Phase tracking for sub-shot-noise-limited receivers

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

Nonconventional receivers for phase-coherent states based on non-Gaussian measurements such as photon counting surpass the sensitivity limits of shot-noise-limited coherent receivers, the quantum noise limit (QNL). These non-Gaussian receivers can have a significant impact in future coherent communication technologies. However, random phase changes in realistic communication channels, such as optical fibers, present serious challenges for extracting the information encoded in coherent states. While there are methods for correcting random phase noise with conventional heterodyne detection, phase tracking for non-Gaussian receivers surpassing the QNL is still an open problem. Here we demonstrate phase tracking for non-Gaussian receivers to correct for time-varying phase noise while allowing for decoding beyond the QNL. The phase-tracking method performs real-time parameter estimation and correction of phase drifts using the data from the non-Gaussian discrimination measurement, without relying on phase reference pilot fields. This method enables non-Gaussian receivers to achieve higher sensitivities and rates of information transfer than ideal coherent receivers in realistic channels with time-varying phase noise. This demonstration makes sub-QNL receivers a more robust, feasible, and practical quantum technology for classical and quantum communications.

I. INTRODUCTION

Optical communication with coherent states can achieve the highest rate of information transfer through lossy and noisy channels [1–3]. Coherent optical communications encode information in the coherent properties of the electromagnetic field, allowing for using high-spectral efficiency modulation and high-sensitivity coherent detection [4,5]. Efficient coherent modulation and detection can dramatically increase the rate of information transfer beyond the reaches of intensity encodings [5–7]. Moreover, the intrinsic nonorthogonality of coherent states can enable quantum communications [8–10] including quantum key distribution [11–16] for secure communications over optical networks [17,18]. However, coherent encodings are highly susceptible to phase noise and random phase variations in real-world devices and communication channels [5,6]. To ensure the expected advantage of coherent communications over intensity modulation and direct detection, communication protocols require efficient methods for phase estimation and phase tracking to correct for random phase changes induced by the channel [5–7], while being compatible with existing communication technologies. Moreover, practical scenarios in low-power and quantum communications require phase tracking based only on the transmitted signal state, without relying on transmissions of strong pilot phase reference pulses [19–25].

Conventional coherent receivers that realize Gaussian measurements, such as heterodyne receivers, can perform phase tracking based on signal postprocessing in the digital domain with diverse and efficient methods for channel and phase estimation [26–30]. These methods renewed interest in coherent communications for increasing information transfer, and have made coherent communications more practical for future realizations of high-capacity communication networks [5,31,32].

Further developments in optical communication will seek to approach the ultimate limits of information transfer in realistic communication channels. Quantum information science (QIS) provides the basis for approaching the fundamental limits in receiver sensitivities [33] and information transfer in communications [1–3]. Receivers based on Gaussian measurements, Gaussian operations, and local operations and classical communication have been investigated for information processing, phase estimation, and state discrimination [34]. The optimal Gaussian receiver for the discrimination of two nonorthogonal coherent states is the simple homodyne receiver [35]. Furthermore, measurements based on adaptive homodyne detection can provide advantages for single-shot phase estimation of coherent states [36–38]. However, the ultimate limits of receivers based on Gaussian operations for state discrimination are still under investigation [34,39]. Among technologies enabled by QIS, nonconventional receivers, termed quantum receivers, use optimized non-Gaussian measurements based on photon counting [35,40–58] to provide sensitivities surpassing quantum noise limit (QNL) of coherent receivers [5], and approach the true quantum-mechanical limit, the
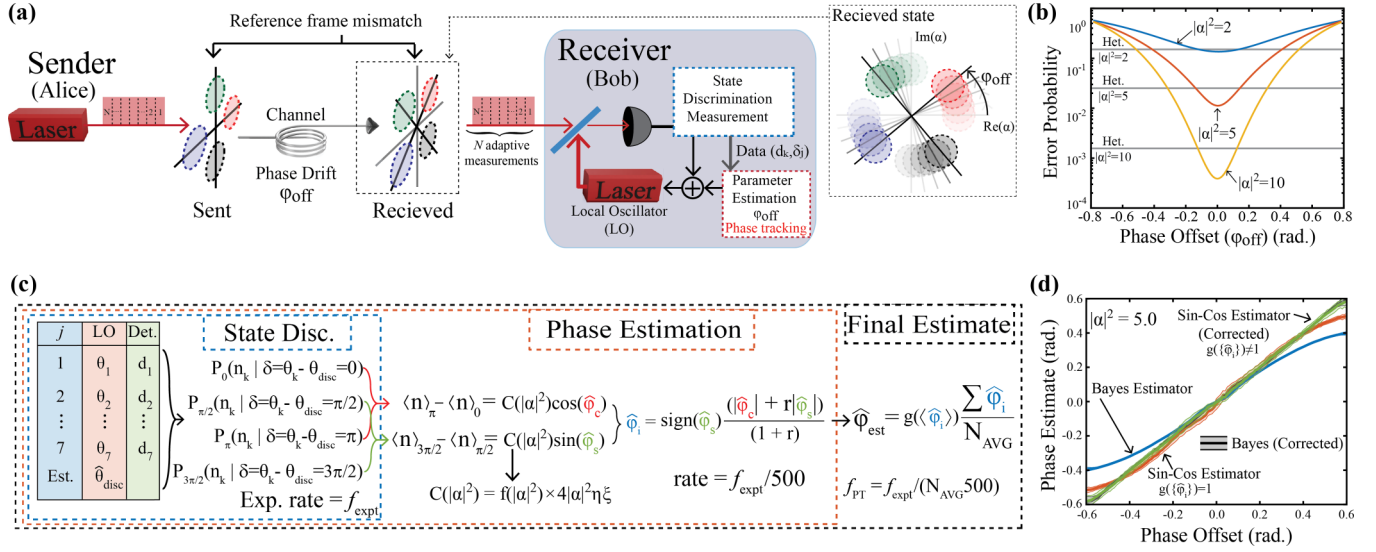

FIG. 1. Phase tracking for non- Gaussian receivers surpassing the QNL. (a) A sender (Alice) prepares a coherent state $|\alpha_k\rangle$ with a phase $\theta_k \in \{0, \pi/2, \pi, 3\pi/2\}$ by phase modulation of lasers. The pulses propagate though a channel inducing random phase drifts ϕ_{off}. The receiver (Bob) uses a local oscillator (LO) to perform optimized discrimination non-Gaussian measurements [53]. The mismatch between Alice's and Bob's phase reference frames caused by the channel increases the discrimination error for decoding information. (b) Probability of error for the adaptive non-Gaussian state discrimination measurement from Ref. [53] as a function of phase offset ϕ_{off} between the signal and LO, for signal mean photon number $|\alpha|^2 = 2.0$, 5.0, and 10.0, together with the ideal heterodyne limit (Het.) for each mean photon number. (c) Flowchart of the algorithm followed for phase tracking. The discrimination measurement provides samples for phase estimation containing detection results d_j for the relative phase between input state and the LO $\delta_j \in \{0, \pi/2, \pi, 3\pi/2\}$. After 500 channel transmissions, the pairs $\{d_j, \delta_j\}$ are used generate two estimates: $\hat{\phi}_c$ and $\hat{\phi}_s$. A weighted average $\hat{\phi}_i$ combines these estimates to increase accuracy. The final estimate $\hat{\phi}_{est}$ is the average over N_{avg} phase estimates $\{\hat{\phi}_i\}$ multiplied by a gain function $g(\langle\hat{\phi}_i\rangle)$. The phase-tracking method feeds forward the estimate $\hat{\phi}_{est}$ to the LO at a rate of $f_{PT} = f_{expt}/(500N_{avg})$ Hz for real-time phase tracking, with experimental repetition rate $f_{expt} \approx 12$ kHz. (d) Expected phase estimate as a function of applied phase for $|\alpha|^2 = 5.0$ for the Sin-Cos estimator $\hat{\phi}_{est}$ without (red line) and with (green line) optimized gain function $g(\langle\hat{\phi}_i\rangle)$, and for a Bayesian estimator (blue line) and the corrected Bayesian estimator (black line). The solid lines represent the mean of 100 Monte Carlo samples and shaded regions correspond to one standard deviation.

Helstrom bound [33]. Moreover, non-Gaussian receivers performing joint measurements over coherent-state codewords hold promise to bridge the gap between the Shannon and the Holevo limits in capacity [1,59]. However, making non-Gaussian receivers practical for coherent communications in realistic channels will require novel approaches for performing efficient phase tracking. These approaches will be fundamentally different from those based on conventional heterodyne detection using digital signal processing post measurement [26–30], and will require realizing active phase estimation [54,60] and correction in real time, while ensuring performance beyond the QNL.

Here we demonstrate a phase-tracking method for non-Gaussian receivers for quadrature phase-shift-keyed (QPSK) coherent states [53] based on coherent displacement, adaptive measurements, and photon counting. The phase-tracking method performs phase estimation and correction in real time using the data collected from the non-Gaussian discrimination measurement [53] without relying on strong phase-reference pilot pulses. This method enables the non-Gaussian receiver to overcome random phase variations encountered in realistic communication channels, while allowing the receiver to perform decoding measurements with sensitivities beyond the QNL, the shot-noise limit of conventional coherent receivers. This demonstration makes non-Gaussian receivers a more robust, feasible, and practical quantum technology for optical communications, and represents a significant advance for re-

alizing low-power communications approaching the quantum limits in realistic communication channels.

II. PHASE TRACKING FOR NON-GAUSSIAN RECEIVERS

Figure 1(a) shows the concept of phase tracking for a non-Gaussian measurement surpassing the QNL over a channel inducing random phase variations. The sender (Alice) uses laser pules to encode information in four coherent states with phases $\theta_k \in \{0, \pi/2, \pi, 3\pi/2\}$. The pulses propagate through the channel, which induces random phase shifts. The receiver (Bob) uses a laser as a local oscillator (LO) phase reference and performs a non-Gaussian discrimination measurement that surpasses the QNL for decoding the information [53]. The finite linewidths of the lasers in the transmitter and the receiver and the random channel phase variations cause mismatch between the phase space reference frames of Alice and Bob. These random phase drifts severely affect the expected performance of the state discrimination measurement. Figure 1(b) shows the probability of error for the adaptive non-Gaussian measurement for discriminating four nonorthogonal states $|\alpha_k\rangle \in \{|\alpha\rangle, |i\alpha\rangle, |-\alpha\rangle, |-i\alpha\rangle\}$ below the heterodyne limit [53], the QNL (see Appendix A), as a function of phase offset ϕ_{off} between the input state $|\alpha_k\rangle$ and the receiver's LO, for mean photon numbers $|\alpha|^2 = 2.0$, 5.0, and 10.0. While the discrimination strategy demonstrated in [53] can

tolerate small phase errors ϕ_{off} without significant degradation, moderate values of ϕ_{off} severely limit its performance, preventing discrimination below the QNL. To keep the expected performance benefit of the non-Gaussian measurement over the QNL, the receiver needs to perform phase tracking to correct for phase drifts induced by the channel. While phase tracking based on heterodyne measurements can be realized with digital signal processing post measurement [22,23,30], non-Gaussian receivers require active phase tracking and correction in real time to maintain performances below the QNL [49,53,55]. Here we demonstrate a method for actively tracking and correcting for time-varying random phases for non-Gaussian receivers to enable sensitivities beyond the ideal heterodyne limit in channels inducing random phase variations.

The phase-tracking method for non-Gaussian receivers builds on a discrimination strategy to discriminate a state $|\alpha_k\rangle \in \{|\alpha\rangle, |i\alpha\rangle, |-\alpha\rangle, |-i\alpha\rangle\}$ implementing N adaptive measurements with photon-number resolution [53]. During each adaptive measurement $j = 1, 2, \ldots, N$, the receiver's LO performs hypothesis testing of the input state $|\alpha_k\rangle$ by adjusting its phase $\theta_j \in \{0, \pi/2, \pi, 3\pi/2\}$ according to a Bayesian discrimination strategy [53]. After N adaptive measurements, the receiver provides an answer to the state discrimination problem θ_{disc} about the phase of the input state $|\alpha_k\rangle$ (see Appendix B). Assuming that this answer θ_{disc} is correct, the data collected during the N adaptive measurements can now be used as samples for estimation of the phase ϕ_{off} induced by the channel. For a discrimination measurement, these data consist of N photon-counting detections $\{d_1, d_2, \ldots, d_N\}$, together with the LO's phases during each adaptive measurement $\{\theta_1, \theta_2, \ldots, \theta_N\}$. Since the answer to the discrimination problem θ_{disc} is a very good estimate of the phase of the input state, it can be used to estimate the relative phases δ_j between the input state and the LO in each adaptive measurement as $\delta_j = \theta_j - \theta_{\text{disc}}$. The data for estimating ϕ_{off} then consist of the pairs $\{d_j, \delta_j\}$. Accumulating data during a moderate number of channel transmissions allows for estimating ϕ_{off} in real time and performing phase tracking simultaneously with the state discrimination measurement [53]. This method enables the receiver to utilize the data from the discrimination measurement to estimate and correct for random phase excursions.

To obtain an estimate of ϕ_{off}, the phase tracking method uses the collected data $\{d_j, \delta_j\}$ from discrimination measurements over 500 channel transmissions. These data consist of photon-counting samples of the interference between the input state and the LO for relative phases $\delta = \{0, \pi/2, \pi, 3\pi/2\}$. In principle, there are different estimators that can produce an estimate from the pairs $\{d_j, \delta_j\}$ (see Appendix C for two possible estimators). However, phase tracking for non-Gaussian receivers requires a simple estimator that can be efficiently calculated in real time, while being robust to the unavoidable errors from the discrimination measurement. A simple estimator can be obtained by using $\{d_j, \delta_j\}$ to generate four photon-number (Poisson) distributions $P_0(n_k|\delta = 0)$, $P_{\pi/2}(n_k|\delta = \pi/2)$, $P_\pi(n_k|\delta = \pi)$, and $P_{3\pi/2}(n_k|\delta = 3\pi/2)$ for $\delta = \{0, \pi/2, \pi, 3\pi/2\}$ [see Fig 1(c)]. Here, n_k is the photon number of detected photons for different distributions. By

calculating the differences between means $\langle n\rangle_\delta$ of these distributions we can form estimates of ϕ_{off} as

$$\langle n\rangle_\pi - \langle n\rangle_0 = C(|\alpha|^2)\cos(\hat{\phi}_c), \tag{1}$$

$$\langle n\rangle_{3\pi/2} - \langle n\rangle_{\pi/2} = C(|\alpha|^2)\sin(\hat{\phi}_s) \tag{2}$$

with

$$C(|\alpha|^2) = f(|\alpha|^2) \times 4|\alpha|^2\eta\xi/N, \tag{3}$$

where η is the detection efficiency, ξ is the interference visibility, and $\hat{\phi}_{c,s}$ are the phase estimates. $f(|\alpha|^2)$ is a factor that is used to reduce a bias in the phase estimates arising from the nonzero probability of error for the state discrimination strategy (see Appendix C). Errors in the state discrimination measurement ($P_E \neq 0$) cause errors in populating the distributions P_δ and thus in their mean values $\langle n\rangle_\delta$. These errors cause $\hat{\phi}_c$ to be biased away from zero by making $\langle n\rangle_\pi - \langle n\rangle_0 < 4|\alpha|^2\eta\xi/N$. The function $f(|\alpha|^2)$ allows for correcting these biases by making $\langle n\rangle_\pi - \langle n\rangle_0 \approx f(|\alpha|^2) \times 4|\alpha|^2\eta\xi/N$, thus reducing the effects of discrimination errors in the phase estimation procedure. These errors also produce biases in $\hat{\phi}_s$ by causing $\langle n\rangle_{3\pi/2} - \langle n\rangle_{\pi/2}$ to be reduced. Appendix C 1 describes the procedure to obtain the optimal values of $f(|\alpha|^2)$ for reducing the effects of discrimination errors. As a second step, the phase tracking method uses a weighted average of estimates $\hat{\phi}_c$ and $\hat{\phi}_s$ with relative weight r to obtain an estimate $\hat{\phi}_i$ of ϕ_{off} within 500 transmissions [see Fig. 1(b)]. The weight r, determined from Monte Carlo simulations, allows for reducing the difference between $\hat{\phi}_i$ and ϕ_{off} at the end points of the capture range for phase tracking (± 0.6 rad in our experimental demonstration). The final estimate $\hat{\phi}_{\text{est}}$ in the phase-tracking method is obtained by averaging over N_{avg} estimates $\hat{\phi}_i$ and multiplying by a gain factor $g(\langle\hat{\phi}_i\rangle)$ that depends on the N_{avg} estimates $\{\hat{\phi}_i\}$ (see Appendix D). This gain factor reduces the difference between the applied phase ϕ_{off} and the estimated phases $\{\hat{\phi}_i\}$. Figure 1(d) shows the result of Monte Carlo simulations of the final phase estimate $\hat{\phi}_{\text{est}}$ with $g(\langle\hat{\phi}_i\rangle) = 1$ (red line) and $g(\langle\hat{\phi}_i\rangle)$ optimized to approach the true phase ϕ_{off} (green line). This final estimate $\hat{\phi}_{\text{est}}$ is used to feed forward to the receiver's LO every $500 \times N_{\text{avg}}$ channel transmissions at a rate f_{PT} for phase drift correction. This method enables real-time phase tracking for correcting time-varying phases while enabling the non-Gaussian receiver to surpass the QNL.

III. EXPERIMENTAL CONFIGURATION

Figure 2 shows the experimental configuration for the demonstration of phase tracking of adaptive non-Gaussian state discrimination measurements for QPSK states $\{|\alpha\rangle, |i\alpha\rangle, |-\alpha\rangle, |-i\alpha\rangle\}$. The measurement strategy consists of $N = 7$ adaptive measurements via feedback in the phase of the LO [53]. A helium-neon (HeNe) laser at 633 nm and an acousto-optic modulator (AOM) prepare 35-μs coherent state pulses at a rate of $f_{\text{expt}} \approx 12$ kHz. The light pulses enter an unbalanced Mach-Zender interferometer through a 50/50 beam splitter. We prepare the phases of the input signal state and the LO with two 4:1 multiplexers (MUX) and two phase modulators (PM). The input states and LO interfere in a 99/1 beam splitter, which implements the displacement
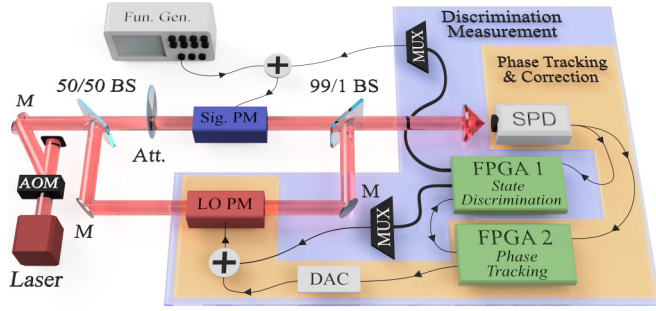
FIG. 2. Experimental configuration. Experimental setup for the demonstration of phase tracking for non-Gaussian state discrimination measurements for QPSK states surpassing the QNL. See main text for details. AOM, acousto-optic modulator; Att., variable attenuator; PM, phase modulator; MUX, multiplexer; FPGA, field-programmable gate array; SPD, single-photon detector; Fun. Gen, function generator; M, Mirror.

operation of the input state [61]. A field-programmable gate array (FPGA), FPGA1 in Fig. 2, implements the discrimination strategy based on adaptive measurements and photon-number-resolving (PNR) detection described in Ref. [53]. This FPGA1 controls the timing of the experiment, processes photon detections, and updates the phase of the LO for each adaptive measurement. The overall detection efficiency of the experiment is $\eta = 72\%$, with interference visibility $\xi = 99.8\%$.

We use a second FPGA (FPGA2) to perform active phase tracking using the data pairs $\{d_j, \delta_j\}$ sent from FPGA1 generated from the state discrimination measurement, as described above. FPGA2 performs real-time phase estimation to obtain an estimate of ϕ_{off}, and feeds forward this information to the receiver to adjust the LO phase to perform phase tracking and correction. Controllable phase offsets and phase noise in the input state are prepared with an arbitrary waveform function generator (FG) [58] for investigating the phase-tracking method in channels inducing random phase variations. We use an 8-bit digital to analog converter (DAC) to feed forward the estimated phase offset to the LO. We chose a finite capture range R of phases for feed forward to the phase of the LO equal to $R = [-0.6, +0.6]$ rad. This choice results in a phase resolution of about $1.2 \text{ rad}/2^8 = 5$ mrad for phase tracking, while having a large enough capture range.

The absolute power of the input state is calibrated using a photodiode-based light-trapping (TRAP) detector with a 0.05% uncertainty tied to an absolute spectral response scale [62]. This TRAP detector was used to calibrate a series of attenuators to lower the power of a power-stabilized 633-nm laser to the single-photon level with a combined 1σ uncertainty of 1.8%, and the transmission of the optical elements from where the state is prepared to where it is detected with transmittance $T = 92.5(2)\%$. This results in a total uncertainty for the calibration of the absolute average photon number per pulse of $\sigma \approx 2\%$. The FPGAs used for implementing the state discrimination strategy and phase tracking were both Altera Cyclone II FPGAs, model EP2C5T144C8 with 4608 logical elements, base clock of 48 MHz, and 158 digital I/O pins.
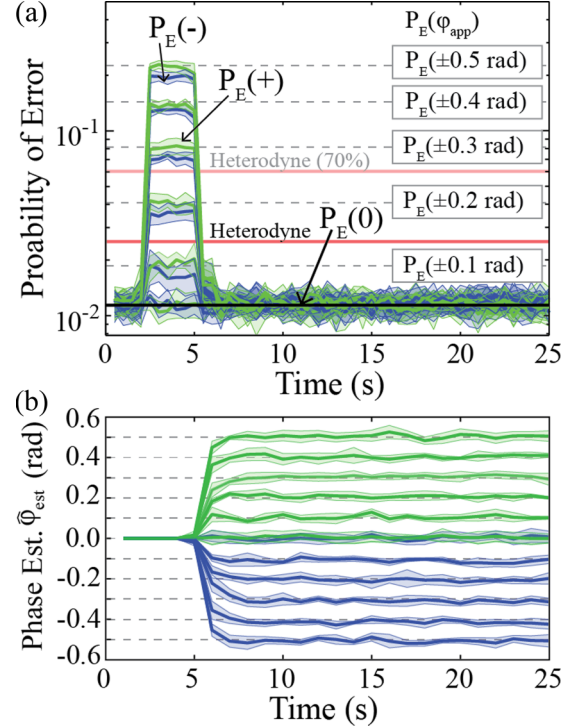

FIG. 3. Phase tracking under constant phase offsets. (a) Probability of error P_E as a function of time (t) shown for every 0.5 s time bin. From $t = 0$ to 2 s no phase offset is applied, $\phi_{app} = 0$. At $t = 2$ s, the input state experiences a constant phase offset ϕ_{app} producing an increase in P_E. At $t = 4$ s, the phase tracking turns on and corrects for ϕ_{app}. (b) Phase-tracking estimates $\hat{\phi}_{est}$ as a function of time estimated every $T_{PT} \approx 0.85$ s. Dashed lines show expected error probabilities P_E in (a) and applied phase offsets ϕ_{app} in (b).

IV. RESULTS

We investigate the performance of the phase-tracking method under different scenarios. In the first scenario, the input state experiences a sudden constant phase offset and the phase-tracking method needs to estimate and correct for large phase offsets. The second scenario aims to simulate a realistic channel inducing time-varying phase noise, where the input state experiences Gaussian random walks in phase at different diffusion rates. This scenario allows us to investigate phase tracking of random phase drifts in the channel and the impact of tracking bandwidth on the performance of non-Gaussian receivers.

A. Phase tracking under constant phase offsets

Figure 3 shows the performance of the phase-tracking method under sudden constant phase offsets of $\phi_{app} = \{\pm 0.1, \pm 0.2, \pm 0.3, \pm 0.4, \pm 0.5\}$ rad of the input state with mean photon number $|\alpha|^2 = 5.0$. Figure 3(a) shows the probability of error P_E calculated for time bins of about 0.5 s. Figure 3(b) shows the phase estimates $\hat{\phi}_{est}$ generated by the phase-tracking method as a function of time. Thick lines represent the average over five independent experimental runs, each time bin corresponds to about 5×10^3 ($\approx f_{expt} \times 0.5$ s) independent experiments, and shaded regions represent one standard deviation. Green (blue) lines correspond to
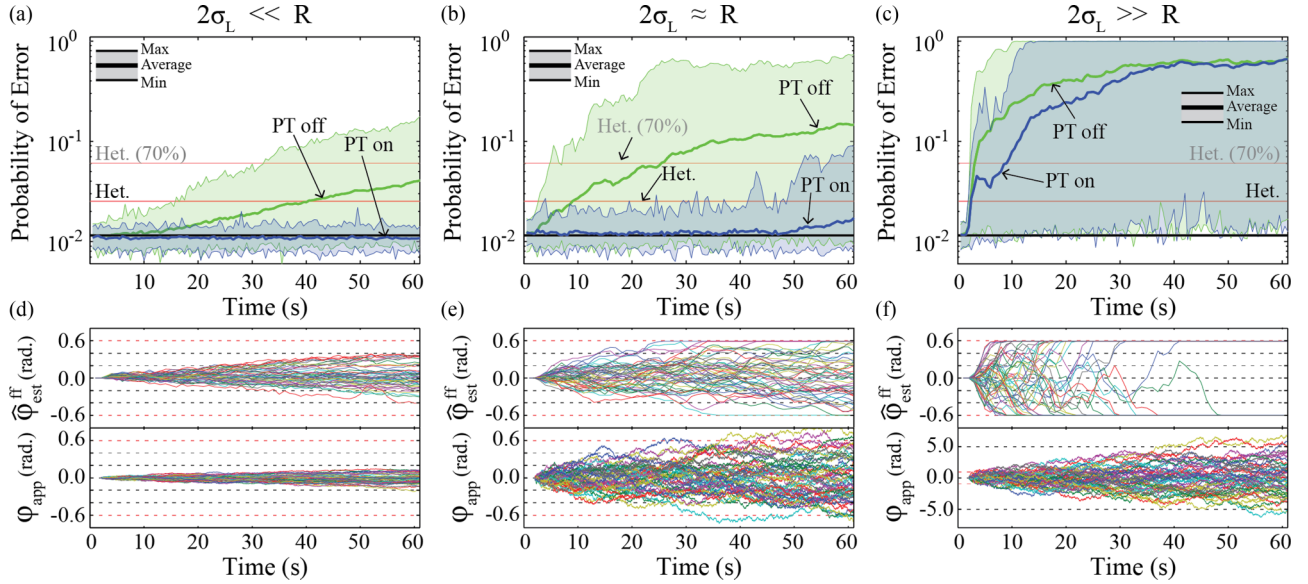
FIG. 4. Phase tracking (PT) of Gaussian random walks in phase. (a)–(c) Probability of error P_E for $|\alpha|^2 = 5.0$ as a function of time for 50 realizations of Gaussian random walks in phase applied to the input signal for (a) $\sigma_1 = 1$ mrad ($= 0.2 \times R/2\sqrt{L}$), (b) $\sigma_1 = 5$ mrad ($= 1 \times R/2\sqrt{L}$), and (c) $\sigma_1 = 5$ mrad ($= 25 \times R/2\sqrt{L}$), with total variances $\sigma_L^2 = L\sigma_1^2$ for each case. Here, $L = 6500$ corresponds to the total steps in the random walks, and $R = [-0.6, 0.6]$ rad is the capture range of phase tracking in our experiment. Bold lines show the average and shade regions the spread for these walks with (blue) and without (green) phase tracking. (d)–(f) Applied phases ϕ_{app} during Gaussian random walks and the phase estimates $\hat{\phi}_{est}^{ff} = \hat{\phi}_{est}$ generated by the phase-tracking method for cases (a)–(c), respectively. The phase estimates $\hat{\phi}_{est}^{ff}$ used to feed forward to the LO for phase correction are bounded by the experimental capture range R, as can be seen in (e) and (f).

positive (negative) applied phase offsets ϕ_{app}. The phase-tracking method here uses $N_{avg} = 20$, so that each estimate $\hat{\phi}_{est}$ was obtained at a phase-tracking rate $f_{PT} = 1/T_{PT} \approx 1/0.85$ s [see Fig. 1(c)]. For this investigation, the relative phase of the signal and LO was locked between each experimental trial, similar to Ref. [53], and the phase offset ϕ_{app} was applied during each state discrimination measurement.

From $t = 0$ to 2 s, we verify that the performance of the receiver is 3.4 dB below the heterodyne limit (red line) without the applied phase offset $\phi_{app} = 0$. At $t = 2$ s, we apply a constant phase offset ϕ_{app} to the input state, producing sudden jumps in the probability of error P_E depending on the value of ϕ_{app}. At $t = 4$ s, the phase-tracking method is turned on. After an estimation cycle $T_{PT} \approx 0.85$ s, the phase tracking produces an estimate $\hat{\phi}_{est}$ and corrects for ϕ_{app}, allowing the receiver to perform below the heterodyne limit for all phase offsets. Figure 3(b) shows the phase estimates $\hat{\phi}_{est}$ as a function of time, demonstrating that the phase-tracking method accurately identifies and corrects for large phase offsets. We observe that this method enables the non-Gaussian receiver to keep its expected advantage of 3.4 dB over an ideal heterodyne measurement.

B. Phase tracking of random walks in phase

1. Phase tracking with different noise strengths

In coherent optical communications, the receiver is usually required to decode information encoded in coherent state signals in the presence of time-dependent random variations in phase, which severely limits the receiver's ability to recover the information. We investigate the phase-tracking method for situations where the input state experiences Gaussian phase

noise [28], which could include effects of phase noise in the LO and the transmitter laser or random fluctuations arising from different processes [27,28,63–66]. Gaussian random walks in phase have been broadly used as an acceptable model for phase noise in optical communications and phase drift between the sender and receiver in classical [5,27,67,68] and quantum communications [22,30,30]. We note that the algorithm used for phase tracking is independent of the choice of phase noise model, as it makes no assumptions about the dynamics of the noise and the noise model is not used to obtain the phase estimate (see Sec. II).

For this study, we do not stabilize the relative phase between the input signal state and the LO. Under these conditions, the receiver experiences the drift of the experimental setup and induced random walks in phase. This situation is analogous to having a LO whose phase is constantly drifting and a channel that induces phase noise. This experimental configuration aims to mimic more realistic situations where the signal and LO are generated from different lasers [28]. In this investigation, the phase noise in the signal is implemented by preparing discrete Gaussian random walks in phase with $L = 6500$ steps, each distributed according to a zero-mean Gaussian distribution with standard deviation σ_1 [28].

Figures 4(a)–4(c) show the probability of error P_E for $|\alpha|^2 = 5.0$ as a function of time for 50 realizations of discrete Gaussian random walks in phase at a rate of $f_{RW} = 100$ Hz for walks with (a) $\sigma_1 = 0.1$ mrad, (b) $\sigma_1 = 5$ mrad, and (c) $\sigma_1 = 25$ mrad. Thick lines show the average over 50 walks and shaded regions represent the spread for these walks with (blue) and without (green) phase tracking (PT) with $N_{avg} = 20$, so that $f_{PT} \approx 1.2$ Hz. After L steps the total variance of the Gaussian random walks are $\sigma_L^2 = L\sigma_1^2$, so that situations

in Figs. 4(a)–4(c) correspond to different regimes: (a) small ($2\sigma_L < R$), (b) moderate ($2\sigma_L \approx R$), and (c) severe ($2\sigma_L \gg R$) phase noise. Here, $R = [-0.6, 0.6]$ rad is the capture range for phase tracking in our experiment, which is experimentally chosen. Figures 4(d)–4(f) show the applied phase ϕ_{app} during the Gaussian random walks for cases in Figs. 4(a)–4(c), respectively, and the phase estimate $\hat{\phi}_{\text{est}}^{ff}$ that is the actual phase sent to the LO for phase tracking via feed forward. Note that the forwarded phase $\hat{\phi}_{\text{est}}^{ff}$ is restricted to be within the experimental capture range R so that $\hat{\phi}_{\text{est}}^{ff} = \hat{\phi}_{\text{est}}$ under the condition $|\hat{\phi}_{\text{est}}^{ff}| < |R|$.

We observe that, in general, Gaussian random walks in phase severely degrade the performance of the non-Gaussian receiver precluding any advantage over the heterodyne limit. However, in situations with small ($2\sigma_L < R$) and moderate ($2\sigma_L \approx R$) levels of noise in Figs. 4(a) and 4(b), respectively, the phase-tracking method accurately estimates and corrects for phase noise, enabling the receiver to maintain its performance 3.4 dB below the heterodyne limit. For small phase noise in Fig. 4(d) with $2\sigma_L < R$, the applied phase ϕ_{app} is smaller than the phase drifts in the experiment. The estimated phase $\hat{\phi}_{\text{est}}^{ff}$ captures the contributions of ϕ_{app} and of these drifts showing a larger variance than ϕ_{app}. For situations with moderate phase noise with $2\sigma_L \approx R$, the applied walks in phase ϕ_{app} contain walks that exceed the capture range R at some point in time, as shown in Fig. 4(e). After this point, the estimated phases $\hat{\phi}_{\text{est}}$ for these walks are clamped at $|R|$ to generate $\hat{\phi}_{\text{est}}^{ff}$ to feed forward to the LO. This procedure produces a slight increase in P_E after 50 s, as shown in Fig. 4(b).

For situations with large phase noise in Fig. 4(c) for which $2\sigma_L \gg R$, the receiver's performance degrades above the heterodyne limit within a short time. The phase-tracking method can reduce the effects of phase noise. However, since the applied phases ϕ_{app} rapidly exceed the capture range R, the estimated phases $\hat{\phi}_{\text{est}}^{ff}$ that are fed forward to the LO are clamped at R for many cases, as can be seen in Fig. 4(f). This procedure limits the performance for phase tracking in our current implementation. However, increasing the resolution of the electronic controller and the DAC to 10 bits used to feed forward to the LO phase can allow for increasing the capture range to $R = \pm\pi$ rad, while maintaining good phase resolution of ≈ 6 mrad for phase tracking. In this case, whenever ϕ_{app} reaches this range, the estimate $\hat{\phi}_{\text{est}}^{ff}$ would wrap around from $\pm\pi$ to $\mp\pi$. This procedure would allow for tracking phase walks exceeding R and maintaining the receiver's performance below the heterodyne limit under any level of phase noise.

2. Phase tracking with different input powers

The performance of the phase-tracking method for non-Gaussian receivers critically depends on the performance of the state discrimination measurement. The information used for phase estimation and tracking $\{d_j, \delta_j\}$ assumes the answer to the discrimination problem ϕ_{disc} to be correct, which is true only with probability $P_C = 1 - P_E$. Since P_E depends strongly on the mean photon number $|\alpha|^2$ of the input state [53], the receiver's ability to perform phase tracking will also
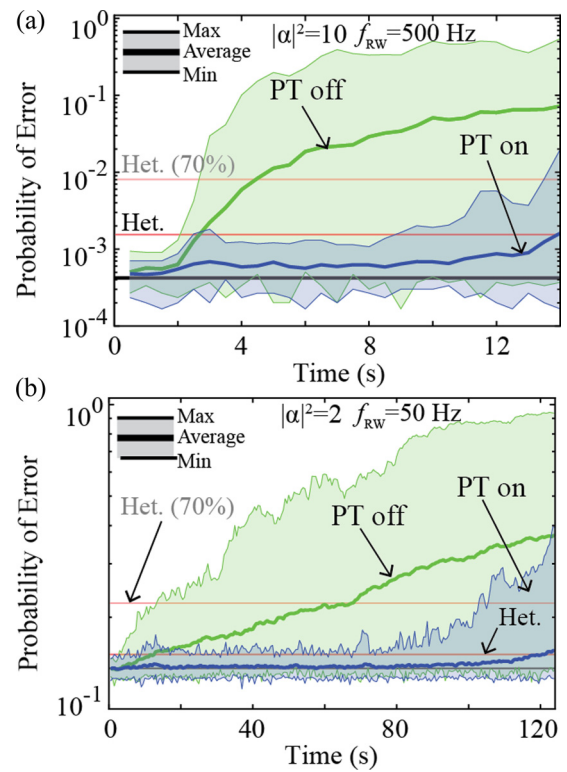

FIG. 5. Phase tracking with different mean photon numbers. Probability of error as a function of time under Gaussian phase noise with (blue) and without (green) phase tracking for (a) $|\alpha|^2 = 10.0$ and (b) $|\alpha|^2 = 2.0$. Thick lines show the averages and shaded regions show the spread in P_E over 50 Gaussian random walks. Parameters for phase tracking f_{PT} and for the Gaussian random walks f_{RW} and σ_i are chosen to satisfy $2\sigma_L \approx R$, analogous to the situation in Fig. 4(b) for $|\alpha|^2 = 5.0$. Black lines show the expected performance of the receiver in the absence of Gaussian phase noise: 6 and 0.45 dB below the heterodyne limit for $|\alpha|^2 = 10.0$ and 2.0, respectively [53]. Note that the phase-tracking method enables the receiver to perform below the heterodyne limit under Gaussian phase noise in the high- and low-input power regimes.

depend on $|\alpha|^2$. Larger input powers $|\alpha|^2$ result in lower error probabilities P_E, and can allow the phase-tracking method to perform phase estimation with higher accuracy, achieve higher tracking rates f_{PT}, and correct for phase noise with higher bandwidths f_{RW}. On the other hand, for low powers $|\alpha|^2$ the performance of the phase-tracking method is affected due to higher P_E. However, achieving phase tracking in these two power regimes is required for both low-power classical [6,27,28,63] and quantum [13,22,23] communications.

Figure 5 shows the performance of the phase-tracking method for $|\alpha|^2 = 10.0$ (a) and (b) $|\alpha|^2 = 2.0$. Phase tracking in these high- and low-input power regimes can be implemented at different rates f_{PT} to correct for noise with different strengths and bandwidths. In the two plots in Fig. 5, the phase-tracking parameters and Gaussian phase noise are chosen to satisfy the condition $2\sigma_L = 2\sqrt{L}\sigma_1 = 2f_{\text{RW}}T\sigma_1 \approx R$, so that these situations are analogous to the one shown in Fig. 4(b) for $|\alpha|^2 = 5.0$. Here, $\sigma_1 = 5$ mrad, and T is the displayed period: $T = 13$ s for $|\alpha|^2 = 10.0$; and $T = 120$ s for $|\alpha|^2 = 2.0$. For

$|\alpha|^2 = 10.0$, phase tracking achieves higher estimation accuracy, and can be reliably implemented with $N_{avg} = 4$, enabling phase-tracking rates of $f_{PT} \approx 5.8$ Hz, five times faster than for $|\alpha|^2 = 5.0$. As a result, the receiver can track and correct for random Gaussian phase noise with a rate of $f_{RW} = 500$ Hz, while performing below the heterodyne limit [see Fig. 5(a)]. Phase tracking for $|\alpha|^2 = 2.0$ requires more samples to obtain accurate phase estimates, and can be implemented reliably with $N_{avg} = 40$ with a tracking rate $f_{PT} \approx 0.5$ Hz. In this case, the receiver can track and correct for phase noise at a rate of $f_{RW} = 50$ Hz, while performing below the heterodyne limit [see Fig. 5(b)].

3. Phase tracking of noise with different bandwidths

The phase-tracking method has a strong dependence on the noise bandwidth present in the communication channel and how it compares to the rate at which phase tracking can be implemented [28]. We have studied the performance of the phase-tracking method for non-Gaussian receivers for tracking random phase noise with different noise bandwidths. This study is described in Appendix E. In our findings we observe that for a fixed f_{PT}, the phase-tracking method can correct for noise with different bandwidths f_{RW}. We note, however, that f_{PT} has to be high enough to keep the receiver's performance below the heterodyne limit for extended periods of time. As one example, we observe that for a non-Gaussian receiver with $|\alpha|^2 = 5.0$ and $f_{PT} = 1.2$ Hz, reliable phase tracking of random noise can be performed for noise bandwidths $f_{RW} = 100$ Hz, and sub-QNL sensitivity can be kept for $f_{RW} = 500$ Hz for ≈ 10 s. Tracking noise with higher bandwidths can be achieved with higher experimental rates $f_{expt} > 11$ kHz to increase f_{PT} or with larger mean photon numbers $|\alpha|^2$ to generate more accurate phase estimates $\hat{\phi}_{est}$.

V. DISCUSSION

Receivers with sensitivities surpassing the QNL of ideal conventional receivers have a large potential for enabling efficient and reliable low-power communications at the single- and few-photon levels. The phase-tracking method demonstrated here for non-Gaussian receivers allows for tracking random phase variations and noise with different strengths and bandwidths. This method provides the much needed robustness to enable non-Gaussian receivers to perform below the QNL in channels with phase noise for a wide range of powers. We note that the phase-tracking bandwidth in our proof-of-principle demonstration was implemented at low rates because of experimental constraints, and used a single laser shared between transmitter and receiver. However, using an estimator with higher estimation accuracy, such as the Bayesian estimator, combined with high-bandwidth electronics and efficient single-photon detectors [69], would allow the receiver's measurement and phase tracking to be realized at much higher bandwidths. This in turn would enable non-Gaussian receivers to overcome realistic noise in communication channels with independent lasers at the receiver and transmitter [6], while outperforming ideal shot-noise-limited coherent receivers [5].

We note that transmissions of high-power pulses interleaved with the input states could be used with a heterodyne detection for phase tracking [20,23], without relying on knowledge of the power of the input coherent states to be discriminated and the visibility of the interference with the LO. The phase-tracking method presented here uses only the data directly collected from the non-Gaussian measurement that assumes a known intensity and visibility. However, the data from the state discrimination strategy could in principle be used to estimate the input intensity and visibility in addition to the phase offset, and allow for tracking of multiple time-varying parameters without the need for dedicated light pulses for estimation and tracking. We also note that it may be possible to split the power of the input state to use a fraction of light to perform phase estimation with a heterodyne measurement. However, the estimation precision of these split-and-estimate methods for phase tracking will depend on the fraction of power used for phase estimation, and there will be an increase in the probability of error in the state discrimination due to the reduced power entering the sub-shot-noise receiver.

In the future, enabling coherent communication technologies that can approach the quantum limits in sensitivity and information transfer in realistic channels at low powers will require the ability to track other impairments in the channel including polarization rotation, background noise, and power variations. While we demonstrated a method for tracking to correct phase drifts induced by a channel, we believe that the data from the state discrimination measurement that are used for phase tracking can be leveraged for estimation and tracking of other sources of noise in the channel, such as amplitude noise (see Appendix E). Moreover, we anticipate that this technique for phase tracking can be applied to optimized non-Gaussian measurements surpassing the QNL in the single-photon regime [55]. This possibility can enable phase tracking in quantum key distribution for secure communications at very low powers without requiring strong phase reference pilot pulses [22].

VI. CONCLUSIONS

We demonstrate a phase-tracking method for non-Gaussian receivers [53] for phase-encoded coherent states surpassing the sensitivity limits of shot-noise-limited coherent receivers: the quantum noise limit (QNL) [5]. The phase-tracking method performs phase estimation and correction in real time using the data from the non-Gaussian discrimination measurement [53], without continuously relying on phase reference pilot fields from the transmitter. Our experimental demonstration shows that the phase-tracking method provides non-Gaussian receivers with the required robustness to overcome random phase noise encountered in realistic communication channels, and enables the receiver to perform measurements beyond the QNL under diverse conditions with different noise strengths and bandwidths. Moreover, since the phase-tracking method uses the data from a measurement surpassing the QNL at very low power levels, this method is well suited for assisting quantum communication protocols based on weak coherent states for efficient [10,70] and secure [8,9,11–16] communications. Our demonstration of phase tracking for non-Gaussian receivers makes sub-shot-

noise-limited receivers a more robust, feasible, and practical quantum technology for low-power communications based on coherent states for approaching the quantum limits in realistic communication channels.

APPENDIX A: QUANTUM NOISE LIMIT

The quantum noise limit (QNL) for the discrimination of coherent states in a given encoding scheme is obtained through the probability of error in discrimination:

$$P_E = 1 - P_C = 1 - \sum_{k=1}^{M} P(\alpha_k)P(\alpha_k|\alpha_k), \quad (A1)$$

where M is the number of states in the alphabet, and $P(\alpha_k)$ is the prior probability of state $|\alpha_k\rangle$ which is equal to $\frac{1}{M}$ for equiprobable states. $P(\alpha_k|\alpha_k)$ is the probability of guessing state $|\alpha_k\rangle$ given that state $|\alpha_k\rangle$ was sent, i.e., the probability of correct discrimination.

For the discrimination of two coherent states in the binary phase shift keying (BPSK) format $|\alpha_k\rangle \in \{|\pm\alpha\rangle\}$, the homodyne measurement along the x quadrature is the optimal Gaussian measurement [71]. The probability of error for the homodyne measurement corresponds to the QNL for the BPSK alphabet. $P(\alpha_k|\alpha_k)$ for a homodyne measurement is [72]

$$P(\alpha_k|\alpha_k) = \frac{1}{\sqrt{\pi}} \int_R e^{-(x-\sqrt{2}\alpha_k)^2} dx = \frac{1}{2}[1 + \text{erf}(\sqrt{2}\alpha)], \quad (A2)$$

where R is the region where $|\alpha_k\rangle$ is the most likely state, and erf(.) is the error function.

Using Eq. (A1), the total probability of error is

$$P_E = 1 - \frac{1}{2}[1 + \text{erf}(\sqrt{2}\alpha)] = \text{QNL}_{\text{BPSK}}. \quad (A3)$$

For QPSK states $|\alpha_k\rangle \in \{|\alpha e^{ik\frac{\pi}{2}}\rangle\}$, where $k \in \{0, 1, 2, 3\}$, the QNL corresponds to the probability of error of an ideal heterodyne measurement [34], which performs a projection onto coherent states and measures both quadratures of the input state simultaneously. The probability of correct discrimination of state $|\alpha_k\rangle$ is given by [72]

$$P(\alpha_k|\alpha_k) = \frac{1}{4}\left[1 + \text{erf}\left(\frac{\alpha}{\sqrt{2}}\right)\right]^2. \quad (A4)$$

Then, the QNL for QPSK states is [5,34]

$$P_E = 1 - \frac{1}{4}\left[1 + \text{erf}\left(\frac{\alpha}{\sqrt{2}}\right)\right]^2 = \text{QNL}_{\text{QPSK}}. \quad (A5)$$

While the homodyne measurement is known to be the optimal Gaussian measurement for the discrimination of two coherent states, the ultimate Gaussian limit for coherent multistate discrimination is not known. Therefore, there may be strategies based on Gaussian operations and measurements [34] that provide advantages over the heterodyne measurement [37].

In a general M-PSK encoding $|\alpha_k\rangle \in \{|\alpha e^{ik\frac{\pi}{2}}\rangle\}$, where $k \in \{0, 1, \ldots, M-1\}$, the probability of correct discrimination can be found through [72]

$$P(\alpha_k|\alpha_k) = \frac{1}{\pi} \iint_R e^{-|re^{i\theta}-\alpha_k|^2} r\, dr\, d\theta. \quad (A6)$$

The QNL$_{\text{MPSK}}$ is then obtained by using Eq. (A1).

APPENDIX B: STATE DISCRIMINATION STRATEGY

The phase-tracking method builds on the adaptive measurement strategy for QPSK states with PNR detection described in detail in Ref. [53]. In this strategy, the receiver performs $N = 7$ adaptive measurements on the input state $|\alpha_k\rangle \in \{|\alpha\rangle, |i\alpha\rangle, |-\alpha\rangle, |-i\alpha\rangle\}$. In each adaptive measurement j ($j = 1, 2, \ldots, N$), the LO is prepared in a state hypothesis $|\beta_j\rangle$, and displaces the input state $|\alpha_k\rangle$ to $\hat{D}(-\beta_j)|\alpha_k\rangle$. Note that for a correct hypothesis $\beta_j = \alpha_k$, the input state $|\alpha_k\rangle$ is displaced to the vacuum state $|0\rangle$. The displaced state $\hat{D}(-\beta_j)|\alpha_k\rangle$ is then detected with a PNR detector with number resolution m, ideally described by operators $\hat{\Pi}_n = |n\rangle\langle n|$ for $n = 1, 2, \ldots, m-1$ and $\hat{\Pi}_m = \hat{I} - \sum_{i=0}^{m-1} \hat{\Pi}_i$. The strategy uses a maximum *a posteriori* probability (MAP) criterion and a recursive Bayesian updating [53]. Given a photon-number detection d_j and the hypothesis β_j in adaptive measurement j, the strategy estimates the posterior Bayesian probabilities for input states and the most likely state. In subsequent adaptive measurements, the LO tests this most likely state, and prior probabilities are updated according to Bayes' theorem. Recursive application of this method during all adaptive measurements results in a final estimate θ_{disc} of the possible input state, which corresponds to the most likely state at the end of the last adaptive measurement N, β_{N+1}. This most likely state corresponds to the answer to the state discrimination problem, and the discrimination strategy allows for surpassing the QNL. After a discrimination measurement, the data collected during N adaptive measurements consists of N photon-counting detections $\{d_1, d_2, \ldots, d_N\}$, together with the phases of the most likely states β_j in each adaptive measurement $\{\theta_1, \theta_2, \ldots, \theta_N\}$. Assuming that the answer to the state discrimination problem is correct, the phase $\theta_{\text{disc}} = \arg\{\beta_{N+1}\}$ then corresponds to the phase of the input state, so that $\delta_j = \theta_j - \theta_{\text{disc}}$ are the relative phases of the input state and the LO during each adaptive measurement. The pairs $\{d_j, \delta_j\}$ correspond to samples of phase space that can be used to estimate the phase offset caused by the channel for performing phase tracking.

APPENDIX C: PHASE ESTIMATOR AND PERFORMANCE

The method for phase tracking for non-Gaussian receivers uses the data collected from the state discrimination measurement, consisting of the pairs $\{d_j, \delta_j\}$, to estimate and correct for the relative phase between the input state and the local oscillator (LO) in real time. This method works in conjunction with the state discrimination strategy and requires no extra resources such as strong phase reference pilot pulses or performing additional measurements for phase estimation. For the adaptive non-Gaussian discrimination measurement in Ref. [53] with photon-number resolution

(PNR) of 3, PNR(3), the receiver samples four photon-number distributions $P_0(n_k|\delta = 0)$, $P_{\pi/2}(n_k|\delta = \pi/2)$, $P_\pi(n_k|\delta = \pi)$, and $P_{3\pi/2}(n_k|\delta = 3\pi/2)$ for $\delta = (0, \pi/2, \pi, 3\pi/2)$. These photon-number distributions can be used to obtain different estimators for the phase offset ϕ_{off} (or the applied phase ϕ_{app} in the experimental investigation described in the main paper). We note that the photon-number distributions $P_0(n_k|\delta)$ can represent the rows of a 4×4 matrix. In general, these distributions can be arranged as rows of a (PNR + 1) $\times M$ detection matrix for M-ary shift keyed states and PNR of the measurement. Below, we describe two estimators: one based on the differences of the mean photon numbers $\langle n \rangle_\delta$ of these distributions referred to as "sine-cosine estimator" that is implemented in our demonstration, and one that is a Bayesian estimator.

1. Sine-cosine estimator

The sine-cosine estimator, as implemented in the experimental demonstration described in the main text, uses the differences of the average of detected photon numbers $\langle n \rangle_\delta$ to obtain a final estimate $\hat{\phi}_{\text{est}}$ of the phase offset ϕ_{off} (or the applied phase ϕ_{app} in the actual experiments) based on the collected data from $N_{\text{avg}} \times 500$ channel transmissions.

As a first step, the estimator obtains two initial estimates: $\hat{\phi}_c$ and $\hat{\phi}_s$. These estimates are obtained from the photon-number distributions $P_\delta(n_k|\delta)$ of the observed data from the state discrimination measurement for the relative phases between the input state and LO $\delta = \{0, \pi/2, \pi, 3\pi/2\}$ (see Sec. II of the paper). Under a situation where there is a phase offset ϕ_{off}, interference visibility ξ, dark count rate ν, and $N = 7$, the mean photon numbers of the distributions $P_\delta(n_k|\delta)$ are

$$\langle n \rangle_0 = 2\eta \frac{|\alpha|^2}{N}[1 - \xi \cos(\phi_{\text{off}})] + \nu,$$

$$\langle n \rangle_{\pi/2} = 2\eta \frac{|\alpha|^2}{N}[1 - \xi \sin(\phi_{\text{off}})] + \nu,$$

$$\langle n \rangle_\pi = 2\eta \frac{|\alpha|^2}{N}[1 + \xi \cos(\phi_{\text{off}})] + \nu,$$

$$\langle n \rangle_{3\pi/2} = 2\eta \frac{|\alpha|^2}{N}[1 + \xi \sin(\phi_{\text{off}})] + \nu. \tag{C1}$$

Combining the equations for $\langle n \rangle_0$ and $\langle n \rangle_\pi$ in Eq. (C1) we can obtain samples for the quantity $\cos(\phi_{\text{off}})$ in terms of $\langle n \rangle$, η, $|\alpha|^2$, ξ, ν, and N. In a similar way, samples for $\sin(\phi_{\text{off}})$ can be obtained from the equations for $\langle n \rangle_{\pi/2}$ and $\langle n \rangle_{3\pi/2}$. These samples can be used to obtain an estimate of the expected values from the average over 500 channel transmissions. For a large number of data samples, we expect that the average of $\cos(\phi_c)$ over these channel transmissions approach the cosine of the average $\bar{\phi}_{\text{off}}$ of the actual phase offset ϕ_{off} over these channel transmissions $\cos(\bar{\phi}_{\text{off}})$. We define this average $\bar{\phi}_{\text{off}}$ from the cosine function as the estimate $\hat{\phi}_c$. Similarly, the estimate $\hat{\phi}_s$ is obtained from the samples of $\sin(\phi_{\text{off}})$. These estimates $\hat{\phi}_c$ and $\hat{\phi}_s$ can be expressed in terms of the estimates of the mean photon numbers $\langle n \rangle_\delta$ for 500 channel

transmissions:

$$\langle n \rangle_\pi - \langle n \rangle_0 = C(|\alpha|^2)\cos(\hat{\phi}_c),$$

$$\hat{\phi}_c = \arccos\left[\frac{\langle n \rangle_\pi - \langle n \rangle_0}{C(|\alpha|^2)}\right] \tag{C2}$$

and

$$\langle n \rangle_{3\pi/2} - \langle n \rangle_{\pi/2} = C(|\alpha|^2)\sin(\hat{\phi}_s),$$

$$\hat{\phi}_s = \arcsin\left[\frac{\langle n \rangle_{3\pi/2} - \langle n \rangle_{\pi/2}}{C(|\alpha|^2)}\right] \tag{C3}$$

with

$$C(|\alpha|^2) = f(|\alpha|^2) \times 4|\alpha|^2 \eta \xi /N, \tag{C4}$$

where $f(|\alpha|^2)$ is a factor arising from nonzero probability of error of the discrimination strategy. Here, $N = 7$ is the number of adaptive measurements, η is the detection efficiency, and ξ is visibility of the displacement operation by interference. As a second step, the two initial estimates $(\hat{\phi}_c, \hat{\phi}_s)$ are combined in a weighted average to form a phase estimate $\hat{\phi}_i$ every 500 pulses:

$$\hat{\phi}_i = \text{sign}(\hat{\phi}_s)\frac{|\hat{\phi}_i| + r(|\alpha|^2)|\hat{\phi}_s|}{1 + r(|\alpha|^2)}. \tag{C5}$$

The weight factor $r(|\alpha|^2)$ is used to increase the linearity of the final estimate $\hat{\phi}_{\text{est}}$ as a function of ϕ_{off}, while reducing its variance near the edges of the capture range R in our experiment $R = \pm 0.6$ rad. As a final step, the final estimate $\hat{\phi}_{\text{est}}$ is obtained from the average of N_{avg} estimates $\hat{\phi}_i$ with a gain factor $g(\hat{\phi}_i)$:

$$\hat{\phi}_{\text{est}} = g(\langle \hat{\phi} \rangle)\frac{1}{N_{\text{avg}}} \sum \hat{\phi}_i = g(\langle \hat{\phi} \rangle)\langle \hat{\phi}_i \rangle. \tag{C6}$$

The gain factor $g(\langle \hat{\phi} \rangle)$ depends on the average of the N_{avg} estimates $\{\hat{\phi}\} = \{\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_{N_{\text{avg}}}\}$ and is used to further increase the linearity with respect to the actual phase offset ϕ_{off}, as described below.

To obtain the final phase estimate $\hat{\phi}_{\text{est}}$ of ϕ_{off}, the estimator aims to find the optimal values for the factors $r(|\alpha|^2)$, $f(|\alpha|^2)$, and $g(\hat{\phi}_i)$, which depend on the input mean photon number $|\alpha|^2$, the estimates $\hat{\phi}_i$, and the experimental detection efficiency η, visibility ξ, and dark counts. We find the optimal values of $r(|\alpha|^2)$, $f(|\alpha|^2)$, and $g(\hat{\phi}_i)$ using numerical approaches based on Monte Carlo simulations of the experiment with the following steps:

(1) Find the optimal value of $r(|\alpha|^2)$ [$r_{\text{opt}}(|\alpha|^2)$] that minimizes the difference $|\hat{\phi}_i - \phi_{\text{off}}|$ at the extreme points of the capture range $R = \pm 0.6$ rad.

(2) Given $r_{\text{opt}}(|\alpha|^2)$, find the optimal value of $f(|\alpha|^2)$ [$f_{\text{opt}}(|\alpha|^2)$] by minimizing the χ^2 between the estimated phase $\hat{\phi}_i$ and the actual phase offset ϕ_{off}.

(3) Given $r_{\text{opt}}(|\alpha|^2)$ and $f_{\text{opt}}(|\alpha|^2)$, find the gain parameter $g(\langle \hat{\phi} \rangle)$ that makes the final estimate $\hat{\phi}_{\text{est}}$ as linear as possible with respect to the applied phase offset ϕ_{off}.

This procedure yields the final estimate $\hat{\phi}_{\text{est}}$, and the optimal parameters $r_{\text{opt}}(|\alpha|^2)$, $f_{\text{opt}}(|\alpha|^2)$, and $g(\langle \hat{\phi} \rangle)$ from Monte Carlo simulations that we use for the experimental demonstration of phase tracking for non-Gaussian receivers.
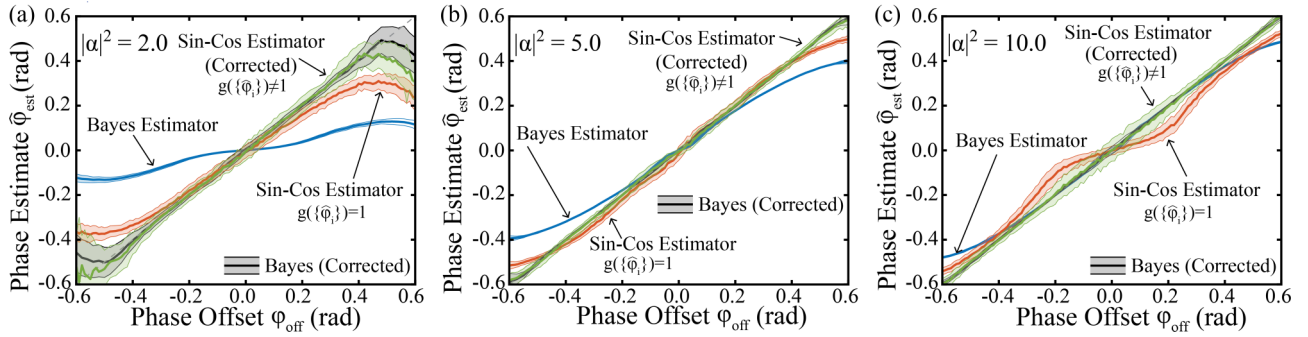
FIG. 6. Phase estimation and estimator performance. (a)–(c) Phase estimates $\hat{\phi}$ as a function of applied phase ϕ_{off} for $|\alpha|^2 = 2.0$, 5.0, and 10.0, respectively, for the sine-cosine (Sin-Cos) estimator $\hat{\phi}_{\text{est}}$ implemented in our demonstration, and for a Bayesian estimator $\hat{\phi}_B$ (blue). The green line shows the corrected $\hat{\phi}_{\text{est}}$ with the optimal gain function $g(\{\hat{\phi}_i\}) \neq 1$. The orange line shows the uncorrected $\hat{\phi}_{\text{est}}$ with $g(\{\hat{\phi}_i\}) = 1$. For the Bayesian estimator, the blue line shows the uncorrected $\hat{\phi}_B$, and the black line shows the corrected $\hat{\phi}_B$. The solid lines represent the average of 100 Monte Carlo samples and shaded regions correspond to one standard deviation for each value of the applied phase offset ϕ_{off}.

Figure 6 shows the final estimate $\hat{\phi}_{\text{est}}$ for mean photon numbers $|\alpha|^2 = 2.0$, $|\alpha|^2 = 5.0$, and $|\alpha|^2 = 10.0$ with optimized gain parameter $g(\langle\hat{\phi}\rangle) \neq 1$ (green) and with $g(\langle\hat{\phi}\rangle) = 1$ (blue). Note that the estimate $\hat{\phi}_{\text{est}}$ with $g(\langle\hat{\phi}\rangle) \neq 1$ shows a closer linear relation with ϕ_{off} compared to the case with $g(\langle\hat{\phi}\rangle) = 1$. Below, we describe the procedure to obtain $r_{\text{opt}}(|\alpha|^2)$, $f_{\text{opt}}(|\alpha|^2)$, and $g(\langle\hat{\phi}\rangle)$.

This procedure reduces the bias of the final estimate $\hat{\phi}_{\text{est}}$ with respect to the true value of the phase offset ϕ_{off}. The estimates $\hat{\phi}_c$ and $\hat{\phi}_s$ are initially biased. $\hat{\phi}_s$ is biased toward zero phase as the magnitude of ϕ_{off} increases. $\hat{\phi}_c$ cannot provide information about the sign of the phase offset ϕ_{off}, and biases the estimates toward positive values. These biases result in a bias of the combined estimator ϕ_i, and $r(|\alpha|^2)$ aims to reduce this bias. The final estimate $\hat{\phi}_{\text{est}}$ is also biased for large values of the phase offsets, as can be seen in Fig. 6, but unbiased for phase offsets near zero. The optimization of the gain function $g(\langle\hat{\phi}\rangle)$ allows for minimizing the overall bias of $\hat{\phi}_{\text{est}}$, which can be mostly suppressed for $|\alpha|^2 \geqslant 5$.

Step 1. Optimal value of $r(|\alpha|^2)$. The parameter $r(|\alpha|^2)$ is a weight factor for the contributions of the initial estimates $\hat{\phi}_c$ and $\hat{\phi}_s$ to the estimate $\hat{\phi}_i$, and its optimal value is chosen to reduce the variance of the final estimate $\hat{\phi}_{\text{est}}$ near the end points of the experimental capture range range $R = \pm 0.6$ rad. The collected data $\{d_j, \delta_j\}$ during the discrimination measurement provide samples for phase estimation which are mostly $\delta_j = 0$. This is because the discrimination strategy is based on hypothesis testing by displacements to the vacuum state [53], and the displaced state spends most of the time in the vacuum state. This means the distribution $P_0(n_k|\delta = 0)$ is populated at a higher rate than $P_{\pi/2}$, P_π, and $P_{3\pi/2}$, and provides more data for the initial estimate $\hat{\phi}_c$ compared to $\hat{\phi}_s$ [see Eqs. (C2) and (C3)]. As a result, $\hat{\phi}_c$ gives a much better estimate with smaller variance. However, $\hat{\phi}_c$ does not give any sign information about the applied phase ϕ_{off}, and is less sensitive to small phase offsets around $\phi_{\text{off}} = 0$. On the contrary, $\hat{\phi}_s$ is more sensitive to small phase offsets and contains the sign information of ϕ_{off}, but is a worse estimate with a much larger variance. The optimal value of $r(|\alpha|^2)$ seeks to balance the contribution of these two initial estimates to minimize $|\hat{\phi}_i - \phi_{\text{off}}|$ at $R = \pm 0.6$ rad. This optimization has

the overall effect of reducing the variance of the final estimate $\hat{\phi}_{\text{est}}$ at these points, where this estimator shows the greatest variance.

To find the optimal value of $r(|\alpha|^2)$ [$r_{\text{opt}}(|\alpha|^2)$] we fix the mean photon number $|\alpha|^2$ and N_{avg}. We use Monte Carlo simulations to obtain the weighted average $\hat{\phi}_i$ with $f(|\alpha|) = 1$ as a function of ϕ_{off} for different values of $r(|\alpha|^2)$. We then obtain a final average for the phase estimate $\langle\hat{\phi}_i\rangle = \sum \hat{\phi}_i/N_{\text{avg}}$ after N_{avg} realizations for applied phases ϕ_{off} within the range $R = \pm 0.6$ rad. We observe that $\langle\hat{\phi}_i\rangle$ is in general a nonlinear function of ϕ_{off}. The optimal $r_{\text{opt}}(|\alpha|^2)$ is obtained by finding the value of $r(|\alpha|^2)$ that minimizes the average of $|\langle\hat{\phi}_i\rangle - \phi_{\text{off}}|$ at $\phi_{\text{off}} = \pm 0.6$ rad. This condition increases the linearity of the final phase estimate $\hat{\phi}_{\text{est}}$ with respect to the applied phase ϕ_{off} and reduces its variance.

We note that $\hat{\phi}_{\text{est}}$ is obtained by multiplying $\langle\hat{\phi}_i\rangle$ by a gain function $g(\langle\hat{\phi}_i\rangle)$. As described in step 3, $g(\langle\hat{\phi}_i\rangle)$ is obtained by inverting the relation of $\langle\hat{\phi}_i\rangle$ as a function of ϕ_{off}. Minimizing $|\langle\hat{\phi}_i\rangle - \phi_{\text{off}}|$ at $\phi_{\text{off}} = \pm 0.6$ rad prevents a large value of the gain $g(\langle\hat{\phi}_i\rangle)$ at these points, which would result in a great increase in the variance of the final estimate $\hat{\phi}_{\text{est}}$. Therefore, determining $r_{\text{opt}}(|\alpha|^2)$ that minimizes $|\hat{\phi}_i - \phi_{\text{off}}|$ at (± 0.6) rad results in a gain $g(\langle\hat{\phi}_i\rangle) \approx 1$, reducing the variance of the final estimate $\hat{\phi}_{\text{est}}$.

Step 2. Optimal value of $f(|\alpha|^2)$. The parameter $f(|\alpha|^2)$ in Eq. (C4) aims to reduce the effect of the nonzero probability of error P_E in the state discrimination measurement. Any state discrimination error causes errors when populating the photon-number distributions $P_\delta(n_k|\delta)$, $\delta = \{0, \pi/2, \pi, 3\pi/2\}$. In a situation with no errors, $P_E = 0$, $f(|\alpha|^2) = 1$ and on average $C(|\alpha|^2) = 4|\alpha|^2\eta\xi/N$. However, when $P_E \neq 0$ the distributions $P_\delta(n_k|\delta)$ are sometimes populated incorrectly, which produces changes in their mean photon numbers $\langle n\rangle_\delta$ in Eqs. (C2) and (C3). This makes $C(|\alpha|^2) \neq 4|\alpha|^2\eta\xi/N$, and $f(|\alpha|^2) \neq 1$.

We note that the distribution $P_\pi(n_k|\delta = \pi)$ can only be incorrectly populated with samples from the other three distributions P_0, $P_{\pi/2}$, and $P_{3\pi/2}$, which have smaller mean photon numbers. As a result, any discrimination error causes the estimated $\langle n\rangle_\pi$ to be smaller than the true value on average. Similarly, the distribution $P_0(n_k|\delta = 0)$ can only be

TABLE I. Optimal values $f_{\mathrm{opt}}(|\alpha|^2)$ and $r_{\mathrm{opt}}(|\alpha|^2)$ for $|\alpha|^2 = 2.0$, 5.0, and 10.0 with $N_{\mathrm{avg}} = 40$, 20, and 4, respectively.

α^2	N_{avg}	$f_{\mathrm{opt}}(|\alpha|^2)$	$r_{\mathrm{opt}}(|\alpha|^2)$
2.0	40	1.210	0.250
5.0	20	0.895	0.333
10.0	4	0.650	0.250

incorrectly populated with samples from distributions $P_{\pi/2}$, P_π, and $P_{3\pi/2}$ with larger mean photon numbers. Then, $P_{\mathrm{E}} \neq 0$ causes the estimated $\langle n \rangle_0$ to increase on average. The overall effect of having $P_{\mathrm{E}} \neq 0$ is to reduce the difference $\langle n \rangle_\pi - \langle n \rangle_0$, such that $\langle n \rangle_\pi - \langle n \rangle_0 < 4|\alpha|^2 \eta \xi / N$. As a result, the estimate $\hat{\phi}_c$ in Eq. (C2) with $f(|\alpha|^2) = 1$ will have a nonzero value when $\phi_{\mathrm{off}} = 0$ that depends on the probability of error P_{E}.

The effect of having $P_{\mathrm{E}} \neq 0$ can be reduced by finding the value of the parameter $f(|\alpha|^2)$ that makes the estimates $\hat{\phi}_c$ and $\langle \hat{\phi}_i \rangle$ close to zero when $\phi_{\mathrm{off}} = 0$. The procedure to find the optimal value $f_{\mathrm{opt}}(|\alpha|^2)$ consists of using Monte Carlo simulations for different values of $f(|\alpha|^2)$ with $r_{\mathrm{opt}}(|\alpha|^2)$ found in step 1. The optimal value $f_{\mathrm{opt}}(|\alpha|^2)$ is the one that minimizes the χ^2 between $\langle \hat{\phi}_i \rangle$ and ϕ_{off}, where a linear dependence is expected. To verify this optimal value we use $\approx 10^6$ Monte Carlo runs with $\phi_{\mathrm{off}} = 0$ to obtain the expected difference $\mathrm{E}[\langle n \rangle_\pi - \langle n \rangle_0 | \phi_{\mathrm{off}} = 0]$. The value $f_{\mathrm{opt}}(|\alpha|^2)$ should then be approximately

$$f_{\mathrm{opt}}(|\alpha|^2) \approx \frac{\mathrm{E}[\langle n \rangle_\pi - \langle n \rangle_0 | \phi_{\mathrm{off}} = 0]}{4|\alpha|^2 \eta \xi / N}. \qquad (C7)$$

Table I shows examples of the optimal values of $r(|\alpha|^2)$ and $f(|\alpha|^2)$ for different mean photon numbers of the input state $|\alpha|^2$ and N_{avg}.

Step 3. Gain function $g(\langle \hat{\phi}_i \rangle)$. The final estimate $\hat{\phi}_{\mathrm{est}}$ is obtained from the product of the average of N_{avg} estimates $\hat{\phi}_i$ and a gain factor $g(\langle \hat{\phi}_i \rangle)$, as shown in Eq. (C6). The gain function $g(\langle \hat{\phi}_i \rangle)$ is solely a function of the average phase estimate $\langle \hat{\phi}_i \rangle$, and ideally maps $\langle \hat{\phi}_i \rangle$ onto the phase offset ϕ_{off} with a linear dependence with unit slope. We obtain the gain function $g(\langle \hat{\phi}_i \rangle)$ by using Monte Carlo simulations with the optimal values of $r_{\mathrm{opt}}(|\alpha|^2)$ and $f_{\mathrm{opt}}(|\alpha|^2)$ to obtain the dependence of the quantity $(\phi_{\mathrm{off}}/\langle \hat{\phi}_i \rangle)$ as a function of $\langle \hat{\phi}_i \rangle$. The quantity $(\phi_{\mathrm{off}}/\langle \hat{\phi}_i \rangle)$ shows in general a nonlinear dependence with $\langle \hat{\phi}_i \rangle$, and this dependence corresponds to $g(\langle \hat{\phi}_i \rangle)$. We fit the quantity $(\phi_{\mathrm{off}}/\langle \hat{\phi}_i \rangle)$ using a smoothing spline, and this spline is defined as the gain function $g(\langle \hat{\phi}_i \rangle)$.

The gain function $g(\langle \hat{\phi}_i \rangle)$ obtained in this way allows to linearize the phase estimator with respect to known applied phase offsets ϕ_{off}. Figure 6 shows $\hat{\phi}_{\mathrm{est}}$ for input mean photon numbers $|\alpha|^2 = 2.0$, 5.0, and 10.0 with optimal values of $g(\langle \hat{\phi}_i \rangle)$ ($\neq 1$) in green; and with $g(\langle \hat{\phi}_i \rangle) = 1$ in orange. In our experimental demonstration, the method for phase tracking is set to generate estimates every $500 \times N_{\mathrm{avg}}$ transmissions through the channel, and subsequently applies a phase correction to the local oscillator every $500 \times N_{\mathrm{avg}}$ shots of the experiment, allowing to perform phase tracking and phase correction at a rate of $f_{\mathrm{PT}} \approx 23/N_{\mathrm{avg}}$ Hz.

2. Bayesian estimator

A second possible estimator of ϕ_{off} based on the collected data (d_j, δ_j) from the state discrimination measurement is the Bayesian estimator. For a Bayes estimator, the photon-number distributions $P_\delta\{n_k|\delta\} = P(n)$ are converted into distributions over the phase $P(\phi|n)$ through Bayes' theorem

$$P(\phi|n)P(n) = \mathcal{L}(n|\phi)P(\phi), \qquad (C8)$$

where $\mathcal{L}(n|\phi)$ are likelihood functions and $P(\phi)$ a prior phase distribution. Given the collected data from the state discrimination measurement $\{\mathrm{data}\} = \{d_j, \delta_j\}$ and assuming some prior distribution $P(\phi)$, a phase estimate can be obtained by forming the posterior probability distribution over phase given by:

$$P(\phi|\{\mathrm{data}\}) = \mathcal{N}P(\phi)\prod_{n=0}^{3}\prod_{m=0}^{3}\mathcal{L}(n|\phi - \delta_m)^{N_{n,m}}, \qquad (C9)$$

where $\mathcal{N}$ is a normalization factor and $N_{n,m}$ is the number of times that n photons were detected given that $\delta_m = \theta_m - \theta_{\mathrm{disc}} = m\pi/2$. Here, θ_{disc} is the answer to the state discrimination problem about the input state. The values of $N_{n,m}$ correspond to elements of the matrix of the photon detections for given $\{\delta_m\}$.

The likelihood function $\mathcal{L}(n|\phi - \delta_m)$ is given by

$$\mathcal{L}(n|\phi - \delta_m) = \frac{\langle n \rangle^n}{n!}e^{-\langle n \rangle}$$

with

$$\langle n \rangle = 2\eta|\alpha|^2[1 - \xi\cos(\phi - \delta_m)] + \nu, \qquad (C10)$$

where η, ξ, and ν are the detection efficiency, interference visibility, and dark counts, respectively. The phase estimate $\hat{\phi}_B$ for the Bayesian estimator is then given by

$$\hat{\phi}_B = \arg\left(\int_{-\pi}^{\pi}e^{i\phi}P(\phi|\{\mathrm{data}\})d\phi\right). \qquad (C11)$$

The Bayesian estimate $\hat{\phi}_B$ provides a more precise estimate of the phase offset ϕ_{off} with smaller variance than the sine-cosine estimate $\hat{\phi}_{\mathrm{est}}$, as shown in Fig. 6. However, this estimator is far more computationally difficult to implement experimentally in real time. While in our current experimental setup such a complex estimator cannot be implemented, the sine-cosine estimator described above produces similar results for estimating ϕ_{off} while remaining computationally inexpensive, and allows for real-time estimation and implementation of the phase-tracking method.

APPENDIX D: ESTIMATOR PERFORMANCE AS A FUNCTION OF N_{avg}

The performance of the phase-tracking method critically depends on the variance of the estimates $\hat{\phi}_{\mathrm{est}}$. In general, increasing the number of samples N_{avg} to obtain an estimate of the applied phase ϕ_{app}[1] improves (reduces) the variance of

---

[1] We denote ϕ_{app} phases that are actually applied in the experimental studies, and we denote ϕ_{off} phases that are used to find theoretical
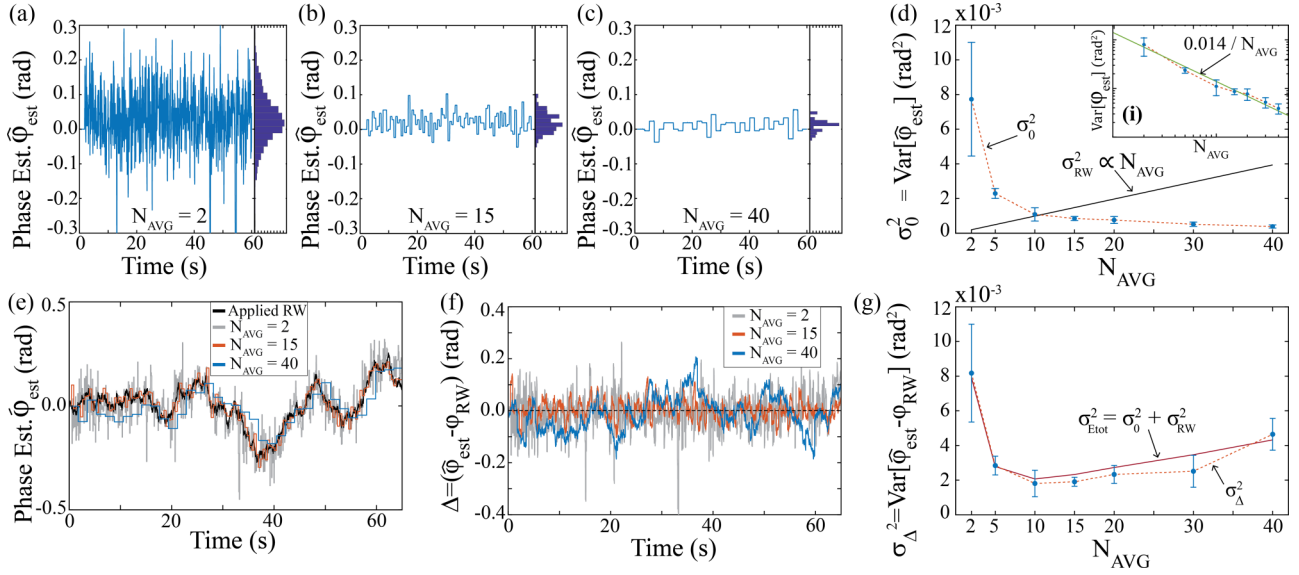
FIG. 7. Estimator variance of the phase-tracking method. Experimentally obtained variance of the estimates $\hat{\phi}_{est}$ without (σ_0^2) and with (Var[$\hat{\phi}_{est} - \phi_{RW}$]) added Gaussian phase noise for $|\alpha|^2 = 5.0$ for different N_{avg} in the algorithm for phase tracking. (a)–(c) Phase-tracking estimates $\hat{\phi}_{est}$ with zero applied phase ($\phi_{app} = 0$) for (a) $N_{avg} = 2$, (b) $N_{avg} = 15$, and (c) $N_{avg} = 40$ as a function of time from $t = 0$ to 60 s. (d) Variance σ_0^2 over 60 s with $\phi_{app} = 0$ (blue points) as a function of N_{avg}. Error bars represent one standard deviation over five different experimental runs. The black solid line shows the expected accumulated variance of random walks σ_{RW}^2 in phase with $\sigma_1 = 5$ mrad, similar to Fig. 4(e) in the main paper, for different estimators with different N_{avg}. Inset (i) shows σ_0^2 on a log-log scale with a linear fit indicating a $1/N_{avg}$ scaling, as expected for statistical uncertainties. (e) Estimated phase for a single applied random walk ($\phi_{app} = \phi_{RW}$) (black line) for three estimators with $N_{avg} = 2$, 15, and 40, in gray, orange, and blue, respectively. (f) Difference between estimated and applied phases $\Delta = \hat{\phi}_{est} - \phi_{RW}$ in (e). (g) Variance of the total estimates σ_Δ^2 for the difference Δ in (f) as a function of N_{avg}. Error bars represent one standard deviation from five different experimental runs. The red solid line shows the expected total variance $\sigma_{Etot}^2 = \sigma_0^2 + \sigma_{RW}^2$ which is composed of σ_0^2 and σ_{RW}^2 from (d), showing good agreement with the observed σ_Δ^2 in the experiment. Note that there is an optimal operation point at $N_{avg} \approx 10$, where the total variance σ_Δ^2 is minimum.

the estimates $\hat{\phi}_{est}$. However, increasing N_{avg} also increases the time required to obtain such estimate, thus reducing the phase-tracking bandwidth f_{PT}. In situations with random Gaussian phase noise with bandwidth $f_{RW} > f_{PT}$, this reduction in f_{PT} can significantly increase probability of error P_E in the state discrimination measurement, and produce estimates that are far less accurate. As a result, there is a tradeoff in the performance of the estimator as a function of N_{avg}. While larger values of N_{avg} provide better estimates for constant phases, in situations where the phase is not constant these estimates may not be accurate, limiting the performance of the phase-tracking method. To investigate this tradeoff, we experimentally study the estimator variance as a function of N_{avg} in situations with Gaussian-distributed random phase noise.

Figures 7(a)–7(c) show the estimates from the experiment using the phase-tracking method as a function of time from $t = 0$ to 60 s with zero applied phase ($\phi_{app} = 0$) for $|\alpha|^2 = 5.0$, for cases with (a) $N_{avg} = 2$, (b) $N_{avg} = 15$, and (c) $N_{avg} = 40$, and the corresponding histograms of estimates. We observe that while smaller N_{avg} increases phase-tracking bandwidth f_{PT}, the variance of the estimates for $\phi_{app} = 0$, denoted as σ_0^2, also increases. On the other hand, larger

N_{avg} reduces f_{PT}, but also reduces σ_0^2. Figure 7(d) shows the variance σ_0^2 for estimators with different N_{avg} (blue points) for five experimental runs with $\phi_{app} = 0$. The inset (i) shows the variance σ_0^2 on a log-log scale with a best-fit line showing good agreement with a $1/N_{avg}$ scaling, which is consistent with the statistical uncertainty for a process with random noise.

For situations with random Gaussian phase noise ϕ_{RW} with variance σ_{RW}^2, the total variance of the estimates will contain contributions from σ_0 and σ_{RW}. The solid black line in Fig. 7(d) shows the expected accumulated variance σ_{RW}^2 for Gaussian random walks between the times to obtain phase estimates, which is proportional to N_{avg}. Then, the total expected variance σ_{Etot}^2 for situations with Gaussian phase noise will be approximately the sum in quadrature of σ_0 (related to the variance of the estimator with different N_{avg}) and σ_{RW}.[2]

Figure 7(e) shows the performance of the estimator for a given applied Gaussian random walk in phase (black line) for $N_{avg} = 2$, 15, and 40, with $\sigma_1 = 5$ mrad (σ_1 as defined

---

values for the optical parameters $r_{opt}(|\alpha|^2)$, $f_{opt}(|\alpha|^2)$, and $g(\langle\phi\rangle)$ for phase tracking.

---

[2]We note that the estimator $\hat{\phi}_{est}$ uses the data from the state discrimination measurement of four coherent states to efficiently estimate a phase offset ϕ_{off}. This task is different from the problem of estimating an unknown phase of a known coherent state probe, for which the fundamental lower bound on the variance is known [2], and the variance of $\hat{\phi}_{est}$ is higher than the fundamental lower bound for that problem [2].

in the main text) from $t = 0$ to 60 s. Figure 7(f) shows the difference between estimated $\hat{\phi}_{est}$ and the applied phase ϕ_{RW}, $\Delta = \hat{\phi}_{est} - \phi_{RW}$ from Fig. 7(e). The variance of Δ will now contain two contributions: one from the estimator with different N_{avg}, ideally given by σ_0^2, and one due to the random walks in phase with variance σ_{RW}^2. For $N_{avg} = 2$ we expect a large variance of the estimates $\hat{\phi}_{est}$, as shown in Fig. 7(a), and due to the relatively high tracking bandwidth f_{PT}, the effect of the random walks is relatively small. On the other hand, for $N_{avg} = 40$, the variance of estimates $\hat{\phi}_{est}$ is expected to be small, as seen in Fig. 7(c), but due to the low f_{PT} relative to f_{RW} ($f_{RW} = 100$ Hz), the accumulated variance from random walks becomes dominant, resulting in larger deviations Δ.

Figure 7(g) shows the total variance σ_Δ^2 of $\Delta = \phi_{est} - \phi_{RW}$ as a function of N_{avg}. Error bars represent one standard deviation over five different random walks. The red solid line shows the expected total variance $\sigma_{Etot}^2 = \sigma_0^2 + \sigma_{RW}^2$, which contains the contributions from the estimator σ_0^2 and the applied random walks σ_{RW}^2. The good agreement between σ_Δ^2 and σ_{Etot}^2 indicates that drifts in the experiment are not significant. We observe that there is an optimal value for $N_{avg} \approx 10$ that minimizes the total variance σ_Δ^2. This optimal value of N_{avg} provides a good phase estimate $\hat{\phi}_{est}$ by increasing the number of samples for parameter estimation, while reducing the effects of errors in state discrimination caused by drifts in phase due to the random walks. We note that optimal values for N_{avg} are larger for smaller values of f_{RW}, and vice versa.

APPENDIX E: PHASE TRACKING WITH DIFFERENT NOISE BANDWIDTHS

Methods for phase tracking should be able to track phase noise with different bandwidths. In general, the performance of any phase-tracking method to track fast noise depends on how fast reliable phase estimates can be generated and how fast correction can be applied [28], which defines the phase-tracking bandwidth f_{PT}. We investigate the performance of the phase-tracking method for the sub-QNL non-Gaussian receiver to track time-varying random phase noise with different bandwidths f_{RW} for $|\alpha|^2 = 5.0$, while keeping the same phase-tracking bandwidth $f_{PT} = f_{expt}/(500N_{avg}) \approx 1.2$ Hz, with $f_{expt} \approx 12$ kHz and $N_{avg} = 20$.

Figure 8(a) shows the probability of error P_E as a function of time from $t = 0$ to 14 s when applying random walks in phase with noise bandwidths $f_{RW} = 100$ Hz (blue) and $f_{RW} = 500$ Hz (green) for two different realizations of 50 random walks. Thick lines show the averages and shaded regions show the spread in P_E over the 50 random walks. In this study, the random walks in phase and the phase tracking are enabled at $t = 2$ s in both cases [see Fig. 8(b) for the applied phase ϕ_{app} and phase estimates $\hat{\phi}_{est}$ for $f_{RW} = 500$ Hz]. We observe that in the presence of phase noise with $f_{RW} = 100$ Hz, phase tracking allows the receiver to maintain a 3.4-dB advantage over the ideal heterodyne limit (Het.), which corresponds to the expected performance without phase noise [53] at $|\alpha|^2 = 5.0$. On the other hand, for phase noise with bandwidth $f_{RW} = 500$ Hz, the average probability of error increases from this ideal case, showing a much larger spread in P_E compared to the case with $f_{RW} = 100$ Hz. Performing phase tracking for non-Gaussian receivers under phase noise
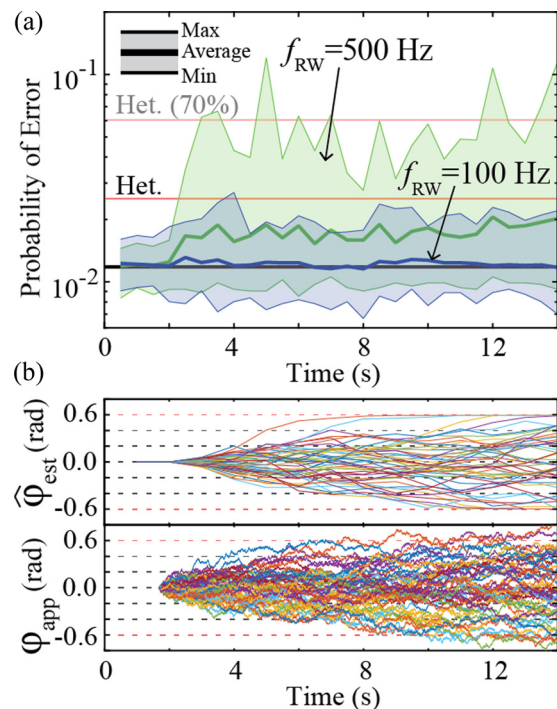

FIG. 8. Phase tracking for different noise bandwidths. (a) Probability of error P_E as a function of time for $|\alpha|^2 = 5.0$ for noise bandwidths $f_{RW} = 100$ Hz (blue) and $f_{RW} = 500$ Hz (green) with different random walks. Thick lines show the averages and shaded regions show the spread in P_E over 50 Gaussian random walks. For $f_{RW} = 100$ Hz, the phase-tracking method allows the non-Gaussian receiver to perform 3.4 dB below the heterodyne limit (Het.), which is the expected performance in the absence of phase noise (black line). Noise with $f_{RW} = 500$ Hz causes P_E to increase compared to $f_{RW} = 100$ Hz. However, phase tracking allows receiver to perform below Het. for 14 s. (b) Applied random walks in phase for $f_{RW} = 500$ Hz (lower) and the phase-tracking estimates (upper). Applied phase $\phi_{app} = \phi_{RW}$ and estimates $\hat{\phi}_{est}$ for $f_{RW} = 100$ Hz are shown in Fig. 4(e) in the main paper over 60 s.

with higher bandwidths f_{RW} requires achieving higher f_{PT} to generate accurate phase estimates $\hat{\phi}_{est}$ at a sufficiently high rate compared to f_{RW}. This can be achieved by increasing f_{expt} resulting in higher sampling rates for phase estimation. Alternatively, increasing $|\alpha|^2$ would reduce errors in state discrimination and increase the accuracy of $\hat{\phi}_{est}$, effectively reducing the effects of low phase-tracking bandwidths f_{PT}. A combination of higher rates f_{PT} and low discrimination errors would make phase tracking reliable for enabling the receiver to maintain discrimination below the heterodyne limit under phase noise with high noise bandwidths.

APPENDIX F: AMPLITUDE NOISE

In addition to phase noise, random amplitude fluctuations may occur in communication channels. Amplitude fluctuations will affect the discrimination strategy and will result in increased discrimination errors if the noise is relatively large. Table II shows the effects of amplitude fluctuations characterized by the standard deviation of the relative amplitude noise σ_{amp} for $|\alpha|^2 = 5$ as a case study, which is

TABLE II. Effect of amplitude noise in error discrimination and phase offset estimation for $|\alpha|^2 = 5$ and $N_{\text{avg}} = 1$, including experimental imperfections.

σ_{amp}	$P_E \times 10^{-2}$	$\sigma^2_{\text{Sin-Cos}} \times 10^{-2}$	$\sigma^2_{\text{Bayes}} \times 10^{-2}$
0.00	1.128	1.523	0.072
0.05	1.171	1.748	0.087
0.10	1.315	2.313	0.160
0.20	1.973	4.720	0.606
0.25	2.460	5.713	0.944

assumed to be Gaussian such that $|\alpha|^2 \rightarrow \mathcal{N}(1, \sigma_{\text{amp}}) \times |\alpha|^2$. This increase in the probability of error for state discrimination causes higher errors in populating the probability distributions $P_\delta(n_k|\delta)$, from which the phase estimator $\hat{\phi}_{\text{est}}$ is formed, and affects the phase-tracking performance, which can be characterized by the estimator variances. Table II

shows the variances for the Sin-Cos and Bayesian estimators for different levels of amplitude noise. We observe that the presence of amplitude noise with levels from $\sigma_{\text{amp}} = 5\%$–$25\%$ has very moderate effects on the error of state discrimination and on the variances of the estimators. This highlights the robustness of the phase-tracking method to amplitude noise.

For situations with slow amplitude noise, it may be possible to perform amplitude estimation and tracking based on the collected data from state discrimination measurement. Specifically, the information contained in the photon-number distributions $P_\delta(n_k|\delta)$ may be sufficient to estimate both phase and amplitude, which can eventually be used for phase and amplitude tracking. The receiver could use the Bayesian estimator for a two-dimensional estimation yielding a simultaneous estimate of phase offset and amplitude of the input state. Alternatively, the receiver could perform phase estimation with the method described here, and sequentially realize amplitude estimation based on Eq. (B1), or vice versa.

[1] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, Phys. Rev. Lett. **92**, 027902 (2004).

[2] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Photonics **5**, 222 (2013).

[3] A. Mari, V. Giovannetti, and A. S. Holevo, Nat. Commun. **5**, 3826 (2014).

[4] P. J. Winzer, J. Lightwave Technol. **30** (2012).

[5] K. Kikuchi, J. Lightwave Technol. **34**, 157 (2016).

[6] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, Opt. Express **16**, 753 (2008).

[7] G. Li, Adv. Opt. Photon. **1**, 279 (2009).

[8] J. M. Arrazola and N. Lütkenhaus, Phys. Rev. A **90**, 042335 (2014).

[9] P. J. Clarke, R. J. Collins, V. Dunjko, E. Andersson, J. Jeffers, and G. S. Buller, Nat. Commun. **3**, 1174 (2012).

[10] F. Xu, J. M. Arrazola, K. Wei, W. Wang, P. Palacios-Avila, C. Feng, S. Sajeed, N. Lutkenhaus, and H.-K. Lo, Nat. Commun. **6**, 8735 (2015).

[11] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).

[12] B. Huttner, N. Imoto, N. Gisin, and T. Mor, Phys. Rev. A **51**, 1863 (1995).

[13] F. Grosshans, G. van Assche, J. Wenger, R. Tualle-Brouri, N. J. Cerf, and P. Grangier, Nature (London) **421**, 238 (2003).

[14] M. Takeoka, S. Guha, and M. M. Wilde, Nat. Commun. **5**, 5235 (2014).

[15] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Nat. Commun. **8**, 15043 (2017).

[16] S. Ghorai, P. Grangier, E. Diamanti, and A. Leverrier, Phys. Rev. X **9**, 021059 (2019).

[17] Y.-L. Tang, H.-L. Yin, Q. Zhao, H. Liu, X.-X. Sun, M.-Q. Huang, W.-J. Zhang, S.-J. Chen, L. Zhang, L.-X. You et al., Phys. Rev. X **6**, 011024 (2016).

[18] M. Sasaki, M. Fujiwara, H. Ishizuka, W. Klaus, K. Wakui, M. Takeoka, S. Miki, T. Yamashita, Z. Wang, A. Tanaka et al., Opt. Express **19**, 10387 (2011).

[19] A. G. Armada and M. Calvo, IEEE Commun. Lett. **2**, 11 (1998).

[20] B. Qi, L.-L. Huang, L. Qian, and H.-K. Lo, Phys. Rev. A **76**, 052323 (2007).

[21] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, Nat. Photonics **7**, 378 (2103).

[22] B. Qi, P. Lougovski, R. Pooser, W. Grice, and M. Bobrek, Phys. Rev. X **5**, 041009 (2015).

[23] D. B. S. Soh, C. Brif, P. J. Coles, N. Lütkenhaus, R. M. Camacho, J. Urayama, and M. Sarovar, Phys. Rev. X **5**, 041010 (2015).

[24] D. Huang, P. Huang, D. Lin, C. Wang, and G. Zeng, Opt. Lett. **40**, 3695 (2015).

[25] A. Marie and R. Alléaume, Phys. Rev. A **95**, 012316 (2017).

[26] J. R. Barry and J. M. Kahn, J. Lightwave Technol. **10** (1992).

[27] D. S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, J. Lightwave Technol. **24**, 12 (2006).

[28] E. Ip and J. M. Kahn, J. Lightwave Technol. **25**, 2675 (2007).

[29] M. Morsy-Osman, Q. Zhuge, L. R. Chen, and D. V. Plant, Opt. Express **19**, B329 (2011).

[30] T. Wang, P. Huang, S. Wang, and G. Zeng, Phys. Rev. A **99**, 022318 (2019).

[31] J. He, R. Norwood, M. Brandt-Pearce, I. Djordjevic, M. Cvijetic, S. Subramaniam, R. Himmelhuber, C. Reynolds, P. Blanche, B. Lynn et al., Comp. Elec. Eng. **40**, 216 (2014).

[32] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton et al., J. Opt. **18**, 063002 (2016).

[33] C. W. Helstrom, Quantum Detection and Estimation Theory, Mathematics in Science and Engineering (Academic, New York, 1976), Vol. 123.

[34] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Rev. Mod. Phys. **84**, 621 (2012).

[35] C. Wittmann, U. L. Andersen, M. Takeoka, D. Sych, and G. Leuchs, Phys. Rev. Lett. **104**, 100505 (2010).

[36] H. M. Wiseman, Phys. Rev. Lett. **75**, 4587 (1995).

[37] H. M. Wiseman and R. B. Killip, Phys. Rev. A **57**, 2169 (1998).

[38] G. M. D'Ariano, M. G. A. Paris, and R. Seno, Phys. Rev. A **54**, 4495 (1996).

[39] G. Chesi, S. Olivares, and M. G. A. Paris, Phys. Rev. A **97**, 032315 (2018).

[40] R. S. Kennedy, A Near-Optimum Receiver for the Binary Coherent State Quantum Channel, MIT Research Laboratory of Electronics Quarterly Progress Report No. 108, 219 (1973).

[41] S. J. Dolinar, An optimum receiver for the binary coherent state quantum channel, research Laboratory of Electronics, MIT, Quarterly Progress Report No. 111, 1973, p. 115 (unpublished).

[42] R. S. Bondurant, Opt. Lett. **18**, 1896 (1993).

[43] R. L. Cook, P. J. Martin, and J. M. Geremia, Nature (London) **446**, 774 (2007).

[44] C. Wittmann, M. Takeoka, K. N. Cassemiro, M. Sasaki, G. Leuchs, and U. L. Andersen, Phys. Rev. Lett. **101**, 210501 (2008).

[45] C. Wittmann, U. L. Andersen, and G. Leuchs, J. Mod. Opt. **57**, 213 (2010).

[46] K. Tsujino, D. Fukuda, G. Fujii, S. Inoue, M. Fujiwara, M. Takeoka, and M. Sasaki, Phys. Rev. Lett. **106**, 250503 (2011).

[47] F. E. Becerra, J. Fan, G. Baumgartner, S. V. Polyakov, J. Goldhar, J. T. Kosloski, and A. Migdall, Phys. Rev. A **84**, 062324 (2011).

[48] C. R. Müller, M. A. Usuga, C. Wittmann, M. Takeoka, C. Marquardt, U. L. Andersen, and G. Leuchs, New J. Phys. **14**, 083009 (2012).

[49] F. E. Becerra, J. Fan, G. Baumgartner, J. Goldhar, J. T. Kosloski, and A. Migdall, Nat. Photonics **7**, 147 (2013).

[50] S. Izumi, M. Takeoka, K. Ema, and M. Sasaki, Phys. Rev. A **87**, 042328 (2013).

[51] R. Nair, S. Guha, and S.-H. Tan, Phys. Rev. A **89**, 032318 (2014).

[52] C. R. Müller and C. Marquardt, New J. Phys. **17**, 032003 (2015).

[53] F. E. Becerra, J. Fan, and A. Migdall, Nat. Photonics **9** (2015).

[54] M. Bina, A. Allevi, M. Bondani, and S. Olivares, Sci. Rep. **6**, 26025 (2016).

[55] A. R. Ferdinand, M. T. DiMario, and F. E. Becerra, npj Quantum Inf. **3**, 43 (2017).

[56] M. T. DiMario, E. Carrasco, R. A. Jackson, and F. E. Becerra, J. Opt. Soc. Am. B **35**, 568 (2018).

[57] M. T. DiMario and F. E. Becerra, Phys. Rev. Lett. **121**, 023603 (2018).

[58] M. T. DiMario, L. Kunz, K. Banaszek, and F. E. Becerra, npj Quantum Inf. **5**, 65 (2019).

[59] S. Guha, Phys. Rev. Lett. **106**, 240502 (2011).

[60] S. Izumi, M. Takeoka, K. Wakui, M. Fujiwara, K. Ema, and M. Sasaki, Phys. Rev. A **94**, 033842 (2016).

[61] M. G. Paris, Phys. Lett. A **217**, 78 (1996).

[62] T. R. Gentile, J. M. Houston, and C. L. Cromer, Appl. Opt. **35**, 4392 (1996).

[63] K. Kikuchi and S. Tsukamoto, J. Lightwave Technol. **26**, 1817 (2008).

[64] G. Xie, A. Dang, and H. Guo, Effects of atmosphere dominated phase fluctuation and intensity scintillation to DPSK system, in *2011 IEEE International Conference on Communications (ICC)* (IEEE, Kyoto, 2011), pp. 1–6.

[65] H. Ghozlan and G. Kramer, On Wiener phase noise channels at high signal-to-noise ratio, in *2013 IEEE International Symposium on Information Theory* (IEEE, Istanbul, 2013), pp. 2279–2283.

[66] M. Reza Khanzadi, Phase noise in communication systems: Modeling, compensation, and performance analysis, Ph.D. thesis, Chalmers University of Technology, Gothenberg, Sweden, 2015.

[67] G. Goldfarb and G. Li, Opt. Express **14**, 8043 (2006).

[68] J. Salz, IEEE Commun. Mag. **24**, 38 (1986).

[69] I. Holzman and Y. Ivry, Adv. Quantum Technol. **2**, 1800058 (2019).

[70] 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 *et al.*, Phys. Rev. Lett. **116**, 240502 (2016).

[71] M. Takeoka, and M. Sasaki, Phys. Rev. A **78**, 022320 (2008).

[72] J. Proakis and M. Salehi, *Digital Communication,* 5th ed. (McGraw-Hill, New York, 2000).