **25**, 17767 (2017).
- [61] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, Modulation format recognition and OSNR estimation using CNN-based deep learning, *IEEE Photonics Technol. Lett.* **29**, 1667 (2017).
- [62] We note that in conventional optical communications, the use of amplification prior to detection can be used to reduce the probability of error. In those situations, the use of non-Gaussian receivers can further reduce the error rates far beyond what could be achieved by heterodyne measurements. For example, given an input pulse power corresponding to $\langle \hat{n} \rangle_0 = 5.0$, the error reduction for a non-Gaussian receiver ($P_{E,NG}$) compared to the heterodyne receiver at the QNL equals a factor of $QNL/P_{E,NG} = 17$. At this initial power, the heterodyne receiver would require 3 dB of noiseless gain to reach the nonamplified non-Gaussian receiver. On the other hand, if the same 3 dB amplifier is used with the non-Gaussian receiver, and compared to the amplified heterodyne receiver, then this ratio grows to $QNL/P_{E,NG} = 210$.
- [63] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. (Springer-Verlag, New York, 1998).
- [64] A. L. Maas, A. Y. Hannun, and A. Y. Ng, in *Proceedings of the 30th International Conference on International Conference on Machine Learning*, Vol. 28 (JMLR.org, 2013).
- [65] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, Tensorflow: A system for large-scale machine learning, in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation* (USENIX Association, Savannah, GA, 2016) pp. 265–283 (software available from tensorflow.org).
- [66] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, Efficient processing of deep neural networks: A tutorial and survey, *Proc. IEEE* **105**, 2295 (2017).
- [67] A. Jain, J. Mao, and K. Mohiuddin, Artificial neural networks: A tutorial, *Computer* **29**, 31 (1996).

- [68] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT, Cambridge, MA, 2016).
- [69] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow* (O'Reilly Media, Inc., Sebastopol, CA, 2017).
- [70] M. R. Khanzadi, Phase noise in communication systems: Modeling, compensation, and performance analysis, Ph.D. thesis, Chalmers University of Technology, 2015.
- [71] I. Holzman and Y. Ivry, Superconducting nanowires for single-photon detection: Progress, challenges, and opportunities, *Adv. Quantum Technol.* **2**, 1800058 (2019).
- [72] G. E. Uhlenbeck and L. S. Ornstein, On the theory of the Brownian motion, *Phys. Rev.* **36**, 823 (1930).
- [73] D. T. Gillespie, The mathematics of Brownian motion and Johnson noise, *Am. J. Phys.* **64**, 225 (1996).
- [74] R. E. Kalman, A new approach to linear filtering and prediction problems, *J. Basic Eng.* **82**, 35 (1960).
- [75] M. Bilkis, M. Rosati, R. M. Yepes, and J. Calsamiglia, Real-time calibration of coherent-state receivers: Learning by trial and error, *Phys. Rev. Research* **2**, 033295 (2020).
- [76] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, Reinforcement Learning with Neural Networks for Quantum Feedback, *Phys. Rev. X* **8**, 031084 (2018).
- [77] Q. Liao, Y. Guo, D. Huang, P. Huang, and G. Zeng, Long-distance continuous-variable quantum key distribution using non-Gaussian state-discrimination detection, *New J. Phys.* **20**, 023015 (2018).
- [78] J.-Y. Guan, F. Xu, H.-L. Yin, Y. Li, W.-J. Zhang, S.-J. Chen, X.-Y. Yang, L. Li, L.-X. You, T.-Y. Chen, Z. Wang, Q. Zhang, and J.-W. Pan, Observation of Quantum Fingerprinting Beating the Classical Limit, *Phys. Rev. Lett.* **116**, 240502 (2016).
- [79] github.com/UNM-QOLab/phase_amp_tracking_nn.
- [80] T. Tieleman and G. Hinton, Lecture 6.5 - Rmsprop, coursera: Neural networks for machine learning, 2012.
- [81] G. Bishop and G. Welch, *An introduction to the Kalman filter, Proceedings of the 28th ACM SIGGRAPH, Course 8* (ACM, Los Angeles, CA, 2001).