HETERODYNE DETECTION OF RANDOM GAUSSIAN SIGNALS IN THE OPTICAL AND INFRARED: OPTIMIZATION OF PULSE DURATION

Marek ELBAUM ‡

Riverside Research Institute, New York, 10023, USA

and

Malvin Carl TEICH ‡

Columbia Radiation Laboratory, Department of Electrical Engineering and Computer Science, New York, 10027, USA

Received 21 July 1978

The performance of a heterodyne system in estimating the mean intensity of a gaussian random signal depends on the mean number of photocarriers released by the signal radiation in its coherence volume (degeneracy parameter). In a pulsed radar, this parameter can be manipulated by varying the pulse duration while keeping the mean number of signal photocarriers constant. Furthermore, in a number of situations of practical interest, an optimal pulse duration exists, corresponding to a degeneracy parameter of unity. Heterodyne and direct detection are compared for this case, and direct detection is found to be superior in the strong-signal limit.

An important application of heterodyning in the infrared [1–3] and visible [4] is the detection and estimation of a weak radiation signal embedded in a strong background. In this paper we consider the estimation problem when the signal and background fields are both zero-mean, stationary, complex spatio-temporal gaussian stochastic processes.

For stochastic signal and noise fields, a suitable performance measure for such a system is the ratio of the estimator of the mean signal intensity to the standard deviation of the observable (which, in our case, is the total current at the output of the system). This quantity, which we denote $\mathrm{SNR_H}$, has been used as a figure of merit for both heterodyne [5] and direct [6] detection in optical spectroscopy. It has also been applied to laser radar systems [7], and indeed, a similar quantity has been proposed for describing the performance of homodyne systems [8]. The traditional signal-to-noise ratio (denoted SNR), on the other hand, is defined as the ratio of the average ac signal power to the average ac noise power measured in the absence of signal [9]. Though widely used in laser radar and in optical communications, this quantity does not adequately account for random signal fluctuations and it must therefore be used with care [8].

An unbiased estimator of the mean signal intensity can be obtained from the current $i(t)$ at the output of the heterodyne detection system shown in fig. 1. The quantity $i(t)$ is generated by passing the photocurrent $i_{\mathrm{L}}(t)$ successively through a narrow bandpass IF filter with transfer function $H_{\mathrm{IF}}(\omega)$, a square-law device, and a low-pass filter (integrator) with transfer function $H_{\mathrm{T}}(\omega)$. The dimensionless figure of merit, $\mathrm{SNR_H}$, is defined as

$$\mathrm{SNR_H} = \frac{\langle i(t)\rangle - \langle i(t)_0\rangle}{[\mathrm{Var}\{i(t)\}]^{1/2}}, \qquad (1)$$

where $\langle i(t)\rangle$ and $\langle i(t)_0\rangle$ represent average values of the output current in the presence and in the absence of signal, respectively. The angular brackets in eq. (1) represent ensemble averaging over all spatial and tempo-

‡ Work supported by DARPA under Contract No. DAAK40-76-C-0500.

‡ Work supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract No. DAAG29-77-C-0019

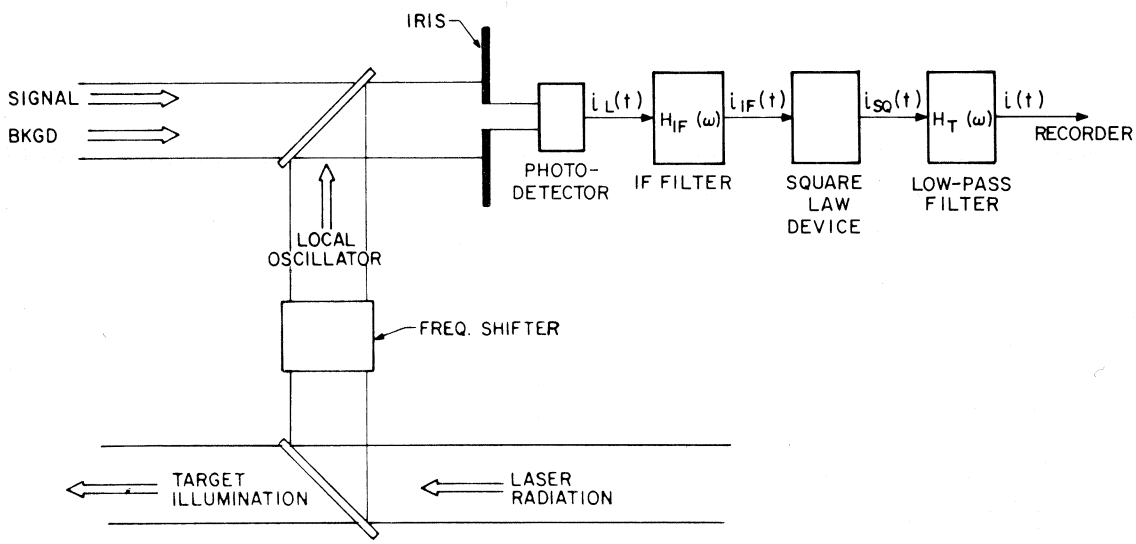

Fig. 1. Block diagram of a laser radar version of the heterodyne system.

ral realizations of the signal and background fields, and Var {·} is read "the variance of {·}". The traditional SNR, on the other hand, is defined as

$$\mathrm{SNR} = \frac{i(t) - i(t)_0}{i(t)_0} = \frac{\eta P_S}{h\nu \Delta_{IF}}. \tag{2}$$

Here η is the detector quantum efficiency, P_S is the detected signal radiation power, h is Planck's constant, ν is the frequency of the radiation, and Δ_{IF} is the bandwidth of the IF filter. In obtaining eq. (2) the usual assumption has been made that $i(t)_0$ is dominated by local-oscillator generated shot noise. SNR and ANR_H are related when the signal is deterministic. In general, $i(t)$ fluctuates randomly in response to the fluctuations in the signal power P_S, so that SNR is a stochastic quantity.

By directing our attention to the quantity SNR_H for a pulsed signal version of the laser radar system shown in fig. 1, we find that there is an optimal integration time for which SNR_H is maximized, provided that the average received signal energy per pulse (or mean number of signal counts per pulse N_S) is fixed and that the integration time is chosen to be approximately equal to the pulse duration T. In this case T is shown to be optimum if N_S is equal to the number of degrees of freedom in the signal radiation, or equivalently, if the degeneracy parameter of the signal radiation is unity. For the optimal pulse duration, SNR_H turns out to be proportional to $N_S^{1/2}$. Finally, the relative merits of heterodyne and direct detection are weighed by comparing SNR_H with its analog for direct detection.

A useful and intuitively appealing formula for SNR_H should relate SNR_H directly to the signal intensity and coherence function, and to the following system parameters: the bandwidths of the IF and integration filters, Δ_{IF} and Δ_T, and the detector aperture A. The key steps in the derivation of such a formula are straightforward: the photocurrent $i_L(t)$ and its power spectrum $P_L(\omega)$ are computed from Mandel's semiclassical photodetection theory [10,11]. The mean and variance of the observed current $i(t)$ are then related to the mean and variance of $i_L(t)$ by means of a systems analysis such as that performed by Cummins and Swinney [5].

The total instantaneous semiclassical field $V(\boldsymbol{x}, t)$ incident on the detector at the point $\boldsymbol{x}$ is a sum of linearly polarized, statistically independent signal, background, and local oscillator (LO) fields, $V_1(\boldsymbol{x}, t)$, $V_2(\boldsymbol{x}, t)$, and $V_{LO}(t)$, respectively, i.e.,

$$V(\boldsymbol{x}, t) = V_1(\boldsymbol{x}, t) + V_2(\boldsymbol{x}, t) + V_{LO}(t). \tag{3}$$

$V_{LO}(t)$ is a deterministic plane wave field. The signal and the background fields are assumed to satisfy the conditions for cross-spectral purity [11]. It follows therefore that the second-order spatio-temporal coherence functions for these fields, γ_1 and γ_2, can be factored as follows:

$$\gamma_k(\boldsymbol{x}_1-\boldsymbol{x}_2,\tau) = \langle V_k(\boldsymbol{x}_1,t)\, V_k^*(\boldsymbol{x}_2,t+\tau)\rangle/\langle I_k\rangle \quad (4)$$

$$= \gamma_k(\boldsymbol{x}_1-\boldsymbol{x}_2,0)\, g_k(\tau), \quad (4a)$$

where $k = 1, 2$ [see eq. (3)]. The first and second factors in eq. (4a) represent the spatial and temporal coherence functions, respectively. The quantity $I_k = V_k(\boldsymbol{x},t)\, V_k^*(\boldsymbol{x},t)$ represents the intensity of V_k, and $\langle I_k\rangle = \langle I_k(\boldsymbol{x}_1,t)\rangle = \langle I_k(\boldsymbol{x}_2,t+\tau)\rangle$. The corresponding coherence areas A_k, and degeneracy parameters δ_k are given respectively by

$$A_k = \int_{-\infty}^{\infty} \gamma_k(\boldsymbol{\rho},0)\,\mathrm{d}^2\boldsymbol{\rho}, \quad \boldsymbol{\rho} = \boldsymbol{x}_1-\boldsymbol{x}_2, \quad (5)$$

$$T_k = \int_{-\infty}^{\infty} g_k(\tau)\,\mathrm{d}\tau, \qquad \delta_k = \eta\langle I_k\rangle T_k A_k. \quad (6, 7)$$

The coherence time and the degeneracy parameter of the signal radiation are assumed to be larger by several orders of magnitude than the corresponding quantities for the background radiation ($T_1 \gg T_2$, $\delta_1 \gg \delta_2$).

With the usual condition of a strong coherent local oscillator ($I_{\mathrm{LO}} \gg \langle I_1\rangle, \langle I_2\rangle$) the power spectrum $P_{\mathrm{L}}(\omega)$ of the photocurrent $i_{\mathrm{L}}(t)$ is the Fourier transform of the photocurrent correlation function, $R_{\mathrm{L}}(\tau) = \langle i_{\mathrm{L}}(t)\, i_{\mathrm{L}}(t+\tau)\rangle$. In calculating $R_{\mathrm{L}}(\tau)$, the appropriate joint probability function is obtained from Mandel's photodetection theory [10], in which the photocurrent is modeled as a series of discrete randomly localized, infinitely narrow pulses represented by Dirac delta functions. It follows that [5,10,11]

$$P_{\mathrm{L}}(\omega) = \int_{-\infty}^{\infty} R_{\mathrm{L}}(\tau)\exp(\mathrm{j}\omega\tau)\,\mathrm{d}\tau$$

$$\approx 2ei_{\mathrm{LO}} + 2\pi i_{\mathrm{LO}}^2\delta'(\omega)$$

$$+ 2i_{\mathrm{LO}}\langle i_{\mathrm{S}}\rangle QP_1(\omega - |\omega_{\mathrm{S}}-\omega_{\mathrm{LO}}|), \quad (8)$$

where $i_{\mathrm{LO}} = e\sigma I_{\mathrm{LO}}A$ and $\langle i_{\mathrm{S}}\rangle = e\sigma\langle I_1\rangle A$ are the mean currents generated by the local oscillator and by the signal beams, respectively, e is the electronic charge, and $\sigma = \eta/h\nu$ is a suitably normalized quantum efficiency. The quantity Q depends on both the detector aperture A and the signal field coherence area A_1, and is given by

$$Q \approx \begin{cases} A_1/A & \text{for } A_1 \ll A, \\ 1 & \text{for } A_1 \gg A, \end{cases} \quad (9)$$

in the two limiting cases indicated. The function $P_1(\omega - |\omega_{\mathrm{S}}-\omega_{\mathrm{LO}}|)$ is the normalized spectrum of the signal-LO interference term, ω_{S} is the center frequency of the signal spectrum, and ω_{LO} is the frequency of LO field. Though $P_{\mathrm{L}}(\omega)$ may be assumed to be symmetric about zero frequency, eq. (8) represents it in terms of positive frequencies only ($\int_0^\infty \delta'(\omega)\,\mathrm{d}\omega = 1$).

The three terms on the right-hand-side of eq. (8) represent, respectively, white shot noise arising from the LO, dc, and the fluctuating (excess noise) component of the photocurrent. The signal and background radiation contributions to the first two terms are neglected in accordance with the assumption that the strong local oscillator at the detector swamps these other contributions. The background-LO interference term does not appear in eq. (8) since the spectrum of the background is generally much broader than that of the signal. Thus it is easily discriminated against by passing $i_{\mathrm{L}}(t)$ through a narrowband IF filter $H_{\mathrm{IF}}(\omega)$ centered at $|\omega_{\mathrm{S}}-\omega_{\mathrm{LO}}|$ with bandwidth Δ_{IF}. The IF filter also rejects the dc photocurrent and reduces the LO shot noise relative to the ac component.

The current at the output of the IF filter, $i_{\mathrm{IF}}(t)$, is a narrowband, zero-mean stationary gaussian process with a variance given by

$$\mathrm{Var}\{i_{\mathrm{IF}}(t)\} = \int_0^\infty P_{\mathrm{L}}(\omega)|H_{\mathrm{IF}}(\omega)|^2\,\mathrm{d}\omega/2\pi. \quad (10)$$

For simplicity, we assume that $|H_{\mathrm{IF}}(\omega)|^2$ and $P_1(\omega - |\omega_{\mathrm{S}}-\omega_{\mathrm{LO}}|)$ are both of rectangular shape and are centered at the difference frequency $|\omega_{\mathrm{S}}-\omega_{\mathrm{LO}}|$; the former has unity height over the bandwidth $2\pi\Delta_{\mathrm{IF}}$, whereas the latter has height $1/\Delta_{\mathrm{S}}$ over its bandwidth $2\pi\Delta_{\mathrm{S}}$ (in accordance with the imposed normalization condition). The value of Δ_{IF} should not be larger than Δ_{S}. Under these conditions

$$\mathrm{Var}\{i_{\mathrm{IF}}(t)\} = 2ei_{\mathrm{LO}}\Delta_{\mathrm{IF}} + 2i_{\mathrm{LO}}\langle i_{\mathrm{S}}\rangle Q\Delta_{\mathrm{IF}}/\Delta_{\mathrm{S}}. \quad (10a)$$

For an integration time T larger than $1/\Delta_{\mathrm{IF}}$, it can be readily shown that the mean and variance of the observed current $i(t)$ at the output of the low-pass (integration) filter are [5,12]

$$\langle i(t)\rangle \approx \mathrm{Var}\{i_{\mathrm{IF}}(t)\}; \qquad \langle i(t)\rangle \approx 2ei_{\mathrm{LO}}\Delta_{\mathrm{IF}}, \quad (11)$$

and

$$\mathrm{Var}\,\{i(t)\} = (\mathrm{Var}\,\{i_{\mathrm{IF}}(t)\})^2/\Delta_{\mathrm{IF}}T, \qquad (12)$$

respectively. Hence, using eq. (1) for $\mathrm{SNR_H}$, which is the quantity of interest, we obtain

$$\mathrm{SNR_H} = \langle i_\mathrm{S}\rangle(\Delta_{\mathrm{IF}}/\Delta_\mathrm{S})Q \times \{[e\Delta_{\mathrm{IF}} + \langle i_\mathrm{S}\rangle(\Delta_{\mathrm{IF}}/\Delta_\mathrm{S})Q](\Delta_{\mathrm{IF}}T)^{-1/2}\}^{-1} = \langle\mathrm{SNR}\rangle/[\Delta_\mathrm{S}/\Delta_{\mathrm{IF}}Q + \langle\mathrm{SNR}\rangle](\Delta_{\mathrm{IF}}T)^{-1/2}. \qquad (13)$$

The quantity $\langle\mathrm{SNR}\rangle = \langle i_\mathrm{S}\rangle/e\Delta_{\mathrm{IF}}$ is the signal-to-shot-noise ratio and is equivalent to eq. (2) for random signal fields. In the limit of large $\langle\mathrm{SNR}\rangle$, $\mathrm{SNR_H}$ depends only on the parameters of the system and is given by $(\Delta_{\mathrm{IF}}T)^{1/2}$. The integration time T is usually chosen to be much greater than $1/\Delta_\mathrm{S}$. This corresponds to increasing $\mathrm{SNR_H}$ by noncoherent averaging.

We now apply these results to a pulsed laser radar system, addressing the practical situation in which single-mode amplitude-statilized laser radiation is scattered from an optically rough target [13–15]. In this case the signal is well represented by the complex gaussian process considered above [7,14,15]. The signal bandwidth Δ_S can be identified with the Doppler spread arising from the rotation of the target, whereas the difference frequency $|\omega_\mathrm{S} - \omega_{\mathrm{LO}}|$ is associated with the translational motion of the target [15]. The pulse duration is assumed to be approximately equal to the integration time T and $T > 1/\Delta_{\mathrm{IF}}$. If $N_\mathrm{S} = \langle i_\mathrm{S}\rangle T/e$ is the mean number of signal photocarriers released in the time T, and $T_1 \sim 1/\Delta_\mathrm{S}$ is the coherence time of V_1 then, from eq. (13),

$$\mathrm{SNR_H} \approx \frac{N_\mathrm{S}}{N_\mathrm{S} + (T/QT_1)}(\Delta_{\mathrm{IF}}T)^{1/2} = \frac{\delta_1}{\delta_1 + 1}(\Delta_{\mathrm{IF}}T)^{1/2}. \qquad (14)$$

In particular, for N_S, T_1, Q, and Δ_{IF} fixed, there exists an optimal pulse duration, $T = T_{\mathrm{opt}}$, for which $\mathrm{SNR_H}$ is maximized. Setting $\partial\mathrm{SNR_H}/\partial T = 0$ and solving for T_{opt}, we obtain

$$T_{\mathrm{opt}} = N_\mathrm{S}T_1Q, \qquad (15)$$

so that

$$\mathrm{SNR_H}(T_{\mathrm{opt}}) = \tfrac{1}{2}\left(N_\mathrm{S}\frac{\Delta_{\mathrm{IF}}}{\Delta_\mathrm{S}}Q\right)^{1/2}. \qquad (16)$$

The significance of T_{opt} is apparent from an examination of eq. (13). The first term in the denominator arises from the LO shot noise whereas the second term is excess noise arising from the random fluctuations of the signal field V_1. Under the constraints indicated above, an increase in T decreases the signal-to-shot-noise ratio $\langle\mathrm{SNR}\rangle$, as well as the excess noise. Thus, the undesired decrease of $\langle\mathrm{SNR}\rangle$, caused by the linearly decreasing signal power, is accompanied by the desired decrease of the excess noise resulting from noncoherent averaging of the mutually independent samples (realizations) of the signal intensity. The number of realizations increases linearly with T at the rate T/T_1. The total number of realizations can be as large as $M = AT/A_1T_1$, viz., the number of degrees of freedom in the received signal radiation. If T is optimum, then the corresponding number of degrees of freedom is equal to the mean number of signal counts N_S, or equivalently, the signal light degeneracy parameter δ_1 is unity.

We therefore conclude that if on the average there is less than one signal photocarrier per coherence volume ($\delta_1 < 1$ and consequently $\langle\mathrm{SNR}\rangle < 1$), the system performance is limited by the shot noise. Indeed, this can be understood as follows: when $\delta < 1$ Mandel [10] has shown that the counting distribution governing the signal photocarriers approaches Poisson so that there is no excess noise. In this case, the conventional SNR also provides a suitable measure of system performance [8]. When the degeneracy parameter is larger than one ($\delta_1 > 1$ and consequently $\langle\mathrm{SNR}\rangle > 1$), on the other hand, the performance is limited by the excess noise. In fig. 2 we present several plots of $\mathrm{SNR_H}/\mathrm{SNR_H}(T_{\mathrm{opt}})$ as a function of T/T_1 for different values of N_S assuming $Q = 1$ and $\Delta_{\mathrm{IF}} = \Delta_\mathrm{S}$. The curves demonstrate that the loss in system performance incurred under suboptimal conditions may be substantial.

Finally, it is instructive to compare the performance of heterodyne and direct detection. In the latter case, the estimator of the signal intensity is obtained at the output of a low-pass integrating filter that immediately follows the photodetector. The total number of photocarriers, n_T, collected during a pulse of duration T, is a sum of the signal- and background-generated counts with mean values N_S and N_B. It has been shown [7] that the ratio of the signal intensity estimator, $N_\mathrm{S} = \langle n_\mathrm{T}\rangle - N_\mathrm{B}$, to the standard deviation of the observable n_T is

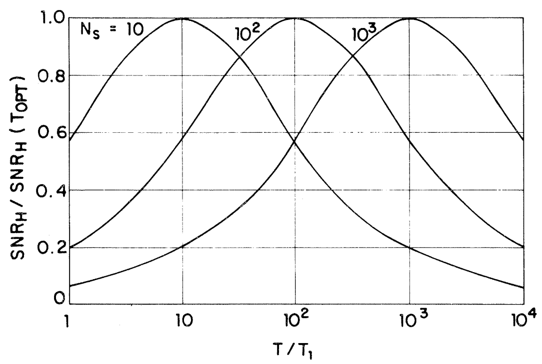

Fig. 2. $SNR_H/SNR_H(T_{opt})$ versus T/T_1.

$$SNR_D = N_S/(N_S + N_B + N_S^2/M')^{1/2}. \qquad (17)$$

The quantities SNR_D and SNR_H are analogous. Eqs. (14) and (17) are valid under the same constraints and M' is the number of degrees of freedom in the signal radiation, which is given by $AT/A_1'T_1'$ for a sufficiently large detector aperture A. For the direct detection case, the coherence area and coherence time of the signal, denoted by A_1' and T_1', respectively, are appropriately defined as $A_1' = \int_{-\infty}^{\infty} |\gamma_1(\boldsymbol{\rho}, 0)|^2 \mathrm{d}^2\boldsymbol{\rho}$ and $T_1' = \int_{-\infty}^{\infty} |g_1(\tau)|^2 \mathrm{d}\tau$. The first two terms in the denominator of eq. (17) represent the signal and background shot noise, respectively, while the third term is the excess noise.

A comparison of eqs. (14) and (17) leads one to conclude the following: (1) In the large signal region, where the performance of both techniques is limited by excess noise, only in direct detection can this noise virtually be reduced to zero by spatial averaging (in addition to temporal averaging) so that the received signal can be made to exhibit a very small degeneracy parameter δ_1' ($\delta_1' = N_S/M' \ll 1$), leading to $SNR_D = N_S^{1/2}$; (2) In the small-signal region, where background dominates ($N_S < N_B$), heterodyne detection is superior. This is evident from a comparison of the optimal performance: $SNR_D \to N_S/N_B^{1/2}$ whereas $SNR_H(T_{opt}) \to \frac{1}{2} N_S^{1/2}$, assuming that $Q = 1$ and $\Delta_{IF} = \Delta_S$ in eq. (16).

Finally we note that although our treatment has dealt specifically with optimization in the temporal domain, similar arguments can be made for the spatial domain and for signal statistics other than gaussian [16]. A number of steps along this path have already been taken [17–20].

References

[1] A. Javan, W.R. Bennet and D.R. Herriott, Phys. Rev. Lett. 6 (1961) 106.
[2] M.C. Teich, R.J. Keyes and R.H. Kingston, Appl. Phys. Lett. 9 (1966) 357.
[3] M.C. Teich, Proc. IEEE 56 (1968) 37.
[4] B.J. McMurtry and A.E. Siegman, Appl. Opt. 1 (1962) 51.
[5] H.Z. Cummins and H.L. Swinney, Light beating spectroscopy, in: Progress in Optics VIII, ed. E. Wolf (North-Holland, Amsterdam, 1967) pp. 134–200.
[6] J. Bures, J. Opt. Soc. Am. 64 (1974) 1598.
[7] M. Elbaum and P. Diament, Appl. Opt. 15 (1976) 2268.
[8] E. Jakeman, C.J. Oliver and E.R. Pike, Adv. Phys. 24 (1975) 349.
[9] H.A. Haus, C.H. Townes and B.M. Oliver, Proc. IRE 50 (1962) 1544.
[10] L. Mandel, Proc. Phys. Soc. 74 (1959) 233.
[11] L. Mandel, J. Opt. Soc. Am. 56 (1966) 1200.
[12] J.S. Bendat and A.G. Piersol, Random data: analysis and measurement procedures (Wiley, New York, 1972) p. 184.
[13] M. Elbaum, P. Diament, M. King and W. Edelson, Appl. Opt. 16 (1977) 1982.
[14] J.W. Goodman, Proc. IEEE 53 (1965) 1688.
[15] M.C. Teich, Proc. IEEE 57 (1969) 786.
[16] J.K. Jao and M. Elbaum, Proc. IEEE 66 (1978).
[17] A.E. Siegman, Proc. IEEE 54 (1966) 1350.
[18] D.L. Fried, Proc. IEEE 55 (1967) 57.
[19] D.L. Fried, IEEE J. Quant. Electron, QE-3 (1967) 213.
[20] L. Mandel and E. Wolf, J. Opt. Soc. Am. 65 (1975) 413.