General modeling framework for quantum photodetectors

Steve M. Young, Mohan Sarovar, and François Léonard

Sandia National Laboratories, Livermore, California 94551, USA

Photodetection plays a key role in basic science and technology, with exquisite performance having been achieved down to the single-photon level. Further improvements in photodetectors would open new possibilities across a broad range of scientific disciplines and enable new types of applications. However, it is still unclear what is possible in terms of ultimate performance and what properties are needed for a photodetector to achieve such performance. Here, we present a general modeling framework for photodetectors whereby the photon field, the absorption process, and the amplification process are all treated as one coupled quantum system. The formalism naturally handles field states with single or multiple photons as well as a variety of detector configurations and includes a mathematical definition of ideal photodetector performance. The framework reveals how specific photodetector architectures introduce limitations and tradeoffs for various performance metrics, providing guidance for optimization and design.

I. INTRODUCTION

Photodetectors are used extensively in a broad range of scientific experiments and for numerous technology applications. Pushing the limits of photodetection is important to allow new phenomena to be explored and to improve the performance in applications. In the realm of single-photon detection, records are constantly being reported for detector performance [1–7], and these detectors are being used to increase understanding of the the fundamental properties of light [8–12]. Still, photodetectors are complex, and determining the fundamental limits of their performance and how to design their internal structure is not straightforward because of several simultaneous performance requirements. In the case of single-photon detectors, these performance metrics include efficiency, jitter, dark count rate, number resolution, and bandwidth but also include aspects such as operating temperature, size, and power requirements. Determining the best possible photodetector that can optimize all of these considerations is challenging because many of these are interrelated.

Since the formulation of quantum mechanics, there have been several theoretical models of photodetection. The pioneering work of Glauber [13], Mandel *et al.* [14], and Kelly and Kleiner [15] established the relationship between the counting statistics of pointlike detectors and states of the electromagnetic field, and this theory has provided the foundation for much of the proceeding work. Refinements of the theory by Scully and Lamb [16], Srinivas and Davies [17], and Ueda *et al.* [18] accounted for the backaction of the detection process on the field, which is important to capture the statistics of continuous photocurrents in the limit of weak fields. Further important refinements of photodetection theory include the relaxation of approximations in the field-matter interaction, e.g., the rotating-wave approximation [19,20], and the incorporation of variations in detector architecture, e.g., multiplexed arrays [21].

While the theoretical models and methods resulting from this large body of literature are useful for understanding photodetection phenomena in many contexts, they generally do not provide a framework to design photodetectors from the ground up. Such an endeavor might have been experimentally unfeasible in the past, but with recent progress in nanoscale fabrication and engineering one can now ask the question of how to design an optimal photodetector starting from the atomic scale. To answer this question, one needs a theory that models the dynamics of the electromagnetic field and some general model of the detector's internal degrees of freedom, with as few assumptions as possible. Such a design-oriented approach is essential to establish the ultimate limits of photodetectors and to identify from general principles new optimal designs or perhaps even radically new photodetector designs. For example, recent progress in developing such an approach to modeling photodetectors has shown that in principle there exist no trade-offs between some detector metrics [22,23], suggesting that improved photodetectors are possible.

In this paper, we build on theories for light-matter interaction with weak fields, open quantum systems, and quantum measurement to develop a holistic approach to modeling photodetectors that allows one to directly relate general criteria for performance to internal photodetector structure, and moreover, optimize this internal structure to meet performance metrics. Our formalism allows one to evaluate the state of the matter system during and after interaction with the field as

$$\hat{\rho}_{\text{MATTER}}(t) = \text{Tr}_{\text{LIGHT}}[\mathcal{P}(t, t_0)\hat{\rho}_{\text{TOT}}(t_0)], \tag{1}$$

where $\mathcal{P}$ is an operator determined by the internal structure of the system and its coupling to both the incident field and amplification processes, and $\hat{\rho}_{\text{TOT}}(t_0) = \hat{\rho}_{\text{LIGHT}}(t_0) \otimes \hat{\rho}_{\text{MATTER}}(t_0)$ represents the initial density operator for the combined matter and field quantum state. Tr_{LIGHT} represents a partial trace over the field degrees of freedom. We show how to explicitly calculate $\mathcal{P}$ in a wide variety of cases, and additionally we show that in most cases of interest, this allows
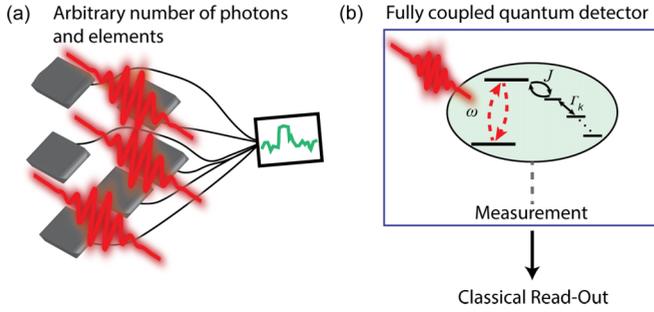
FIG. 1. Illustration of the general photodetection situation being considered. (a) An arbitrary number of photons impinges on an arbitrary number and configuration of elements. An element is an object that creates a signal upon photon absorption. It can consist of a few atoms or macroscopic collections of atoms. (b) A fully coupled quantum detector considers the photon field, the optical absorption, and the measurement as part of one quantum system. The result of the measurement is fed to a classical readout.

us to represent measurement outcome probabilities $\Pi(t)$ as

$$\Pi(t) = \text{Tr}[\mathcal{K}(t, t_0)\hat{\rho}_{\text{TOT}}(t_0)], \qquad (2)$$

which can be used to determine average performance. This abstraction enables us to both intuitively understand detection as propagation of an input pulse to an outgoing signal and analyze the effects of detector internal structure on average photodetector performance. In particular, it naturally furnishes a definition for ideal detection that places conditions on $\mathcal{P}$ (or $\mathcal{K}$)—and, consequently, the detector architecture—that must be satisfied to achieve it, ultimately allowing us to identify new photodetector designs with superior predicted performance.

II. MODELING APPROACH

We seek to develop an approach that can describe a general photodetection system as illustrated in Fig. 1. By a general photodetection system, we mean one that is composed of a number of elements, where the function of each element is to generate a signal upon photon absorption. An element could be composed of a few atoms or could be microscopic in size. For example, a single molecule on a few-atom quantum transport channel could be an element; a macroscopic semiconductor composed of interacting atoms could also be an element. In both cases, one signal is generated. The general photodetection system is composed of many such elements, each interacting with the field and each generating its own signal. The only limitation we impose on the photodetection system is that it is smaller than the photon wavelength.

In Ref. [24], it was shown that the Ito Langevin equation for the interaction of a pointlike (i.e., within a photon wavelength) matter system with a few-photon pulse can be solved by rewriting it as a set of master equations constructed according to the initial incoming field in the Fock basis. In this paper, we consider a single-mode wave packet, which is relevant to detectors in single-mode waveguides, but the approach can be generalized to more complex fields. For a single-field-mode wave packet with frequency ω_E and light-field temporal profile at the detector of $E(t)$, the overall

density matrix for the field is

$$\hat{\rho}_{\text{LIGHT}}(t) = \sum_{N,M} c_{N,M}(t)|N\rangle\langle M|,$$

where $|N\rangle$ is a Fock state of the mode with occupation N; see Appendix A. We note that in the case of specific numerals in place of N and M separating commas will be omitted. For a field initially in a state defined by a set of $c_{N,M}(t_0)$, the reduced density matrix characterizing the matter degrees of freedom at any time t is

$$\hat{\rho}_{\text{MATTER}}(t) = \sum_{N,M} c_{N,M}(t_0)\hat{\varrho}^{N,M}(t),$$

where $\hat{\varrho}^{N,M}(t)$ are a set of auxiliary density matrices obeying the equations

$$\dot{\hat{\varrho}}^{N,M}(t) = \mathcal{V}_{\text{SYS}} + \mathcal{V}_{\text{L-M}} + \mathcal{V}_{\text{AMP}}, \qquad (3)$$

with

$$\mathcal{V}_{\text{SYS}} = -i[\hat{H}, \hat{\varrho}^{N,M}(t)] + \sum_i^{\text{BATHS}} \mathcal{D}[\hat{Y}_i]\hat{\varrho}^{N,M}(t),$$

$$\begin{aligned}\mathcal{V}_{\text{L-M}} = &\sqrt{N}E(t)e^{-i\omega t}[\hat{S}\hat{\varrho}^{N-1,M}, \hat{L}_i^\dagger] \\ &+ \sqrt{M}E^*(t)e^{i\omega t}[\hat{L}_i, \hat{\varrho}^{N,M-1}\hat{S}^\dagger] + \mathcal{D}[\hat{L}]\hat{\varrho}^{N,M} \\ &+ \sqrt{MN}|E(t)|^2(\hat{S}\hat{\varrho}^{N-1,M-1}\hat{S}^\dagger - \hat{\varrho}^{N-1,M-1}),\end{aligned}$$

$$\mathcal{V}_{\text{AMP}} = \sum_i^{\text{AMPS}} \mathcal{D}[(2k_i)^{1/2}\hat{X}_i]\hat{\varrho}^{N,M}(t), \qquad (4)$$

where "BATHS" and "AMPS" refer to sums over the number of baths and amplification channels, respectively. It is straightforward, if cumbersome, to extend this formalism to multiple field modes, including additional spontaneous emission channels.

Equation (3) has three parts: the internal dynamics of the system, the light-matter interaction due to the incoming field excitation, and the amplification of internal states of the detector, modeled as a weak measurement of some internal states. $\mathcal{V}_{\text{SYS}}$ describes the internal evolution according to $\hat{H}$, the noninteracting Hamiltonian, and the influence of external baths $\hat{Y}_i$, described using the Lindblad superoperator

$$\mathcal{D}[\hat{O}]\hat{\rho} = \hat{O}\hat{\rho}\hat{O}^\dagger - \tfrac{1}{2}\hat{O}^\dagger\hat{O}\hat{\rho} - \tfrac{1}{2}\hat{\rho}\hat{O}^\dagger\hat{O}.$$

$\mathcal{V}_{\text{L-M}}$ describes the light-matter interaction, including spontaneous emission, and is mediated by the dipole coupling $\hat{L}$ and the quadratic coupling $\hat{S}$. It is important to note that these terms dictate that the evolution of a given auxiliary density matrix relies on the evolution of auxiliary density matrices interacting with fields with reduced initial excitations. We emphasize that these field indices do not directly correspond to the state of the field at a time t; they correspond to the state of an incoming field before interaction at time t_0, and the density matrices that contain the correct information about the system are the ones that correspond to the actual incoming field state given by $c_{N,M}(t_0)$. So, for an initial field with Fock state in a single mode containing two photons, the density matrix that contains the true dynamics of the system is $\hat{\varrho}^{22}$. The others are auxiliary density matrices required to propagate $\hat{\varrho}^{22}$ and are in some sense fictitious. This is shown schematically in
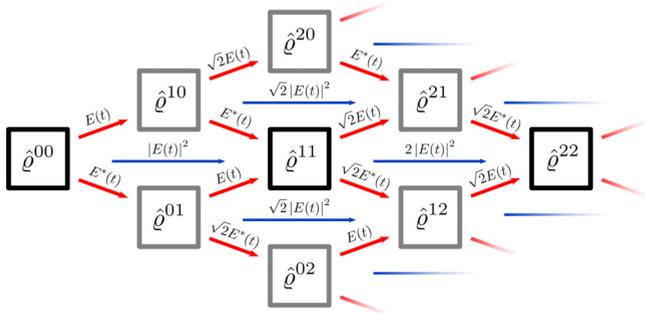
FIG. 2. The relationship between matrices $\hat{\varrho}^{N,M}(t)$ evolved by the system of equations in (3). Each matrix depends on the evolution of matrices with lower superscript index; the real system density matrix is given by the matrix with superscript indices corresponding to the initial state of the incoming field.

Fig. 2. Each diagonal density matrix ($N = M$) is initialized to the state of the system $[\hat{\varrho}^{N,N}(t_0) = \hat{\rho}_{\text{MATTER}}(t_0)]$, and the off-diagonal ones ($N \neq M$) are zero.

The amplification process $\mathcal{V}_{\text{AMPS}}$ is mediated by operators $\hat{X}_i = \sum_j \chi_{ij} \hat{x}_{ij}$, which we constrain to be Hermitian and therefore expressible as sums of projectors onto internal states $\hat{x}_{ij} = |v_j\rangle\langle v_j|$. These operators also influence the dynamics through the Lindbladian. These operators can be interpreted as a weak measurement process with amplification strengths χ_{ij} and rate k_i [25]. However, Eq. (3) as written only describes the averaged dynamics associated with amplification. It may be unraveled [25] into individual measured trajectories by conditioning the dynamics on the prior results of the measurement imposed by the amplification. This is accomplished by adding the term

$$\sum_i^{\text{AMPS}} \frac{dW_{i,t}}{dt}(2k_i)^{1/2}(\hat{X}_i\hat{\varrho}^{N,M} + \hat{\varrho}^{N,M}\hat{X}_i^\dagger - 2\langle\hat{X}_i\rangle\hat{\varrho}^{N,M}) \quad (5)$$

to Eq. (3), where W_i are Wiener processes for each amplification channel that correspond to particular measurement records, whose increments are explicitly

$$dI_i(t) = \langle\hat{X}_i\rangle(t)dt + \frac{1}{(8k)^{1/2}}dW_{i,t}. \quad (6)$$

We assume the Wiener processes across measurement channels are uncorrelated, and for later use define the integrated measurement records

$$I_i(t) = \frac{1}{t_m}\int_{t-t_m}^{t}dt'\left(\langle\hat{X}_i\rangle(t') + \frac{1}{(8k)^{1/2}}\frac{dW_{i,t'}}{dt'}\right), \quad (7)$$

where t_m defines an integration time that sets the temporal resolution of the detector. As an example, $I_i(t)$ could be the time-dependent current measured with an external electronic read-out. We note the appearance of the quantity $\frac{1}{(8k)^{1/2}}\frac{dW_{i,t}}{dt}$, which is noise arising from the continuous measurement process.

The machinery introduced thus far allows for determination of the average and unraveled dynamics of the system, and in principle, computation of measurement outcomes. However, it cannot be written in the form of the direct relations that we desire as expressed by Eqs. (1) and (2). We will now

restructure and extend this machinery so that we may write the system dynamics and performance in this form, considerably enhancing its power and utility.

The system of equations in Eq. (3) is linear (and inhomogeneous), since it may be solved in sequence and the time-dependent field modifies only the input arising from previously solved systems of equations. However, solving these differential equations in this form is inconvenient; they are expressed in terms of rank 4 linear operators, instead of the typical matrix operators. It is therefore preferable to recast these equations by rewriting, e.g., each $\hat{\rho}$ as a vector $\bar{\rho}$, the internal and amplification superoperators as matrices, and the coupling to other density matrices as an input vector. A common scheme for this conversion is to take $\bar{\varrho}_{i\cdot n+j} = \hat{\varrho}_{j,i}$; in this case, superoperators may be converted using the relation $(\hat{O}\hat{\rho}\hat{Q}) = (\hat{Q}^{\mathrm{T}} \otimes \hat{O})\bar{\rho}$, where $\otimes$ is the Kronecker product. However, this transformation is not necessarily optimal, and we will employ others as is convenient. For evaluation of the output records from Eq. (7), the amplification operators $\hat{X}_i$ become $\bar{X}_i = \sum_j \chi_{ij}\bar{x}_{ij}$, where $\bar{x}_{ij}$ is the vectorized form of $\hat{x}_{ij}$ and $\langle\hat{X}_i\rangle(t) = \bar{X}_i \cdot \bar{\rho}_{\text{MATTER}}(t) = \bar{X}_i^\dagger\bar{\rho}_{\text{MATTER}}(t)$. For later convenience, we also define $\hat{x}_i = \sum_j \hat{x}_{ij}$ and its vectorized form $\bar{x}_i$, which are projectors onto the monitored internal states.

Thus, we have for the average dynamics

$$\dot{\bar{\varrho}}^{N,M}(t) = \bar{\mathcal{A}}\bar{\varrho}^{N,M}(t) + \bar{\beta}^{N,M}(t)$$

with

$$\begin{aligned}\bar{\beta}^{N,M}(t) &= |E|^2\sqrt{MN}\bar{\mathcal{S}}\bar{\varrho}^{N-1,M-1}(t) \\ &+ e^{-i\omega t}E(t)\sqrt{M}\bar{\mathcal{L}}^+\bar{\varrho}^{N-1,M}(t) \\ &+ e^{i\omega t}E^*(t)\sqrt{N}\bar{\mathcal{L}}^-\bar{\varrho}^{N,M-1}(t).\end{aligned}$$

Here $\bar{\mathcal{A}}$ contains the coefficients of the matrix elements as written above, and the matrices $\bar{\mathcal{S}}$ and $\bar{\mathcal{L}}^\pm$ contain the field couplings to density matrices with lower mode indices. The use of $+$ and $-$ superscripts in $\bar{\mathcal{L}}^\pm$ is intended only to distinguish the two objects rather than denote any operation. Importantly, $\bar{\mathcal{A}}$ contains the information on the internal structure of the photodetector. In the following, we will assume that $\bar{\mathcal{A}}$ is time independent for simplicity. Assuming the system is in an initial state $\bar{\rho}(t_0)$ before the wave packet arrives, the solution is

$$\bar{\varrho}^{N,M}(t) = \delta_{N,M}e^{\bar{\mathcal{A}}(t-t_0)}\bar{\varrho}^{N,M}(t_0) + \int_{t_0}^{t}d\tau\, e^{\bar{\mathcal{A}}(t-\tau)}\bar{\beta}^{N,M}(\tau). \quad (8)$$

Unless otherwise noted, we will assume that $\bar{\varrho}(t_0)$ is an eigenstate of $\bar{\mathcal{A}}$ (e.g., a ground state), so that $e^{\bar{\mathcal{A}}(t-t_0)}\bar{\varrho}^{N,M}(t_0) = \bar{\varrho}^{N,M}(t_0)$, and take $t_0 = -\infty$ for convenience. Computing $\bar{\varrho}^{N,M}(t)$ iteratively from $\bar{\varrho}^{00}(t_0)$, we can see that our final expression is a set of nested integrals propagating the initial system state through successive interactions with the field. While this may appear similar to a perturbative expansion in an interaction with a semiclassical field, we emphasize that the physical meaning of the series is distinct. The computed evolution is exact, having been derived from the complete evolution of the combined, fully quantized light-matter system. For each interaction, the field is implicitly modified

to account for changes in the field occupation arising from the interaction. Thus, the expansion terminates due the finite occupation of the field, rather than as an approximation. We note that the only approximation made thus far, beyond the treatment of the detector as pointlike, is regarding a timescale separation between the degrees of freedom of the detector that interact with light and the ones that carry the amplified information. This enables us to make the Markov approximation and describe the amplification as a weak measurement. Crucially, no timescale separation is assumed between the field dynamics and initial detector states, as is typical in traditional photodetection theory [13].

Writing $\bar{\mathcal{G}}(t-\tau) = e^{\bar{\mathcal{A}}(t-\tau)}$, we note that $\bar{\mathcal{G}}(t-\tau)$ propagates the input from the field-system interaction at τ to t, and thus defines a set of Green's functions that are independent of the field degrees of freedom. Therefore, they characterize the internal modes and amplification of our detector. In particular, the elements of $\bar{\mathcal{G}}$ will take the form of sums of exponentials with eigenvalues of $\bar{\mathcal{A}}$ as factors of time in the arguments. Thus, the eigenvalues of $\bar{\mathcal{A}}$ play an important role in characterizing the detector performance, as we will see below.

In many cases, $\bar{\mathcal{A}}$, and therefore $\bar{\mathcal{G}}(t)$, will be block diagonal, and $\bar{\mathcal{L}}^{\pm}$ will be block off-diagonal. This occurs when elements coupled by the field interaction are not coupled by internal processes. As a result, states within blocks will only evolve under the action of $\bar{\mathcal{G}}(t)$ within their blocks, while $\bar{\mathcal{L}}^{\pm}$ will map one block of states to another. This can simplify both analysis and solution significantly.

We note that this Green's function formalism can be related to the conventional positive operator valued measure (POVM) formulation of a quantum measurement on the incoming field; see Appendix B.

III. PERFORMANCE

To obtain detector performance metrics, typically the stochastic master equation consistent with Eqs. (8) (this is explicitly given in Appendix C) and the measurement record in Eq. (7) must be numerically integrated. These are stochastic trajectories of the system, and thus one has to resort to Monte Carlo averaging over these trajectories to obtain average detector metrics, which can be expensive and cumbersome. However, in many practical parameter regimes it is possible to exploit the properties of Green's functions to obtain estimates of these metrics using only the average evolution equation, Eq. (8).

In particular, we desire an expression for the probability $\Pi_i(N,t)$ that at time t the ith channel has recorded a hit for an incoming field with N photons. A hit is recorded when the output $I_i(t)$ exceeds a threshold $I_{\mathrm{HIT},i}$. In the strong amplification regime, the stochastic trajectories become jumplike [23], and the signal portion of the measurement current, Eq. (7), dominates. In this case, $\Pi_i(N,t)$ can be estimated as the cumulative probability that the monitored states are populated by the internal dynamics and stay populated long enough to register a hit.

In Appendix C, we present in detail how these probabilities are determined from the average dynamics, which determine the probability of a photoexcitation being transduced into the monitored subsystem, and then the likelihood that the created

population will persist long enough to record a hit. We find that the upper bound for a single detection channel is given by

$$\Pi_i(t) = \bar{x}_i^{\dagger} \bar{\mathcal{G}}(t_{\mathrm{MIN}}) \bar{x}_i \left[\bar{x}_i^{\dagger} \bar{\rho}(t - t_{\mathrm{MIN}}) \right.$$
$$\left. - (\bar{x}_i^{\dagger} \bar{\mathcal{A}} \bar{x}_i) \int_{t_0}^{t - t_{\mathrm{MIN}}} d\tau \, \bar{x}_i^{\dagger} \bar{\rho}(\tau) \right], \quad (9)$$

where t_{MIN} is the minimum time for a detection event to be registered.

If the monitored subsystem is stable and population loss from it can be neglected, $\bar{x}_i^{\dagger} \bar{\mathcal{G}} \bar{x}_i = 1$ and $\bar{x}_i^{\dagger} \bar{\mathcal{A}} \bar{x}_i = 0$. In this case, the above reduces to

$$\Pi_i(t) = \bar{x}_i^{\dagger} \bar{\rho}(t - t_{\mathrm{MIN}}). \quad (10)$$

Thus, in this case, the performance can be directly approximated from the average population in the monitored subsystem, consistent with previous results [23].

A. Efficiency

The total probability that N photons are detected given M incoming photons is $P_N(M,t) = p_N[\Pi_i(M,t)]$, where p_N is a function that maps the outcomes of all output channels into a detection probability. In the case of a single photon, $P_1(1,t) = \sum_{i=1}^{n} \Pi_i(1,t)$, where n is the number of detector elements. In that case, the mapping function p_1 is a simple summation, but for multiple photons it is more complex, as discussed in Appendix C. In general, the total probability of detecting exactly N photons in a field containing N photons is then $P_N(N,\infty)$; in this case, we will often simply write $P_N(N)$.

B. Dark counts

Dark counts for a detector element occur when a hit is recorded in the absence of a field due to total noise exceeding $I_{\mathrm{HIT},i}$. This noise may include thermal fluctuations of the system (i.e., the system has a finite probability of entering the monitored state in the dark) and fundamental noise to the amplification process, as well as noise arising after the amplification due to additional (classical) signal processing and transduction. Here, we consider only the former contributions, both of which are captured in Eq. (7), and ignore the dark counts due to the classical signal processing chain. Since the amplification noise is Gaussian, the dark count rate r_i can be obtained straightforwardly from the amplification and integration time t_m using Eq. (7) as

$$r_i = \frac{\Pi_i(0, t_m + t_0)}{t_m} + \frac{0.5}{t_m} \mathrm{erfc}(2\sqrt{kt_m} \Delta I_{\mathrm{HIT},i})$$

where $\Pi_i(0, t_m + t_0)$ is the probability of obtaining a hit in time t_m due to noise (when no photons are present in the field) and $\Delta I_{\mathrm{HIT},i}$ is the difference in signal between hit and nonhit states of the detector. When amplification noise is the dominating contributor to the dark counts, minimization of the dark count rate requires stronger amplification and/or longer integration times, which in some cases will limit performance. The overall dark count rate for the photodetection system is given by $R_N = p_N(r_i)$, where p_N is the same mapping function as for the efficiency.

C. Jitter and latency

The ultimate limits to jitter and latency in a detector are imposed by the temporal spread of an electromagnetic pulse. While these quantities can be simply defined in the case of a single-photon pulse, more care is required when defining them for multiphoton pulses. We shall define jitter and latency with respect to a temporal distribution determined by the pulse profile $E(t)$,

$$f(t) = N|E(t)|^2 \left[\int_{t_0}^t d\tau |E(\tau)|^2 \right]^{N-1},$$

which we show below to correspond to the behavior of an ideal intensity detector and arrival time it would register for the Nth photon of a multiphoton pulse.

We compare this quantity to the distribution of detection times for the Nth photon in an M photon pulse. Since $P_N(M, t)$ gives the cumulative probability of having registered such a hit at time t, $\dot{P}_N(M, t)\Delta t$ represents the probability of a hit being obtained in a short interval Δt centered at t, which, when normalized to the total probability of a hit, yields a distribution of detection times

$$g(t) = \frac{\dot{P}_N(M, t)}{P_N(M, \infty)}.$$

For example, in the case of a single detection element $i = 1$, this can be expressed as

$$g(t) = \frac{\bar{x}_i^\dagger \bar{\mathcal{G}}(t_{\text{MIN}}) \bar{x}_i \bar{x}_i^\dagger [\dot{\bar{\rho}}(t - t_{\text{MIN}}) - \bar{\mathcal{A}} \bar{x}_i \bar{x}_i^\dagger \bar{\rho}(t - t_{\text{MIN}})]}{\Pi_1(M, \infty)}.$$

The latency in the detection is then defined as the difference between the mean detection time from this distribution and the mean time from $f(t)$, i.e.,

$$\mu = \int_{t_0}^\infty dt\, t[g(t) - f(t)].$$

Similarly, the standard deviation of the detection time for the Nth photon gives the jitter, i.e.,

$$\sigma = \sqrt{\int_{t_0}^\infty dt\, t^2 g(t) - \left(\int_{t_0}^\infty dt\, t g(t) \right)^2}.$$

For convenience, we define σ_{SYS} as the jitter originating from the detector so that

$$\sigma = \sqrt{(\sigma_0)^2 + (\sigma_{\text{SYS}})^2}$$

with

$$\sigma_0 = \sqrt{\int_{t_0}^\infty dt\, t^2 f(t) - \left[\int_{t_0}^\infty dt\, t f(t) \right]^2}.$$

IV. IDEAL DETECTION

The above naturally leads to a definition of ideal detection: a pulse arriving at time t is immediately and fully transduced to a monitored state, such that $g(t) = f(t)$, and thus

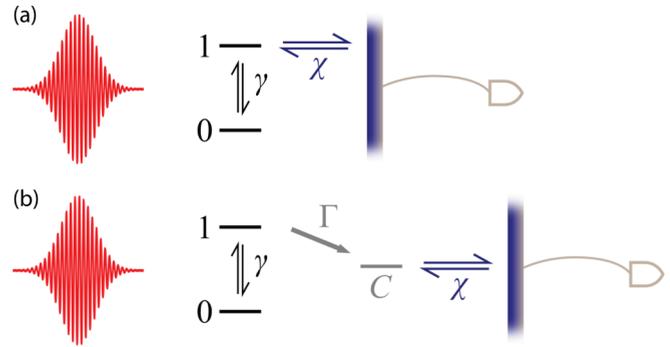$$P_N(N, t) = \left[\int_{-t_0}^t d\tau |E(\tau)|^2 \right]^N. \tag{11}$$

FIG. 3. The two- and three-state systems analyzed in Sec. V. A single mode couples the ground state 0 and excited state 1 with strength γ. (a) In the two-state system, the 1 state is directly amplified with strength χ and rate k. (b) In the three-state system, there is an incoherent decay from state 1 to a state C, which is amplified.

Since $|E(\tau)|^2$ is normalized to 1 (see Appendix A), we have $P_N(N, \infty) = 1$ in this case, corresponding to 100% efficiency. Additionally, since the distribution of the detection times is equivalent to the temporal distribution of the photon(s) in the pulse, $\mu = 0$ and $\sigma = \sigma_0$. Furthermore, one can choose an amplification rate, $k \gg 1/(\chi^2 t_m)$, to make $R_N \approx 0$, and achieve dark count rates that are arbitrarily close to zero.

The above conditions on the metrics can be translated into conditions on the Green's functions governing the dynamics and the underlying architecture. In general, $\bar{\mathcal{G}}$ will have two effects on the signal propagation: One, it may directly attenuate the signal, and two, it may act to alter the shape of the signal that is passed on to the next integral, which will reduce the efficiency. To obtain Eq. (11) from Eq. (10), $\bar{\mathcal{G}}$ must act as a δ function when acting on density vectors diagonal in the field and must be a constant (i.e., only comprise modes with zero eigenvalues) when acting on density vectors off-diagonal. Additionally, the overall magnitude must be unity. This will be shown concretely in the examples.

V. SINGLE-PHOTON, SINGLE-ELEMENT DETECTOR

To illustrate the Green's functions formalism we have developed, we first consider the simplest model for a detector, a two-state system [see Fig. 3(a)]. A single mode couples states 0 and 1 with strength γ, while state 1 is amplified with strength χ and rate k. The matrices governing the dynamics are

$$\hat{L} = \begin{bmatrix} 0 & \gamma \\ 0 & 0 \end{bmatrix}, \quad \hat{X} = \begin{bmatrix} 0 & 0 \\ 0 & \chi \end{bmatrix}, \quad \hat{H} = \begin{bmatrix} 0 & 0 \\ 0 & \omega_1 \end{bmatrix}.$$

Then, with

$$\bar{\rho} = \begin{bmatrix} \hat{\rho}_{00} \\ \hat{\rho}_{11} \\ \hat{\rho}_{01} \\ \hat{\rho}_{10} \end{bmatrix},$$

we have

$$\bar{\mathcal{A}} = \begin{bmatrix} 0 & \gamma^2 & 0 & 0 \\ 0 & -\gamma^2 & 0 & 0 \\ 0 & 0 & -\frac{\gamma^2+2k\chi^2}{2} + i\omega_1 & 0 \\ 0 & 0 & 0 & -\frac{\gamma^2+2k\chi^2}{2} - i\omega_1 \end{bmatrix},$$

which can be exponentiated to obtain

$$\bar{\mathcal{G}}(t) = \begin{bmatrix} 1 & 1 - e^{-\gamma^2 t} & 0 & 0 \\ 0 & e^{-\gamma^2 t} & 0 & 0 \\ 0 & 0 & e^{\left(i\omega_1 - \frac{2k\chi^2+\gamma^2}{2}\right)t} & 0 \\ 0 & 0 & 0 & e^{\left(-i\omega_1 - \frac{2k\chi^2+\gamma^2}{2}\right)t} \end{bmatrix}.$$

The field coupling is described by

$$\bar{\mathcal{L}}^+ = -\bar{\mathcal{L}}^{-\dagger} = \begin{bmatrix} 0 & 0 & 0 & -\gamma \\ 0 & 0 & 0 & \gamma \\ \gamma & -\gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

For an incident resonant single-photon field, the density matrix is given by $\bar{\varrho}^{11}$, and we have

$$\bar{\varrho}^{11}(t) = \int_{-\infty}^{t} d\tau \bar{\mathcal{G}}(t-\tau)\bar{\mathcal{L}}^+ E(\tau) e^{-i\omega\tau}$$
$$\times \int_{-\infty}^{\tau} d\tau' \bar{\mathcal{G}}(\tau-\tau')\bar{\mathcal{L}}^- E^*(\tau') e^{i\omega\tau'} \bar{\rho}(t_0) + \text{H.c.},$$

or explicitly in terms of the population of the excited state,

$$\hat{\varrho}_{11}^{11}(t) = 2\gamma^2 \int_{-\infty}^{t} d\tau e^{-\gamma^2(t-\tau)} E(\tau)$$
$$\times \int_{-\infty}^{\tau} d\tau' e^{-\frac{2k\chi^2+\gamma^2}{2}(\tau-\tau')} E^*(\tau').$$

In order to obtain estimates for the detector metrics, we apply the estimation scheme described by Eq. (9). Using the above definitions for $\bar{\mathcal{G}}(t)$ and $\bar{X}$,

$$P_1(1,t) = e^{-\gamma^2 t_{\text{MIN}}} \int_{-\infty}^{t-t_{\text{MIN}}} d\tau 2\gamma^2 E(\tau)$$
$$\times \int_{-\infty}^{\tau} d\tau' e^{-\frac{2k\chi^2+\gamma^2}{2}(\tau-\tau')} E^*(\tau'). \quad (12)$$

With this, we can compare against the results obtained in Ref. [23] via direct simulation. Using a Gaussian pulse with $\sigma_E = 1$ ns and $\gamma^2 = 1$ ns^{-1}, we compute $P_1(1,\infty)$ using the intervals t_{min} corresponding to thresholds that yield the dark count rate in that work (i.e., a probability of 1% for obtaining a dark count in a 15-ns interval) and plot it against the efficiency obtained from direct simulation (Fig. 4). In this case, the estimation scheme provides a good approximation of the efficiency. The estimation deviates more for low $(2k)^{1/2}\chi$, which is expected: The approximation of the dynamics being jumplike is less valid, the contribution of noise is increased, and the longer integration times increase the likelihood that a direct simulation signal will be split across two integration windows and fail to produce a hit, an event that is not captured by the estimation scheme as presented. We also plot the
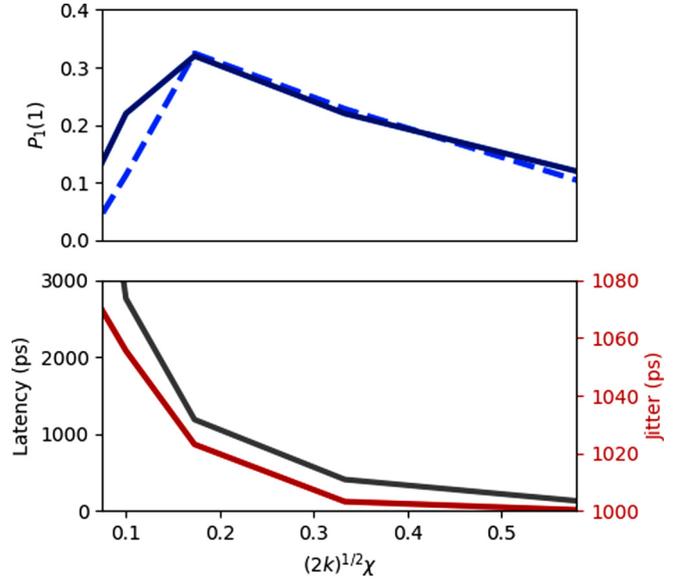

FIG. 4. The estimated efficiency vs simulated efficiency (upper panel) and the estimated jitter and latency (lower panel) for the system in Fig. 3(a). The simulation results are taken from Ref. [23] and the estimation is performed using the same parameters and Eq. (11). The estimated quantity deviates from the calculated quantity in the small $(2k)^{1/2}\chi$ region because the approximations that go into the estimated quantity (i.e., jumplike trajectories) are not valid in this very weak measurement regime.

estimated jitter and latency, shown in the lower panel. These are reduced for stronger amplification, consistent with the trends observed in Ref. [23].

We next consider the system containing three states: 0, 1, and C [Fig. 3(b)]. A single, resonant mode couples 0 and 1 with strength γ, while state 1 decays incoherently to state C according to Γ, which is amplified according to χ and k. We previously demonstrated that this configuration can lead to ideal detection with 100% efficiency, negligible dark counts, and minimal jitter, provided that certain conditions are met for γ and Γ [23].

For this system, we have

$$\hat{L} = \begin{bmatrix} 0 & \gamma & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \hat{Y} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \Gamma & 0 \end{bmatrix},$$

$$\hat{X} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \chi \end{bmatrix}, \quad \hat{H} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \omega_1 & 0 \\ 0 & 0 & \omega_C \end{bmatrix}.$$

We construct the vectorized density matrix as

$$\bar{\rho} = \begin{bmatrix} \hat{\rho}_{00} \\ \hat{\rho}_{11} \\ \hat{\rho}_{CC} \\ \hat{\rho}_{01} \\ \hat{\rho}_{10} \end{bmatrix}.$$

Coherences with the C state are omitted as they will always be zero for the given interactions and initial state $\hat{\rho}(t_0) = |0\rangle\langle 0|$.

Then,

$$\bar{\mathcal{A}} = \begin{bmatrix} 0 & \gamma^2 & 0 & 0 & 0 \\ 0 & -(\gamma^2+\Gamma^2) & 0 & 0 & 0 \\ 0 & \Gamma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{\gamma^2+\Gamma^2}{2}+i\omega_1 & 0 \\ 0 & 0 & 0 & 0 & -\frac{\gamma^2+\Gamma^2}{2}-i\omega_1 \end{bmatrix}$$

and

$$\bar{\mathcal{L}}^+ = -\bar{\mathcal{L}}^{-\dagger} = \begin{bmatrix} 0 & 0 & 0 & 0 & -\gamma \\ 0 & 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & 0 & 0 \\ \gamma & -\gamma & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This leads to

$$\bar{\mathcal{G}}(t) = \begin{bmatrix} 1 & \gamma^2 \frac{1-e^{-(\Gamma^2+\gamma^2)t}}{\Gamma^2+\gamma^2} & 0 & 0 & 0 \\ 0 & e^{-(\Gamma^2+\gamma^2)t} & 0 & 0 & 0 \\ 0 & \Gamma^2 \frac{1-e^{-(\Gamma^2+\gamma^2)t}}{\Gamma^2+\gamma^2} & 1 & 0 & 0 \\ 0 & 0 & 0 & e^{\left(i\omega_1-\frac{\Gamma^2+\gamma^2}{2}\right)t} & 0 \\ 0 & 0 & 0 & 0 & e^{\left(-i\omega_1-\frac{\Gamma^2+\gamma^2}{2}\right)t} \end{bmatrix}.$$

We note two features of the above matrices: First, the top 3×3 diagonal block of $\bar{\mathcal{A}}$ only has a single nonzero eigenvalue, a decay mode, and second, none of these matrices have a dependence on the overall amplification $(2k)^{1/2}\chi$. The separation of the unmonitored and monitored subspaces by an incoherent decay, and the fact that there are only incoherent processes within the monitored subspace implies that the amplification does not directly effect the dynamics.

For a single photon at the resonant field frequency and a system initially in state 0, we find that the probability of a hit, which is equal to the population in the C state in this case, is

$$P_1(1,t) = \hat{\varrho}_{CC}^{11}(t) = 2\frac{\Gamma^2\gamma^2}{\gamma^2+\Gamma^2}$$
$$\times \int_{-\infty}^{t} d\tau (1 - e^{-(\gamma^2+\Gamma^2)(t-\tau)})E(\tau)$$
$$\times \int_{-\infty}^{\tau} d\tau' e^{-\left(\frac{\gamma^2+\Gamma^2}{2}\right)(\tau-\tau')}E^*(\tau'). \quad (13)$$

When $\gamma^2 + \Gamma^2 \gg 1/\sigma_E$, where σ_E is the width of a smooth pulse described by $E(t)$, the exponential decay can be approximated by a δ function. Then,

$$P_1(1,t) = 2\frac{\gamma^2\Gamma^2}{\gamma^2+\Gamma^2}\int_{-\infty}^{t} d\tau |E(\tau)|^2 \frac{2}{\gamma^2+\Gamma^2},$$
$$P_1(1) = 4\frac{\gamma^2\Gamma^2}{(\gamma^2+\Gamma^2)^2}.$$

This expression satisfies Eq. (11) except for the prefactor. It is evident by inspection that this prefactor is unity when $\gamma = \Gamma$. Thus, Eq. (11) can be fully satisfied and perfect efficiency is achieved when the two rates are equal; the large coupling limit ($\gamma^2 + \Gamma^2 \gg 1/\sigma_E$) ensures that the signal is not distorted as

it is processed by the system, while the coupling matching condition ($\gamma = \Gamma$) ensures the photon is fully converted to population in C. We emphasize that as long as the pulse is sufficiently long, this result is obtained regardless of pulse shape.

We also note that the fast incoherent process (large Γ) will widen the detection bandwidth and decrease sensitivity to the resonance condition. Repeating the above calculation for an off-resonant pulse, we obtain

$$P_1(1,t) = 2\frac{\Gamma^2}{\gamma^2+\Gamma^2}\int_{-\infty}^{t} d\tau E(\tau)\int_{-\infty}^{\tau} d\tau'\gamma^2$$
$$\times \cos[\Delta\omega(\tau - \tau')]e^{-\left(\frac{\gamma^2+\Gamma^2}{2}\right)(\tau-\tau')}E^*(\tau'),$$
$$P_1(1) = 4\frac{\gamma^2\Gamma^2}{(\gamma^2+\Gamma^2)^2+4\Delta\omega^2},$$

where $\Delta\omega$ is the detuning from resonance. The bandwidth of the detector is therefore determined by the rate $\gamma^2 + \Gamma^2$. The requirement that the pulse be temporally wide can also be understood as requiring that the pulse frequency distribution be narrow compared to this bandwidth, such that it is approximately constant over the frequencies in the pulse.

It is also evident from this calculation that any additional sources of decoherence, such as amplification or decays to unmonitored states, will reduce the efficiency; they will appear in the denominator of the prefactor and the maximum efficiency will no longer be unity. For example, suppose that we have additional dephasing of the excited state, modeled by the Lindblad operator $\hat{Y}_1 = \kappa|1\rangle\langle 1|$. This will inhibit the formation of optical coherence but will not affect the populations. Adding this dynamics gives

$$P_1(1,t) = 2\frac{\Gamma^2}{\gamma^2+\Gamma^2}\int_{-\infty}^{t} d\tau E(\tau)\int_{-\infty}^{\tau} d\tau'\gamma^2$$
$$\times e^{-\left(\frac{\gamma^2+\Gamma^2+\kappa^2}{2}\right)(\tau-\tau')}E^*(\tau'),$$
$$P_1(1) = 4\frac{\gamma^2\Gamma^2}{(\gamma^2+\Gamma^2)^2}\frac{\gamma^2+\Gamma^2}{\gamma^2+\Gamma^2+\kappa^2}.$$

These expressions indicate that the efficiency is the efficiency obtained without the dephasing dynamics multiplied by an additional factor dependent on κ. This factor—and the total efficiency—will always be less than 1.

Our formalism also reveals the critical role of the zero eigenvalue mode of $\bar{\mathcal{A}}$ in the dynamics; the presence of a zero eigenvalue indicates a component of the population that is persistent; the state will be populated indefinitely. Decays or oscillations would prohibit an expression in the form of Eq. (11) from being achieved. For example, suppose that population decays from the C state to the 0 state at a rate δ^2, so that the detector effectively resets after time $\approx 1/\delta^2$. Then, even with $\gamma = \Gamma$,

$$P_1(1,t) = \int_{-\infty}^{t} d\tau e^{-\delta^2 t_{\text{MIN}}}|E(\tau - t_{\text{MIN}})|^2,$$
$$P_1(1) = e^{-\delta^2 t_{\text{MIN}}}, \quad (14)$$

so that the overall efficiency is limited by the possibility that population in C decays before sufficient time to record a hit

has passed. Maximizing efficiency requires both separation of the amplification from the absorption and propagation of the signal into such stationary system mode where it can be amplified at leisure, i.e., shelving. However, this extends the natural reset time, limiting count rate in the absence of an active reset mechanism.

In practice, the time required for the system to record a hit places additional constraints on the amplification $(2k)^{1/2}\chi$. The maximum threshold that can produce a hit is $\Delta I_{\mathrm{HIT}} = \chi$. For a desired dark count rate R_1, then $(2k)^{1/2}\chi \geqslant \frac{\mathrm{erfc}^{-1}(2t_m R_1)}{\sqrt{2t_m}}$. This is especially relevant if relaxation from the C state to the ground state is present. Maximizing the efficiency in Eq. (14) requires increasing the amplification so that t_m may be reduced, or reducing the signal threshold and necessarily permitting more dark counts.

VI. OTHER DETECTOR STRUCTURES

To demonstrate the broad applicability of the formalism, we now apply it to several cases with other detection mechanisms or detector configurations.

A. Two-state quadratic coupling detector

In some cases, the matter system couples to the field quadratically; the interaction does not alter the photon number but the phase of the field. For example, in circuit-QED architectures in the dispersive regime, the interaction between matter and field takes the form $H_I = a^\dagger a \hat{\sigma}_z$ [26]. In this case, the natural way to model the field-matter interaction within our framework is to couple through $\hat{S}$ rather than $\hat{L}$ in Eq. (4). Instead of creating optical coherence, the field scatters off the matter subsystem and causes a change. Consider a two-level system such as that in Fig. 3(a), with the field scattering causing a general unitary operation on the matter degrees of freedom. Then, the system is described by the operators

$$\hat{S} = e^{-i\left(a_x\hat{\sigma}_x + a_y\hat{\sigma}_y + a_z\hat{\sigma}_z\right)\theta/2}, \quad \hat{H} = \begin{bmatrix} 0 & 0 \\ 0 & \omega_1 \end{bmatrix},$$

$$\hat{X} = \begin{bmatrix} 0 & 0 \\ 0 & \chi \end{bmatrix},$$

where $\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z$ are Pauli matrices and $a_x^2 + a_y^2 + a_z^2 = 1$. This gives

$$\bar{\mathcal{G}}(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\omega_1 t} & 0 & 0 \\ 0 & 0 & e^{i\omega_1 t} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\bar{S} = \sin^2(\theta/2)(\hat{\alpha} \otimes \hat{\alpha}^{\mathrm{T}} - \mathbf{I})$$
$$+ i\sin(\theta/2)\cos(\theta/2)(\hat{\alpha} \otimes \mathbf{I} - \mathbf{I} \otimes \hat{\alpha}^{\mathrm{T}}),$$

where $\hat{\alpha} = (a_x\hat{\sigma}_x + a_y\hat{\sigma}_y + a_z\hat{\sigma}_z)$. This yields

$$P_1(1, t) = \bar{\varrho}_{11}^{11} = \sin^2(\theta/2)\int_{-\infty}^{t-t_{\mathrm{MIN}}} \left(1 - a_z^2\right)d\tau |E(\tau)|^2,$$

so ideal detection is achieved when $a_z = 0$ and $\theta = \pi$. The quadratic field coupling does not produce a field-driven decay and can directly transfer population. As a consequence,
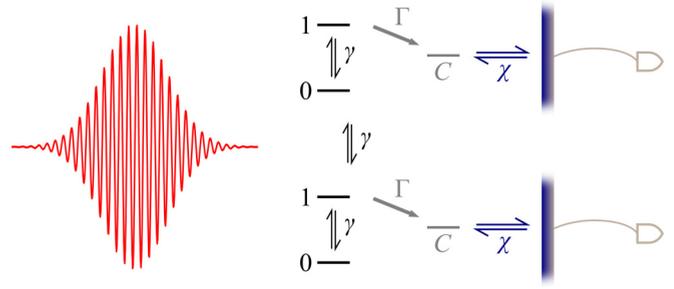

FIG. 5. System comprising two elements described by Fig. 3(b). The elements collectively interact with the field and may exchange population through the field. In systems with more elements, the field couples each element with each other element.

there are no losses and the requirement for ideal detection is straightforwardly satisfied without requiring a shelving state. Even when ideal efficiency is not obtained, the quadratic coupling ensures that $\dot{P}_1(1, t) = |E(t - t_{\mathrm{MIN}})|^2$, meaning jitter is minimized and latency is limited by the integration time of amplification.

B. Single photon, multiple degenerate elements

We next consider n copies of the elements of Fig. 3(b) collectively interacting with the incoming light pulse, illustrated for two elements in Fig. 5. Each absorbing element is assumed to be degenerate in the sense that all elements have the same transition energy between states 0 and 1, and each absorbing element has its own associated readout channel. The system Hilbert space is the direct product of n three-state Hilbert spaces. There is a single operator coupling the system with the field and a bath coupled to each subsystem such that

$$\hat{L} = \sum_i \gamma |0_i\rangle\langle 1_i|, \quad \hat{Y}_i = \Gamma |C_i\rangle\langle 1_i|,$$

where 1_i and C_i denote the 1 and C states of the ith element. As a result, the field introduces coupling between array elements whereas the baths $\hat{Y}_i$ do not; following the procedure introduced in Sec. II, we find that

$$P_1(1, t) = \sum_{i=1}^n \Pi_i(1, t) = \sum_i \bar{x}_i \cdot \bar{\varrho}^{11}(t)$$

$$= 2n\frac{\gamma^2\Gamma^2}{n\gamma^2 + \Gamma^2}\int_{-\infty}^t d\tau (1 - e^{-(n\gamma^2 + \Gamma^2)(t-\tau')})$$

$$\times E(\tau)\int_{-\infty}^\tau d\tau' e^{-(\frac{n\gamma^2 + \Gamma^2}{2})(\tau-\tau')}E^*(\tau').$$

This is the same expression as for the single-element detector with $\gamma \to \sqrt{n}\gamma$, indicating that a multiple-copy detector can be made equivalent to a single-copy detector with stronger field coupling, and the ideal detector is now characterized by $\Gamma = \sqrt{n}\gamma$. The presence of multiple degenerate elements enhances the effective field coupling, a phenomenon recognized as superradiance [27].

It is important to note that the need to monitor multiple states may introduce practical constraints on detector performance. In particular, there are three main amplification schemes that are consistent with the above description: (i) a
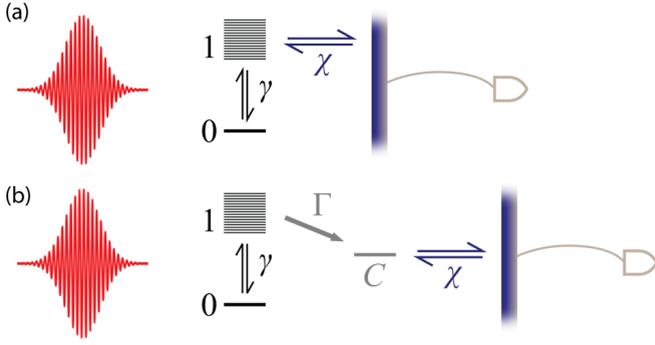
FIG. 6. Systems featuring excitation into a dispersive band. A single mode couples the ground state 0 and all excited states with strength γ. In panel (a), the states in the band are directly amplified with strength χ and rate k, while in panel (b) the excited band states incoherently decay to a state (or potentially multiple states) C, which are amplified.

single amplification process sensitive to all the C states, corresponding to a single $\hat{X} = \chi \sum_i^n |C_i\rangle\langle C_i|$, (ii) an amplification process monitoring each C state, so that there are n separate $\hat{X}_i = \chi |C_i\rangle\langle C_i|$, all of which are processed to determine hits individually, and (iii) amplification processes for each C state, the combined signal of which is processed to determine whether a hit has occurred. For a single photon, both (i) and (ii) result in the same dark count rate for given amplifications and times, since the hits are determined based on the noise provided by a single amplification process. However, for (iii), the noise is the total contributed by all amplification processes and will be larger than the single-process noise by a factor of $\sqrt{n}$, increasing the difficulty of discriminating hits and dark counts.

C. Single-photon, dispersive band

In most cases, the detector will contain multiple absorbing elements that are not degenerate, resulting in a band of possible excitations from the ground state. This is the case in ensembles of atoms or solid-state systems. It is therefore essential to understand how the above results can be extended to such systems.

We consider systems like those appearing in Fig. 6, which have a single ground state that may be excited with strength γ to n independent states with energies ω distributed according to a normalized function $D(\omega)$. First, we consider the case with no amplification dynamics in order to isolate the effect of having a band of states that are coupled to the field.

In the limit $n \to \infty$, we treat the distribution of excited states as a continuous density, and in the case where the distribution of these states is Lorentzian, we can analytically compute this excited-state population. For a Lorentzian with full-width-half-maximum $\zeta^2/2$ and centered at the incoming pulse frequency ω_E, the state distribution is

$$D(\omega) = \frac{1}{\pi} \frac{\zeta^2/2}{(\zeta^2/2)^2 + (\omega - \omega_E)^2},$$

and the total population of the band is

$$u(t) = \sum_i \bar{\varrho}_{ii}^{11}(t) = \frac{2n\gamma^2}{n\gamma^2 + \zeta^2}$$

$$\times \int^t d\tau (n\gamma^2 e^{-(n\gamma^2+\zeta^2)(t-\tau)} + \zeta^2) E(\tau)$$

$$\times \int^\tau d\tau' e^{-\frac{n\gamma^2+\zeta^2}{2}(\tau-\tau')} E^*(\tau'),$$

where i indexes the states in the band.

The form of this expression is that of Eq. (13) and the Green's functions have the required form for ideal detection. The energy dispersion results in dephasing that suppresses emission, allowing the excitation into the band to be long lived. From the above expression, it is clear that the bandwidth can be tuned to obtain full collection of the photon when $\zeta^2 = n\gamma^2 \gg \sigma_0$. However, this does not necessarily indicate ideal performance; unlike the case of incoherent decay to a shelving state, amplification will influence the absorption process with consequences for performance. If the excited states are directly amplified according to $\hat{X} = \chi \sum_i^n |i\rangle\langle i|$ with rate k, then we can write

$$P_1(1,t) = u(t) = \frac{2n\gamma^2}{n\gamma^2 + \zeta^2}$$

$$\times \int^t d\tau (n\gamma^2 e^{-(n\gamma^2+\zeta^2)(t-\tau)} + \zeta^2) E(\tau)$$

$$\times \int^\tau d\tau' e^{-\frac{n\gamma^2+\zeta^2+2k\chi^2}{2}(\tau-\tau')} E^*(\tau')$$

and in the limit $\zeta^2 = n\gamma^2 \gg \sigma_0$

$$P_1(1,t) = \frac{4n\gamma^2\zeta^2}{(n\gamma^2+\zeta^2)^2} \times \frac{n\gamma^2 + \zeta^2}{n\gamma^2 + \zeta^2 + 2k\chi^2}.$$

The amplification introduces decoherence that suppresses absorption and will limit efficiency, i.e., the Zeno effect; see Ref. [23]. However, unlike the two-state system, the band shelving-like effect allows for long measurement intervals and correspondingly lower amplifications; this suggests that higher efficiencies may be obtained at the expense of increased latency.

We may circumvent such tradeoffs as before by introducing incoherent decays from each state in the band to states that are not optically coupled (dark states) that are then amplified. It is straightforward to alter the above treatment to account for the additional dynamics. For a decay rate Γ^2 to a single state C (or multiple C states), the band population becomes

$$u(t) = \int^t d\tau e^{-\Gamma^2(t-\tau)} \frac{n\gamma^2 e^{-(n\gamma^2+\zeta^2)(t-\tau)} + \zeta^2}{n\gamma^2 + \zeta^2}$$

$$\times 2n\gamma E(\tau) \int^\tau d\tau' e^{-\frac{n\gamma^2+\zeta^2+\Gamma^2}{2}(\tau-\tau')} \gamma E^*(\tau')$$

and

$$P_1(1,t) = \bar{\varrho}_{CC}(t) = \Gamma^2 \int^t d\tau u(\tau)$$

$$= \int^t d\tau \frac{\Gamma^2}{n\gamma^2 + \zeta^2} \left[\frac{n\gamma^2 (1 - e^{-(n\gamma^2+\zeta^2+\Gamma^2)(t-\tau)})}{n\gamma^2 + \zeta^2 + \Gamma^2} \right.$$

$$+ \frac{\zeta^2(1 - e^{-\Gamma^2(t-\tau)})}{\Gamma^2}\Bigg]$$

$$\times 2n\gamma E(\tau) \int^\tau d\tau' e^{-\frac{n\gamma^2+\zeta^2+\Gamma^2}{2}(\tau-\tau')} \gamma E^*(\tau').$$

When $\Gamma^2 + \zeta^2 + n\gamma^2 \gg \sigma_0$, after the pulse passes

$$P_1(1) = 4\frac{n\gamma^2(\Gamma^2+\zeta^2)}{(n\gamma^2+\zeta^2+\Gamma^2)^2},$$

which is unity when $\Gamma^2 + \zeta^2 = n\gamma^2$. Thus, in the presence of a band, the optimal tuning of the decay rate Γ^2 has contributions from the bandwidth as well as the optical coupling. However, the jitter and latency both may still be impacted. For the above parameters,

$$P_1(1, t) = \int^t d\tau \left[1 - \frac{e^{-\Gamma^2(t-\tau)}}{2 + \Gamma^2/\zeta^2} \right] |E(\tau)|^2.$$

If Γ is small, the second term in brackets is only zero at long times, which delays and stretches the distribution of arrival times. This adds latency

$$\mu = \frac{1}{\Gamma^2(2 + \Gamma^2/\zeta^2)}$$

and increases the jitter by

$$\sigma_{\text{SYS}} = \mu\sqrt{3 + 2\Gamma^2/\zeta^2}.$$

D. Multiple photons, multielement detector

We can now ask how an ideal single-photon detector performs in resolving photon number. For the array of degenerate elements, we note that in the limit where $n\gamma^2 + \Gamma^2 \gg 1/\sigma_0$, the populations of the 1 states and coherences between them are close to 0. Thus, for two photons, we have

$$\bar{\varrho}_{n:C_iC_iC_jC_j}^{22}(t) \approx y_{n-1}(t)\bar{\varrho}_{n:C_iC_i}^{11},$$

where

$$y_n(t) = \frac{4\gamma^2\Gamma^2}{(n\gamma^2+\Gamma^2)^2} \int_{-\infty}^t d\tau |E(\tau)|^2.$$

Here $\bar{\varrho}_{n:C_iC_iC_jC_j}^{22}$ and $\bar{\varrho}_{n:C_iC_i}^{11}$ represent the total system density matrix in a compact notation; see Appendix D for details. We can express $\bar{\varrho}^{11}(t)$ as

$$\varrho_n^{11}(t) = \bar{\varrho}_{n-1}^{00}(t) \otimes \begin{bmatrix} 1 - ny_n(t) \\ 0 \\ y_n(t) \\ 0 \\ 0 \end{bmatrix}$$

$$= \bar{\rho}_{n-1}(t_0) \otimes \begin{bmatrix} 1 - ny_n(t) \\ 0 \\ y_n(t) \\ 0 \\ 0 \end{bmatrix},$$

and thus for a system illuminated by a two-photon pulse

$$\bar{\varrho}_n^{22} = \bar{\rho}_n(t_0) + \left[\int_{-\infty}^t d\tau e^{\bar{A}(t-\tau)} \sqrt{2} E(\tau) e^{-i\omega\tau} \bar{\mathcal{L}}^+ \int_{-\infty}^\tau d\tau' e^{\bar{A}(\tau-\tau')} \sqrt{2} E^*(\tau') e^{i\omega\tau'} \bar{\mathcal{L}}^- \bar{\varrho}_n^{11}(\tau') + \text{H.c.} \right].$$

We note that in order to determine photon number it is necessary for the amplification scheme to distinguish the states with different numbers of C states occupied. At present, we take each C_i state to be amplified by an operator $\hat{X}_i = \chi|C_i\rangle\langle C_i|$, so that, as discussed in Appendix C,

$$P_2(2, t) = \sum_{i>j}^n (\bar{x}_i \times \bar{x}_j) \cdot \bar{\varrho}_n^{22}(t) = \frac{8(n-1)\gamma^2\Gamma^2}{[(n-1)\gamma^2+\Gamma^2]^2} \frac{4n\gamma^2\Gamma^2}{(n\gamma^2+\Gamma^2)^2} \frac{\left[\int_{-\infty}^t d\tau |E(\tau)|^2\right]^2}{2}, \tag{15}$$

where $\times$ represents element-by-element multiplication.

Repeating this procedure, we may find the population of the state corresponding to collection of all N photons at any time t. Accounting for the degeneracy of this state via a binomial coefficient, we obtain the following expression for the probability of complete absorption of all incoming photons:

$$P_N(N) = \binom{n}{N} \prod_{k=0}^{N-1} \frac{4n(k+1)}{(2n-k)^2}. \tag{16}$$

It is evident that this detector is only a perfect *single*-photon detector; multiple photons are detected increasingly inefficiently as the number of photons increases. This may be

understood by thinking about the absorption of photons as a successive process. Absorption of the first photon proceeds with maximal efficiency, filling state C. However, the occupation of this state blocks further collection by that state: If an element of the array absorbs a photon, it is no longer available to absorb further photons. This is seen in Eq. (15) in the appearance of $n-1$ instead of n multiplying γ^2. Thus, the detector interacting with successive photons is increasingly detuned from the optimal ratio, as the effective field coupling is reduced but Γ is not. This suggests that high-efficiency photon resolution requires an array of many more elements than potential photons detected. In Fig. 7, the number of detector elements required to detect all photons in a pulse with efficiencies of 80%, 95%, and 99% is plotted against
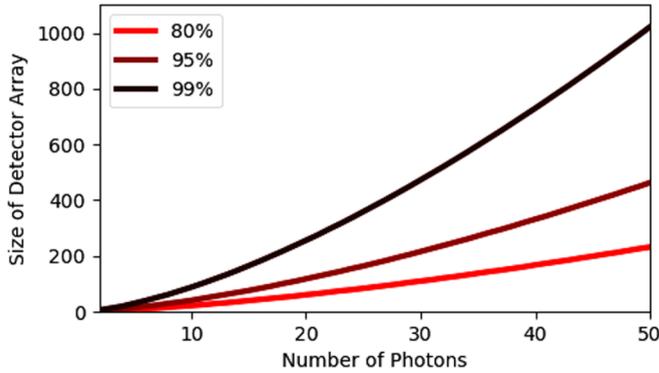
FIG. 7. Number of elements in a detector array required to detect a number of photons with a certain efficiency. The detector elements have the configuration in Fig. 5.

the number of photons in the pulse. As shown, the minimum number of detector elements increases dramatically as more stringent requirements for efficiency are imposed.

This reveals an important design consideration for high-performance number-resolving detectors: The effective parameters of the detector will be altered by interaction with the photon packet, and an ideal detector architecture is one that continues to be optimally tuned for interaction with the remainder of the wave packet. For example, to avoid the need for excessive numbers of elements, the array structure analyzed above must be reconfigured so that either the absorption properties are unchanged as the system absorbs photons or the decay process is also modified so that the effective decay and effective optical coupling remain matched following absorption.

VII. CONCLUSION

We developed a general framework, valid for any pointlike detector system under weak light coupling, for simulating, analyzing, and engineering one- and few-photon detection. Based on theoretical treatments of light-matter interactions and open quantum systems, we constructed a unified approach to modeling photodetection, deriving an expression that yields an output signal generated through the propagation of the incoming photon occupation by a Green's function representing the matter system. This expression provides a means of both explicitly calculating detection trajectories for determination of performance metrics as well as efficiently estimating these metrics from averaged solutions. It naturally furnishes a definition of ideal detection that places constraints on the system Green's function and ultimately the detector structure. Using this, we analytically analyze detector schemes and show that they conform to this definition with appropriate parameters. We also generalize the description of an ideal detector to systems with multiple excited levels, with and without energy dispersion, as well as multiple-photon resolution.

We emphasize that this framework describes detection abstractly and may encompass detectors utilizing very different physical systems, allowing comparative analysis of different detector architectures in a unified way. While we find, in principle, that ideal detector performance is possible, the structure of energy levels and couplings between them or the

physically allowed parameters may prohibit the realization of this limit. By representing a detector implementation within the framework, it may be evaluated against the ideal limit, clarifying the constraints and tradeoffs imposed by the system physics. This can guide engineering of the system architecture and parameters to improve the detector performance as well as facilitate the design of entirely novel detectors.

ACKNOWLEDGMENTS

APPENDIX A: LIGHT FIELD CONSTRUCTION

We use fields constructed using the quantum noise formalism where the field annihilation and creation operators $\hat{b}(t)$ and $\hat{b}^\dagger(t)$ satisfy the commutation relation $[\hat{b}^\dagger(t), \hat{b}(t')] = \delta(t - t')$. From this, one may define the objects

$$dB_t = \int_t^{t+dt} d\tau\, \hat{b}(\tau),$$

$$d\Lambda_t = \int_t^{t+dt} d\tau\, \hat{b}^\dagger(\tau)\hat{b}(\tau),$$

$$|N\rangle = \frac{1}{\sqrt{N!}}\left[\int d\tau E(\tau)\hat{b}(\tau)^\dagger\right]^N |0\rangle,$$

$$dB_t|N\rangle = dt\sqrt{N}E(t)|N-1\rangle, \tag{A1}$$

where N is the number of photons in the ith mode with temporal field profile $E(t)$ defined so that $\int_{-\infty}^\infty dt |E(t)|^2 = 1$. Then the light field is in general $\hat{\rho}_{\text{LIGHT}}(t) = \sum_{N,M} c_{N,M}(t)|N\rangle\langle M|$.

In this paper, we consider only a single mode; however, the consideration of multiple arbitrary modes is straightforward, if cumbersome. In this case, each mode will be associated with an index i denoting the mode temporal profile $E_i(t)$ and occupation N_i. The component field states are written as $|\mathbf{N}\rangle = \prod_i |N_i\rangle$, where $\mathbf{N}$ is a vector containing the N_i, so that the overall field density matrix is $\hat{\rho}_{\text{LIGHT}}(t) = \sum_{\mathbf{N},\mathbf{M}} c_{\mathbf{N},\mathbf{M}}(t)|\mathbf{N}\rangle\langle\mathbf{M}|$. As a result, the matter state is $\hat{\rho}_{\text{MATTER}}(t) = c_{\mathbf{N},\mathbf{M}}(t_0)\hat{\varrho}^{\mathbf{N},\mathbf{M}}(t)$ and the $\hat{\varrho}^{\mathbf{NM}}(t)$ are dependent on all $\hat{\varrho}^{\mathbf{N}-1_i,\mathbf{M}}(t)$, $\hat{\varrho}^{\mathbf{N},\mathbf{M}-1_i}(t)$, and $\hat{\varrho}^{\mathbf{N}-1_i,\mathbf{M}-1_i}(t)$ where $\mathbf{N}-1_i$ is the overall occupation $\mathbf{N}$ less one photon in the ith mode.

APPENDIX B: CONNECTION TO POVM

Any quantum measurement can be mathematically represented as a positive operator-valued measure (POVM) [28,29], which is a set of operators defined on the Hilbert space of the system being measured. Each of these operators is labeled by a measurement outcome, and in our context where one is

measuring a continuous current, these POVM elements are labeled by a continuous measurement record.

In this section, we show how the Green's functions we have defined are related to a POVM on the field degrees of freedom. We consider a single mode of the field for simplicity, but the results can be generalized to more complex fields.

Using Eq. (A1), we obtain a recursive operator

$$\mathcal{P}(t) = 1 + \int_{-\infty}^{t} \bar{\mathcal{G}}(t-\tau)(\bar{\mathcal{S}} d\Lambda_\tau + \bar{\mathcal{L}}^+ dB_\tau + \bar{\mathcal{L}}^- dB_\tau^\dagger)\mathcal{P}(\tau)$$

such that our final system state $\bar{\rho}(t)$ can be expressed as

$$\bar{\rho}_{\text{MATTER}}(t) = \text{Tr}_{\text{LIGHT}}[\mathcal{P}(t)\bar{\rho}_{\text{MATTER}}(t_0) \otimes \hat{\rho}_{\text{LIGHT}}(t_0)],$$

where $\hat{\rho}_{\text{LIGHT}}(t_0)$ is our initial field state constructed as indicated above. We can now write operator expectation values as a function of the input state

$$\langle \hat{X}_i \rangle(\hat{\rho}_{\text{LIGHT}}(t_0), t)$$
$$= \text{Tr}_{\text{LIGHT}}[\bar{X}_i \cdot \mathcal{P}(t)\bar{\rho}_{\text{MATTER}}(t_0) \otimes \hat{\rho}_{\text{LIGHT}}(t_0)].$$

For a complete set of appropriately normalized amplification operators $\bar{X}_i$ obtained from vectorization of POVM elements $\hat{X}_i$ on the system, when an initial state is assumed, $\bar{X}_i \cdot \mathcal{P}(t)\bar{\rho}_{\text{MATTER}}(t_0)$ forms a coarse-grained POVM on the field.

APPENDIX C: ESTIMATION OF EFFICIENCY FROM AVERAGE DYNAMICS

The stochastic term given by Eq. (5) unfortunately contributes both explicit time dependence and nonlinearity due to the normalizing third term of the superoperator $\mathcal{H}$, meaning that it cannot be described efficiently using the machinery presented in the main text. Furthermore, since detection events are defined as the output exceeding some threshold, which is a nonlinear condition on the current, one usually has to resort to stochastic trajectory simulations in order to estimate efficiencies [23].

However, as mentioned in the main text, under commonly prevailing conditions, measurement metrics can be estimated from the averaged dynamics. Under sufficiently strong amplification, the trajectories become jumplike, with the system being in either a monitored or unmonitored state. We may thus write the probability of the system being in the monitored subsystem at time t as

$$P_{\text{MON}}(t) = P_{\text{MON}|\text{MON}}(t, t')P_{\text{MON}}(t')$$
$$+ P_{\text{MON}|\text{UNMON}}(t, t')P_{\text{UNMON}}(t'), \quad \text{(C1)}$$

where $P_{x|y}(t, t')$ is the probability of the system being in subsystem x at t given that it was in subsystem y at t'. Here, the subsystems are labeled MON and UNMON, which correspond to the monitored and unmonitored states. Recall that $\hat{x}$ projects onto the subspace of states being monitored through $\hat{X}$. Therefore, $P_{\text{MON}}(t) = \text{Tr}[\hat{x}\hat{\rho}(t)] = \bar{x}^\dagger\bar{\rho}(t)$, and hence rewriting the above gives

$$P_{\text{MON}}(t) = P_{\text{MON}|\text{MON}}(t, t')\bar{x}^\dagger\bar{\rho}(t')$$
$$+ P_{\text{MON}|\text{UNMON}}(t, t')(\mathbf{I} - \bar{x})^\dagger\bar{\rho}(t'). \quad \text{(C2)}$$

Now, we formulate a probability quantity that will serve as an approximation of detection efficiency. Consider a single

detection channel for simplicity. Our proxy for efficiency is based on the time-dependent probability that a detection event is registered in the channel of interest, which we denote by $\Pi(t)$. This is the probability of registering a detection event in the interval $(t_0, t]$ given a photon in the field at time t_0. We write this explicitly as

$$\Pi(t) = \int_{t_0+t_{\text{MIN}}}^{t} dt' s_{\text{MON}|\text{UNMON}}(t' - t_{\text{MIN}})$$
$$\times P_{\text{MON}|\text{MON}}(t', t' - t_{\text{MIN}}), \quad \text{(C3)}$$

where $s_{\text{MON}|\text{UNMON}}(t)$ is the probability of transitioning from the unmonitored subspace to the monitored subspace at time t. The integrand in this expression represents the joint probability of population transferring from the unmonitored subspace (which includes the field degrees of freedom) to the monitored subspace at time t' and that this population stays in the monitored subspace for time t_{MIN}, which we take as the minimum time that the monitored subspace has to be populated before a detection is registered. (A more experimentally motivated condition for registering a detection is that the measured current exceed some threshold value I_{HIT}; however, since it is difficult to develop an expression based on this condition, we proceed with the temporal condition. The approximate relationship between the two is that $t_{\text{MIN}} \approx I_{\text{HIT}}t_m/\chi$) where t_m is chosen to yield a certain dark count rate.

In order to proceed, we require an expression for $s_{\text{MON}|\text{UNMON}}(t)$. To obtain this, we return to Eq. (C1) and consider the incremental quantity

$$\Delta P_{\text{MON}}(t) \equiv P_{\text{MON}}(t) - P_{\text{MON}}(t - \Delta t)$$
$$= [P_{\text{MON}|\text{MON}}(t, t - \Delta t) - 1]P_{\text{MON}}(t - \Delta t)$$
$$+ P_{\text{MON}|\text{UNMON}}(t, t - \Delta t)P_{\text{UNMON}}(t - \Delta t).$$

Now, the first quantity in this expression corresponds to the rate at which population is staying in the monitored subspace while the second quantity corresponds to the rate at which population is entering the monitored subspace from the unmonitored subspace at time t. This is precisely the quantity $s_{\text{MON}|\text{UNMON}}(t)\Delta t$, and given this expression for it, we can compute it explicitly in terms of the density matrix and dynamics:

$$s_{\text{MON}|\text{UNMON}}(t)$$
$$= \frac{d}{dt}(P_{\text{MON}|\text{UNMON}}(t, t')P_{\text{UNMON}}(t - dt))|_{t'=t}$$
$$= \frac{d}{dt}(P_{\text{MON}|\text{UNMON}}(t, t')(\mathbf{I} - \bar{x})^\dagger\bar{\rho}(t'))|_{t'=t}$$
$$= \frac{d}{dt}(\bar{x}^\dagger\bar{\rho}(t) - P_{\text{MON}|\text{MON}}(t, t')\bar{x}^\dagger\bar{\rho}(t'))|_{t'=t}$$
$$= \bar{x}^\dagger\left(\frac{d\bar{\rho}(t)}{dt} - \frac{dP_{\text{MON}|\text{MON}}(t, t')}{dt}|_{t'=t}\bar{\rho}(t)\right),$$

where in the third line we have used the identity in Eq. (C2). $P_{\text{MON}|\text{MON}}(t, t')$ is typically a function of the internal dynamics and can then be readily determined from the system $\bar{\mathcal{G}}$. When $\bar{A}$ is time independent, it can further be written as a function of $\tau = t - t'$ as

$$P_{\text{MON}|\text{MON}}(t, t') = \bar{x}^\dagger\bar{\mathcal{G}}(\tau)\bar{x}.$$

Hence,

$$\frac{dP_{\text{MON|MON}}(t,t')}{dt}\Big|_{t=t'} = \bar{x}^\dagger \frac{d\bar{\mathcal{G}}(t)}{dt}\Big|_{t=0}\bar{x} = \bar{x}^\dagger \bar{\mathcal{A}}\bar{x}.$$

Finally, using this expression for $s_{\text{MON|UNMON}}(t)$ in Eq. (C3) results in

$$\begin{aligned}
\Pi(t) &= \int_{t_0+t_{\text{MIN}}}^{t} dt' s_{\text{MON|UNMON}}(t'-t_{\text{MIN}}) \\
&\quad \times P_{\text{MON|MON}}(t',t'-t_{\text{MIN}}) \\
&= \int_{t_0}^{t-t_{\text{MIN}}} d\tau s_{\text{MON|UNMON}}(\tau) \\
&\quad \times P_{\text{MON|MON}}(\tau+t_{\text{MIN}},\tau) \\
&= \int_{t_0}^{t-t_{\text{MIN}}} d\tau P_{\text{MON|MON}}(t_{\text{MIN}}) \\
&\quad \times \bar{x}^\dagger \left[\frac{d\bar{\rho}(\tau)}{d\tau} - \bar{x}^\dagger \bar{\mathcal{A}}\bar{x}\bar{\rho}(\tau) \right] \\
&= \bar{x}^\dagger \bar{\mathcal{G}}(t_{\text{MIN}})\bar{x}\bigg[\bar{x}^\dagger \bar{\rho}(t-t_{\text{MIN}}) \\
&\quad - (\bar{x}^\dagger \bar{\mathcal{A}}\bar{x})\bar{x}^\dagger \int_{t_0}^{t-t_{\text{MON}}} d\tau \bar{\rho}(\tau) \bigg],
\end{aligned}$$

where we have assumed that $\bar{x}^\dagger \bar{\rho}(t_0) = 0$.

If the monitored subsystem is stable and population loss from it can be neglected, $P_{\text{MON|MON}}(t,t') = 1$, and therefore $\bar{x}^\dagger \bar{\mathcal{G}}\bar{x} = 1$ and $\bar{x}^\dagger A\bar{x} = 0$. In this case, the above reduces to

$$\Pi(t) = \bar{x}^\dagger \bar{\rho}(t-t_{\text{MIN}}). \tag{C4}$$

These expressions are the probability that a detection event is registered in the measurement channel. In general, there may be multiple measurement channels, each associated with an operator $\hat{X}_i$, projection subspace $\hat{x}_i$, and $\Pi_i(t)$. The probability that a given number of photons N are detected will depend in some fashion on the combined behavior of all the channels, so that P_N is some function p_N of all Π_i.

As an example, we refer to the system discussed in Sec. VI. Since each element has an associated measurement channel, we must consider $\Pi_i, i \in 1..n$. Since the absorption of a photon corresponds to one measurement channel registering a hit, we take the registering of a hit on N channels to correspond to detection of an N photon wave packet. Thus, for a single photon, $p_1(\Pi_i) = \sum_i \Pi_i$, and for, e.g., three photons, $p_3(\Pi_i) = \sum_{i>j>k} \Pi_i|\Pi_j|\Pi_k$, where $\Pi_i|\Pi_j$ is the probability of a detection event on the ith channel given one on the jth channel. Since this system satisfies the conditions of Eq. (C4), this can be expressed as

$$\Pi_i|\Pi_j|\Pi_k... = (\bar{x}_i \times \bar{x}_j \times \bar{x}_k...)^\dagger \bar{\rho}(t-t_{\text{MIN}}),$$

where $\times$ represents an element-wise product.

APPENDIX D: DEGENERATE MULTIELEMENT

For a system of n elements, we can represent the total system density matrix using the notation

$$\hat{\rho}_{n:m_il_im_jl_jm_kl_k...} = \hat{\rho}^i_{m_il_i} \otimes \hat{\rho}^j_{m_jl_j} \otimes \hat{\rho}^k_{m_kl_k} \otimes \cdots,$$

where m_i and l_i are one of 0_i, 1_i, C_i, corresponding to the states of subsystem i. For example, $\hat{\rho}_{3:0_11_12_02_03_13}$ represents a system with three subsystems and coherences between the 0 and 1 states in subsystems 2 and 3. Since the subsystems are all degenerate, for brevity only the subsystems not in the ground state will be noted in the subscript and we use, for example, $\hat{\rho}_{n:1001}$ to represent all of the $n(n-1)/2$ configurations with this description. The operators on the vectorized density matrix are then

$$\bar{\rho} = \begin{bmatrix} \hat{\rho}_{00} \\ \hat{\rho}_{n:11} \\ \hat{\rho}_{n:1001} \\ \hat{\rho}_{n:CC} \\ \vdots \\ \hat{\rho}_{n:10} \\ \hat{\rho}_{n:01} \\ \vdots \end{bmatrix}$$

so that

$$
\bar{\mathcal{A}} = \begin{bmatrix}
0 & n\gamma^2 & 0 & n(n-1)\gamma^2 & & 0 & 0 & \\
0 & -(\gamma^2+\Gamma^2) & 0 & -(n-1)\gamma^2 & & 0 & 0 & \\
0 & \Gamma^2 & 0 & 0 & \cdots & 0 & 0 & \cdots \\
0 & -\gamma^2 & 0 & -(n-1)\gamma^2 & & 0 & 0 & \\
& \vdots & & & \ddots & & & \\
0 & 0 & 0 & 0 & & i\omega - \frac{n\gamma^2+\Gamma^2}{2} & 0 & \\
0 & 0 & 0 & 0 & & 0 & -i\omega - \frac{n\gamma^2+\Gamma^2}{2} & \\
& \vdots & & & & & & \ddots
\end{bmatrix},
$$

$$
\bar{\mathcal{L}}^+ = \begin{bmatrix}
0 & 0 & 0 & 0 & & -n\gamma & 0 & \\
0 & 0 & 0 & 0 & & \gamma & 0 & \\
0 & 0 & 0 & 0 & \cdots & (n-1)\gamma & 0 & \cdots \\
0 & 0 & 0 & 0 & & 0 & 0 & \\
& \vdots & & & \ddots & & & \\
0 & 0 & 0 & 0 & & 0 & 0 & \\
\gamma & -\gamma & -(n-1)\gamma & 0 & & 0 & 0 & \\
& \vdots & & & & & & \ddots
\end{bmatrix}
$$

APPENDIX E: DISPERSIVE BAND

Much of the machinery developed in other contexts for manipulating Green's functions may be brought to bear as well. In some cases, it is convenient to separate $\bar{\mathcal{A}}$ into an $\bar{\mathcal{A}}^0$ and $\bar{\mathcal{A}}^1$, where the former can be easily exponentiated (e.g., it is diagonal) so that

$$
\bar{\mathcal{G}}(t) = e^{\bar{\mathcal{A}}t} = e^{(\bar{\mathcal{A}}^0 t + \bar{\mathcal{A}}^1 t)} = \lim_{dt\to 0}[e^{\bar{\mathcal{A}}^0 dt} + \bar{\mathcal{A}}^1 dt]^{t/dt}
$$

$$
= \hat{g}(t) + \int_0^t d\tau\, \hat{g}(t-\tau)\bar{\mathcal{A}}^1 \hat{g}(\tau) + \int_{-\infty}^t d\tau\, \hat{g}(t-\tau)\bar{\mathcal{A}}^1 \int_0^\tau d\tau'\, \hat{g}(\tau-\tau')\bar{\mathcal{A}}^1 \hat{g}(\tau') + \cdots
$$

$$
= \hat{g}(t) + \int_0^t d\tau\, \hat{g}(t-\tau)\bar{\mathcal{A}}^1 \bar{\mathcal{G}}(\tau). \tag{E1}
$$

An example of this is provided by a system comprising a ground state labeled 0 and a dispersive band of n excited states labeled with the indices $1..n$. All excitations are characterized by the same light-matter coupling γ. For clarity, we represent elements of $\bar{\rho}$ with the notation $\bar{\rho}_{(ij)} = \hat{\rho}_{ij}$ and the element of a matrix $\bar{\mathcal{O}}$ coupling density matrix elements $\hat{\rho}_{ij}$ and $\hat{\rho}_{rs}$ with the notation $\bar{\mathcal{O}}_{(ij)(rs)}$.

Then $\bar{\mathcal{A}}^0$ and $\bar{\mathcal{A}}^1$ are, giving only nonzero elements,

$$
\bar{\mathcal{A}}^0_{(ij)(ij)} = i(\omega_i - \omega_j), \qquad\qquad i,j \in 0\ldots n,
$$

$$
\bar{\mathcal{A}}^1_{(00)(ij)} = \gamma^2,
$$

$$
\bar{\mathcal{A}}^1_{(0i)(0j)} = \bar{\mathcal{A}}^1_{(i0)(j0)} = -\gamma^2/2, \qquad\qquad i,j,r,s \in 1\ldots n,
$$

$$
\bar{\mathcal{A}}^1_{(ij)(rs)} = -\frac{\gamma^2}{2}(\delta_{ir} + \delta_{js})
$$

The total population of the band is given by

$$
u(t) = \sum_i \bar{\varrho}^{11}_{(ii)}(t) = \sum_{ijrs} 2\int_0^t d\tau\, \gamma \bar{\mathcal{G}}_{(ii)(rj)}(t-\tau)E(\tau)\int^\tau d\tau'\, \gamma \bar{\mathcal{G}}_{(r0)(s0)}(\tau-\tau')E^*(\tau').
$$

Expanding $\sum_s \bar{\mathcal{G}}_{(r0)(s0)}$ according to Eq. (E1) gives

$$\sum_s \bar{\mathcal{G}}_{(r0)(s0)}(t) = e^{-i\omega_r t} - \sum_r \frac{\gamma^2}{2} \int_0^t d\tau e^{-i\omega_r(t-\tau)} e^{-i\omega_s \tau} + \sum_{qs} \frac{\gamma^4}{4} \int_0^t d\tau e^{-i\omega_r(t-\tau)} \times \int_0^\tau d\tau' e^{-i\omega_q(\tau-\tau')} e^{-i\omega_s \tau'} + \cdots.$$

If we take the continuum limit $n \to \infty$ and assume a Lorentzian distribution of states $D(\omega) = \frac{n}{\pi} \frac{\zeta^2/2}{(\zeta^2/2)^2 + (\omega - \omega_E)^2}$ around a center frequency ω_E, then $\sum_r e^{i\omega_r t} = \int \sigma(\omega_r) d\omega_r e^{i\omega_r t} = n e^{-(\zeta^2/2)t}$ and the series can be evaluated analytically as

$$\sum_s \bar{\mathcal{G}}_{(r0)(s0)}(t) = e^{-i\omega_r t} - \sum_s \frac{\gamma^2}{2} \int_0^t d\tau e^{-i\omega_r(t-\tau)} n e^{-(\zeta^2/2)\tau} + \sum_{qs} \frac{\gamma^4}{4} \int_0^t d\tau e^{-i\omega_r(t-\tau)} \int_0^\tau d\tau' n e^{-(\zeta^2/2)\tau'} + \cdots$$

$$= e^{-i\omega_r t} - \frac{n\gamma^2}{2} \int_0^t d\tau e^{-i\omega_r(t-\tau)} \times e^{-\zeta^2/2\tau} \left(1 - \frac{n\gamma^2}{2} + \frac{n^2\gamma^4}{4} \cdots\right) = e^{-i\omega_r t} - \frac{n\gamma^2}{2} \int_0^t d\tau e^{-i\omega_r(t-\tau)} e^{-\frac{\zeta^2 + n\gamma^2}{2}\tau}.$$

Following a similar procedure for $\sum_r \bar{\mathcal{G}}_{(pq)(r0)}(t)$, we can ultimately write

$$u(t) = \frac{2n\gamma^2}{n\gamma^2 + \zeta^2} \int^t d\tau (n\gamma^2 e^{-(n\gamma^2 + \zeta^2)(t-\tau)} + \zeta^2) E(\tau) \int^\tau d\tau' e^{-\frac{n\gamma^2 + \zeta^2}{2}(\tau - \tau')} E^*(\tau').$$

[1] J. C. Bienfang, A. J. Gross, A. Mink, B. J. Hershman, A. Nakassis, X. Tang, R. Lu, D. H. Su, C. W. Clark, C. J. Williams, E. W. Hagley, and J. Wen, Opt. Express **12**, 2011 (2004).

[2] M. E. Woodson, M. Ren, S. J. Maddox, Y. Chen, S. R. Bank, and J. C. Campbell, Appl. Phys. Lett. **108**, 081102 (2016).

[3] W. H. P. Pernice, C. Schuck, O. Minaeva, M. Li, G. N. Goltsman, A. V. Sergienko, and H. X. Tang, Nat. Commun. **3**, 1325 (2012).

[4] F. Marsili, F. Bellei, F. Najafi, A. E. Dane, E. A. Dauler, R. J. Molnar, and K. K. Berggren, Nano. Lett. **12**, 4799 (2012).

[5] F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, Nat. Photon. **7**, 210 (2013).

[6] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, Rev. Sci. Instrum. **82**, 071101 (2011).

[7] R. H. Hadfield, Nat. Photon. **3**, 696 (2009).

[8] K. N. Cassemiro, K. Laiho, and C. Silberhorn, New J. Phys. **12**, 113052 (2010).

[9] F. Piacentini, A. Avella, E. Rebufello, R. Lussana, F. Villa, A. Tosi, M. Gramegna, G. Brida, E. Cohen, L. Vaidman, I. P. Degiovanni, and M. Genovese, Nat. Photon. **13**, 1191 (2017).

[10] J. Řeháček, Z. Hradil, O. Haderka, J. Peřina, and M. Hamar, Phys. Rev. A **67**, 061801 (2003).

[11] D. Achilles, C. Silberhorn, C. Śliwa, K. Banaszek, and I. A. Walmsley, Opt. Lett. **28**, 2387 (2003).

[12] J. Sperling, T. J. Bartley, G. Donati, M. Barbieri, X.-M. Jin, A. Datta, W. Vogel, and I. A. Walmsley, Phys. Rev. Lett. **117**, 083601 (2016).

[13] R. J. Glauber, Phys. Rev. **130**, 2529 (1963).

[14] L. Mandel, E. C. G. Sudarshan, and E. Wolf, Proc. Phys. Soc. **84**, 435 (1964).

[15] P. L. Kelley and W. H. Kleiner, Phys. Rev. **136**, A316 (1964).

[16] M. O. Scully and W. E. Lamb, Phys. Rev. **179**, 368 (1969).

[17] M. D. Srinivas and E. B. Davies, Opt. Acta **28**, 981 (1981).

[18] M. Ueda, N. Imoto, and T. Ogawa, Phys. Rev. A **41**, 3891 (1990).

[19] P. D. Drummond, Phys. Rev. A **35**, 4253 (1987).

[20] M. Fleischhauer, J. Phys. A **31**, 453 (1998).

[21] J. Sperling, W. Vogel, and G. S. Agarwal, Phys. Rev. A **85**, 023820 (2012).

[22] S. J. van Enk, J. Phys. Commun. **1**, 045001 (2017).

[23] S. M. Young, M. Sarovar, and F. Léonard, Phys. Rev. A **97**, 033836 (2018).

[24] B. Q. Baragiola, R. L. Cook, A. M. Branczyk, and J. Combes, Phys. Rev. A **86**, 013811 (2012).

[25] K. Jacobs and D. A. Steck, Contemp. Phys. **47**, 279 (2006).

[26] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. A **69**, 062320 (2004).

[27] M. Gross and S. Haroche, Phys. Rep. **93**, 301 (1982).

[28] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, Lecture Notes in Physics Vol. 190, edited by A. Böhm, J. D. Dollard, and W. H. Wootters (Springer, Berlin, 1983).

[29] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Springer, Berlin, 2010).